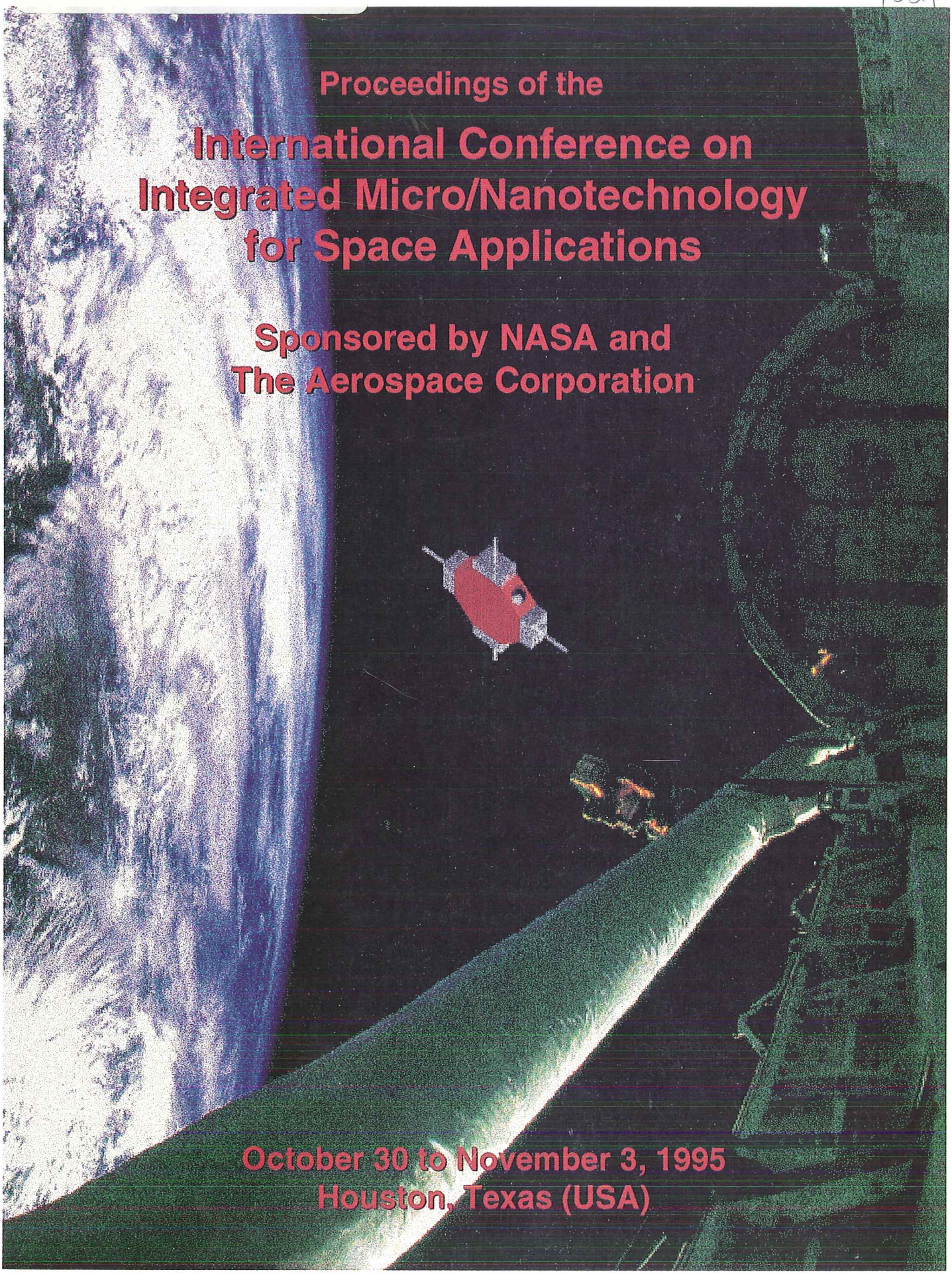
NASA-CR-201963

226673
702P.



Proceedings of the

International Conference on Integrated Micro/Nanotechnology for Space Applications

Sponsored by NASA and
The Aerospace Corporation

October 30 to November 3, 1995
Houston, Texas (USA)

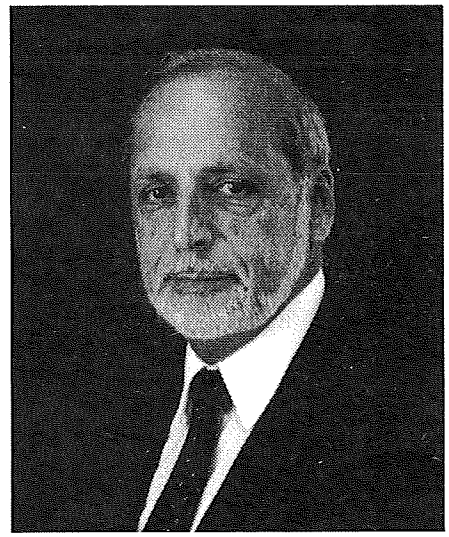
International Conference on Integrated Micro-Nanotechnology for Space Applications

The First International Conference on Integrated Micro/Nanotechnology for Space Applications was hosted jointly by the Aerospace Corporation and Johnson Space Center in Houston from 30 October through 3 November. The purpose of the conference was to bring together scientists and engineers from the fields of microtechnology, nanoelectronics and space technology to explore the possibilities for applying newly emerging capabilities in microtechnology to space operations.

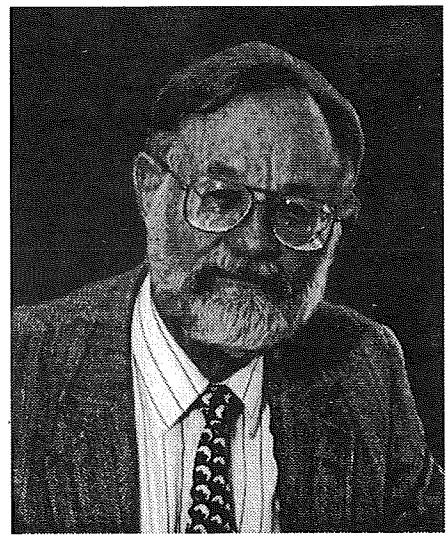
The evolution of microelectronic technology coupled with the growth of microelectro-mechanical systems (MEMS) in the past 4-5 years has had a significant impact in the commercial terrestrial sector. This influence can be evidenced particularly in sensor, optical switching and mass data storage applications that have been inserted into major industries such as transportation, medicine, telecommunications and computers. The focus of this conference was to anticipate and extend the incorporation of Nano-electronics and MEMS into Application Specific Integrated Microinstruments (ASIMs) in order to revolutionize the development of space systems.

The conference attracted over 175 participants from ten different countries including Belgium, China, France, Germany, Japan, The Netherlands, Sweden, Switzerland, and the United Kingdom. NASA and the European Space Agency were represented in addition to the U. S. Air Force Laboratories. The delegates heard presentations from 65 authors on topics ranging from mission applications of nanosatellites to silicon micromachining for photonic applications.

Following the plenary sessions, the last two days of the conference featured workshops in nine different areas of microtechnology. National and international experts worked to assemble point designs for space applications of hardware, software and systems in future manned and unmanned space missions. The Conference venue also served as the initial meeting place of a National Steering Committee constituted from government program directors with responsibilities in microtechnology. Their meetings were used to coordinate programs and optimize the use of government resources to further space applications. These Proceedings include both the presented papers and the workshop reports.



Dr. Seymour Feuerstein
Co-host
The Aerospace Corporation



Dr. Kenneth Cox
Co-host
Johnson Space Center

Seymour Feuerstein, Principal Director
Mechanics and Materials Technology Center
The Aerospace Corporation

Kenneth Cox, Assistant to the Director
Engineering Directorate
Johnson Space Center

Conference Officials

Conference Directors

Kenneth Cox
NASA/JSC
Houston, Texas

Seymour Feuerstein
The Aerospace Corporation
El Segundo, California

Technical Program Cochairs

Robert H. Stroud
The Aerospace Corporation
Herndon, Virginia

Mark Holderman
NASA/JSC
Houston, Texas

Technical Director

Steve Watson
Director, Space Systems Group
Scientific Research and Development Office
Washington, DC

Technical Advisors

Sumner Matsunaga
The Aerospace Corporation
Herndon, Virginia

Steve Edwards
USAF
Washington, DC

Technical Program Committee

David Sutton (Chairman)
The Aerospace Corporation
El Segundo, California

Ernie Robinson
The Aerospace Corporation
El Segundo, California

Henry Helvajian
The Aerospace Corporation
El Segundo, California

Siegfried Janson
The Aerospace Corporation
El Segundo, California

Local Arrangements

Elaine Hirahara
The Aerospace Corporation
El Segundo, California

George Nield
NASA/JSC and AIAA (Houston Chapter
Houston, Texas

Local Sponsor

American Institute of Aeronautics and Astronautics, Houston Chapter

Proceedings of the

International Conference on Integrated Micro/Nanotechnology for Space Applications

Sponsored by:

and

THE AEROSPACE CORPORATION

October 30 - November 3, 1995

Houston, Texas (USA)

Table of Contents

Conference Officials

Opening Session (Monday, AM)
Session Chair: Robert Stroud, The Aerospace Corporation

Strategic Thinking for Space
by Kenneth Cox, NASA/JSC

The Impact of Microengineering Technology on Space System Development
by David G. Sutton, The Aerospace Corporation

Nanoelectronics: Opportunities for Future Space Applications
by Gary Frazier, Texas Instruments

Microelectromechanical Systems
by Kaigham Gabriel, ARPA (Presented by David Nagel, Naval Research Labs)

Session 1: Space Systems & Requirements (Monday, AM)
Session Chair: Charles Price (NASA/JSC) and Siegfried Janson (Aerospace)

Considerations for Micro- and Nano-Scale Space Payloads
by David A. Altemir, NASA/JSC

Sprint: The First Flight Demonstration of the External Work System Robots
by Charles Price and Keith Grimm, NASA/ JSC

Silicon Satellites: Picosats, Nanosats, and Microsats
by Siegfried Janson, The Aerospace Corporation

Implications of Gun Launch to Space for Nanosatellite Architectures
by Miles R. Palmer, Science Applications International Corporation

Vehicle Tracking System Using Nanotechnology Satellites and Tags
by Dino Lorenzini, SpaceQuest, Ltd.,and Chris Tubis, National Semiconductor Corp.

Teledesic Global Wireless Broadband Network: Space Infrastructure Architecture, Design Features and Technologies
by James Stuart, Teledesic Corporation

Session 2: Data Processing and Communications (Monday, PM)
Session Chairs: George Haddad (UM) and Michael Gorlick (Aerospace)

Separation and Detection of Toxic Gases with a Silicon Micromachined Gas Chromatography System
by Edward S. Kolesar, Jr., Texas Christian Univ., and Rocky R. Reston, USUHS

Table of Contents (continued)

Session 2: Data Processing and Communications (continued)

A 32-bit Ultrafast Parallel Correlator Using Resonant Tunneling Devices – 11
by Shriram Kulkarni, Pinaki Mazumder and George I. Haddad, U. of Michigan

Highly-parallel, Highly-Compact Computing Structures Implemented in Nanotechnology – 12
by D. G. Crawley, M. J.B. Duff, T. J. Fountain, C. D. Moffat, and C. D. Tomlinson, University College, London, UK

Technologies and Designs for Electronic Nanocomputers – 13
by Michael Montemerlo, J. Christopher Love, Gregory Opiteck, David Goldhaber, and James C. Ellenbogen, MITRE Corporation

Resonant Tunneling Analog-to-Digital Converter – 14
by T.P.E. Broekaert, A. C. Seabaugh, J. Hellums, A. Taddiken, H. Tang, J. Teng, and J.P.A. van der Wagt, Texas Instruments

MEMS-based Communication Systems for Space-Based Applications – 15
by Héctor J. De Los Santos and Robert A. Brunner, Hughes Space and Communications, and Juan F. Lam, Le Roy H. Hackett, Ross F. Lohr, Jr., Lawrence E. Larson, Robert Y. Loo, Mehran Matloubian, and Gregory L. Tangonan, Hughes Research Laboratoriess,

Silicon Micromachining in RF and Photonic Applications – 16
by Tsen-Hwang Lin, Phil Congdon, Gregory Magel, Lily Pang, Chuck Goldsmith, John Randall, and Nguyen Ho, Texas Instruments

Software Component Technologies and Space Applications – 17
by Don Batory, University of Texas at Austin

Averting Denver Airports on a Chip
by Kevin J. Sullivan, University of Virginia

Session 3: Posters and Demonstrations (Monday, PM)

Session Chairs: Mark Holderman (NASA/JSC) and Robert Smith (Aerospace)

AIAA Spacecraft GN&C Interface Standards Initiative: Overview – 19
by A. Dorian Challoner, Hughes

A High Average Power Free Electron Laser for Microfabrication and Surface Processing Applications – 20
by H. F. Dylla, S. Benson, J. Bisognano, C. L. Bohn, L. Cardman, D. Engwall, J. Fugitt, K. Jordan, D. Kegne, Z. Li, H. Liu, L. Merminga, G. R. Neil, D. Neuffer, M. Shinn, C. Sinclair, and M. Wiseman, CEBAF, L. J. Brillson, Xerox, D. Henkel, Henkel Metall. Tech., H. Helvajian, The Aerospace Corporation, M. J. Kelley, DuPont, and Shanti Nair, University of Massachusetts

MEMS in Space Systems
by J.C. Lyke, M.A. Michalicek, and Babu Singaraju, Air Force Phillips Laboratory – 21

Table of Contents (continued)

Session 3: Posters and Demonstrations (continued)

Subminiaturization for Environmental Research Aircraft and Sensor Technology (ERAST) Instrumentation
by Marc Madou, U. C. Berkeley, M. Loewenstein and S. Wegener, NASA/Ames

DoD Space Test Program (STP)
by Major Llwyn Smith, U.S. Air Force Space and Missile Systems Center

Weaves as an Interconnection Fabric for ASIMs and Nanosatellites
by Michael M. Gorlick, The Aerospace Corporation

Performance Thresholds for Application of MEMS Inertial Sensors in Space
by Geoffrey N. Smit, The Aerospace Corporation

Advanced Modeling of Micromirror Devices
by M. Adrian Michalicek, Darren E. Sene, and Victor M. Bright,
Air Force Institute of Technology

Monitoring Composite Material Pressure Vessels with a Fiber-Optic/Microelectronic Sensor System
by Charles Klimcak and B. Jaduszliwer, The Aerospace Corporation

Micromechanical switches on GaAs for Microwave Applications
by John Randall, C. Goldsmith, D. Denniston, and T-H Lin, Texas Instruments

High Density Memory Based on Quantum Device Technology
by Paul van der Wagt, Gary Frazier, and Hao Tang, Texas Instruments

InGaAlAsPN: A Materials System for Silicon Based Optoelectronics and Heterostructure Device Technologies
by T. P. E. Broekaert, S. Tang, R.M. Wallace, E. A. Beam III, W. M. Duncan, Y.-C. Kao, and H.-Y. Liu, Texas Instruments

The Fundamentals of Using the Digital Micromirror Device (DMD™) for Projection Display
by Lars A. Yoder, Texas Instruments

Aercam Demonstration
by Charles Price, NASA/JSC

MEMS Spaceborne Testbed
by Robert H. Stroud, The Aerospace Corporation, and
Mark Holderman, NASA/JSC

Low Volume Packaging for a Microinstrumentation System
by Andrew Masonand Ken Wise, University of Michigan

Table of Contents (continued)

Session 3: Posters and Demonstrations (continued)

Pyroelectric Applications of the VDF-TrFE Copolymer
by J.J. Simonne, Laboratoire d'Analyse et d'Architecture des Systèmes, Ph. Bauer, SOFRADIR, L. Audaire, CEA/DTA/LETI/DOPTS/S-LIR, and F. Bauer, Institut de Recherche Franco-Allemand de St Louis

Stereo-Based Region-Growing using String Matching
by Robert Mandelbaum and Max Mintz, University of Pennsylvania

Session 4: Microinstruments (Tuesday, AM)

Session Chairs: Bart van der Schoot (U. Neuchatel, Swiss) and William Trimmer (Belle Mead Research)

Little Spaces for Outer Space (Invited)
by William Trimmer, Belle Mead Research, Inc.

Micro-Electromechanical Instrument and Systems Development at the Charles Stark Draper Laboratory
by J. H. Connelly, J. P. Gilmore, and M. S. Weinberg, C. S. Draper Laboratory, Inc.

PZT Thin Film Piezoelectric Travelling Wave Motor
by Shen Dexin, Zhang Baoan, Yang Genqing, Jiao Jiwei, Lu Jianguo, and Wang Weiyuan, State Key Laboratories of Transducer Technology, Shanghai Institute of Metallurgy

Demonstration of a Monolithic Micro-Spectrometer System
by S. Rajic and C. M. Egert, Oak Ridge National Laboratory

Silicon Micromachined Sensor for Broadband Vibration Analysis
by Adolfo Gutierrez, Daniel Edmans, Chris Cormeau, and Gernot Seidler, InterScience Inc.

A Microinstrumentation System for Remote Environmental Monitoring
by Andrew Mason, Wayne G. Baer, and Kensall D. Wise, University of Michigan

Spacecraft Applications of Compact Optical and Mass Spectrometers
by N. M. Davinic and D. J. Nagel, Naval Research Laboratory

High-performance Electronics at Ultra-Low Power Consumption for Space Applications: From Superconductor to Nanoscale Semiconductor Technology
by Robert V. Duncan, Sandia Corp.

Microrefrigeration by a Pair of Normal Metal/Insulator/Superconductor Junctions
by M. M. Leivo and J. P. Pekola, University of Jyväskylä, and D. V. Averin, State University of New York (SUNY) at Stony Brook

Table of Contents (continued)

Session 5: Materials Processing, Packaging and Architectures (Tuesday, PM)
Session Chairs: James C. Lyke (Air Force Phillips Laboratory) and Henry Helvajian (The Aerospace Corporation)

Laser Material Processing for Microengineering Applications
by Henry Helvajian, The Aerospace Corporation

Recent Developments in Microsystems Fabricated by the LIGA-Technique
by J. Schulz, K. Bade, A. El-Kholi, H. Hein and J. Mohr,
Institute für Mikrostrukturtechnik, Karlsruhe, Germany

Micromechanical Machining Processes and Their Application to Aerospace Structures, Devices, and Systems
by Craig Friedrich and Robert O. Warrington, Louisiana Technical University

Advancing MEMS Technology Usage through the MUMPs Program
by David A. Koester, K. W. Markus, V. Dhuler, R. Mahadevan, and A. Cowen,
MCNC MEMS Technology Applications Center

Method of Making Large Area Microstructures
by Alvin M. Marks, Advanced Research Development, Inc.

Surface Micro-Machining: Progress Towards Micro-Electro Mechanical Systems
by James Doscher, Analog Devices Inc.

Advanced Packaging for Integrated Micro-Instruments
by James C. Lyke, U.S. Air Force Phillips Laboratory

The Impact of Space Radiation Requirements and Effects on ASIMs
by C. Barnes, A. Johnston and G. Swift, Jet Propulsion Laboratory (NASA)

3D Thin Film Microstructures for Space Microrobots
by Isao Shimoyama, University of Tokyo

Session 6: Space Applications I (Wednesday, AM)
Session Chair: Charles Sawin (NASA/JSC)

Development of a Space Bioreactor Using Microtechnology
by Philippe Arquint, Marc A. Boillat, Nico F. de Rooij, Sylvain Jeanneret and Bart H. van der Schoot, University of Neuchâtel,
Birgitt Bechler, Augusto Cogoli, and Isabelle Walther, Space Biology Group ETHZ,
Volker Gass and Marie-Thérèse Ivorra, Mecanex SA

Integrated Microreactor for Chemical and Biochemical Applications
by Norbert Schwesinger, L. Dressler, Th. Frank, and H. Wurmus,
Technical University of Ilmenau, Germany

Table of Contents (continued)

Session 6: Space Applications I (continued)

Fluid Flow Volume Measurements Using a Capacitance/Conductance Probe System
by T. Nguyen and G. Arndt, NASA/Johnson Spacecraft Center,
and J. Carl, Lockheed

Biomorphic Architectures for Autonomous "Nanosat" Designs
by Brosl Hasslacher and M. Tilden, Los Alamos National Laboratory

Session 7: Space Applications II (Wednesday, AM/PM)

Session Chair: Ernest Y. Robinson (Aerospace)

MEMS Technology for Space Applications
by A. van den Berg, V. L. Spiering, T. S. J. Lammerink, M. Elwenspoek, and P. Bergveld, MESA Research Institute, Univ. of Twente, The Netherlands

Thermal Switch for Satellite Temperature Control
by H. Ziad, P. Van Gerwen, and K. Baert, IMEC, Belgium,
T. Slater, Wildcat Micromachining,
E. Masure and F. Preud'homme, Verhaert Design & Development, Belgium

Optically-Switched Resonant Tunneling Diodes for Space-Based Optical Communication Applications
by T. S. Moise, Y.-C. Kao, D. Jobanovic, and P. Sotirelis, Texas Instruments Inc.

A **Micromechanical INS/GPS System for Small Satellites**
by N. Barbour, T. Brand, R. Haley, M. Socha, J. Stoll, P. Ward, and M. Weinberg, Charles Stark Draper Laboratory

Micro Scanning Laser Range Sensor for Planetary Exploration
by Ichiro Nakatani, Hirobumi Saito, Takashi Kubota, Takahide Mizuno, Hiroshi Katoh, Satoru Nakamura, Kenji Kasamura, Institute of Space and Astronautical Science (ISAS), and Hiroshi Goto, OMRON Corp, Japan

NanoSatellite Power Systems Considerations
by Michael Robyn, Larry Thaller, and D. Scott, The Aerospace Corporation

Low-Mass Inflation Systems for Inflatable Structures
by Daniel Thunnissen, Mark S. Webster, and Carl S. Engelbrecht,
Jet Propulsion Laboratory (NASA)

Miniature Telerobots in Space Applications
by S. C. Venema and B Hannaford, University of Washington

Table of Contents (continued)

Forward to Workshop Reports

Workshop Reports

Appendix I, Attendance List

Opening Session (Monday, AM)

Session Chair: Robert Stroud, The Aerospace Corporation

Strategic Thinking for Space
by Kenneth Cox, NASA/JSC

The Impact of Microengineering Technology on Space System Development
by David G. Sutton, The Aerospace Corporation

Nanoelectronics: Opportunities for Future Space Applications
by Gary Frazier, Texas Instruments

Microelectromechanical Systems
by Kaigham Gabriel, ARPA (Presented by David Nagel, Naval Research Labs)

The Impact of Microtechnology on Space System Development*

David G. Sutton

The Aerospace Corp.
P. O. Box 92957
Los Angeles, CA 90009
Phone (310) 336-5049
e-mail: dave_sutton@qmail2.aero.org

Microtechnology has the potential for a great beneficial impact on both the launch and operation of space systems. The reasons for this include savings in the mass, power consumption, volume, and cost of manufacture and testing of space systems. Less apparent, but equally valuable, are the advantages in reliability to be gained by increased redundancy and the reduction of complexity that are inherent in the fabrication processes. Despite the leveraged gains to be had by "microengineering" space systems, the conservatism of the aerospace community will retard the rapid incorporation of this technology into both new and existing systems. This is more true of government space programs where success is measured by lack of launch failures and less true of commercial ventures where success may be measured by other criteria. A successful program for the development and insertion of microtechnology into government systems will need to consider these factors. U. S. Air Force launches have the highest success rate in the world. One can hardly expect an organization to abandon a successful strategy, especially when the risks of failure include increased costs and a loss of capability that is vital to national security. However, there is a strategy for evolving microtechnology in space systems that fits both the risk avoidance culture and parallels the expected development of microtechnology. It starts with the development of autonomous, unobtrusive systems for launch environment measurements and for the determination of health and welfare of both the launch vehicle and payload. As microtechnology progresses and experience is gained, drop-in subsystems can be employed to initially increase redundancy and eventually replace current subsystems. These flightworthy systems can be combined to produce parasitic spacecraft hosted on larger satellites for specialized missions such as the Untethered Flying Observer that is the subject to be considered by one of the Conference Workshops. Finally, truly autonomous microsatellites can be developed as the systems mature and advantageous missions are defined.

* This Study was conducted in support of the Technology Development and Applications Directorate of the Aerospace Corporation.

Microtechnology has the potential for a great beneficial impact on both the launch and operation of space systems. The reasons for this are mostly apparent. They include savings (due to miniaturization of components and subsystems) in the mass, power consumption, volume and cost of manufacture and testing of satellites. Less apparent but equally valuable are the advantages in reliability to be gained by increased redundancy and the reduction of complexity that are inherent in the fabrication processes used to produce micro-electromechanical systems. Despite the inherent, leveraged gains to be had by "microengineering" space systems the conservatism of the aerospace community will retard the rapid incorporation of this technology into both new and existing systems. This is more true of government space programs where success is measured by lack of launch failures and less true of commercial ventures where success may be measured by other criteria. A successful program for the development and insertion of microtechnology into government systems will need to consider these factors. This paper suggests a strategy for evolving microtechnology in space systems that fits both the risk avoidance culture and parallels the expected development of microtechnology. Reduced launch costs alone offer substantial initiative for the replacement of traditional space systems with their microengineered equivalents. A Titan IV (SRMU) can launch a payload of 40,000 pounds to low earth orbit at a cost of $200 million.[1] Without including the costs of payload development, manufacture, testing and integration, this cost is roughly $5,000 per pound. Further savings can be achieved with microtechnology by reducing on-orbit power consumption. The marginal cost of adding power to a satellite solar array is approximately $4 million per square meter[2] or about $20,000 per watt. In addition to these demonstrable savings, there are tangible but less quantifiable savings to be had from the increased reliability of microsystems. This reliability results directly from the increased redundancy, built-in test capability and simplified fabrication technology that is microtechnology's legacy from solid state electronics. Finally there are additional savings to be had from the smaller test facilities that will be used to qualify these microsystems.

In addition to savings resulting from modification to existing spacecraft and missions, microtechnology can and will be an enabling technology for whole new ways of doing things. These include both new ways of doing old missions and completely new missions. For example, Janson has outlined a proposal for launching a complete earth observing constellation of nanosatellites with a single Pegasus.[3] Other examples include seeding the moon or mars with seismic microsensors and utility meter reading from space. See the paper by D. Lorenzini and D. Tubis in these Proceedings.

All organizations conducting business in space stand to benefit from the savings and enhanced capacity to be found in applications of microtechnology. The U. S. Air Force has one of the oldest, largest and most successful operations in space. Its launch vehicles and satellites have the highest success rate in the world as shown in the following tables.[4]

Table 1: USA launch vehicle success/failure record (1984-1994)

Record	DOD Programs	Non-DOD
Success/Failure	101/5	82/8
Success Rate	95.3%	91.1%

Table 2: USA satellite success/failure record (1984-1994)

Record	DOD Programs	Non-DOD
Success/Failure	100/1	69/13
Success Rate	99.0%	84.1%

This record was achieved as the result of a focused program to insure a reliable, uninterrupted, space defense capability and to protect large investments in launch vehicles, payload development and acquisition. If a single launch (booster and payload) costs $1 billion and there are 5-10 launches per year, then the demonstrated >10% advantage in combined launch and satellite reliability over the non-DoD record is worth >$1.0 billion in savings per year. Of course, this is not really savings, but a return on the money and effort invested in building to high standards of reliability, exhaustive testing, and flight qualification of hardware.

The salient point is that one can hardly expect an organization to change a successful strategy, especially when the risks of failure include increased costs and a loss of capability that is vital to national security. Recent experience with small launch vehicle failures gives emphasis to this point.[5]

We can expect that this risk averse strategy will be continued in the future. Several features of this strategy limit development opportunities for new systems, including microsystems. Among them the following:

1) Flight qualification of all new systems

2) Large test and flight costs added on top of any development costs

3) Class 1 changes in vehicle configuration that cost >$1 million

4) Space test opportunities that are limited (See the papers on the STP program and the MEMS Testbed that appear later in these Proceedings)

5) Technology is frozen at beginning of long acquisition cycles

6) Infrequent block changes in existing systems

The Space Test Program has been highly innovative and successful (see the paper by Maj. L. Smith in these Proceedings), but would have to be greatly expanded to provide the increased opportunities for flight qualification necessary to sustain a rapidly evolving program in space applications of microtechnology. Other strategies for initiating new programs either within DoD or with NASA run counter to current downsizing efforts.

There is an alternate strategy for evolving microtechnology in space systems that fits both the risk avoidance culture and parallels the expected development of microtechnology. It starts with the development of autonomous, unobtrusive systems for launch environment measurements and for the determination of health and welfare of both the launch vehicle and payload. Microengineered systems can be used in this way to reduce risk directly and to gather information for design improvements of existing systems. As microtechnology progresses and experience is gained, drop-in subsystems can be employed to initially increase redundancy and eventually replace current subsystems. These flight worthy systems can be combined to produce parasitic spacecraft hosted on larger satellites for specialized missions, such as the Untethered Flying Observer that is the subject of a report by one of the Conference Workshops. Finally, truly autonomous microsatellites can be developed as the platform and its systems mature and advantageous missions are defined.

Let us examine an example suitable for the first leg of this strategy. Titan launch vehicles currently employ the Wideband Instrumentation System (WIS) for inflight monitoring of acoustics and vibration. This system is limited by the availability of telemetry channels to providing data from 23 locations. To move the location of one sensor invokes a class one change with a cost approximating $0.5 million. Checkout and calibration of the WIS are known causes of launch delays and their incumbent costs.

The environments inferred from these limited WIS measurements establish design requirements for avionics modules and other launch vehicle components. The uncertainties that result from limited data require conservative designs with higher costs and higher weights. Despite this conservatism data from nearly every flight prompt redesign and requalification of hardware to meet measured environments that exceed calculated or inferred design environments.

The WIS function can be greatly enhanced with virtually no impact on the vehicle or its operation by inserting autonomous microsensors at critical points where more data are required. The technology exists for making these devices truly self contained with their own power and communications capability.[6] Three dimensional vibration and shock measuring instruments with very high dynamic ranges and self check capability can be assembled in wrist watch sized packages. They can be simply mounted on or next to critical assemblies with virtually no impact on the environment to be measured or the vehicle's power and telemetry systems. An independent monitoring system on the ground would suffice to collect the generated data. Insertion of these devices in parallel with the existing system would be very attractive to those requiring additional data for model development, would result in cost savings by reducing the design margins currently required for instrument packages and would not be subject to the constraints placed on conventional configuration changes. Due to their enhanced measurement capability and flexible deployment features these devices would rapidly prove to be indispensable for launch vehicle design and payload environment definition.

Just such an instance illustrating how operations can come to depend on systems meant only to provide awareness exists in the literature. The PAX, 3 axis accelerometer package for vibration measurements, was mounted on-board the Olympus telecommunications spacecraft in order to establish baseline vibration data associated with various functions. The data from this instrument were being gathered primarily for use in the design of a laser communications system.[7] However, PAX was also intended to monitor the evolution of mechanical systems over the life of the spacecraft. It consisted of a 2.3 kilogram package. The sensors were manufactured by the Centre Suisse pour Electronique et Microtechnique in Neuchatel, Switzerland using silicon microfabrication technology.

In 1991 satellite power and attitude control were lost for over two months resulting in on-board temperatures down to -70 °C. Once control was reestablished, comparison of vibration data from the PAX with baseline data accumulated before the failure was used to identify and assess faulty systems.[8] In this manner a scanning infrared earth sensor was found to be the cause of a severe knocking and was turned off before it could cause further damage. Similarly,

a reaction wheel bearing was found to be the cause of a recurring "screech." When this noise event was eventually correlated with ambient temperature fluctuations local heaters were used to eliminate it, thereby extending the lifetime of this system. In this manner a monitoring system proved to be essential to the recovery and life extension of an orbiting satellite following a severe anomaly. It is expected that similar events will prove microengineered monitoring systems to be invaluable to launch vehicle and spacecraft operations and will eventually make them required for all spacecraft.

A second example of a current application of microtechnology comes from an entirely different sphere. GaAs used in high speed spacecraft electronics evolve hydrogen gas. When sealed in hermetic packages the gradual build up of gas is sufficient to poison the circuits. It is therefore necessary to test stored units for hydrogen accumulation before they are built into payloads. For this purpose an integral chemical microsensor was constructed that provides accurate, *in situ*, nondestructive monitoring of H_2 in the ambient atmosphere of sealed electronics packages.[9] Packages provided with this self-test capability are inherently more reliable, since faulty units can be eliminated before launch and on-orbit degradation can be diagnosed and isolated.

The second leg of this strategy is based on the experience and capability acquired in developing and employing diagnostics and extends to drop-in subsystems. These systems will be initially employed to increase redundancy, but as experience and confidence grows will eventually replace current subsystems. The incentive to incorporate microsystems in existing spacecraft will be the reduction in weight and power, but the vast enhancement in redundancy and its associated reliability will be an equally valuable gain.

Some of these microengineered subsystems will become available through commercial developments. For example, accelerometers, chemical microsensors, GPS based guidance systems, and microoptics are being rapidly developed for applications in the automobile, chemical, shipping and communications industries. However, those applications that are specific to space will require investment and development sponsored by the end user, if they are to keep pace with the concurrent activity stimulated by the commercial markets. Examples of these later subsystems include propulsion, star and earth sensors and radiation hard microelectronics. For examples of current developments in both commercial and space-specific arenas see the papers in these Proceedings by J. Gilmore, A. Mason, D. Nagel, I. Nakatani, G. Smit, L. Thaller, A. van den Berg, K. Wise and others.

The convergence of the concurrent efforts in the commercial and government spheres will eventually enable microsatellites to be designed as assemblies of subsystems. The first operational microsatellites are likely to be hosted on larger spacecraft and have functions that are limited to diagnostics and local environmental sensing. Small satellites and robots to serve these functions are already under development at Johnson Space Center for shuttle and space station operations[10]. See the paper by C. Price and K. Grimm in these Proceedings. In addition, the Jet propulsion Laboratory, Diamler-Benz Aerospace[11] and Space Industries,[12] have designs for small satellites that fit this category. All of these efforts are fertile ground for microsatellite development.

The first autonomous microsatellites are already being conceptually designed for applications in monitoring terrestrial shipments (see the paper by D. Lorenzini and D. Tubis) and earth observing missions.[13] In order to be effective these satellites will necessarily be deployed in constellations that require cooperative behavior for orbital phasing, drop-out compensation, and potentially phased detection and cellular communications. These capabilities will require advances in communications, navigation, computation and software that will only partially be achieved by commercial enterprises. Government users will need to make focused investments in microtechnology in order to meet their specific missions. However, the payoff in terms of low cost, secure, robust systems that are deployable on demand and can meet old missions in new ways and enable entirely new missions will be sufficient enticement to continue the odyssey.

---

1 Steven J. Isakowitz, International Reference Guide to Space Launch Systems, AIAA, Washington, DC (1991), p. 268.

2 Robert L. Abramson, "Cost Trends in Satellite Miniaturization," in *Micro- and Nanotechnology for Space Systems: An Initial Evaluation*, H. Helvajian and E. Y. Robinson, ed., Aerospace Technical Report ATR-93(8349)-1 (31 March 1993).

3 S. W. Janson, "Chemical and Electric Micropropulsion Concepts for Nanosatellites," AIAA Paper 94-2998, AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Indianapolis, IN (June 27-29, 1994).

4 I-Shih Chang, , "Investigation of Space-Related Mission Failures," TOR94(3530)-4, The Aerospace Corporation (April 1994).

5 *Space News*, June 26-July2 (1995).

6 E. Y. Robinson, "ASIM Applications in Current and Future Space Systems," in *Microengineering Technology for Space Systems*, H. Helvajian, ed., Aerospace Technical Report ATR-95(8168)-2 (30 September 1995).

[7] N. S. Ferguson, J. N. Pinder, and D. E. L. Tunbridge, "Spacecraft Vibrations Due to Mechanisms; Measurements from Olympus On-Station," Fifth European Space Mechanisms and Tribology Symposium, ESTEC, Noordwijk, The Netherlands, 28-30 October (1992), p. 221.

[8] D. Tunbridge, "The Olympus PAX, Measurement of Mechanism Induced Vibration," Fifth European Space Mechanisms and Tribology Symposium, ESTEC, Noordwijk, The Netherlands, 28-30 October (1992), p. 359.

[9] B. H. Weiller, J. D. Barrie, K. A. Aitchison, and P. D. Chaffee, "Chemical Microsensors for Satellite Applications," *Materials Research Society Symposium Proceedings,* **360**, 535 - 540 (1995) and Aerospace Report No. ATR-95(8061)-1.

[10] *Space News*, April 24-30 (1995).

[11] Space News, April 24-30 (1995).

[12] Space News, August 14-27 (1995).

[13] S. W. Janson, "Spacecraft as an Assembly of ASIMs," in *Microengineering Technology for Space Systems,* H. Helvajian, ed., Aerospace Technical Report ATR-95(8168)-2 (30 September 1995).

STRATEGIC THINKING FOR SPACE

An Unfolding Story

Dr. Kenneth J. Cox

November 1995

OUTLINE

- Strategic Thinking for Space/Update
- Future JSC Outreach Roles
- Outreach Forums
  - Operability Workshop - September 1995 (Local)
  - SATWG Conference - October 1995 (National)
  - Integrated Micro-Nano Technology Space Applications Conference - November 1995 (International)

BUILD BRIDGES

TEAR DOWN WALLS

DANCE OF
PERMANENT WHITE WATER MANAGEMENT

THE CHALLENGE

- Manage and lead within a continuous, turbulent, change upon change environment with high stakes

THE FEELINGS

- Similar to being on a raft in white water with only a pole
  - Raft is virtually unmanageable without a rudder
  - Pole is reactive; must fend against the rocks
  - If successful, must do the same thing tomorrow
  - Do not know what will be ahead

THE MYTH

- Someone in upper management is in calm water and always knows the big picture

THE RESPONSE

- Organizations must adapt to increasing levels of continuous change to remain viable and to properly evolve

THE FOCUS

- Develop discipline, integrity, and risk taking
- Utilize creative intuitive skills
- Encourage leadership development
- Honor diversity
- Value commitments to service
- Utilize open dialogue and active listening skills
- Facilitate partnering
- Create meaning in the work environment

A VIEWPOINT OF SPACE

- Space is Like a Multi-faceted Diamond
- The Major Challenge - How to understand the essence of space and to appropriately organize our country's vision and action
- We must Transition from a Space Program to a Space Movement
  - Space is a place
  - Major Goals should be to put society in space, and to put space into society
- Examples
  - Space Exploration - Vision Driven
  - Space Commerce - Market Driven
  - Space Security - Defense Driven
  - Space Science - Knowledge Driven
  - Space Technology - Applications Driven
  - Space Settlements - Evolution Driven
  - Planetary Sustainability - Human Survival Driven
  - Search for Life Forms - Curiosity Driven

VISION FOR THE FUTURE

- The NASA Administrator has suggested that the Space Program needs a unifying vision
  - Now is the time to crystallize the general public constituency for a long range space movement
  - We need to focus on an "outward looking" process to better understand what our citizens really value and want to support
  - We must engage in new ways to make space activity more relevant to local and regional communities

- Proposed Third Millennium Vision (2000 - 3000)
  - Initially, humans will create a presence throughout the solar system and seek out value for mankind on Earth
  - Then, humans will establish permanent settlements in low Earth orbit and on the Moon, and will explore the planets
  - Later, humans will live, work, and prosper in multiple space communities within the solar system

FUNDAMENTAL PRINCIPLES

- The frontier of space provides the opportunity to achieve higher goals
  - promotes continuous inquiry and enlightened discovery
  - provides for new realms of experience and enhanced perspectives
  - promotes global cooperation of nations
  - potentially binds generations together
  - allows us to understand our beginnings and to consciously create our future
  - provides an opportunity to better humankind

- We must use the power of science, technology, and engineering for the benefit of humankind by involving capabilities of
  - awareness
  - knowledge
  - wisdom
  - consciousness

MAJOR FACTORS OF INFLUENCE

- Transition from the Cold War to the Global Village
- Constrained Fiscal Environment
- Development of Global Concepts
- Public and Private Sector Shift in Responsibilities and Emphasis
- Transition from Limited Information Access to Global Information Access
- Earth/Space Ecological and Sustainability Issues

SIGNIFICANT TRENDS

- Shift emphasis from "Cold Warriors" to "Trade Negotiators"
- Develop entrepreneurial Government/Industry/Academia Interactions
  - Improve partnering and community building skills
  - Encourage enhanced relationships - respect and trust
  - Search for common ground
  - Work for the collective improvement
- International space cooperation arena
  - Exploration
  - Planetary Sustainability
- International space competition arena
  - Launch commercial services
  - In-space commercial services

EVOLUTION OF SPACE

- Public Sector Investment/Government Expenditures
  - Research Science and Technology Development
  - Exploration, Development, and Settlement
  - National Security
  - Planetary Sustainability
- Private Sector Investment/Market Driven Revenues
  - Commercial Goods and Services
  - Industrial Manufacturing
- Volunteer Sector Investment/Resources and Contributions
  - Nonprofit organizations
  - Media and Education
  - Futurist organizations
- Industry should lead, and public/voluntary sectors should contribute in developing commercial markets as national and regional enterprises

PROJECTED SPACE ACTIVITY

- America as a Nation will be a leader and a pioneer in space
- The Government will phase out of direct launch vehicle ownership and operations
- The Government will improve it's capability to act as a catalyst for opening the space commerce frontier
  - An affordable, high access to space capability will be imperative
  - Opportunities for industry-led commercial investment and industrialization will increase exponentially
  - Development of space-based businesses such as manufacturing, product development and tourism will grow significantly
- The "Humans in Space" movement will shift to include the vision of a broad cross-section of common citizens living, working, and thriving in space communities

STRATEGIES FOR THE FUTURE

- Utilize the International Space Station as a Science, Engineering Development, and Cultural Field Center
  - Study Natural Processes in Low Gravity Environments
  - Initiate an International Organization to pursue Scientific Research,Technology Development, and Cultural Integration studies
  - Allow appropriate processes to leverage the ideas, discoveries, and developments into innovative and unique commercial products

- Integrate Space Operational Services
  - Consolidate and streamline operations support in the fields of mission planning, communications, data and tracking networks, and control center facilities
  - Establish Initiatives to develop and apply information/ knowledge and decision/logic technologies to lower the cost of space operations and logistics support
  - Utilize space commercial satellite capabilities as they evolve

STRATEGIES FOR THE FUTURE (CONT)

- Support affordable/reliable access to space initiatives
  - Focus on lower cost and high access
  - Implement Reusable Launch Vehicle technology and demonstration programs
  - Continue to evolve Shuttle cost streamlining initiatives
  - Implement significant improvements in business processes
  - Blend evolutionary and revolutionary approaches

- Organize a System Analysis Staff to provide product assessment, assistance and development services for space
  - Provide integrated system perspectives in the areas of customers, operations, engineering, logistics, development, science and autonomous/automated systems
  - Create transition plans for evolving present capabilities to future systems including boosters, upper stages, platforms, satellites , orbit transportation, rescue, communication, and tracking
  - Act as agents to make space a place of opportunity for entrepreneurial activities and commercial market applications

STRATEGIES FOR THE FUTURE (CONT)

- Act as a catalyst for the development of a low earth orbit community settlement
  - Consider a space town including a business park in the vicinity of the International Space Station and Russian Mir
  - Develop Commercial/Industrial platforms and modules
  - Promote Government and commercial tenant occupancies
  - Develop residential housing and storage facilities
  - Provide utility and consumable services including solar power generation and distribution
  - Develop Basic Goods and Services
    - Medical
    - Life Support and Food
    - Security and Legal
    - Communication
    - Space Media
    - Tourism
    - Sports
    - Entertainment Parks
    - Arts and Crafts

STRATEGIES FOR THE FUTURE (CONT)

- Organize lunar exploration and settlement initiatives
  - Utilize Lunar Resources to Import Clean Energy to Earth
  - Develop an Optical Astrophysical Observatory
  - Investigate mining, industrial processes, and manufacturing
  - Provide solar power generation and distribution
  - Consider propulsion propellant production

- Plan International Planetary Space Initiatives
  - Develop Mars exploration initiatives
    - Robotic Missions
    - Human Missions
  - Initiate asteroid exploration concepts
  - Place emphasis on new physical discoveries and phenomenon

STRATEGIES FOR THE FUTURE (CONT)

- Organize an international consortium to appropriately share space infrastructure investments across nations
  - Establish transportation services including rescue/emergency
  - Understand habitable volume requirements for industry/residence
  - Develop standard life support systems
  - Provide power generation and distribution, and other consumable services
  - Develop biomedical services and products
  - Provide international security

- Develop hybrid robotic missions for space science
  - Explore the Sun, and understand connection with Earth and heliosphere
  - Probe life in the Universe back to its origin
  - Increase fundamental knowledge of the Galaxy/Universe
  - Understand the origin and evolution of planetary systems

STRATEGIES FOR THE FUTURE (CONT)

- Evolve Planetary Sustainability Initiatives
  - Develop Technology for Stabilizing Future Earth Ecological Loads
    - Biotechnology - Modeling Nature and Life Beings
    - Nanotechnology - Dealing with Matter and Physical Sizes
    - Noetictechnology - Enhancing Human Consciousness and Transcendental Ways of Knowing
  - Expand the knowledge of the entire Earth/space system
    - Develop synchronized measurements - space/air/ground
    - Understand changes
    - Improve forecast and assessment abilities
  - Create and validate global climatic models
  - Consider a comet/asteroid collision avoidance initiative

A NEW SPACE ENTERPRISE

- We as a nation must
  - Transition to a planetary/interstellar viewpoint
  - See space as a vast open and promising frontier
  - Integrate vision with reality and commitment

- We must pioneer and develop new skills, new perceptions, and new ways to relate
  - Space Development and Evolution - involves social/technical/economic/political/ecological systems
  - Technology and Science R&D - involves technical, science, and engineering systems
  - Development and Operations - involves Industry and Government systems
  - We must develop cooperative and mature interrelationships between the civil, security, and commercial space domains in order to properly serve the American public

STRATEGIC SPACE FORUMS

- Space forums involving all sectors of American society is needed to focus on Space Development and Evolution
- Potential products include
  - Development of an accepted and acknowledged vision for space
  - Identification of major space thrusts and associated scenerios
    - Space commercialization
    - Space exploration, development, and settlement
    - National and global security
    - Planetary sustainability
  - Development of shared space infrastructure roadmap
  - Enactment of supportive space policy
  - Establishment of a set of international space initiatives
- Networking of multiple National/International Space Forums is needed to honor diverse viewpoints and develop viable options

THE CHALLENGE

- Act as explorers, pioneers, innovators, and entrepreneurs to open the frontier and to expand space as a place
  - Provide leadership in crafting viable scenarios for space development and evolution
  - Utilize space to pioneer and model new behaviors and gain new insights

- Become a model for blending evolutionary and revolutionary concepts
  - Appropriately honor past experience
  - Realistically understand the present environment
  - Creatively align the future

FUTURE JSC OUTREACH ROLES

- Center of strategic development for human space exploration and evolution
  - Lunar initiative
  - Mars initiative
- Center of influence for development of space commerce in Earth orbit
  - Leverage ISSA/Shuttle programs
  - Develop community partnerships
- A cutting edge agency for establishing creative and innovative organizational and cultural change
  - Internal transformations
  - External relationships

INTEGRATED MICRO-NAN0 TECHNOLOGY SPACE APPLICATIONS CONFERENCE

- EXPLICIT THEMES
  - Nano-Electronics
  - Micro-Electro-Mechanical Systems
  - Application Specific Integrated Micro-instruments
  - Space Applications

- IMPLICIT THEMES
  - Space Flight Demonstrations
  - Soft Technology Associated with People, Culture and Community

BUILD BRIDGES
TEAR DOWN WALLS

OPEN AREAS TO EXPLORE

- Space has a vital role in the evolution of humanity
  - Science/technology/commerce tend to push to the future
  - Holistic principles tend to pull to the future
- Reality is created through engagement with other energy sources
  - Open systems engage with their environments
  - Understanding and developing relationships and interconnections is extremely important
- Integration of future technologies (bio, nano, and noetic) may apply to IOU space (inner, outer, and under)
- commercial space may be common ground for integrating civil space and security space in the future
- Must honor both hierarchical and democratic forms of organization in the future
  - Need to unlock and align collective creative energy
  - Need to develop a capability to think globally and act locally

CELEBRATION

It is not the critic who counts; not the man who points out how the strong man stumbles, or where the doer of deeds could have done them better. The credit belongs to the man who is actually in the arena, whose face is marred by dust and sweat and blood; who strives valiantly; who errs, and comes short again and again, because there is no effort without error and shortcoming; but who does actually strive to do the deeds; who knows the great enthusiasms, the great devotions; who spends himself in a worthy cause; who at the best knows in the end the triumph of high achievements, and who at the worst, if he fails, at least fails while daring greatly, so that his place shall never be with those cold and timid souls who know neither victory nor defeat.

Theodore Roosevelt, 1910

71069

Nanoelectronics: Opportunities for future space applications

23059

1P

Gary Frazier

Texas Instruments Incorporated
P. O. Box 655936
Mail Station 134
Dallas, TX 75265
Phone: (214) 995-2844 Fax (214) 995-2836
e-mail: gfrazier@resbld.csc.ti.com

Further improvements in the performance of integrated electronics will eventually halt due to practical fundamental limits on our ability to downsize transistors and interconnect wiring. Avoiding these limits requires a revolutionary approach to switching device technology and computing architecture. Nanoelectronics, the technology of exploiting physics on the nanometer scale for computation and communication, attempts to avoid conventional limits by developing new approaches to switching, circuitry, and system integration.

This presentation will overview the basic principles that operate on the nanometer scale that can be assembled into practical devices and circuits. Quantum resonant tunneling (RT) will be used as the centerpiece of the overview since RT devices already operate at high temperature (120°C) and can be scaled, in principle, to a few nanometers in semiconductors. Near- and long-term applications of GaAs and silicon quantum devices will be suggested in signal and information processing, memory, optoelectronics, and RF communication.

Text of full paper not available at time of publication.

Microelectromechanical Systems

Dr. Kaigham J. Gabriel
Deputy Director
Electronics Technology Office
Advanced Research Projects Agency

ARPA views Microelectromechanical Systems (MEMS) as a revolutionary enabling technology that merges computation and communication with sensing and actuation to change the way people and machines interact with the physical world. Using the same fabrication processes and materials that are used to make microelectronic devices, MEMS conveys the advantages of miniaturization multiple components, and integrated microelectronics to the design and construction of integrated electromechanical systems.

MEMS is a manufacturing technology that will impact widespread application including: miniature inertial measurement units for competent munitions and personal navigation, distributed unattended sensors for asset tracking and environmental/security surveillance, mass data storage devices, miniature analytical instruments, a range of embedded pressure sensors for passenger car, truck and aircraft tires, non-invasive biomedical sensors, fiber-optic components and networks, distributed aerodynamic control, and on-demand structural strength. The long-term goal of ARPA's MEMS program is to merge information processing with sensing and actuation to realize new systems and strategies for both perceiving and controlling systems, processes and the environment. Short-term goals include: demonstration of key devices, processes and prototype systems using MEMS technologies; development and insertion of MEMS products into commercial and defense systems; and lowering the barriers to access and commercialization by catalyzing an infrastructure to support shared, multi-user design, fabrication and testing.

The MEMS program has three major thrusts: advanced devices and processes, systems design and development, and infrastructure. These three thrusts cut across a number of projects and focus application areas including: inertial measurement, fluid sensing and control, electromagnetic and optical beam steering, mass data storage, signal processing, active structural control, precision assembly and distributed networks of sensors and actuators.

Advanced Devices and Processes

Advanced devices and processes will exploit monolithic actuators, sensors and electronics to achieve new functionality, increased sensitivity, wider dynamic range, programmable characteristics, designed-in reliability and self-testing. New device concepts include: the integration of microdynamic devices with communication, control and computation components, miniature electromechanical signal processing elements (tuning elements, antennas, filters, mixers), miniature optomechanical devices (cross-bar switches, fiber-optic interconnects and aligners, deformable gratings and tunable interferometers), force/motion balanced accelerometers and pressure sensors, miniature analytical equipment (gas chromatography and mass spectrographs on a chip, DNA analysis chips), process control (HVAC equipment, mass flow controllers), and simultaneous, multi-parameter sensing with monolithic sensor clusters. The long-range goal of this thrust is to produce technology that allows the system designer to intermingle sensing, actuation, computation, communication and control. Examples of projects in this area include: MEMS surface micromachining processes integrated with all-CMOS microelectronics for low-cost, monolithic motion detection (safing, fusing and arming functions, tamper detection, automotive impact sensors), integrated actuators and optical waveguides for extreme and hazardous environment sensing, polymerase

chain reaction chambers with integrated heaters, thermistors, fluid valves, and channels (field-portable pathogen detection, enhanced and accelerated sequencing) and fluid valves and regulators appropriate for hydraulic and pneumatic systems pressures and flows. Advances in these areas will be paced by, among others, the development of new material deposition, removal and shaping processes, the merger of hybrid processes (e.g. microelectronics and optoelectronics), and control strategies for inertia-negligible, friction/viscous force-dominated structures.

Systems Design and Development

Systems design and development will focus on high-density arrays that achieve their function or macroscopic action through the coordinated microscopic action of multiple, identical, and relatively simple microactuators. The batch fabrication inherent in photolithographic-based processing, makes it possible to fabricate thousands or millions of components, and their interconnections, as easily and at the same time that it takes to fabricate one component. This multiplicity allows additional flexibility in the design of electromechanical systems to solve problems. Rather than designing components, the emphasis can shift to designing the pattern and form of interconnections between thousands or millions of components. The diversity and complexity of function in ICs is a direct result of the diversity and complexity of the interconnections and it is the differences in these patterns that differentiate a microprocessor from a memory chip. MEMS makes available a disturbed approach to design for solving problems in electromechanical systems design.

Requirements for the construction of high-density array MEMS include the development of a high-yield, high-uniformity fabrication processes for constituent components and algorithms for embedded control of multiple (> 100,000) devices to achieve macroscopic function through distributed, coordinated action of individual elements. New concepts in the integration of multiple devices to form microdynamic systems include: distributed, plug-in sensors interconnected by wired or wireless networks sharing a common communication/power bus, high-fidelity inertial sensing from massively-parallel inertial elements integrated with signal processing circuitry, combustion control through distributed and precise control of reactants and conditions, small-area and low-power displays based on electromechanical deferrable gratings, and increased structural strength from distributed detection and adjustment of material buckling. High-density microactuator arrays (>100/cm^2) would find applications in deformable surfaces for underwater or air vehicle control, high-density data storage systems, integrated source/optics projection displays, direct-wire fine-line lithography, distributed and accelerated analytic instrumentation, precision parts handling and assembly, and on-demand structural strength.

Infrastructure

Infrastructure activities are focused on the services, tools, processes and equipment that will accelerate the affordability and manufacturability of MEMS devices and systems. While MEMS is a new way to make electromechanical systems that leverages microelectronics fabrication, significant differences in MEMS devices and fabrication processes, particularly at end-stage processes and interfaces, require new processing and packaging approaches. Conventional electronics packaging and interconnects seek to provide an appropriate electrical, thermal and mechanical environment for networks of electronic devices. Such packaging often also aims to shrink large quantities of electronics into a small volume, with attendant improvements in performance and reliability. The interface/package manufacturing part of this thrust acknowledges that in many cases the electronics and interconnections need to be not only small, but conform to, rather than dictate the system form factor. Examples are the skin of a hypervelocity missile, an unattended sensor, or the knee joint of an exoskeleton. In addition to

providing appropriate isolation or protection from some of the environment, packages for MEMS components and systems need to also allow controlled access to selected physical parameters that are either being sensed or controlled. Examples include deformable gratings that need access to light but need to operate in vacuum, integrated fluid systems (access to samples for analytical instruments, distributed miniature flaps that need to be in the boundary layer of an aerodynamic stream), and silicon carbide sensors that need to access pressure and temperature inside combustion engines.

Design and simulation tools that combine process descriptors; material characteristic data bases; finite-element packages for structural, electromagnetic fields, fluid interactions, electrostatic fields and electronic device modeling are an integral and continuing part of the infrastructure. Also being supported under this thrust area are CAD design and layout tools that incorporate and allow free-form geometries, cell-design libraries (submission protocols, archiving and recall), parameterized cell models, and dynamic visualization simulators. A rich and robust design and simulation environment is necessary to provide an infrastructure for current and future activity in the field.

Complementing and synchronized with the design and simulation tools developments are the support and development of affordable, distributed fabrication services or users in industry, academia and the government. A surface micromachining fabrication service has and is providing hundreds of distributed users from diverse backgrounds affordable access to micromachining processes at regular, scheduled intervals. This access is accelerating both innovation and commercialization of MEMS products and is being used by other Federal and service agencies for education and training programs in MEMS.

Defense and Technology Impact

MEMS components and systems are moving from being scientific curiosities to applications in sensors, displays and data-storage. DoD has been slow to embrace these technologies because of a lack of demonstrated manufacturability, designability and relevant systems concepts. This program has visibly reduced the risk associate with MEMS devices by developing key technologies and merging advances in MEMS with wireless communication and low-power electronics to demonstrate that DoD-relevant devices can be designed and manufactured. Examples include: high dynamic-range accelerometers for precision munitions; portable inertial guidance systems for personal navigation and head-mounted display orientation sensing; on-demand structural strength for lighter weight platforms; passive IFF systems, condition-based maintenance with embedded MEMS devices; hazardous environment (vehicle engines) sensors, multi-parameter unattended ground sensors for wide-area surveillance, and flow-control deformable surfaces for advanced maneuverability of air/undersea vehicles and projectiles.

Because devices and systems that will be produced or enabled by this program can be expected to be deployed in large numbers, affordability and manufacturing issues are key to DoD use. The manufacturing processes resulting from this program would be capable of producing a diversity of components and systems without retooling and to realize near-equivalent unit costs form either prototype or full-scale production quantities. Procurement costs for new components, systems and spare parts would decrease, inventory/storage costs would be reduced and manufacturing capability/facilities could be consolidated. With many of the proposed technologies' starting points based on proven, mature integrated circuit fabrication techniques, the development, acceptance and sourcing of the devices and technologies will be accelerated.

The ability of MEMS to gather and process information, decide on a course of action and control the environment through actuators increases the affordability, functionality and the

number of smart systems. Microdynamic devices and systems will be merged with communication, control and processing components; three-dimensional microfilming/machining technologies; and embedded/flexible packaging techniques. The range of current design and simulation tools for electronic devices will be extended with three-dimensional mechanical and electric field modeling modules and process-related material property modules to improve the design, integration and operation of microdynamic devices, integrated wireless communication components and low-power electronic systems. The enhanced capability enabled by MEMS will increasingly be the product differentiator of the 21st century, pacing the level of both defense and commercial competitiveness.

Session 1: Space Systems & Requirements (Monday, AM)

Session Chair: Charles Price (NASA/JSC) and Siegfried Janson (Aerospace)

Considerations for Micro- and Nano-Scale Space Payloads
by David A. Altemir, NASA/JSC

Sprint: The First Flight Demonstration of the External Work System Robots
by Charles Price and Keith Grimm, NASA/ JSC

Silicon Satellites: Picosats, Nanosats, and Microsats
by Siegfried Janson, The Aerospace Corporation

Implications of Gun Launch to Space for Nanosatellite Architectures
by Miles R. Palmer, Science Applications International Corporation

Vehicle Tracking System Using Nanotechnology Satellites and Tags
by Dino Lorenzini, SpaceQuest, Ltd.,and Chris Tubis, National Semiconductor Corp.

Teledesic Global Wireless Broadband Network: Space Infrastructure Architecture, Design Features and Technologies
by James Stuart, Teledesic Corporation

7107

CONSIDERATIONS FOR MICRO- AND NANO-SCALE SPACE PAYLOADS

2305?6
8 P

David A. Altemir
NASA/Lyndon B. Johnson Space Center

Abstract

This paper collects and summarizes many of the issues associated with the design, analysis, and flight of space payloads. However, highly miniaturized experimental packages, in particular, are highly susceptible to the deleterious effects of induced contamination and charged particles when they are directly exposed to the space environment. These two problem areas are addressed and a general discussion of space environments, applicable design and analysis practices (with extensive references to the open literature), and programmatic considerations is presented.

Introduction

The use of advanced micro- or nano-technology in space payloads will increasingly provide a highly visible avenue by which to advance the state-of-the-art in ultra-large-scale integrated systems. Just as the aerospace industry fueled the development of integrated microelectronics over twenty-five years ago, the unique requirements of space systems to minimize weight and maximize performance will undoubtedly contribute to the extension of our engineering capabilities to the nano-scale. Since launch systems are not expected to employ these new technologies initially, flight experiment payloads (and vehicle diagnostic systems) will provide the first opportunities to demonstrate the practicality of these new technologies.

This paper collects and summarizes many of the issues associated with the design, analysis, and flight of highly miniaturized space payloads. Space environments, applicable design and analysis practices, and programmatic considerations will be discussed. Since many advanced micro- and nano-technology development activities exist outside the mainstream aerospace community, the information related here may serve as a useful introduction to potential payload designers and managers interested in flying space payloads.

Because studies of space environments, design and analysis strategies, and launch vehicle program requirement documents comprise a large body of literature, only a top-level overview of space flight in low earth orbit (LEO) will be given here and extensive reference will be made to the detailed literature.

Space Environment

Radiation

Thermal radiation is the primary mechanism for heat transfer in space. As a result, exposed payloads may be susceptible to very large temperature gradients. In predicting the operating temperatures of space hardware, a detailed numerical analysis is usually performed taking into account material optical properties and sun angles. The tailoring of surface properties for radiative heat transfer is an especially important element of payload design. However, it should be noted that these surface properties may be susceptible to changes due to contamination or interactions with photons, nuclear particles, or electrons, which will be discussed later [1].

In calculating temperatures, an accepted value for the solar heat flux is 1371 W/m^2 with a variance of ± 10 W/m^2 [2]. As a result of this flux, spacecraft in LEO typically experience naturally-induced surface temperatures ranging from -150 F to 250 F.

Some of the solar flux is reflected by a planet and is referred to as albedo. Albedo forms a secondary contribution to the heat flux incident upon spacecraft and is highly dependent upon time of day and other orbital parameters. Calculations of albedo can be made using specialized computer programs. References 3 and 4 are good resources for the thermal analysis of spaceflight systems and orbital thermal analysis software is available through the COSMIC software repository located at the University of Georgia.

Other forms of radiation, such as ultraviolet (UV), nuclear, and cosmic radiation, may also drive important considerations during design. For short duration micro- or nano-scale payloads, UV radiation, in particular, plays a significant role in the degradation of spacecraft materials and has special implications for thermal control. For example, the synergistic effects of UV radiation and induced contaminant films have been known to cause the darkening of optical, electronic, and thermal control equipment [5,6]. For longer duration missions, higher energy photons , such as x-rays, become more important and the fluence of these quanta is such that physical damage to materials can occur.

Upper Atmosphere

Typically, vacuum levels in LEO lie generally between 10^{-6} and 10^{-10} torr largely depending whether the control surface of interest is towards the ram or wake direction of flight. A lesser source of variability is due to fluctuations in solar activity. However, levels of vacuum as high as 10^{-14} torr are possible with the Wake Shield Facility (WSF), for example, which is a free-flying payload experiment carrier which has been flown aboard the Shuttle. For the simulation of on-orbit conditions, MSIS-90, an upper atmospheric database also available through COSMIC, has been widely employed in analyses of flight hardware with the LEO environment.

For payloads mounted on launch or re-entry vehicle external surfaces (such as data acquisition , a standard atmospheric model may be useful in approximating the ambient pressure and gas species to which payloads will be subjected. For modeling re-entry conditions, the GRAM-90 atmospheric model has been widely used and is available through COSMIC [7]. The Space Shuttle program also uses the Revised Range Reference Standard Atmosphere and the 1963 Patrick Air Force Base Standard Atmosphere to represent launch ascent conditions.

Induced Contamination

The contamination of payloads can be significant, especially for highly miniaturized payloads whose operation are more likely to suffer from relatively small contaminant depositions. Optics, radiators, sensors, and antennas are especially at risk since the accidental deposition of foreign materials on their surfaces may impair the successful operation of these devices (see refs. 8 and 9). Material surface properties may also be affected by contamination which has implications for thermal control [10]. Gases evolved (i.e., outgassed) from spacecraft materials are often the source of contamination for which the judicious selection of materials during design is the best prevention.

Material selection should be based on experimentally measured outgassing levels of the candidate materials (see ref. 11). ASTM E 1559-93 is a particularly attractive method for the testing of spacecraft materials [12]. However, it can be said that anodized aluminum and quartz are routinely employed for exposed spacecraft surfaces and may represent convenient materials for the packaging of micro- and nano-scale payloads. In contrast, silicones are problematic from the standpoint of contamination and should be avoided. In many cases, material outgassing may be minimized by conditioning the hardware to vacuum or a purging flow of inert gas prior to assembly (see ref. 13).

An appropriate design analysis may also involve the accounting of the outgassing species based on experimentally measured mass fluxes and a simulation of their interactions with the natural environment and spacecraft surfaces. MOLFLUX (available through COSMIC) is a software package that has been especially useful to this end [14]. However, several detailed assessments of contamination environments are already available for the Shuttle, International Space Station Alpha (ISSA), and Spacelab platforms (e.g., refs. 15, 16, and 17).

Plasmas and Charged Particles

While traveling through the ionosphere which begins approximately 50 to 70 km above the Earth's surface, a spacecraft will encounter effects due to plasma [2]. This plasma is created by the photo ionization of the ambient neutral atmosphere. Charged particles are present in the form of positively charged ions and free electrons that may contain enough energy to penetrate several centimeters of metal. However, charged particles with such high energies are not dominant at LEO altitudes. Nonetheless, the number of charged particles at LEO may be enough to confuse or blind certain sensors.

The characteristics of plasmas are dependent upon plasma density which is expressed as the electron number density. This density is defined by altitude, local time, season, and amount of solar activity. Some variations in plasma density are shown in Figure 1.

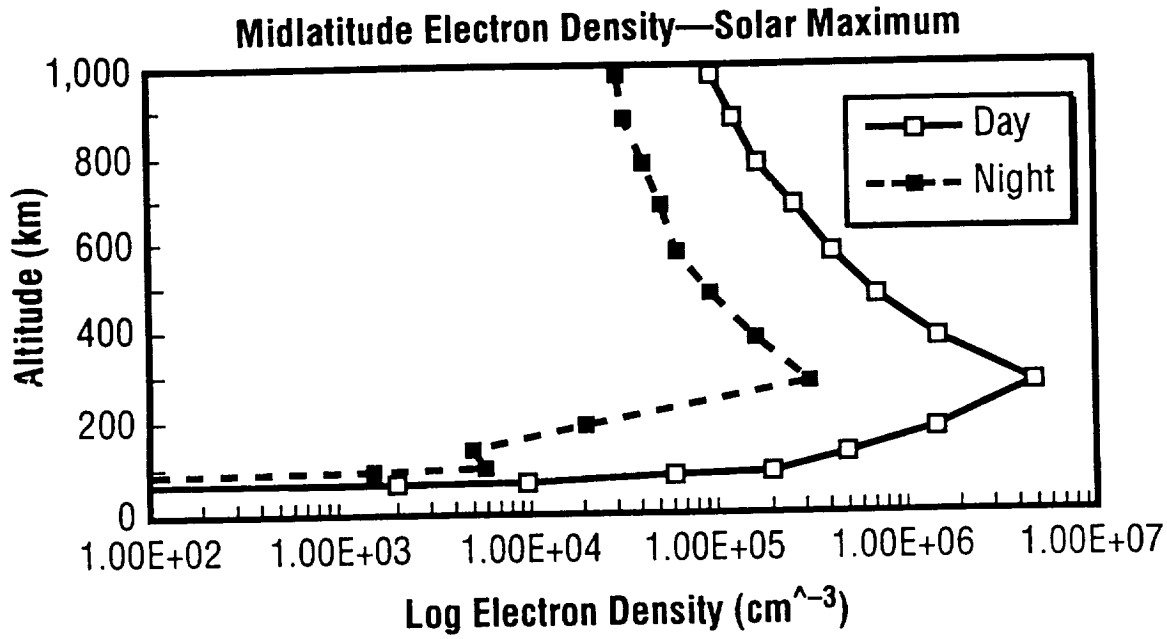


Figure 1. LEO Plasma Density at Solar Maximum [19].

Plasma can result in spacecraft charging, electromagnetic interference (such as radio frequency signals), as well as the erosion of spacecraft surfaces [18]. These phenomena may drive special considerations for the design of micro- and nano-scale payloads even in LEO (see ref. 20). However, through the careful selection of materials, the effects of charging and material erosion can be mitigated. The electrical biasing or shielding of electronic equipment, too, can be of significant help for this area of concern.

Another significant photo ionization effect at LEO altitudes (130 to 190 km) is the splitting of diatomic oxygen into monatomic oxygen. Monatomic oxygen is highly reactive and is responsible for the degradation of many nonmetallic spacecraft materials. To minimize these effects, it is preferable to locate payloads on the wake side of the host spacecraft where the exposure to the monatomic oxygen flux is less [5]. Reference 21 provides an assessment of atomic oxygen and ultraviolet exposures aboard the Shuttle (STS-46) and reference 22 presents an overview of the measured material reactivities for that flight.

Magnetic Fields

The Earth's magnetic field traps charged particles and deflects low-energy cosmic rays. The magnetic field consists of dipoles that result in a field strength at Earth's surface of 0.3 gauss at the equator and 0.6 gauss at the poles [2]. Currents from the magnetosphere cause deviations from the near-Earth field at altitudes greater than 2000 kilometers. Geomagnetic storms caused by solar activity result in fluctuations in the magnetic field strength. However, magnetic fields can be determined for orbital spacecraft by using a spherical harmonic expansion model :

$$\vec{B} = -\nabla \vec{U}$$

for which various expressions for U, the magnetic potential, are available in the literature [e.g., ref. 2].

Microgravity

Gravitational forces upon spacecraft can be approximated to within 0.1 percent using the central-force model:

$$\vec{F} = \frac{\mu_E m}{r^2} \cdot \hat{r}$$

where,

μ_E = Earth's gravitational constant = 3.986012×10^{14} N m^2/kg
m = Mass of spacecraft
r = Distance from center of Earth

However, gravitational models are available that offer accuracies of a few parts in a million (e.g., ref. 23).

Usually, high frequency accelerations induced by the host spacecraft are of more serious concern for the design of highly miniature payloads. Fortunately, the primary vibrational modes of most spacecraft structures lie in the range of 5 to 200 Hz and are too low to affect the operation of most micromechanical devices. However, more significant vibrations in the range of 200 to 2000 Hz may arise as a result of acoustic noise in the payload compartment [24,25]. Unfortunately, detailed vibroacoustic data is usually not comprehensive for most commercial payload environments and, oftentimes, only a few data points are available from the commercial launch vehicle operator in order to characterize a launch vehicle's vibrational environment. However, microgravity assessments are available for the Shuttle (e.g., refs. 26 and 27).

Orbital Debris

Meteroids pose a threat to orbiting spacecraft especially at geosynchronous altitudes (800 to 1000 km) and higher orbital inclinations [19]. In many cases, the concern over orbital impact damage from natural sources will be negligible especially if shielding is employed in some way. However, artificial sources of particle impacts, such as waste water dumps for example, may be of larger concern (see ref. 28). However, the probability of particle impacts is inversely proportional to the size of the particle and, therefore, degradation of micro- and nano-scale payloads is more likely to occur due to impacts of very small (i.e., less than 100 μm) particles. The size of these small payloads offers an additional advantage in that the probability of an impact is further reduced due to their low visibility as a target.

For payloads for which orbital impacts remain a consideration, numerous models are available for predicting the vulnerability of such a threat. Reference 29 is a well-accepted model that characterizes the population of orbital debris. A caveat that has been noted, however, in that this model may underestimate particles larger than 2 cm [30].

Accessing Space

Space Shuttle

Shuttle payloads may be either in the mid-deck or in the payload bay depending on whether access to open space is required. For payload bay payloads, the period between 1998 and 2002 will be marked by Space Shuttle missions dedicated to the delivery of space station elements. These missions will maximize the capacity of the payload bay thereby displacing the smaller packages traditionally used for scientific experimentation in the open space environment. However, by designing ultra small payloads that do not require Shuttle utilities and are largely unobtrusive, the micro- and nano-scale payload designer will have a unique opportunity to couple space science experimentation in the payload bay with the technological advancement of highly miniaturized integrated systems in spite of the hardships that conventional payloads will face.

International Space Station Alpha

At this time, payload manifesting for the International Space Station Alpha (ISSA) has not yet begun. However, the idea of micro- and nano-payloads is being promoted at NASA and, at this time, remains largely conceptual with a significant degree of enthusiasm.

Expendable Launch Vehicles

An important consideration when developing highly miniaturized payloads, especially for flight on commercial launch vehicles, is to employ a high degree of autonomy that minimizes (or eliminates altogether) the need for utilities supplied by the launch vehicle (i.e., power, thermal control, data acquisition and telemetry). It is also important to make sure that payloads do not adversely affect other payload customers by inducing excessive contamination or electromagnetic interference.

Summary

An overview of the LEO space environment and its effects upon spacecraft and space payloads has been given. Also, selected design and analysis guidelines have been referenced and the importance of designing small, unobtrusive micro- and nano-scale payloads has been emphasized. In conclusion, Table I outlines some possible space environmental effects and influences that should be considered during the development of highly miniaturized space systems.

Table I. Possible Space Environment Effects and Influences on Micro- and Nano-Scale Payloads.

	Micromechanical Systems	**Micro- Electronics and Photonics**	**Microfluidic Systems**
Radiation	fatigue;cross-linking/brittle transitions; thermally-induced vibrations; solid diffusion	semiconductor transitions; dielectric properties; solid diffusion	fluid viscosity; surface tension; Brownian motion; Marangoni flow
Upper Atmosphere	pressure equalization; outgassing	pressure equalization; outgassing	pressure equalization; outgassing
Induced Contamination	mass changes; changes in dynamic response; changes in surface properties	changes in surface properties	changes in surface properties
Plasma and Charged Particles	static charging	static charging; electromagnetic interference	static charging
Microgravity	minor effects	negligible effects	minor effects
Orbital Debris	structural damage (minor concern)	structural damage (minor concern)	structural damage (minor concern)

Acknowledgments

The author wishes to express his gratitude to Steven Koontz of the Johnson Space Center and Laura Eadie of Purdue University for their assistance in the compilation of the reference material and also to Lubert Leger of the Johnson Space Center for his review of this work.

References

1. Durcanin, J.T., Chalmers, D.R., Visentine, J.T., "The Definition of the Low Earth Orbital Environment and its Effect on Thermal Control Materials", *AIAA 22nd Thermophysics Conference*, AIAA-87-1599. (1987)

2. Anderson J., *et. al.*; "Natural Orbital Environment Guidelines for Use in Aerospace Vehicle Development", NASA Technical Memorandum 4527. (1994)

3. Gilmore D.G., *Satellite Thermal Control Handbook*, Aerospace Corporation Press, El Segundo, CA. (1994)

4. Griffin, M.D., French, J.R., *Space Vehicle Design*, American Institute of Aeronautics and Astronautics. (1991)

5. Koontz, Steven L., NASA/Johnson Space Center. Personal communication.

6. Hall, D.F., Fote, A.A., "Thermal Control Coatings Performance at Near Geosynchronous Altitude", *Journal of Thermophysics and Heat Transfer.* (1992)

7. Justus C.G., Alyea F.N., Cunnold D.M., Jeffries III W.R., Johnson D.L.; "The NASA/MSFC Global Reference Atmospheric Model - 1990 Version (GRAM-90)", NASA Technical Memorandum 4268 Part I. (1991)

8. O'Donnell, T., "Evaluation of Spacecraft Materials and Processes for Optical Degradation Potential", *SPIE Technical Symposium Proceedings.* (1982)

9. Johnson, R. B., Connatser, R., *Contamination Analysis of SSF Candidate Materials, Final Report*, NASA Contractor Report 184431. (1991)

10. Glassford, A.P., "Spacecraft Contamination Database", *Proceedings of SPIE Optical System Contamination: Effects, Measurement, Control II.* (1990)

11. Bertrand, W.T., Wood, B.E., "Solar Absorptance of Optical Surfaces Contaminated with Spacecraft Material Outgassing Products", Arnold Engineering Development Center, Arnold Air Force Base report AEDC-TR-93-7. (1993)

12. ASTM E 1559-93, "Standard Test Method for Contamination Outgassing Characteristics of Spacecraft Materials", American Society of Testing and Materials. (1993)

13. Scialdone, J.J., "Spacecraft Thermal Blanket Cleaning: Vacuum Baking or Gaseous Flow Purging", *Journal of Spacecraft and Rockets*, v. 30, no. 2. (1993)

14. Rios, E.R., Rodriguez, R.T., Hakes, C.L., Ehlers, H.K.F., Laskey, K.J., Russ, T.W., "MOLFLUX: Molecular Flux User's Manual, Revision 2", NASA/Johnson Space Center publication JSC-22496. (1992)

15. Ehlers, H.K.F., Jacobs, S., Leger, L.J., "Space Shuttle Contamination Measurements from Flights STS-1 Through STS-4", *J. Spacecraft*, v. 21, no. 3. (1984)

16. Ehlers, H.K.F., Hakes, C.L., Soares, C.E., Rios, E., Theall, J.R., *Internal Space Station Alpha External Induced Contamination Assessment Report in Support of the Incremental Design Review #1*, NASA/Johnson Space Center publication JSC-26954. (1995)

17. Bareiss, L.E., Payton, R.M., Papazian, H.A., *Shuttle/Spacelab Contamination Environment and Effects Handbook*, NASA Contractor Report 4053. (1987)

18. Tascione, T.F., *Introduction to the Space Environment*, Orbit Book Company. (1988)

19. James, B.F., Norton, O.W., Alexander, M.B.; *The Natural Space Environment: Effects on Spacecraft*, NASA Technical Reference Publication 1350. (1994)

20. Purvis, C.K., Garrett, H.B., Whittlesey, A.C., Stevens, J.N., *Design Guidelines for Assessing and Controlling Spacecraft Charging Effects*, NASA Technical Paper 2361. (1984)

21. Koontz, S.L., Leger, L.J., Rickman, S.L., Hakes, C.L., Bui, D.L., Hunton, D.E., Cross, J.B., "Oxygen Interactions with Materials III-Mission and Induced Environments", *Journal of Spacecraft and Rockets*, v. 32, no. 3. (1995)

22. Koontz, S.L., Leger, L.J., Visentine, J.T., Hunton, D.E., Cross, J.B., Hakes, C.L., "EOIM-III Mass Spectroscopy and Polymer Chemistry: STS-46, July-August 1992", *Journal of Spacecraft and Rockets*, v. 32, no. 3. (1995)

23. Marsh, J.G., Lerch, F.J., Putney, B.H., Christodoulitis, D.C., Smith, D.E., Pelsentreger, T.L., and Sanchez, B.V.: "A New Gravitational Model for the Earth from Satellite Tracking data: GEM-T1." *J. Geophys. Res.*, vol. 93(B6), 1988, pp. 6169-6215.

24. Lauriente, M., Hoegy, H. *eds.*: *ELV Payload Environment: Proceedings of the NASA/Industry Conference on Launch Environments of ELV Payloads*, University Research Foundation. (1991)

25. Warden, R.M., "Design Criteria for Mechanisms Used in Space", American Vacuum Society Series 7, conference proceedings No. 192. (1989)

26. Thomas, D., "Microgravity Evaluation of Columbia Middeck During STS-32", NASA Technical Paper 3140. (1992)

27. Baugher, C.R., Martin G.L., DeLombard, R., "Review of the Shuttle Vibration Environment", American Institute of Aeronautics and Astronautics publication AIAA-93-0832. (1993)

28. Zumwalt, L.C., Lee, A.L., "Assessing Particle Impacts from Orbiter Waste Dumps", American Institute of Aeronautics and Astronautics publication AIAA-94-0251. (1994)

29. Kessler, D., Reynolds, R., Anz-Meador, P., *Orbital Debris Environment for Spacecraft Designed to Operate in Low Earth Orbit*, NASA Technical Memorandum 100 471. (1989)

30. Kessler, D., "Orbital Debris Environment for Spacecraft in Low Earth Orbit", *Journal of Spacecraft and Rockets*, v. 28, no. 3. (1991)

Sprint: The first flight demonstration of the external work system robots

Charles R. Price and Keith Grimm

NASA Johnson Space Center
Houston, TX

The External Work Systems "X Program" is a new initiative by NASA's Code XS, Spacecraft Systems Division of the Office of Spacecraft and Technologies that will, in the next ten years, develop a new generation of space robotics for active and participative support of zero g external operations. The EWS robotics development will center on three areas of emphasis: The Assistant Robot, the Associate Robot, and the Surrogate Robot that will support EVA (External Vehicular Activities) prior to and after, during, and instead of spacesuited human external activities, respectively. The EWS robotics program will be a combination of technology developments and flight demonstrations for operational proof of concept. The first EWS flight will be a flying camera called "Sprint" that will seek to demonstrate operationally flexible, remote viewing capability for EVA Operations, Inspections, and Contingencies for the Space Shuttle and Space Station. This paper will describe the need for Sprint and its characteristics.

Text of full paper not available at time of publication.

71078
030523
16 P

Silicon Satellites: Picosats, Nanosats, and Microsats

Siegfried W. Janson, Mechanics and Materials Technology Center
The Aerospace Corporation, El Segundo, CA, 90009

Abstract

Silicon, the most abundant solid element in the Earth's lithosphere, is a useful material for spacecraft construction. Silicon is stronger than stainless steel, has a thermal conductivity about half that of aluminum, is transparent to much of the infrared radiation spectrum, and can form a stable oxide. These unique properties enable silicon to become most of the mass of a satellite; it can *simultaneously* function as structure, heat transfer system, radiation shield, optic, and semiconductor substrate.

Semi-conductor batch-fabrication techniques can produce low-power digital circuits, low-power analog circuits, silicon-based radio-frequency circuits, and micro-electromechanical systems (MEMS) such as thrusters and acceleration sensors on silicon substrates. By exploiting these fabrication techniques, it is possible to produce highly-integrated satellites for a number of applications. This paper analyzes the limitations of silicon satellites due to size. Picosatellites (~ 1 gram mass), nanosatellites (~1 kg mass) and highly-capable microsatellites (~10 kg mass) can perform various missions with lifetimes of a few days to greater than a decade.

I. Why Silicon?

I.A. Silicon as a Structural Material

Mechanical properties of silicon, aluminum, stainless steel, titanium, and diamond are given in table 1. Silicon is already used as the structural material for microelectromechanical systems (MEMS) due to it's high strength[1] and should be considered for spacecraft structure as well. Single crystal silicon is stronger than aluminum, stainless steel, or titanium, yet it is less dense. Materials with intrinsically high strength-to-density ratios are particularly valuable for spacecraft due to launch costs of ~$10,000 per kg to Low Earth Orbit (LEO) and ~$50,000 per kg to geosynchronous Earth orbit (GEO). Note that titanium has a higher strength-to-density ratio than aluminum or stainless steel, yet it is still more than an order-of-magnitude less than what silicon provides. Single crystal diamond has the best strength-to-density ratio, but the material cost (~$10,000 for a man-made 2 gram crystal) is many orders-of-magnitude higher than the launch cost. Single crystal silicon carbide (see ref. 1) has a strength-to-density ratio between silicon and diamond, and should also be considered for some spacecraft applications.

Table 1. Mechanical properties of silicon compared to other structural materials.
(Data from references 1, 2, 3, 4, and 5)

	Density ($g\ cm^{-3}$)	Yield Strength (MPa)	Strength/Density (MN-m/kg)	Young's Modulus (GPa)
Silicon (single crystal)	2.3	7,000	3.0	~170
Aluminum (2024-T3)	2.8	350	0.13	73
Stainless Steel (304)	8.0	1,000 (cold worked)	0.13	200
Titanium (Ti-6Al-4V)	4.4	900	0.20	115
Diamond (single crystal)	3.5	50,000	14.3	~1,000

While metals are available in a variety of shapes and sizes, single crystal silicon is available in 10, 12.5, 15, and 20 cm diameter cylinders up to 2 meters long. Present manufacturing costs are about $185 per kg which could drop to ~$50 per kg by adopting new refinement techniques.[6] Even though the material cost (per kg) of single crystal silicon is much more expensive than typical spacecraft structural materials, it is still more than an order-of-magnitude less than the launch cost. Silicon is a viable structural material, especially for small satellites.

Thermal properties of silicon, aluminum, stainless steel, titanium, and diamond are given in table 2. Silicon has a high specific heat and a much higher melting point than aluminum. Silicon is a very good heat conductor; it has a higher thermal conductivity than stainless steel and titanium, and about 50% that of aluminum. Silicon also has a thermal expansion coefficient that is ~4 times lower than titanium and about an order-of-magnitude lower than aluminum and stainless steel. Once again, diamond has superior properties, but it must be used sparingly due to its high cost.

Table 2. Thermal properties of silicon compared to other structural materials. Specific heat, thermal conductivity, and thermal expansion coefficient at 300 K.

	Thermal Conductivity (W/m °K)	Thermal Expansion Coefficient (cm/cm °K)	Melting Temperature (K)
Silicon (single crys.)	150	2.5×10^{-6}	1700
Aluminum (2024-T3)	240	2.2×10^{-5}	~850
Stainless Steel (304)	16	1.7×10^{-5}	~1700
Titanium (Ti-6Al-4V)	8	9×10^{-6}	~2100
Diamond (single crys.)	2,100	1.0×10^{-6}	4200

I.B. Silicon as an Optical Material

Silicon can be used as an optical material throughout most of the infrared spectrum. As shown in figure 1, it is transparent between 1.4 and 7 microns (wavelength) and from 25 to beyond 100 microns. The maximum transmission of 54% results from the index-of-refraction of ~3.5 over this wavelength range. Index-matching coatings or surface texturing can be used to bring transmission efficiencies close to 100% over selected ranges of interest.

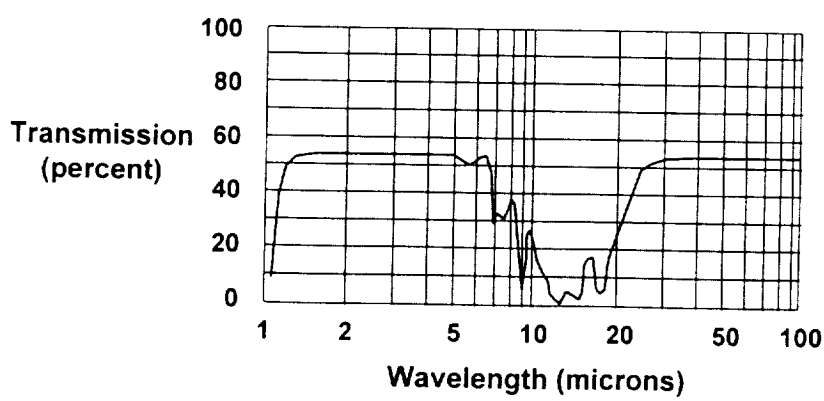


Figure 1. Optical transmission of silicon in the infrared.

I.C. Silicon as a Substrate for Electronics

Microprocessors, microcontrollers, memory, and other digital integrated circuits (ICs) are batch-fabricated primarily on silicon wafers. These wafers are cut from larger single crystals which have been ground into a cylindrical shape. Hundreds of identical devices are fabricated on a single side of a 0.5 to 1 mm thick, 10 to 20 cm diameter, silicon wafer which is cut apart to release the individual rectangular silicon dice. These dice are then mounted in plastic or ceramic carriers which provide hermetic sealing against the environment, mechanical rigidity, improved heat dissipation, and electrical connections which can be soldered or plugged into a circuit board.

Almost the entire mass and volume of spacecraft electronic systems is determined by packaging; i.e. enclosures ("boxes"), connectors, circuit boards, and chip carriers. Multi-chip modules (MCMs) place multiple electrically-connected die in a single carrier to reduce wasted space and increase active device area density. MCMs can achieve 10 to 15 times greater IC packaging densities on circuit boards than conventional single-die packages with greater system performance and reliability at lower cost.[7] The next step in reducing packaging size and mass is to produce wafers with interconnected dice. This wafer-scale integration eliminates unnecessary dicing, wiring, and packaging.

How much silicon "real estate" (area) is required for typical functions? Silicon die areas currently range from ~0.2 cm^2 for microcontrollers to ~2 cm^2 for high end microprocessors. A PIC16C71 microcontroller dice is 0.4 cm x 0.5 cm in size[8] while a PowerPC 604 dice is 1.24 cm x 1.58 cm.[9] Dynamic random access memory (DRAM) die, suitable for use with microcontrollers and microprocessors, require ~0.03 cm^2 of silicon per million bits of storage capacity. Feature sizes are steadily decreasing with time and high-end integrated circuits are becoming more complex. Projections of dice size for DRAMs and high-end microprocessors over the next 15 years, given in table 3, show the interesting result that dice size for these products is expected to increase with time. By the year 2010, high-end microprocessor and DRAM dice could be several cm on a side. To support efficient packing of rectangular dice on circular silicon wafers during fabrication, 30 cm diameter and larger silicon wafers will become common.

Table 3. Roadmap projections for semiconductor technology. (From reference 10)

Year	Smallest Feature (μm)	Dynamic RAM: Dice Size (cm^2)	Dynamic RAM: Billions of Bits per Dice	Microprocessors: Dice Size (cm^2)	Microprocessors: Millions of Transistors per cm^2
1995	0.35	1.9	0.064	2.5	4
1998	0.25	2.8	0.256	3.0	7
2001	0.18	4.2	1	3.6	13
2004	0.13	6.4	4	4.3	25
2007	0.10	9.6	16	5.2	50
2010	0.07	14.0	64	6.2	90

A single ~4 cm x 4cm DRAM dice in the year 2010 could hold 8 *gigabytes* of information; more capacity than is currently available on CD-ROMs and most personal computer hard disk drives. A single DRAM could hold an uncompressed 1000 km x 400 km image with 10-meter spatial resolution and 16 bits of intensity resolution. The command and data handling (C&DH) system for a future satellite could reside in 1 or 2 dice and fit on a 10 cm diameter wafer with extra silicon area for communications systems, power controllers, etc.. If circuit complexity remains fixed, circa 2010 dice could be much smaller than today's versions due to a factor of 5 reduction in feature size. A future microcontroller

equivalent to the PIC16C71, for example, would fit on a dice smaller than 1 mm on a side. The total silicon area required for the C&DH system in a current generation microsatellite would shrink from ~ 2 cm^2 to ~ 0.1 cm^2.

Microprocessors and microcontrollers have been steadily increasing in performance due to increases in operating speed, partially enabled by smaller feature sizes, and improvements in processing architectures. Microprocessor clock speeds were a few MHz during the early 1970's, are in the low 100 MHz range today, and are expected to reach the GHz range by 2010. The INTEL 8086 microprocessor, introduced in 1978, had a 0.33 million-instructions-per-second (MIPS) performance with a 5 MHz clock speed while the INTEL Pentium, introduced in 1993, had a 112 MIPS performance with a 66 MHz clock speed.[11] By extrapolating the exponential performance increases in microprocessors over the last 15 years (about a 1.5 times increase in performance per year), the predicted performance by 2010 is 100,000 MIPS. This is quite staggering when one realizes that NASA's Galileo probe uses a mere 0.5 MIPS processor for C&DH.

While silicon has been the substrate of choice for commercial digital circuits, operating frequencies of ~1 GHz and higher will require special wafers or another substrate such as gallium-arsenide. Currently, silicon offers lower cost and simplified power requirements while gallium arsenide offers higher frequency operation and higher efficiency. A comparison of silicon and gallium-arsenide transistor circuits is given in table 4. The gallium arsenide cost in table 4 is based on a total manufacturing cost of $1500 per 10 cm diameter wafer with a 75% yield.[12]

Table 4. Basic comparison between gallium-arsenide and silicon transistor integrated circuits. Data from reference 12.

Parameter	Gallium-Arsenide MESFET	Silicon Bipolar
Cost per mm^2	$0.25 (10 cm diameter wafer)	$0.10 (20 cm diameter wafer)
Cutoff frequency	18 to 25 GHz	12 to 18 GHz
Breakdown Voltage	15 to 20 V	5 to 9 V
Substrate	Insulating	Conductive
Noise figure @ 5 GHz	1.5 dB	6.0 dB
Bias requirements	positive and negative	positive
Power added efficiency	60%	40%

While data processing circuits may not yet run at GHz clock speeds, satellite communications circuits routinely operate from ~100 MHz to greater than 20 GHz. Below ~1 GHz, conventional CMOS technologies on conventional silicon wafers can provide inexpensive radio-frequency integrated circuits. One group at UCLA is investigating 2-chip and single-chip fully-integrated (rf, analog, and digital) CMOS transceivers for operation in the 900 MHz industrial, scientific, and medical (ISM) band.[13] Their goal is to produce a frequency-hopped spread-spectrum transceiver with up to 160 kilobit/sec data transfer rates at 20 mW power output and 100 milliampere, 3 Volt DC input.[14]

Above 1 GHz, gallium arsenide offers increased DC-to-RF conversion efficiency for power amplifiers and decreased noise for preamplifiers (low noise amplifiers at the receive antenna) used in communications systems. Standard silicon microwave monolithic integrated circuits (MMICs) usually perform poorly above a few GHz because commercial wafers have a resistivity on the order of 10 Ω-cm,

which is too low to act as a good dielectric.[15] Better performance can be achieved using high-resistivity silicon substrates (10,000 Ω-cm and higher for up to 40 GHz operation) or silicon-on-insulator (SOI) construction. A more recent development is silicon-germanium (SiGe) technology that offers higher than 70 GHz operation on silicon substrates.[16,17] Today's silicon technology can be used for satellite communications bands from VHF (very high frequency; ~140 MHz) to S-band (2.5 to 2.7 GHz). Future advancements in silicon technology will push the operating frequency range higher.

I.D. Silicon as a Radiation Shield

The near-Earth space environment is much harder on electronics than the surface environment due to the presence of high-energy electrons, protons, and heavier ions. Elastic scattering of high energy protons and ions results in displacements of stationary target atoms while inelastic scattering results in secondary particle "showers" of lower-energy target atoms or fission daughter products. High-energy electrons, protons, and ions all leave ionizing tracks behind them. Anomalous effects in semiconductor circuits due to these interactions range from a temporary change in logic state, due to the sudden appearance of charge, to permanent substrate atom and charge dislocations which produce altered current-voltage characteristics and possible device failure.[18] Single-event upsets (SEUs) are particle-induced "bit-flips" while latchups are more serious high-current flow conditions generated by new low-resistance paths created by particle-induced ionization trails. Both SEUs and latchups can be controlled by appropriate choice of semiconductor technology, "watchdog" and error-correction circuits, and error-correction software. Continual accumulation of radiation damage, however, ultimately results in device failure.

Table 5, adapted from reference 19, gives rough radiation hardness levels for different types of semiconductor devices. A *rad* is the amount of particle radiation that deposits 100 ergs of energy per gram of target material and the radiation hardness level represents total dose required for device failure. Typical low-power consumer electronic components (CMOS) are designed to operate in our low-radiation biosphere (roughly 0.3 rad/year) but can tolerate 1 to 10 kilorad integrated radiation doses. Unfortunately, the radiation tolerance varies widely from design to design so radiation testing should be performed on selected components. Transistor-transistor logic (TTL) and emitter-coupled logic (ECL) circuits are inherently more radiation hard than CMOS, but they require more power. NMOS, PMOS, I^2L, and silicon-on-sapphire MOS circuits can be fully immune to latchup. Radiation hardening requires a balance between power and circuit availability for choice of technology, mass requirements for shielding, and circuit complexity for latchup and SEU control.

How much silicon radiation shielding is required for a given mission? Dose rates for a silicon target are usually given as a function of grams/cm^2 or thickness of spherical aluminum shielding for a given orbit and given solar conditions (i.e. minimum or maximum solar activity). Silicon and aluminum are next to each other on the periodic table; their nuclei and average atomic masses differ by only one proton, and they have similar densities. Both materials have similar ability to slow down incident energetic electron and protons while silicon generates a slightly higher level of bremsstrahlung X-rays because bremsstrahlung is proportional to the square of the atomic number (14^2 for Si vs. 13^2 for Al). Aluminum and silicon shielding thicknesses in grams/cm^2 are equivalent within the uncertainties of radiation environment estimates.

Table 5. Radiation hardness levels for semiconductor devices

Technology	Total Dose in rads (silicon)
CMOS (soft)	10^3 - 10^4
CMOS (hardened)	5×10^4 - 10^6
CMOS (silicon-on-sapphire: soft)	10^3 - 10^4
CMOS (silicon-on-sapphire: hardened)	> 10^5
ECL	10^7
I^2L	10^5 - 4×10^6
Linear integrated circuits	5×10^3 - 10^7
MNOS	10^3 - 10^5
MNOS (hardened)	5×10^5 - 10^6
NMOS	7×10^2 - 7×10^3
PMOS	4×10^3 - 10^5
TTL/STTL	> 10^6

Figure 2 shows the yearly dose rate due as a function of aluminum shielding thickness (full sphere shielding) for 700 km altitude orbits with inclinations of 28.5° and 98.2°. CMOS circuits with an assumed total radiation dose tolerance of ~3000 rads will require at least 0.3 g/cm^2 aluminum (or 1.3 mm of silicon thickness) shielding for a 1 year on-orbit lifetime in a 700 km, 28.5° inclination orbit. For the more interesting sun-synchronous (98.2° inclination) orbit, about 0.8 g/cm^2 (or 4 mm silicon thickness) is required for a 1 year lifetime and about 3 g/cm^2 (1.3 cm silicon) for a 10 year lifetime. At lower altitudes, significantly less shielding is required, while at higher altitudes, significantly more shielding may be required. Use of more radiation-resistant technologies is the only solution for some orbits.

Figure 3 shows the dose rate dependence as a function of circular equatorial orbit altitude inside spherical aluminum shields with densities of 0.5 g/cm^2 (0.18 cm thick aluminum or 0.21 cm thick silicon) and 3.0 g/cm^2 (1.1 cm thick aluminum or 1.3 cm thick silicon). Note the rapid rise in dose rate with altitude below 1000 nmi (1850 km), the existence of a hard-to-shield proton belt at ~2000 nmi (3700 km), and the existence of an easier-to-shield electron belt at ~10,000 nmi (18,500 km). At geosynchronous Earth orbit (GEO; 35,786 km or 19,320 nmi altitude and 0° inclination) with a maximum dose of 3,000 rads, 0.5 gm/cm^2 (0.22 cm silicon) and 3.0 gm/cm^2 (1.3 cm silicon) shielding give lifetimes of roughly 11 days and 3 years, respectively.

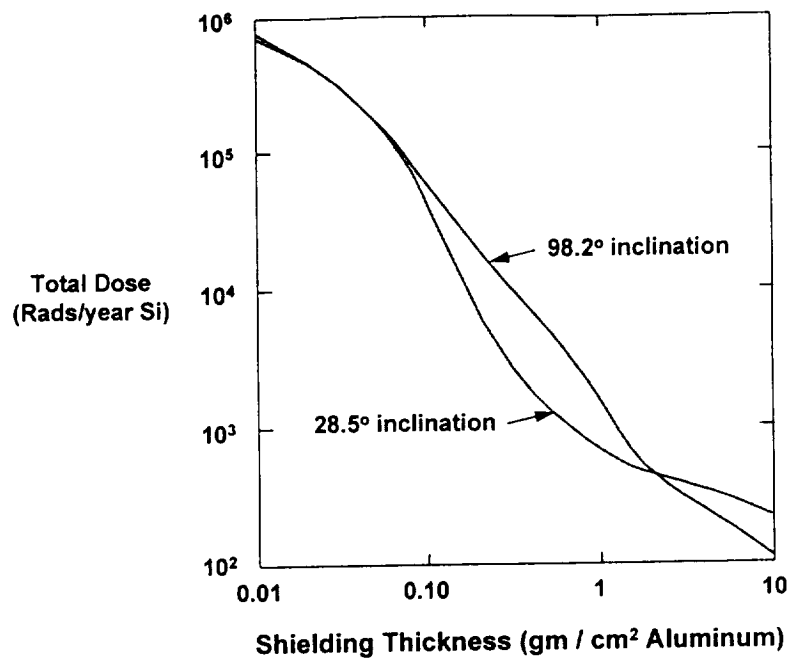


Figure 2: Total yearly dose, under solar maximum conditions, in silicon as a function of aluminum shielding thickness for 700 km circular orbits. Data adapted from reference 20.

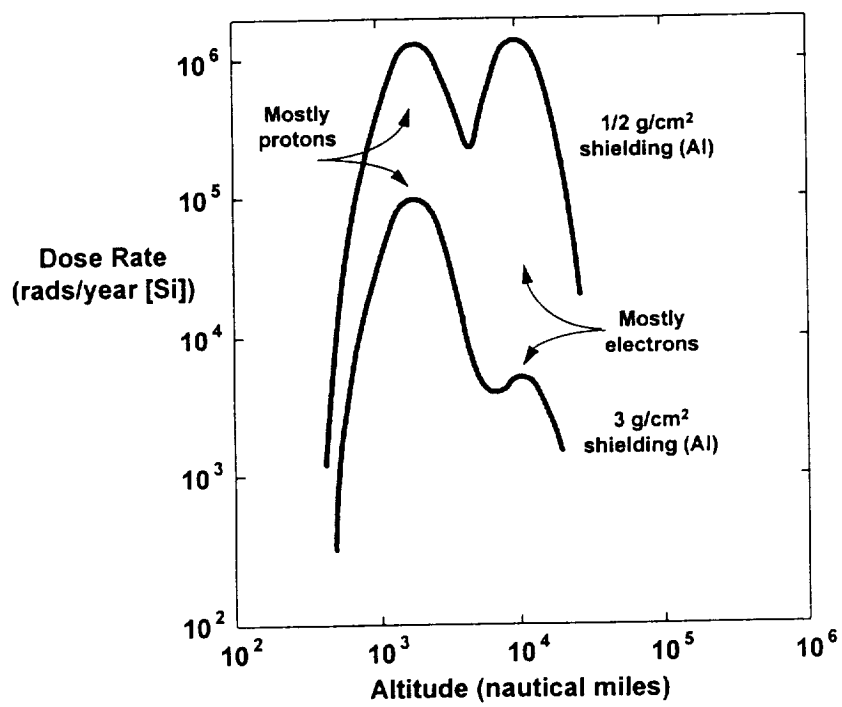


Figure 3. Radiation environment for circular equatorial orbits. (Adapted from ref. 19)

I.E. Silicon as a Substrate for Spacecraft Systems

Microelectro-mechanical systems (MEMS) such as micron-scale silicon diaphragm pressure sensors, acceleration sensors, chemical sensors, and valves have already been demonstrated.[21,22,23,24] Concepts for silicon-based, batch-fabricated chemical and electric propulsion systems suitable for small satellites have also been presented.[25] Integration of MEMS with microelectronics for data processing, memory, signal conditioning, power conditioning, and communications results in a stand-alone "application-specific integrated microinstrument" (ASIM). Examples of MEMS and ASIM applications for spacecraft can be found in reference 26 and in other papers from this conference.

II. Silicon Satellites

II.A. Introduction

The silicon satellite, as introduced in references 27 and 28, presented a new paradigm for space system design, construction, testing, architecture, and deployment. Integrated spacecraft complete with some degree of attitude and orbit control can be designed for mass-production using batch-fabrication techniques. Integrated circuits for C&DH, communications, power conversion and control, on-board sensors, attitude sensors, and attitude control devices can be manufactured on thick silicon substrates that provide structure, radiation shielding, and thermal control. Some conventional components such as batteries and individual solar cells will still be required, but the total number of parts and assembly time will be drastically reduced. The spacecraft, as shown in figure 4, is essentially a multi-ASIM module.

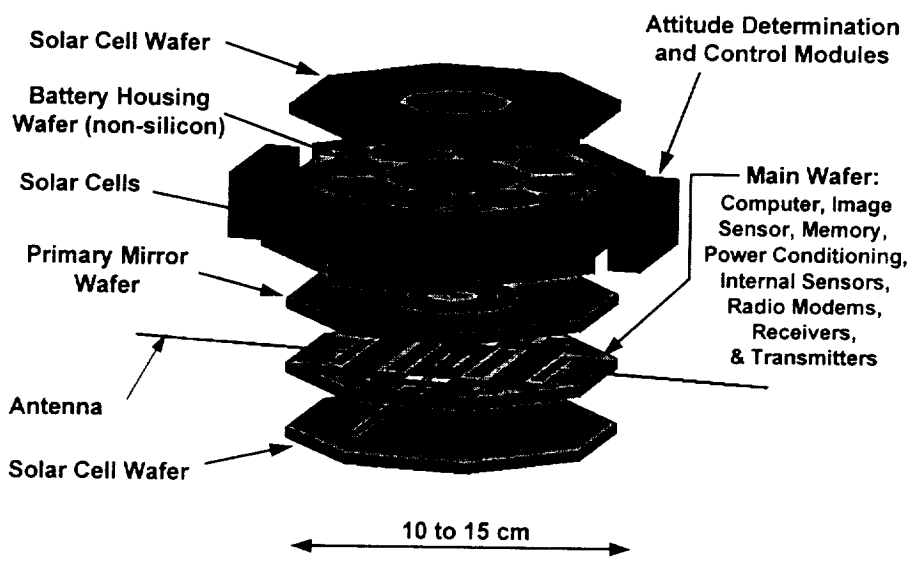


Figure 4. A hypothetical silicon satellite.

Silicon wafers are routinely produced with diameters up to 20 cm which will increase to 30 cm within a decade. Low-volume (10 to 1000 wafers) production of custom circuits and MEMS uses wafers with diameters less than 15 cm. Simple ASIM-based integrated satellites will have dimensions of 10 to

20 cm while more complex configurations using additional non-silicon mechanical structure (i.e. truss beams, honeycomb panels and inflatable structures) will be much larger.

The benefits of batch-fabricated silicon satellites are:

1. Reduced parts count due to integrated electronics, sensors, and actuators on a single substrate,
2. The ability to add redundancy and integrated diagnostics without significantly impacting production cost,
3. Decreased material variability and increased reliability due to rigid process control,
4. Rapid prototype production capability using electronic circuit, sensor, and MEMS design libraries with existing (and future) CAD/CAM tools and semiconductor foundries,
5. Elimination of labor-intensive assembly steps (welding, wiring cable harnesses, etc.)
6. Automated testing of systems and subsystems, and
7. Paper less documentation of designs, fabrication processes, and testing.

Low cost per function is a direct result of the fabrication process; semiconductor batch fabrication techniques evolved within the constraints of consumer-driven market economics. Low mass and volume are simply byproducts of the fabrication process that can be exploited for space applications.

II.B. Satellite Classification

The term "microsatellite" has traditionally been used for satellites with masses between about 100 kg and 10 kg. Recently, the terms "nanosatellite" and "picosatellite" have been used almost interchangeably for 1 kg class vehicles. In an attempt to standardize these names and still keep within the spirit of the prefixes "micro", "nano", etc., I propose a new classification scheme given in table 6.

Table 6. Satellite classification by mass

Classification	Mass Range
Microsatellite	1 kg to 100 kg
Nanosatellite	1 gram to 1 kg
Picosatellite	1 milligram to 1 gram
Femtosatellite	1 microgram to 1 milligram

II.C. Power Considerations

How much solar power can small satellites produce? Assuming that all of the available surface area is covered by solar cells, the extremes occur for a cubic satellite and for a satellite spread out into a ~10 micron-thick sheet; i.e. a flat all-solar-array satellite. Figure 5 shows the power extremes for 20% solar conversion efficiency, random pointing for a cubic satellite, and optimum pointing for a thin sheet satellite with average density equal to silicon. Picosatellites through microsatellites can produce power levels in the 1 to 100 Watt range while femtosatellites are in the microwatt to milliwatt range.

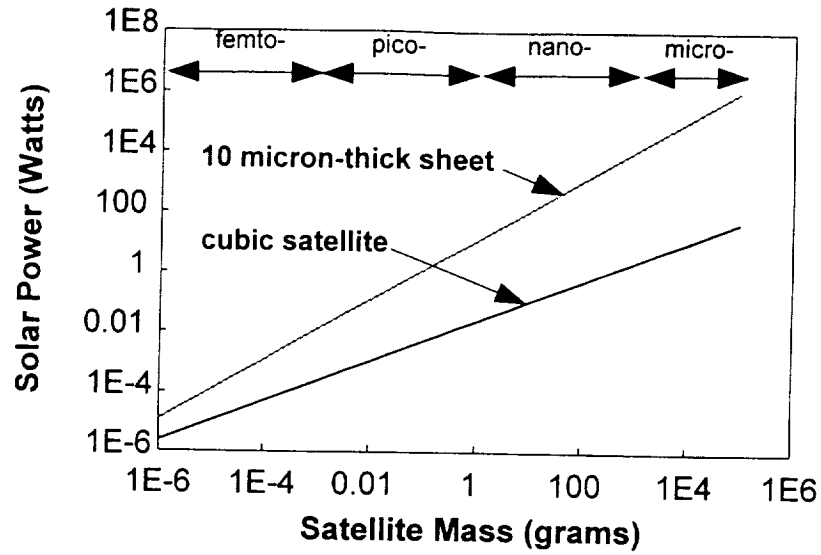


Figure 5. Solar power output ranges, in full on-orbit sunlight, for satellites with an average density equal to silicon and 20% power conversion efficiency over all exterior surfaces.

II.D. Thermal Considerations

The overall thermal balance of a spacecraft is determined by its orbit, its geometry, its surface properties, and its internal design. Solar flux (~1370 W/m^2) has an effective blackbody temperature of 5800 K. The Earth reflects ~30% of this incoming solar radiation back to space; the average reflectance or albedo ranges from 23% at the equator to 74% over Antarctica.[29] Over 95% of direct and reflected solar energy is carried by photons with wavelengths between 0.2 and 2.5 microns. Thermal radiation from the Earth has an effective blackbody temperature of ~300 K which has 98% of the energy carried by photons with wavelengths greater than 5 microns. The primary energy input to a satellite, averaged over an orbit, is direct and reflected solar radiation while the primary energy outflow is infrared emission from the spacecraft at wavelengths greater than 5 microns. The average emissivity (or absorptivity) α in the visible and near infrared wavelength range (0.4 to 2 microns) controls the major heat input to spacecraft surfaces while emissivity ε in the medium to long infrared wavelength range (5 to 50 microns) controls the heat rejection capability of spacecraft surfaces.

At thermal equilibrium, the heat radiated by a satellite to space is just the sum of the energy absorbed plus heat generated internally:

$$\alpha A_s G_s + \alpha A_e G_r + \varepsilon A_e G_e + Q = \varepsilon \sigma T^4 A_r \qquad (1)$$

where A_s is the surface area for absorption of solar energy, A_e is the surface area for absorption of solar energy reflected by the Earth, G_r is the local flux of sunlight reflected from the Earth, G_e is the local flux of thermal energy radiated by the Earth, Q is the internal heat generation rate, σ is the Stefan-Boltzmann constant (5.67 x 10^{-16} W/(m^2*T^4), T is temperature, and A_r is the surface area for heat radiation. As spacecraft shrink in size, surface-area-to-volume ratios and hence surface-area-to-mass ratios increase. Small satellites with body-mounted solar arrays can have power-to-mass ratios equivalent to large satellites with deployable solar arrays, yet still be power-limited. High fractional surface coverage for

solar cells is generally the rule, which results in spacecraft whose thermal balance is determined by solar cell absorptivity and emissivity. This does not allow much latitude in controlling spacecraft temperature ranges for small satellites.

Consider thermal control of a 10 cm diameter, 1.2-kg mass spherical silicon microsatellite at an altitude of 700 km. The IR flux from the Earth G_e is ~200 W/m^2 over the entire orbit and the reflected solar flux G_r is ~260 W/m^2 over about half the orbit. If we assume that the surface is completely covered by 20% efficient solar cells, the orbit average electric power is ~1.5 Watts. With an absorptivity of 0.8 and an emissivity of 0.83, the maximum equilibrium temperature (full sunlight + reflected sunlight + Earth IR + internal heat generation) from eq. (1) is 306 K while the minimum equilibrium temperature (Earth IR + internal heat generation) is 209 K. Conventional spacecraft electronics and batteries cannot tolerate these temperature extremes. Fortunately, the satellite's thermal mass and appropriate insulation techniques can be used to control temperature fluctuations for key spacecraft systems.

Figure 6 shows spacecraft temperature, as a function of time, for 4-cm-diameter, 10-cm-diameter, and 20-cm-diameter solid silicon spheres (nanosatellites and microsatellites) fully covered by solar cells. Their masses are 80 grams, 1.2 kg, and 9.8 kg, respectively. The orbit is a 700 km altitude circular equatorial orbit (0° inclination) and the solar cells have the same efficiency, emissivity, and absorptivity used in the previous equilibrium temperature calculations. The dynamic spacecraft temperature is calculated in one-minute time intervals by numerically integrating a lumped-heat-capacity energy equation:

$$\alpha A_s G_s + \alpha A_e G_r + \varepsilon A_e G_e + Q - \varepsilon\sigma T^4 A_r = \frac{d}{dt}\left(mc_p T\right) \tag{2}$$

where d/dt is the first time derivative, m is the spacecraft mass, and c_p is the constant-pressure heat capacity of silicon (736 joules/kg*K). As diameter increases, mass and thermal capacity increase, which results in reduced temperature swings over an orbit.

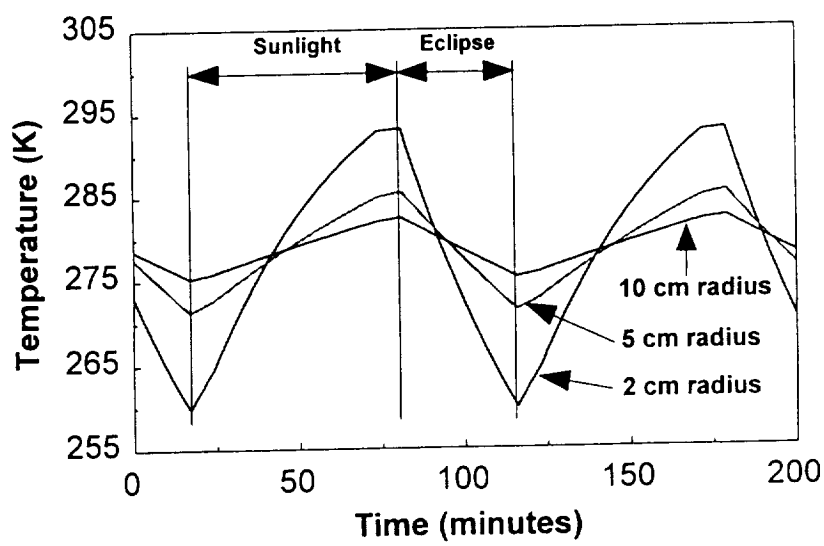


Figure 6: Spacecraft temperature as a function of time for a solid spherical silicon spacecraft covered completely by solar cells with absorptivity of 0.8 and emissivity of 0.83. These temperatures were calculated using a lumped heat-capacity model assuming a circular orbit at an altitude of 700 km.

Figure 6 indicates that passive thermal control is possible for nearly spherical nanosatellites and microsatellites. When dimensions drop below 2 cm, the temperature extremes exceed typical electronics and battery limits. Femtosatellites, with their extremely low mass, can reach the equilibrium sunlight (or eclipse) temperature within minutes.

II.E. Orbit Lifetime Considerations

The Earth's atmosphere effects spacecraft motion even at altitudes beyond 1000 km. The main effects are atomic oxygen erosion and orbital decay due to atmospheric drag. The drag force F_D on a satellite is given by

$$F_D = \frac{1}{2}\rho V^2 S C_D \tag{3}$$

where ρ is the local atmospheric density, V is the satellite velocity, S is effective cross-sectional area of the spacecraft, and C_D is the satellite drag coefficient (C_D~2 for most satellites). Orbit decay rates are often parameterized by introducing the ballistic coefficient W, defined as the satellite mass M divided by the cross-sectional area S times the drag coefficient C_D. A spherical solid silicon satellite with a diameter of 10 cm would have a ballistic coefficient of ~150 kg/m^2. Figure 7 shows the maximum ballistic coefficient as a function of mass for a cubical satellite and a 10-micron-thick sheet satellite. Note how the ballistic coefficient is constant (and extremely small!) for the thin sheet satellite while it is a function of mass for a cubic satellite.

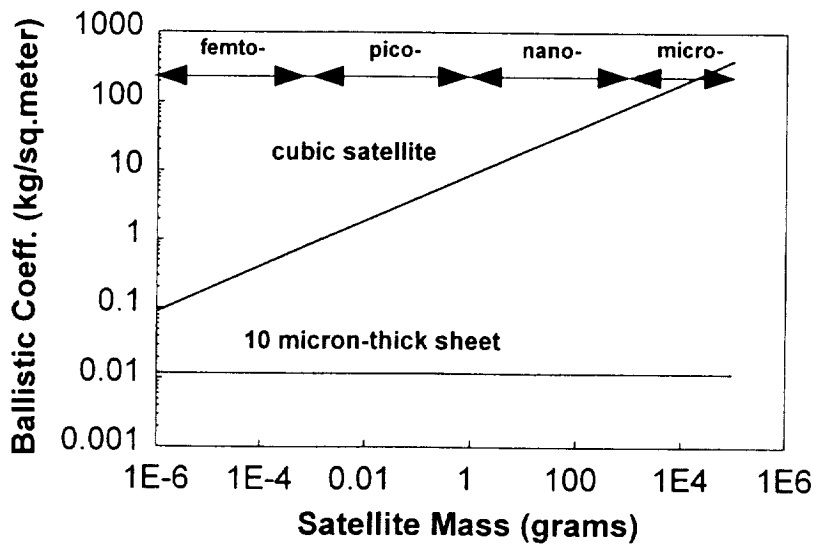


Figure 7. Maximum ballistic coefficient as a function of mass for solid silicon satellites.

Figure 8 shows calculated orbital decay rates as a function of ballistic coefficient and altitude for worst-case solar-maximum conditions (4 to 6 years from now). Currently, we are in solar-minimum conditions, which produces decay rates about an order-of-magnitude lower at 300 km altitude and about 3 orders-of-magnitude lower at 700 km altitude. For satellites with ballistic coefficients of ~100 kg/m^2, note that at 700 km altitude the solar-maximum orbit decay rate is only ~10 km per year, while at 500 km altitude the decay rate is high enough to produce an orbital lifetime of only ~1 1/2 years. Roughly spherical (or cubical) picosatellites will have worst-case orbit decay rates of 100 to 1000 km per year even at an altitude of 700 km. For a 500 km altitude, orbit lifetimes would be from a few days to a few months. If orbit altitudes below 500 km and/or mission lifetimes of

greater than a few years are required, drag make-up propulsion for femtosatellites and picosatellites will be mandatory.

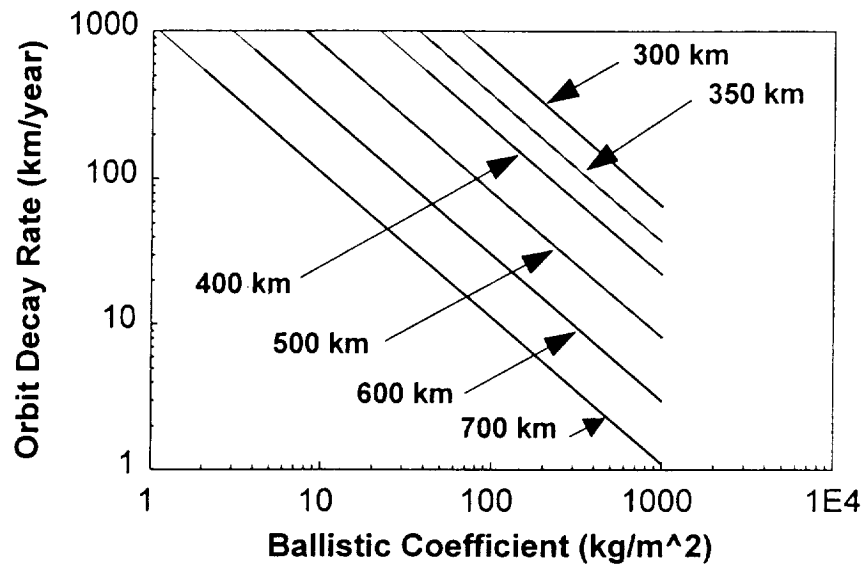


Fig. 8. Orbit decay rates as a function of ballistic coefficient and altitude for circular orbits under very active solar conditions.

III. Conclusions

III.A. Femtosatellites and Picosatellites

Femtosatellites don't have enough radiation shielding or a high enough ballistic coefficient to survive more than a week on-orbit. At altitudes below 500 km where radiation shielding (0.38 mm maximum for a 1 milligram cubic femtosatellite) may be adequate for radiation-hardened electronics, the high ballistic coefficient limits lifetimes to a few days. At higher altitudes, rapidly increasing radiation levels also limit lifetime to a few days. Femtosatellites should be nearly spherical in shape to minimize air drag and maximize radiation shielding. Maximum power generation levels will therefore be in the sub-milliwatt range. Active femtosatellites are an extremely difficult challenge due to their low thermal mass and wild temperature swings.

Picosatellites are the smallest useful satellites, but active thermal control will be required. A thermally passive picosatellite will have temperature swings of 90 K between sunlight and eclipse in low Earth orbit. Cubic picosatellites can have as much as 0.18 cm silicon radiation shielding and a ballistic coefficient of ~9 kg/m^2. Orbit lifetimes can be several years at 700 km altitude under solar-maximum conditions and several years at 500 km under solar-minimum conditions. Nearly spherical satellites are needed to provide radiation shielding, and if low-inclination orbits are used (below 700 km altitude), radiation-soft CMOS electronics may be feasible. Power outputs will be in the 10's of milliwatts range. Picosatellites may be good for disposable or short-duration (i.e. 1-week) missions.

III.B. Nanosatellites

Nanosatellites are the smallest satellites that don't require active thermal control. Silicon radiation shielding thicknesses greater than 1 cm are possible with power outputs in the Watt range. Nearly spherical nanosatellites can operate for several years at altitudes below 500 km in LEO due to their modest ballistic coefficients, and in GEO due to increased radiation shielding. Flattened nanosatellites, i.e. 2 cm thick disks, can produce several Watts of solar power and still retain good thermal control, radiation shielding, and modest ballistic coefficient.

III.B. Microsatellites

Microsatellites have the best power, radiation shielding, orbit lifetime, and thermal characteristics. Flattened silicon microsatellites offer mission lifetimes of years or decades with power levels of 10 to 100 Watts.

Acknowledgments

This work was supported by the Technology Development and Applications Directorate of the Aerospace Corporation. I gratefully acknowledge their support and for previous support provided by the Aerospace Sponsored Research Program.

References:

1. Kurt E. Petersen, "Silicon as a Mechanical Material," Proceedings of the IEEE, **70** #5, p.420-427, May 1982.
2. ASM Specialty Handbook: Aluminum and Aluminum Alloys, edited by J.R. Davis and Associates, p. 68 - 73, ASM International, Materials Park, Ohio, May 1994.
3. ASM Specialty Handbook: Stainless Steels, edited by J.R. Davis and Associates, p. 7-10, ASM International, Materials Park, Ohio, Dec. 1994.
4. Metals Handbook Volume 2: Nonferrous Alloys and Special-Purpose Materials, p. 620-622, ASM International, Materials Park, Ohio, Oct. 1990.
5. ASM Engineering Materials Reference Book; Second Edition, edited by Michael Bauccio, p. 281-282, ASM International, Materials Park, Ohio, March 1995.
6. L.C. Rogers, "Chapter 2: Polysilicon Preparation" in Handbook of Semiconductor Silicon Technology, edited by W.C. O'Mara et al., Noyes Publications, Park Ridge, NJ, 1990.
7. Ron Schneiderman, "Multichip Modules Gain Ground at Higher Frequencies," Microwaves and RF, p.49-50, May 1993.
8. Direct measurement on UV-erasable unit (quartz window shows dice size)
9. Peter H. Singer, "PowerPC 604 in Production," Semiconductor International, p. 17, **17**, #11, October 1994.
10. Linda Geppert, "Technology 1996: Solid State", IEEE Spectrum, **33** #1, p. 51 - 55, January 1996.
11. Michael Feibus and Michael Slater, "Pentium Power," PC Magazine, p. 108-120, **12** #8, April 1993.
12. Charles Huang, "GaAs MMIC Power Amplifiers Drive Cellular Systems," Microwaves and RF, p. 85-86, October 1994.
13. Asad A. Abidi, "Low-Power Radio-Frequency IC's for Portable Communications," Proceedings of the IEEE, **83**, #4, p. 544-569, April 1995.

14. Jonathan Min et al., "An All-CMOS Architecture for a Low-Power Frequency-Hopped 900 MHz Spread-Spectrum Transceiver," Proceedings of the IEEE Custom Integrated Circuits Conference, p. 379-382, San Diego, CA, May 1-4, 1994.

15. Susan Reinecke Taub and Samuel A. Alterovitz, "Silicon Technologies Adjust to RF Applications," p.60-74, Microwaves and RF, **33**, #10, October 1994.

16. Johann-Friedrich Luy et al., "Si/SiGe MMIC's," p. 705-714, IEEE Transactions on Microwave Theory and Techniques, **43**, #4, Aprl 1995.

17. G. Dawe et al., "SiGe Technology: Application to Wireless Digital Communications," p. 14-23, Applied Microwave and Wireless, Summer 1994.

18. Robert D. Rasmussen, "Spacecraft Electronics Design for Radiation Tolerance," Proceedings of the IEEE, **76** #11, p. 1527-1537, November 1988.

19. Michael D. Griffin and James R. French, Space Vehicle Design, p. 70, American Institute of Aeronautics and Astronautics, Washington D.C, 1991.

20. Michele M. Gates, Mark J. Lewis, and William Atwell, "Rapid Method of Calculating the Orbital Radiation Environment," Journal of Spacecraft and Rockets, **29** #5, p. 646-652, Sept.-Oct. 1992.

21. J.B. Angell et al., "Silicon Micromechanical Devices," *Scientific American*, p.44-55,**248**, #4, April 1983.

22. R.T. Howe et al., "Silicon Micromechanics: Sensors and Actuators on a Chip," *IEEE Spectrum*, p.29-35, **27**, July 1990.

23. K. Wise and K. Najafi, "Microfabrication Techniques for Integrated Sensors and Microsystems," *Science*, p. 1335-1342, **254**, 29 Nov. 1991.

24. M. Mehregany, "Microelectromechanical Systems," *IEEE Circuits & Devices*, p.14-22, **9**, July 1993.

25. S.W. Janson, "Chemical and Electric Micropropulsion Concepts for Nanosatellites," AIAA Paper 94-2998, 30th AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Indianapolis, IN, June 1994.

26. S.W. Janson, "Spacecraft As an Assembly of ASIMs," p. 181-258 in *Microengineering Technology for Space Systems*, edited by H. Helvajian, Aerospace Technical Report ATR-95(8168)-2, The Aerospace Corporation, El Segundo, CA, Sept. 1995.

27. S.W. Janson, "Mini-, Micro-, and Nanosatellite Concepts," p. 67 - 75, in *Micro- and Nanotechnology for Space Systems: An Initial Evaluation*, edited by H. Helvajian and E. Robinson, Aerospace Technical Report ATR-93(8349)-1, The Aerospace Corporation, El Segundo, CA, March 1993.

28. S.W. Janson, H. Helvajian, and E.Y. Robinson, "The Concept of Nanosatellite for Revolutionary Low-Cost Space Systems," Paper IAF-93-U.5.573, 44th Congress of the International Astronautics Federation, Graz Austria, October 1993.

29. David G. Galimore, "Satellite Thermal Environments," Chapter 2 in Satellite Thermal Control Handbook, edited by D.G. Gilmore, The Aerospace Corporation Press, El Segundo, CA, 1994.

IMPLICATIONS OF GUN LAUNCH TO SPACE FOR NANOSATELLITE ARCHITECTURES

7/080

030586

6P.

Miles R. Palmer
Science Applications International Corporation
1710 Goodridge Drive, McLean, Virginia 22102
(703) 749-5143

Abstract

Engineering and economic scaling factors for Gun Launch to Space (GLTS) systems are compared to conventional rocket launch systems. It is argued that GLTS might reduce the cost of small satellite development and launch in the mid to far term, thereby inducing a shift away from large centralized geosynchronous communications satellite systems to small proliferated low earth orbit systems.

INTRODUCTION

The present space launch industry is oriented toward large geosynchronous satellites due to favorable scaling of rocket launch and satellite costs with size. The technical and economic factors behind this scaling will be explored and contrasted to the potential scaling of GLTS systems. It will be shown that GLTS systems appear to yield a maximum return on investment for 100-1000 kg satellite payloads.

The microprocessor induced a shift in the computer industry from mainframes to small computers by enabling small computers to become cost effective. Because it enables small payloads to become cost effective, GLTS might induce such a shift in communications satellite architectures.

ECONOMICS OF PRESENT LAUNCH SYSTEMS

The commercial space launch industry is oriented toward ever larger geosynchronous satellites (Agrawal 1986). The reason for this is shown in the return on investment model data of Figure 1, which shows that economic payoffs increase strongly with satellite size (Agrawal 1986). This trend derives from the fact that both launch and satellite prices per kilogram decline steeply with satellite size. Since revenues are almost linear with satellite mass, the return on investment increases strongly with satellite mass.

For Figure 1, launch costs are assumed as $\$550{,}000M^{-.45}$ per kg and satellite costs are assumed as $10 million for design plus $\$840{,}000M^{-.25}$ per kg. Revenues are assumed to be $15,000 per kg of satellite per year excluding 50 kg of satellite weight for parasitic mass. Revenues are calculated as a simple fraction of initial investment costs for the satellite and launch. M is the satellite or payload mass in kg.

Historical launch prices per kg to LEO are plotted in Figure 2 versus payload size (Isakowitz 1991). Although there is a great deal of scatter due to differences in accounting methods and government subsidies, there is clearly a strong downward trend. Nonlinear regression analysis indicate that the best geometric fit to this data is for a launch price per kg proportional to the -0.45 power of mass with a launch cost at one kilogram of $540,000.

Satellite prices show a similar behavior as illustrated in the data of Figure 3 derived from an Aerospace Corporation cost model (Lenard, 1990). A major basis for both these trends is that many of the same functions are necessary when designing, fabricating, launching, and operating launch vehicles and satellites, semi-independent of their size. The larger the satellite or launch vehicle, the better these costs are amortized. Most of these costs are also better amortized with higher launch rates.

Another factor which strongly influences launch prices is useful payload fraction. Larger vehicles such as the space shuttle or a Titan-4 have larger payload fractions to orbit that small vehicles such as the Scout (Isakowitz 1991). A vehicle which launches more payload for a given vehicle size should be able to deliver that payload at a lower price since the vehicle cost component of the price would be lower.

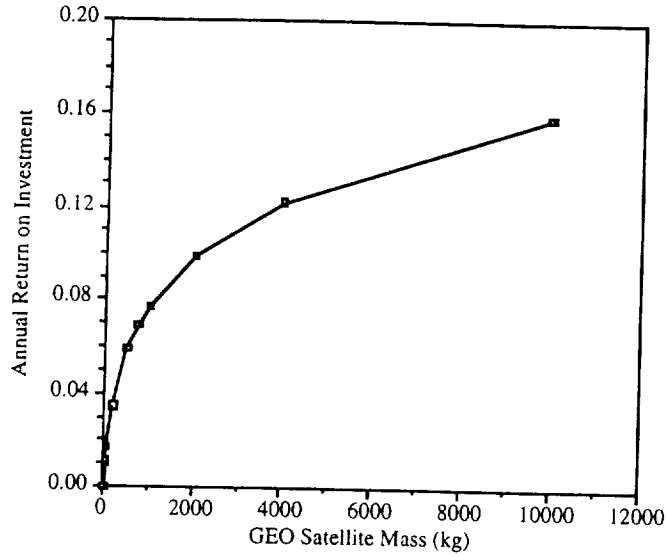


FIGURE 1. Return on Investment Model for Present Satellite Communications Systems

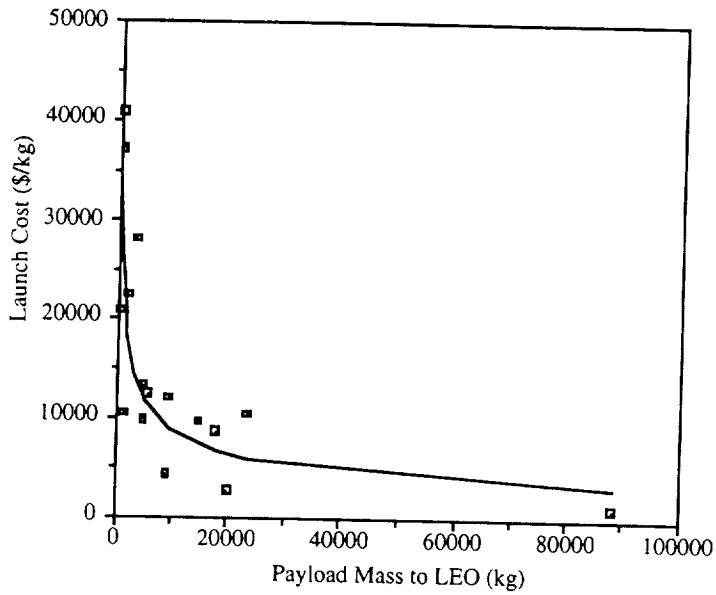


FIGURE 2. Historical Launch Costs versus Payload Mass (Isakowitz, 1991)

The effect of vehicle size on payload fraction is most easily seen in performance models for single stage to orbit (SSTO) vehicles (Hampsten 1994). Figure 4 shows a predicted payload fraction for a SSTO as a function of vehicle size. Even quite large SSTO vehicles (1000-2000 tons) have payload fractions well below the 3-5% typical for large multistage vehicles (Isakowitz 1991). A SSTO vehicle must be quite large to be cost effective, since payload fractions much below 1% would probably not be cost competitive in operation (Hampsten 1994).

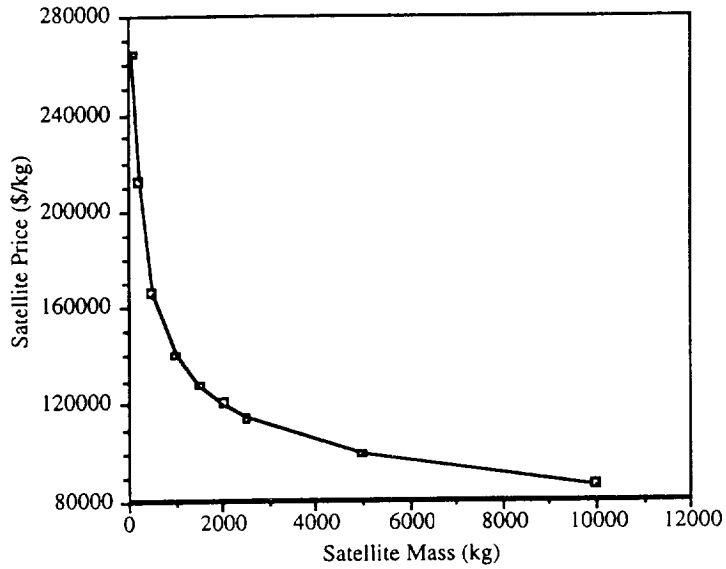


FIGURE 3. Model for Dependence of Satellite Prices on Mass

Development of such large vehicles is very expensive. Development costs for a 1000 ton class SSTO have been estimated at $6B-$37B (Aviation Week 1994) (Space News,1994). Such large investments are difficult to motivate in either the government or commercial sectors.

Development of much smaller, less costly vehicles would be much easier to initiate, but cost ineffective and technically difficult due to the vanishingly low payload fractions achievable with very small vehicles. Resorting to a totally new technological approach such as gun launch to space has the potential to resolve this dilemma.

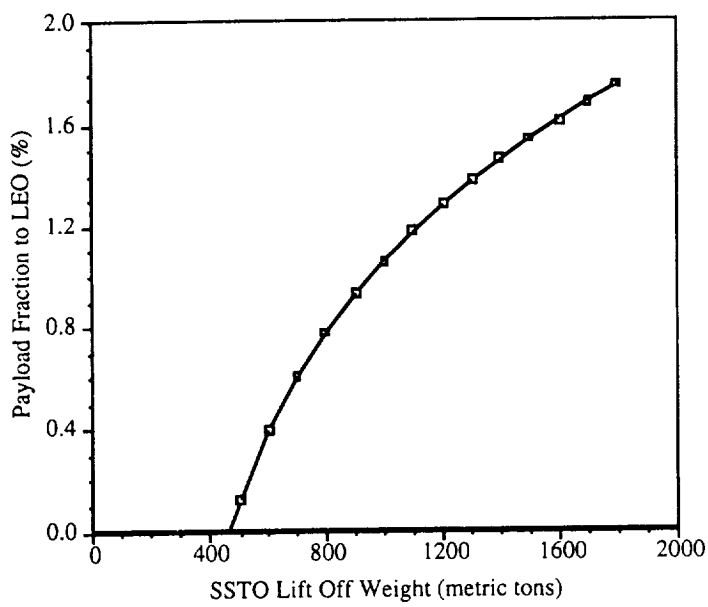


FIGURE 4. Effect of Rocket Launch Vehicle Scale Size on Payload Fraction.

Small launch vehicles have small payload fractions for two reason. First, the proportion of parasitic masses such as valves, tanks, etc., grows as vehicle size diminishes. Second, aerodynamic losses and total velocity increment to orbit grow rapidly with the vehicle surface area to volume ratio.

ECONOMICS OF GUN LAUNCH SYSTEMS

Gun launch systems largely avoid both these problems since the propulsion system parasitic mass and fuel are mostly not accelerated with the payload, remaining in the rest frame of the launcher. Since the propulsion system and fuel are not accelerated, the gun can achieve high effective payload fractions even for tiny payloads. Since bulky fuel is not carried with the payload, the payload is physically much smaller, and aerodynamic losses are greatly reduced.

A gun launched payload does have significant parasitic masses for sabot, armature, structural reinforcement, and thermal shielding. These parasitic masses cannot be determined precisely without rigorous engineering, but estimates of 10-40% have been made (Castle, 1990). If the gun provides almost the entire velocity increment necessary to achieve orbit, this translates to payload fractions to orbit of 60-90%.

A good means of comparing the payload fraction achieved with guns with that of rockets is to calculate an effective specific impulse for gun launch analogous to that for a rocket stage (1). This allows direct comparison of the payload mass fraction obtainable from a gun to that obtainable from a rocket.

Gun effective specific impulse = $I_{SPe} = -V/g_0 \mathrm{Log}_e (M_P/M_L)$ (1)

V = gun launch velocity
g_0 = acceleration of gravity
M_P = payload mass = M_L - parasitic masses
M_L = gun launch accelerated mass (does not include gun propellant)
λ = mass efficiency ratio = M_P/M_L

Figure 5 shows effective specific impulse as a function of gun velocity for several ratios of payload to total launch mass. Conventional military powder guns launching standard shells at 800-1800 m/sec attain mass efficiency ratios, λ, of 0.7-0.9. These guns demonstrate effective specific impulses in the 500-2000 second range, yielding much higher mass efficiency ratios than could be obtained with rockets.

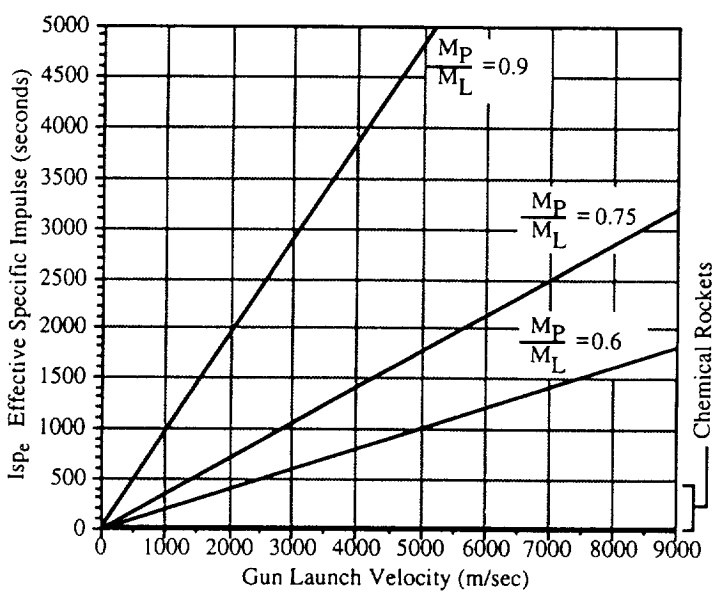


FIGURE 5. Effective Specific Impulses Achievable with Gun Launch Systems

Unfortunately, conventional powder guns cannot efficiently produce velocities over about 2000 m/sec. Launch to orbit requires much higher velocities, at least 7900 m/sec orbital velocity, plus another 500-2000 m/sec for aerodynamic losses, plus small gravity losses.

Research and development has been conducted on many gun technologies over the years. Figure 6 shows a taxonomy of the various gun types which have been explored (Palmer, 1991).

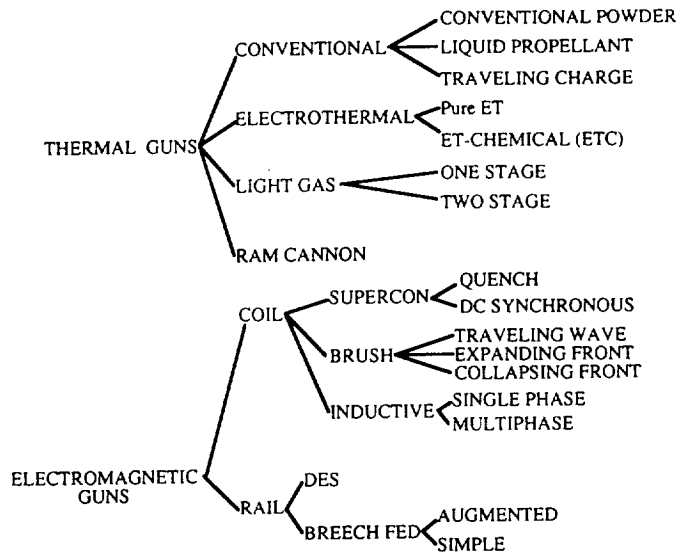


FIGURE 6. Taxonomy of Guns Types

Thermal guns are relatively low in cost to develop, but have thus far had practical upper velocity limits of 2000-4000 m/sec. Such a gun would require an additional 4500-8000 m/sec rocket boost to achieve orbit, and the booster would have to be designed to withstand high accelerations. At the 4500 m/sec end of this range, such a rocket might achieve orbit with a single stage less expensively than conventional small rockets, which require three or four stages.

Electromagnetic guns have been much less well developed due to the very high cost of the electrical power supplies necessary to drive them. If the satellite payload mass is assumed to be 50% of the gun launched mass, and the electromagnetic launcher is assumed to be 30% efficient at a launch velocity of 9000 m/sec, an electromagnetic launcher to propel a 1000 kg payload into orbit might require a capacitor power supply costing over $12B. Capacitor costs are assumed at a $0.63/Joule first unit cost with a learning curve cost factor exponent of 0.9.

Velocities over 7 km/sec have been achieved with electromagnetic railguns. However, these velocities have been obtained for very small masses at impractically low efficiencies. Development of higher efficiency launchers and lower cost power supplies will be necessary to achieve practical GLTS. Steady progress has been achieved in both areas in the last several years.

Assuming that an electromagnetic launcher could be developed with 30% efficiency at 9000 m/sec, and that power supply energy storage costs could be reduced to $0.001 per joule, a highly cost effective launch system would result. Such a system could launch many payloads at low cost and, over time, eliminate present high cost satellite design and construction practices.

Assuming that satellites would eventually be designed and constructed at prices similar to those for tactical missiles, the economics of satellite communications would be revolutionized. Figure 7 shows model data for the return on investment as a function of satellite size in such a scenario.

For Figure 7, a 100,000 kg total on orbit mass is assumed for the satellites. The number of satellites would then equal 100,000 divided by the satellite mass. Satellite costs are assumed as $10M for design plus $840,000$M^{-.25}$ per kg for the first satellite plus $10,000$M^{-.15}$ for all additional satellites. Revenues are assumed as $15,000 per kg of useful satellite mass per year with the useful mass fraction calculated as $0.3M^{-1}$. Launcher construction costs are assumed as $8,000,000$M^{-.3}$ per kg of launched mass. Launcher operations costs are estimated at 25% of launcher construction cost plus $100,000 per launch. Revenues are calculated as a simple fraction of initial investment costs of satellite and launcher. M is the satellite or payload mass in kg.

The returns on investment would be very high at current spaceborne communications prices, probably resulting in a dramatic reduction in space communications prices world wide. At some point, space communications could become lower in price than terrestrial wireless or even land line communications.

An important feature of Figure 7 is that return on investment is optimal for satellites in the range of 100-1000 kg. This is due to the low functional mass fraction of very small satellites and the high cost of very large gun launch systems. This feature of GLTS could greatly reduce initial investment costs and incentivize development of small launchers capably of cheaply launching small, mass produced communication satellites.

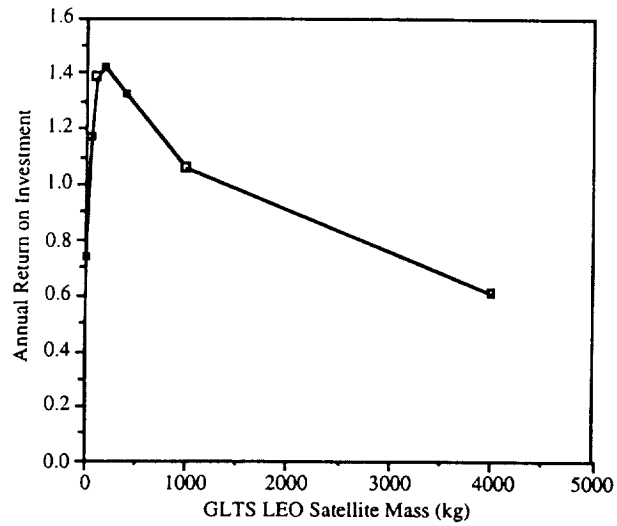


FIGURE 7. Return on Investment for Gun Launched Low Earth Orbit Satellite Systems

References

Agrawal, B., (1986)*Design of Geosynchronous Spacecraft*, Prentice Hall, Englewood Cliffs, New Jersey.

Aviation Week, (1994), "Access to Space", March 14, 1994, p. 32.

Castle, M., (1990), Personal Communication, GE Re-entry Systems, Philadelphia, Pennsylvania, 1990.

Hampsten, K., (1994), Personal Communication, Launch Vehicle Office, USAF, Albuquerque, NM, January, 1994.

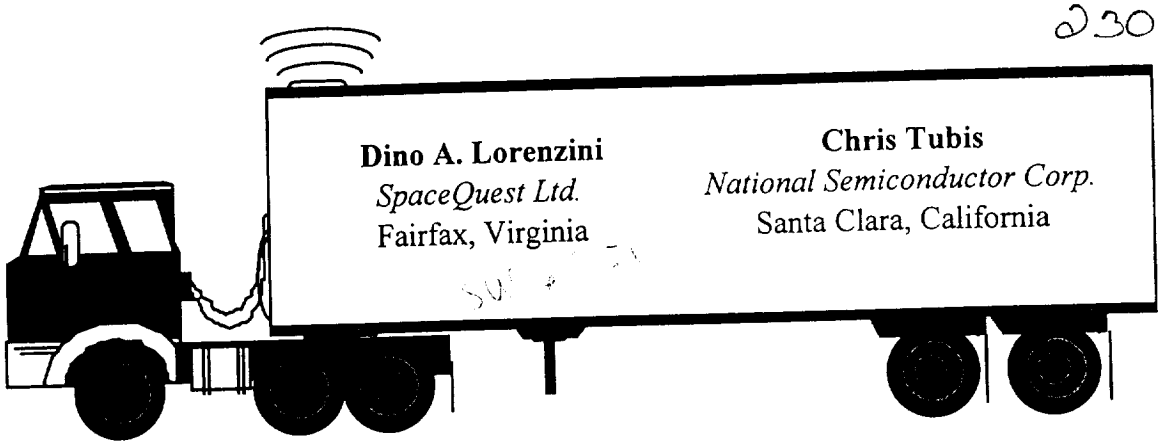
Isakowitz, S. J., (1991) *International Reference Guide to Space Launch Systems*, AIAA, Washington, D.C.

Lenard, R., (1990), Personal Communication, SDIO, The Pentagon, Washington, D.C., March, 1990.

Palmer, M. (1991), "A Revolution in Access to Space", IEEE Trans Mag, 27(1), January 1991, p. 11.

Space News, (1994), "Department of Defense Space Launch Modernization Report", May 16, 1994, p. 3.

VEHICLE TRACKING SYSTEM USING NANOTECHNOLOGY SATELLITES AND TAGS



Dino A. Lorenzini
SpaceQuest Ltd.
Fairfax, Virginia

Chris Tubis
National Semiconductor Corp.
Santa Clara, California

ABSTRACT

This paper describes a joint project by SpaceQuest Ltd. and National Semiconductor Corporation to design, develop and deploy a satellite-based tracking system incorporating micro-nanotechnology components. The system consists of a constellation of "Nanosats," a satellite command station and data collection sites, and a large number of low-cost electronic "tags." Two prototype Nanosats are currently under construction by SpaceQuest for launch on a Russian booster in July, 1996. The miniature tags are being developed by National Semiconductor using advanced micro-nanoelectronic components.

Both government and commercial applications are envisioned for the satellite-based tracking system. Government users are interested in keeping track of high value assets, including military hardware, trucks, containers, and high-priority shipments. Commercial users are interested in tracking the location of trailers, intermodal containers, fishing vessels, and high-value cargo. The projected low price for the tracking services is made possible by the lightweight Nanosats and inexpensive tags which use high production volume single chip transceivers and microprocessor devices.

The NanoSat structure consists of a five-inch aluminum cube with body-mounted solar panels on all six faces. A UHF turnstile antenna and a simple, spring-release separation mechanism complete the external configuration of the spacecraft. The Nanosat uses a passive, magnetic stabilization system without orbital station-keeping. High efficiency, low power consumption electronics are used to conserve the one watts of average orbital power generated by the Nanosat solar panels. Additional energy conservation features are incorporated into the microprocessor electronics.

The GaAs solar cells will average 2.4 watts of power between sunlight and eclipse. Total power consumption for all the spacecraft bus functions, including one fixed-frequency and one agile receiver, is expected to be less than one watt, leaving 1.4 watts to operate the satellite transmitter. A single transmitter with variable data rates up to 9,600 bps and adjustable levels up to 7 watts or RF power is used for polling the tags, downloading collected data, and transmitting satellite telemetry.

A V-53A microprocessor with two megabytes of random access memory serves as the communications control center and data storage device. The position information collected from the "tagged" items are stored on board the Nanosat for later transmission to a data collection site.

At a projected selling price of less than $100, the low-cost tags being developed by National Semiconductor can be considered disposable for many applications. The high-performance, frequency-controlled, single-chip transceivers using BiCMOS technology have extremely low power consumption. High levels of silicon integration are used to achieve a low manufacturing cost at high production volumes. Computer-controlled circuits and functions keep the power consumption of the unit low so that a small, integrated, lithium battery can be used effectively for many of the intended applications.

A prototype system consisting of two Nanosats and several hundred tags should achieve its initial operational capability sometime during 1997.

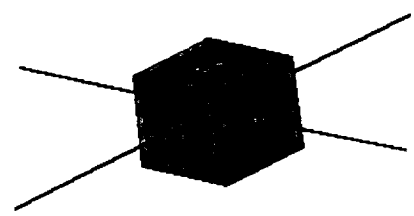
1. INTRODUCTION

The idea of putting a softball-size satellite into orbit was suggested by a government client. Our previous experience with small satellite designs was the construction and launch of a 10 kg, 9-inch cube in September, 1993. This commercial Low Earth Orbit Satellite (LeoSat) has multiple receivers, modems and transmitters. Its intended purpose was to validate the feasibility of using a small, inexpensive microsatellite to communicate directly with an inexpensive tag having a low gain antenna. Several satellites of this 10 kg class are currently in orbit performing useful data communications missions.

The advantages of a reduced spacecraft size are lower cost, lower launch weight (thus, launch costs), lower power consumption, and a shorter integration and test time. Reducing the size of the satellite electronics is a straight forward process using commercial grade microelectronic components. However, the physical size of the satellite's surface area limits the amount of solar energy that can be collected with body-mounted solar panels, thus restricting the number and types of missions that can be performed by a Nanosat. Nevertheless, several specialized applications have been identified where a miniature, low power Nanosat can provide useful space communications services.

2. NANOSAT CONCEPT

Rational



Nanosats smaller than a 6-inch cube are a natural evolution to the generation of microsats that have already proven their worth in space. Small size does not necessarily translate into the equivalent of reduced capability. Used in combination for special-purpose applications, Nanosat can provide benefits well beyond their small size for specialized mission applications. Because of their relatively low cost, these space devices can be manufactured in quantity and stored for immediate launch when needed. Their small size, rigid structure, simple separation device, and lack of deployable, explosives or fuels, simplifies the task of finding a compatible piggyback launch vehicle, provided that the required orbit is not critical to mission success. Also, Nanosats can be launched in clusters of 24 or more spacecraft for a total system cost of $10 million or less. Although Nanosats cannot do everything, they can be a cost-effective solution for filling an important niche in future space requirements for both government and commercial users.

System Architecture

Because of the low cost of producing Nanosats in quantity, and the availability of piggyback launches, it is economically feasible to tradeoff performance for quantity. For example, total system communication capacity can be achieved by launching many Nanosats, rather than only a few larger satellites. Proliferating a single orbital plane with 20 to 30 randomly-spaced Nanosats becomes a feasible alternative to using 6 to 8 Microsats with station-keeping capabilities.

Simplicity of operation, highly-specialized functions, and self-monitoring features can dramatically reduce the amount of ground command and control required to operate a Nanosat. Automated network control procedures to download mission data and upload revisions to the tag polling schedule keep operating costs to a minimum, despite the large number of Nanosats that may comprise the space segment of a satellite-based vehicle tracking system.

A typical Nanosat communications system would consist of 2 to 100 spacecraft in orbit, a Network Operations Center to coordinate all message traffic, several regional data collection sites to service international users effectively, and tens of thousands of tags distributed worldwide. A real-time data relay mode can be implemented on a region-by-region basis where both the sender and receiver are in a common Nanosat footprint (approximately 5,000 km diameter). For transferring messages globally, a store-and-forward mode can be used as illustrated in Figure 1 below. In this case, a message or data file is collected by a Nanosat and stored in the flight computer memory for later distribution to the intended receiving station. Two megabytes of on-board data storage is sufficient for most Nanosat applications.

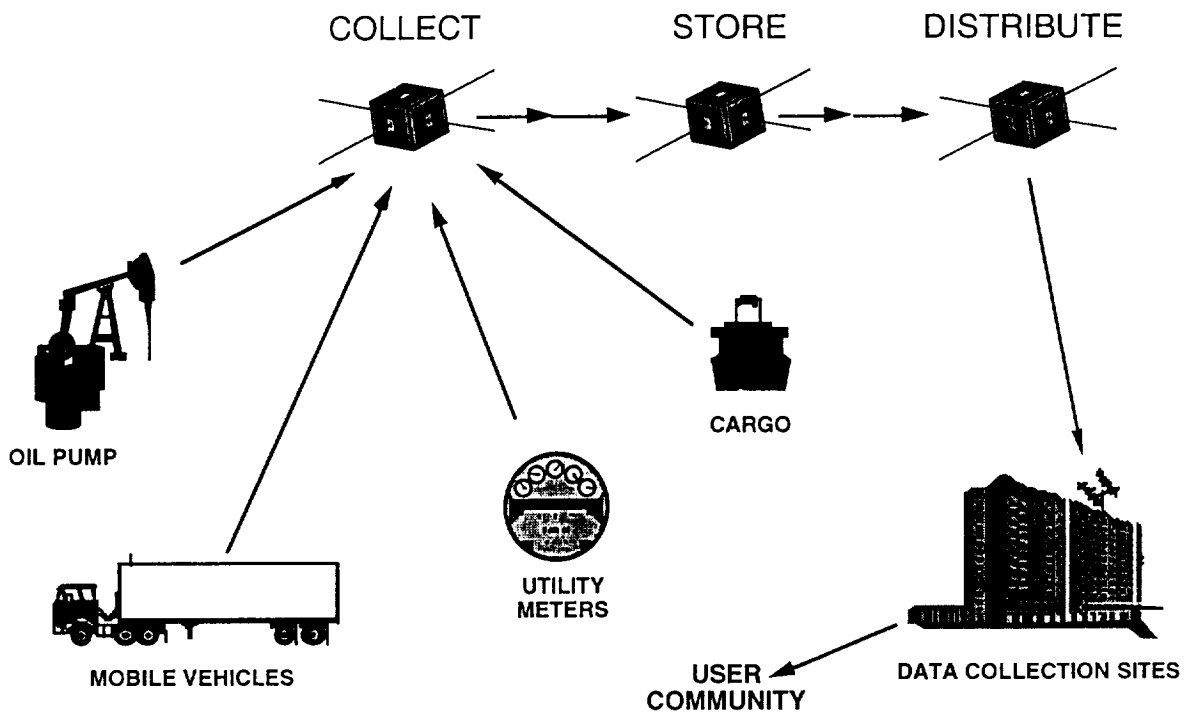


Figure 1. Nanosat Store-And-Forward Concept

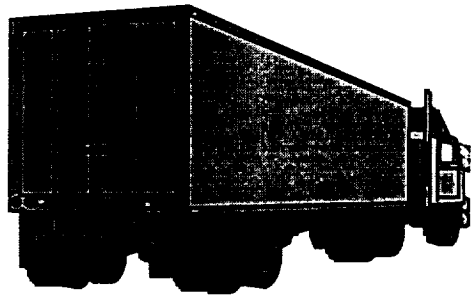
3. NANOSAT APPLICATIONS

Because of the limited power available for a Nanosat transmitter, the selection of potential applications favors those that maximize the satellite's receiver operation and minimizes the satellite's transmitter operation. Thus, those applications where the user desires to "pick up" data from many small ground transmitters distributed worldwide, and to "deliver" that data to a few ground stations equipped with a large tracking antenna, are best suited for the use of Nanosats. Only limited polling by the satellite is needed to coordinate the tag transmission activities. Among the possible applications for Nanosats are: (1) reporting the location of mobile vehicles or high-value assets, (2) transferring critical environmental data from remote monitoring stations to a central location, and (3) reading hard-to-access gas and electric meters.

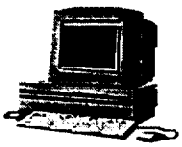
Reporting Mobile Vehicle Location

On any given day aboard ships, trains, and trucks around the world, some five million shipping containers are on the move, packed with textiles, appliances, computer parts, and every other kind of merchandise. Most shipments arrive safely, but some fall prey to the hazards that lurk between factory and market. The goods show up smashed, baked, soaking wet, or don't show up at all. Manufacturers and insurance companies try to keep tabs on shipments and protect them from harm, but it is hard to know precisely where a container is at any moment during its journey, let alone what is going on inside that container.

No one knows exactly how much cargo is lost or stolen, but in the United States alone, nearly 10 percent of all cargo shipped is lost or delayed in transit due to misrouting. Losses due to theft and vandalism account for an additional 6 to 8 percent. These problems are faced every day by transporters of all varieties of cargo. A shipping order may indicate where the ship, truck, or train is supposed to be, but how can it be certain that the container is where it should be? Was it shipped at the right time? Was it sent to the right place? Using a Nanosat tracking system, these questions are much easier to answer, by telling the user exactly where his cargo is, and what condition it is in.



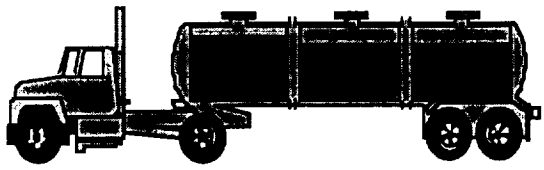
By discreetly mounting a small GPS-equipped transmitter tag, a vehicle can be tracked when stolen or reported missing. Using an automatic activation scheme or satellite polling, the tag can be initialized and begin transmitting its position. It is then a simple matter to track down the missing vehicle. This system can be used by rental companies to keep track of large vehicle fleets, or by government agencies to manage their mobile resources.



With a small electronic tag attached to each container of a shipping company's fleet, the location and status of every container can be determined at nearly any time of the day. A shipping manager can track the progress of an important shipment right on a computer screen, showing not only where the shipment is, but what its condition is, and whether it has been tampered with.

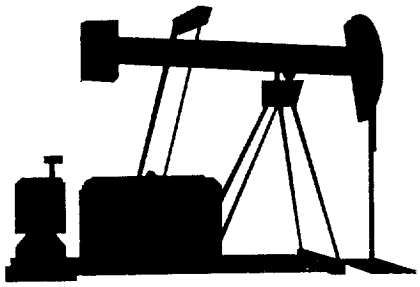
The problem of losing cargo becomes even more significant when one considers the number of vehicles that contain hazardous cargo. The safety risk posed by the loss of a toxic or flammable chemical container is much greater than the monetary loss. Tracking the location and status of hazardous materials is essential for reducing the dangers that these cargoes present.

Using a GPS receiver integrated into a mobile transmitter tag, a Nanosat system can track cargo in near-real time. If the hazardous cargo is lost, enters a high-risk area, or has its schedule interrupted, the operators know it immediately and may act to reduce the risk of an environmental disaster. Additionally, with increasing public and government concern for safer transportation of hazardous cargo, a Nanosat system can provide a solution for better risk management.



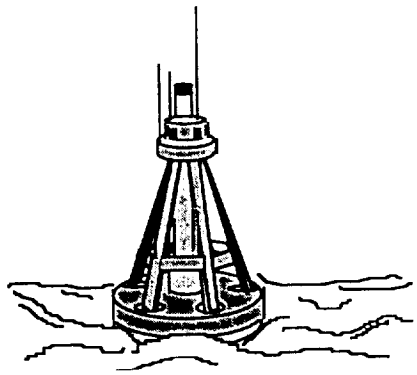
Transferring Remote Environmental Data

This application focuses on the active monitoring of natural resources and man-made facilities that affect the environment. In recent years, Federal, State, and local governments have strengthened regulations requiring the monitoring of air and water quality and pollution levels. The location of many facilities prohibits constant monitoring by traditional methods. The use of remote transmitters and a Nanosat data relay system can provide an effective and cost efficient means for monitoring these facilities. For example, placing terminals along an oil or gas pipeline to measure corrosion, pressure and flow rate, or at an unmanned reservoir dam to measure water level, allows the status of the site to be relayed to a control center several times each day, providing advanced notice of impending problems.



The requirements are widespread for global environmental remote-sensing to support the scientific community. Ground and sea surface-based environmental data can be of use in understanding the nature of Earth's changing environment. The availability of distant sensing capability through the use of remotely-located tags can enhance the effectiveness of research in this area, offering wide coverage and timely data retrieval. Environmental applications suitable for a Nanosat system include monitoring such variables as:

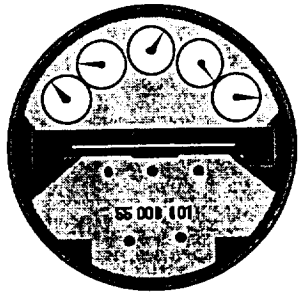
- air or ground temperature
- atmospheric pressure
- ocean conditions
- snow pack levels
- iceberg movement
- river or reservoir level
- air or water pollution conditions
- seismic or volcanic activity
- oil spill movement



Reading Gas And Electric Meters

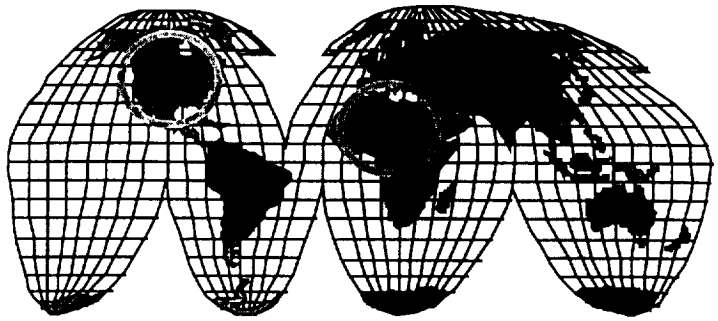
Another application for a Nanosat system includes the gas and electric utility industry, with a specific focus on the hard-to-access meter locations. As the world economies develop, the need for accurate and timely data and information is becoming increasingly important. Industries require communications systems which constrain cost while contributing to productivity gains.

There are approximately 120 million electric meter locations in the United States. Of these 15-20% are considered remote and hard-to-access. These meters represent the most costly metering sites for utility companies. The meters provide no information regarding the quality of service being provided to the customer, nor do they allow the utility company to detect tampering or abnormal usage patterns associated with theft of energy. In addition, the only way that traditional meters can be read or electrical service connected or disconnected is through a personal visit by a meter reader or customer service representative.



A Nanosat system can provide the means to monitor and record a large number of electrical service parameters by providing the communications network over which such information can be retrieved and communicated to any number of customers. The estimated size of the addressable U.S. metering market is approximately 17 million meters.

4. SYSTEM CONCEPT



The concept of operation for the Nanosat system is that the user contracts to receive cargo status and tracking information services for his specific application. The user is provided with a tag, which is installed at a given site or on a mobile asset. Once activated, the tags transmits information (which may include GPS determined position, local temperature, water or snow level, short messages, or any other applicable transducer output) during specific periods each day. The information is received by one or more orbiting Nanosat, and re-transmitted to one of the data collection sites for processing. The processed data is then made available to the user through a variety of electronic channels.

The overall Nanosat system is comprised of four interrelated segments: (1) Space Segment (Nanosat constellation), (2) User Segment (tags and user software), (3) Control Segment (command stations and data collection sites), and (4) Launch Segment (Russian booster rocket).

5. SPACE SEGMENT

Nanosat Requirements

Due to its small physical size, a Nanosat has very little area for mounting solar panels. Deployable arrays can increase the overall solar collection area, but comes at the expense of increased complexity and risk. The use of GaAs solar cells with conversion efficiencies as high as 20% will maximize the amount of power generated by the Nanosat with no loss in reliability. Because there will be insufficient solar energy generated during sunlight to power the Nanosat transmitter, a large capacity battery is needed. Only a limited amount of transmit power per day will be available and must be used judicially to perform the required mission. A typical 800 km, circular, sun-synchronous orbit experiences a 35% eclipse cycle on every 100 minute orbit.

The small size of the Nanosat body presents certain problems for the transmit and receive antenna at lower frequencies. Operation in the UHF band or higher is desired to keep the size of the antennas within reasonable dimensions. At least one of the Nanosat receivers should use a fixed crystal frequency to insure that the satellite control center can gain access to it at all times. Power consumption of the spacecraft bus must be keep to an absolute minimum, and the transmit power must be used judiciously.

Nanosat Design Characteristics

The Nanosat is a very small "microsatellite" carefully designed and optimized for data relay with low power consumption. Excluding antennas, the satellite is a mere 5 inch cube and weighs approximately six pounds. It has an average power consumption of less than 1 watt, and uses high-efficiency GaAs solar cells to recharge the battery subsystem. Despite its small size, the Nanosat receiver is just as sensitive as satellites ten times its size. Each Nanosat has two receivers, power-agile transmitter, and two megabytes of solid-state data storage. The radios used for communications are tuned to the UHF band. A sketch of the Nanosat configuration is shown in Figure 2.

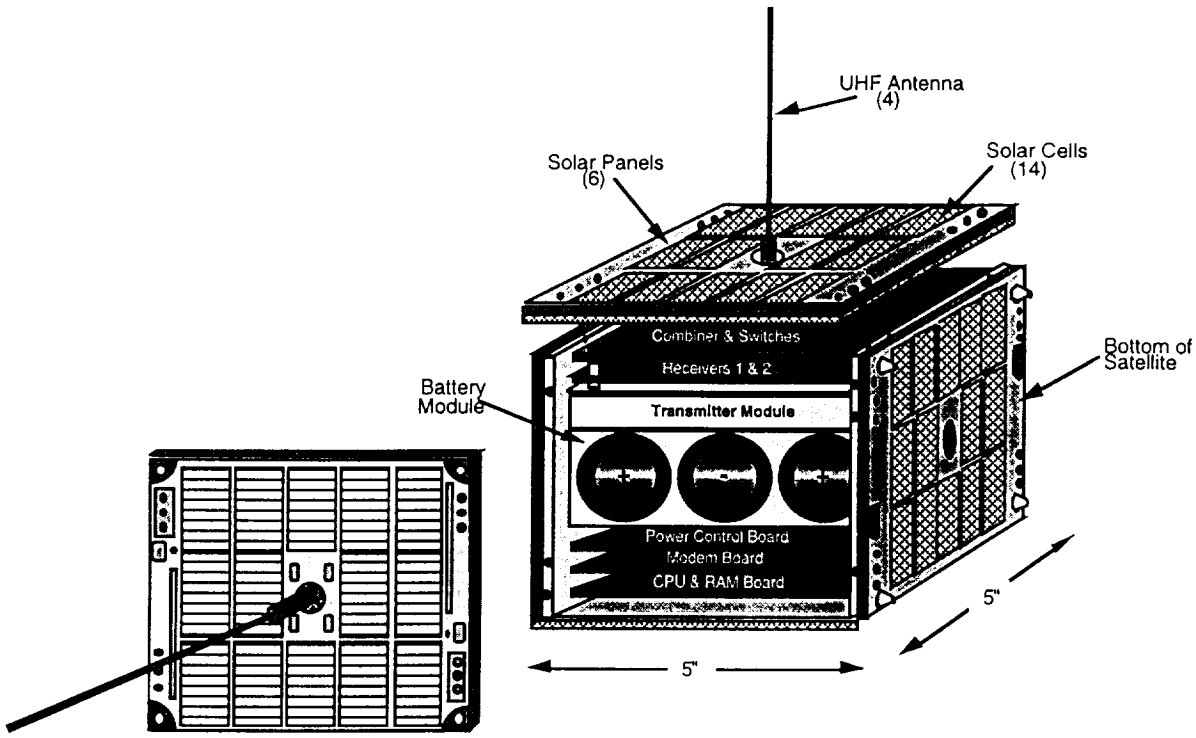


Figure 2. Nanosat Configuration

Because weight is not a significant concern, the Nanosat uses an all aluminum structure to achieve low manufacturing cost. Three F-size NiCd cells provide 6 amp-hours of battery capacity at 4.2 volts. Six multi-layer printed circuit boards with surface-mount components on both sides contain all of the essential Nanosat electronics. Board interconnection is accomplished using flexible Kapton ribbon cables.

The six interchangeable solar panels each contain two serial strings of seven 4 cm x 2 cm GaAs cells having an efficiency of 19%. Each panel produces 2.88 watts of power in sunlight. The six Nanosat panels combined will generate 3.7 watt-hours of energy during its 100 minute orbit after charging and regulation inefficiencies are considered. On a daily basis the Nanosat will collect approximately 54 watt-hours of energy. Since the entire Nanosat bus draws less than one watt continuously, this leaves 30 watt-hours of energy to operate the transmitter in a push-to-talk mode. This is sufficient to operate the Nanosat transmitter at 7 watts of RF power for at least two hours each day. Thus, a single Nanosat can send more than 15,000 polls using one hour of transmit time each day. The polls can be distributed throughout the world as needed to service mobile or fixed tags. In addition, each Nanosat can download over two megabytes of stored data to any data collection site during the remaining one hour of daily transmit time.

The Nanosat communication system consists of two high-sensitivity receivers, one agile and one set at a fixed frequency. The agile receiver also functions as a spectrum analyzer to measure the receive signal strength across a 2 MHz band. A single, power-agile transmitter operates from 1,200 to 9,600 bps. Its DC to RF efficiency is better than 50%. Its output power is software adjustable from 0.5 watt to a maximum of 7 watts. A 4-element, circularly-polarized, turnstile antenna serves for both the received and transmit functions. Each element consists of a short length of piano wire. UHF frequencies are used for both transmit and receive.

The modulation scheme is narrow-band FM, Gaussian-filtered Minimum Shift Keying (GMSK). The two demodulators can be commanded from the ground to operate at 1,200 bps or at 9,600 bps for uploading new flight software from the Network Operations Center in a single pass.

The Nanosat flight computer uses a V-53A 16-bit microprocessor which uses 3 volt logic. Two megabytes of Error Detecting And Correcting static RAM are used for both program and data storage. Additional memory can be included if required for a particular mission. The computer module also contains remote reset circuitry, an analog-to-digital converter for collecting on-board telemetry, direct memory access, and multiple serial communications controllers. The flight computer runs a real time multi-tasking kernel with application programs for polling, telemetry, housekeeping, memory management, subsystem power and functional control, and protocol implementation. A block diagram of the Nanosat electronics is shown in Figure 3.

A completely passive system consisting of a permanent magnet and four soft metal damping rods are used to stabilize the Nanosat. Using this approach, the Nanosat will align itself with the Earth's magnetic field, tumbling two times each orbit. The omni-directional antenna allows communication link closure for almost all orientations. A linear ground-based antenna accommodates both the right-hand and left-hand circularly-polarized Nanosat signals. No station-keeping mechanism is provided, nor needed for the intended applications.

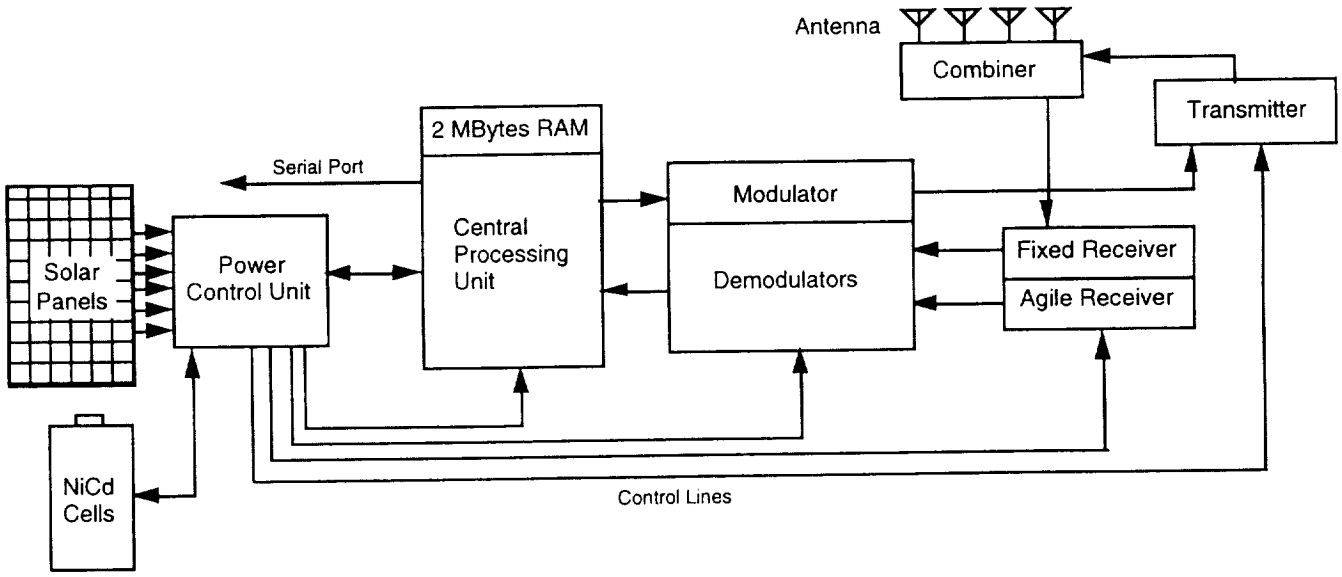
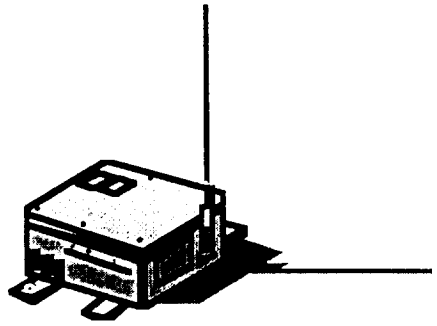


Figure 3. Block Diagram of Nanosat Electronics

6. USER SEGMENT

Tag Requirements



In order to be useful in commercial applications, the tag must be available at very low cost. A high level of functional integration along with large production quantities is necessary to drive the selling price down to levels that are attractive to mariners, shippers, researchers and environmentalists.

Besides its low cost, the tag must consume very little power so that unattended operation for several months is possible using only an internal battery. This requires extensive use of low power GaAs components, an efficient power amplifier design, and an intelligent power switching of the tag's subsystems. The smaller and lighter it is possible to manufacture the tag, the more varied applications it will be possible to address. A small size will allow it to be placed inconspicuously on many objects to be sensed or tracked. In some respects, the tag is as complex as the Nanosat itself.

Design Features

The tag under development by National Semiconductor is a self-contained device incorporating a microprocessor, modem, transmitter, receiver, antenna, and internal battery. Its built-in software handles all functions involved in collecting, processing, and transmitting the data it is programmed to relay. The terminal receives and stores data for transmission to the orbiting Nanosats. Its compact design facilitates easy mounting to a variety of objects, including cargo containers, buoys, pipelines, and boats. The case is weatherproof and rugged, allowing it to survive in harsh environments.

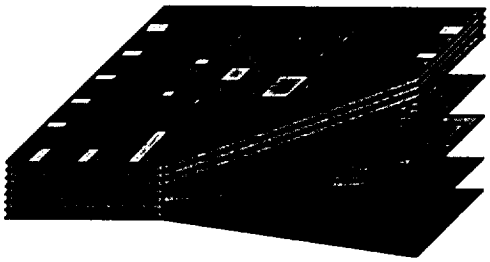
Through its standard RS-232/422 interface, the terminal can be configured to receive a signal from just about any source, and store that digital data. A transducer measuring anything from barometric pressure to ultraviolet light-levels may be incorporated into the terminal. The terminal may even be configured to transmit short text messages from remote sites. The input options make the terminal extremely flexible, because its fundamental design is independent of its specific use.

Each terminal also has the ability to program its transmission cycle for a specific role. For example, the terminal may be told to transmit data packets every time a Nanosat is in view (about 10 times per day for a two satellite constellation), just once a day, or once a month, depending on the time-varying nature of its data. For terminals with both transmit and receive capabilities, data transmission may be configured to take place only when the terminal is "polled" by the satellite. Limiting the tag's transmission time permits efficient use of the tag's battery power. Computer-controlled circuits and functions keep the power consumption of the unit low so that a small, integrated, lithium battery can be used effectively for many of the intended applications.

At a projected selling price of less than $100, the low-cost tags being developed by National Semiconductor can be considered disposable for some applications.

Low Cost Implementation

The high-performance, frequency-controlled, single-chip transceivers using BiCMOS technology have extremely low power consumption. The complete RF components will be embedded into a single multilayer ceramic module occupying an area of less than one square inch. Through the use of low temperature cofired ceramic (LTTC) technology, National Semiconductor is able to create substrate modules that integrate silicon, passive components and ceramic technology. Hundreds of passive components, including filters, are buried inside a multilayer ceramic substrate. This process enables micro-miniaturization and low cost.



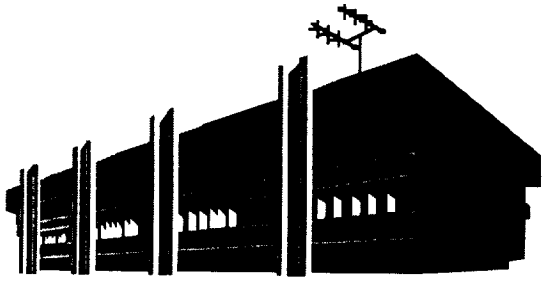
Using advanced lamination techniques and a single firing at 900°C, layers of ceramic tapes imprinted with thick-film materials are combined to form a monolithic structure. The ceramic structure built using the LTCC process contains buried conductors, vias, components such as capacitors, resistors, couplers and filters, and other RF devices to produce a highly integrated module.

8. GROUND SEGMENT

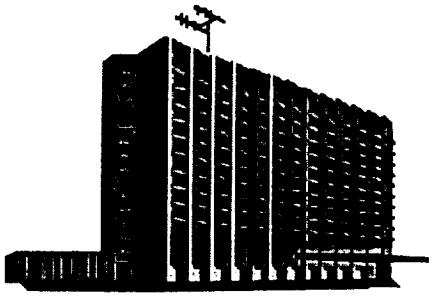
The Nanosat ground segment consists of one or more Network Operations Centers and as many data collection sites as may be required to download user data efficiently and give major users immediate access to the data generated by their deployed tags.

Network Operations Center

The Network Operation Center is responsible for the command and control of all of the Nanosats in the space constellation. This includes making changes to on-board software, monitoring the health status of each satellite via telemetry analysis, and uploading new mission schedules as may be required to satisfy user requirements. The primary Nanosat Network Operations Center is located at SpaceQuest's Headquarters in Fairfax, Virginia. A backup Operations Center is located at National Semiconductor in Santa Clara, California.



Data Collection Sites



A data collection sites consist of a state-of-the-art computer running commercial and custom software packages integrated and modified to meet unique mission requirements. These low-cost, fully-automated data collection sites can be deployed at various places around the world as may be needed by the user community. Ground station equipment for communicating with the Nanosats and downloading user data is located at each data collection site. Moreover, depending on the requirements of a particular user, a data collection site may be set up at the user's location, allowing immediate retrieval of data from the Nanosats.

Operation of the a data collection site is automatic and does not require the full-time presence of a system operator. It is ideally suited for applications where client personnel cannot be attendant. Specifically, the data collection site will download relevant tag data automatically and allow the client to retrieve his data without having to rely on an operator. The data collection sites can also be controlled remotely from SpaceQuest's Network Operations Center.

9. LAUNCH SEGMENT

The first two Nanosats will be launched into a low Earth orbit by a Russian Cosmos launch vehicle. The Nanosats will be a secondary payload on the booster during a planned launch in July, 1996 from Plesetsk Cosmodrome, Kapustin Yar, Russia.

The Cosmos booster, whose primary purpose is to deploy medium-sized satellite payloads, has completed more than 700 orbital missions with a cumulative operational reliability of 97.3%.

The current Nanosat development and launch schedule is shown in Table 1 below.

Nanosat Design & Prototypes	June 95 to Jan 96
Flight Unit Construction	Feb 96 to May 96
Integration and Testing	May 96 to June 96
Launch	July 96
System Checkout with Tags	July 96 to Sept 96
System Operation	Sept 96

Table 1. Nanosat Development Schedule

10. SUMMARY

SpaceQuest Ltd. and National Semiconductor Corp. are using nanotechnology to develop a space-based communications system that can perform several useful operational missions. The low cost and short development time needed to product the Nanosats, combined with inexpensive, low-power tags, will open up new applications which were not economically feasible with more expensive space systems.

A prototype system consisting of two Nanosats and several hundred tags should achieve its initial operational capability sometime in 1997. Vehicle tracking and monitoring will be the first applications to be addressed by this new service.

71083

TELEDESIC GLOBAL WIRELESS BROADBAND NETWORK: SPACE INFRASTRUCTURE ARCHITECTURE, DESIGN FEATURES AND TECHNOLOGIES

Dr. James R. Stuart
Vice President, Space Infrastructure

Teledesic Corporation, Kirkland, WA
(206) 803-1400, fax (206) 803-1404

NASA/Aerospace Corp. First International Conference on Integrated Micro-Nanotechnology for Space Applications

South Shore Harbor Resort and Conference Center, Houston, TX
30 Oct. - 2 Nov. 1995

Agenda

- Low Earth Orbit (LEO) Wireless Communications Revolution
- Teledesic 'Broadband LEO' Services and Network Features
  - Teledesic Corporate and Program Overview
  - Teledesic Services and Applications, Capacity, Coverage and Spectrum Usage
  - Teledesic Network and Architecture Features
- Teledesic Space Infrastructure Architecture and Design
  - Teledesic System/Subsytem Design Features and Key Technologies
  - Teledesic Space Segment Power, Mass, ΔV and Reliability Budgets
  - Teledesic Launch Campaign Features and Debris Mitigation
  - Teledesic Integrated Software and Distributed Control Features
  - Teledesic Constellation Control Operations and Ground Segment Features
- Industrial Impact of LEO Communications Revolution
  - New Service Providers and New Equipment Suppliers
  - Commercial Space Industry Impact
- Emerging Applications for the Shrinking Satellite Evolution
  - Possible Future with Hybrid Networks of GSO's, LEO's and HALE/UAV's
  - General Features of Ideal Micro/Nano-Satellite Networks
  - A Vision of Future Micro/Nano-Satellite Designs for Comm Networks

Teledesic

J.R.Stuart 10/29/95 1

Abstract: Teledesic represents a new paradigm for distributed space systems' design, production and operations. This paper will describe the Teledesic broadband services, applications, global network design and unique features of the new Teledesic space infrastructure, technologies and design approaches. The paper's introduction will discuss the wireless information revolution and the current 'Little' and 'Big' communications LEO's. The current technological and economic trends that drive us inevitably to higher frequency bands and much larger constellations (≥1000 satellites) will be briefly addressed. The Teledesic broadband network, services and architectural features will be described. Then the capabilities of the extremely high-performance and high-power Teledesic LEO satellites will be described (e.g., many kW's, 100's of MIPS, 1000 mps, 100's of beams, etc.).

The Teledesic satellites are a new class of small satellites, which demonstrate the important commercial benefits of using technologies developed for other purposes by U.S. National Laboratories (e.g., Phillips, NRL, JPL, LeRC, etc.). The Teledesic satellite architecture, subsystem design features and new technologies will be described. The new Teledesic satellite manufacturing, integration and test approaches will also be addressed which use modern high volume production techniques and result in surprisingly low space segment costs. The constellation control and management features and attendant software architecture features will be addressed. After briefly discussing the economic and technological impact on the USA commercial space industries of the space communications revolution and such large commercial constellation projects, the paper will conclude with observations on the trends towards future systems architectures using networked groups of much smaller satellites.

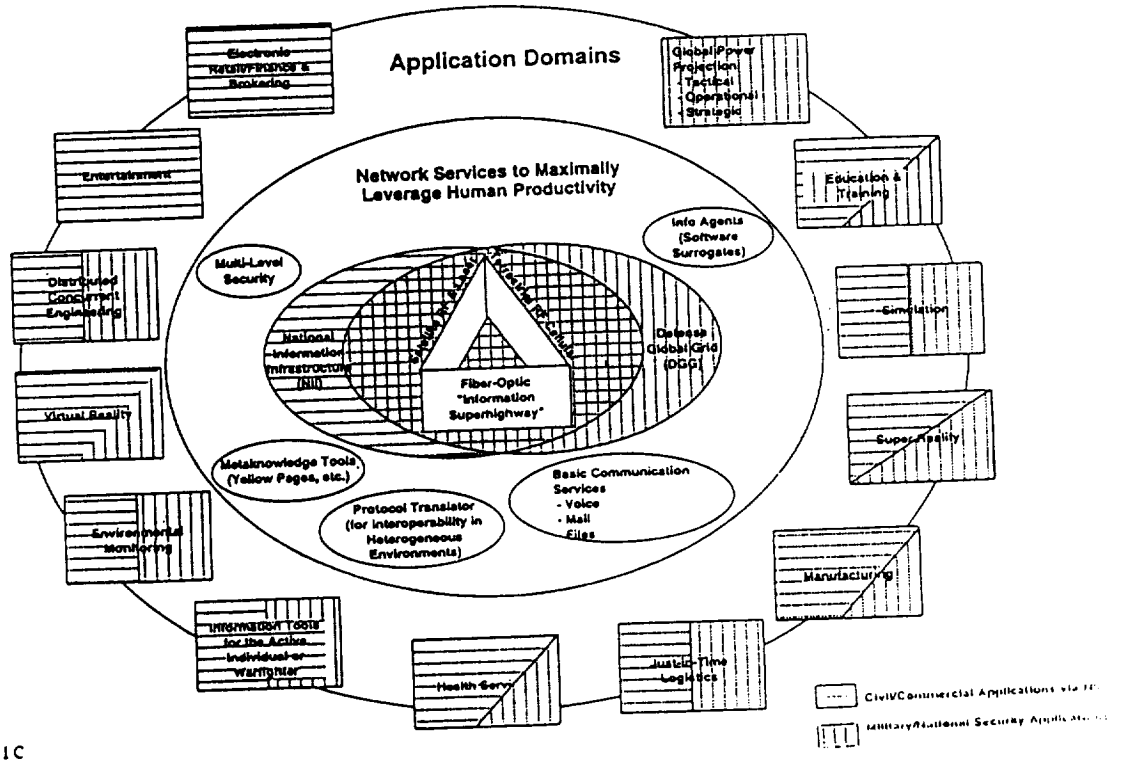
Teledesic

Global Information Infrastructure (GII)

- GII is a Vision of a Universally Accessible Web of Multiple Interconnected Networks
  - Permitting Access to Widely Distributed Private/Public Data Bases
  - Providing Ready Transmission of Information (Voice, FAX, Text, Images, Video, etc.)
    - In Any Format - To Anyone - In Any Place - At Anytime
- GII is an Entire GII System:
  - Human Users (and Developers)
  - User's Information Appliances (Computing and Consumer Electronics)
  - Accessed Information, Data Bases and Computing Resources
  - Networks
- The GII Network Will Be an Intricately Tangled Web of Multiple Overlaid Networks
  - Wired and Wireless
  - Terrestrial and Space
  - Physical and Virtual
  - Private, Commercial and Government
- GII (and Large Evolving Commercial Market) Will Migrate to Efficient Web Elements:
  - Reliable, Ubiquitous, Seamless, Interconnected, Flexible, Cost effective
  - Successful Elements will be Interoperable:
    - 'Open' Interfaces with Accepted Standards
    - Wide Array of Competing Information Appliances and S/W Tools
      - Interoperable and Interchangeable by Design
      - Standard User-Friendly (Easy) Interfaces for H/W and S/W:
        (e.g. Discovery/Recovery Applications, Operating Systems , etc.)
    - Many Interchangeable Competing Service Providers and Equipment Suppliers

J.R.Stuart 10/29/95 3

A Current View of LEOs Role in the National Information Infrastructure (NII)



Teledesic

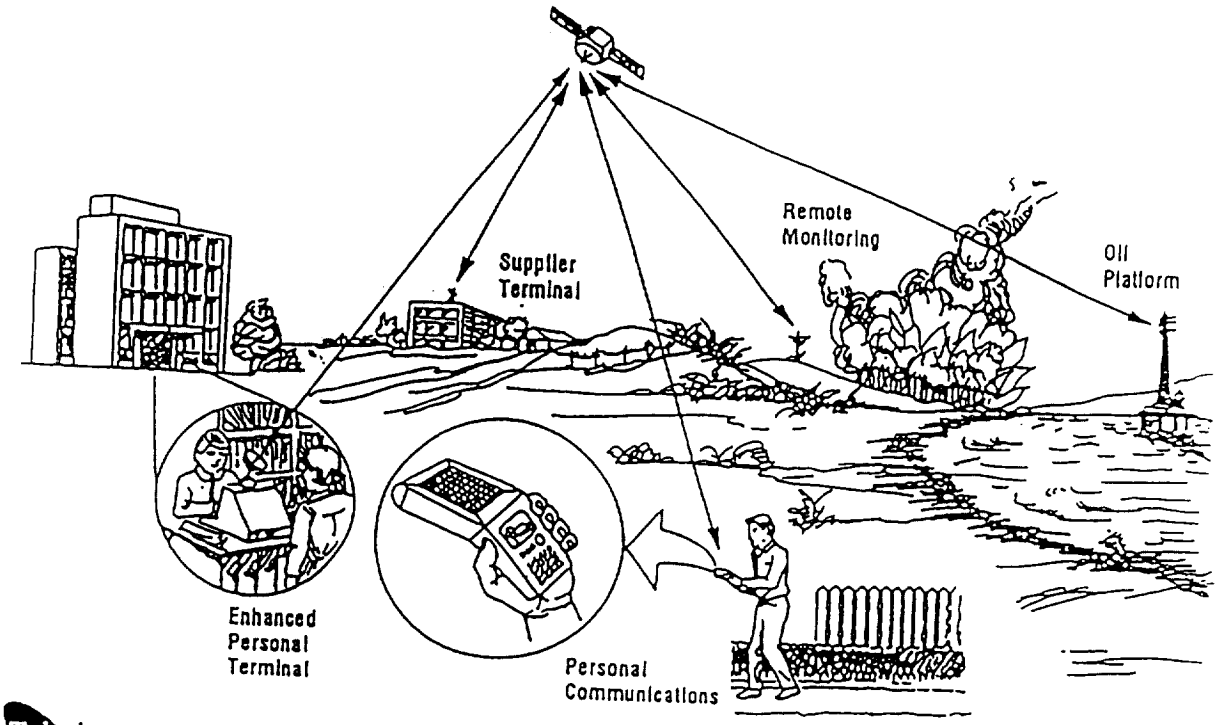
Source: Dr. Robert J. Bonometti (ARPA), Sept. '93

Low Earth Orbit (LEO) Wireless Communications Revolution

- LEO Communications Services Will Be Available Globally and Economically:
  - Voice and Broadband Data, Fixed and Mobile Services, Personal Communications
  - FAX , E-mail, Messages, Monitoring, Alarms, Positioning, Tracking and Location
- Personal Ground Terminal Business Is Enormously Larger Than LEO Space Segments
  - LEO Constellations Enable This Much Larger 'Information Appliance' Business
  - Hottest New Personal Electronic Products Since PC's and VCR's Will Be:
    - Mobile Communicators, Wireless Modems, Pocket Videophones, etc.
- Shift from Last 30 Years of Satellite Communications Evolution:
  - Bigger, More Powerful, Longer Lifed Satellites
  - Hierarchical Point-To-Point Communications Architectures
- Biggest Advance In Satellite Communications In 30 Years:
  - Lightsats, Intersatellite Links, Distributed Networks, New Competitive Multiple-Choices
    - Interconnectivity, Interoperablity, Global Marketplace Determination of 'Best'
  - Global LEO MSS Com. Services: 800M$ in 1992 to 10B$ in 2002 (1993, NASA/NSF)
- Future Will Be Networks Of Hybrid Systems Connecting Everyone To Everyone
  - Overlaid Interconnected and Interoperable Networks
    - Terrestrial Wire, Cellular, Coaxial Cable, Fiber Optic Cable, etc.
    - GSO Large Satellites, and the New LEO, MEO and GSO Lightsats
  - Large, Competitive, Open, Diverse Global Markets
  - Multiple Service Approaches Will Become Available to All Customers
  - Continuous Evolution Of Most Effective Set of Communications Networks
    - 'One Size Fits All' is Victim to More Convenient 2nd-to-Market Choices
    - Bandwidth/Quality/Price/Convenience-On-Demand (Interoperable Choices)

Teledesic

Wireless Satellite Services on the Horizon



Teledesic

LEO Satellite Communications Systems Service Categories

- Mobile (MSS) "Little" LEO's (UHF, VHF)

  Noncontinuous Worldwide Coverage (Periodic to Near-Real Time Availablity)
  "Bent Pipe" and "Store-and-Forward"
  Gateways, PSTN Connections
  Modulations: FDMA/TDMA or CDMA

  Non-RealTime and Near-Real Time Digital Mobile Services (2.4 kbps - 9.6 kbps)
  Digital Messages, Alarms, Monitoring Data, Tracking, E-Mail, FAX , Paging, etc

  Typical Delivery Delay Times
  Within Footprint (~4000 km Diameter): ≤2-10 minutes
  International (e.g., USA-Europe): 30 minutes - 8 hours

  Typical Subscriber Costs
  Terminals: $500-$100 (as low as $25 quoted for meter reading)
  Data: 1.0¢ -0.001¢ per byte

- Mobile (MSS) "Big" LEO's (L-Band)

  Continuous Worldwide Coverage (~ Cellular Dial-tone Availability)
  Either 'Bent Pipe' or via Intersatellite Links
  Gateways, PSTN Connections
  Modulations: TDMA or CDMA

  Local Cellular Company Size (largest: ~250,000 Subscribers at 0.1 Erlang)
  Real Time Mobile Services (~ 4.8 kbps): Digital Voice, Narrowband Data (<Toll Quality)

  Typical Long Distance Delay Times: ~Terrestrial Delays

  Typical Subscriber Costs
  Terminals: $1000 -$500 (and lower for RDSS only)
  Voice/Data: $3.00 - $0.50 per minute

  Typical Time and Cost to Send Daily NY Times (1 MB): 3.47 hours, $60 to $600

Teledesic

J.R.Stuart 6/21/95

LEO Satellite Communications Systems Service Categories (Cont'd)

- Fixed (FSS) and Mobile (MSS) 'Broadband' LEO's (Ka-Band)

  Continuous Worldwide Coverage
  Terrestrial Dial-tone Availability
  Small, Earth-Fixed Cells
  Regional Bell Operating Company Size
  >20,000 simultanous T1 (1.5 Mbps) connections worldwide
  Intersatellite Links
  Gateways, PSTN Connections
  Modulation: FDMA/TDMA
  Real Time Interactive Services (16 kbps - 1.2 Gbps)
  Bandwidth On Demand
  Broadband Data, Video, Digital Voice, etc. (>Toll Quality, 10^{-10} BER)
  Typical Phone Company Services and Features
  Typical Long Distance Delay Times: < Fiber

  Typical Subscriber Costs
  Interface Units: $10,000-$1,000 (falling sharply with volume and competition)
  Data: Comparable to local PTT charges
  Time and Cost to Send Daily NY Times (1 MB): 5 sec., few cents (to ~local PTT charges)

Teledesic

J.R.Stuart

'Little LEO' Satellite Communications Systems on the Horizon

- Mobile (MSS) "Little" LEO's (UHF, VHF)

FCC Construction License Granted

Orbital Communications Corp. (OrbComm)	36 Satellites,	40 kg,	4 year lifetime

FCC Construction License Pending (Experimental Licenses Granted)

Starsys Global Positioning, Inc. (Starnet)	24 Satellites,	125 kg,	5 year lifetime
Volunteers in Technical Assistance (VITA)	2 Satellites,	136 kg,	5 year lifetime

FCC Construction License Pending (2nd Round Applicants)

CTA Commercial Systems, Inc. (GEMnet)	38 Satellites,	45 kg,	5 year lifetime
E-Sat, inc. (E-Sat), USA	6 Satellites,	100 kg,	10 year lifetime
Final Analysis Comunication Services, Inc. (FAIsat) (Rec'd experimental lic. for 1 satellite)	26 Satellites,	100 kg,	7 year lifetime
GE American Communications (Eyetel)	24 Satellites,	15 kg,	5 year lifetime
Leo One USA Corp. (LEO ONE USA)	48 Satellites,	124 kg,	5 year lifetime
Orbital Communications Corp. (OrbComm) (Requesting 12 additional satellites)	48 Satellites,	40 kg,	4 year lifetime
Volunteers in Technical Assistance (VITA) (Requesting 1 additional satellite)	3 Satellites,	128 kg,	5 year lifetime

International 'Little LEO's' (in development/planning), e.g:

Leo One Panamericana (Mexico), ECO-8 (Brazil), Gonetz, Courier, Elekon (Russia), MiniSat(Spain), Safir (Germany), TAOS/S80T (France), Artes (Belgium), Leostar (ESA), KITCOM (Australia), etc.

Teledesic

'Big LEO' Satellite Communications Systems on the Horizon

- Mobile (MSS) "Big" LEO's (L-Band)

FCC Construction License Granted

Globalstar Telecommunications Ltd. (Gobalstar), USA
48 Satellites (+ 8 spares), 426 kg, 7.5 year lifetime, 1.6 B$

Iridium Inc. (Iridium), USA
66 Satellites (+ up to 12 spares), 700 kg, 5 year lifetime, 3.4 B$

Odyssey Worldwide Services, (Odyssey), USA
12 Satellites, 1952 kg, 12 year lifetime, 2.5 B$

FCC License Decision Deferred (Financial qualifications must be met by 1/96)

Constellation Communications, Inc. (ECCO), USA
46 Satellites (+ 8 spares), 500 kg, 6 year lifetime, 1.7 B$

Mobile Communications Holdings, Inc. (Ellipso), USA
16 Satellites, 500 kg, 5-7 year lifetime, 1.1 B$

American Mobile Satellite Corp., (AMSC), USA
12 Satellites, 3.1 B$

+ International 'Big LEO's' (in development/planning), e.g:

·Inmarsat P, UK (10 Satellites, 2.6 B$, 1.4 B$ committed), Russia, France, China, etc.

Teledesic

'Broadband LEO' Satellite Communications Systems on the Horizon

- Fixed and Mobile (FSS/MSS) Broadband LEO's (Ka Band)

Teledesic Corporation (Teledesic), Kirkland, WA, USA

Partners:	Craig O. McCaw, William H. Gates III, McCaw Cellular Communications (AT&T)
Constellation:	840 Satellites (21 polar orbits at 700 km altitude) (+ 84 in-orbit spares)
Satellite Mass, Lifetime:	800 kg, 10 year
Primary Market:	Rural and remote parts of the world that would not be economic to serve through traditional wireline means
Typical User:	Educational institutions, government agencies, health-care and industrial/commerical organizations, and people in remote areas
Typ. Cost per Minute:	Comparable to local PTT charges (includes PSN charges for local, long-distance, Int'l tails)
Initial Interface Unit Cost:	$10,000-$1,000 (falling sharply with volume/competition) (Standard Terminals, 16 kbps to 2 Mbps) ('Gigalink' Terminals, 155 Mbps to 1.2 Gbps)
Total System Cost:	9 B$
Communications:	Satellite switching (FDMA/TDMA)
FCC Status:	FCC Filed 3/94 (FSS), Ammendment 12/94 (MSS)

Teledesic

Teledesic Corporation Background and Status

- Teledesic Company Background
  - Founded in June, 1990
  - Concept Originally Developed (reduced to writing) in 1988
  - Headquaters: Kirkland, WA
- Corporate Mission Statement:

> "Teledesic seeks to organize a broad, cooperative effort to bring affordable access to advanced information services to rural and remote parts of the world that would not be economic to serve through traditional wireline means."

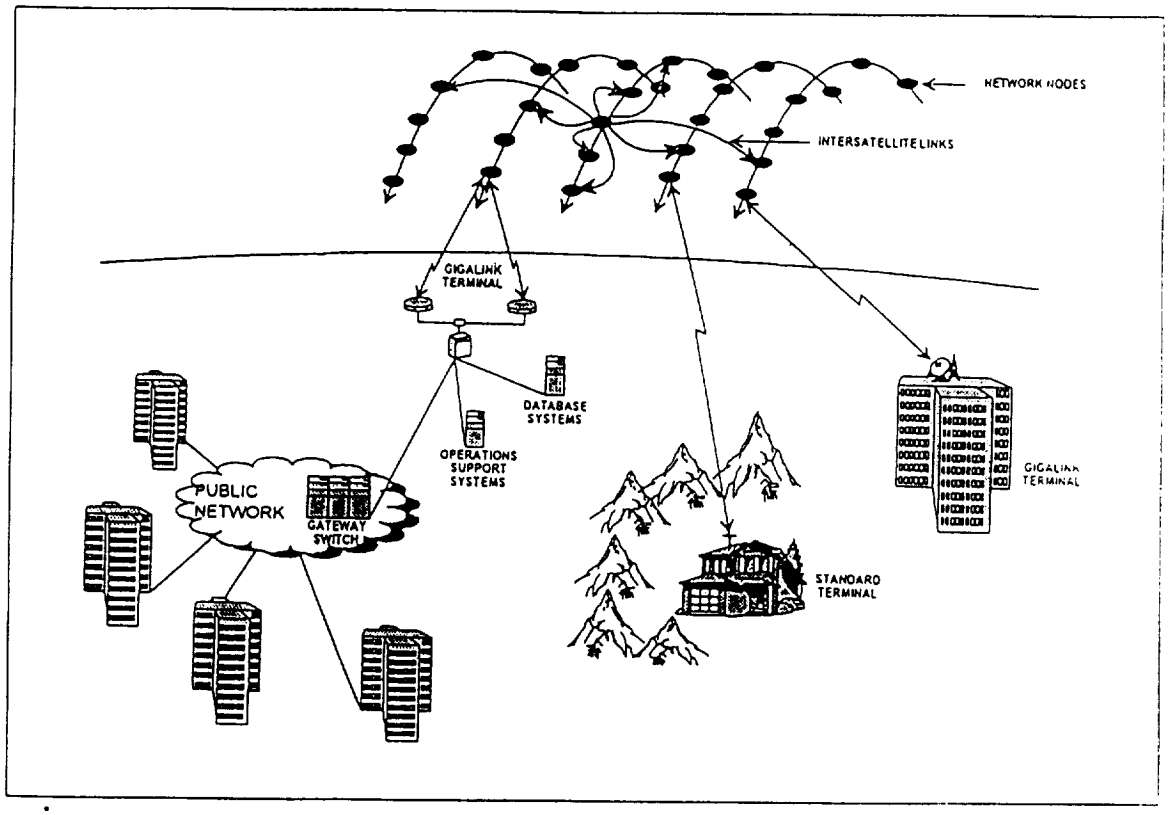
- Teledesic Shareholders

Craig O. McCaw (Founder - McCaw Cellular Communications)	32%
William H. Gates III (Founder - Microsoft)	32%
McCaw Cellular Communications (AT&T)	24%
Others	12%

- Teledesic Status
  - Feasibility Study and Point Design (Phase A) Completed
    - > 5 Years by Extraordinary Team of Full-time Employees, Consultants, and Subcontractors
  - FCC Application Filed 3/94 (FSS) and Ammendment Filed 12/94 (MSS)
  - Currently in Pre-Phase B (Planning and Development)
    - Regulatory Process Support
    - Program Planning and Organizational Development
    - System Requirements Update and Technologies Assessment
    - Key Supplier/Partner Candidates Identification and Selection

Teledesic

Teledesic Network Overview



Teledesic Services and Applications

- Provider (Wholesale) of Telecommunications Services to 'In-Country' Distributors

  Interactive 'Network-Quality' Voice, Data, Video, Multimedia, etc.

  Bandwidth-on-Demand

  - 16 kbps to 2 Mbps (Standard Terminals)
  - 155 Mbps to 1.2 Gbps ('Gigalink' Terminals)

- Switched and Point-to-Point Connections

- Connections Via Gateways to Terminals on Other Networks

- Teledesic Service Quality

  Comparable to Modern Urban Network

  'Fiber-Like' Delays

  16 kbps Basic Channels (Support 'Network-Quality' Voice, Data, etc.)

  1.5 Mbps Channels (Support 'Network-Quality' Data, 'VCR-Quality' Video, etc.)

  1.2 Gbps Channels (Support 'Fibre-Quality' Broadband Applications)

  Bit Error Rates $<10^{-10}$

  High Link Availability (Comparable with Urban Terrestrial Networks)

Teledesic Capacity, Coverage and Spectrum Usage

- Teledesic Network Capacity (Note: Actual user capacity depends on average channel rate and usage)

  Standard Terminals (16 kbps to 2 Mbps)

  >23 Mbps (standard terminal) capacity within Teledesic 53 km x 53 km Cell

  >20,000 simultanous T1 (1.5 Mbps) connections worldwide

  'Gigalink' Terminals (155 Mbps to 1.2 Gbps)

  16 steerable 'Gigalink' spots within Teledesic 1400 km diam. Footprint

  >8,000 simultanous 'Gigalink' connections worldwide

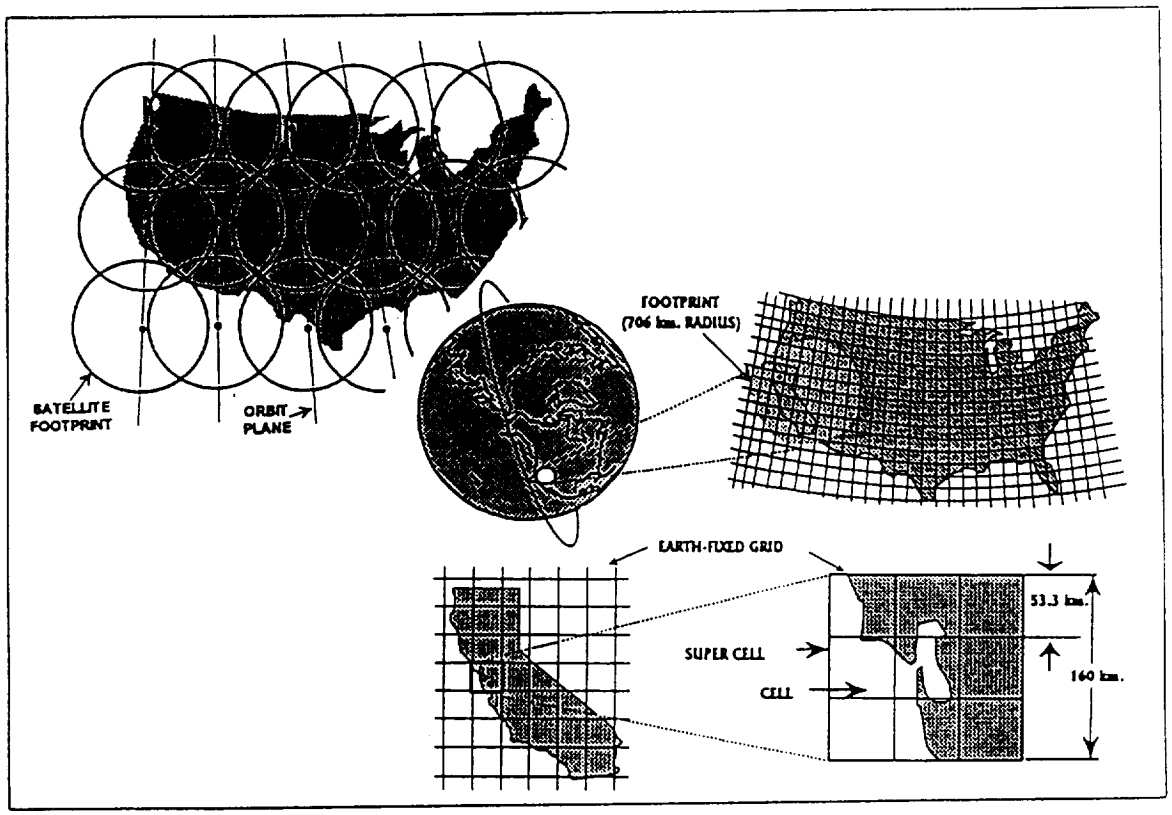
- Teledesic Network Handles Wide Variation in Channel Rates and User Densities

- Teledesic Network Grows 'Gracefully' to Much Higher Capacity

- Teledesic Spectrum Resource Bandwidth Requirements

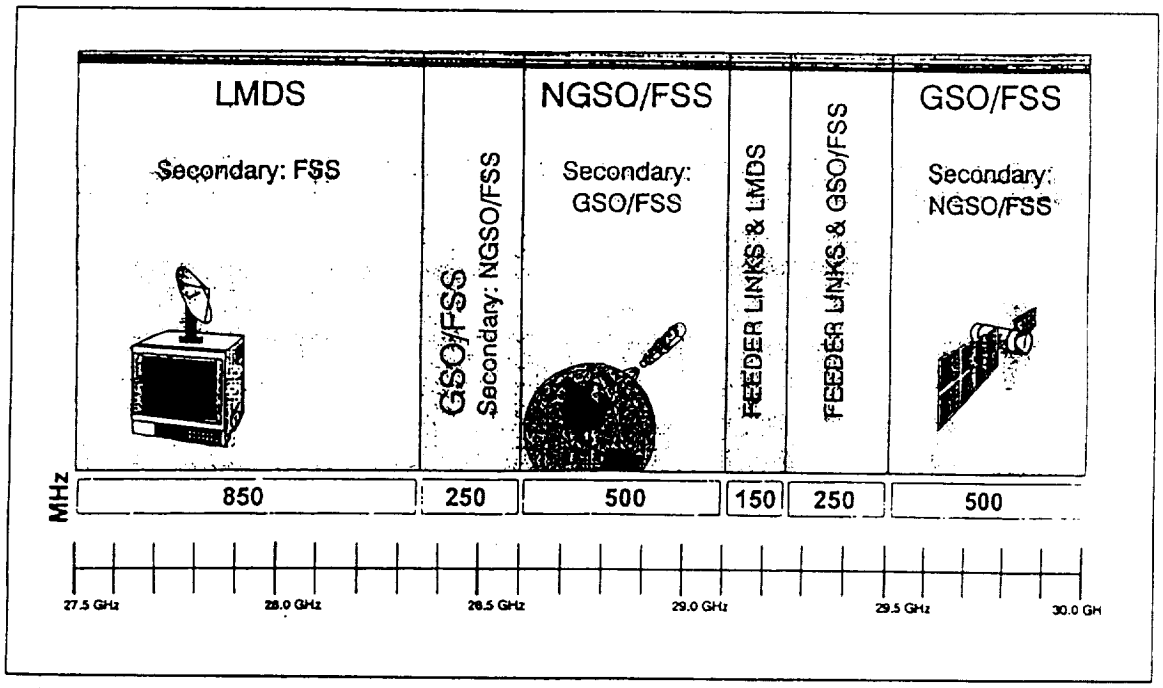
Standard Terminal Uplink (Bandwidth):	500 MHz
Standard Terminal Downlink (Bandwidth):	500 MHz
Gigalink Terminal Uplink (Bandwidth):	800 MHz
Gigalink Terminal Downlink (Bandwidth):	800 MHz
Intersatellite Cross LInks (Bandwidth):	2000 MHz

Teledesic Footprint and Cell Features



FCC Notice of Proposed Rulemaking (NPRM), 13 July 1995

- Proposed NGSO Allocation Can Accommodate Teledesic
  - 500 Mhz: Primary for Broadband LEO Service (NGSO)
  - 750 Mhz: Secondary for NGSO



Source: FCC (NPRM, 7/13/95)

Teledesic

J.R.Stuart 7/17/95 14

'Broadband GEO' Communications Systems on the Horizon

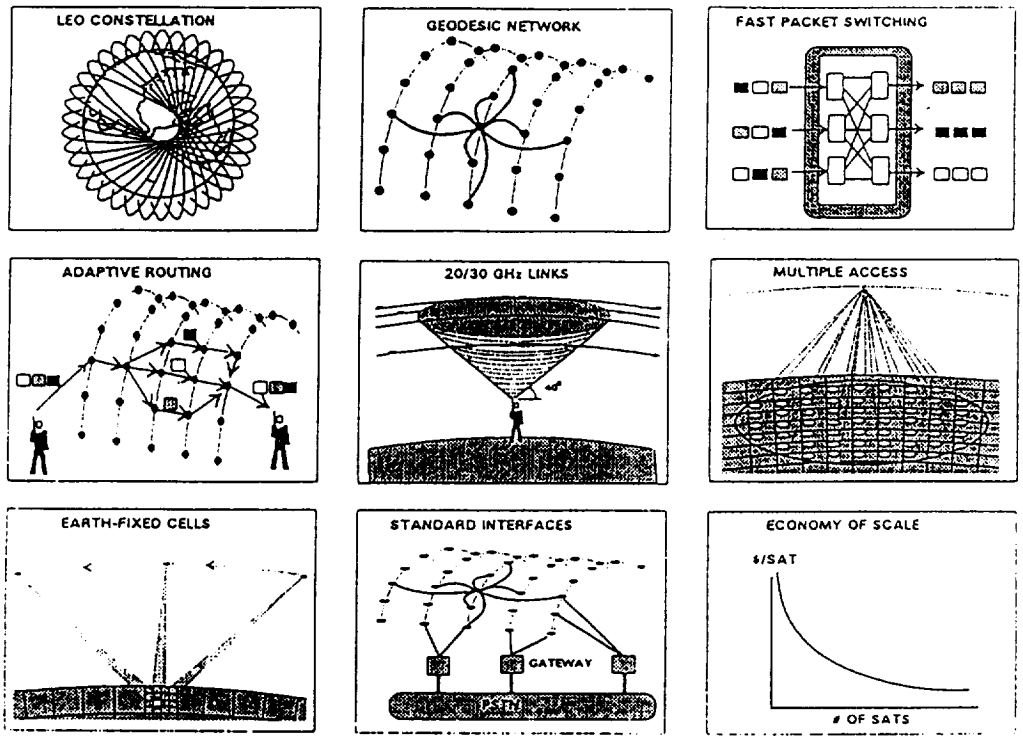
- Broadband GEO Fixed (FSS) (Ka Band)

FCC License Applicants

AT&T (Voicespan), USA	12 Satellites
EchoStar (EchoStar), USA	2 Satellites
GE Americom (xx), USA	9 Satellites
Hughes (Spaceway/Galaxy), USA	15 Satellites
KaStar (KaStar), USA	1 Satellite
Lockheed (AstroLink), USA	9 Satellites
Loral (CyberStar), USA	3 Satellites
Motorola (Millenium), USA	4 Satellites
NetSat 28 (xx), USA	1 Satellite
PanAmSat (PanAmSat),USA	1 Satellite*

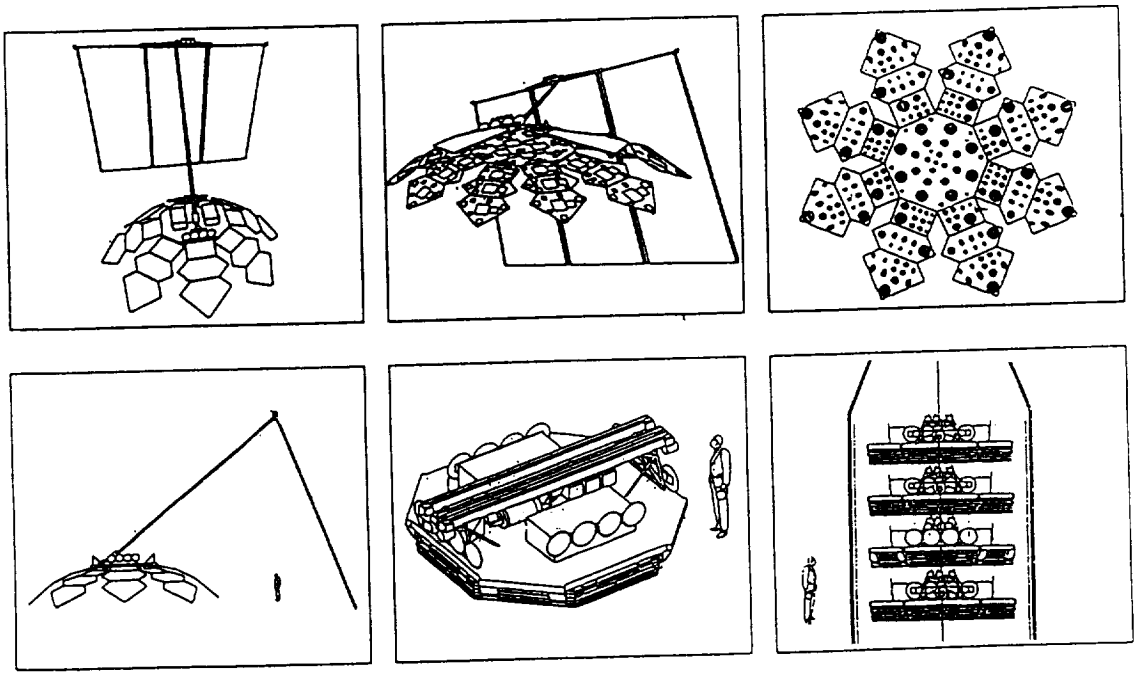
Teledesic

Teledesic Broadband LEO Network and System Features



Teledesic

Teledesic Satellite Configuration Features



Teledesic

Teledesic Space Segment Key Features

- Modern, High Performance, High Power, Mass-Producible Satellite System
  - Identical 3-Axis Stabilized Satellites for All Constellation Positions
  - High Performance, High Reliability, 10 year Lifetime Satellite System
    - High Power (>6.6 kW EOL, >300AH, 15 kW surge capability)
    - High Computational Power (>300 MIPs, >2 Gbytes RAM)
    - High ΔV Low-Thrust Propulsion (>1000 mps)
    - Lightweight (795 kg)
    - Compact Launch Configuration (3.1-3.3 m diameter x 2 m height)
- Design Features Tailored Specifically for Large Constellation
  - High Volume Production of Components
    - Large Economies of Scale
  - Automated Integration and Test of Satellite Systems
    - On-Board Test S/W
    - Autonomatic On-Orbit Health Monitoring and Constellation Control
  - Self-Stacked, Self-Deployed Group Launch by Variety of Launchers
    - Multiple International Launchers and Launch Sites
    - Assembly Facilities at Launch Sites
  - Automatic Orbit Transfer, Insertion and Gap-Filling
  - Active On-Orbit Spares with Routine Block Replenishments
  - Reliable End-of-Life Disposal/Deorbit Capability

Teledesic

Teledesic Space Segment Key Technologies

Baseline Modern Space Technologies	Technology Back-ups	Enhanced Technology Alternatives
Power		
CIS Thin Film Solar Array (Copper Indium Diseleinide, 6% EOL)	Crystal Si, Crystal GaAS Multi-junction, Concentrators	Thin Film CdTe (6% EOL) Thin Film CIGS (8% EOL) Poly-, Amorphous-Si
NH_2 (CPV) Batteries (6x60 AH)	NH_2 (IPV) Batteries NH_2 Batteries NiCad Batteries	Sodium Sulphur Batteries Lithium Ion Batteries Thin Film Polymer Batteries Flywheels (Lightweight, Long-life)
High Voltage Distribution System	28 VDC (DET)	AC Distribution
Propulsion		
Pulse Plasma Electric Thrusters (0.7 mN, 60 kN-s,1200 Isp)	Hall SPT Thrusters (80mN) Arc-Jets	Deflagration Thrusters Zenon Thrusters
Mechanisms		
Shape Memory Solar Array Extension Booms	Bistem Booms Cont. Longeron Booms	Inflatable Solar Array Booms
Paraffin (HOP) Latch/Deploy Mechanisms	Motors, Spring/Dampers	Shape Memory Mechanisms
Vibration Isolation (Passive)	Tuned Static Attachments	Embedded Active Piezo Electrics
Structures		
Advanced Composite Structures	Standard Composites Aluminum	Smart Structures Integrated Cabling/Thermal

Teledesic

Teledesic Space Segment Key Technologies (cont'd)

Baseline Modern Space Technologies	Technology Back-ups	Enhanced Technology Alternatives
Attitude Determination and Control		
Lightweight IFOG IMU's	RLG, QRS IMU's	DQI IMU's
Long-Life Reaction/Momentum Wheels	Multiple Back-up Wheels	Magnetic Suspension Wheels
Data Handling/Electronics		
High Perf. Rad Hard Microprocessors (PC603)	RS3000/6000, 68020	PC604, Pentium, etc.
Optical LAN Data Bus	1773 LAN Data Bus	High Perf. Optical LAN Bus
SC-cut Crystal Oscillators	AT-cut Crystal Oscillators	- -
GaAs A/D Converters	ECL A/D's	CMOS A/D's
GaAs VLSI Digital Signal Processors	ECL DSP's	CMOS DSP's
GaAs Fast Packet Switches	ECL FPS's	CHFET, Optical FPS's
Multi-chip (MCM) Packaging	Advanced Hybrids	UHDI, 3-D Packaging
Software		
Automated Prod., Ass'y, Test, On-orbit Ops S/W	Partially Automated S/W	Autonomous IA&T/COCC S/W
Communications		
PHEMT GaAs MMIC's: HPA's and LNA's	HBT MMICs	InP MMICs
20/30 GHz Phased Array Antennas	Gimballed Arrays/Reflectors	Multi-Beam Lens
60 GHz Intersatellite Phased Arrays	Gimballed 60 GHz Reflectors	Optical Intersatellite Links

Teledesic

Teledesic Satellite Resource Budgets

SATELLITE SUBYSTEM RESOURCE BUDGETS	Mass Kg	Power Average W	Volume cm3	Reliability %
Structure	87 kg	0 W	9,846 K cm3	99.9986 %
Mechanisms	52 kg	0 W	1,139 K cm3	95.3037 %
Cabling	22 kg	0 W	8 K cm3	99.9996 %
C&DH/TT&C	9 kg	8 W	11 K cm3	98.8493 %
Temperature Control	37 kg	24 W	153 K cm3	99.9990 %
Attitude/Orbit Det. and Control	12 kg	19 W	51 K cm3	96.6997 %
Propulsion	60 kg	0 W	250 K cm3	99.9999 %
Power	239 kg	2288 W	85 K cm3	98.9488 %
Communications Payload	144 kg	3000 W	3,557 K cm3	80.0488 %
Contingency (20%)	132 kg	1068 W	3,020 K cm3	-
SATELLITE SYSTEM:	795 kg	6,407 W	18.1 m3	72.2 %

Component Volume Inside Satellite Bus (3): 0.553 m3
Bus Fill Factor: 25%

Teledesic

Teledesic Satellite Propulsion ΔV Budget

PROPULSION ΔV BUDGET (10 yr)	Velocity Increment m/s
Orbit Transfer and Insertion	272 m/s
Orbit Drag Maintenance	47 m/s
Sunsync. Orbit Maintenance	30 m/s
Gap-Filling Maintenance	87 m/s
Deorbit Retro Maneuver	65 m/s
Contingency (20%)	100 m/s
TOTAL ΔV REQUIREMENT:	**601 m/s**
Total ΔV Capability	1010 m/s
TOTAL ΔV MARGIN (68%):	**409 m/s**

Teledesic

Teledesic Constellation Deployment

- Diverse Set of International Launchers (and Launch Sites) Baselined
  - Launch Site Throughput Capacity Assured
    - 973 satellites Launched in 24 months
    - ~1 Launch per Month from 4-6 International Launch Sites (~6 Pads)
    - Multiple Satellite Stacked Launches
  - Avoids Single-Point Interruptions (Launcher, Launch Site)
    - Launcher Production Problems
    - Launch Delays
    - Launch Failures
  - Assures Stable Launcher Supply (Capacity)
  - Assures Stable Launcher Economics (Competition)
- > 30 Viable International Candidate Launchers Identified
  - Expected Phase B Design Results:
    - Satellite Stowed Dimensions: 3.1-3.3 m diam x 2 m
    - Satellite Launch Mass: 800 kg
    - Stack Dispenser/Tug (1100 kg, 3.1-3.3 m diam x 2 m)
- Initial Deployment and Replenishment
  - Initial Launch of 973 Satellites
    - 840 Satellites — Constellation On-Orbit
    - 84 Satellites — On-Orbit Active Spare Satellites
    - 49 Satellites — Launch Failure Margin
  - Routine Replenishment of 195 Satellites
  - Autonomous Deployment, Orbit Raising and Positioning
    - Injection by Dispenser/Tug ~600km, Near-Polar
    - Low Thrust Spiral to Final Orbit (700km)
    - Drift to Adjacent Orbit Planes (12-16 weeks), as required

Teledesic

Teledesic Debris Mitigation

- Teledesic Mangement is Committed to Debris Mitigation
  - Early Establishment as Top Level Design Requirement
  - Long Term Self Interest
- Teledesic Debris Mitigation Requirements

> Risk of Teledesic generated debris on Teledesic constellation and other Space Assets must be small compared to risk from ambient debris environment.

- Teledesic Debris Mitigation Actioins
  - Early Establishment Unique Government/Industry Debris Experts for Debris Mitigation Analyses and Trades
    - Air Force Phillips Lab, The Aerospace Corp.
    - Lockheed-Martin, Orion Int'l, Teledesic
- Completed Phase 1 of Two Phase Study
  - Phase 1 focus: establish environments and requirements
  - Phase 2 focus: formulate design rules and validation methodology
- Completed NASA/JSC Review of Phase 1 Study Results (29-30 Sept. 1994)

> " Teledesic debris mitigation policy of limiting and managing the generation of debris to less than background is achievable."

IAF paper (IAA-94IAA.6.2.702 (12 Oct 1994)
Teledesic Debris Mitigation Phase 1 Study Report (Lockheed-Martin, Apr 1995)

Teledesic

Teledesic Ground Segment Key Elements

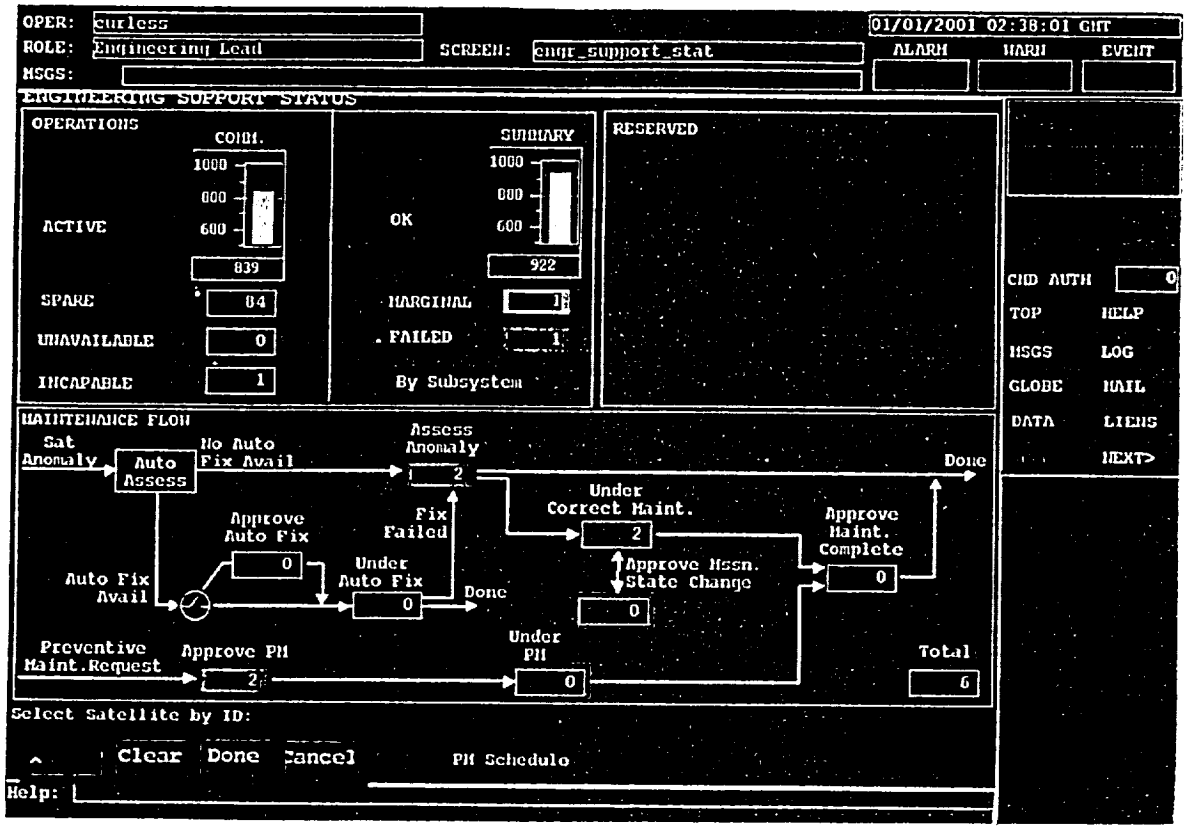
- Terminals

Standard Terminals:	16 kbps to 2 Mbps
'Gigalink' Terminals:	155 Mbps to 1.2 Gbps
COCC, NOCC, SPAC Gateways	155 Mbps to 1.2 Gbps

- Network Operations and Control Centers (NOCC)
  - Redundant Facilties, providing e.g.,
    - Feature Processors
    - Network Management
    - Subscriber and Network Databases
    - Global Administration and Billing Systems
  - Owned and Operated by Teledesic
- Service Provider Administration Centers (SPAC)
  - Redundant Gateway Antennas
  - Regional Administration, Billing Systems and Regional Network Control
  - Owned and Operated by Service Provider
- Constellation Operations and Control Centers (COCC)
  - Redundant Facilties for 4 Teams
    - Health Monitoring/Failure DetectionTeam ('Front Room')
    - Diagnostic/Failure Isolation Team ('Back Room')
    - Disposal/Deorbit Team ('Back Room')
    - Launch/Initialization/ReplacementTeam ('Back Room')
  - Owned and Operated by Teledesic

Teledesic

Teledesic COCC Display Example



J.R.Stuart 7/17/95 29

Large LEO Projects Will Stimulate the Commercial Space Industry and Global Competitiveness

- Global Information System Infrastructure
  - Wireless Bandwidth on Demand (16 kbps to 1.2 Gbps)
- Space Communications Technology
  - 20/30 GHz Phased Arrays, GaAs Receivers/Transmitters
  - 60 GHz or Optical Gigabit Intersatellite Links (>1 Gbps)
  - Gigabit Modems and Multi-Gigabit Packet Routing
- Low Cost, High Capacity User Terminals (rates up to 1.2 Gbps)
- Volume Satellite Component Production, e.g.:

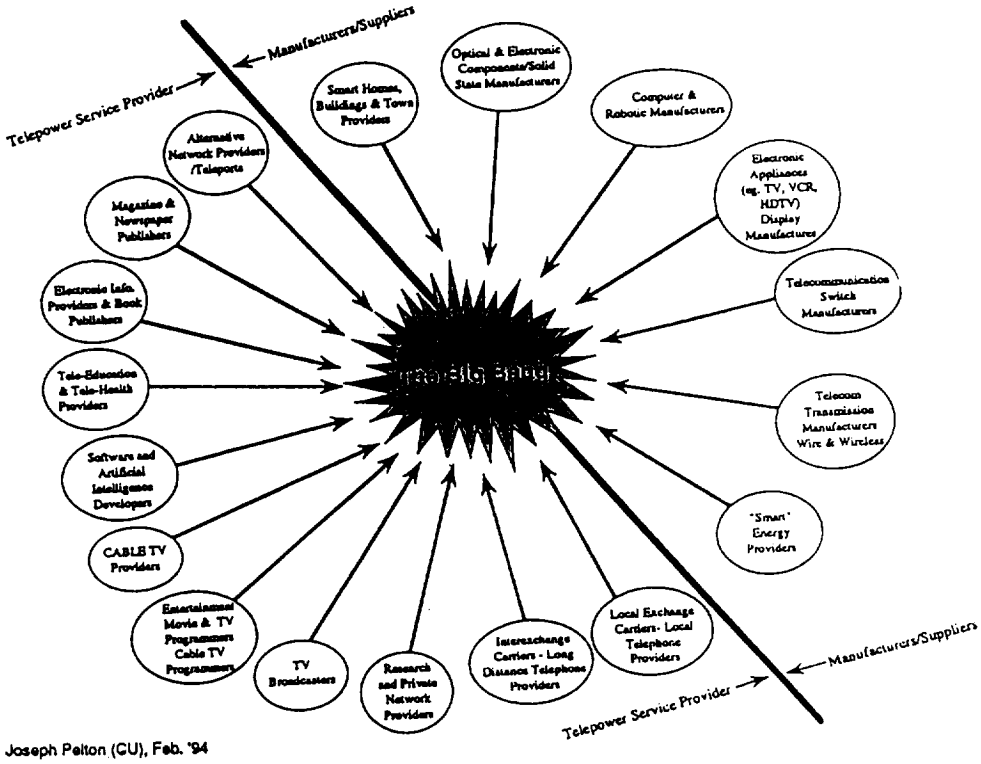
10 Million	Watts of Solar Cells
300,000	Amp-hours of Batteries
24,000	Gigabits Modems
8,000	Electric Thrusters
8,000	Gigabits Crosslinks
3,000	Space Computers (with Peripherals)

- Automatic Satellite Production, Assembly, Test and Constellation Operations
  - State-of-the-Art Software Engineering Techniques
  - Production, Assembly, Test and Operations S/W
    - Standard Operating System with Applications (3rd Party)
- Robust Launch Campaign
  - 1 Million Kilograms to Low-Earth Orbit

Teledesic

J.R.Stuart 8/24/95 32

Wireless Revolution in New Service Providers and New Equipment Suppliers



Source: Dr. Joseph Pelton (CU), Feb. '94

Teledesic

Emerging Applications for Smaller (Light-, Micro- and Nano-) Satellites

- Current revolution in size and capabilities of space components and systems
  - Driving from Lightsats (100's-10 kg) to Microsats (10-1kg) to Nanosats (<1 kg)
- Shrinking satellites' attributes (small, light, focussed, high performance, quick development, producibility, distributed control, high-tech front-end investments, etc.) causing fundamental changes in choices:
  - New space systems capabilities, affordability and availability
  - New Industry structure and business approach
  - New technologies expand marketplace, applications, opportunities
- New space applications within reach of unprecedented number of people, e.g.:
  - Scientists, battlefield commanders, farmers, businessmen, researchers, etc.
- Difficult to predict market by extrapolating from 'mainframe' satellite experience
  - e.g., Apple couldn't foresee spreadsheets while developing Apple II
  - <u>New unforeseeable powerful applications undoubtedly coming</u>
- 5 early markets for increasingly smaller satellites apparent now:
  - Space science research
  - Environmental monitoring
  - Tactical military applications
  - Technology testbeds
  - <u>Communications</u>
    - Commercial space dominated by sat comm goods/services
    - Evolutionary sat comm improvements over past 30 years
      - 1000 times more cost effective
      - 100 times higher power
      - 50 times higher frequency use efficiency
      - 10 times longer lifetimes

USA Satellite Communications Technology 'Firsts'

First satellite with broadcast transmission capability from space (SCORE) -- 1958
First teletype relay by satellite (Courier 1B) -- 1958
First passive communications satellite (ECHO) -- 1960
First active communications satellite (Telstar) -- 1962
First communications satellite to transmit TV worldwide (Relay) -- 1962
First geosynchronous communications satellite (Syncom II) -- 1963
First operational military communications satellite (IDSCS) -- 1965
First operational commercial communications satellite (INTELSAT I, "Early Bird") -- 1965
First communications satellite capable of multiple access transmissions (INTELSAT II) -- 1967
First satellite to provide UHF mobile communications (TACSAT) -- 1968
First satellite with a despun antenna (INTELSAT III) -- 1968
First satellite with high-power spot-beam antennas (INTELSAT IV) -- 1971
First communications satellite to achieve frequency reuse (INTELSAT IVA) -- 1975
First communications satellite to provide commercial mobile satellite services (MARISAT) -- 1976
First complex hybrid communications satellite capable of operating in multiple frequency bands with multiple frequency reuse (INTELSAT V) -- 1980

Source: ITRI NASA/NSF Panel Report on Int'l Satellite Communications, 7/93

A Conservative Projection of the Annual Communications Service Business for the Next Decade

SATELLITE SERVICE	1992	2002
Fixed Satellite Services		
o INTELSAT	$4.5 billion	$8.5 billion
o Regional and Other International Sat. Systems	1.8 billion	3.6 billion
o U.S./Canada Nat'l Systems	2.3 billion	4.5 billion
o Other National Systems	1.4 billion	3.4 billion
Fixed Satellite Service (Total)	$10.0 billion	$20.0 billion
Mobile/Low Orbit Services	$0.8 billion	$10.0 billion
Broadcast Satellite Services	$0.5 billion	$8.0 billion
Military Satellite Services	**	**
Other (e.g., Data Relay, etc)	$0.1 billion	$0.3 billion
Total Services	$11.4 billion	$38.3 billion

* Table: does not include equipment sales (e.g., satellites, launch vehicles, ground stations, etc.), which were about $5 billion in 1992, and are predicted to double in the next decade.

** No accurate or meaningful figures for military services are readily available.

Source: Pelton, Edelson, Helm (ITRI NASA/NSF Panel Report on Int'l Satellite Communications) 7/93

Space Communications Today

- Space Communications is a Big Business
  - First and Still the Only Big Commercial Pay-off in Space
  - 160 Countries and Territories Involved with GSO Systems
    - >100 Satellites in GSO
    - > 20 Operational International, Regional and National Systems
  - 10 Countries have Significant Satellite Communications Industry Capabilities
    - > 10 B$/year in Revenues from Space Communications
    - > 5 B$/year Equipment Market (Satellites, ELV's, Terminals, etc.)
- US has Dominated the Space Communications Business for Past 25 Years
  - R&D from NASA and DOD Played Key Role in USA Satellite Communications Industry Development
  - USA Lead the World in Satellite Communications Technology Development
- Satellite Communications Business is Changing Fast and about to Explode
  - Global Market will Expand Rapidly into Personal Communications
  - Large Economic and World Power Stakes are Involved for Dominant Nation(s)
- USA Leadership (Technological and Economic) is Being Challenged
  - Over Past 2 Decades Many Other Nations have Invested Heavily Sat Com R&D
  - Dominant Role Played by USA in Past 25 years is Clearly Now Over
  - Engaged in Global Competition for Dominance (Technological and Economic)

Trends to Micro/Nano-Satellite Networks

- Commercial space dominated by sat Communications goods/services
  - Evolutionary GSO sat comm improvements over past 30 years
    - 1000 times more cost effective
    - 100 times higher power
    - 50 times higher frequency use efficiency
    - 10 times longer lifetimes
  - LEO's and HALE's will be next revolution in communicaitons
    - Shrinking size and distance to user
  - Nano-satellites in clusters and constellations will be following wave
    - Distributed, networked/interlinked, virtual missions
    - Driven by comms (and remote sensing) applications
- Bandwidth, Availability, Interoperability and Mobility will drive future comms
  - Small user terminals require high power or large apertures on satellite
  - Low power, distributed, large aperture, interlinked network of nano-sats
- General Features of Ideal Micro/Nano-Satellite Networks
  - Low-Cost, Disposable, Low Power Highly Efficient Nodes
  - Large, Distributed Aperture, Small Steerable Beams
  - Capable Inter-satellite links with precision position determination
  - Reliable Distributed Control and High Autonomy
  - Shared Mu.ti-Network Operating Systems and Interoperable Control
  - etc., etc. . .
- A Vision of Micro/Nano-Satellite Designs for Comm Networks

Teledesic

Biography: Dr. James R. Stuart is the Vice President, Space Infrastructure for Teledesic Corporation (a global, wireless, broadband, LEO network), and has been an independent, international consultant specializing in advanced commercial space systems. He has layed an important role in the creation and development of LEO and GSO communications lightsats, and is currently an active principal and board member in several entrepreneurial technology and space companies involved with communications satellites and small launch vehicles. Dr. Stuart also acts as Chief Technical Advisor for two 'Little LEO' store-and-forward programs (a licensed Mexican constellation, and a recently filed U.S. constellation). Dr. Stuart previously held positions as Chief Scientist and Chief Engineer at Ball Space Systems Division in Boulder, CO. He was also founding Chief Engineer of Orbital Sciences Corporation, Assistant Laboratory Director of the Laboratory for Atmospheric and Space Physics, and creator and first Project Manager of the Mars Observer Project at NASA/Jet Propulsion Laboratory, where he was also Manager of Advanced Planetary Programs. Dr. Stuart has been on various graduate faculties of the University of Colorado at Boulder for over 15 years: in the Electrical Engineering, Telecommunications and Aerospace Engineering Sciences Departments, as well as in the Center for Space Construction. He received his Ph.D. in Systems Engineering (1979), M.S. in Operations Research (1977), and M.S. in Electrical Engineering (1974) from the University of Southern California, and his B.S. in Physics (1968) from the University of Washington. Dr. Stuart has received numerous professional awards, including NASA's Exceptional Service Medal for his project management of the Solar Mesosphere Explorer Project, JPL's highly successful, first modern small satellite project. He is also listed in Via Satellite's "Top 100 Executives in the Satellite Communications Industry". Dr. Stuart has published over 90 professional papers on the topics of small satellite systems, space technologies and communications satellite economics.

Teledesic

Session 2: Data Processing and Communications (Monday, PM)

Session Chairs: George Haddad (UM) and Michael Gorlick (Aerospace)

Separation and Detection of Toxic Gases with a Silicon Micromachined Gas Chromatography System
by Edward S. Kolesar, Jr., Texas Christian Univ., and Rocky R. Reston, USUHS

A 32-bit Ultrafast Parallel Correlator Using Resonant Tunneling Devices
by Shriram Kulkarni, Pinaki Mazumder and George I. Haddad, U. of Michigan

Highly-parallel, Highly-Compact Computing Structures Implemented in Nanotechnology
by D. G. Crawley, M. J.B. Duff, T. J. Fountain, C. D. Moffat, and C. D. Tomlinson, University College, London, UK

Technologies and Designs for Electronic Nanocomputers
by Michael Montemerlo, J. Christopher Love, Gregory Opiteck, David Goldhaber, and James C. Ellenbogen, MITRE Corporation

Resonant Tunneling Analog-to-Digital Converter
by T.P.E. Broekaert, A. C. Seabaugh, J. Hellums, A. Taddiken, H. Tang, J. Teng, and J.P.A. van der Wagt, Texas Instruments

MEMS-based Communication Systems for Space-Based Applications
by Héctor J. De Los Santos and Robert A. Brunner, Hughes Space and Communications, and Juan F. Lam, Le Roy H. Hackett, Ross F. Lohr, Jr., Lawrence E. Larson, Robert Y. Loo, Mehran Matloubian, and Gregory L. Tangonan, Hughes Research Laboratoriess,

Silicon Micromachining in RF and Photonic Applications
by Tsen-Hwang Lin, Phil Congdon, Gregory Magel, Lily Pang, Chuck Goldsmith, John Randall, and Nguyen Ho, Texas Instruments

Software Component Technologies and Space Applications
by Don Batory, University of Texas at Austin

Averting Denver Airports on a Chip
by Kevin J. Sullivan, University of Virginia

SEPARATION AND DETECTION OF TOXIC GASES WITH A SILICON MICROMACHINED GAS CHROMATOGRAPHY SYSTEM

Edward S. Kolesar, Jr. and Rocky R. Reston*

Texas Christian University, Department of Engineering
Fort Worth, TX 76129

*Uniformed Services University of Health Sciences
School of Medicine, Bethseda, MD 20853

Abstract

A miniature gas chromatography (GC) system has been designed and fabricated using silicon micromachining and integrated circuit (IC) processing techniques. The silicon micromachined gas chromatography system (SMGCS) is composed of a miniature sample injector that incorporates a 10 µl sample loop; a 0.9-m long, rectangular-shaped (300 µm width and 10 µm height) capillary column coated with a 0.2-µm thick copper phthalocyanine (CuPc) stationary-phase; and a dual-detector scheme based upon a CuPc-coated chemiresistor and a commercially available, 125-µm diameter thermal conductivity detector (TCD) bead. Silicon micromachining was employed to fabricate the interface between the sample injector and the GC column, the column itself, and the dual-detector cavity. A novel IC thin-film processing technique was developed to sublime the CuPc stationary-phase coating on the column walls that were micromachined in the host silicon wafer substrate and Pyrex® cover plate, which were then electrostatically bonded together. The SMGCS can separate binary gas mixtures composed of parts-per-million (ppm) concentrations of ammonia (NH_3) and nitrogen dioxide (NO_2) when isothermally operated (55-80°C). With a helium carrier gas and nitrogen diluent, a 10 µl sample volume containing ammonia and nitrogen dioxide injected at 40 psi (2.8 x10^5 Pa) can be separated in less than 30 minutes.

Introduction

Gas chromatography (GC) is a popular analytical chemistry tool commonly employed in the laboratory setting to analyze gas mixtures. With a GC system, the components of a gas mixture can be separated, identified, and their concentrations quantified. In their most common configuration, GC systems tend to be large, fragile, expensive, and bulky pieces of instrumentation. In consonance with the Environmental Protection Agency (EPA) and National Institute of Occupational Safety and Health (NIOSH) federal mandates for accomplishing on-site chemical analyses, several investigators have recently focused their attention toward realizing portable and robust GC systems[1-10]. Consistent with this motivation, this research realized a functional GC system by applying conventional integrated circuit (IC) processing techniques in conjunction with the concept of micromachining the column and integrating the key components on a single-crystal silicon wafer. The evolution of this technology will ultimately afford the opportunity to realize a complete miniaturized GC system that is inherently smaller, less massive, and highly portable compared to the presently available hardware (envisaged to be similar in size to a pocket calculator).

As depicted in Figure 1, the silicon micromachined gas chromatography system (SMGCS), consistent with conventional gas chromatography systems, consists of five fundamental components: (1) carrier gas supply, (2) sample injection system, (3) separation column, (4) detector, and (5) data processing element. In this research, the separation column and detector were microfabricated. To separate the components of a gas sample, a precise and reproducible volume needs to be extracted from the environment of interest with the sample injection valve, where it is then injected into the capillary column via an inert carrier gas. The stationary-phase thin-film which coats the surfaces of the capillary column adsorbs and desorbs each component of the sample gas depending upon its unique activation energy.

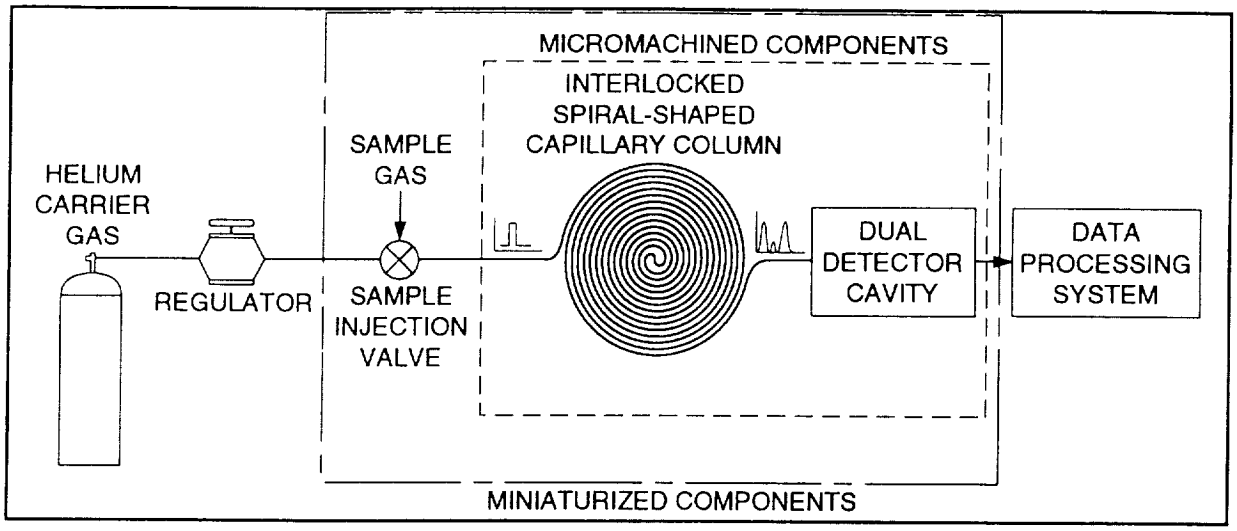


Figure 1. Functional block diagram of the micromachined gas chromatography (GC) system.

The differential propagation rate (velocity) of each gas component through the capillary column depends upon several factors, including the sample injection pressure, carrier gas velocity, the temperature, and the affinity of the individual gas components relative to the column's thin-film stationary-phase. As a consequence of this phenomenon, the gas components comprising the injected sample pulse mixture emerge at the capillary column's output as a time-resolved series of peaks that are separated from each other in the inert carrier gas. To detect the peaks associated with each separated gas component, the capillary column's gas effluent is analyzed by one or more detectors, whose functions are to measure a particular property of the gas components (for example, thermal conductivity, electrical conductivity, etc.). The magnitude of a detector's response to a particular gas component can correspondingly be related to its concentration in the injected sample.

SMGCS Design and Fabrication

In this research, a commercially available, electronically-actuated sample injector (Valco Instruments Company, Inc., product E6N6W, Houston, TX) incorporating a sample loop with a 10 μl volume was utilized to satisfy the critical requirement for injecting a reproducible pulse of the sample gas into the GC column. The volume of the sample loop was established experimentally by analyzing the conditions under which the capillary column would become saturated if it were to be filled with the highest conceivable concentration of any of the analyte gas components.

As shown in Figure 1, the miniaturized portion of the SMGCS is composed of a micromachined, interlocking, spiral-shaped capillary column integrated with a dual-detector arrangement, which consists of a separately batch fabricated copper phthalocyanine (CuPc) coated integrated circuit (IC) chemiresistor and a commercially available, 125-μm diameter thermal conductivity detector (TCD) bead (Thermometrics, Inc., model B05, Edison, NJ). The lower portion of the micromachined GC column is fabricated in a 3-inch diameter (100)-oriented silicon wafer (Polycore Electronics, *n*-type, Newberry Park, CA), and the matching upper portion of the column is etched in a 4-inch square Pyrex® cover plate (Schott America, Pyrex® 7740, Yonkers, NY). Before these two components are electrostatically bonded together, the GC column walls are coated with an α-phase CuPc thin-film[11], and then the TCD is mounted in the micromachined detector cavity. The independently batch fabricated IC chemiresistor (MOSIS - Metal-Oxide-Semiconductor Implementation System, Marina del Rey, CA) is coupled with the SMGCS along with the interface structure for the sample injector. The electrical response of the SMGCS results from the fundamental gas chromatography process which occurs in the separating column, and the corresponding performance of the non-specific TCD and the highly-specific CuPc-coated chemiresistor.

To quantify the efficiency of a candidate GC column design, a *separation factor* (*SF*) performance parameter can be calculated based upon the following equation[12]:

$$SF = \frac{L}{h(D, z_o, v_o, k)\, g} \left(\frac{k}{k+1}\right)^2 \tag{1}$$

where L is the column's length, k is the column's partition ratio, g is a pressure correction factor, and h is the theoretical plate height. The theoretical plate height, h, is further defined by[13]:

$$h = 2\frac{D}{v_o} + \frac{4(1+9k+51k^2/2)}{105(1+k)^2}\frac{v_o z_o^2}{D} + \frac{2k^3}{3(1+k)^2}\frac{v_o z_o^2}{F^2 c^2 D_1} \tag{2}$$

where D is the diffusivity of the sample gas in the mobile phase, v_o is the effluent output velocity, z_o is the column height, F is the ratio of the effective surface area of the stationary-phase relative to the actual area, c is the partition coefficient, and D_1 is the diffusivity of the stationary-phase. The numerical values and relationships utilized in the design of the SMGCS are summarized in Tables 1 and 2.

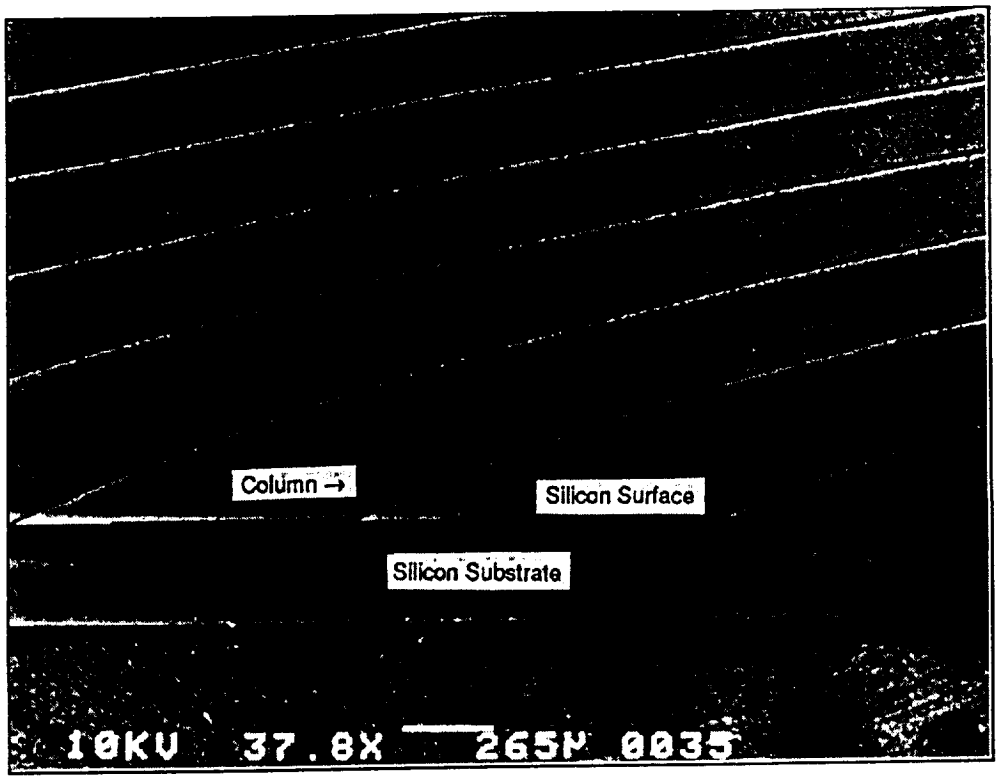
Table 1. Fundamental Micromachined Gas Chromatograph Design Relationships.

GC System Operational/Performance Parameter	Analytical Relationship
Plate Height	$h = 2\frac{D}{v_o} + \frac{4(1+9k+51k^2/2)}{105(1+k)^2}\frac{v_o z_o^2}{D} + \frac{2k^3}{3(1+k)^2}\frac{v_o z_o^2}{F^2 c^2 D_1}$
Partition Ratio	$k = k_o e^{(-E_a/k_b T)}$
Partition Coefficient	$c = k / F$
Diffusion Coefficient (Gas)	$D = \frac{v_{av}\lambda}{2}$
Output Velocity	$v_o = \frac{z_o(P_i^2 - P_o^2)}{6\mu L P_o}$
Pressure Correction Factor	$g = \frac{9\left[\left(\frac{P_i}{P_o}\right)^4 - 1\right]\left[\left(\frac{P_i}{P_o}\right)^2 - 1\right]}{8\left[\left(\frac{P_i}{P_o}\right)^4 - 1\right]^2}$

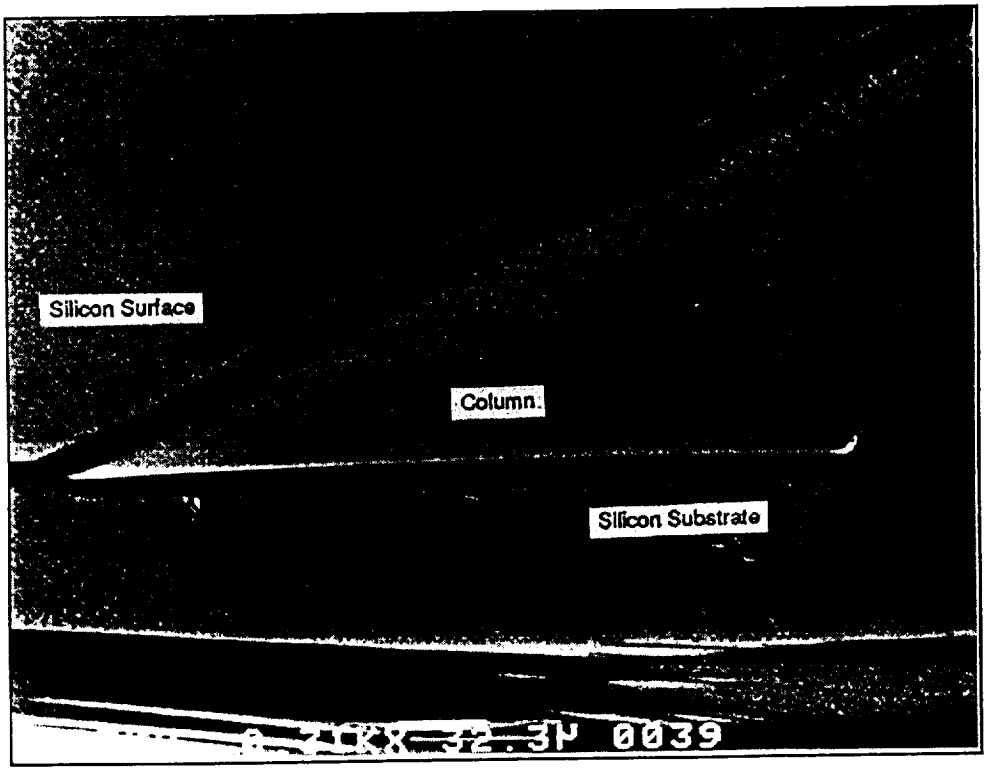
Table 2. Physical, Operational, and Experimental Parameters for the Silicon Micromachined Gas Chromatograph Design.

Physical Parameters		
Parameter	***Symbol***	***Value***
Boltzmann's Constant	k_b	1.38×10^{-23} joules/Kelvin
Helium Viscosity	μ	200 poise
Diffusion Coefficient	D	10^{-6} m^2/sec
Mean Free Path	λ	5×10^{-9} m
Average Molecular Velocity	v_{av}	400 m/sec
Operational Parameters		
Parameter	***Symbol***	***Value***
Input Pressure	P_i	40 psi (2.8×10^5 Pa)
Output Pressure	P_o	1 atmosphere (1×10^5 Pa)
Column Length	L	0.9 m
Column Width	Y	300 µm
Column Height	$2z_o$	10 µm
Temperature	T	80 °C
Column Permeability	q	2.6×10^7 poise/m^2
Pressure Correction Factor	g	1.08
Experimental Parameters		
Parameter	***Symbol***	***Value***
Heat of Adsorption	E_a	0.38 eV
Partition Ratio Constant	k_o	4×10^{-4}
Adsorption Lifetime	$1/\beta$	93 sec
CuPc Diffusion Coefficient	D_1	6×10^{-19} m^2/sec
Partition Coefficient	c	2400
Effective Surface Area Ratio	F	14
Crystallite Radius	r_σ	300 Å
Crystallite Height	h_σ	2000 Å

The critical component in the SMGCS is the micromachined capillary gas separation column, which is 0.9-m long and has a rectangular-shaped cross-section (300 µm width and 10 µm height). The aspect ratio of the capillary column's cross-section was designed using the analytical model developed by Golay[13] in conjunction with the limitations imposed by the IC photolithography and wet chemical etching processes. As depicted in Figure 2, the silicon wafer portion of the column was fabricated using conventional IC negative photoresist (Olin Hunt Specialty Products, Inc., Waycoat HR100 negative photoresist, West Patterson, NJ), an SiO_2 etch mask, and wet chemical isotropic etching



(a)



(b)

Figure 2. Scanning electron microscopy (SEM) micrograph of the isotropically etched portion of the GC capillary column micromachined in the silicon wafer (300 μm width and 9 μm depth). (a) expanded view. (b) detailed view.

($HF:HNO_3:CH_3COOH$, 2:15:5). The height of the column patterned in the silicon wafer was 9 μm. The balance of the column's height (1 μm) was correspondingly etched in the Pyrex® cover plate with a buffered hydrofluoric acid solution ($HF:NH_4F$, 1:4). Before depositing the column's CuPc stationary phase and electrostatically bonding the two components together, the silicon wafer was anisotropically etched (20 % wt KOH at 50 °C) to realize the injection port interface structure (Figure 3) and the dual-detector cavity (Figure 4), which was designed to maximize its performance by minimizing its dead volume (20 nl dead volume after the TCD was positioned).

The SMGCS column's CuPc (Fluke Chemical Corporation, Ronkonkoma, NY) stationary-phase (0.2 μm thickness) was sublimed (3 Å/sec deposition rate) under vacuum [He cryo-pumped, 10^{-6} Torr (1.3 x 10^{-4} Pa)] onto the surfaces of the etched silicon wafer and Pyrex® cover plate (Denton Vacuum, Inc., model DV-602, Cherry Hill, NJ). As shown in Figure 5, a commercially available polishing medium (PSI Testing Systems, Inc., 0.3 μm Al_2O_3 particle size, product number 16.3-6, Houston, TX), lubricated only with deionized water (to minimize contaminating the CuPc thin film) and secured on a marble plate, was used to selectively remove the CuPc thin film from the surfaces of the silicon wafer and Pyrex® cover plate, while leaving behind the desired stationary-phase on the column walls. The morphology of the resulting α-phase CuPc column coating was verified with scanning electron microscopy (SEM), transmission electron microscopy (TEM), and transmission electron diffraction (TED). Infrared (IR) spectroscopy was independently utilized to verify the chemical structure of the deposited CuPc coating (after deposition and after it was mechanically polished) and to investigate the selective adsorption of the ammonia (NH_3) and nitrogen dioxide (NO_2) analytes.

After the TCD was mounted in the dual-detector cavity, the silicon wafer and Pyrex® cover plate were electrostatically bonded[14] together (1800 V, 150 °C, 24-hour duration). The CuPc-coated chemiresistor was then aligned with its port on the dual-detector cavity. To complete the SMGCS fabrication, the sample injector (Figure 3) was connected to the column's input port with a short length (20 cm) of micro-capillary tubing (794 μm o.d., 254 μm i.d.; Valco Instruments Company, Inc., product T20N10D).

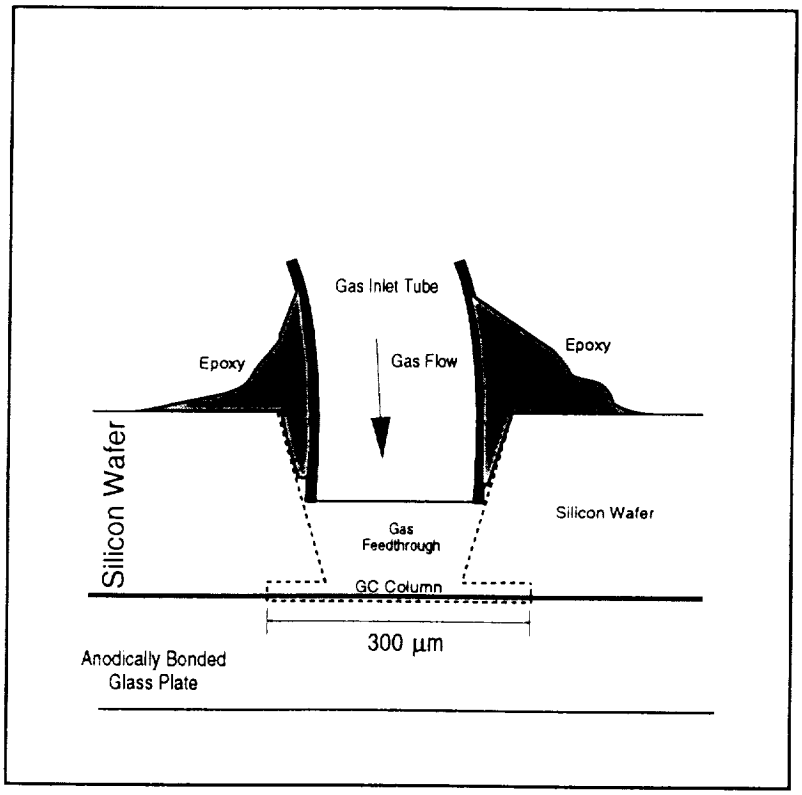


Figure 3. Side view of the anisotropically etched injection port interface structure design.

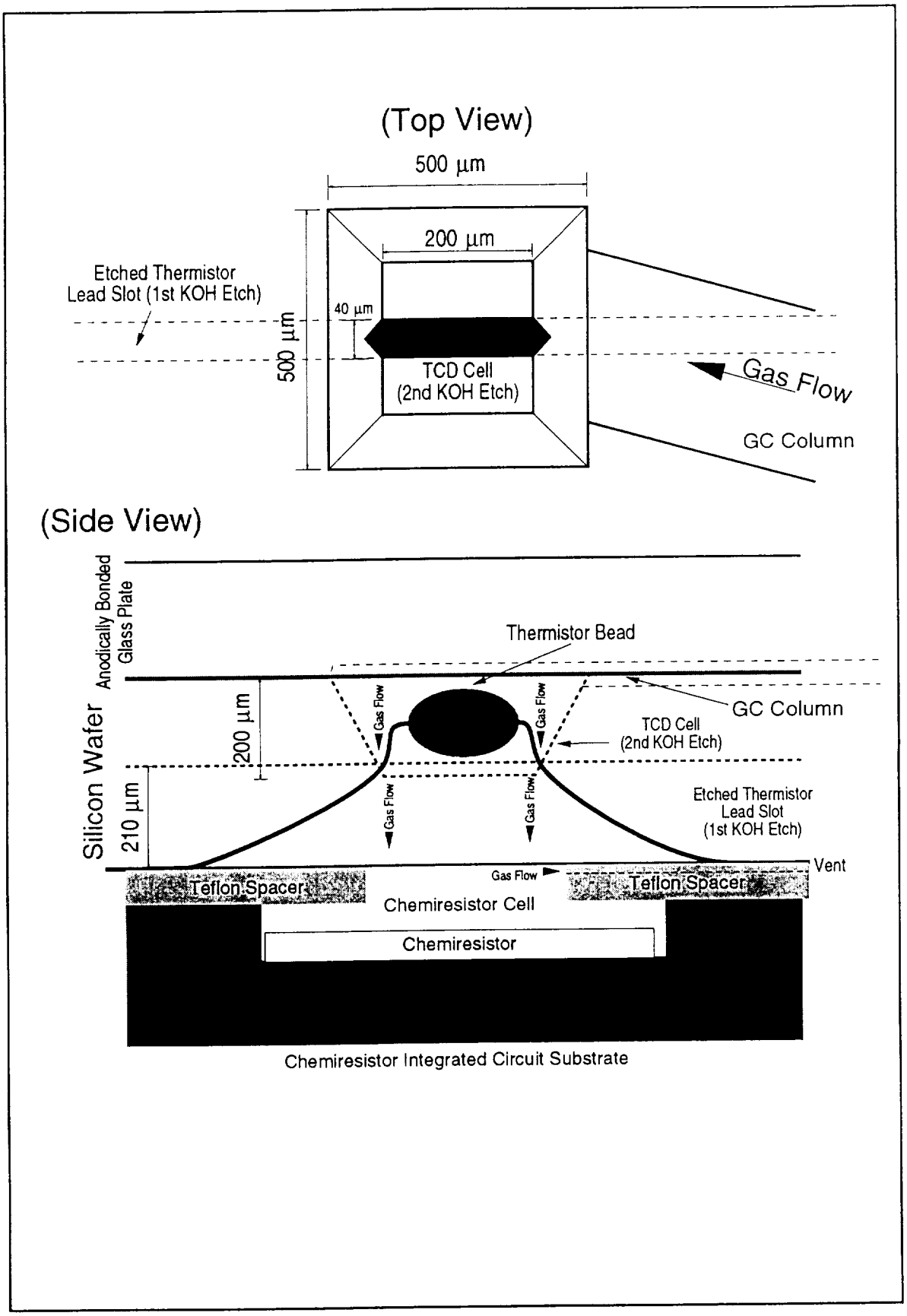
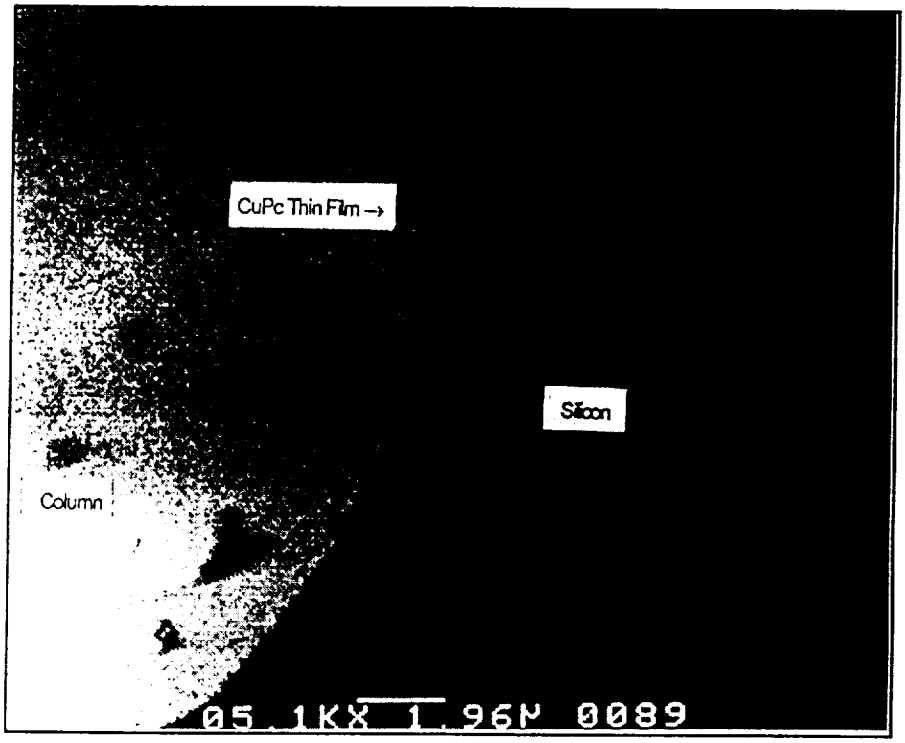
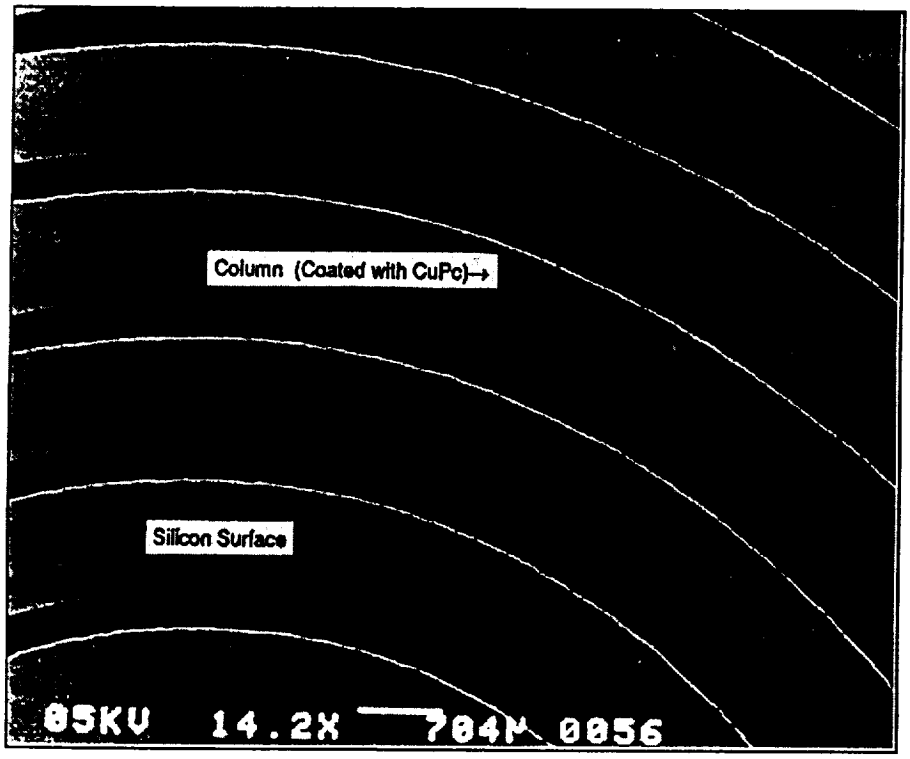


Figure 4. Thermal conductivity detector (TCD) cell and chemiresistor interface design.



(a)



(b)

Figure 5. SEM micrograph of the CuPc-coated portion of the GC column micromachined in the silicon wafer. (a) Column before the CuPc coating was removed from the flat surface regions (intra-column) of the silicon wafer. (b) Column after the CuPc coating was removed from the flat surfaces required to accomplish the electrostatic bonding process with the Pyrex® cover plate. (Similar results are achieved with the portion of the capillary column etched into the Pyrex® cover plate).

SMGCS Performance Evaluation

To evaluate the efficiency and separating power of the SMGCS, helium was used as a carrier gas, and binary mixtures of parts-per-million (ppm) concentrations of ammonia (NH_3) and nitrogen dioxide (NO_2) were realized using commercially available permeation tubes (GC Industries, Inc., models 23-7014 and 23-7052, Fremont, CA) and nitrogen (N_2) as the diluent. The 10 µl gas sample volumes were injected into the column with a pressure of 40 psi (2.8 x10^5 Pa) to maximize its theoretical separation factor[13]. To maintain the integrity of the electrostatic bond between the silicon wafer and the Pyrex® cover plate, isothermal operation of the SMGCS was limited to temperatures spanning 55 °C to 80 °C. In operation, the TCD was used to capture the diluent nitrogen gas peak, and since the gas specific, CuPc-coated chemiresistor behaves as an integrating detector, its response was differentiated to establish the peaks associated with NH_3 and NO_2. Since NH_3 is an electron-donor gas, and NO_2 is an electron-acceptor species, their converse electrical interactions with the chemiresistor's *p*-type CuPc semiconductor coating motivated generating chromatograms that depict the absolute value of the detector's differentiated response (to facilitate their interpretation consistent with the format of conventional GC data).

Under isothermal operating conditions, Figures 6 and 7 illustrate the performance of the SMGCS to separate and detect NH_3 and NO_2 when their concentrations are systematically varied. Isothermal operation at 80 °C requires less than 30 minutes to process a gas sample. Using the numerical data summarized in Tables 1 and 2, along with equations (1) and (2), the theoretically calculated separation factor (SF), as shown in Table 3, agrees reasonably well with the values extracted from the measured chromatograms.

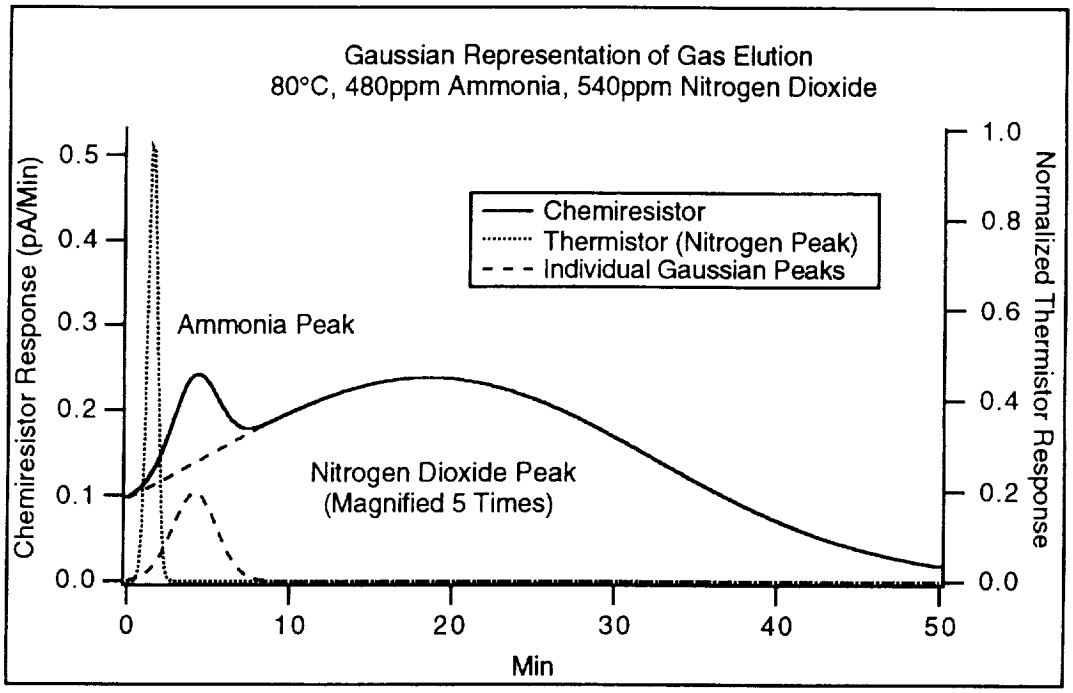
Table 3. Comparison of the SMGCS Theoretical and Experimental Values of the Separation Factor (SF).

Operating Temperature (°C)	Theoretical SF	Experimental SF
55	4.4	3.8
66	3.3	3.4
76	2.8	2.7
80	2.1	1.9

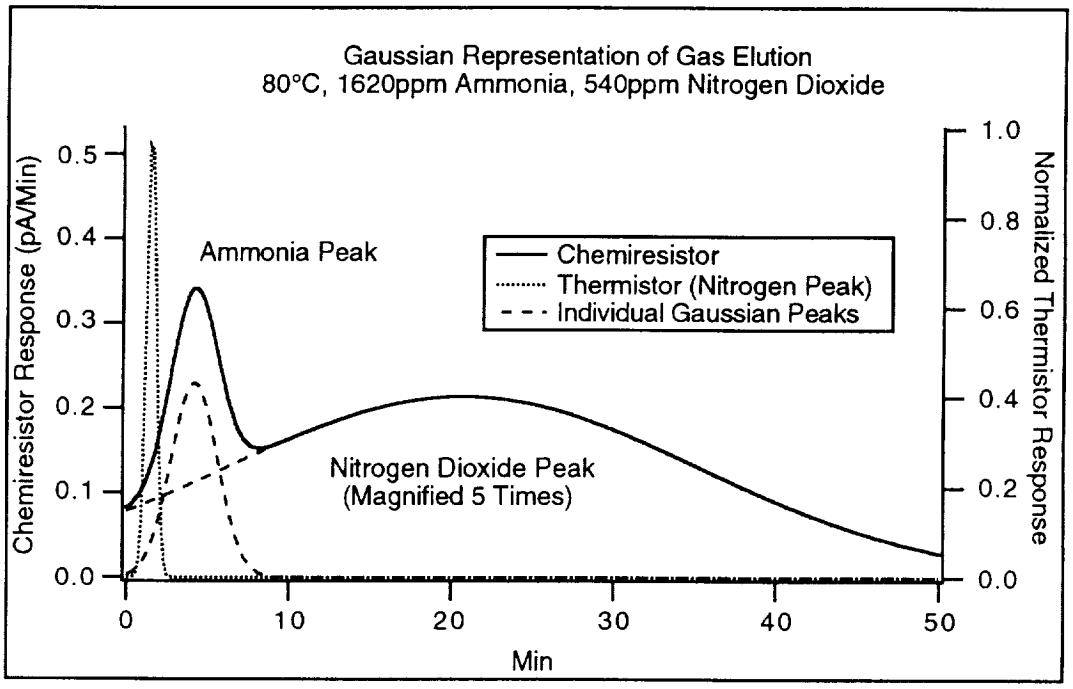
Conclusion

The results of this investigation demonstrate the viability of using single-crystal silicon micromachining for implementing a GC system and using a CuPc thin-film stationary-phase which is capable of separating and detecting NH_3 and NO_2. The significant accomplishments ascertained from this research can be summarized in two areas: micromachining and chemical sensing.

In the micromachining area, a novel TCD cell design was implemented. The critical features of this design, compared to those previously reported[1-3], include: reduced dead volume (less than 20 nl), ease of thermistor insertion, and the ability to pass the column effluent to another detector (e.g., the chemiresistor). Also, a new technique was developed, enabling, for the first time, the deposition of a nearly-homogeneous thin-film stationary-phase (2000 Å thick) within the micromachined GC column. Finally, the ability to perform a low temperature (less than 300 °C) anodic bond (1800 V for 24 hours)

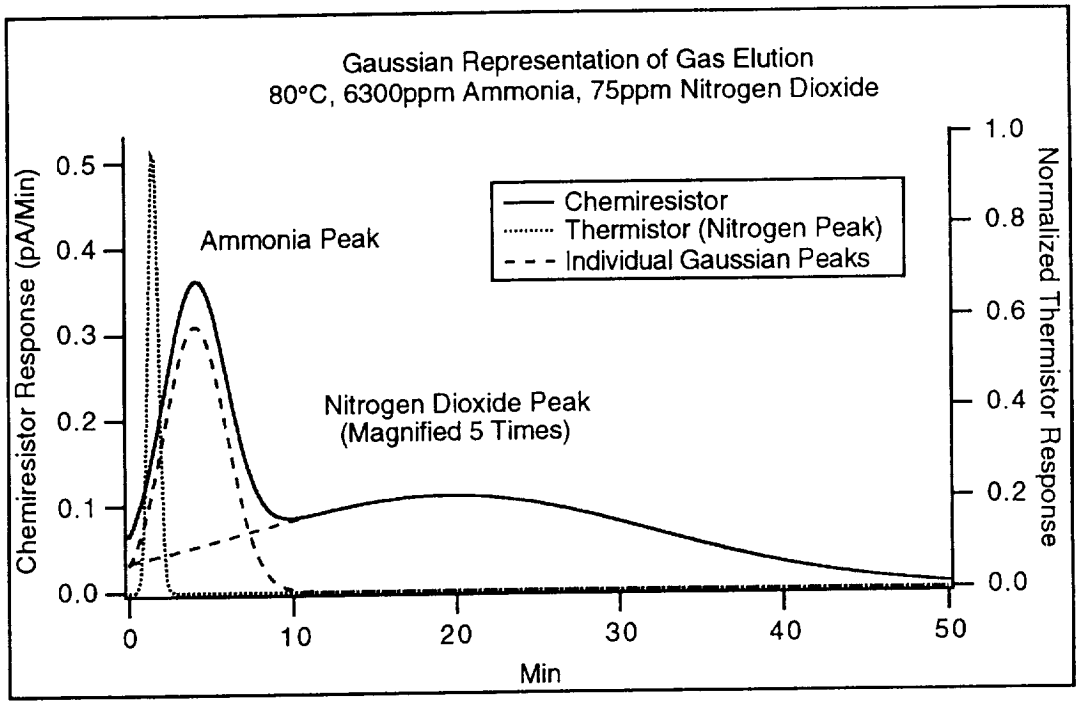


(a)

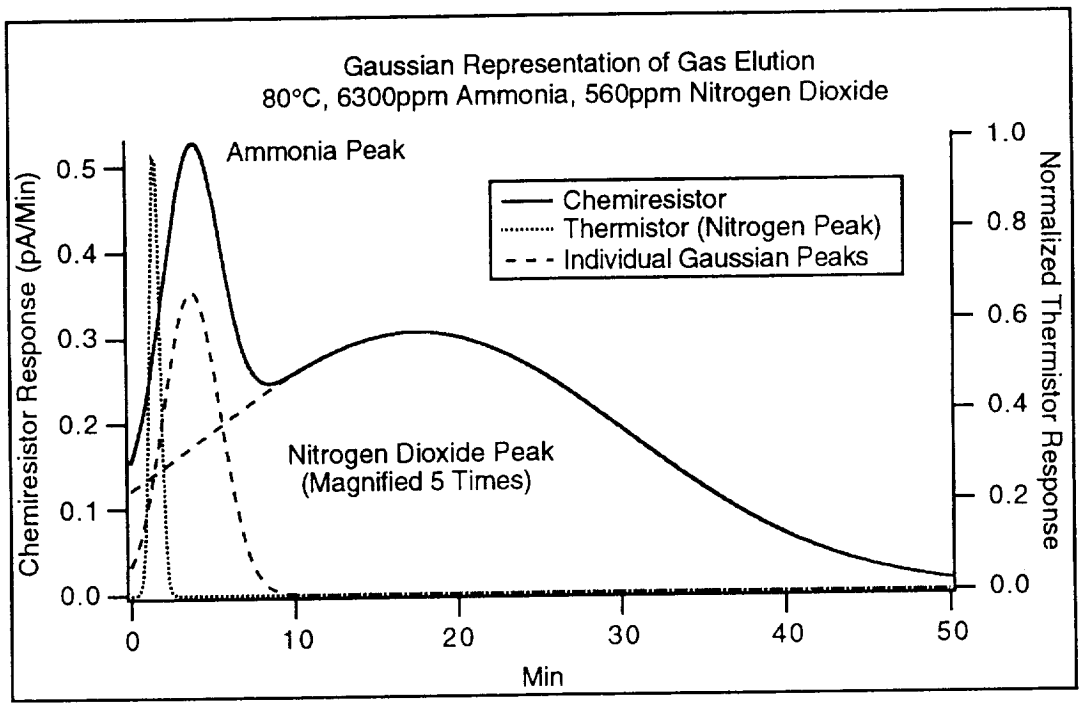


(b)

Figure 6. Chromatograms obtained with the micromachined GC system operated at 80 °C for two different NH_3 concentrations for a fixed NO_2 concentration (540 ppm). (a) 480 ppm NH_3. (b) 1620 ppm NH_3.



(a)



(b)

Figure 7. Chromatograms obtained with the micromachined GC system operated at 80 °C for two different NO_2 concentrations for a fixed NH_3 concentration (6300 ppm). (a) 75 ppm NO_2. (b) 560 ppm NO_2.

was demonstrated, which is inherently important because it is compatible with the thermally-sensitive thin films and other bulk materials.

The primary accomplishment was the separation and detection of NH_3 and NO_2. The SMGCS was capable of separating NH_3 and NO_2 at parts-per-million concentration levels in less than 30 minutes when operated at 80 °C. Furthermore, this research served as a proof-of-concept for using a SMGCS to investigate the adsorptive properties of other thin films.

With respect to future research, improvements are being implemented concerning the column design and detector configuration. The micromachined GC column's length can readily be increased by decreasing the inter-column spacing without significantly sacrificing yield, resulting in a proportional increase in the separation factor (SF). Also, incorporating the chemiresistor IC directly within the detector cell, similar to the cavity specifically micromachined for the TCD, should improve the sensitivity of the chemiresistor (a smaller dead volume implies a higher localized analyte concentration). An integral heater (with controller) and micromachined sample injection valve could also be incorporated into the SMGCS design using standard IC fabrication techniques, further reducing the requirement for external equipment. Finally, investigations concerning the adsorptive properties of other thin films (in particular, other metal-doped phthalocyanines), using the SMGCS as a tool, should be of significant value to those developing chemical sensors based upon these materials.

Acknowledgments

The authors acknowledge the support received from Texas Christian University, Research and Sponsored Projects Program, grant 5-23762, Fort Worth, TX.

References

1. S.C. Terry and J.B. Angell, "A Column Gas Chromatography System on a Single Wafer of Silicon," in P.W. Cheung, D.G. Fleming, M.R. Neuman, and W.H. Ko (eds.), Theory, Design, and Biomedical Applications of Solid State Chemical Sensors, pp. 207-218, CRC Press, Boca Raton, FL, 1978.

2. S.C. Terry, J.H. Jerman and J.B. Angell, "A Gas Chromatographic Air Analyzer Fabricated on a Silicon Wafer," *IEEE Transactions on Electron Devices,* vol. ED-26, pp. 1880-1886, December 1979.

3. J.H. Jerman and S.C. Terry, "A Miniature Gas Chromatograph for Atmospheric Monitoring," *Environmental International,* vol. 5, pp. 77-83, 1981.

4. S. Saadat and S.C. Terry, "A High-Speed Chromatographic Gas Analyzer," *American Laboratory,* vol. 5, pp. 90-101, 1984.

5. J.B. Angell, S.C. Terry and P.W. Barth, "Silicon Micromechanical Devices," *Scientific American,* vol. 248, p. 44, 1983.

6. S.C. Terry and J.H. Jerman, "Miniature Gas Chromatograph Apparatus," U.S. Patent 4,474,884, October 2, 1984.

7. A. van Es, J. Janssen, R. Bally, C. Cramers and J. Rijks, "Sample Introduction in High Speed Capillary Gas Chromatography - Input Band Width and Detection Limits," *J. High Resolution Chromatography and Chromatography Communications,* vol. 10, pp. 273-279, 1987.

8. G. Lee, C. Ray, R. Siemers and R. Moore, "Recent Developments in High Speed Gas Chromatography," *American Laboratory,* vol. 21, pp. 108-119, 1989.

9. R. Siemers, D. Heigel and A. Spilkin, "Rapid Micro-GC Analysis of Permanent Gases," pp. 44L-44R, *American Laboratory,* vol. 23, 1991.

10. S. Santy, A. Spilkin and J. Strauss, "Rapid Analysis of Chlorofluorocarbons Using a Micro Gas Chromatograph," *American Laboratory*, vol. 23, p. 34, 1991.

11. E.M. Vary, "Investigation of Phthalocyanines Using Gas Chromatography," PhD Dissertation, University of California, Los Angeles, CA, 1966.

12. J.H. Purnell, "The Correlation of Separating Power and Efficiency of Gas Chromatographic Columns," *Chemical Society Journal*, vol. 61, pp. 1268-1274, 1960.

13. M.J. E. Golay, "Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular Cross-Sections," in D.H. Desty (ed.), Gas Chromatography, p. 36, Academic Press, NY, 1958.

14. G. Wallis and D.I. Pomerantz, "Field-Assisted Glass-Metal Sealing," *J. Appl. Phys.*, vol. 40, pp. 3946-3949, 1969.

A 32-bit Ultrafast Parallel Correlator Using Resonant Tunneling Devices

Shriram Kulkarni, Pinaki Mazumder, and George I. Haddad
Department of Electrical Engineering and Computer Science
The University of Michigan
1301 Beal Avenue, Ann Arbor, MI 48109-2122

Abstract

An ultrafast 32-bit pipelined correlator has been implemented using resonant tunneling diodes (RTDs) and hetero-junction bipolar transistors (HBTs). The negative differential resistance (NDR) characteristics of RTDs is the basis of logic gates with the self-latching property that eliminate pipeline area and delay overheads which limit throughput in conventional technologies. The circuit topology also allows threshold logic functions such as minority/majority to be implemented in a compact manner resulting in reduction of the overall complexity and delay of arbitrary logic circuits. The parallel correlator is an essential component in CDMA transceivers used for the continuous calculation of correlation between an incoming data stream and a PN sequence. Simulation results show that a nano-pipelined correlator can provide an effective throughput of one 32-bit correlation every 100 ps, using minimal hardware, with a power dissipation of 1.5 watts. RTD+HBT based logic gates have been fabricated and the RTD+HBT based correlator is compared with state of the art CMOS implementations.

1. Introduction

Space based communication systems experience high signal-noise (S/N) ratio in the transmission channel and have inherently low power budgets for communication. An added constraint is the requirement of high reliability and security for space to earth transmissions due to their vital nature in supporting military and civilian systems. Spread spectrum communication increases transmission bandwidth by distributing the data signal energy over a large frequency band by use of a pseudo-noise (PN) spreading sequence. The uniqueness of the PN sequence results in receivers being able to detect the transmitted signal even in the high noise environments due to low cross correlation with extraneous transmissions. Thus, the required transmitter power is reduced in spread spectrum systems. Spread spectrum signals have low probability of detection by unintended receivers and hence provide good security. Similarly, the redundancy in the spread spectrum signal allows for reliable communication. Hence, spread spectrum modulation satisfies the constraints imposed by space based communication systems.

The parallel correlator forms an essential component in a digital communication system. Typically, in spread spectrum systems, a parallel correlator computes the correlation of the incoming data stream with a pre-determined pseudo-noise (PN) sequence of a fixed length. This correlation value is used to estimate the output data. For a binary input data stream, the result of such an operation essentially determines whether the output should be 0, 1 or indeterminate. An indeterminate output is primarily caused due to the receiver PN sequence not being the same as the transmitter sequence. Thus, communication between different transceivers can be regulated on the

This work was supported by the U.S. Army Research Office under URI program contract DAAL03-92-G-0109 and by ARPA under contract DAAH04-93-G-0242

basis of PN sequence uniqueness. This provides the capability of rejecting interference from multiple transmission paths and jamming [1].

The correlator as described in this paper is particularly suited for direct sequence spread spectrum systems that use binary phase shift keying as the digital modulation. Figure 1 shows the essential function of the spread spectrum demodulator along with waveforms for desired signal reception and jamming signal rejection. The serial input data stream is shifted with each clock cycle and correlation is performed between the fixed PN sequence and as many stored bits of the input data stream as the length of the PN sequence. The ability of the system to respond only to the spreading code while rejecting others makes it useful in systems that experience jamming and multipath interference. The same feature is the basis of code division multiple access (CDMA) systems that allow multiple users to carry out independent messaging in a single spectrum band. The correlation value between the incoming data stream and the PN sequence has to be generated at each clock cycle. If a purely combinational circuit along with a shift register were chosen to implement the correlator, for long PN sequences, it would result in extremely slow operation due to many levels of logic required for computation of correlation. However, in a bit serial communications application as described in this paper, there is no data dependence and hence deep pipelining schemes can be effectively used to improve the throughput of the correlator.

2. Theoretical development

From a hardware viewpoint, correlation between the two binary streams can be represented as follows.

$$\psi(\tau) = \Sigma(f(t) \oplus g(t-\tau)) \tag{1}$$

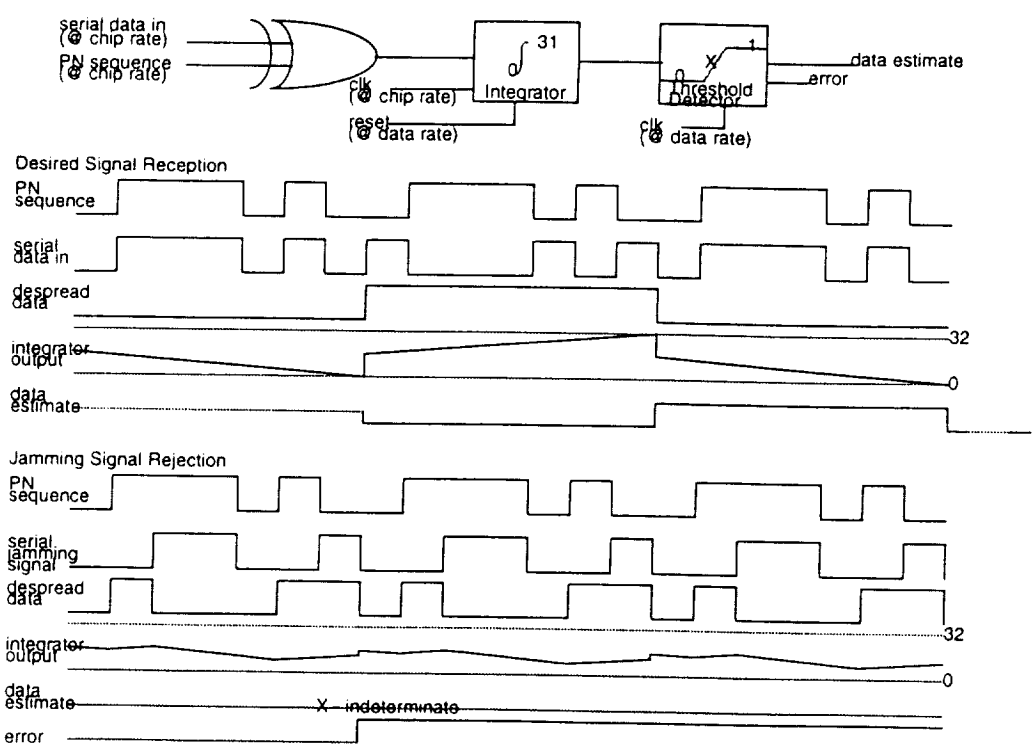


Figure 1. Spread spectrum demodulator

Here, $f(t)$ and $g(t)$ are binary data streams which specifically represent the PN sequence and the input data stream for discussion of the parallel correlator. The XOR operator correlates two binary inputs i.e. it produces a logic 1 output only if the two signals are unlike. The summation of the XOR outputs over the length of the signals gives a measure of the likeness between the two signals. The difference between the number of 1s and the number of 0s in the correlation vector will result in a number that ranges from the negative of the PN sequence length, through 0, up to the positive of the PN sequence length reflecting a 0, indeterminate and 1 output respectively. Thresholds can be set for 0 and 1 detection to account for noise in the channel. The above number, henceforth referred to as the *correlation value*, can also be written as follows.

$$\textit{Correlation Value} = 2\cdot(\Sigma \textit{ of 1s}) - (\textit{PN sequence length}) \tag{2}$$

3. RTD-HBT logic family

The current-voltage characteristics of an RTD can be approximated by the piecewise linear form shown in Figure 2.

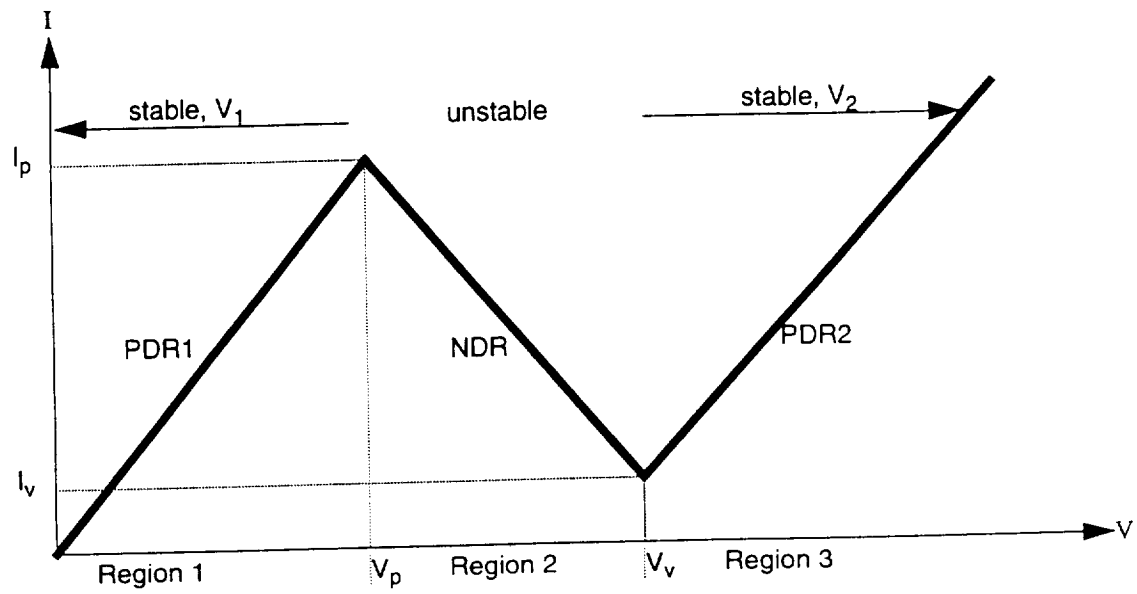


Figure 2. Piecewise approximation of RTD characteristics

As the voltage applied across the device terminals is increased from zero, the current increases until the V_p, the peak voltage of the RTD. The corresponding current is call the peak current, I_p of the RTD. As the voltage across the RTD is increased beyond V_p, the current through the device drops abruptly due to tunneling until the voltage reaches V_v, the valley voltage. The current at this voltage is the valley current, I_v. Beyond V_v, the current starts increasing again. For a current in $[I_v, I_p]$ there are two possible stable voltages; $V_1 \leq V_p$ or $V_2 \geq V_v$. The tunneling characteristic of the RTD facilitates implementation of self latching circuits.

3.1 Bistable mode operation

A binary logic circuit is said to operate in bistable mode when its output is latched, and any change in the input is reflected in the output only when a clock or other evaluation signal is applied. The bistable mode has been used in several earlier technologies, notably in superconducting logic [2]. Superconducting logic typically uses a multi-phase AC power source to periodically

reset/evaluate each gate. Similar logic using resonant tunneling devices has been proposed by several authors [3, 4, 5]. The chief disadvantage of these circuits is the requirement of an AC power source whose frequency determines the maximum switching frequency. The RTD+HBT logic circuits described below use a DC power supply and multiphase clocks but the clock signals are not required to supply large amounts of power as in the case of the earlier circuits.

The operating principle of the new bistable element may be understood by considering the circuit shown in Figure 3. There are m input transistors and one clock transistor driving a single RTD load. The input transistors can be in either of two states - On, with a collector current of I_H or Off, with no collector current. The clock transistor can be in one of two states - High, with collector current I_{CLKH}, and Quiescent with collector current I_{CLQ}. In addition, there is a global reset state where all the collector currents are 0. When the clock transistor current is at I_{CLKQ}, the load lines in Fig. 1 show that the circuit has two possible stable operating points for every possible input combination. When the clock current is I_{CLKH}, there is exactly one stable operating point for the circuit when n or more inputs are high and the sum of the collector currents is $nI_H + I_{CLKH}$. This operating point corresponds to a logic 0 output voltage. Hence this circuit can be operated sequentially to implement any non-weighted threshold logic function $f(x_1, x_2, ..., x_m;n)$, where $f(x_1, x_2, ..., x_m)$ is 1 if and only if $(x_1 + x_2 + ... + x_m) < n$, and $x_1, x_2, ..., x_m$ take on values of either 0 or 1.

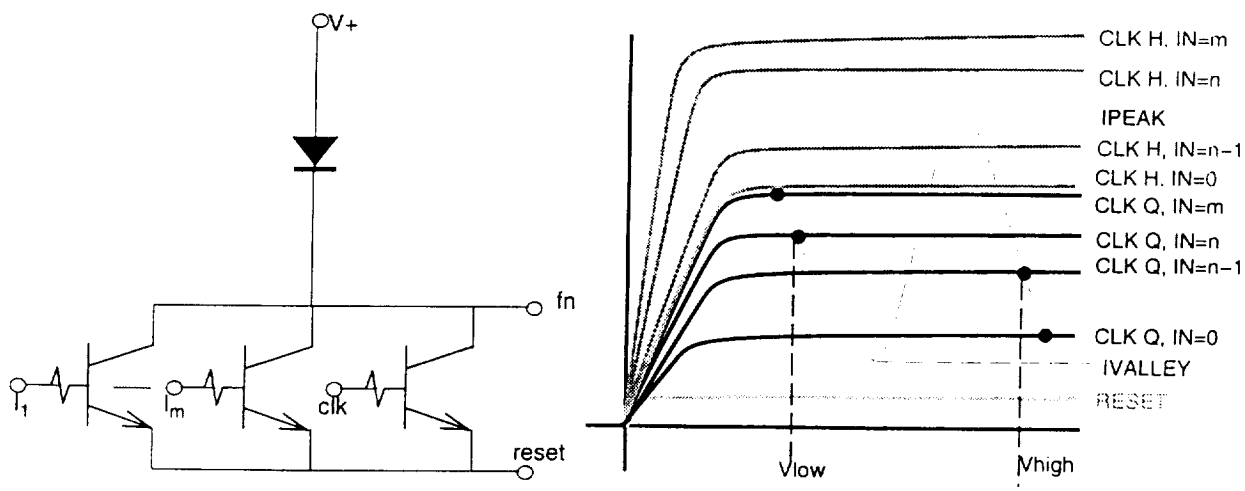


Figure 3. RTD+HBT bistable logic gate operating principle

The operating sequence is as follows:

1. Inputs I_1 through I_m change.
2. The *reset* line goes high forcing all transistors into cut-off. The current through the RTD falls below the valley current, and the *fn* node is pulled high.
3. The *reset* line goes back to 0. The *fn* node remains high.
4. The *clk* signal goes high, causing the total current through the RTD to increase. If more than n inputs are high, the current through the RTD exceeds the peak current causing a jump to the second positive differential resistance (PDR) region of the RTD characteristic corresponding to $V_{RTD} > V_{VALLEY}$, where V_{RTD} is the voltage across the RTD and V_{VALLEY} is the valley voltage of the RTD. This results in the *fn* node going low. If less than n inputs are high the current through the RTD does not exceed the peak current and the operating point remains in the first PDR region of the RTD,

where $V_{RTD} < V_{PEAK}$, and V_{PEAK} is the RTD peak voltage. Thus, *fn* remains high.

5. The *clk* signal goes to its quiescent state so that the current through the clock transistor is I_{CLKQ}. The output voltage at node *fn* reaches a stable level corresponding to whether the RTD was in the first PDR region or the second PDR region in the previous step of the sequence.

For a three input circuit, three non-trivial threshold functions can be implemented for the cases where $n = 1, 2, 3$. For $n = 1, f_1(x_1, x_2, x_3) = 0$ if and only if 1 or more inputs are high. This corresponds to a NOR function. For $n = 3, f_3(x_1, x_2, x_3) = 0$ if and only if all 3 inputs are high. This corresponds to a NAND function. For $n = 2, f_2(x_1, x_2, x_3) = 0$ if and only if 2 or more inputs are high. This corresponds to an inverted majority or inverted carry function.

Figure 4 shows the simulated traces obtained from NDR-SPICE [6] for an inverter, a three input NOR, and a three input MINORITY gate designed using RTDs and HBTs. It can be seen that the outputs change only on arrival of the clock pulse and hence the circuits are operating in bistable mode. Input and output voltage swings are matched to enable cascaded circuits to function correctly. The signal levels are 1V for logic zero and 2V for logic one.

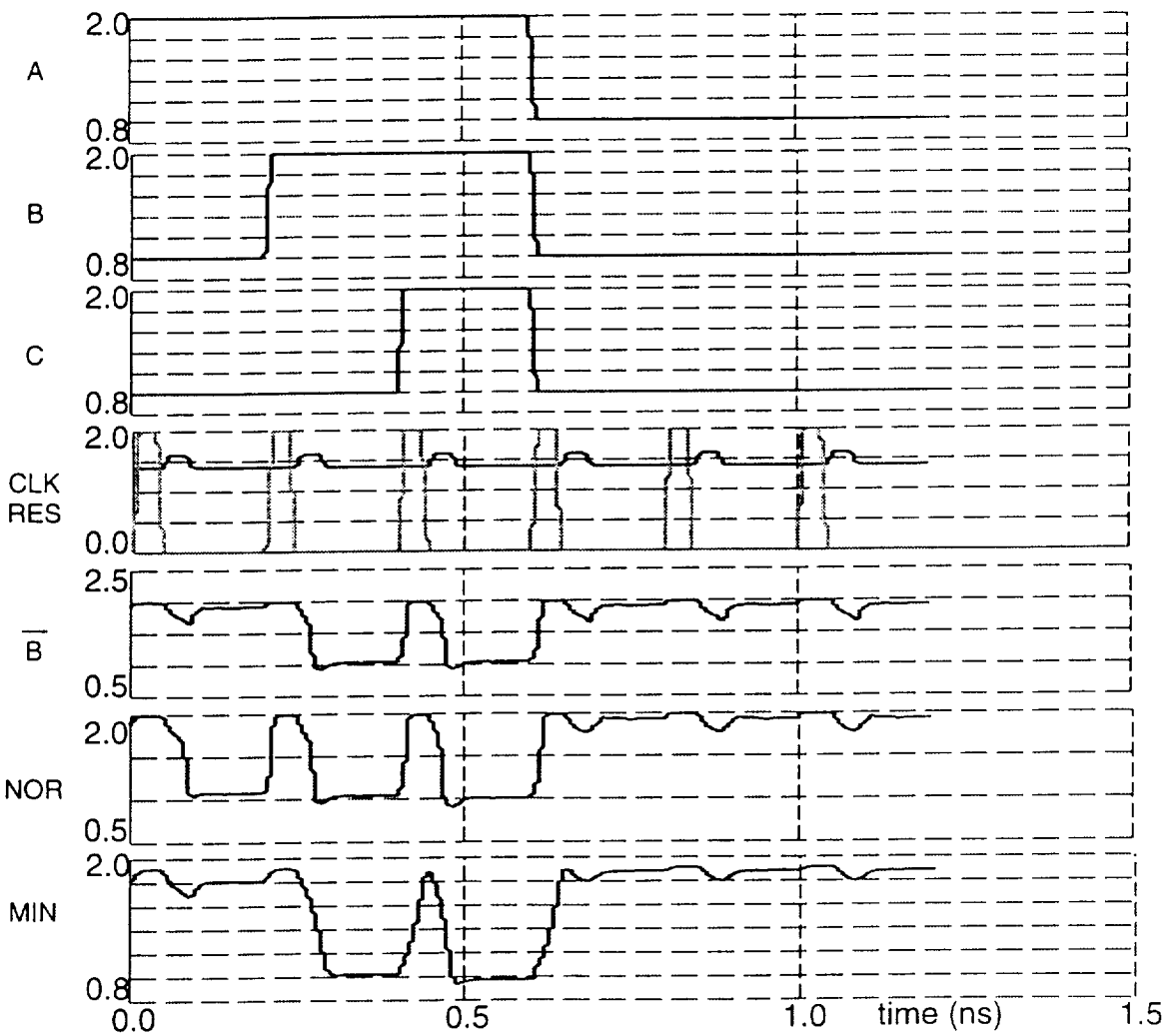


Figure 4. RTD+HBT basic gates simulation

3.2 Design constraints

We now present the design equations for a k input threshold gate with a threshold value of n. Let m be the are of the RTD used and let J_P and J_V represent the peak and valley current densities of the RTD, respectively. I_H, I_{CLKH} and I_{CLKQ} are defined as in section 3.1. The design constraints for the aforementioned gate can be written as:

$$h = mJ_P - (I_{CQ} + kI_H) > 0 \tag{3}$$

$$l = I_{CQ} - mJ_V > 0 \tag{4}$$

$$hh = mJ_P - (I_{CH} + (n\text{-}1)I_H) > 0 \tag{5}$$

$$hl = I_{CH} + nI_H - mJ_P > 0 \tag{6}$$

$$I_{CH} > I_{CQ} > 0 \tag{7}$$

where,

h = quiescent clock, logic high switching margin

l = quiescent clock, logic low switching margin

hh = high clock, logic high switching margin

hl = high clock, logic low switching margin

The design process begins by choosing the input high and low voltages. The input high and low voltages must respectively turn the input transistors on or off. To maintain good noise margins, signal voltage swings should be maximized. However, for cascaded logic stages to operate correctly without resorting to use of level shifters, it is necessary to match the input and output voltage swings. An optimum match resulted in the signal voltage levels being set to 1V for logic 0 and 2V for logic 1. The input transistor size determines the value of I_H. I_H should be small to minimize power consumption and area, but should be large enough to have good switching margins hh and hl. The value of I_{CLKQ} and the area of the RTD are determined from the equations involving I_{CLKQ}. The peak and valley current densities (J_P and J_V) are determined by the growth process, and the RTD area factor m determines the actual currents. The simulations in this paper use an RTD with peak current of 100 μA and a valley current of 25 μA for m = 1. Setting $I_{CLKQ} = (m(J_P + J_V) - kI_H)/2$ satisfies both equations (3) and (4), when m is chosen such that $m > 3I_H/(J_P - J_V)$. This also results in the equalization of the switching margins h and l. Choosing $I_{CLKH} = mJ_P - (n - 0.5)I_H$ satisfies the remaining design equations and also equalizes the switching margins hh and hl. The clock line voltages and the clock transistor sizes are determined from the values of I_{CLKQ} and I_{CLKH}. The switching margins for the circuits are $0.5I_H$ or a 50% variation is allowable in the drain current of any one input transistor. When all transistors are systematically larger or smaller, the allowable variation before the circuit malfunctions is $0.5I_H/n$. For a NOR gate, n = 1 and the allowable variation is 50%. For a 3-input inverted majority gate the allowable variation is 25% and for a 3-input NAND gate it is 16%. Thus, the switching margin of a NOR gate remains constant with increase in the number of inputs whereas, the switching margin of a NAND gate degrades rapidly with increase in the number of inputs. Thus, the best design margins are provided by the NOR function and the NAND function should be avoided in so far as possible.

3.3 Co-integration of RTDs and HBTs

RTDs and HBTs were integrated on the same wafer to build a 3-input threshold gate with

the same topology as the circuit shown in Figure 3. Figure 5 shows a photomicrograph of the integrated circuit. The functionality of the circuit is determined by the input and clock voltages as discussed in section 3.2. By adjusting the values of the supply voltage, input high voltage and clock voltages NAND, NOR and MINORITY functions were tested and the oscilloscope traces are shown in Figure 6. It should be noted that in the correlator design, the signal voltages are fixed and hence functionality of the gates is determined by the device sizes.

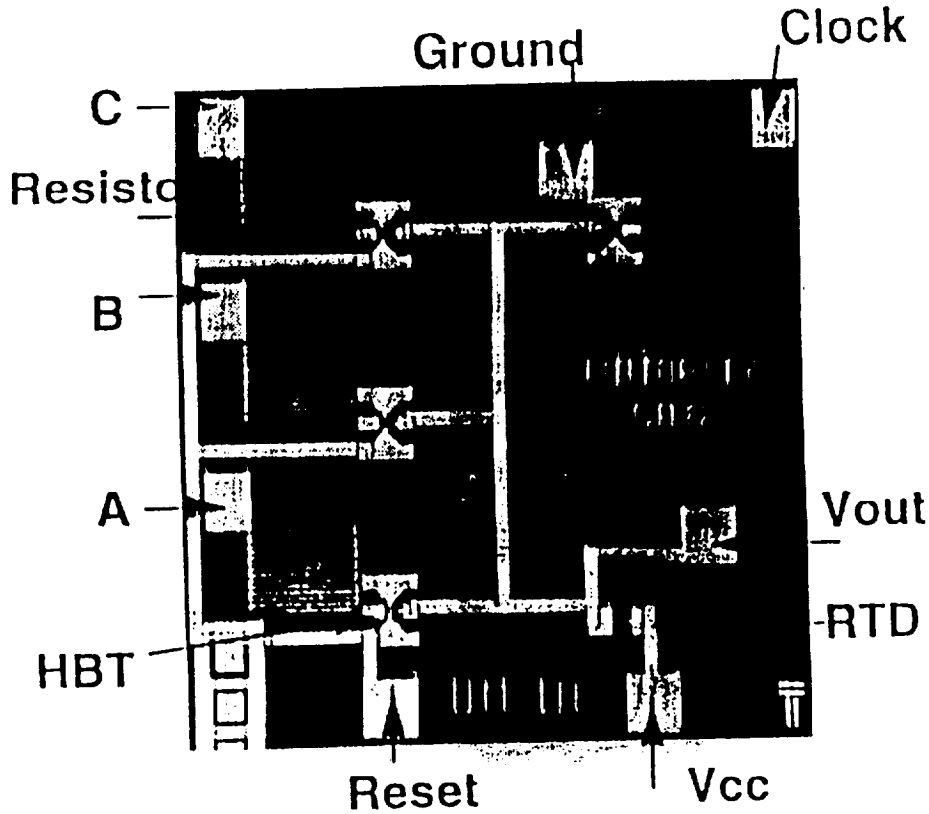


Figure 5. Photomicrograph of RTD+HBT bistable gate

3.4 Pipelined computation

Pipelining is a well studied means of speeding up any computation. An existing combinational block is divided into several sequential stages such that each stage performs a different operation during a particular clock cycle. The drawback of pipelining is that each computation takes the same or more time as nanopipelining [7] but there is an added penalty in the area devoted to the pipeline latches in the circuit.

Consider a combinational block that is composed of n stages with each stage having a delay of t_c. This results in a total delay of $n \cdot t_c$. We could partition the combinational block into k stages from 1 to n, where each stage output is latched. If we assume a latch delay of t_l, the maximum delay of the circuit is now $(n \cdot t_c/k + t_l)$. The throughput of the circuit increases from $1/(n \cdot t_c)$ to $1/(n \cdot t_c/k + t_l)$ but the latency increases from $n \cdot t_c$ to $n \cdot (t_c + t_l)$. Also, if a_c is the area of the combinational block; after pipelining, the area of the circuit increases to $a_c + k \cdot m \cdot a_l$, where a_l is the area of a latch and m is the number latches at each stage. The best possible theoretical throughput would be $1/t_c$ when we have latches at the output of each combinational stage. However, if all combinational stages don't have the same delay, then the maximum achievable throughput with the use of separate pipeline latches is $1/(b \cdot t_c + t_l)$ where $b \cdot t_c$ is the longest combinational stage

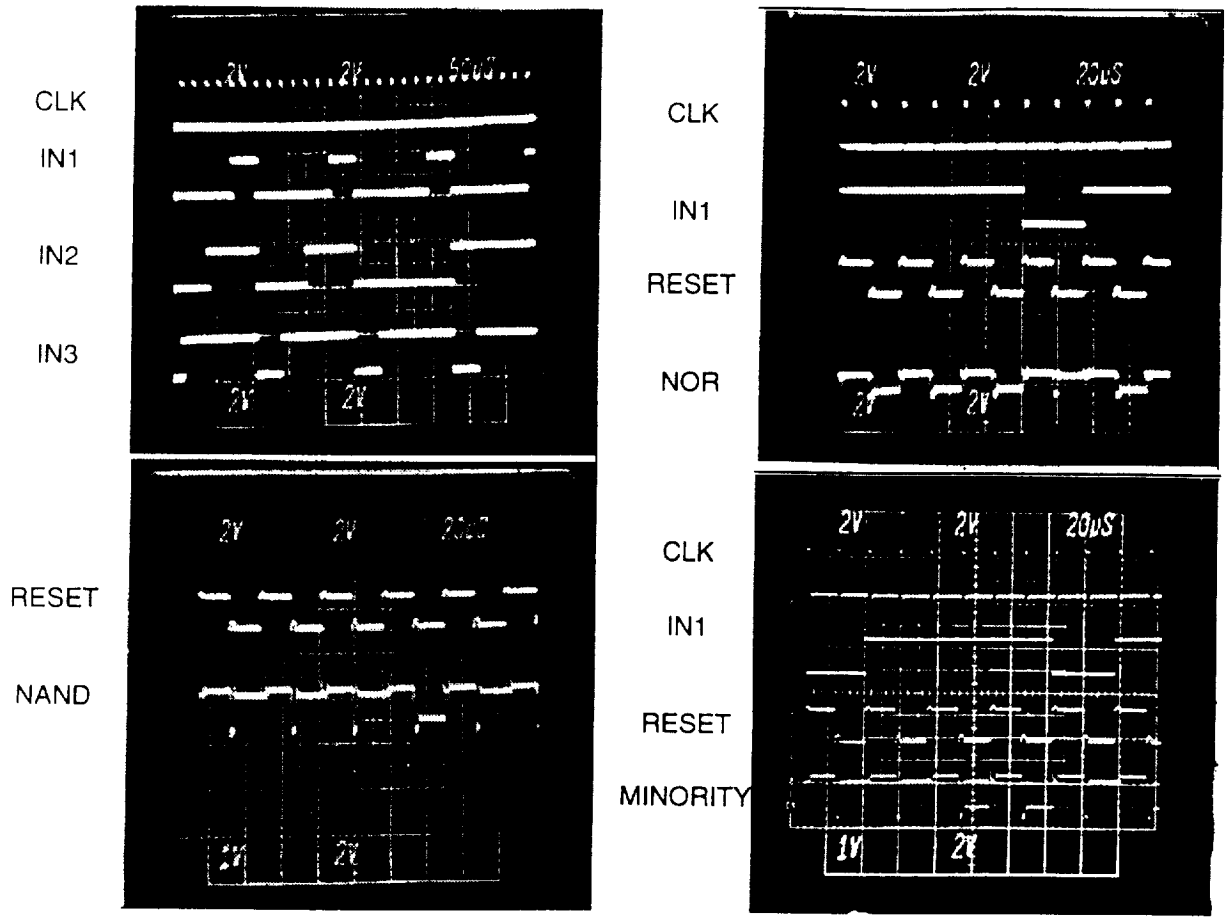


Figure 6. Oscilloscope traces for fabricated RTD+HBT primitive gates

delay. If the latch delay t_l is much larger than the longest combinational delay $b \cdot t_c$, it places an upper bound on the maximum achievable throughput of the pipelined circuit. Thus, we see that pipelining using conventional logic results in direct trade-offs between the area of the pipeline latch and the achievable throughput. The use of bistable NDR devices in designing circuits improves the performance of nanopipelined circuits over conventional pipelined circuits because the latch delay, t_l=0. Also, if latency is not of concern, each logic gate can operate in the bistable mode resulting in maximum possible throughput.

3.5 Nanopipelined full adder implementation

The basic bistable logic gates mentioned previously are used to build a nanopipelined full adder that best illustrates the advantages of the NDR logic family. For the parallel correlator, we prefer an adder with complementary sum and carry outputs in order to reduce the number of pipeline stages and hence the latency of the circuit. The complementary sum and carry functions for a 1-bit full adder are written as follows.

$$\overline{S} = \overline{a \oplus b \oplus c_{in}} \tag{8}$$

$$\overline{C} = \overline{a \cdot b + b \cdot c_{in} + c_{in} \cdot a} \tag{9}$$

The $\overline{S}$ function is implemented as a three level nanopipelined circuit whereas the $\overline{C}$ func-

tion is implemented using a single minority gate. The circuit for the 1-bit full adder is shown in Figure 7. It is apparent that the $\overline{S}$ and $\overline{C}$ outputs are not synchronized with each other. For a single stage of addition, we would need to add two bistable buffers at the $\overline{C}$ output to synchronize the $\overline{S}$ and $\overline{C}$ outputs. However, in the correlator we perform several successive stages of addition and synchronization at each adder will result in increased latency. Hence, synchronization is performed after all stages of addition are complete. For correct operation of the true-bistable logic gates a reset and evaluate pulse is required as mentioned previously. However, when multiple gates are cascaded, as in the implementation of the full adder, a gate must be evaluated only after all its inputs have been correctly evaluated. This requires a two-phase evaluation scheme in which each gate is evaluated in a different phase than its fan-ins and fan-outs. An example timing relationship between phases of consecutive logic blocks for the parallel correlator is illustrated in Figure 8.

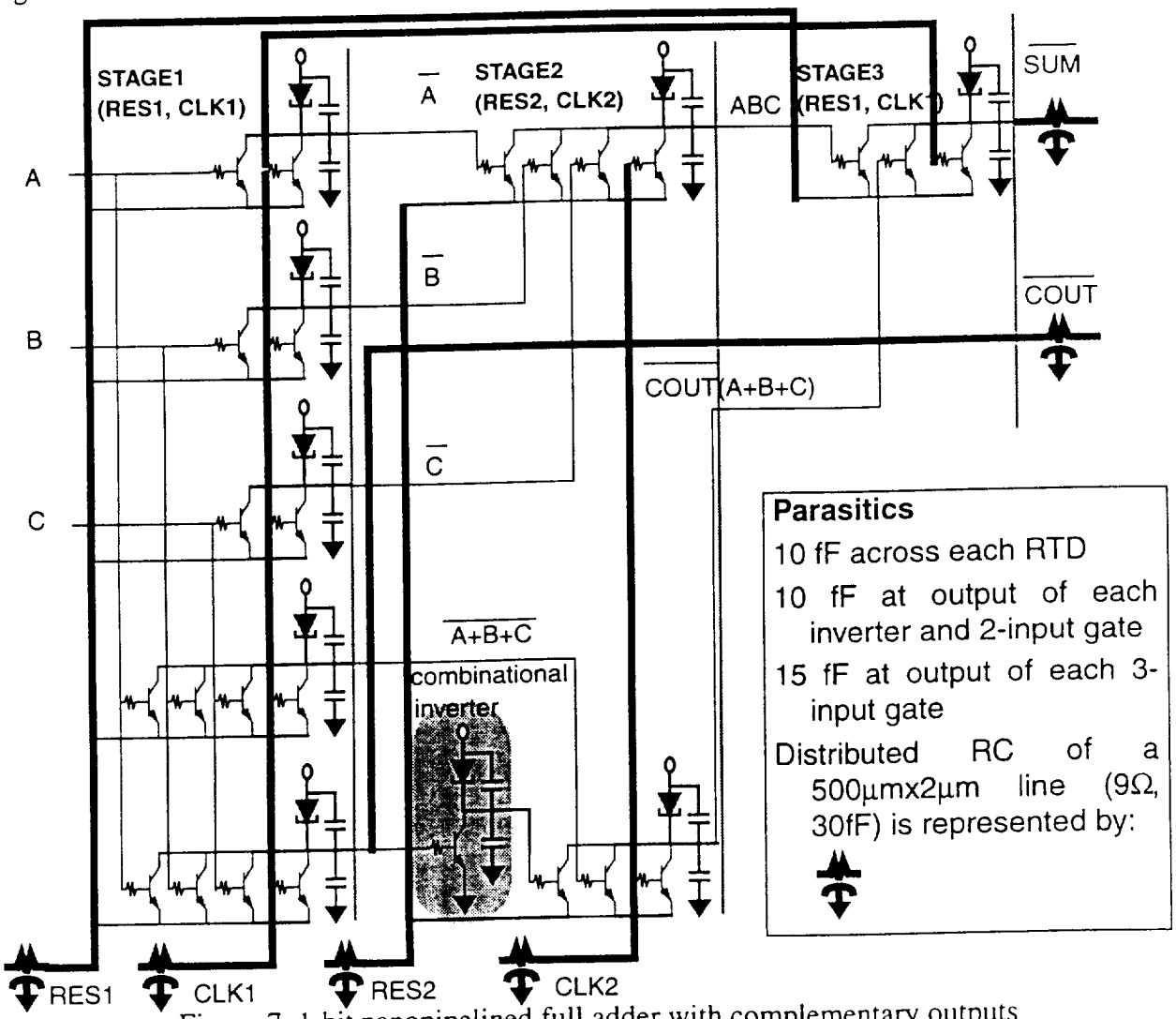


Figure 7. 1-bit nanopipelined full adder with complementary outputs

The *res1* and *clk1* signals form phase1 of the clock whereas *res2* and *clk2* form phase2 of the clock. The two phases of the clock must be non-overlapping. However, the reset and clock signals of a phase may partially overlap as shown in Figure 8. A large overlap period between the aforementioned signals is not desirable since the circuit output is not valid during this time. The

simulated output for the 1-bit nanopipelined adder is shown in Figure 9. To project realistic performance, load capacitances and parasitics have been added to the RTDs and HBTs used in the circuit. Also, clock and reset lines are assumed to be global lines with a distributed RC parasitic elements as shown in Figure 7. The circuit outputs are assumed to drive global bus lines across the chip. The two phase clock consisting of *reset1-clock1* and *reset2-clock2* operates at 10 GHz.

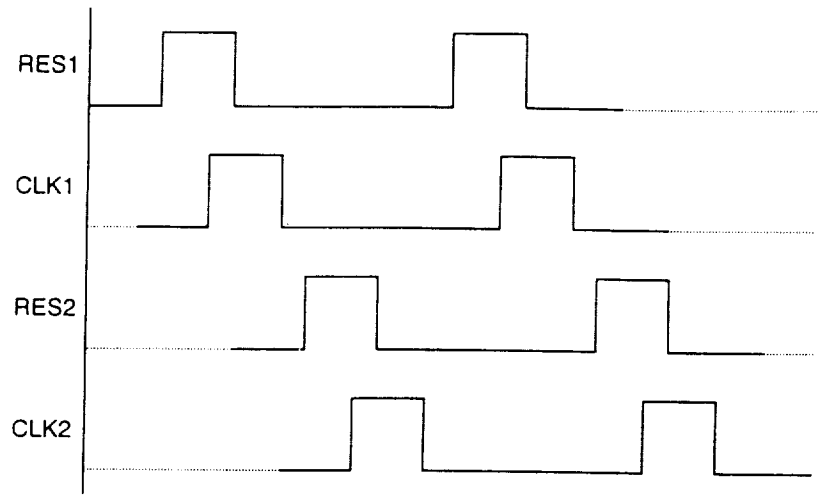


Figure 8. Multiphase timing scheme

4. Correlator implementation

The block diagram of the pipelined correlator is illustrated in Figure 10. A 32-bit latch holds the PN sequence. The input is a serial bit stream which is fed to a 32-bit shift register. The 32-bit latch and 32-bit shift register are each composed of 64 bistable inverters. A pair of cascaded bistable inverters each operating on single, separate phases of the two-phase clock form the basic 1-bit latch. The 32-bit raw correlation vector is generated by performing a bitwise XOR operation on the PN sequence latch output and the most recent 32 bits of the sampled signal available at the shift register output. The raw correlation vector is registered and this forms the input to the pipelined adder network that determines the difference between the number of 1s and 0s in the raw correlation vector. This is the correlation value between the incoming signal and the resident PN sequence and is determined for the 32 most recent data bits at every clock cycle. This value ranges from -32 to +32. The functional description of the correlator is illustrated in the equations (10) through (14).

$$data[31 \leftarrow 0] = \left\{ D^{32}(d_{in}), D^{31}(d_{in}), \ldots, D^{1}(d_{in}) \right\} \tag{10}$$

$$code[31 \leftarrow 0] = \left\{ D^{1}(PN_{31}), D^{1}(PN_{30}), \ldots, D^{1}(PN_{0}) \right\} \tag{11}$$

$$corr[31 \leftarrow 0] = code[31 \leftarrow 0] \oplus data[31 \leftarrow 0] \tag{12}$$

$$sum[5 \leftarrow 0] = \sum_{i=0}^{31} corr[i] \tag{13}$$

$$diff[6 \leftarrow 0] = 32_d - 2 \bullet sum[5 \leftarrow 0] \tag{14}$$

Here, $D_i(s)$ represents the value of signal s, i clock cycles prior to the current input.

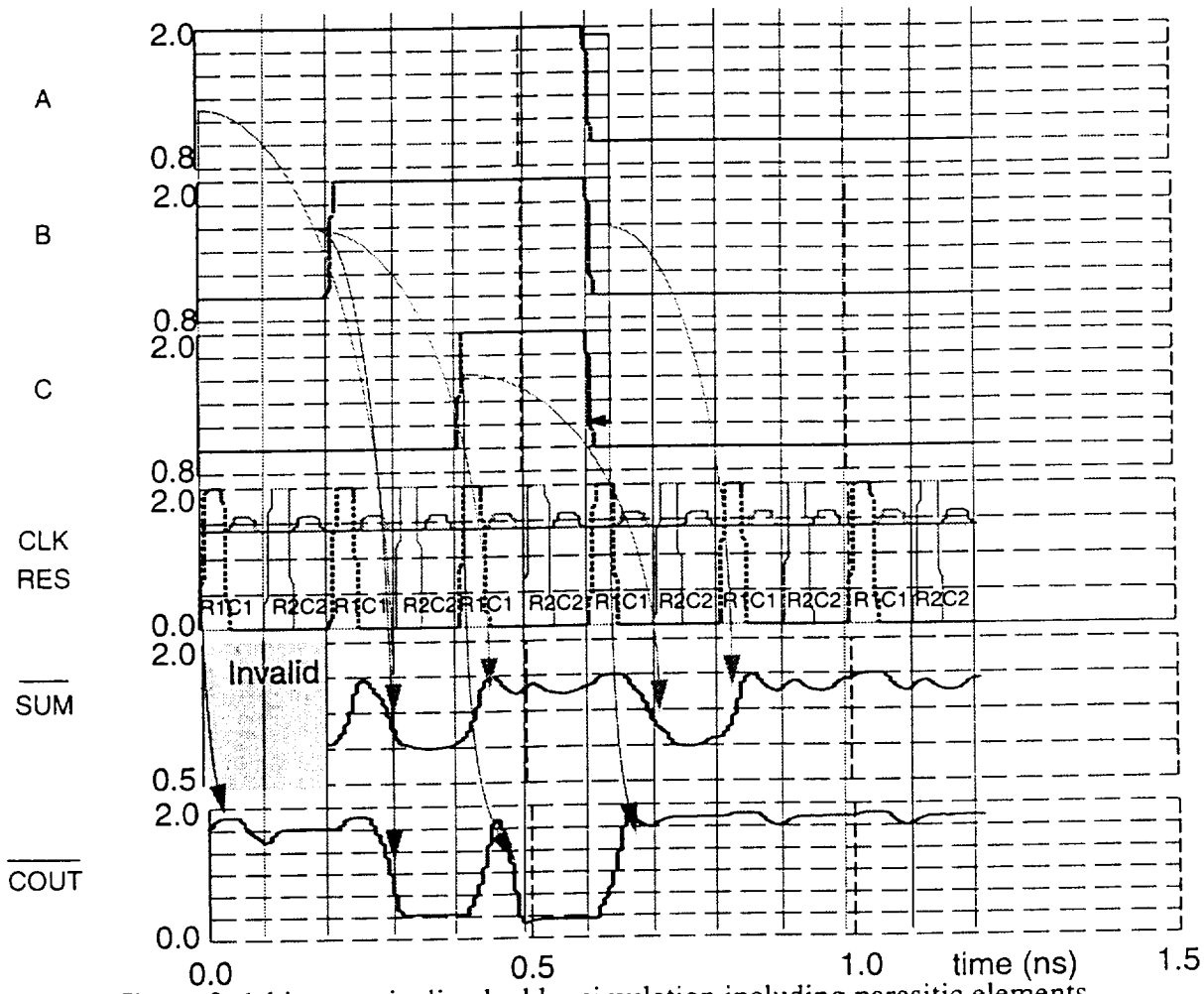


Figure 9. 1-bit nanopipelined adder simulation including parasitic elements

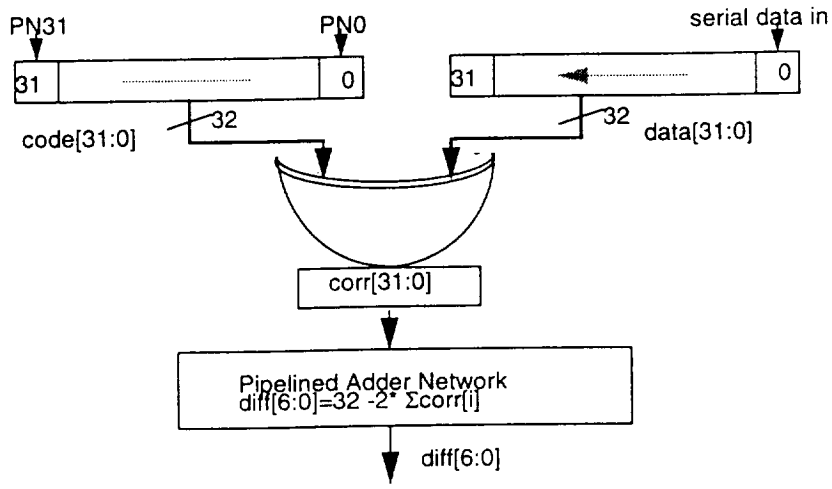


Figure 10. Pipelined correlator block diagram

4.1 Pipelined Adder Network

The adder network consisting of 26 nanopipelined full adders, 11 nanopipelined half

adders, and 36 bistable inverters is illustrated in Figure 11. The adders used in the design have complemented sum and carry outputs in order to reduce pipeline latency. The input to the adder network is the raw correlation vector generated by the 32-bit bistable XOR network. The circuit performs eighteen stages of addition to generate a 7-bit result which is the difference between the number of 1s and number of 0s in the correlation vector. Since each stage is nano-pipelined due to use of self latching gates in the bistable adders, the throughput of the circuit is one 32-bit correlation every cycle. However, since the seven bits of the adder network output are not simultaneously generated, bistable inverters are required to synchronize the bits such that all seven bits of a correlation appear in order at the output of the correlator. The least significant bit of the correlation value is always 0 since the difference between the number of 1s and number of 0s in a 32-bit vector is always even. The pipelined adder network essentially sums up the number of 1s in the correlation vector. Bits 0, 1, 2, 3 and 4 of the sum of 1s directly translate to bits 1, 2, 3, 4 and 5 of the difference between number of 1s and number of 0s. Bit 6 of the correlation value is computed while bit 5 of the sum of 1s is being generated by connecting the carry input of the final full adder to V_{dd}. This achieves the 2s complement subtraction required for computing the difference between the number of 1s and number of 0s in the correlation vector. No additional pipe stages are required for this conversion.

The functional simulation of the 32-bit parallel correlator is shown in Figure 12. The PN sequence for this simulation is chosen to be AAAAAAAA Hex. Note, that this is not an optimum PN sequence but rather is chosen for the ease of illustration of the functionality of the correlator. The input is a pattern of alternating 1s and 0s which results in the 32-bit shift register output toggling between AAAAAAAA Hex and 55555555 Hex at each cycle. This causes the raw correlation vector to alternate between all 1s (FFFFFFFF Hex) and all 0s (00000000 Hex) with each cycle. Thus, the desired correlation difference should be +32 decimal and -32 decimal respectively for the two cases mentioned above. This is seen to be the case in the simulation output. It should be noted that the simulation output reflects changes in the input 10 cycles prior to the output due to pipeline latency. However, the same input pattern has been maintained and is shown in the current plot for the purpose of illustration.

4.2 Comparison with CMOS technology

The correlator designed using RTDs and HBTs is compared with a CMOS implementation using 0.5 micron process technology. The results of the comparison for three circuits - the basic bistable majority gate, the bistable full adder and the 32-bit parallel correlator - are presented in Table 1.

Table 1: Comparison of RTD+HBT circuits with CMOS implementations

Parameter	Bistable Majority		Bistable full adder		32-bit Parallel Correlator	
	CMOS (0.5μ)	RTD+HBT	CMOS (0.5μ)	RTD+HBT	CMOS (0.5μ)	RTD+HBT
Device count	20	5	68	34	6000	2060
Power dissipation	0.7 mW	2 mW	2.3 mW	12 mW	600 mW	1.5 W
Speed	400 MHz	20 GHz	400 MHz	10 GHz	400 MHz	10 GHz
Power-Delay product	1.75 pJ	0.1 pJ	5.75 pJ	1.2 pJ	1.5 nJ	0.15 nJ

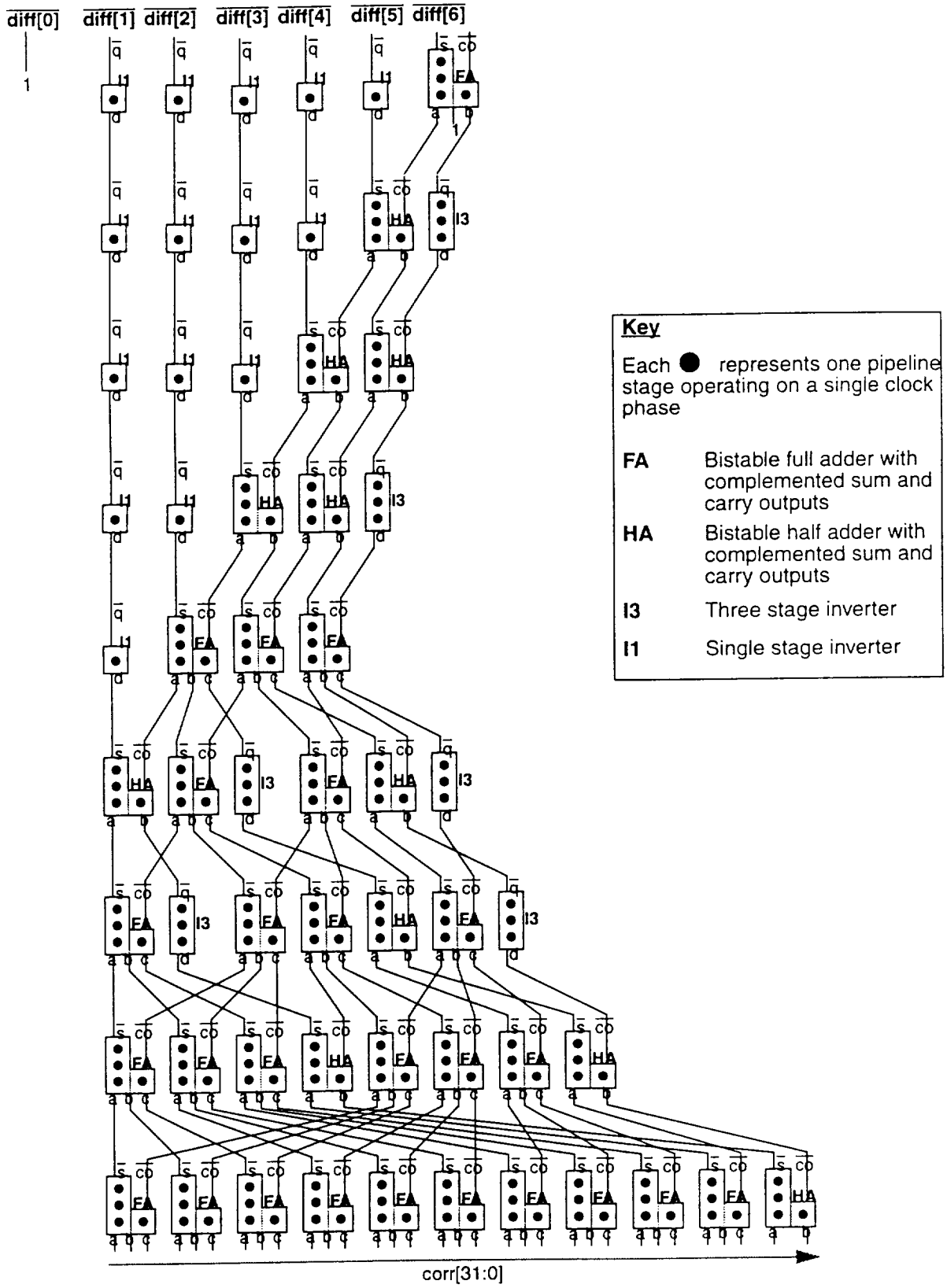


Figure 11. Pipelined Adder Network

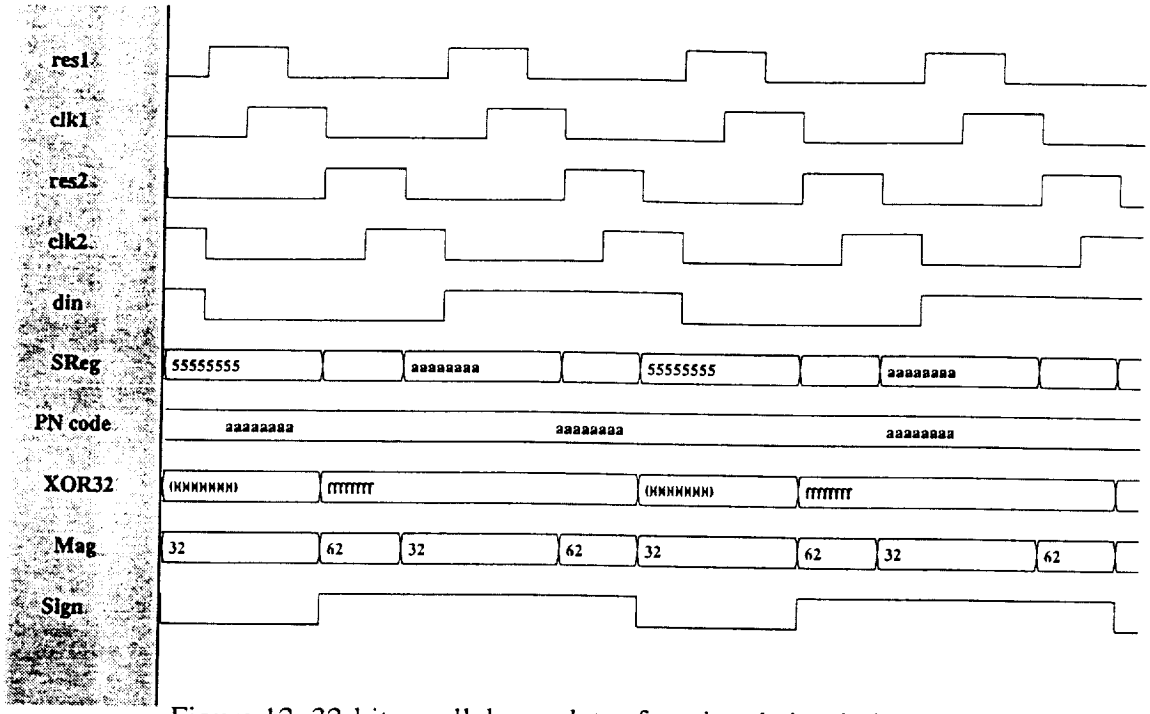


Figure 12. 32-bit parallel correlator functional simulation

The RTD+HBT based correlator offers a tenfold improvement in power-delay product even though it consumes greater absolute power. The fewer number of devices used in the correlator also imply a reduction in wiring lengths and hence parasitics and delays associated with interconnects are much smaller in the RTD+HBT correlator.

Conclusions

The synchronous, sequential nature of true-bistable gates using RTDs and HBTs has been exploited to build a very high speed and compact parallel correlator. Design equations and constraints have been studied and a design methodology for RTD+HBT bistable logic gates is proposed. The bistable nature of the logic gates has demonstrated advantages over conventional logic families by eliminating pipeline area and delay overheads in deep pipelined logic systems resulting in improved throughput and smaller circuit size. The compact implementation of threshold functions allows a single gate carry function which facilitates design of high speed arithmetic and logic functions used in the correlator. Reduction in device count has led to shorter interconnections resulting in reduced parasitic delays. The nanopipelined correlator offers a tenfold lower power-delay product as compared to a state of the art CMOS implementation. The proposed design style has applications in the development of high speed digital communication system architectures to achieve several Gb/s data throughput. In particular, for space based communication systems, nanopipelined RTD+HBT based logic designs offer compact solutions with very low power-delay products.

References

[1] R. E. Ziemer, and R. L. Peterson, *Digital Communications and Spread Spectrum Systems*, New York: Macmallan, 1985

[2] T. VanDuzer, "Superconductor digital ICs," in *VLSI Handbook* (J. DiGiacomo, ed.), New York: McGraw-Hill, 1989.

[3] K. Maezawa and T. Mizutani, "A new resonant tunneling logic gate employing monostable-bistable transition," *Japanese Journal of Applied Physics*, vol. 32, pp. L42-L44, Jan 1993.

[4] S. Mohan, P. Mazumder, R. K. Mains, J. P. sun, and G. I. Haddad, "Logic design based on negative differential resistance characteristics of quantum electronic devices," *IEE Proceedings-G*, vol. 140, pp. 383-391, Dec. 1993.

[5] W. Williamson, III, "High-speed RTD-based logic for systems applications," in *ULTRA electronics program review*, ARPA, Oct. 1994.

[6] P. Mazumder, J. P. Sun, S. Mohan and G. I. Haddad, "DC and transient simulation of resonant tunneling devices in NDR-SPICE," in *21st International Symposium on Compound Semiconductors*, 1994

[7] S Mohan, P. Mazumder, and G. I. Haddad, "Ultra-fast pipelined arithmetic using quantum electronic devices," *IEE Proceedings-E*, vol. 141, pp. 104-110, Mar. 1994.

HIGHLY-PARALLEL, HIGHLY-COMPACT COMPUTING STRUCTURES IMPLEMENTED IN NANOTECHNOLOGY

D G Crawley, M J B Duff, T J Fountain, C D Moffat, C D Tomlinson

Department of Physics & Astronomy
University College London

230602
10P.

Abstract

In this paper, we describe work carried out as part of the ARPA ULTRA program, in which we are evaluating how the evolving properties of nano-electronic devices could best be utilized in highly parallel computing structures. Because of their combination of high performance, low power and extreme compactness, such structures would have obvious applications in spaceborne environments, both for general mission control and for on-board data analysis. However, the anticipated properties of nano-devices mean that the optimum architecture for such systems is by no means certain. Candidates include SIMD arrays, neural networks and MIMD assemblies.

We explain that, because the propagation of signals through large arrays of nano-devices is almost certain to be a source of difficulty, our initial investigations center on the most regularly structured locally-connected architecture, the SIMD mesh-connected processor array. This structure offers the additional advantages of minimum external interfacing, conceptually simple redundancy schemes for fault-tolerance, and moderate memory requirements.

We describe the current phase of our program, in which we are simulating structures based on both resonant tunnelling devices and quantum cellular automata. In addition, we describe a novel architecture (the propagated instruction processor) which removes the requirement for other than near-neighbour connections for both data and control lines.

We calculate that the minimum anticipated device dimensions (in the order of a few tens of nanometres) would allow an array of 1000x1000 processors to be constructed on a few square mms of semiconductor. This would be equivalent in general performance to a system of at least 1000 DEC Alpha devices but would be optimally suited to the on-board analysis of sensor data, particularly in the form of images. Such a structure would offer ample computing power for the autonomous control of robotic systems.

We report initial results from the performance evaluation of our simulated structures using a comprehensive suite of algorithms. Our future program involves completing this evaluation for all of the SIMD-based systems and then, hopefully, extending our investigations to include the two other promising architectural configurations - neural networks and MIMD systems.

1. Introduction

In the next decade we are likely to see the emergence of several families of semiconductor devices with characteristic dimensions of the order 10^{-9}m. These so-called nanoelectronic devices will provide circuit elements which are several orders of magnitude smaller and several orders of magnitude faster than are currently available and will therefore present new challenges to computer architects. The potential use of these nanoelectronic devices to construct highly-parallel, highly-compact computing structures is being studied as part of the ARPA Ultra program [1,2].

The extremely small size of the devices will make it possible to incorporate millions of logic gates in a single chip and, whilst this offers enormous potential for high-speed, low-power computing, it does at the same time present major difficulties of organisation and control which will need to be studied with extreme care.

For ease of fabrication and in order to provide a comprehensible computer architecture which can be both tested and programmed intelligibly, a regular structure is to be preferred. Fortunately, a natural candidate architecture is available and has been studied in depth for the last twenty-five years: the Single Instruction stream, Multiple Data stream (SIMD) mesh-connected processor array [3].

SIMD arrays consist of assemblies of identical, simple processor elements (PEs), usually connected each to its nearest neighbours in a square array and each capable of only simple logic operations on single-bit data. Instructions are fed in a parallel stream to every PE and each instruction is executed simultaneously by every PE. Currently, arrays are in operation with between 32^2 and 256^2 PEs. Memory is distributed uniformly across the array so that each PE has access to, typically, several kilobytes of local data storage. Memory in this form is clearly well suited to image data structures (each pixel residing in the local memory of one PE) and both this and the parallel processing strategy which can easily be implemented in arrays have led to SIMD mesh-connected arrays being applied particularly successfully to image processing tasks [4,5].

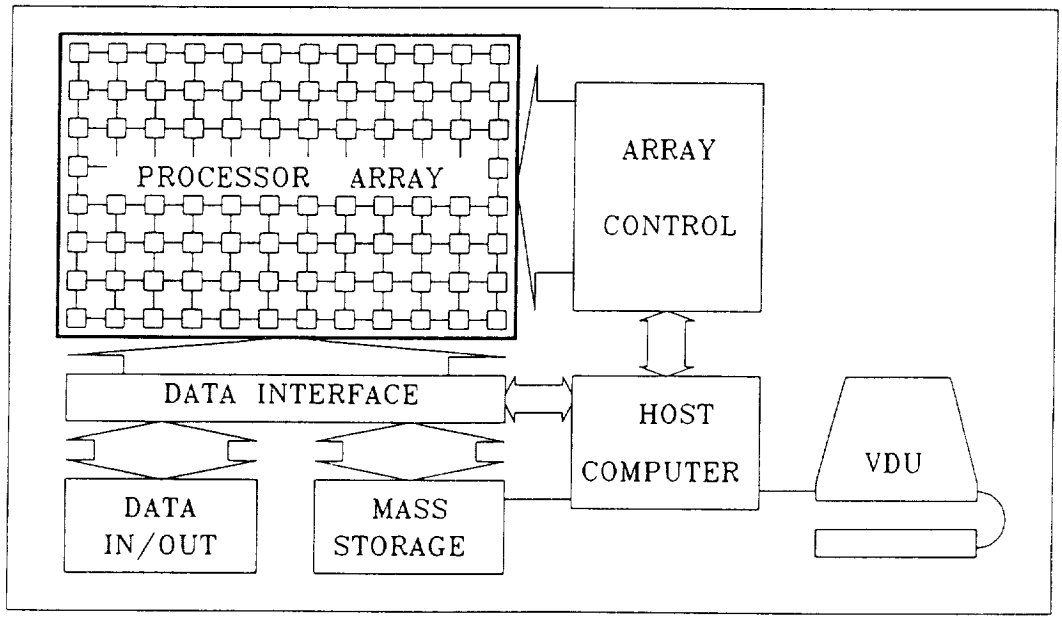


Figure 1 General architecture of SIMD systems

The large number of PEs in even the smaller arrays leads to a high degree of parallelism and hence to potentially large computing speed gains over single processors. However, the PEs are inherently simple and individually much less powerful than the processors found in, for example, workstations, so the gains from parallelism are offset against the losses due to simplicity of the PEs. Further losses are the result of underlying lack of parallelism in many of the tasks to be executed which make it difficult to make effective use of a major fraction of the PEs at any one time. The successful use of arrays in image processing stems from the fact that many image processing tasks are basically parallel in nature, especially considering those derived from convolution, which can be decomposed into local neighbourhood operations (i.e. operations in which the result value at any address (i,j) in the array is a function of the

original values at all addresses in the immediate neighbourhood of (i,j), typically for all points (x,y) in which $i-1 \leq x \leq i+1$ and $j-1 \leq y \leq j+1$).

One disadvantage of SIMD arrays is the comparatively inefficient manner in which data has to be transported across the array when the algorithm so requires it. A good example of this problem is image rotation in which, ultimately, it might be necessary to move pixels from one side of the array to the other. If the only connections available are those between neighbouring PEs, then this process has to be executed in a sequence of steps from PE to PE along the required path and may therefore involve a very large number of system clock cycles. A similar problem occurs in relation to inputting data to and outputting data from the array, although in conventional arrays, this is usually achieved by providing additional data paths along the principal array directions.

2. Nanoelectronic Arrays

Arrays constructed from nanoelectronic devices will have, in general, all the benefits and disadvantages of the arrays constructed using present day technology. The expected additional benefits are larger array dimensions (permitting, for example, the processing of larger images and other data sets) [6], higher speeds (leading to increasingly complex programs which will run in real time) and lower power consumption and weight (resulting in higher system portability). As the technology matures, it can be expected that the benefits will be obtained with no significant increase in cost.

On the other hand, it can also be anticipated that larger array sizes and the nanoelectronic technology will introduce additional operational difficulties illustrated in Figure 2. The current indications are that it will not be possible to incorporate many (if any) 'wire-like' connections to and between PEs over distances much greater than those between PEs; the dimensions of the active elements are so small that loss-free metal links between them would occupy too great a fraction of the substrate space. Reducing the dimensions of the wire connections to the nanometre range would introduce unacceptable losses. This implies that the SIMD instruction stream could not be carried on a word parallel bus running to every PE.

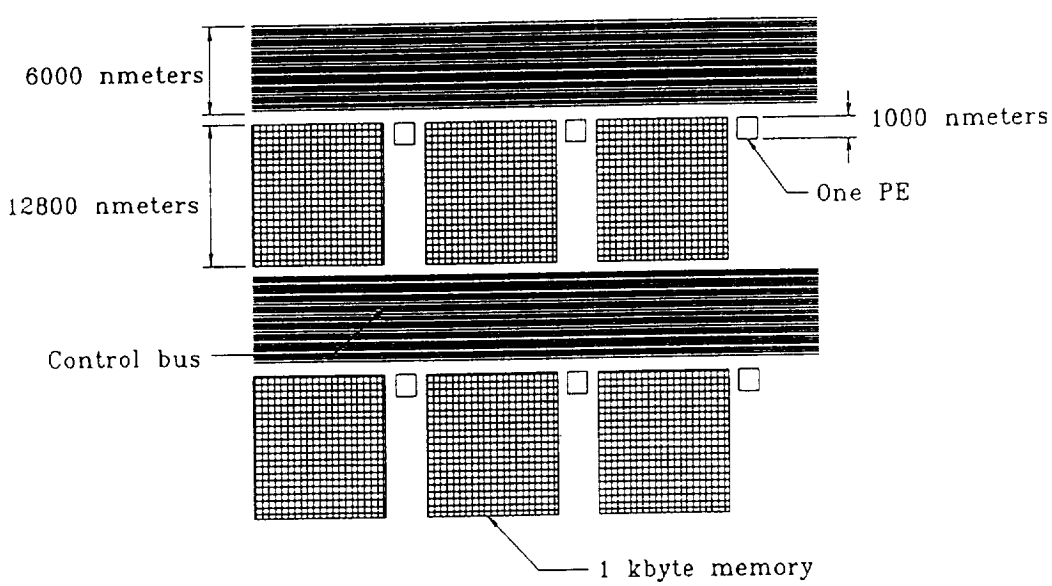


Figure 2 A conceptual floor plan illustrating the relative areas occupied by PEs, memory and control bus

In the same way, data transport from outside the array into the local memory in the array would also have to be serialised. Under these circumstances, in which the PEs are used as the instruction or data paths, then PE control would no longer be simultaneous and $O(N)$ clock cycles (for an $N \times N$ array) would be needed for each step of a parallel operation [6].

A further problem which might be expected to arise in large arrays is that material and process defects could make it almost impossible to construct an array with all its PEs fully operational. Test techniques will have to be devised to pinpoint the defective PEs and circuit redundancy (with error correction) incorporated into the design.

Finally, preliminary design calculations have indicated that for an array with conventional computing capability, the area occupied by the local memory completely dwarfs that taken up by the PEs. Unless between-chip optical interconnect can be achieved, it will not be feasible to remove local memory from the array chip and array sizes will not be as large as had been hoped.

3. Circuit Design and Simulation

Nanoelectronic technology is in a very early stage of development and it is by no means clear which of the various proposed types of device will be the first to become available for practical use in computing circuits. The current indications are that both *resonant tunnelling devices* (RTDs) and *quantum cellular automata* (QCAs) are worth serious consideration, although the former are likely to become available rather sooner than the latter [7].

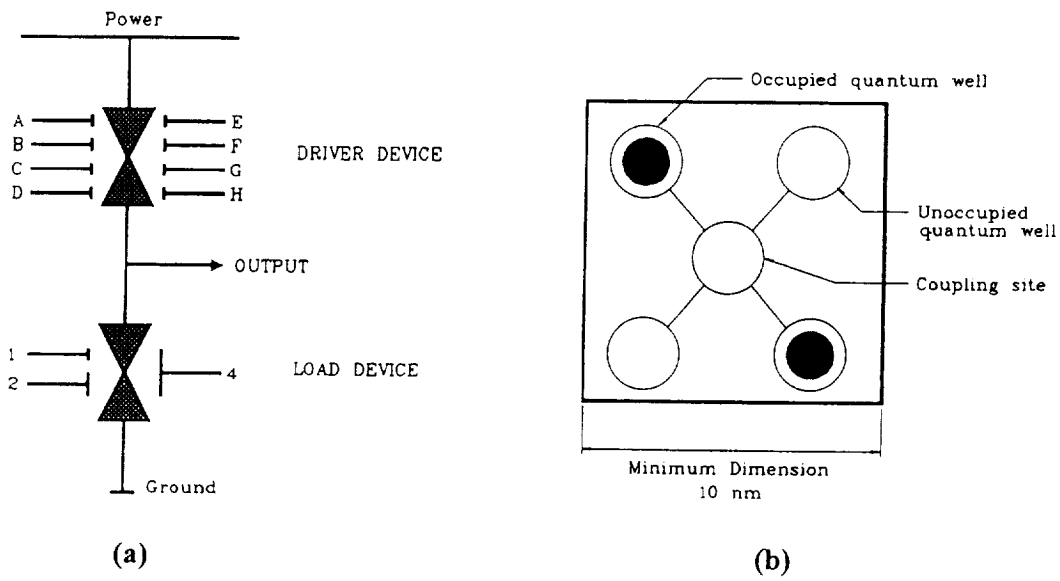


Figure 3 Nanoelectronic devices: (a) A resonant tunelling threshold logic element, (b) a quantum cellular automaton

3.1 RTD-based circuits

Devices based on resonant tunnelling effects can be scaled down to very small dimensions since their operation depends on the ability of current to tunnel through potential barriers; it is this same tunnelling effect which causes ordinary transistor operation to break down (through current leakage) when attempts

are made to scale them to comparably small dimensions. In the present programme, an element has been devised by placing two RTDs in series, one of which acts as a driver and the other as a load. By suitably adjusting bias currents on the two devices, the elementary circuit can be switched between two stable states. A further extension of this proposal is to provide the biasing by means of multiple contacts to both RTDs so that the circuit produced behaves as a general purpose threshold gate shown in Figure 3(a). This can then be tuned to implement thresholding functions which, in turn, can be used as the basis for producing a comprehensive selection of Boolean logic functions. The universal nature of these implementations leads to the possibility of the fabrication of complex circuits with a high degree of regularity, whilst the reduction in number of devices (from the hundred or so required in a conventional implementation) to two offers further benefits of compactness.

3.2 QCA arrays

An alternative nanoelectronic structure is based on single electrons occupying quantum wells. These also scale to extremely small dimensions, an element being a small as 10^{-8}m square. Two state logic elements can be envisaged which comprise five quantum wells in a configuration shown in Figure 3(b).These elements can be assembled into processor circuits such as that shown in Figure 4, which will execute all the computational functions required in a general purpose array processor.

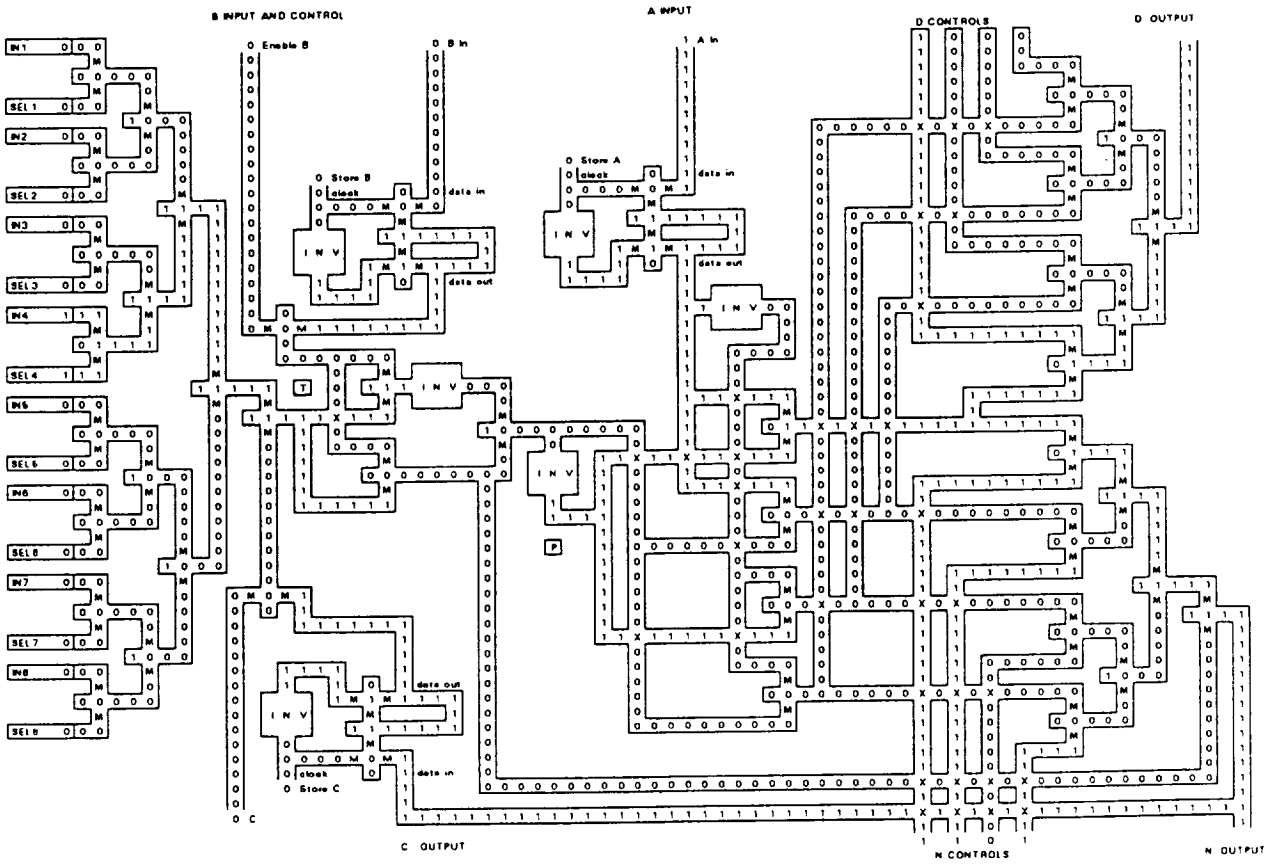


Figure 4 An EXCEL simulation of the CLIP4 processing element

In the case of QCA arrays, the technology is less well advanced than that for RTDs and, at this stage, it is only possible to explore potential circuit designs without attempting to obtain hard information as to

expected performance or device size. A simulation based on an EXCEL spreadsheet is being used for this purpose and yields some performance information in terms of clock cycles [8]. The simulation also gives some idea as to the probable complexity of the proposed circuits (i.e. the number of devices required) and indicates the type of layout that might be involved.

3.3 Array simulation

Enough is known about the expected properties of RTDs to make it worthwhile simulating RTD arrays both in hardware (employing large scale RTDs and heterojunction bipolar transistors) and in software in order to evaluate candidate array architectures. A versatile, semiconductor technology independent, software array simulator has been written as part of this project and is being used in conjunction with a

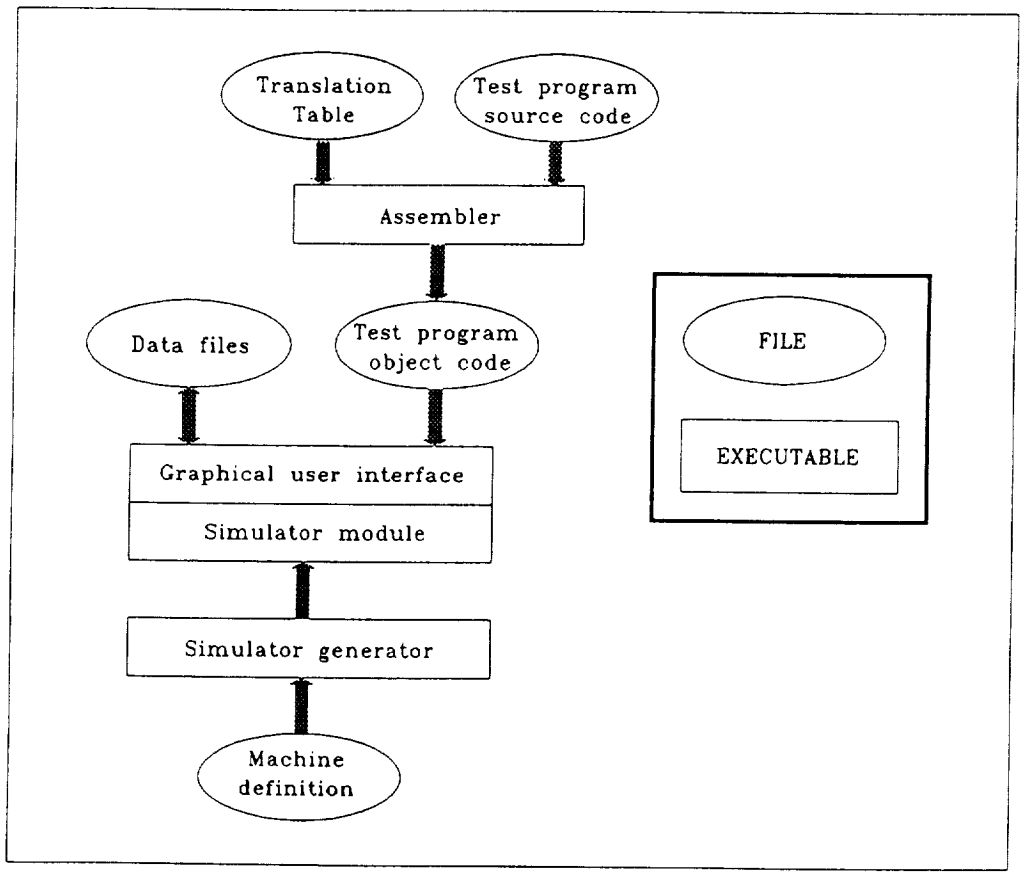


Figure 5 The array simulator software

specially selected suite of image processing algorithms [9]. This allows an assessment to be made of the performance to be expected for appropriate combinations of technology, PE circuit design, array architecture and detailed algorithmic implementation, noting here that efficient algorithms defined at the task level (e.g a Fourier Transform) can be expected to require a detailed design which takes into account the computer structure on which the computation is to be performed. At the very least, full account must be taken of the parallelism to be encountered. The structure of this simulator is illustrated in Figure 5.

As a first step, simulations are being based on the CLIP3 array processor which was developed by some of the authors in 1973. This array was constructed and, in a slightly modified form (CLIP4), used for image processing continuously over a ten year period [10]. It is therefore well understood and a good

vehicle for the studies now being made. Experience in programming such arrays is invaluable in order to maximise the probability that the best possible array performance is being achieved for a given algorithmic task.

4. The Propagated Instruction Processor

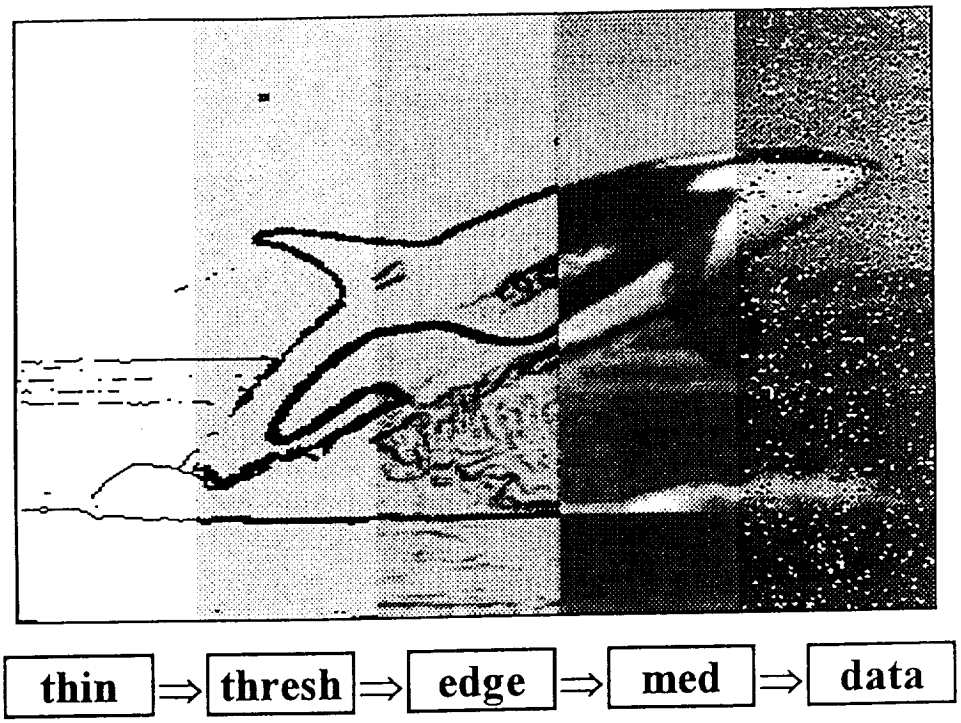


Figure 6 An illustration of propagated instruction processing

If the difficulty of distributing PE instructions (due to the non-availability of wide instruction busses) is as severe as has been feared, it might be necessary to consider a very different way of controlling the processor arrays. Since it will be argued that data (and therefore instruction) paths will have to be stepped from PE to PE via the nearest neighbour connections, then it is important to investigate whether this process can be carried out without losing large numbers of clock cycles. One solution is suggested: the instructions are clocked into the array from left to right along each row. As each PE receives an instruction (or as each group of columns of PEs receive an instruction), the instruction is executed and the results stored in local memory. A sequence of instructions is therefore executed in a travelling wave across the array and the array behaves as a type of pipeline in the manner illustrated in Figure 6 [6]. With careful design of the instruction stream, it has been shown that apart from some latency in the pipeline, programs can be executed in a time which is not significantly longer than that achieved in a conventional SIMD array. An exception to this occurs when the operations being performed are not confined to local neighbourhoods and involve propagating data in various directions across the array. Strategies for dealing with this situation are being evolved in simulation and will be evaluated later in the project. At the worst, instructions involving propagation of data can be executed by first propagating the instruction across the entire array and then allowing time for global propagation as the processed data steps through PEs to all parts of the array. Once again, $O(N)$ time is involved but it is considered that this type of operation forms only a small fraction of the total in typical programs.

5. Applications for Nanoelectronic Computing Arrays

The minimum anticipated device dimensions (in the order of a few tens of nanometres) would allow an array of 1000 x 1000 PEs to be constructed on a few square mm of semiconductor. Although it is too early to state with certainty what will be the switching frequency of any particular device, a target figure of around 100 GHz (for both RTDs and QCAs operating at room temperature) would seem to be feasible [11,12]. The functionality obtainable with the more complex nanoelectronic devices can be compared with simple circuits constructed in present day technology and requiring possibly 10 clock cycles at switching frequencies of 100MHz. This would indicate a speed gain of at least 10^4.

If these preliminary figures are considered in relation to the use of nanoelectronic SIMD arrays for image processing, the potentially large array size would allow high resolution images to be processed at television frame rates at which the complexity of the processes being performed could be increased from the currently attainable simple level, such as skeletonizing binary images, to the complex vision level involving the detection and recognition of objects in an uncontrolled environment. A more precise evaluation of the performance to be expected from such systems is now being carried out, using the simulator and the test suite of image processing algorithms discussed above.

It is clear that the use of nanoelectronic based vision systems in conjunction with robotic systems would lead to the possibility of autonomous navigation of exploratory vehicles under conditions in which response rates are of importance and where the transmission delay for signals sent back to a control station would prevent the possibility of real-time steering and vehicle control. More generally, in any circumstances in which the conditions are hazardous to human life, one of the main reasons for requiring human operators to be present can be eliminated by equipping robotic systems with a vision capability. Current research in this area [13,14] illustrates that neither presently available on-board systems nor remote control by powerful off-line computers offers sufficient intelligence for this task.

Although the main emphasis of this discussion has been on the use of nanoelectronic SIMD arrays for image analysis, conventional technology SIMD arrays have already been applied to a wide range of non-image problems. However, in general, SIMD arrays do not perform optimally on high-level tasks in vision systems and in more general control applications where many varying data types from a range of sensors may be involved. In this type of application and as part of the Ultra programme, the authors are also considering the use of nanoelectronic devices in other computer architectures, especially **M**ultiple **I**nstruction stream, **M**ultiple **D**ata stream (MIMD) systems and neural networks.

MIMD systems can be envisaged as small arrays (typically 32 - 128 PEs) of relatively complex PEs, sometimes assembled on a high speed bus to which is also connected memory, some of it being local to each PE and some accessible by all the PEs. Alternative connection structures are also employed in which direct paths are supplied between selected subsets of PEs (thus enabling especially high performance over a chosen range of commonly occurring algorithms). Systems of this type do not assume data to have an intrinsic two-dimensional structure (as is ideally the case with SIMD array processors) and each PE will usually operate with relatively little direct communication with other PEs (compared with mesh-connected arrays). SIMD arrays obtain their parallelism by operating in parallel on different parts of the data whereas MIMD arrays can also obtain parallelism by splitting the program into parallel sections. Provided the task to be performed is not essentially serial in its structure, MIMD systems can achieve $O(P)$ performance gains (compared with single processors), where P is the number of PEs in the system. The lack of dependence of MIMD systems on the data structure makes them most suitable for processes in which many different data types are involved.

The compact nature of nanoelectronic systems gives the opportunity, in principle, of constructing, and employing in a space environment, hybrid systems combining both SIMD and MIMD subsystems, thereby acquiring the special capabilities of both subsystems with negligible increase in loading on the spacecraft power supplies or payload weight. The problems encountered in operating hybrid systems are being studied by the authors and their collaborators in another research programme [15].

The use of nanoelectronic devices to construct neural networks is also being investigated by the authors. Neural networks have been shown to provide powerful solutions to a broad spectrum of problems ranging from natural speech recognition to financial forecasting. These computing circuits are trained by example, rather than explicitly programmed. On the negative side, whereas the networks exhibit the capability of generalisation (e.g. by classifying objects *similar* to ones in the training set but not *included* in the training set), each network is designed to work on a specific problem or, at the most, a specific class of problems. However, in a situation in which autonomous behaviour in a previously unexplored environment is wanted, it could be that a neural network solution might be the most likely to succeed.

Two alternative approaches are being considered. In the first, neural network algorithms can be embedded in mesh-connected SIMD arrays [16] so the array is programmed to act as a neural network (which then has to be trained for its specific function). In the second approach, nanoelectronic devices directly implementing the thresholding function employed in neural networks are connected together in a network structure. This architecture may well prove difficult to construct as many neural networks which are currently being investigated require large numbers of interconnections within the assembly of neural elements. The difficulties already described in attempting to provide long-distance connections between nanoelectronic components of a circuit may well render this approach unworkable. Nevertheless, the well publicised successes of neural networks in some areas of application make it worthwhile not to discard these studies too readily.

6. Conclusion

The potential offered by nanoelectronic devices for the construction of highly-compact computing structures with extremely high clock frequencies, coupled with lower power consumption and low weight, makes them ideal candidates for space-borne computing systems in which communication times back to the control centre are too long to permit effective real-time control.

Circuits are envisaged in which millions of nanoelectronic devices will be combined on one semiconductor substrate and it is realised that the control and programming of such circuits will present new challenges to software engineers and programme designers. The experience gained over the past two decades in working with SIMD arrays is expected to be invaluable in the studies which must now be made and is influencing the design of nanoelectronic circuits in the direction of array architectures.

Although no fully scaled nanoelectronic devices operating at room temperature have yet been fabricated, the progress in that direction is sufficiently encouraging to stimulate plans for employing the devices which will emerge in the next decade and to consider them as excellent candidates for space applications.

7. References

[1] ARPA ULTRA Electronics Program Review, Presentation Summaries, Santa Fe New Mexico, (1993)

[2] ARPA ULTRA Electronics Program Review, Presentation Summaries, Santa Fe New Mexico, (1994)

[3] Flynn M. J., Very high-speed computing systems, Proc. IEEE 54, pp. 1901-1909, (1966)

[4] Cellular Logic Image Processing, Eds. Duff M. J. B. and Fountain T. J. , Academic Press, (1986)

[5] Strong J. P., Computations on the MPP at the Goddard Space Flight Center, in Proc. IEEE Vol. 79 No. 4, Eds. Maresca M. and Fountain T. J., pp. 548-558, (1991)

[6] Fountain T. J. and Tomlinson C., The Propagated Instruction Processor, in Proc of BMVC 95, pp. 563-572, BMVA Press, (1995)

[7] Moffat C., A Survey of Nanoelectronics, Internal Report 94/2, Image Processing Group, Dept. of Physics, University College London, (1994)

[8] Tougaw P. D. and Fountain T. J., The Design of Computer Memory Elements using Adiabatically-switched Quantum Cellular Arrays, in preparation, (1995)

[9] Crawley D., Flexible Simulation of Processor Arrays, submitted to 4th Euromicro Workshop on Parallel & Distributed Processing, Braga, Portugal (1996)

[10] Fountain T. J., Processor Arrays: Architectures and Applications, Academic Press, (1987)

[11] Tougaw P. D. and Lent C. S., Dynamic Behaviour of Coupled Quantum Dot Cells, in Proc. 3rd Int. Conf. on Computational Electronics, Ed. Goodnick S., (1994)

[12] Chang C. E. et al., Analysis of heterojunction bipolar transistor/resonant tunneling diode logic for low-power and high-speed digital applications, IEEE Trans. on Electron Devices 40 4, pp. 685-691, (1993)

[13] Carnapete L. S. et al., A miniature retina-based AGV called VAMPIRE, in Proc.CAMP95, IEEE Computer Society Press, (1995)

[14] Danese G. et al., PAVIA: A Control System for Active Vision, in Proc. CAMP95, IEEE Computer Society Press, (1995)

[15] Architecture and Programming Model of an SIMD-MIMD system for Image Processing, Esprit BRA Project 8894 Deliverable A45, (1994)

[16] Forshaw M. R. B., Implementation of Neural Network and Optimisation Algorithms on Mesh Arrays, Internal Report 95/1, Image Processing Group, Dept. of Physics, University College London, (1995)

71090

Technologies and designs for electronic nanocomputers

Michael S. Montemerlo, J. Christopher Love, Gregory J. Opiteck,
David J. Goldhaber, and James C. Ellenbogen

Nanosystems Group
The MITRE Corporation
McLean, VA 22102
Internet: ellenbgn@mitre.org

Diverse space-related applications have been proposed for microscopic and sub-microscopic structures, mechanisms, and "organisms." To govern their functions, many of these tiny systems will require even smaller, nanometer-scale programmable computers--i.e., "nanocomputers"--on-board. The speaker will provide an overview of the results of a nearly two-year study of the technologies and designs that presently are in development for electronic nanocomputers. Strengths and weaknesses of the various technologies and designs will be discussed, as well as promising directions for remedying some of the present research issues in this area. The presentation is a synopsis of a longer MITRE review article on the same subject.

Text of full paper not available at time of publication.

RESONANT TUNNELING ANALOG-TO-DIGITAL CONVERTER

T. P. E. Broekaert, A. C. Seabaugh, J. Hellums, A. Taddiken, H. Tang, J. Teng, and J. P. A. van der Wagt

Texas Instruments Incorporated
P.O. Box 655936, MS 134
Corporate Research and Development
Dallas, Texas 75265

(214) 995-4312 (phone), (214) 995-2836 (fax)
tombo@resbld.csc.ti.com

71091
030607
6P

Abstract

As sampling rates continue to increase, current analog-to-digital converter (ADC) device technologies will soon reach a practical resolution limit. This limit will most profoundly effect satellite and military systems used for e.g. electronic countermeasure, electronic and signal intelligence, and phased-array radar. New device and circuit concepts will be essential for continued progress. We will describe a novel, folded architecture ADC which could enable a technological discontinuity in ADC performance. The converter technology is based on the integration of multiple resonant tunneling diodes (RTD) and heterojunction transistors on an indium phosphide substrate. The RTD consists of a layered semiconductor heterostructure AlAs/InGaAs/AlAs (2/4/2 nm) clad on either side by heavily-doped InGaAs contact layers. Compact quantizers based around the RTD offer a reduction in the number of components and a reduction in the input capacitance. Because the component count and capacitance scale with the number of bits N, rather than by 2^N as in the flash ADC, speed can be significantly increased. A 4-bit, 2-GSps quantizer circuit is now under development to evaluate the performance potential. Circuit designs for ADC conversion with a resolution of 6-bits at 25 GSps may be enabled by the resonant tunneling approach.

Introduction

Since the initial investigations of resonant tunneling by Chang, Esaki, and Tsu [1] and the measurements of terahertz response in resonant-tunneling diodes (RTDs) by Sollner et al. [2], the physics and applications of resonant tunneling devices have received increasing attention [3,4]. Today, a wide range of applications are being explored including high speed and multistate memory [5-7], shift registers/correlators [8], and logic [9-13], where the number of interconnects and transistor delays are reduced by the use of the multi-state tunneling device. Other applications for tunneling devices include analog-to-digital convertors [14-16], optical receivers [17], samplers [18], and triggering circuits [19], all with microwave to millimeter wave bandwidths. In addition tunneling device architectures based on universal logic gates [20,21] could provide solutions to the interconnect bottleneck of post ULSI ICs.

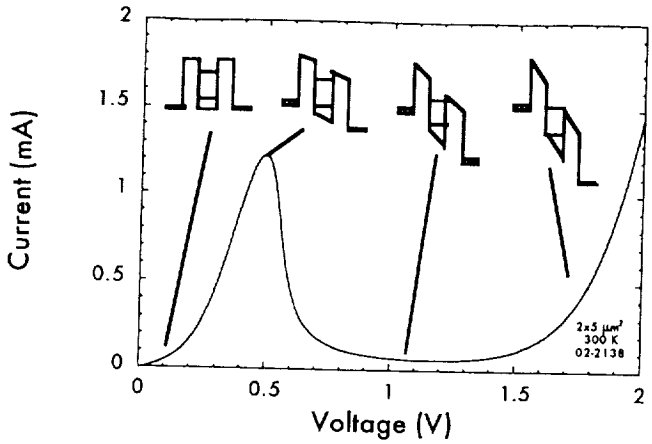


Figure 1. Current-voltage characteristic of an AlAs / $In_{0.53}Ga_{0.47}As$ / InAs resonant-tunneling diode. The curve shown is the result of a *SPICE* simulation with close agreement to measured data at room temperature. In the insets, energy band diagrams indicate the relative alignments of electrons and quantum well states at selected bias conditions.

The RTD, in its simplest form, is a trilayer heterojunction device consisting of two wide bandgap electronic barriers cladding an interior quantum well region as depicted in the inset of Fig. 1. The total thickness of the structure is typically 10 nm or less. A peak in the current-voltage characteristic occurs when electrons in the emitter are energetically aligned with the lowest quasi-bound well state, while the valley (minimum) occurs when the lowest quantum-well state is energetically below the conduction band minimum (see again Fig. 1).

Multiple-Peak Resonant-Tunneling Diodes and Transistor Integration

Beyond the single peak I-V characteristic of Fig. 1, the RTD can be combined epitaxially in series to create multiple current peaks. Shown in Fig. 2 is a *SPICE* simulation illustrating the I-V characteristics of a 4-diode stack. In combination with a transistor, the multiple current peaks of the series RTD combination

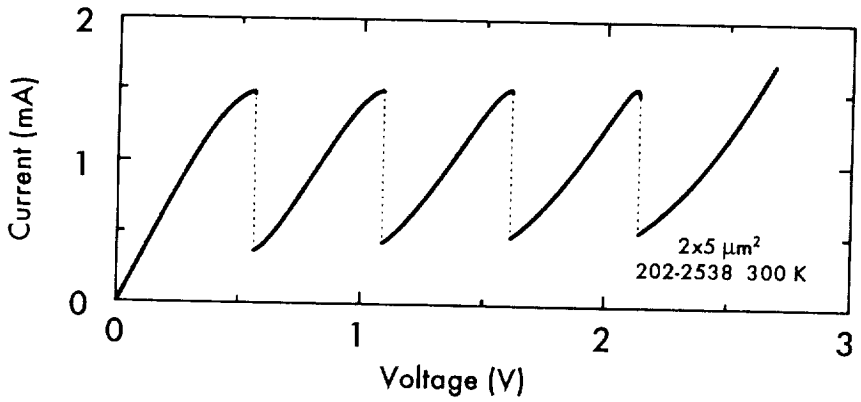


Figure 2. Formation of multi-peak current-voltage characteristics by series combination of resonant tunneling diodes; *SPICE* simulation based on a fit to experimental data.

can be translated into a multi-state transfer characteristic as illustrated in Fig. 3. In this configuration using a 3-RTD stack in the source of a heterojunction field-effect transistor (HFET), a binary output characteristic can be obtained for a multi-valued input characteristic. As we will show this characteristic is naturally suited to analog-to-digital convertor applications. The combination of RTD and transistor occupies less space than the two devices fabricated separately since the epitaxy for the two devices, one above the other, allows vertical integration with only a single additional metallization.

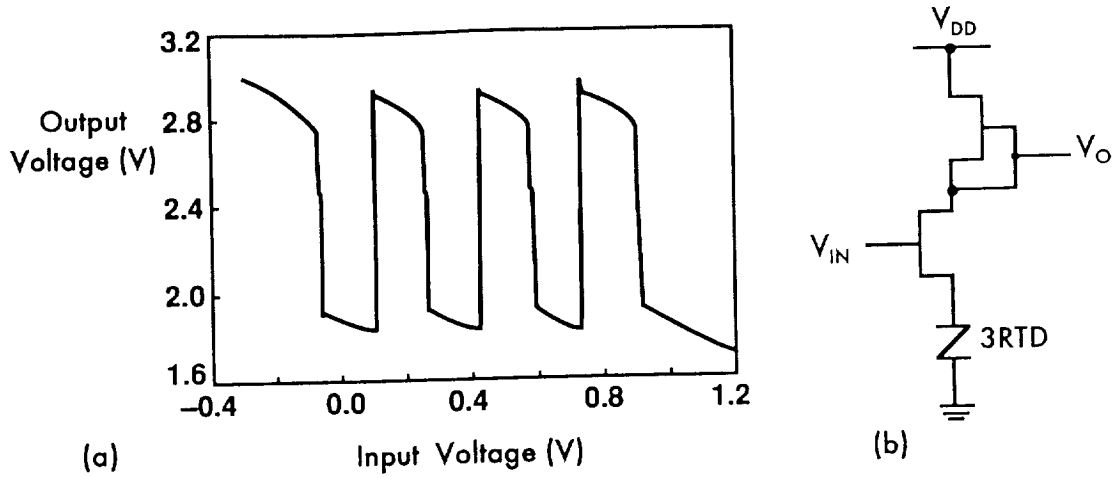


Figure 3. *SPICE* simulated transfer characteristic of a 3-peak AlAs/InGaAs resonant tunneling diode and an InAlAs/InGaAs heterostructure field-effect transistor circuit to yield multistate output.

Resonant Tunneling Quantizer

A simplified schematic diagram of a novel 4-bit quantizer that uses RTDs, HFETs, and resistors is shown in Fig. 4. Four identical multilevel folding amplifiers, each consisting of an HFET with a four-peak RTD

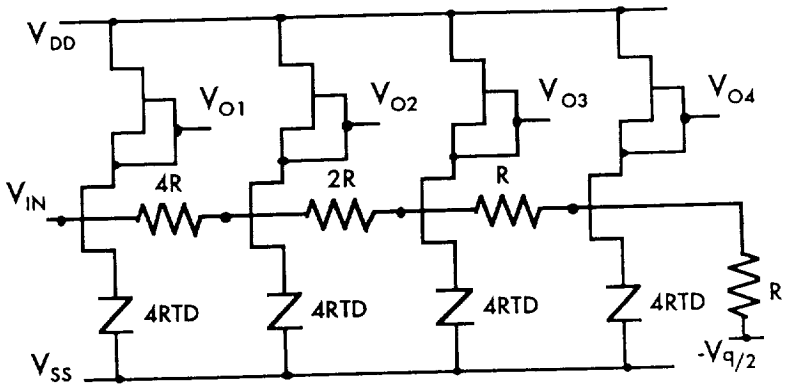


Figure 4. Four bit folding quantizer using 4-peak resonant-tunneling diodes (4RTD) and field effect transistors.

in series with the source and a load resistor, are connected by a binary-weighted resistor ladder. The gate current to the HFETs is at all times much less than the current through the resistor chain; the sum of the resistances, 8R, can be 50 Ω. Four-peak RTDs are used here for a four-bit quantizer (however, only two-peak RTDs are needed for a 3-bit quantizer). The first multilevel folding amplifier generates the LSB as follows. As the gate voltage of the HFETs, V_{IN}, is swept upward from the off-state, the output voltage, V_{o1}, starts high and goes low as V_{IN} exceeds the threshold voltage of the HFET, V_t. As the gate voltage continues to increase, the RTD switches from its peak current to its first valley current thereby restricting the HFET current and forcing V_{o1} high again. For further increases in V_{IN}, this cycle repeats and the input/output relation is as shown in the top trace of Fig. 5.

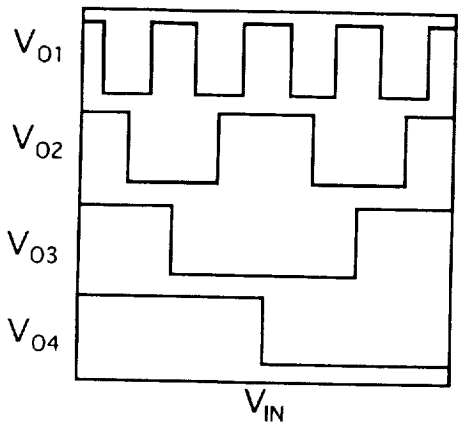


Figure 5. Output voltages of a folded four-bit resonant-tunneling quantizer for the circuit shown in Figure 4.

The more significant bits, V_{o2}, V_{o3}, and V_{o4}, are generated in a similar fashion; however, the binary-weighted resistor ladder divides V_{IN} so that 2, 4, and 8 times the LSB are required to generate the first switching transition, respectively. The four outputs in Fig. 5 are an inverted Gray code representation of the analog input. This novel circuit topology fully folds the analog input directly to a digital output. Since both the number of components and the input capacitance scale as the number of bits N, rather than as 2^N for the conventional 4-bit quantizer, the speed and component count can be significantly reduced. We estimate that the full 4-bit ADC operating at 2 GHz can be constructed with fewer than 100 components and with power dissipation less than 500 mW. Both number of components and power are reduced by an order of magnitude over conventional approaches.

Resonant Tunneling Memory

In addition this resonant tunneling technology can also be used for memory [4-8]. Shown in Fig. 6 are two cells which can be used to latch and store data at speeds as high as 25 GHz. In combination with the necessary drive circuitry these enable the direct storage of digitized data at the full sampling rate of the ADC. The cell shown in Fig. 6 (a) is used for latch/registers and shift-registers whereas the cell shown in Fig. 6 (b) is typically used in SRAM type arrays.

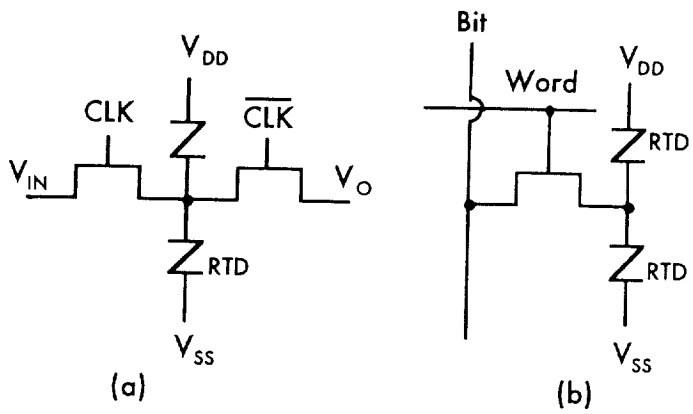


Figure 6. Resonant tunneling diode memory cells: (a) latch and (b) SRAM.

Conclusions

We have outlined the critical elements of a resonant tunneling analog-to-digital converter technology which can provide 4 bits resolution at 2 GHz with potential for 6 bits resolution at 25 GHz. In addition, memory based on RTDs can also provide data storage at frequencies as high as 25 GHz.

Acknowledgments

We would like to acknowledge useful discussions with G. Frazier, T. Moise, and I. Obeid.

References

[1] L. L. Chang, L. Esaki, and R. Tsu, *Appl. Phys. Lett.* **24**, 593-595 (1974).

[2] T.C.L.G. Sollner, W.D. Goodhue, P.E. Tannenwald, C.D. Parker, and D.D. Peck, *Appl. Phys. Lett.* **43**, 588 (1983).

[3] F. Capasso, S. Sen, and F. Beltram, in *High Speed Semiconductor Devices* (S. M. Sze, ed.), pp. 465-520, John Wiley & Sons, New York, 1990.

[4] A. C. Seabaugh and M. A. Reed, "Resonant Tunneling Transistors," *Heterostructures and Quantum Devices* (a volume of VLSI Electronics), eds. N. G. Einspruch and W. R. Frensley, (Academic Press, Orlando, FL 1994) pp. 351-383.

[5] A. C. Seabaugh, Y.-C. Kao, and H. T. Yuan, *IEEE Electron Dev. Lett.* **13**, 479-481 (1992).

[6] C. H. Mikkelson, A. C. Seabaugh, E. A. Beam, J. H. Luscombe, and G. A. Frazier, *IEEE Trans. Electron Dev.* **41**, 132-137 (1994).

[7] M.-H. Shieh and H. C. Lin, *IEEE J. Solid-State Circ.* **29,** 623-630 (1994).

[8] T. C. L. G. Sollner, E. R. Brown, C-L. Chen, C. G. Fonstad, W. D. Goodhue, R. H.Mathews, and J. P. Sage, *Proc. 1993 Int. Semicond. Dev. Res. Conf.*, p. 307, Charlottesville, Va.

[9] T. S. Moise, A. C. Seabaugh, E. A. Beam III, and J. N. Randall, *IEEE Electron Dev. Lett.* **14**, 441-443 (1993).

[10] A. C. Seabaugh, A. H. Taddiken, E. A. Beam III, J. N. Randall, Y.-C. Kao, and B. Newell, *Int. Electron Dev. Meeting Tech. Dig.* 419-422 (1993).

[11] L. J. Micheel, A. H. Taddiken, and A. C. Seabaugh, *23rd Int. Symp. on Mult. Val. Logic* (IEEE Comp. Soc. Press, Los Alamitos, CA) 164-169 (1993).

[12] C. E. Chang, P. M. Asbeck, K.-C. Wang, and E. R. Brown, *IEEE Trans. Electron Dev.* **40**, 685-691 (1993).

[13] T. S. Moise, Y.-C. Kao, A. C. Seabaugh, and A. H. Taddiken, *IEEE Electron Dev. Lett.* **15**, 254-256 (1994).

[14] T.-H. Kuo, H. C. Lin, R. C. Potter, and D. Shupe, *IEEE J. Solid-State Circuits* **26**, 145-149, (1988).

[15] S.-J. Wei, H. C. Lin, R. C. Potter, and D. Shupe, *IEEE J. Solid-State Circuits* **28** 697-700 (1993).

[16] G. Frazier, A. Taddiken, A. Seabaugh, and J. Randall, *Dig. Tech. Papers Solid-State Circ. Conf.* (IEEE cat. no. 93CH3272-2) 174-175 (1993).

[17] T. Moise, Y.-C. Kao, L. Garrett, and J. Campbell, *Appl. Phys. Lett.* **66**, 1104-1107 (1995).

[18] A. Miura, T. Yakihara, S. Uchida, S. Oka , S. Kobayashi, H. Kamada, and M. Dobashi, *IEEE Trans. Microwave Techniq.* **38**, 1980-1985 (1990).

[19] L. Yang, S. D. Draving, D. E. Mars, and M. R. T. Tan, *IEEE J. Solid-State Circ.* **29**, 585-595 (1994).

[20] J. N. Randall, *Nanotechnology* **4** 41-48 (1993).

[21] X. Deng, T. Hanyu, and M. Kameyama, *IEICE Trans. Inf. & Sys.* **E78-D**, 951-958 (1995).

71093

MEMS-Based Communications Systems for Space-Based Applications

230606

Héctor J. De Los Santos and Robert A. Brunner

Hughes Space and Communications Company, Los Angeles, CA 90009

Juan F. Lam, Le Roy H. Hackett, Ross F. Lohr, Jr., Lawrence E. Larson, Robert Y. Loo, Mehran Matloubian, and Gregory L. Tangonan

Hughes Research Laboratories, Malibu, CA 90265

1.0 Introduction

As user demand for higher capacity and flexibility in Communications Satellites increases, new ways to cope with the inherent limitations posed by the prohibitive mass and power consumption, needed to satisfy those requirements, are under investigation [1]. Recent studies suggest that while new satellite architectures are necesary to enable multi-user, multi-data rate, multi-location satellite links [2], these new architectures will inevitably increase power consumption and in turn, spacecraft mass, to such an extent, that their successful implementation will demand novel lightweight/low power hardware approaches.

In this paper, following a brief introduction to the fundamentals of Communications Satellites, we address the impact of MicroElectroMechanical Systems (MEMS) technology, in particular, MEM switches, to mitigate the above mentioned problems and show that low-loss/wide bandwidth MEM switches will go a long way towards enabling higher capacity and flexibility space-based communications systems.

2.0 Fundamentals of Communications Satellites

The fundamental function of a communications satellite is that of providing a communications link between two distant locations on the Earth, namely, a transmitting station and a receiving station [1]. The transmitting station launches a carrier signal, of a certain frequency f_1, up into space in the direction of the satellite. The receiving station, on the other hand, is equipped to receive a carrier signal of another frequency f_2, from the direction where the satellite floats in space. In the simplest case, the satellite receives the carrier signal at the *uplink* frequency f_1, translates it to the *downlink* frequency f_2, and transmits it down to the receiving station. In general, the more uplink and downlink frequencies, and the more transmitting and receiving stations the satellite can link, the greater is its capacity and usefulness, but unfortunately, also the bigger is its required physical size (mass) and power consumption. Since mass is the primary driver of satellite costs [1], it becomes tightly coupled to the satellite's performance, ultimately playing a limiting role on it.

The Communications Satellite could be viewed as composed of two main parts [3], namely, a *platform* and a *payload*. The platform includes the following subsystems:

1-The physical structure
2-The Electric Power Supply
3-Temperature Control
4-Attitude and Orbit Control
5-Propulsion Equipment
6-Telemetry, Tracking and Command Equipment.

The payload, on the other hand, includes:

1-The receiving antenna
2-The transmitting antenna
3-All electronic equipment supporting transmission of carriers.

An examination of the mass and power distribution in a conventional geosynchronous satellite [2], reveals that the communications payload accounts for roughly one quarter of the spacecraft's dry mass, and that it consumes the most power of any single subsystem. Furthermore, within the payload, the

transmitters and antennas account for the bulk of the mass and the power consumption. These fractions are bound to increase even more when we consider satellite architectures which are now emerging as solutions to increasing user demands for satellite capacity and flexibility [1].

3.0 Satellite Architectures

An examination of the evolution of satellite architectures [1] shows a steady increase in sophistication, from the conventional single data rate relay satellite, consisting of a simple downconverter, to regenerative satellites with on-board signal processing and routing capabilities. This increased capacity and flexibility comes at the expense of increased hardware, hence, power, mass, and cost.

We believe that MEMS, through their impact on the Antenna System, will be an enabling technology for the realization of the wireless future paradigm, Figure 1, in which space-based systems will become part of a global communications grid. In this grid the satellite will link airborne users, naval/maritime users, home users, remote office users, public phone booths, VSAT terminals, digital cellular, personal communications network microcells, mobile users, etc, as well as be a node for intersatellite communications. The key to this future versatility lies on agile multi-beam, multi-frequency, lightweight antennas. The key antenna technology that will permeate the space-based wireless future is the phased-array antenna. In the next section we discuss how MEMS will impact such an antenna.

4.0 MEMS-Based Phased-Array Antenna

In simple terms, a phased-array antenna consists of a set of phase shifters or true time delay units that control the amplitude and phase of the excitation to an array of antenna elements in order to set the beam phase front in a desired direction. In a typical Ku-band 5-bit state-of-the-art phasor downlink module, Figure 2, each channel consists of a phase shifter, an attenuator, and a solid state power amplifier (SSPA), implemented in Microwave and Millineter Wave Integrated Circuit (MMIC) technology, driving an array of antenna patches.

An examination of the measured performance of the individual blocks making up a channel reveals a number of technology-related limitations, the most prominent of which has to do with the huge insertion loss, e.g. 10.5dB @14.5GHz introduced by a 5-bit phase shifter. The phase shifter, a state-of-the-art FET-based MMIC chip, despite offering wideband performance and small size, dominates the loss of the chain, thus placing an undue burden on the SSPA by demanding a higher gain and power consumption. These, in turn, drive the unit's power supply capability, heat sinking, and weight requirements. When one considers that typical full-scale phased-array antennas contain thousands of channels, it becomes obvious that such losses as exhibited by FET-based phase shifters are prohibitive. The fundamental reason for the high insertion loss associated with FET-based phase shifters lies on the inevitable device channel resistances: both the "open" channel resistance for the case of the low-impedance state, and the residual series resistance in the pinch-off channel, for the high-impedance case [4].

We believe what is needed are switches with ultralow insertion loss to minimize the use of SSPAs; which are capable of broadband operation to achieve versatility for diverse and simultaneous tasks; that possess high electrical isolation to minimize crosstalk effects; that possess ultralight weight (mass) to effect a lower cost per payload; and whose manufacturing is inexpensive. A realization of such a switch is the Deformable Microwave Micromachined Switched [5], Figure 3. The structure consists of a cantilever beam which, with no voltage applied to it, interrupts the path along a transmission line, and that when deflected, closes it. Preliminary results show that, in the DC-45GHz frequency range, the switch possesses an isolation greater than 25dB in the open mode, while keeping the insertion loss below 0.5 dB in the closed mode, Figure 4. Furthermore, a comparison of the performance of MEM and PIN-diode switches shows that MEM switches are far superior both in terms of insertion loss and isolation, as well as bandwidth, Figure 5. Therefore, it appears clear that brought to maturity, MEM switches, of the type shown here, are posed to become a dominant technology, in the not too distant future.

5.0 Conclusion

An example of the impact that MEMS technology can exert on space-based systems, in particular, Communications Satellites, has been presented. One

area where such impact has been illustrated is that of phased-array antennas. The low-insertion loss, high-isolation, and broadband properties of cantilever beam-type deformable microwave micromachined switches will have a tremendous impact on reducing the power consumption, mass, and indeed on the feasibility of future phased-array systems, thus enabling the realization of the wireless future paradigm.

6.0 References

[1] Edward W. Brugel, *Space Communications* 12 (1994) pp. 121-174.

[2] Christophe E. Mahle, Bernard D. Geller, J. Ram Potukuchi, Geoffrey Hyde, *IEEE AES Magazine*, 3 [11], pp. 3-10 (November 1988).

[3] G.Maral and M. Bousquet, *Satellite Communications Systems: Ststems, Techniques and Technology*, 2nd Edition, (1993).

[4] V. Sokolov, P. Bauhan, J. Geddes, T. Contolatis, and C. Chao, "A GaAs Monolithic Phase Shifter for 30ghz Applications," *IEEE* MTT Symposium, 1983, pp. 40-44.

[5] R.H. Hackett, L.E. Larson, and M.A. Melendes, "The Integration of Micro-Machine Fabrication with Electronic Device Fabrication on III-V Semiconductor Materials", *IEEE* TRANSDUCERS 91 Conference, 1991 INTERNATIONAL CONFERENCE ON SOLID-STATE SENSORS AND ACTUATORS. IEEE cat. # 91CH2817-5. Page 51.

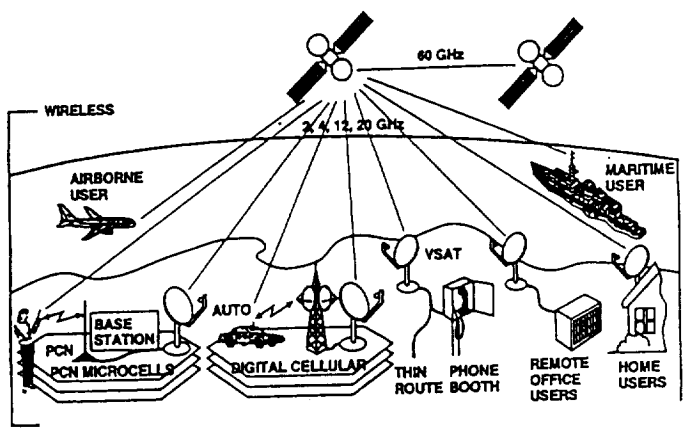


Figure 1. Wireless future paradigm.

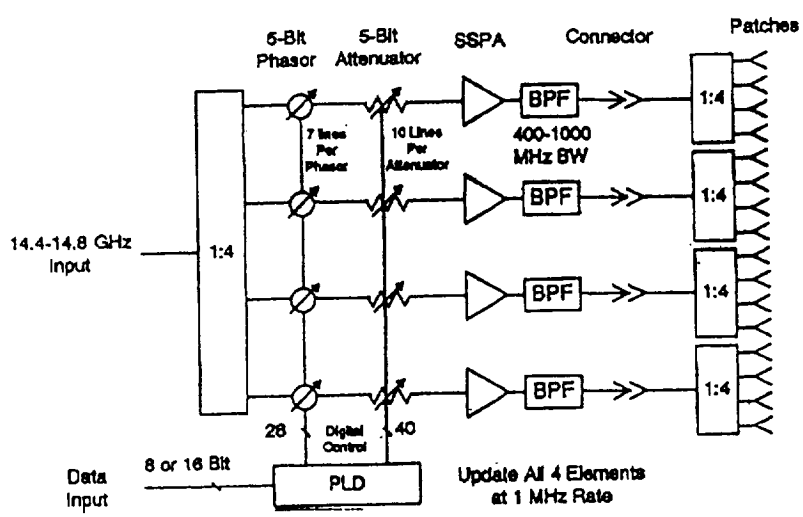


Figure 2. Five-bit phasor downlink module.

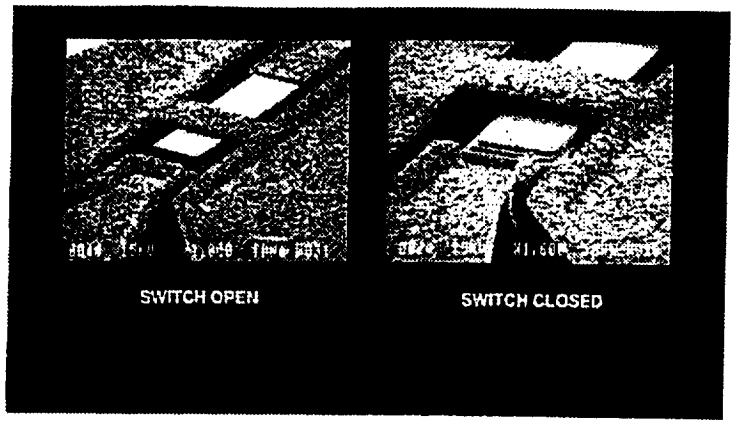


Figure 3. Deformable microwave micromachined switch.

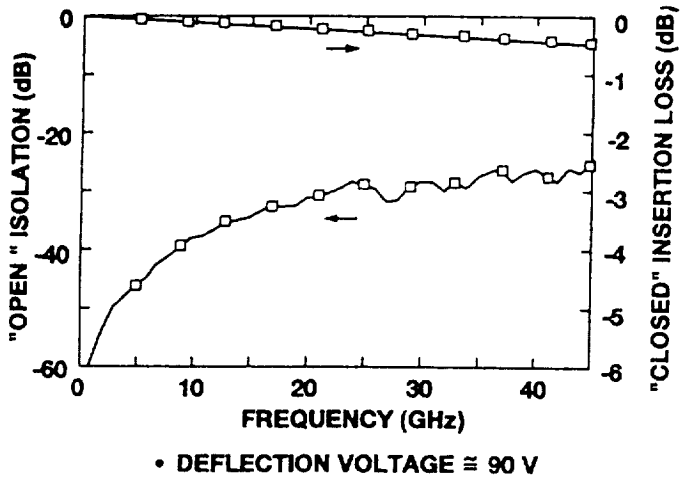


Figure 4. Experimental results of deformable microwave switch.

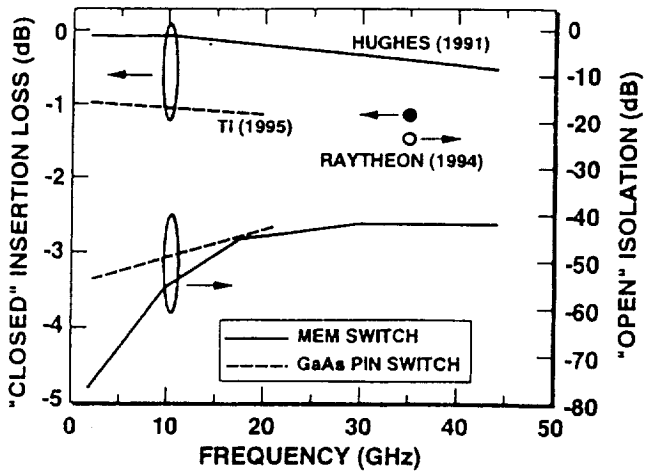


Figure 5. Comparison of insertion loss and isolation of MEM and PIN switches.

SILICON MICROMACHINING IN RF AND PHOTONIC APPLICATIONS

Tsen-Hwang Lin, Phil Congdon, Gregory Magel, Lily Pang, Chuck Goldsmith, John Randall, and Nguyen Ho

Texas Instruments
Dallas, TX USA

1
31094
030607
101

ABSTRACT

Texas Instruments has developed membrane and micromirror devices since the late 1970s. An eggcrate-spacer membrane was used as the spatial light modulator in the early years. Discrete micromirrors supported by cantilever beams created a new era for micromirror devices. Torsional micromirror and flexure-beam micromirror devices were promising for mass production because of their stable supports. TI's digital torsional micromirror device is an amplitude modulator (known as the digital micromirror device or DMD) and is in production development, discussed elsewhere. We also used a torsional device for a 4 × 4 fiber-optic crossbar switch in a 2 cm × 2 cm package. The flexure-beam micromirror device is an analog phase modulator and is considered more efficient than amplitude modulators for use in optical processing systems. TI also developed millimeter-sized membranes for integrated optical switches for telecommunication and network applications. Using a membrane in RF switch applications is a rapidly growing area because of the micromechanical device performance in microsecond-switching characteristics. Our preliminary membrane RF switch test structure results indicate promising speed and RF switching performance. TI collaborated with MIT for modeling of metal-based micromachining.

This talk will provide an overview of developments in metal-based micromechanical devices for RF and photonic applications at Texas Instruments.

I. INTRODUCTION

In the 1970s, TI start using a microelectronic process to develop micromachining structures for spatial light modulators (SLMs). The concept of micromachining was just beginning,[1] and researchers hoped its application to spatial light modulators would help implement optical systems that would provide a breakthrough for information processing.[2,3] In a 10-year period, TI successfully demonstrated several versions of spatial light modulator based on microelectronics/micromirror technology. In the early 1990s, TI invested heavily in production display and printing systems. Meanwhile, other micromirrors and structures were studied within TI to adapt to a wide variety of applications. Larger torsional mirrors were used in fiber-optic switches.[4] Millimeter-sized membranes were used in integrated optic switches.[5,6] Flexure-beam micromirror devices were developed for highly efficient optical correlators.[7] Also, broader aspects of micromachining technology were revisited and prospered in the late 1980s.[8,9] Applications for micromachining in sensors and actuators were introduced at explosive speed. The concept of creating a three-dimensional structure enable by microelectronic batch production brought enormous opportunity to replace existing macrostructures or to develop a new device for new system solutions. TI is also looking into using the same material system for non-optical applications such as an RF switch. During development of the membrane-based RF switches, our test structures gave encouraging results. Micromachine modeling is an important link to transition our research into production. TI established a partnership with MIT to model the devices we fabricated.

DMD applications in projection displays and hard copy systems are described in other publications.[10,11] Torsion micromirror devices have a similar structure to DMDs, and can be used in fiber-optics switches, which are described in Section II. The flexure-beam micromirror device is another promising micromirror structure for production because of its symmetrical support architecture.[12] This device can be used in many

optical information processing applications, as elaborated in Section III. The millimeter-sized flexure-beam membrane can be used in modulating cladding layers of waveguides and changing the propagation properties of integrated optics, as explained in Section IV. Section V provides information about RF switch applications for drumhead devices. Section VI describes the newly established MEMCAD system imported from MIT.

II. TORSIONAL MICROMIRROR IN FIBER-OPTIC APPLICATIONS

TI has developed several versions of fiber-optic crossbar switches using micromirror devices.[4] The first version was set up on an optical table and demonstrated video conferencing. The 4 × 4 crossbar fiber-optic switches consist of 16 switching nodes. Each node uses a pair of fibers, microlenses, and one torsional micromirror (Figure 1). The beam from the incoming fiber focuses on the torsional mirror. The mirror is operated at a binary position: the "on" position reflects the beam into the outgoing fiber through microlenses, and the "off" position deflects the light away. The first version took too much space and was not portable. In the second version, we moved toward manufacturability and portability. The incoming and outgoing fibers are aligned through lithography and orientation-dependent etch that form "V"-grooves for fiber positioning and reflection facets for optical beams (Figure 2). The micromirror becomes the only non-integrated part. The substrate is 2 cm × 2 cm silicon material. The third version is under Advanced Research Projects Agency (ARPA) contract to integrate micromirrors on the substrate. In the "off" state, we use a large asymmetrical micromirror to block the beam path that is parallel to the surface (Figure 3).

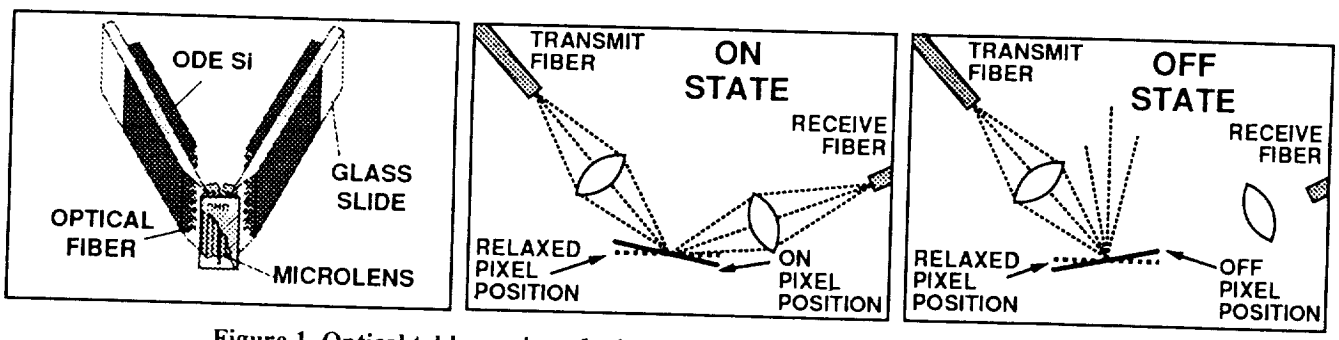


Figure 1. Optical table version of micromirror-based fiber-optic crossbar switch.

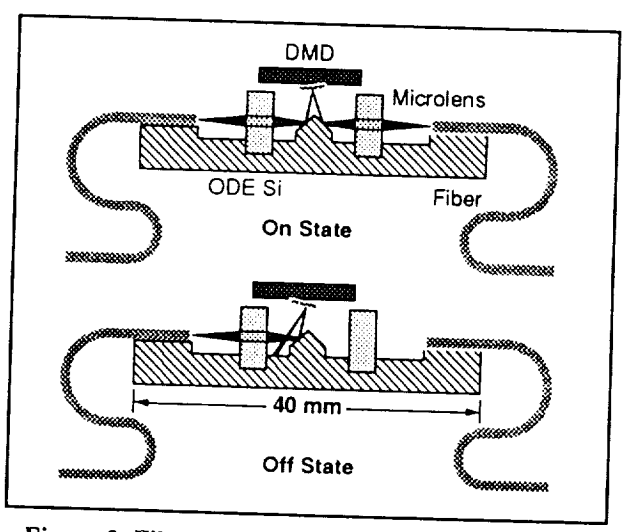


Figure 2. Fiber-optic crossbar switches with fiber on defined-groove substrate and micromirror on a separate chip.

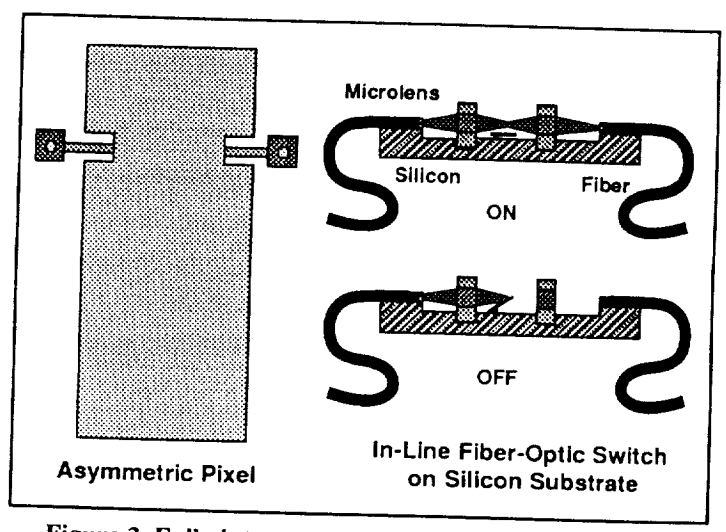


Figure 3. Fully integrated fiber-optic crossbar switches using asymmetric micromirror as shutter.

III. FLEXURE-BEAM MICROMIRROR DEVICES AND APPLICATIONS

The basic structure of the flexure-beam micromechanical element that we implement in this research is shown in Figure 4. The mirror element consists of a square reflecting plate attached at four points to L-shaped flexure hinges. The hinges are attached to four posts that provide the main mechanical support and electrical contact for the element. Each post also mechanically supports four hinges that attach to four different mirrors. The pixel is attracted downward by electrostatic force when it is addressed. The vertical motion of the mirror changes the length of the optical path at the pixel and hence the phase information.

A long list of applications has been waiting for this device. The best known application for the flexure-beam micromirror device is optical correlation.[13,14] Figure 5[15] illustrates a joint transform correlator using two micromirror devices, one as input and one as filter plan, and fold the optical path to use one set of Fourier lenses.

Another application that uses Fourier transform is the spectrum analyzer.[16] Figure 6 shows a spectrum analyzer. A receiver converts a serial incoming signal containing a mixture of different frequencies. A peripheral electronic sample-and-hold circuit maintains the signal at a high enough rate to recognize the waveform. The discretized analog signal is then loaded on the 2-D micromirror SLM. A Fourier lens reveals spots at locations associated with the frequencies.

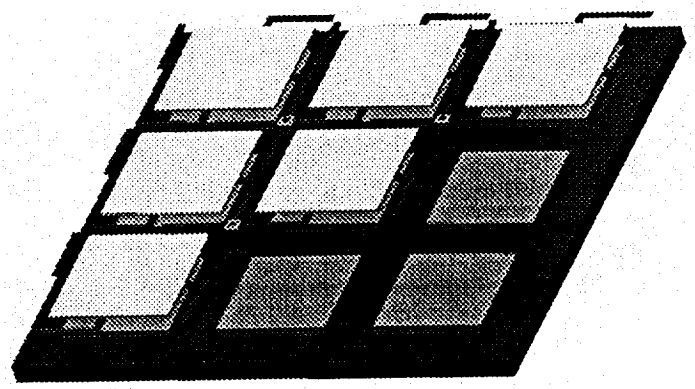


Figure 4. Perspective view of the flexure-beam DMD element.

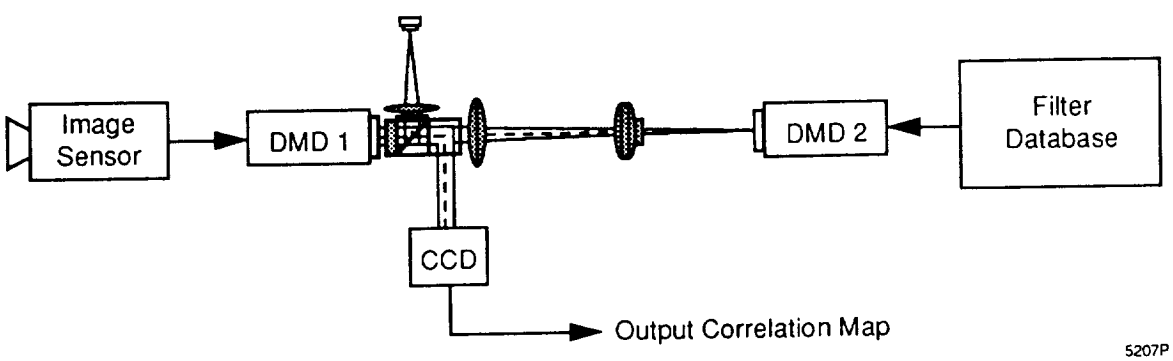


Figure 5. DMD optical correlator.

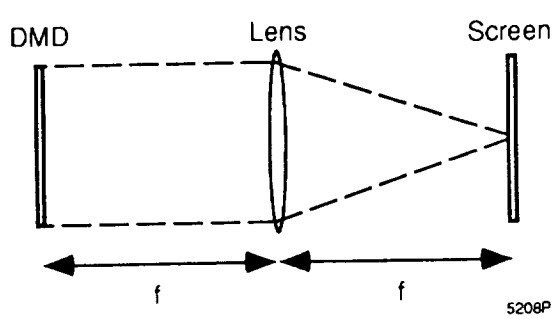


Figure 6. Spectrum analyzer.

Optical interconnects using micromirrors and computer-generated holograms (CGHs)[17] will provide a solution to the bottleneck of massively parallel computing. Figure 7 illustrates the basic concept of the micromirror/CGH interconnect system. The proposed scheme uses the interference property of coherent light. Figure 7 shows the basic interconnection scheme with an example of two processing elements (PEs). The laser in PE1 sends a beam to spot "a" on the CGH. The CGH diffracts this beam into three beams that impinge on the IC plane: one is sent to the SLM and reflected to spot "b" in the CGH plane; two others, a1 and a2, are sent to detectors 1 and 2, respectively. The CGH is constructed so that there is no phase change in a1, a2, and b1 and a 180-degree phase change in b2. If the SLM does not modulate the phase of the beam reflected to "b", the beams interfere constructively at detector 1 and destructively at detector 2. If the SLM modulates the phase on the beam reflected to "b" by 180 degrees, destructive interference occurs now at detector 1, and detector 2 receives the signal associated with constructive interference. If we expand communication to a network of four PEs, we need two SLMs in each PE. There are a total of four different combinations to select the optical paths with two modulators. The number of SLMs, M, needed for interconnecting N PEs is:

$$M = 2 \cdot (\sqrt{N-1})$$

For 64 processors, each PE needs 14 SLMs.

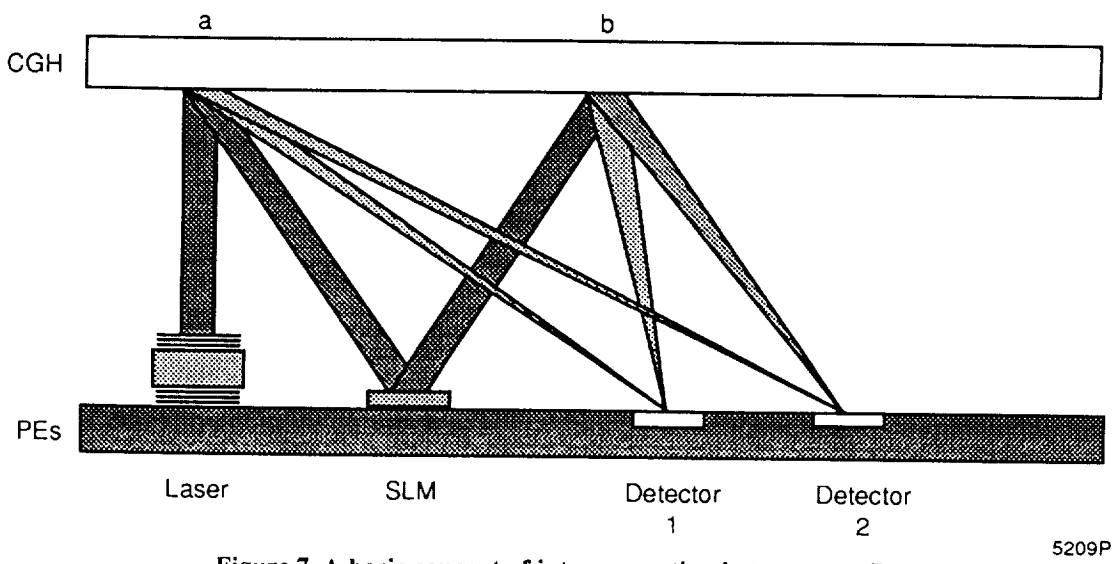


Figure 7. A basic concept of interconnection between two Pes.

Using phase modulators in beam forming and beam steering is an emerging field that is garnering attention.[18] Figure 8 demonstrates a method using phase modulated micromirrors to produce necessary beam shape and direction. Also, the phase modulator can be used in a holographic data storage system as a phase-encoded reference beam.[19]

Figure 9 illustrates the scheme of writing to and reading from a photorefractive medium. The input data beam interferes with the phase-encoded beam and is recorded at one plane of the photorefractive material. By adjusting the optics, another array of data can be stored at another plane. Using an array of the photorefractive fibers, we can expect a storage density up to 10^{13} bit cm^{-3}.

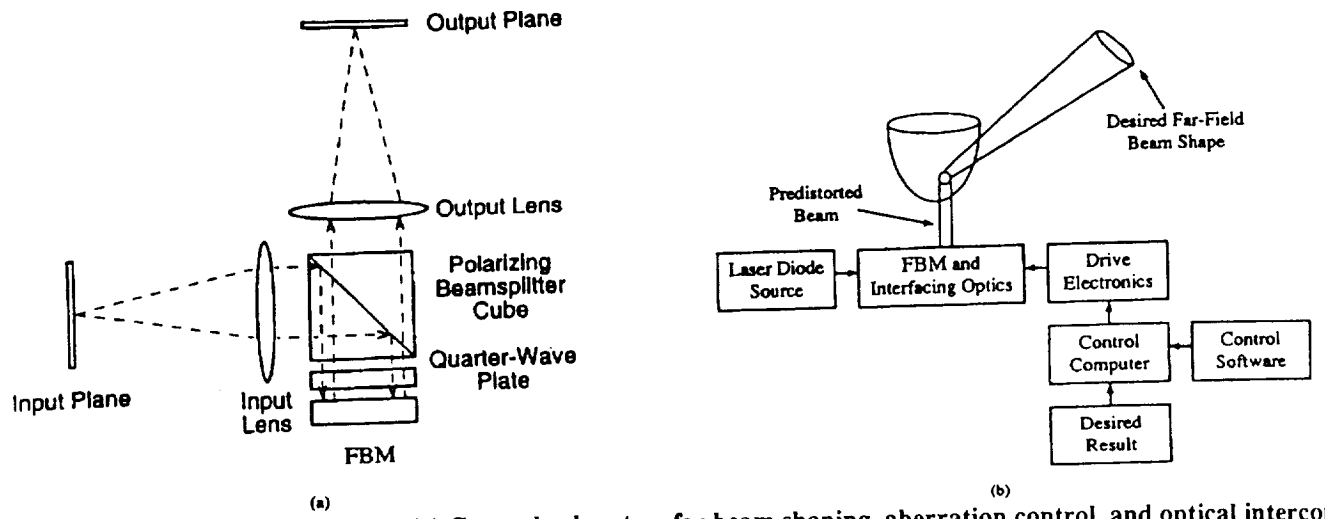


Figure 8. Active diffractive optics: (a) General subsystem for beam shaping, aberration control, and optical interconnection. (b) Example system for optical antenna aberration correction.

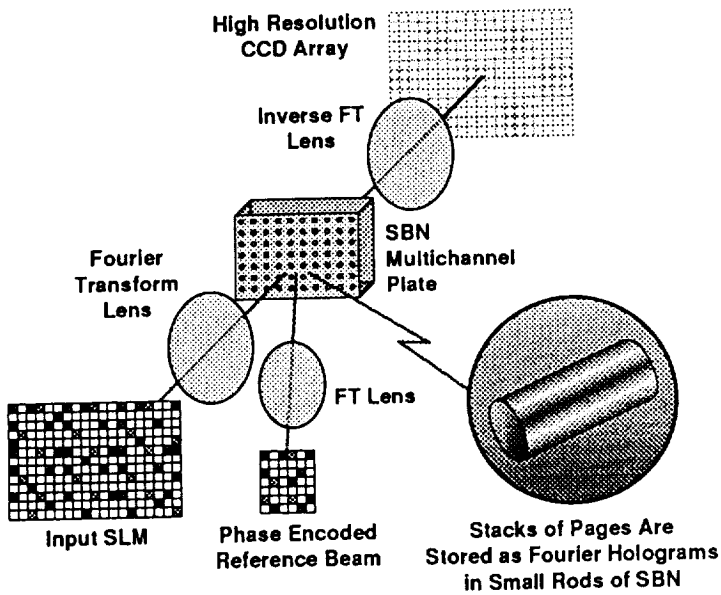


Figure 9. Schematic of holographic data storage system using binary amplitude SLM as data input, phase SLM as phase-encoded reference beam, SBN as holographic storage medium, and CCD as readout device.

IV. MEMBRANE DEVICES FOR INTEGRATED OPTIC APPLICATIONS

Photonic ON/OFF and routing switches can be made using alterations of the modal effective index caused by changing the cladding on a dielectric waveguide. Drumhead-like aluminum membranes are formed suspended over passive waveguides. The membranes are electrostatically pulled into contact with the waveguides, with operating times of tens to hundreds of microseconds. A routing switch can be made using the change in the real part of the modal index, as shown schematically in Figure 10. A directional coupler is fabricated in the waveguide layer to have a coupling ratio of unity; i.e., all the light is coupled from the input to the "cross" channel when the membrane is undeflected. Another way of saying this is that the air-clad directional coupler is made exactly one coupling length long. When a membrane is pulled into contact with the cross-channel waveguide, the effective index of that channel changes so as to destroy the synchronism of the coupler. If the index change is sufficiently large, negligible cross-coupling occurs, and all the light remains in the "bar" output channel. Although the synchronism could be altered sufficiently to cause switching by deflecting a membrane over either of the coupled guides, it is preferable to pull the cross-channel membrane down so that any losses caused by the absorption of the metal film will occur in

the cross channel, thus enhancing the crosstalk performance in the bar state. A simple 2-D beam propagation method (BPM) model has shown that complete switching is obtained for the index changes available in this system, even with an air gap of 0.1 μm under the membrane.

Figure 11 shows the membrane Mach-Zehnder interferometer optical switch. It consists of two 3 dB couplers connected by two equal-length straight waveguides. An aluminum membrane suspended a few micrometers above one of the two interferometer arms acts as a phase shifter. The input optical light is divided into two equal components by the first 3 dB coupler. One of the two components is then modulated in phase by the effect of an aluminum membrane. With the membrane up, the two wave components experience an identical phase shift; the optical power crosses over and leaves through the "cross" output channel. As the metal membrane is electrostatically pulled into contact with one of the two arms of the interferometer, the effective refractive index of the guided mode in that arm changes. If the index change is sufficiently large to introduce a π phase shift difference between the two wave components, all the optical power will go straight across, emerging from the "bar" output channel.

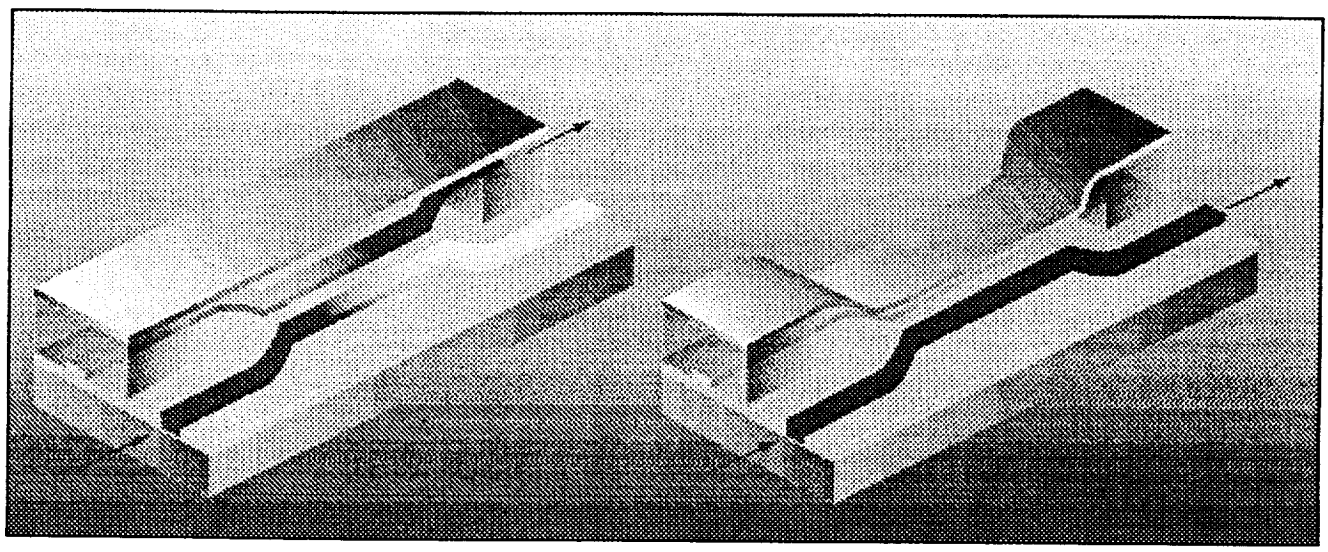


Figure 10. Routing switch based on a directional coupler. (a) Membrane undeflected: output directed to cross channel by the directional coupler; (b) Membrane deflected to touch cross channel: coupler unbalanced, light remains in the bar channel.

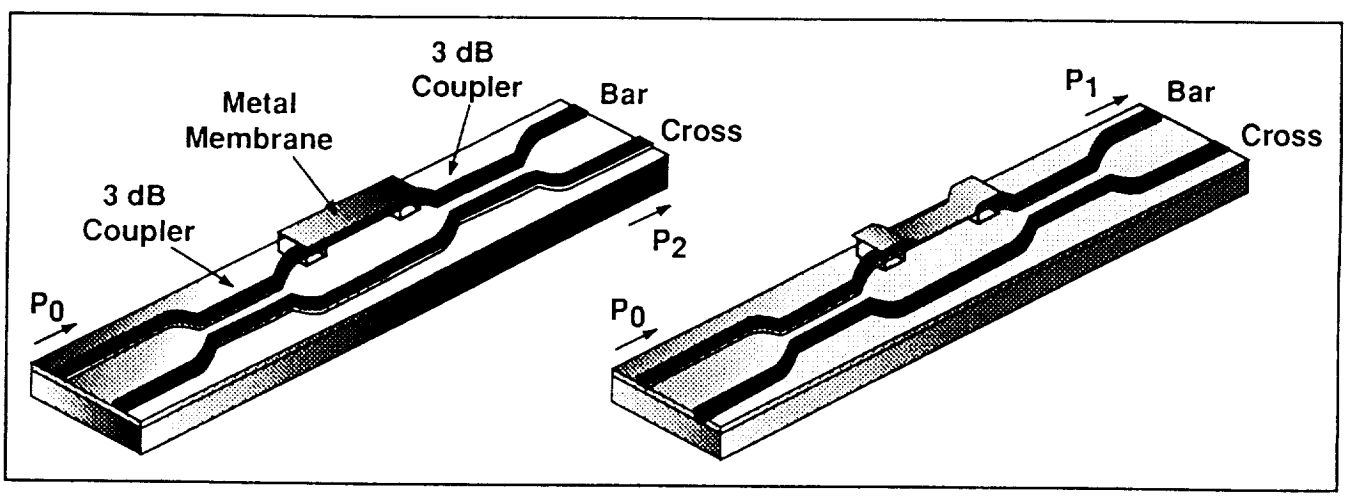


Figure 11. Metal membrane optical switch based on a Mach-Zehnder interferometer: (a) membrane up, (b) membrane down. Membrane may be placed over either channel in the center region of the switch.

V. DRUMHEAD DEVICE FOR RF APPLICATIONS

The first micromechanical switch was introduced in 1979.[20] However, micromechanical switches were not intensively considered for system applications until the early 1990s.[21-24] Surface micromachining and batch process technology became more mature and cost-effective, which made the device more attractive. TI investigated intensively in the early 1990s also, and obtained encouraging results.[24] Our switching elements are all based on the same metal membrane material as TI's micromirrors. The membranes are actuated by dc potentials to make or break the path of RF or microwave signals. By selecting proper materials and dimensions for the membrane and electrodes, the switches can be used to switch the signals at reasonable switching voltages.

From a functional point of view, the micromachined switches can be divided into resistive and capacitive switches (Figure 12). Our resistive switch yielded 1.5 to 2.5 ohms of dc resistance. For our long-term goals, capacitive switches are simpler and use lower voltage and power for switching because they do not need to recess the actuating electrodes, and a larger percentage of voltage and power for electrostatic actuation. The first capacitive switch (Figure 13) shows switching action in capacitive RF reflection measurement.

The compatibility of membrane switch construction with silicon CMOS processing makes this RF switch an attractive candidate for microwave integration with other passive RF devices. Low-cost microwave phase shifters can be fabricated using this technology and incorporated on the same substrate with other active or passive components.

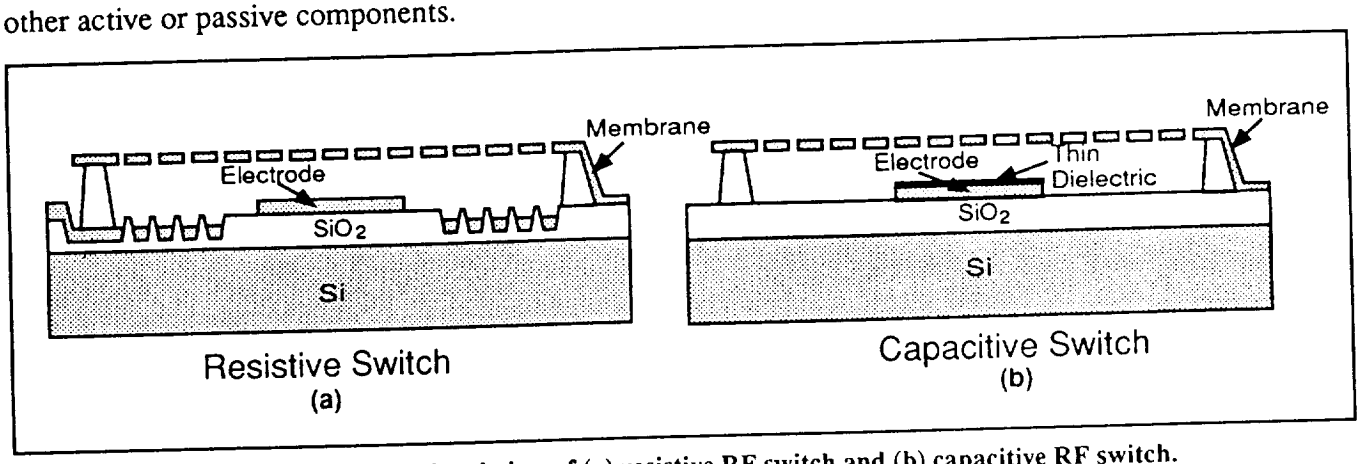


Figure 12. Cross-sectional view of (a) resistive RF switch and (b) capacitive RF switch.

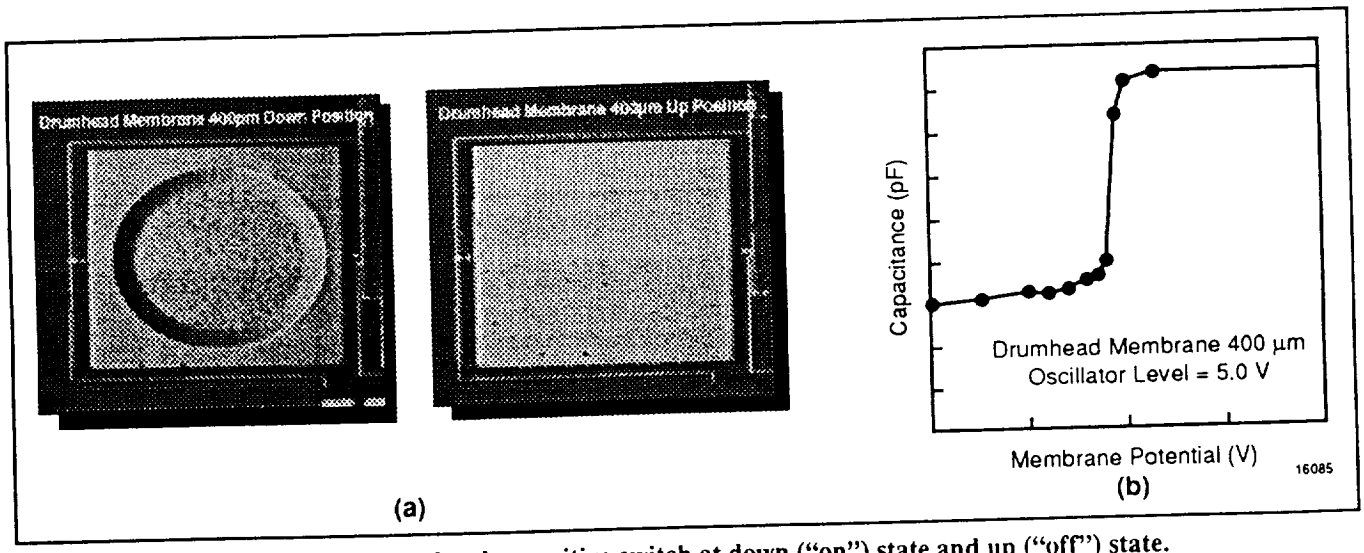


Figure 13. (a) Drumhead capacitive switch at down ("on") state and up ("off") state. (b) Result of the capacitance change versus actuation voltages.

VI. COMPUTER-AIDED DESIGN

Fabrication equipment and processes are very expensive. To minimize trial-and-error or major mistakes, computer-aided design (CAD) tools have been widely used in many industries (e.g., SPICE and SUPREM for semiconductors and ABAQUS for mechanical design). Micromachining is in its infancy, but demands for understanding of electromechanical and magnetomechanical micromechanisms are growing rapidly. Simulation tools were developed at different institutes based on specialized needs. For general-purpose micromachining, MIT's MEMCAD[25] and the University of Michigan's CAEMEMS[26] are the best known CAD tools. TI became one of the first two industrial beta sites[27] for MIT's MEMCAD. TI also worked closely with MIT to design test structures to characterize and monitor material properties such as Young's modulus, stress, and Poisson ratio. These parameters can then be used in both micromechanical simulation and process control monitoring. (We have obtained some preliminary results but they are inconclusive.) Design and process revision are under development. Figure 14 shows the MEMCAD is capable of taking a layout directly and converting it into a mechanical structure through process description file. MEMCAD can then mesh the structure and use finite element methods to simulate mechanical behavior of the device.

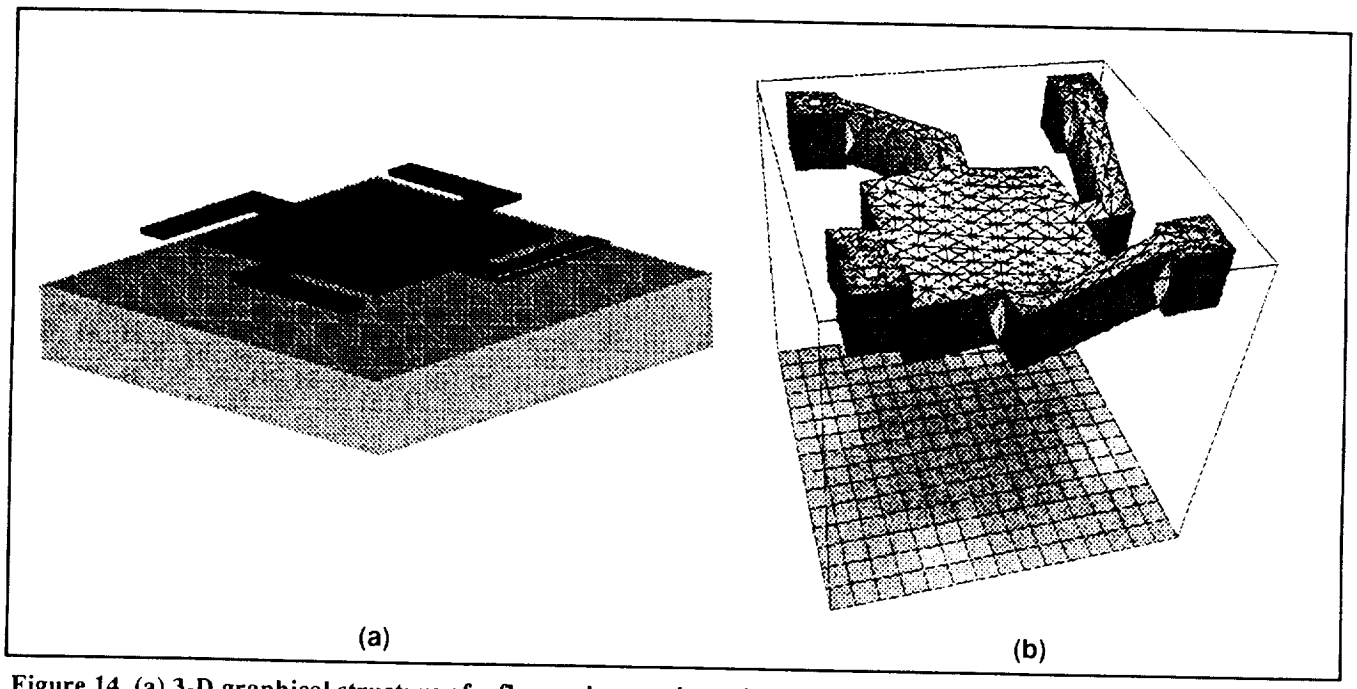


Figure 14. (a) 3-D graphical structure of a flexure-beam micromirror converted from layout and process description file. (b) Finite element method and electromechanical simulation of the micromirror in MEMCAD.

VII. SUMMARY

TI has developed IC process compatible micromechanical devices for more than 15 years. Some devices were successfully transferred to productization, while many opportunities remain to be explored. We covered in this paper: (1) the development of high-performance fiber-optic crossbar switches from optical table version to integrated packaging version; (2) new flexure-beam micromirror phase modulators and applications for image processing, optical interconnect, beam forming, and holographic data storage; (3) membrane devices for integrated-optic application by using the contact of membrane to the waveguide to change the effective refractive index; (4) RF switch as the only electrical application that can impact the telecommunication technology through its favorable performance and cost effectiveness; and (5) collaboration with MIT to boost our simulation capability using MEMCAD. We expect the micromirror/membrane technology to be an important technology in the 21st century. Opportunity and applications are waiting for development.

REFERENCES

1. K.E. Peterson, "Micromechanical Light Modulator Array Fabricated on Silicon," *Appl. Phys. Lett.*, Vol. 31, p. 521, 1977.
2. A. Vanderlugt, "Signal Detection by Complex Spatial Filtering," *IEEE Trans. Info. Theory,* Vol. IT-10, p. 139, 1964.
3. J.L. Homer, *Optical Signal Processing.* San Diego, CA, Academic Press, 1987.
4. T.G. McDonald, R.M. Boysel, and J.B. Sampsell, "4 × 4 fiber-optic crossbar switch using deformable mirror device," *OSA Tech Digest,* Vol. 14: Spatial Light Modulators and Applications, p. 80, 1990.
5. G.A. Magel, T.H. Lin, L.Y. Pang, and W.R. Wu, "Integrated Optic Switch for Phased Array Applications Based on Electrostatic Actuation of Metallic Membrane," *Proc. SPIE,* Vol. 2155 p. 107, 1994.
6. L. Pang. J. Leonard, G.A. Magel, S. Eshelman, and T.H. Lin, "Integrated Optical Switch for Radar Phase Control," *Proc. SPIE*, Vol. 2236, p. 96, 1994.
7. J.M. Florence, T.H. Lin, W.R. Wu, and R.D. Juday, "Improved DMD Configuration for Image Correlation," *Proc. SPIE*, Vol. 1296, p. 101, 1990.
8. R.S. Muller, R.T. Howe, S.D. Senturia, R.L. Smith, and R.M. White, ed. *Microsensors,* IEEE Press, NY, 1991.
9. *Proc. of 8th International Conf. on Solid State Sensors and Actuators, and Eurosensors IX.* Stockholm, June 1995.
10. J.M. Younse, "Mirrors on a Chip," IEEE Spectrum, p. 27, November 1993.
11. W.E. Nelson and L.J. Hornbeck, "Micromechanical Spatial Light Modulator or Electrophotographic Printers," *Proc. SPSE 4th Int'l Cong. Adv. in Non-Impact Printing Tech.*, p. 427, 1988.
12. T.-H. Lin, "Implementation and Characterization of a Flexure-Beam Micromechanical Spatial Light Modulator," Opt. Eng., Vol. 33, p. 3643, 1994.
13. J.M. Florence and R.O. Gale, "Coherent optical correlator using a deformable mirror device spatial light modulator in Fourier plan," *App. Opt.*, Vol. 27, p. 2091, 1988.
14. J.M. Florence, T.H. Lin, W.R. Wu, and R.D. Juday, "Improved DMD configuration for image correlation," *Proc. SPIE,* Vol. 129, p. 101, 1990.
15. J.M. Florence, "Joint transform correlator systems using deformable mirror spatial light modulators," Opt. Lett, Vol. 14, p. 341, 1989.
16. R.M. Boysel, "A 128 ¥ 128 frame-addressed deformable mirror spatial light modulator," *Opt. Eng.*, Vol. 30, p. 1622, 1991.
17. T.H. Lin and M.R. Feldman, "Reconfigurable optical interconnects for multichip module using spatial light modulator and computer-generated holograms," *Proc. Symposium on High Density Integration in Comm. and Comp. Syst.*, Vol. 44, 1991.
18. T.A. Rhoadarmer, S.C. Gustafson, G.R. Little, and T.H. Lin, "Flexure-beam micromirror spatial light modulator for acquisition, tracking, and pointing," *Proc. SPIE,* Vol. 2291, p. 13, 1994.
19. L. Hesselink and M.C. Bashaw, "Optical memories implemented with photorefractive media" Opt. and Quantum Elect., Vol. 25, p. 5611, 1993.
20. K.E. Peterson, "Micromechanical membrane switches on silicon," *IBM J. Res. Develop.*, Vol. 23. p. 376, 1979.
21. H. Hosaka, H. Kuwano, and K. Yanagisawa, "Electromagnetic microrelays: concepts and fundamental characteristics," *IEEE Microelectromechanical Systems Workshop. Tech Digest,* p. 12, 1993.

22. L.E. Larson, R.H. Hackett, and M.A. Melendes, "Micromachined microwave actuator (MIMAC) technology—A new tuning approach for microwave integrated circuits," *Proc. IEEE Microwave and Millimeter Wave Monolithic Circuit Symp.*, p. 27, 1991.

23. J.J. Yao and M.F. Chang, "A surface micromachined miniature switch for telecommunications with signal frequency from dc up to 4 GHz," *8th Int'l Conf on Solid-State Sensors and Actuators Tech Digest*, Vol. 2, p. 384, 1995.

24. C. Goldsmith, T.-H. Lin, B. Power, W.-R. Wu, and B. Norrell, "Micromechanical membrane switches for microwave applications," *IEEE MTT-S Digest*, p. 91, 1995.

25. S.D. Senturia, R.M. Harris, B.P. Johnson, S. Kim, K. Nabors, M.A. Shulman, and J.K. White, "A computer design system for microelectromechanical systems," *IEEE J. MEMS*, Vol. 1, No. 1, p. 3, 1992.

26. S.B. Crary and Y. Zhang, "CAEMEMS: An integrated computer-aided engineering workbench for micro-electro-mechanical systems," *Proc. IEEE MEMS*, p. 113, 1990.

27. S.D. Senturia and J.R. Gilbert, MIT, March 1995.

Software Component Technologies and Space Applications

Don Batory
Department of Computer Sciences
The University of Texas
Austin, Texas 78712

71096
230609
6P

Abstract

In the near future, software systems will be more reconfigurable than hardware. This will be possible through the advent of software component technologies, which have been prototyped in universities and research labs. In this paper, we outline the foundations for these technologies and suggest how they might impact software for space applications.

1 Introduction

Software component technologies will fundamentally change the way complex and customized software systems will be designed, developed, and maintained. Well-understood domains of software (e.g., avionics software, communication networks, operating systems, etc.) will be standardized as libraries of plug-compatible and interoperable components. A software system in these domains (e.g., a particular avionics system, a particular operating system, etc.) will be specified as a composition of components. High-performance source code that implements these systems will be generated automatically. Application developers will purchase component libraries for the domains of interest, and will configure components to build the target systems/platforms that their applications need. The evolution of software, once a formidable problem, is radically simplified: an updated version of a system is defined as a composition of components and its software is generated automatically. It is in this manner that future software engineers will leverage off of existing componentry to "mass produce" complex and customized software quickly and cheaply.

For this vision to become a reality requires basic changes in the way we understand and write software. First and foremost, a software component technology requires us to address the following:

- **encapsulation** - what should a building block (i.e., component) of software systems encapsulate?
- **composition** - what does composition mean?
- **paradigm** - what model of programming supports software component technologies?
- **scalability** - how can large families of systems be expressed by a small number of components?
- **verification** - how can one verify that a composition of components is consistent and implements the specifications of the target system?

Answers to these questions lie at the confluence of a number of independent research areas: transformation systems, object-oriented programming, parameterizing programming, domain-specific compiler optimizations, and domain modeling and the design of reusable software. Research on software system generators lies at the heart of this intersection, where specific and practical answers to these questions have been found.

2 A Paradigm Shift

The evolution of customized software is the bane of most projects: it is difficult to achieve and is horrendously costly. There is always the need to develop new variants of existing systems, each variant/version offers new features that are specific to the class of applications that are to be supported. But often the effort needed to make even minor changes to a system is far out of proportion to the changes themselves [Par79].

The problem is that source code is the most detailed and concrete realization of a software design. The most critical changes (and hence the most important evolutionary changes) to a software system are modifications to its design. Minimal design changes often require major software rewrites. Rather than maintaining and evolving source code, an alternative is to maintain and evolve the *design* and to *generate* the corresponding source code automatically. This is the concept of *design maintenance* [Bax92].

Design maintenance asserts that the designs of software systems are quantized; there are primitive components of software design in every domain. A component encapsulates a domain-specific capability that software systems of that domain can exhibit. The design of a software system is therefore expressed as a composition of components, where a composition defines the set of capabilities that a target system is to have. Furthermore, the evolution of a system's design occurs in quantum steps and these steps correspond to the addition or removal of domain-specific capabilities (i.e., components) from the target system.

Design maintenance has two important implications. First, conventional methods of software design must change because they view software systems as one-of-a-kind products. Reusing previous designs or source code is largely an ad hoc and fortuitous activity. Design maintenance, in contrast, requires the identification of primitive components of design for a large family (or domain) of software. A primitive component, by definition, is reusable because it is used in the design of many family members. *Domain modeling* is the name given to software design methodologies that identify primitive components of software designs for a specific domain [Pri91, Gom94, Bat95a].

A second implication is that a primitive component of software design need not correspond to a primitive code module or package in a generated system. In general, the introduction of a component to a system's design might require incremental modifications to many parts (e.g., object-oriented classes) of a system's software. Furthermore, the modifications that a component makes to the source code of one system might be different than that made to another; such differences arise because certain domain-specific optimizations could be applied to one system, but not in the other. Thus a component must encapsulate more than just algorithms: it must also encapsulate *reflective* computations, i.e., domain-specific decisions about when to use a particular algorithm and/or when to apply a domain-specific optimization. For most domains, reflective computations are critical for generating efficient code [Bat93].

What programming paradigm supports such componentry? The rallying cry of object-orientation is that "everything is an object". Object-oriented design methodologies and programming languages are indeed powerful, but they are insufficient for software components. A programming paradigm that has been found to encompass object-orientation, in addition to providing the generality needed, is that of *program transformation systems*. The rallying cry of transformation systems is "everything is a transformation", or more specifically a *forward refinement program transformation (FRPT)*. The connection between components and FRPTs is direct: an FRPT elaborates a high-level program by introducing details (e.g., source code) that efficiently implement a domain-specific capability. Such elaborations can occur in many parts (e.g., classes) of a system's software. Moreover, an inherent part of an FRPT is the ability to perform reflective computations, so that only the most efficient algorithms are generated. Composing components is equivalent to composing transformations [Bax92, Bat92].

The key ingredient that enables components to be composed is due to a disciplined design that standardizes the abstractions (and their programming language interfaces) of a domain. Simply put, domain modeling ensures that components are designed to be interoperable, interchangeable, and plug-compatible and thus can be used as software building blocks; components with ad hoc interfaces that are not interoperable, interchangeable, and plug-compatible are not building blocks.

Among the benefits of software componentry is that few components are actually needed to assemble large families of systems. We expect most domain-specific libraries to have a few hundred components, where domain experts can easily identify components to be used (without requiring elaborate library classification and searching methods). Another benefit of software componentry is that there are simple algorithms to determine automatically if a composition of components is consistent and that it implements the specifications of a target system. While demonstrating consistency falls short of formal verification, it is an major step forward in making software system generation practical [Per89, Bat95b].

Software component technologies and generator technologies have been developed for the domains of avionics, database systems, file systems, network protocols, and data structures. Related composition/encapsulation technologies in software architectures are [Gor91, Per92, Gar93]. Readers who are interested in the technical details of these discussions are urged to consult the cited references.

3 Relevance to Software Development for Space Applications

In the following, I address the community of software developers for space applications. However, I admit that there is very little in my comments that are specific to space applications; the problems that I address and the benefits that can be reaped are applicable to software in general.

The main obstacles I foresee in the promulgation of software component technologies and software system generators are not technical in nature. To be sure, there are plenty of difficult technical problems ahead, but I am confident that these problems are solvable. My intuition for this not-very-bold statement is that domain modeling takes a retrospective view of software systems that have been built. Thus, solutions to thorny design problems have already been devised in a multiplicity of contexts. The activities of domain modeling - the basis of software component technologies - are to show how these specific solutions fit into a more general (i.e., building blocks) context. It is not the case that entirely new solutions to domain-specific problems (e.g., space applications) must be invented for software component technologies to work. Very little "invention" of new algorithms, etc. is needed. Hence my optimism.

The real challenge will be the acceptance of software component technologies by the space application community. The primary obstacle is that programmers and system designers are reluctant to change the way they understand and view problems in software. More specifically, this is the "not-invented-here" syndrome. If software componentry for space applications were invented in-house, it would have a much greater chance of being used. But even in-house development would be a major step from conventional approaches.

The reluctance to change has its consequences: researchers will be encouraged to seek a "silver bullet" that will miraculously solve intractable problems that have been brought on by traditional and established methods of software production. The difficulties of software evolution; the infeasibility of implementing competing, possibly radically different, designs for evaluation; the inexpensive development of product families are examples. Experience has shown that enough (minimal) progress and enough clever ideas will be demonstrated by researchers to keep the "silver bullet" hopes of software managers alive for years to come. However, I am skeptical that incremental progress will ever lead to a satisfac-

tory and economical solution. To address the major problems of software development today will ultimately require a paradigm shift.

Paradigm shifts occur when there is a general perception that major benefits will ensue. The shift from structured programming and C-like programming languages to object-oriented design methods and programming languages is being paved by a wide spectrum of realized benefits and good salesmanship of object-orientation. The same will be needed for software component technologies. The benefits of componentry are real and substantial, but are not yet that well understood or appreciated. Not surprisingly, number of advocates for software componentry needs to be enlarged.

Despite my enthusiasm for software componentry, I don't believe software component technologies are silver bullets. These technologies do not solve problems, but only simplify some problems (e.g., evolution). For example, there are there are many performance-related parameters in avionics software whose values must be determined through extensive testing and simulation. Avionics source code with such performance-related parameters is typically easy to generate. However, how one determines the values to be assigned to these parameters (e.g., aircraft-specific parameters) does not seem to be fully automatable, and the tried-and-true processes of testing and simulation still need to be performed. Thus, many of the existing activities of software development will still remain.

Another point to be made is that most "new" systems always include new and unprecedented functionality. It has been estimated that upwards of 80% of a "new" system can be built from available components. This means that 20% of the "new" system will need to be added. While a factor of five reduction in the amount of software to be written is a substantial savings, software development will certainly not cease.

But there will also be unique opportunities that software component technologies provide that would otherwise be difficult or impractical. For example, synthesis from specifications makes it feasible to evaluate radically different software designs. As another example, self-tuning and self-reorganizing software is possible: components can be added to systems to monitor their performance. Periodically, the system can reconfigure itself automatically, based on known usage patterns, to enhance its performance.

In conclusion, if software evolution, the cost-effective creation of product families, the need to experiment and retrofit system designs, and improving programmer productivity are critical to future software for space applications, then the design and use software component technologies should be made a top priority.

4 References

[Bax92] I. Baxter, "Design Maintenance Systems", *CACM* April 1992, 73-89.

[Bat92] D. Batory and S. O'Malley, "The Design and Implementation of Hierarchical Software Systems with Reusable Components", *ACM TOSEM*, October 1992.

[Bat93] D. Batory, V. Singhal, M. Sirkin, and J. Thomas, "Scalable Software Libraries", *ACM SIGSOFT* 1993.

[Bat95a] D. Batory, L. Coglianese, M. Goodwin, and S. Shafer, "Creating Reference Architectures: An Example From Avionics", *ACM SIGSOFT Symposium on Software Reusability*, Seattle, 1995, 27-37.

[Bat95b] D. Batory and B.J. Geraci, "Validating Component Compositions in Software System Generators", Department of Computer Sciences, University of Texas at Austin, TR-93-03, February 1995.

[Gar93] D. Garlan and M. Shaw, "An Introduction to Software Architecture", in *Advances in Software Engineering and Knowledge Engineering*, Volume I, World Scientific Publishing Company, 1993.

[Gom94] H. Gomaa, L. Kerschberg, V. Sugumaran, C. Bosch, and I. Tavakoli, "A Prototype Domain Modeling Environment for Reusable Software Architectures", *Third International Conference on Software Reuse, Rio de Janeiro*, November 1-4, 1994, 74-83.

[Gor91] M.M. Gorlick and R.R. Razouk, "Using Weaves for Software Construction and Analysis", *Proc. ICSE* 1991, 23-34.

[Par79] D.L. Parnas, "Designing Software for Ease of Extension and Contraction", *IEEE Transactions on Software Engineering*, March 1979.

[Per89] D.E. Perry, "The Logic of Propagation in The Inscape Environment", *ACM SIGSOFT 1989*, 114-121.

[Per92] D.E. Perry and A.L. Wolf, "Foundations for the Study of Software Architecture", *ACM SIGSOFT Software Engineering Notes*, October 1992, 40-52.

[Pri91] R. Prieto-Diaz and G. Arango (ed.), *Domain Analysis and Software Systems Modeling*, IEEE Computer Society Press 1991.

AVERTING DENVER AIRPORTS ON A CHIP

Software Engineering Research for Space Applications of MEMS
(position paper)

Kevin J. Sullivan
Department of Computer Science
University of Virginia
Charlottesville, VA 22903

sullivan@virginia.EDU
(804)982-2216

71098
10 p

INTRODUCTION

Recent years have seen important advances in our software engineering capabilities. As a result we are now in a more stable environment than in the recent past. *De facto* hardware and software standards are emerging. Work on software architectures and design patterns [GS95,GHJV95] signals a consensus on the importance of early system-level design decisions, and agreement on the uses of certain paradigmatic software structures. We recognize the non-existence of simple solutions to complex software development problems. We now routinely build systems that would have been risky or infeasible a few years ago.

Unfortunately, technological developments threaten to destabilize software design again. Systems designed around novel computing and peripheral devices will spark ambitious new projects that will stress current software design and engineering capabilities. Micro-electro mechanical systems (MEMS) and related technologies provide the physical basis for new systems with the potential to produce this kind of destabilizing effect. Software will, with high probability, be the "pacing," high-risk item for many such systems.

One important response to anticipated software engineering and design difficulties is carefully directed engineering-scientific research. Although no single advance in software engineering will suffice for fast cycle time production of dependable software for future systems, the coordinated application of results from several lines of research should be of substantial value. Recent research has identified promising lines of attack on key problems. Two specific problems meriting substantial research attention are the following:

- We still lack sufficient means to build software systems by generating, extending, specializing, and integrating *large-scale reusable components*.
- We lack adequate computational and analytic tools to extend and aid engineers in maintaining *intellectual control* over complex software designs.

DISCLAIMER

In this paper I elaborate on the claim that advances in MEMS and related technologies will destabilize software design. I offer practical advice on how to deal with the problem. And I discuss research at the University of Virginia that is relevant to the above problem formulations.

My research does not address MEMS *per se*. I am not expert in the area. The positions in this paper are informed opinion. It is impossible even for experts to predict the future impacts of given technological developments, and it is far too easy to be way too optimistic or pessimistic. This position paper is thus an exploration of ideas, not a wager on future outcomes.

While not an expert, I do work in a context in which MEMS is an emerging issue. My research falls within the scope of an *end-to-end systems* research program joint between the electrical engineering and computer science departments. In the next phase of this program, we will target MEMS applications. Work to date has been at the device level: We are designing a sensor to detect chlorine ions to help address the problem of corrosion of the rebar in concrete bridges. The devices will be placed in the concrete. When readings exceed a threshold, electricity will be applied to drive off the chlorine.

RESEARCH OVERVIEW

As this simple device and related devices begin to mature, design and implementation of software to manage them will become a more important issue. As I discuss below, rapid advances in devices will require new work in software engineering. At present, proposed devices are quite limited. The devices we are proposing have a sensor, simple CPU, small memory, and small antenna: hardly a challenge to the best software engineers. These simple devices will be good for initial, small-scale software engineering research for MEMS.

As we scale up the complexity of systems based on new technologies, it will be critical to address the difficult software engineering issues already holding us back in the most demanding current efforts (e.g., advanced, space-based telecommunications systems). We will shortly require major improvements in both software development throughput (productivity), latency (cycle time in response to new requirements), and dependability.

My current work comprises a number of promising attacks on these problems. My primary focus has been on the integration of *independent components* into systems [Sul92,Sul94]. Recent variants focus on integration of *large-scale components*, including, *commercial off-the-shelf components* [Sul95a]. Productivity and dependability demand the reuse of certified, large-scale components; dependability and flexibility demand advanced integration techniques. In the tools area, I focus on the need for rapid development of custom, high-confidence *computational and analytic tools*. Such tools are needed to aid engineers in obtaining and maintaining intellectual control over complex software systems [Sul95b]. Advances in the two broad areas of component-based development and tools promise to significantly benefit future system development projects.

TECHNOLOGICAL DISCONTINUITIES DEMAND RESEARCH

Technological discontinuities demand engineering research. If the technological situation changes by an order of magnitude, you have a whole new set of research problems [Wulf95]. Old solutions to known problems have to be rethought. New problems emerge. Yet, while order-of-magnitude change is costly, it also presents enormous opportunities.

Petroski presents the pencil as a paradigm of technological discontinuity [P89]. In 1793, unavailability of English graphite due to the onset of war forced Continental pencil manufacturers to engage in engineering research and development to find alternative ways of making pencil leads [P89]. The French Minister of War, Carnot, commissioned Nicolas-Jacques Conte to develop alternatives. Taking a "deliberately innovative" approach that united "the scientific method... with experience and with the tools and products of craftsmen," Conte developed the modern ceramic method for making pencil leads.

Two hundred years later, the computer revolution—a six-orders-of-magnitude improvement in computing machines over thirty years, produced in large part by "the transformation of computer manufacture from an assembly industry into a process industry [B95]"—drove the need for a significant new engineering research program. As Dijkstra noted, "as long as there were no machines, programming was no problem at all; when we had a few weak computers, programming became a mild problem, and now we have gigantic computers, programming has become an equally gigantic problem [D72]." Software design ambitions scale with device capabilities; but software design abilities do not!

The resulting "software crisis," was characterized by many costly engineering failures. As in 1793, so in 1968 the need was countered by an aggressive engineering research program: The NATO Science committee established software engineering as a discipline. Problems with large systems are still a serious concern [G94], but research has produced made much progress toward understanding software system design in the small and large.

DO MEMS BETOKEN A NEW DISCONTINUITY?

The question I ponder in which paper is this: Do MEMS devices betoken a similar technological discontinuity—one that will destabilize software design? Will ambitious future systems based on MEMS and related technologies exceed our software engineering capabilities? If so, what is the proper response?

A case for an affirmative answer includes several points. First, MEMS interact with the physical world, requiring complex real-time behaviors; should MEMS become common, they will raise the average complexity of programs to be designed, increasing demands for scarce good software designers. Second, MEMS do seem destined to become both more common and much more complex, for the same reasons that computers became common and complex: transformation of manufacture from an assembly industry to a process industry. Third, the ability to produce MEMS at low price and in high volumes will encourage the design of systems incorporating many complex devices. This—in the area of large, distributed systems—is where real software engineering difficulties will begin.

The title of this position paper suggests "Denver Airports on a Chip" as a metaphor for the difficulty of programming advanced future systems. As a paradigm of failure owing to software difficulties, one can do little better than the Denver Airport baggage system. That system provides a complexity baseline for assessing potential future software difficulties.

The key problem at Denver was the difficulty of programming a "central nervous system of some 100 computers networked to one another and to 5,000 electric eyes, 400 radio receivers and 56 bar code scanners [G94]." The ambitiousness of the concept outstripped the software engineering capabilities of the developers. My point is *not* to criticize the developers but to pose the question, how do applications now being envisioned compare to Denver; can we use a comparison of physical characteristics (numbers and kinds of devices and interconnects) as a rough way to gauge potential difficulties?

At least some of the applications now being discussed appear to match or surpass "Denver" in complexity. A great deal of the complexity of ambitious future systems will be in the software; and if miniaturization succeeds, we're going to have many more such systems. As a case study in failure due to software engineering difficulties, Denver Airport is invaluable to future system designers: The *key* is to avert software engineering failures.

SOFTWARE ENGINEERING RESEARCH FOR MEMS

How can the risk of software engineering failure be managed? The most important point is to recognize that software engineering difficulties are *a* top risk, and probably *the* top risk, facing advanced technology projects. These risks must be managed aggressively.

A two-pronged approach is needed. First, use best current practices. Take an iterative approach to risk management in which you continually reevaluate risks and address the most serious ones first [Boehm76]. Understand that good management is essential. Hire great software designers [B95]. Recognize there is no silver bullet: No single technique or advance will radically simplify the software problem. Second, understand that we still haven't resolved certain key foundational software engineering research issues (e.g., how to build systems from large-scale reusable components). Progress in these research areas (properly employed) will contribute significantly to success.

Fortunately, in my view, past software engineering research has laid the foundations for new syntheses that will help us to meet the demands for software for future systems with increasing confidence. In the rest of this paper, I discuss my research in two areas. The first attack—building software systems by integrating independent, large-scale components—targets the problem of *throughput, latency, cost in development, and dependability in the resulting product.* The second attack—rapid development of high-confidence analytic and computational software engineering tools—targets the need to help engineers to extend and maintain *intellectual control* over complex software designs.

Component-based software development

Building software by integrating independent components is critical for at least three reasons. First, it achieves a separation of concerns essential for both intellectual and managerial control of complex systems. Second, it amortizes development costs to the

extent that components are reused. Third, in some sense it permits one to meet the need for exponential demands for software by combining a small number of parts in different ways.

Many component-based approaches have been devised and used, e.g., procedures, pipes and filters, and objects. Structured programming does not support composition of *large-scale* components—e.g., commercial off-the-shelf (COTS) application-sized parts. Unix Pipes-and-filters are inadequate for *interactive* systems. Classical object-oriented approaches do not support very well the integration of visible, stand-alone objects [Sul94].

I explore relevant engineering issues in the context of a specific design approach, called the mediator method [Sul94]. In this method, requirements are mapped to an architecture called a behavioral entity-relationship (ER) model. The nodes in such a model represent independent, visible *behaviors;* the edges represent *behavioral relationships*—potentially complex ways in which the behavioral components are required to interact. The next step is to map the behavioral ER model onto a set of components—such as C++ objects or COTS applications—in a way that preserves the structure of the model. Components that implement behaviors are independent and visible; while separate components (called *mediators*) implement the behavioral relationships.

This approach embodies a view of both the static structure of an integrated system and the way that it evolves. In particular, this design approach is intended to provide an unusual degree of flexibility in composing and evolving integrated systems, overcoming a serious problem with common design methods: that they throw integration and evolution into conflict. The architectural model accommodates evolution by the addition, change and deletion of behaviors and behavioral relationships: one adds, changes, or deletes nodes and edges in the behavioral ER model, with corresponding changes to the implementation. One integrates a new component into a system, for example, by adding the component as an independent part, then adding new mediators to make it work with the existing parts. The existing parts don't have to change to work with the new one. The mediators take care of integration separately from the parts being integrated.

Results to date are encouraging. We have used the approach to develope a radiation treatment planning system for cancer patients [Sul95c] now in clinical use at several large research hospitals. The components—Common Lisp objects—model anatomy, radiation fields, graphical views, etc. The mediators integrate the component so that, for example, a change to a view results in a corresponding change the anatomical model. The approach was key to producing Prism on a modest budget—about eight person years [Sul95c].

I am now exploring the mediator integration of COTS components in a case study involving the design of a commercial-grade fault-tree analysis tool supporting a novel analysis techniques developed by my colleague Joanne Dugan. Components include Visio for drawing, Microsoft Access for storing and generating reports on fault trees, and other large-scale components. The mediators are in Visual Basic. I developed a prototype in about a week, and am now building a complete system. Delegating the interface and bookkeeping functions to volume-priced components will permit us to deliver, for a cost of under $1000, a serious new computational tool in a rich package comprising multiple millions of lines of code. This work is shedding much light on component integration issues.

While the applicability of the detailed mediator work to MEMS applications is not clear, the basic insights and component-integration structures will certainly be valuable. I am continuing research on component integration, pursuing topics including the following:

- identification of useful components and component types,
- techniques for specifying and implementing large-scale reusable components, and for specifying and implementing interconnection structures to integrate them [2,3],
- a discipline of programming with large-scale components, i.e., techniques for mappings application concepts onto integrated sets of components [5], and for characterizing the engineering tradeoffs involved in making the mapping decisions often required by shortcomings in components and integration mechanisms, and
- techniques for integrating components in the context of the heterogenous hardware environments of future systems, at the applications and operating systems levels.

Computational and analytic software engineering tools

Intellectual control is the cornerstone of dependability. This was true in Denver, and it will be true for future systems. Engineers have for centuries used computational and analytic tools to extend their understanding of complex structures. In the absence of such tools, the best choice may be not to build at all. Stephenson decided to employ a tubular bridge instead of a suspension bridge to span the Menai straights because he lacked the tools needed to understand why suspension bridges had failed in the past. His tubular design was an environmental, economic, and aesthetic failure; but he showed excellent engineering judgement by not trying to build a critical system beyond his intellectual reach [Sull95b].

In the future we will increasingly be asked to handle systems that stretch or exceed our intellectual abilities. The need is acute for tools to extend the intellect to help us assure the dependability of future systems.

Current tools exhibit at least two problems. First, it is too difficult to develop custom tools to answer specific but often fairly simple questions about software. Second, it is hard to validly interpret the outputs of many analysis tools. The latter fact is not as appreciated as it ought to be.

In a recent empirical study [MNL95], Murphy, Notkin and Lan report on significant divergences in the outputs of tools that compute static call graphs from C programs. Not only are the outputs different, but there is little or no indication that they should be; the input-output specifications are not clearly spelled out; and it is hard to infer or deduce what the specifications are. In the absence of such information, it is hard for the engineer to have much confidence in decisions made on the basis of tool results. That is, the tools give an impression of intellectual control, but they don't give real intellectual control—dangerous.

Figure 1 depicts a simple, prototype framework for the rapid development of custom program analysis tools. The "probe" is a commercial compiler front end plus an abstract "visitor class" in C++. Specializing the visitor enables extraction of selected information from the front end-generated intermediate representation (as indicated in the Figure by specializations for extraction of the includes hierarchy, call graphs, and global variable uses). The front end, owned by the Edison Design Group [EDG95], can be configured to parse many important variants of C and C++, and is highly optimized.

I typically define visitor subclasses to format and output Prolog facts [CM94] about the program to be analyzed—e.g., whether there is a call from procedure P to Q. The output is piped to the European Community Research Centre's Common Logic Programming System [E95]. This Prolog system includes a transactional relational database useful for storing large databases of facts; and it supports logic programming-based inference. After massaging the data produced by the probe, Eclipse constructs a "dot" program, which it then pipes to the "dot" program. "Dot" is a sophisticated, programmable graph layout system developed at ATT. The framework supports rapid generation of simple, custom analysis and structure visualization tools.

Instantiating the framework with less than 100 lines of C++ code and a few lines of Prolog enabled me to implement a static call graph extraction program. In contrast to current programs (whose outputs differ substantially and whose precise input-output specifications are unclear), my program realizes the following simple specification:

- if there is an arc in the extracted call graph from a node P to a node Q then there is a call instruction in procedure P with target procedure Q, and that if that instruction is executed there will be a runtime call from P to Q, and
- if there is no arc in the graph from P to Q then there is no possibility of a runtime call from P to Q, unless P is annotated to indicate that P contains indirect calls through pointers, in which case further (possibly manual) investigation is needed.

The function is not complex, and indeed it is simpler than the function of more sophisticated tools that perform more complex semantic analysis to refine the call graph. But no matter how sophisticated, if the engineer doesn't really understand how to interpret the tool output, the sophistication is not very useful. The key, of course, is not this particular tool or its component parts, but in the ability to rapidly generate families of customized tools to answer a range of questions about industrial systems. As I take this relatively new research forward, I am particularly emphasizing the following objectives:

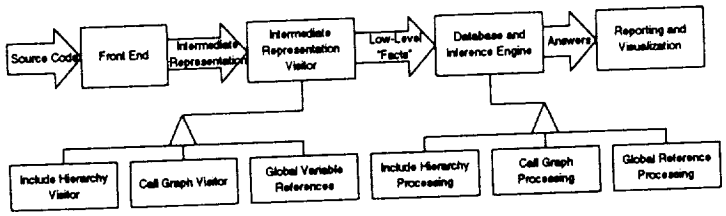


Fig. 1. - Architecture of the Tool Framework.

- To develop *probes* to gather signals from software systems (measurements, both static and dynamic), much as oscilloscope probes sample signals in electrical circuits. An example is a probe that extracts static call graphs from programs.
- To develop *analysis and presentation facilities* to prepare and display information derived from the captured signals. For example, I use a Prolog inference engine to draw conclusions from the raw outputs produced by probes.
- To develop architectural approaches that support rapid construction and operation of customized tools through *integration of large-scale reusable components*.
- To elaborate on the concepts of *evaluatability of tool results* as a key quality criterion, and of actual evaluation of tool results as a key responsibility of an engineer.

Again, the direct applicability of the specific work to MEMS-based applications is not clear. The point is that the computational and analytic tool situation for software engineering in general is woefully inadequate. The consequence is unnecessary limitation of our intellectual control over complex (or even just plain large) systems. Research and development in this area are important as we move into a domain of very ambitious projects.

CONCLUSION

Technology advances in both evolutionary and revolutionary ways. Most change is evolutionary, and the impacts of change are often less dramatic than predicted. Nevertheless, rapid advances can revolutionize the circumstances in which design is done, often requiring significant, new engineering research and development activities. The development of MEMS devices and related technologies seem likely to do this by sparking ambitious application concepts that will stress our software engineering capabilities. Best practices can help; but inadequacies in the foundations of software engineering demand that we also engage in aggressive software engineering research as a major part of our strategy to manage considerable software-related risks. The good news is that the basis for substantial progress has been laid by past research. The potential for technological change to catalyze major advances in software engineering is quite exciting.

The French had understood the risk of depletion of English graphite for many years before the war finally and quickly shut off the strategic material. Crisis finally prompted decisive research. We can anticipate and brace for the problems likely to be produced by new MEMS technologies, while ensuring that resources are not wasted on problems that don't materialize, with significant investments in software engineering research directed at relevant foundational issues. Mechanisms should be established to assure appropriation of the benefits of such research. Carefully directed investment in emerging MEMS-*specific* software engineering research issues is also warranted.

I have discussed my research in two areas: component-based software development (attacking cost, dependability, throughput and latency); and rapid construction of computational and analytic tools (attacking intellectual control). The technological change underlying this work—carefully developed, validated, and implemented, and thoughtfully combined with related research results—will help to avert "Denver Airports on a Chip."

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation under grant number CCR-9502029. Some of the work described in this document is joint with John Knight and Joanne Dugan.

REFERENCES

[B76] Boehm, B., "A Spiral Model of Software Development and Enhancement," Computer, May, 1988, pp. 61-72.

[B95] Brooks, F., *The Mythical Man Month (Anniversary Edition),* (Reading, Massachusetts: Addison-Wesley), 1995.

[CM94] Clocksin, W.F. and Mellish, C.S., *Programming in Prolog*, Springer Verlag, 1994.

[C90] Chen, Y.F. et al., "The C Information Abstraction System," *IEEE Transactions on Software Engineering,* SE-16(3):325-334, March 1990.

[D72] Dijkstra, E., *Programming Considered as a Human Activity,* in Classics of Software Engineering.

[E95] European Community Research Centre (ECRC), The Eclipse Common Logic Programming System, February, 1995.

[G94] Gibbs, W., "Software's Chronic Crisis," *Scientific American,* Sept., 1994.

[GS95] Garlan, D. and Shaw M., "An Introduction to Software Architecture," in V. Ambriola and G. Tortora, editors, *Advances in Software Engineering and Knowledge Engineering,* pp. 1-39, Singapore, 1993.

[GHJV95] Gamma, E. et al., *Design Patterns,* (Reading, Massachusetts: Addison-Wesley), 1995.

[MNL95] Murphy, G., Notkin, D., and Lan, E., "An Empirical Study of Static Call Graph Extractors," University of Washington Department of Computer Science and Engineering Technical Report UW-CSE-TR-95-08-01, August, 1995 (to appear in the Proceedings of ICSE-18, Berlin, March, 1996).

[P94] Petroski, H., *Design Paradigms,* (Cambridge: Cambridge University Press), 1994.

[P94] Petroski, H., *The Pencil: A History of Design and Circumstance*, (New York: Knopf), 1990.

[SN92] Sullivan, K.J., and D. Notkin, "Reconciling environment integration and software evolution", *ACM Trans. on Software Engineering and Methodology,* 1(3), July 1992.

[Sul94] Sullivan, K.J., "Mediators: Easing the design and evolution of integrated systems", Ph.D. dissertation and technical report CSE-TR 94-08-01, Department of Computer Science and Engineering, University of Washington, Seattle, WA, 1994.

[Sul95a] Sullivan, K.J. and Knight, J.C., "Assessment of an Architectural Approach to Large-Scale, Systematic Reuse," University of Virginia Department of Computer Science Technical Report, To Appear in Proceedings of ICSE18, Berlin, 1996.

[Sul95b] Sullivan, K.J., "Rapid Development of Simple, Customized Program Analysis

Tools," University of Virginia Department of Computer Science Technical Report CS-95-45, 1995 (submitted to the ICSE18 workshop on program comprehension).

[Sul95c] Sullivan, K.J., Kalet, I.J. and Notkin, D. "Mediators in a Radiation Treatment Planning Environment," University of Virginia Department of Computer Science Technical Report CS-95-29, 1995 (submitted to IEEE Transactions on Software Engineering).

[Wulf95] Wulf, W., Personal Communication, University of Virginia, 1995.

Session 3: Posters and Demonstrations (Monday, PM)

Session Chairs: Mark Holderman (NASA/JSC) and Robert Smith (Aerospace)

AIAA Spacecraft GN&C Interface Standards Initiative: Overview
by A. Dorian Challoner, Hughes

A High Average Power Free Electron Laser for Microfabrication and Surface Processing Applications
by H. F. Dylla, S. Benson, J. Bisognano, C. L. Bohn, L. Cardman, D. Engwall, J. Fugitt, K. Jordan, D. Kegne, Z. Li, H. Liu, L. Merminga, G. R. Neil, D. Neuffer, M. Shinn, C. Sinclair, and M. Wiseman, CEBAF, L. J. Brillson, Xerox, D. Henkel, Henkel Metall. Tech., H. Helvajian, The Aerospace Corporation, M. J. Kelley, DuPont, and Shanti Nair, University of Massachusetts

MEMS in Space Systems
by J.C. Lyke, M.A. Michalicek, and Babu Singaraju, Air Force Phillips Laboratory

Subminiaturization for Environmental Research Aircraft and Sensor Technology (ERAST) Instrumentation
by Marc Madou, U. C. Berkeley, M. Loewenstein and S. Wegener, NASA/Ames

DoD Space Test Program (STP)
by Major Llwyn Smith, U.S. Air Force Space and Missile Systems Center

Weaves as an Interconnection Fabric for ASIMs and Nanosatellites
by Michael M. Gorlick, The Aerospace Corporation

Performance Thresholds for Application of MEMS Inertial Sensors in Space
by Geoffrey N. Smit, The Aerospace Corporation

Advanced Modeling of Micromirror Devices
by M. Adrian Michalicek, Darren E. Sene, and Victor M. Bright, Air Force Institute of Technology

Monitoring Composite Material Pressure Vessels with a Fiber-Optic/Microelectronic Sensor System
by Charles Klimcak and B. Jaduszliwer, The Aerospace Corporation

Micromechanical switches on GaAs for Microwave Applications
by John Randall, C. Goldsmith, D. Denniston, and T-H Lin, Texas Instruments

High Density Memory Based on Quantum Device Technology
by Paul van der Wagt, Gary Frazier, and Hao Tang, Texas Instruments

InGaAlAsPN: A Materials System for Silicon Based Optoelectronics and Heterostructure Device Technologies
by T. P. E. Broekaert, S. Tang, R.M. Wallace, E. A. Beam III, W. M. Duncan, Y.-C. Kao, and H.-Y. Liu, Texas Instruments

AIAA spacecraft GN&C interface standards initiative: Overview

71100

A. Dorian Challoner

AIAA Committee on Standards for Guidance, Navigation and Control

The American Institute of Aeronautics and Astronautics (AIAA) has undertaken an important standards initiative in the area of spacecraft Guidance, Navigation and Control (GN&C) subsystem interfaces. The central objective of this effort is to establish standards that will promote interchangeability of major GN&C components, thus enabling substantially lower spacecraft development costs. Although initiated by developers of conventional spacecraft GN&C, it is anticipated that interface standards will also be of value in reducing the development costs of microengineered spacecraft. The standardization targets are specifically limited to interfaces only, including information (i.e., data and signal), power, mechanical, thermal, and environmental interfaces between various GN&C components and between GN&C subsystems and other subsystems. The current emphasis is on information interfaces between various hardware elements (e.g., between star trackers and flight computers). The poster presentation will briefly describe the program, including the mechanics and schedule, and will publicize the technical products as they exist at the time of the conference. In particular, the rationale for the adoption of the AS1773 fiber-optic serial data bus and the status of data interface standards at the application layer will be presented.

Poster not available at time of publication.

71102

A HIGH-AVERAGE-POWER FREE ELECTRON LASER FOR MICROFABRICATION AND SURFACE PROCESSING APPLICATIONS†

230613
10P.

H. F. Dylla, S. Benson, J. Bisognano, C. L. Bohn, L. Cardman, D. Engwall, J. Fugitt, K. Jordan, D. Kehne, Z. Li, H. Liu, L. Merminga, G. R. Neil, D. Neuffer, M. Shinn, C. Sinclair, M. Wiseman, CEBAF,
L. J. Brillson, Xerox, D. P. Henkel, Henkel Metallurgical Tech.,
H. Helvajian, Aerospace Corp., M. J. Kelley, DuPont,
Shanti Nair, University of Massachusetts

Abstract
CEBAF has developed a comprehensive conceptual design of an industrial user facility based on a kilowatt UV (160–1000 nm) and IR (2–25 micron) free electron laser (FEL) driven by a recirculating, energy-recovering 200 MeV superconducting radio-frequency (SRF) accelerator. FEL users—CEBAF's partners in the Laser Processing Consortium, including AT&T, DuPont, IBM, Northrop Grumman, 3M, and Xerox—are developing applications such as metal, ceramic and electronic material microfabrication, and polymer and metal surface processing, with the overall effort leading to later scale-up to industrial systems at 50–100 kW. Representative applications are described. The proposed high-average-power FEL overcomes limitations of conventional laser sources in available power, cost-effectiveness, tunability and pulse structure.

Introduction

The Laser Processing Consortium—a collaboration involving nine U.S. corporations and companies, seven research universities, and a Department of Energy accelerator laboratory (CEBAF)—is planning to take the first of two steps in developing a profitable, production-scale capability to use laser light for high-volume manufacturing processes (1). We propose to develop a cost-effective, high-average-power free electron laser (FEL) that would deliver light at wavelengths fully adjustable across the infrared (IR), ultraviolet (UV), and deep ultraviolet (DUV) portions of the spectrum. Such an FEL would address multibillion-dollar markets by fundamentally improving industry's abilities to

- modify polymer film, fiber, and composite surfaces,
- process metal surfaces and electronic materials,
- micromachine or surface-finish metals, ceramics, semiconductors, and polymers, and
- evaluate materials nondestructively and monitor manufacturing processes.

Laser light offers distinct advantages for the work of manufacturing. Laser light's coherence and high brightness allow delivery of high power densities onto material substrates. Its monochromaticity allows precise matching to typical narrow-band absorption. In short pulses, it can modify surfaces without the counterproductive side effect of bulk heating. Moreover, environmentally benign laser processing can replace wet-chemistry processing methods that produce enormous amounts of dilute aqueous waste. For all of these reasons, industry has become interested in lasers, and in fact is using them widely for cutting and welding. Laser Processing Consortium industrial members now have substantial commercial interest in seeing lasers further developed for a wide range of production applications.

†Supported by U.S. DOE Contract #DE-AC05-84ER40150 and the Commonwealth of Virginia.

But conventional lasers suffer limitations in cost, power, and choice of wavelength. Therefore industry needs a fundamental improvement in laser technology: a laser that can affordably deliver precisely controlled light at average power levels that are orders of magnitude higher than now available, and at wavelengths fully selectable across the IR, the UV, and especially the DUV. An FEL "driven" by electrons from a superconducting radio-frequency (SRF) electron accelerator can meet these cost and performance requirements (2, 3).

A production-scale manufacturing FEL's driver accelerator is its key subsystem and technological challenge. Even though FELs have been in development for nearly two decades, mainly using nonsuperconducting acceleration technologies, FELs currently operating in the U.S. reach only about 10 W of average power. But superconducting accelerators, in contrast to pulsed, room-temperature, copper-cavity-based accelerators, permit continuous-wave (CW) operation, which automatically means high average power in the electron beam. The virtual absence of ohmic losses in the accelerating structures means vastly superior energy efficiency. However, even though SRF has matured as a technology in recent years, no high-average-power SRF-based FELs exist. This consortium believes cost-effective FEL for industrial applications can be built using a comparatively small, CEBAF-type SRF accelerator.

Therefore we propose a two-phase development program. In Phase 1, capitalizing on existing infrastructure and expertise within the consortium, we will design, build, and commission at CEBAF a demonstration-and-development user facility centered on a kilowatt-scale FEL, which we will operate across the IR, UV, and DUV for

- development, analysis, and refinement of commercial applications,
- investigation of opportunities for widening the commercial potential of high-average-power FELs, and
- demonstration of the key subsystem technologies to allow confident scale-up of SRF-based FELs to higher-power, lower-cost operation.

The Phase 1 FEL is hereafter called the *Demo FEL*. In Phase 2, we will scale up from the Demo FEL and build a 50–100 kW version, a prototype for cost-effective production use at industrial sites.

Surface Processing and Microfabrication with Light

Most prospective light-based manufacturing will involve modifying materials' surfaces and will take place in the UV, although applications in the IR will be significant, such as surface processing at IR wavelengths that match strong absorptions in solid materials and modest resolution (2–10 μm) micromachining.

In principle, UV light offers an array of opportunities for substantially increasing both the applications and the commercial value of surface modification. But existing UV sources—including lamps that provide incoherent light, harmonically converted Nd-doped solid state lasers, and conventional excimer lasers—suffer severe limitations for such work. None offers high enough average power or low enough cost per delivered kilojoule of light for general production purposes, and the light output from the coherent sources is not available at—or tunable to—wavelengths overlapping specific absorption bands of interest.

Nonetheless, a few conventional lasers—UV excimer lasers in particular—have seemed to offer promise. A laser is a nonintrusive, *in situ* processing tool which can perform multiple tasks simultaneously, including serving as a process monitor. It is easily amenable to automation and is commonly used in specific-area processing. Furthermore, lasers can be used as diagnostics for monitoring surface character, quantifying the integrity of an embedded interface, and

spectroscopically identifying adsorbates and ablated material. The key point is that a laser can be used to both deposit and remove material while serving as a diagnostic probe, all *in situ*. This multiuse, multirole, *in situ* capability is not offered with other advanced materials-processing techniques like molecular beam epitaxy (MBE), chemical vapor deposition (CVD), fast ion bombardment (FIB), and magnetron/plasma sputtering. And a further benefit is the elimination of the environmental costs of processing with wet chemistry.

Since excimers became available about ten years ago, many researchers have taken advantage of the intense absorption found at their wavelengths to explore surface processing with light. Absorption coefficients in the 10^4–10^5 cm^{-1} range result in essentially all the energy being deposited in the outermost few tenths micron or less of a material. Consequently, novel materials or material states can be created on the surface while leaving the desirable properties of the bulk intact. Depending on the details of wavelength, irradiance, and fluence, UV light can transform chemistry, morphology, and topography. Further, with a proper choice of conditions, laser-induced ablation can remove material, allowing micromachining to the dimensional scale of the light wavelength itself.

To date, a few commercially viable applications have been developed for conventional-laser material processing, mostly limited until very recently to cutting and welding tasks. Newly increased reliability in pulsed lasers, especially pulsed excimers, has resulted in other applications, including lithography, pulsed-laser deposition/etching, and micrometer-scale machining/surface texturing. "Machined" or "grown" materials include metals, semiconductors, superconductors, ceramics, insulators, and biocompatible materials. In addition, a number of multicomponent, device-quality, "tailored" thin films have been grown by pulsed-laser deposition processing.

Certain specialized, high-value-added conventional-laser applications have been notably cost-effective: those requiring limited irradiation doses at one of four fixed excimer laser wavelengths and those requiring relatively few intense pulses of high-fluence ablation. The former approach has limited application to large-area processing given the low duty cycle of current laser systems. The latter approach unnecessarily affects a large volume of the material workpiece by removing, cutting, or altering material. This damages the surrounding area via thermal and plasma effects, generating debris and thereby wasting incident laser energy through absorption in the above-surface plasma. To mitigate these effects, the laser fluence is commonly reduced, which is tantamount to significantly increasing the processing time. Given the low (subkilohertz) repetition rate of current high-power lasers, the additional processing time makes the application too costly. High-fluence laser processing does have applications in advanced-materials development, but more applications become possible if each processing step can be made to affect less material—a measured approach that would add precision to processing, but is not economically viable with present laser technology.

Figure 1 illustrates the fluence and processing-rate limitations of conventional lasers for typical surface transformation and surface melt treatments of metals. The figure compares the approximate minimum fluence requirements for CO_2, Nd:YAG, and excimer lasers with those of the Demo FEL (4). Because the high absorptivity and short (picosecond) pulse length of the Demo FEL light greatly reduce its fluence requirements, the estimated processing rate of the Demo FEL is several orders of magnitude faster than those of the conventional lasers.

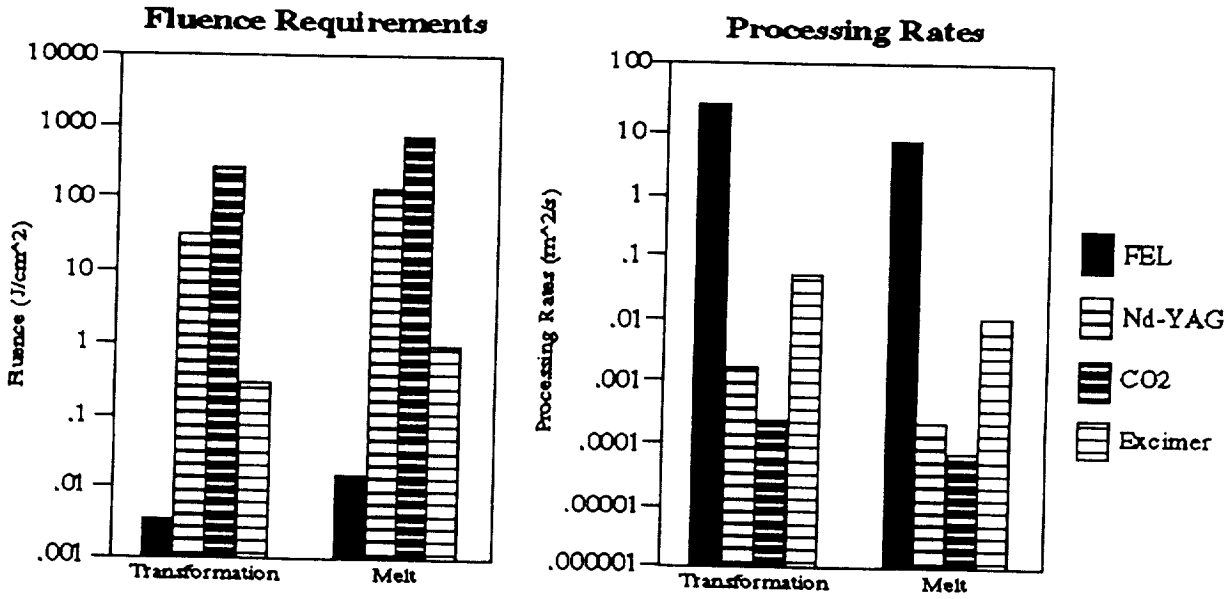


Figure 1. Calculated laser fluence and processing rates for metal surface processing by conventional lasers (Nd:YAG at 1 kW, CO_2 at 10 kW, excimer at 200 W) and the Demo FEL (at 1 kW).

So despite substantial industrial R&D investment and some limited successes, conventional laser technology is not likely to improve U.S. manufacturing capability in any fundamental way. Commercially available excimer lasers present perhaps the best illustration: they remain limited to tenths of kilowatts, tens of cents per kilojoule, and a few isolated specific wavelengths, while large-scale exploitation of UV surface processing will require sources of at least a few tens of kilowatts, light that costs under a cent per kilojoule, and full wavelength tunability.

Therefore the Laser Processing Consortium's industrial members have become increasingly interested in precompetitive FEL R&D to prepare for manufacturing in the twenty-first century.

Phase 1: Technology Demonstration and Development

Applications Development in the Demo FEL User Facility

The proposed Phase 1 project is to build and operate on the CEBAF accelerator site a user facility centered on the Demo FEL—a kilowatt-scale, SRF-driven FEL producing light in the UV (160–1000 nm) and the IR (2500–25,000 nm). In user laboratories in the facility, we intend to exploit the Demo FEL's capabilities to benchmark and extend the industrial utility of FEL manufacturing applications and to investigate the technology required for Phase 2 development of a 50–100 kW production-scale device. The key goal for ultimate commercialization in Phase 2 is to achieve light at a total cost (capital plus operational) of less than a cent per kilojoule.

Each user laboratory will have a particular technology focus and will be equipped by the user industries with exposure stations for large-area samples, vacuum systems for analytic equipment, and optical diagnostics for light-source characterization. Computer control interfaces will be provided in each laboratory to allow the users to control the beam parameters and optimize

performance for each process. Provision will be made not only for proprietary research, but in a few cases for actual commercial use of the facility. Industry members have committed equipment for four of the user laboratories for: (1) large-area surface processing of polymers, (2) micromachining and microfabrication, (3) laser-photochemical processing, and (4) laser-material surface diagnostics.

Industrial applications identified by consortium members include:

- Polymer microtexturing. Polymer film or fiber surfaces can be microughened with exposure to 248 nm UV laser light, a processing treatment that can give the product new friction, filtration, wetting, or visual-appearance characteristics (5). Commercially important applications include better adhesion for forming multicomponent film products or composite structures, more effective fibers for use in filters, and improved "feel" of synthetic fiber fabrics. For commercial viability, the treatment would require a fluence of 200 to 600 mJ/cm^2 at about a penny per kilojoule and at a minimum power of 25 kW.

- Microfabrication. Laser-micromachining can fashion micron-scale structures with nanometer-scale control. Through subthreshold ablation (6) with light from a low-fluence (0.1–100 mJ/cm^2), high-repetition-rate (> Mhz) tunable UV laser (200–300 nm), a host of novel materials processing applications would become possible, such as micro-optics, ultrahigh-density storage media (>5 $Gbits/in^2$), adhesiveless microfasteners (i.e., micro-Velcro), and precision abrasive surfaces.

- Surface conductivity. Delivering from 10 to 40 J/cm^2 at a fluence near the ablation threshold drives polymer decomposition toward graphite, imparting electrical conductivity (7). Stable, durable "wires" as narrow as 30 nm could be directly written onto polyimide substrates for microelectronics applications. The treatment requires 10 kW; a tolerable production cost is pennies per kilojoule.

- Laser annealing. A slow cool resulting from a low-velocity laser scan can alter a metal surface grain structure in a manner similar to bulk furnace annealing, resulting in improved resistance to fatigue-crack nucleation (8). The Demo FEL could be wavelength-tuned for full absorption at possible annealing rates of 10 m^2/sec.

- Large-area diamond coating. Thin-film diamond has lucrative applications in microelectronics packaging (e.g., insulating layers in flip-chip technology), tribology (e.g., bearings and contact surfaces), and flat-panel displays (e.g., field-emission sources). Currently, coatings of amorphous diamond on materials are accomplished by using the 10 nsec pulses from solid state lasers (e.g., Nd:YAG) (9). The Demo FEL's repetition rate is expected to be 10^6 times faster; this means an ability to coat roughly 100 times the area (1 m by 1 m) in one second.

The Demo FEL

High-average-power, wavelength-tunable laser light from the Demo FEL briefly described in this section will allow users to demonstrate and develop FEL applications as discussed above. Figure 2 contrasts Demo FEL performance with that of conventional lasers.

The Demo FEL will provide average powers in the kilowatt range. In addition, the output of the Demo FEL will have the following characteristics:

- **Tunability**: Existing sources are not tunable and/or available at wavelengths overlapping specific absorption bands of interest. The Demo FEL will provide light which is tunable across the UV (160–1000 nm) and the IR (2500–25,000 nm). As a bonus, the entire visible spectrum will be accessible (350–750 nm). This light will have all of the characteristics of high-quality laser emission: narrow bandwidth (typically <0.1%), spatial coherence (1–2 times the diffraction limit), and linear polarization.

- **Temporal Structure**: Existing sources are either CW at low intensity or have pulses that are much too long to enable efficient surface processing. By contrast, the Demo FEL will generate very short pulses (1 psec) that are ideally suited to rapid thermal annealing or ablation of near-surface regions. This pulse length matches the time scales of surface molecular rearrangements and vibrations. In ablation applications using high-power excimer lasers, pulse lengths exceed 10 nsec, and these pulses are long enough to interact with gas-phase ejecta with an associated loss of surface-interaction efficiency. The FEL circumvents this problem.

- **Efficiency/Cost**: The cost per unit energy of light delivered from conventional UV sources is too high (of order $0.10/kJ) for profitable industrial applications. For demonstration and development purposes, the Demo FEL will provide light at about $1/kJ. Prospective goals for the Phase 2 UV FEL are operation at 10% wall-plug efficiency and a cost of delivered light below $0.01/kJ.

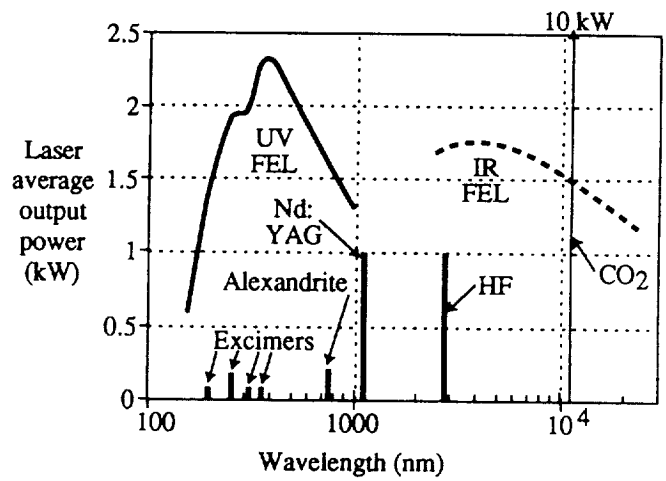


Figure 2. Demo FEL power vs. wavelength. Conventional laser outputs appear as narrow lines at fixed wavelengths.

Figure 3 is the Demo FEL layout. To provide the needed light, the FEL will extract energy from electrons accelerated in two passes through a recirculating, energy-recovering SRF driver linac (linear accelerator) (10). The linac will consist of three cryomodules, cryostats closely similar to those used at CEBAF, each containing four pairs of linked SRF accelerating cavities. For superconducting operation at 2.0 K, the linac will tap the excess capacity of CEBAF's nearby main refrigerator, the Central Helium Liquefier (CHL). The electron beam will originate in an injector with three main elements: an Nd:YLF-laser-driven 500 kV DC photoemission electron source, a copper cavity to bunch the beam, and a two-SRF-cavity quarter-cryomodule. Development work directly useful for the injector—a key technological challenge for Phase 1—is already under way at CEBAF, thanks in large part to support from the Commonwealth of Virginia.

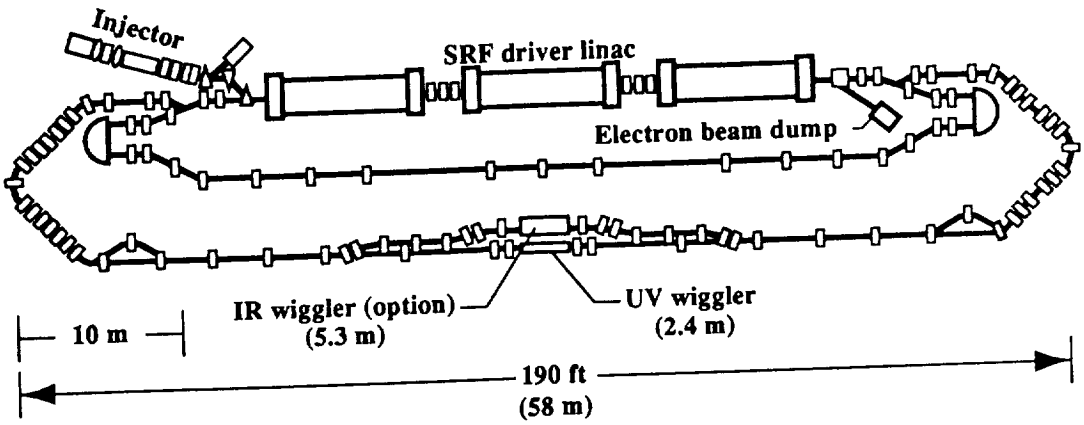


Figure 3. Demo FEL layout.

Demo FEL operation can be summarized as follows. An electron beam at 10 MeV energy from the injector attains 105 MeV in the first of two acceleration passes through the linac. After recirculating back to the injection point clockwise through the low-energy recirculator line (at the center of the machine, between the linac and the wigglers), the beam attains 200 MeV by the end of its second acceleration pass. Then it is directed to an FEL wiggler, where it yields about 0.5% of its power in the form of laser light. In the sinusoidal magnetostatic field of either the UV or the IR wiggler, relativistic accelerated electrons undulate transversely. The resulting light output is initially spontaneous emission, but the light bounces back and forth in an optical cavity, extracting energy from the beam until it is amplified to saturation. Light is outcoupled from the optical cavity and delivered for user applications. The electron beam then decelerates in two energy-recovery passes back through the linac. Energy recovery extracts the substantial energy the electron beam retains after it has transited the wiggler—in effect recycling the energy by converting it back to RF power at the linac cavities' resonant frequency. Finally, about 10 MeV of remaining energy is absorbed in a cooled, shielded copper beam dump.

SRF technology, energy recovery, and electron beam recirculation have already been combined and demonstrated at lower average current at CEBAF. With these key, integrally linked design features we are aiming at overall cost-effectiveness in the Demo FEL, with specific emphasis on developing cost-reducing and reliability-enhancing measures for the Phase 2 production-scale device. An SRF linac is intrinsically efficient, requiring substantially less RF power input than does a room-temperature system equipped with energy recovery. And if the SRF linac itself uses energy recovery, an additional RF efficiency advantage of more than an order of magnitude can be gained. Moreover, energy recovery eliminates the need for large-scale radiation-management measures. A 200 MeV beam at 5 mA without energy recovery would require a megawatt-scale beam dump. But a beam decelerated to approximately its original injection energy requires far less: for the Demo FEL, a 50 kW beam dump. Recirculation lowers capital cost by minimizing the number of superconducting components, lowers operating cost by reducing the cryogenic load, and substantially reduces system footprint size.

The Demo FEL user facility is proposed to be built at CEBAF, the DOE-owned site of a new 4 GeV accelerator, a user facility for nuclear physics research. Beyond ensuring technical, cost, and schedule success, important project management goals include exploiting advantageous synergisms. These synergisms include CEBAF's SRF and electron-source technology expertise and infrastructure, the excess liquid helium capacity of the main refrigerator that serves the 4 GeV accelerator, and CEBAF's already existing environmental and radiation-monitoring permits.

Phase 2: Technology Scale-Up

We expect that successful demonstration and development efforts in Phase 1 will focus and intensify needs already identified by industry for systems operating at much higher power levels, and will also provide hard data and practical experience for meeting these needs. In Phase 2 we will build a prototype 50–100 kW industrial FEL suitable for cost-effective production use at individual industrial sites and at regional processing centers serving multiple manufacturers.

Research and testing will ensure that the device is industrially useful. The prototype must be cost-effective, robust, reliable, and easy to operate. For an operational commercial system, capital cost and operating cost are key considerations which lead to an overall figure of merit, the cost per delivered kilojoule. Present commercially available excimer laser systems cost between 10 and 20 cents per delivered kilojoule when both operating and capital costs are taken into account. A key goal for the Phase 2 program is to produce UV light at around 0.2 cents per kilojoule.

Although detailed analyses of Phase 2 can only be prepared based on actual Phase 1 data and experience, preliminary analyses have been carried out. It is clear that developing the capability for higher powers will require attention to:

- Injector performance. The injector is a challenge because of the desire for high average current with long cathode life, and because of the high brightness (emittance) specifications for the electron beam. Average current more than an order of magnitude higher than that for Phase 1 will place increased demands on the electron source, so a critical goal during Phase 1 will be to develop a high-intensity, high-quality source.

- Optical cavity performance. The optical cavity is a challenge because of the high intracavity intensity exacerbated by relatively high mirror-coating absorption at short wavelengths and tight limits on mirror deformation. A modest R&D program for high-reflectivity coatings will be conducted in conjunction with Phase 1.

- Operating frequency. This choice requires tradeoffs involving not only SRF cavity design, but also transport characteristics of the lattice design, RF source efficiency, and cryogenic system performance.

- Lattice design. The lattice design effort includes optimizing the number of recirculation passes.

Program Status (November 1995)

The Laser Processing Consortium, its objectives, and its resources have grown and evolved steadily since 1991, the year CEBAF organized an advisory group of high-technology corporations to analyze prospects for market-oriented applications of CEBAF technology

A NASA-sponsored peer review panel convened in March 1994 and a similar DOE-sponsored panel convened in May 1995 confirmed industry's need for FELs and strongly endorsed this consortium's approach for developing them.

The Laser Processing Consortium's proposal to develop free-electron lasers for industry already has obtained state and private-sector support. . The Commonwealth of Virginia, where CEBAF is located, is providing substantial support for the FEL enterprise, including matching funds already in use for electron source development and funds for the FEL User Facility building. Industry

and university members of the consortium have made commitments for most of the required end station equipment in the User Facility. The first matching Federal funds to begin constructing the FEL hardware are expected in FY96.

References

1. Laser Processing Consortium Proposal: Free-Electron Lasers for Industry, Vol. 1 and 2 (May 1995).

2. H. F. Dylla et al., Proc. 1995 Particle Accelerator Conf., Dallas (IEEE, in press).

3. G. R. Neil and S. Benson, Proc. 1995 Particle Accelerator Conf., Dallas (IEEE, in press).

4. S. Nair, "Applications of Free Electron Lasers to Metals and Ceramics Processing," CEBAF/DOE TN No. 94-057, (Nov. 1994).

5. W. Kesting, D. Knittel, T. Bahners, and E. Schollmeyer, Appl. Surf. Sci., 54, (1992) 330.

6. H. Helvajian, L. Wiedeman and H.-S. Kim, Adv. Mat. for Optics and Elect: 2, (1993) 31.

7. J. L. Hohman, K. B. Keating and M. J. Kelley, Mat. Res. Soc. Symp. Proc., 354, (1995) in press.

8. B. Brenner et al., in Laser Treatment of Materials, B. L. Mordike et al., (DGM Informationsgesellschatt, Verlag, Germany) (1992), 199.

9. C. B. Collins, F. Davanloc, H. Park, SPIE Proc. 2403, (1995).

10. D. Neuffer et al., Proc. 1995 Particle Accelerator Conference, Dallas (IEEE, in press).

MEMS IN SPACE SYSTEMS

7/104

J. C. Lyke, M. A. Michalicek, and B. K. Singaraju
Space Electronics Division
Air Force Phillips Laboratory
Kirtland AFB, NM 87117-5776

230611

10 P

Abstract

Microelectromechanical Systems (MEMS) provide an emerging technology area that has the potential for revolutionizing the way space systems are designed, assembled, and tested. The high launch costs of current space systems are a major determining factor in the amount of functionality that can be integrated in a typical space system. MEMS devices have the ability to increase the functionality of selected satellite subsystems while simultaneously decreasing spacecraft weight.

The Air Force Phillips Laboratory (PL) is supporting the development of a variety of MEMS-related technologies as one of several methods to reduce the weight of space systems and increase their performance. MEMS research is a natural extension of PL research objectives in microelectronics and advanced packaging. Examples of applications that are under research include on-chip microcoolers, micro-gyroscopes, vibration sensors, and three-dimensional packaging technologies to integrate electronics with MEMS devices. The first on-orbit space flight demonstration of these and other technologies is scheduled for next year.

Introduction

PL research in the Space Electronics Division (PL/VTE) is dedicated to solving pacing problems in existing and emerging USAF space systems. Chief among these problems is the need to provide high density electronics solutions that are within the cost, reliability, and performance requirements of these systems. While much of the PL/VTE research investment is devoted to the development of radiation-hardened processes and components that are normally taken for granted as available by terrestrial systems (such as microprocessors and memory), significant development programs in advanced two- and three-dimensional microelectronics packaging have existed for nearly a decade. These programs initially were concerned with wafer scale integration of complex digital functions to accelerate the miniaturization benefits of radiation-hardened microcircuits, but later expanded in terms of the types of electronics addressed (e.g., analog, microwave, power) and in the approaches for achieving this reduction. Great benefits were later achieved through the exploration of three-dimensional packaging and the significant parallel investments by the Advanced Research Projects Agency (ARPA) in packaging technology in general. The greater density benefits of three-dimensional packaging led to more and more systems functions being addressed within very dense multi-chip modules (MCMs). Within the last five years, the advent of MEMS has promoted the possible integration of MEMS devices as well, leading to provoking systems engineering possibilities, such as enclosing inertial reference units (IRUs) within MCMs. One program, the monolithic interceptor processor (MiP), explored concepts for reducing the complete high-performance electronics system for an interceptor in smaller than a coffee cup (<25 cubic inches), complete with cryogenic imaging sensors and IRUs, and limited prototyping activities based around "many-layered" (>12) 3-D MCMs have been conducted. More recent research has evolved that may establish standards for heterogeneous 3-D packaging in electronic systems.

While PL's exploration of packaging issues had raised the possibility of introducing MEMS devices into systems to further accelerate the size, weight, and power reductions, a number of USAF Small Business Innovative Research (SBIR) projects and PL-sponsored Air Force Institute of Technology (AFIT) research efforts focus MEMS development activities in a complementary sense to provide solutions to certain packaging problems (thermal management) and miniature sensors (vibration sensors). Many innovative concepts that are discussed in this paper are products of these efforts.

Over the last few years, PL/VTE has worked more closely with the Aerospace Corporation, ARPA, and NASA to better define and establish solutions to barriers for the inclusion of MEMS devices in space systems. Aerospace has introduced the notion of Application-Specific Integrated Microinstruments (ASIM), a concept to which PL has contributed several packaging concepts, including the "Constant Floor Plan" MCM, which can greatly accelerate the schedule and reduce cost for early prototypes. Of the great variety of ARPA MEMS research initiatives, three efforts in particular have been identified in which PL provides technical management support. Among these are programs that involve micro-inertial reference systems, advanced micro-instruments, and failure analysis of MEMS devices. NASA's New Millennium Program (NMP) is pursuing a paradigm for satellites beyond 2000 that are not only much smaller than the present genre of civil and military satellites, but built in greater quantity, much more quickly, and at a much lower cost. The possibility of achieving these goals with MEMS was deemed to be so significant a possibility as to warrant NMP establishing an Integrated Product Development Team (IPDT) in MEMS and Instruments, one of only five for the entire initiative. PL/VTE is involved as a member on that IPDT, as well as the microelectronics IPDT, to support the establishment of technology roadmaps to achieve ambitious NMP objectives.

In research to date, one of the most significant barriers encountered to the spaceflight of MEMS devices is a perception (sometimes correct) of technology immaturity. Intrinsically, lack of reliability data on MEMS devices are indicated, particularly in a space environment. The lack of real data on MEMS performance in space environments has been so acute as to promote the development of dedicated experiment and sub-experiment payloads for space missions as an in-house activity in PL/VTE. At least two space experiment payloads with dedicated MEMS devices and objectives are under development at the time of this writing, with prospects of three other experimental missions within the next five years. In each case, a search has been underway to identify the most current relevant and useful examples of MEMS devices that are suitable for spaceflight. In fact, the first two experiments do not contain products of PL/VTE MEMS research programs, since some commercial devices, such as accelerometers, have emerged that are slated for widespread use in the automotive industry. The search continues, with plans to field the most promising candidate of PL/VTE, ARPA, NASA, and commercial MEMS research.

MEMS Programs

This section describes several of the aforementioned MEMS research efforts which are directly relevant to space and missile applications. Many MEMS efforts have been established to find better solutions to advanced packaging of electronics, particularly in the efficient thermal management of 3-D densely packaged MCMs. Active thermal management solutions, ranging from microcoolers with flow-through liquids and pumps to Sterling engine refrigerators are being explored. Another class of MEMS efforts deals with packaging issues associated with MEMS devices. One MEMS research effort, sponsored at AFIT, deals with the compatibility of canonical MEMS structures with representative types of MCM packaging approaches. In a second effort, a 3-D packaging approach is being explored for inclusion of certain MEMS components. Another class of MEMS device research deals with MEMS-based relays which is quite important to space systems. Finally, other MEMS efforts pertain to the establishment of an instrumentation (sensing) function, such as vibration, linear, or rate sensors, in particular for tactical grade inertial measurement.

Thermal Management of Advanced Packaging

It is intuitively clear that as electronics functional densities increase, so too does power density and power dissipation per unit volume. Given that a conventional system achieves a real packaging efficiency of less than one percent, it is not surprising that a thermal management issue might exist when the same electronics are packaged at > 25% efficiency, as in the case of some more recent PL/VTE packaging research. Thermal management has been found a natural role of MEMS technology. The diversity of MEMS approaches to achieve thermal control is quite impressive.

Micro-Encapsulated Liquid Cooling. Two AFIT research efforts[1,2] were sponsored to investigate the benefits of silicon micromachining to the thermal management problem of planar MCMs. These efforts were based on an approach of a plenum system formed directly into silicon wafers through micromachining. A number of configurations and microtextures were examined for thermal performance under various gas and liquid flow-through conditions. The circulating liquid, supplied from a peristaltic pump, in effect provided the effect of an active heat spreader.

The concept of directly micromachining silicon wafers to form thermal management structures was also applied in the second phase of a PL-managed USAF SBIR research program[3] to form an efficient indirect cooling system for distributing liquid nitrogen to a nine-chip prospective MCM based on cryogenic digital semiconductors and multilevel high temperature superconducting interconnections (Intermagnetics General Corporation, Albany, NY). In this case, deep microchannels are formed through a special chemical process to positions of a silicon wafer corresponding to each of nine chip locations. Liquid nitrogen flow-through with this configuration, depicted in simplified form in Figure 1, achieves a more effective cooling system than direct immersion.

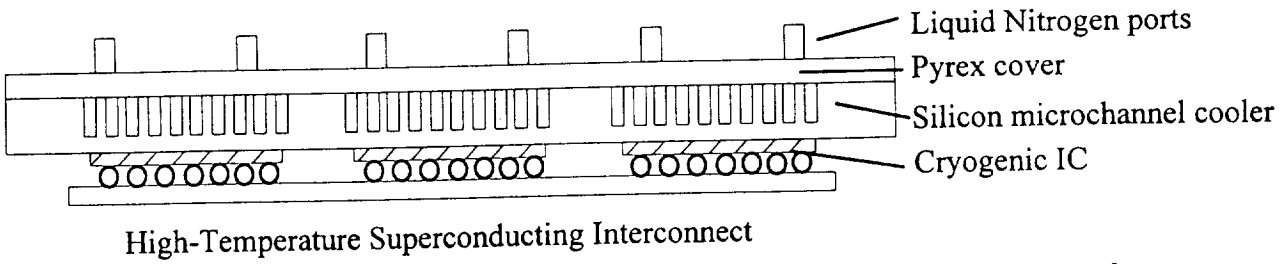


Figure 1. Simplified depiction of microchannel cooler for cryogenic thermal control.

Miniature refrigerators for ICs. Another more direct approach for cooling heat-generating components involves the application of micromachining and micro-engineering technology to create miniature refrigerators, about the size of an IC (one cm^2). An on-going phase II SBIR is investigating the construction of components for such refrigerators based on a Stirling engine, with the goal of cooling heat generating components which normally operate at ambient temperatures (Sunpower, Athens, OH)[4]. A prospective application of such a Sterling cooler involves its strategic placement within locations of MCMs that contain several high-power integrated circuits, in a manner transparent to the end user.

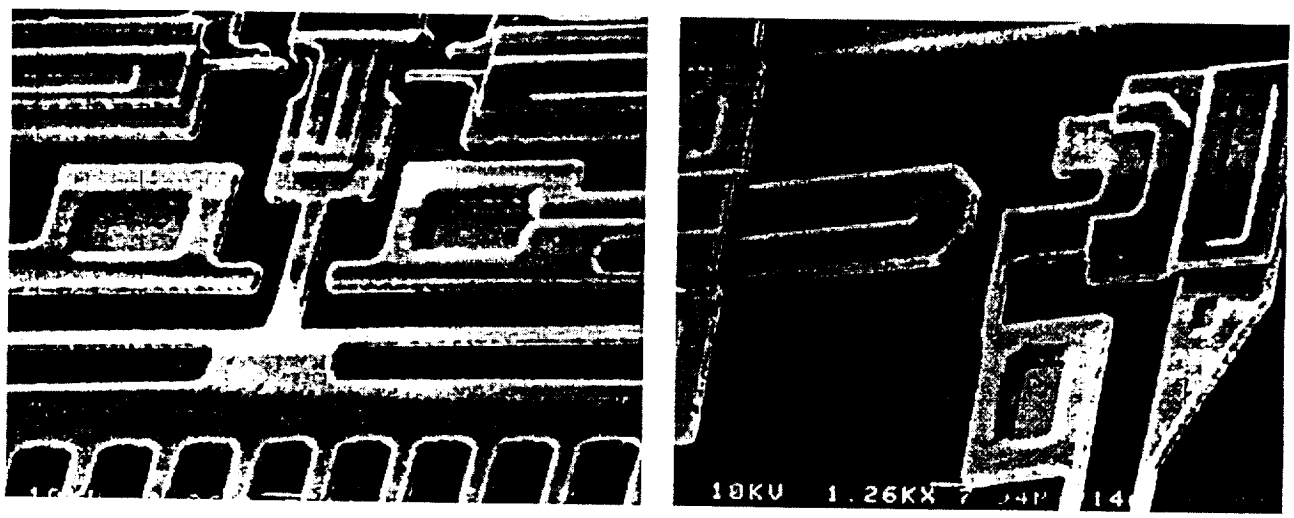
Compatibility of MEMS in Packaging Approaches

Without adequate attention to packaging, MEMS devices lose their full potential to reduce the size, weight, and power of systems. Packaging of MEMS with other components in MCMs logically exploits the benefits of both approaches. A three-dimensional approach to combine certain MEMS devices that employ a simple flip-chip approach is under development based on a "telescopic" stacking of monolithic devices (Optical ETC, Huntsville, AL) to realize a multi-chip assembly. This investigation is supported as a SBIR program funded by the Technology Reinvestment Program[5]. A PL-sponsored AFIT PhD research effort is in progress to investigate the compatibility of various MEMS

devices with thin-film MCM processes. At issue is the sequence of fabrication to realize practical MCMs that contain MEMS devices. Handling MEMS devices after release chemistry steps is often problematic, but performing release chemistry on an ensemble of non-MEMS chips may also be difficult. The approach explored in this effort will create a sort of MEMS-based assembly test chip (ATC), similar to MCM ATCs developed by Sandia National Laboratories for packaging and assembly process condition monitoring.[6] By introducing this ATC into a packaging system, it will be possible to explore the impacts of the fabrication sequence on the MEMS devices as well as examine a typical release process on the MCM or assembly.

Micro-relays

Microrelays have significant potential to improve reliability and performance in space systems. Electro-mechanical relays are generally undesirable in conventional form for space systems due to their bulkiness and potential for failure. The simpler construction and use of tightly controlled batch fabrication processes may result in relays that have better reliability and that may be aggregated to facilitate graceful degradation. Other fault tolerant concepts possible for such devices, by their compact nature, include their use in creating densely packaged, switchable buses for complex spaceborne architectures. Solid state switches are not always adequate in space environments due to the tenet issues of the space radiation environment on semiconductors. Total ionizing dose effects degrade the salient characteristics of many bipolar and FET-based switching device processes, even for digital applications.



(a) Micro-Relay (b) Micro-Latch

Figure 2. Micro-relay and micro-latch designed, fabricated, and tested at AFIT during 1994. (photographs provided courtesy of Dr. Victor Bright, AFIT Department of Engineering).

The inability to service space systems make attractive the possibility of incorporating on-board spare assemblies that are positively and physically "dis-connectable" from the system. Even the best conventional solid state switching devices cannot achieve this performance for the most sensitive instrumentation systems, due to small but finite leakage currents, the lack of signal excursion capability and dynamic resistance non-idealities. These approaches are also not possible with bulky conventional relays, but could be easily achieved with micro-relays. For power reduction, micro-latches (relays with memory) are also desirable. Various AFIT research projects have produced several designs of micro-relays and micro-latches for advanced MEMS switching techniques (Figure 2)[7].

MEMS-based Motion and Inertial Sensing

Under ARPA sponsorship, Draper Laboratory is conducting a $6M effort to design and build micro-inertial reference systems components using flip-chip bonding technology. A monolithically packaged rate sensor is shown in **Figure 2** based on the Draper approach. It is conceivable that, when combined with MEMS-based accelerometers, a complete inertial reference unit based on three orthogonally mounted pairs of gyroscopes and accelerometers could be realized within one cubic inch.

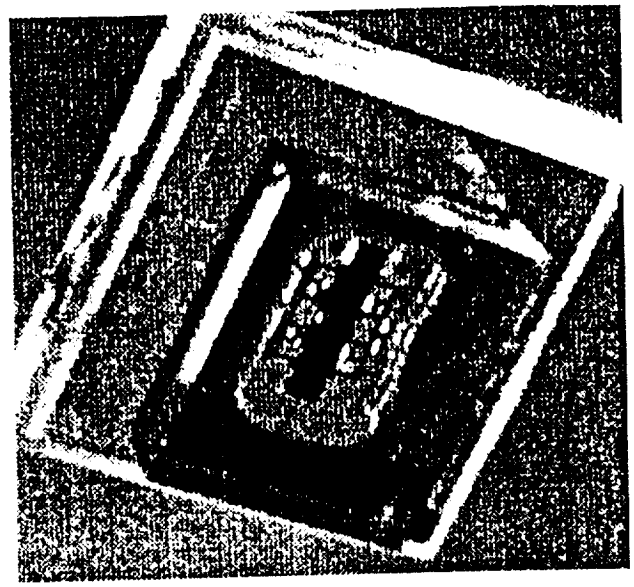


Figure 2. MEMS-based monolithic rate sensor.

An additional goal of the research is to produce low-cost MEMS inertial sensors by reducing the cost in two categories. First, the cost of packaging the devices is hoped to decrease with improvements in the stability of the gas environment and device alignment within the vacuum package. Additionally, it is desired to develop an integrated test process that would minimize cost by establishing an improved acceptance and calibration procedure during manufacturing.

Flat-Pack Gyro. In 1993, for the MiP study, a number of micromachined inertial sensors were examined, none of which possessed adequate drift rates. The search for better rate sensors, in conjunction with a published solicitation for SBIR topics in MEMS, led to a very promising concept involving a spinning gyro rate sensor. The current effort, a second phase SBIR, involves a "milli-machined" version of a gyroscope of more conventional origin, in which the drift rate is reduced by several orders of magnitude over standard designs. A millimachined chip-size gyroscope is under developed that may theoretically achieve a 0.5 degree/hour drift rate. Shown in Figure 3, the gyroscope operates like standard gyros with a design variation that uses a spinning disk supported with gas bearings as the inertial reference. The disk is magnetically actuated and its motion is capacitively coupled to sensor electrodes that detect the output axis rotation.[8]

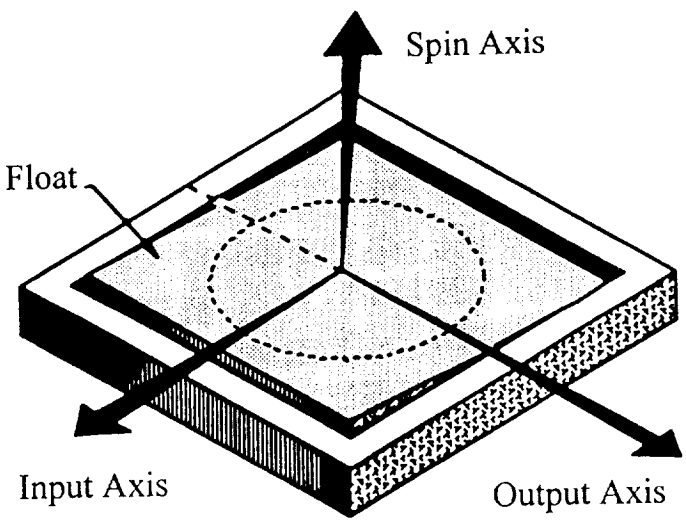
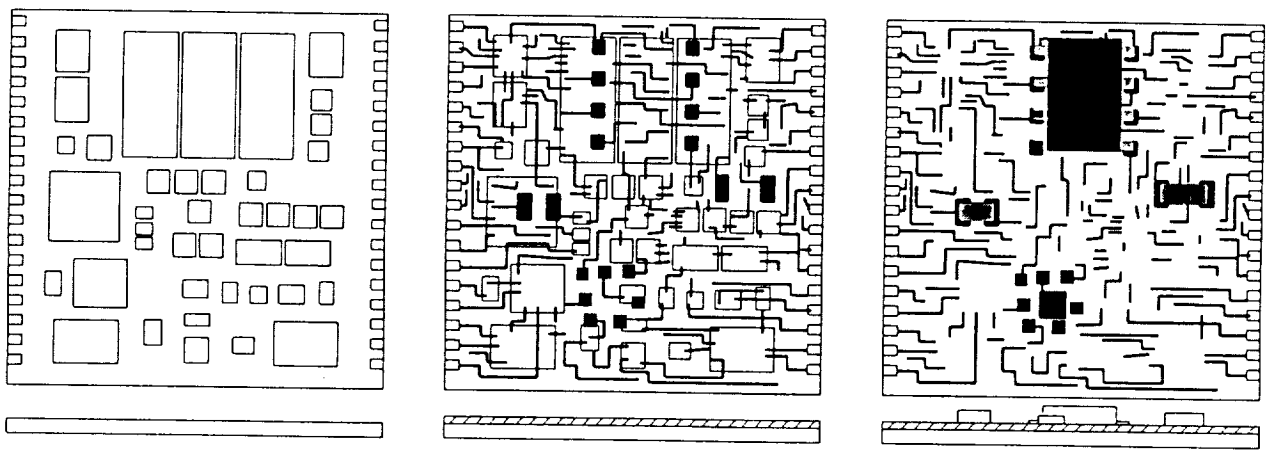


Figure 3. The Flat Pack Gyro (approx. one inch square).

Piezoresistive Vibration Sensor. A SBIR project is under way at InterScience, Inc. in Troy, New York in which an array of cantilever devices is used for vibration sensing. Each cantilever has a mass supported at the end of beams of various length throughout the array. The base of each cantilever is piezo-resistively coupled to a potential source that can detect the motion of the beams. The result is a real-time spectral vibration sensor that has a frequency resolution proportional to the difference in the lengths of adjacent beams.[9]

Micro-Instruments and Actuators

Certain MEMS devices, when combined with microelectronics and advanced packaging, can form microinstruments. One ARPA-funded project (under Broad Agency Announcement 94-40) of direct interest involves an approach to construct various scientific instruments such as gas chromatographs through the introduction of a novel pump concept (Berkeley Microinstruments, Menlo Park, CA). Joint discussions between Aerospace and PL/VTE have further identified an enabling concept for a customizable, common instrument MCM, for which the components are pre-defined, but the interconnects are user-definable. With this concept of a "constant floor plan" MCM (Figure 4), it is possible to realize rapidly prototype-able instrumentation electronics, to which sensors could be surface mounted. Clearly, this concept, combined with the aforementioned research in MCM-MEMS compatibility will significantly advance the ability to create more affordable application-specific integrated microinstruments.



(a) (b) (c)

Figure 4. Constant Floor Plan rapid prototyping concept. (a) Substrate with pre-determined commodity component arrangement. (b) Formation of patterned overlay interconnection to chip bond pads and for surface-mount connections. (c) Introduction of MEMS sensor.

Deformable Micromirrors. Deformable micromirrors have tremendous promise to provide support for applications where dynamic adjustments of phase or angle in certain types of optical trains are required. Several groups within Phillips Laboratory have examined deformable micromirrors for a variety of applications. PL/VTE is planning research to explore micromirror arrays designed to function in a variety of tasks currently reserved for larger, heavier, and more costly equipment.

Two primary styles of micromirrors are under consideration for space systems applications. The first is the phase-mostly piston style flexure-beam device that can be employed for phase manipulation of optical data. This style of device is known for its applications in adaptive optics in which large arrays are individually actuated to discretely lengthen or shorten an optical path to correct for phase aberrations. Such phase aberrations are inherent in optical systems which must operate within or through the earth's atmosphere in which random discontinuities in atmospheric pressure, temperature, and density create discrepancies in propagation along the optical wave front. The use of such devices can extend the limit in resolution of current optical systems or allow space systems to use optical information processing.

Another design of micromirror under consideration is an Axial-Rotation device. This device, or a similar design, can be used to scan a field of view and send the optical information to an image sensor that can be rigidly placed in a space system. For instance, the tracking of ICBMs is typically done using a large gimbaled mirror that is rotated around to select a field of regard. As an alternative, a chip

containing millions of micromirrors could be employed to perform this tracking and scanning function much more efficiently. A proposed plume acquisition system is shown in Figure 5. The approach is analogous to the High-Definition Television system developed by Texas Instruments. Rather than projecting an image, however, the imaging sensor is used to measure the relative infra-red intensity of an array of pixelized scenes from the field of regard[10].

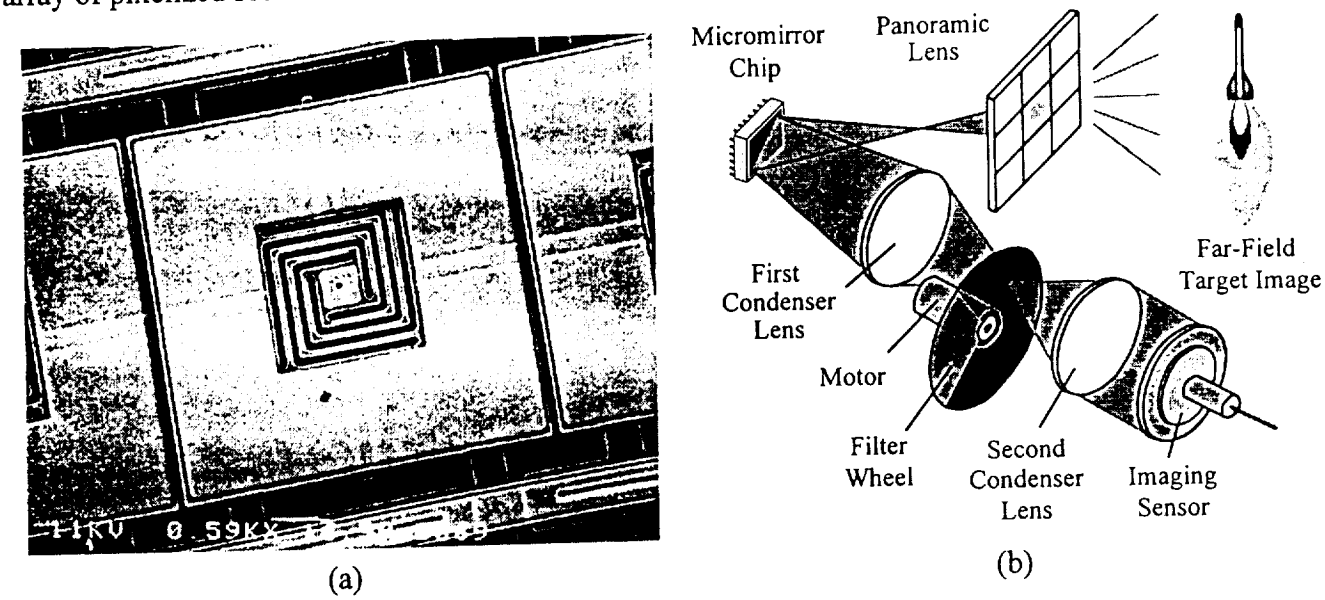


Figure 5. Axial-Rotation micromirrors and applications. (a) Micromirror configuration. (b) Proposed plume acquisition system.

The end result of such an application is a plume acquisition system that is faster, lighter, more reliable, and less costly than its current large-scale counterpart. Since the micromirrors in this system can operate in the tens to hundreds of MHz range, the image can be scanned several times and passed through a filter wheel which is added to allow for multi-spectral analysis of the image to distinguish targets.

On the Insertion of MEMS into Space Systems

Despite the tremendous advantages of MEMS devices, namely the potential to provide dramatic reductions in the size, weight, and power required for space systems, there is little acceptance presently for their application. The primary barriers are the relative immaturity of the technology, the novelty of MEMS, and lack of reliability information on these devices in a space environment. Space systems are notoriously conservative, a paradigm which initiatives such as the NASA New Millennium Program are attempting to shatter. Phillips Laboratory recently commissioned the development of a new series of satellites, called MightySat, to examine new technologies in spaceflight. PL/VTE has established a small but comprehensive payload, currently under development, including a sub-experiment for exploring MEMS technologies. In doing this, it was discovered that despite the large variety of MEMS devices in development and in the literature, that relatively few were ready for simple insertion in a non-critical experimental payload, much less a space system with a critical mission to perform. This finding attests to the present developmental state of MEMS technology. Clearly, given this finding, a relatively sparse database on MEMS reliability exists. Under normal conditions, space systems are hesitant to fly technologies that do not have a considerable maturity and experience base, making insertion of MEMS even more difficult. In an attempt to "short-circuit" some of these issues, PL/VTE is establishing experimental space insertions of MEMS technologies, based on devices developed not only from internal and joint programs, but from commercial sources as well. It is believed that in this manner, by establishing a continued series of space insertion opportunities, it will be possible to establish an early

database on reliability, which, in conjunction with other on-going initiatives in reliability, will accelerate the pace of insertion of viable MEMS-based components and modules into space systems.

Reliability

A primary concern for the use of MEMS in space systems is the reliability of the devices. To better understand fundamental reliability issues, Failure Analysis Associates has been conducting a $1.14 million research effort under ARPA funding to study the failure mechanisms of MEMS devices. Among the most significant failure modes are fractures due to actuation beyond the active range of the device and various fabrication impurities that can increase the likelihood of a particular failure. Figure 6 illustrates the effect of fabrication errors on the structure of a device. The circles highlight the areas where the support structure and the active grid of the device are poorly formed due to any one of several steps in fabrication.

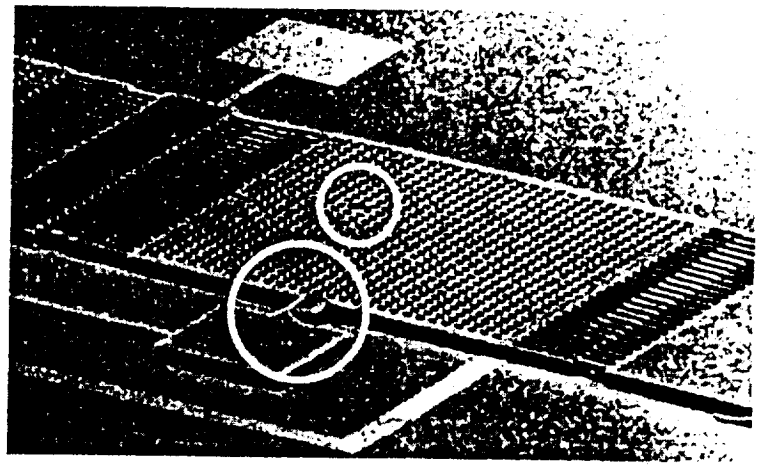


Figure 6. Example of Device Fabrication Errors (photograph from Failure Analysis Associates).

The primary purpose of the research is to reduce the likelihood of failures or the possibility of failures without the need to completely inspect every MEMS device under consideration for space systems applications. The goal of the failure mode analysis is either to show that the most significant of the feasible modes will most likely not cause failure during the mission of interest or that the effect or probability of any given failure mode can be reduced by making improvements in the fabrication process.

MAPLE: A Fledgling Space Experiment Series

The purpose of the Microsystem And Packaging for Low-power Electronics (MAPLE) experiments is to demonstrate aggressive new approaches in electronics, advanced packaging, and MEMS as enablers for next-generation spacecraft. It provides the opportunity for heritage and in-situ reliability instrumentation of controversial new technology components. It also provides the opportunity to examine new semiconductor processes and circuit designs that achieve low power operation and how they will operate in a space environment. The MAPLE concept is a "basket "experiment, based on a collection of subexperiments pertaining to advanced microelectronics, MEMS, and advanced packaging. MAPLE payloads are designed as a low-weight and relatively low power payload (usually under 8W and five pounds).

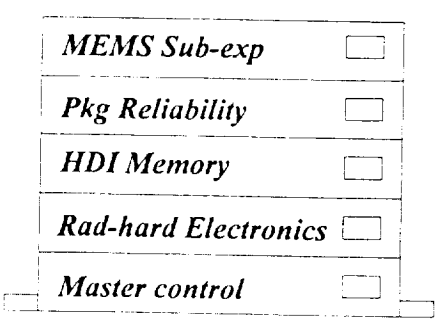
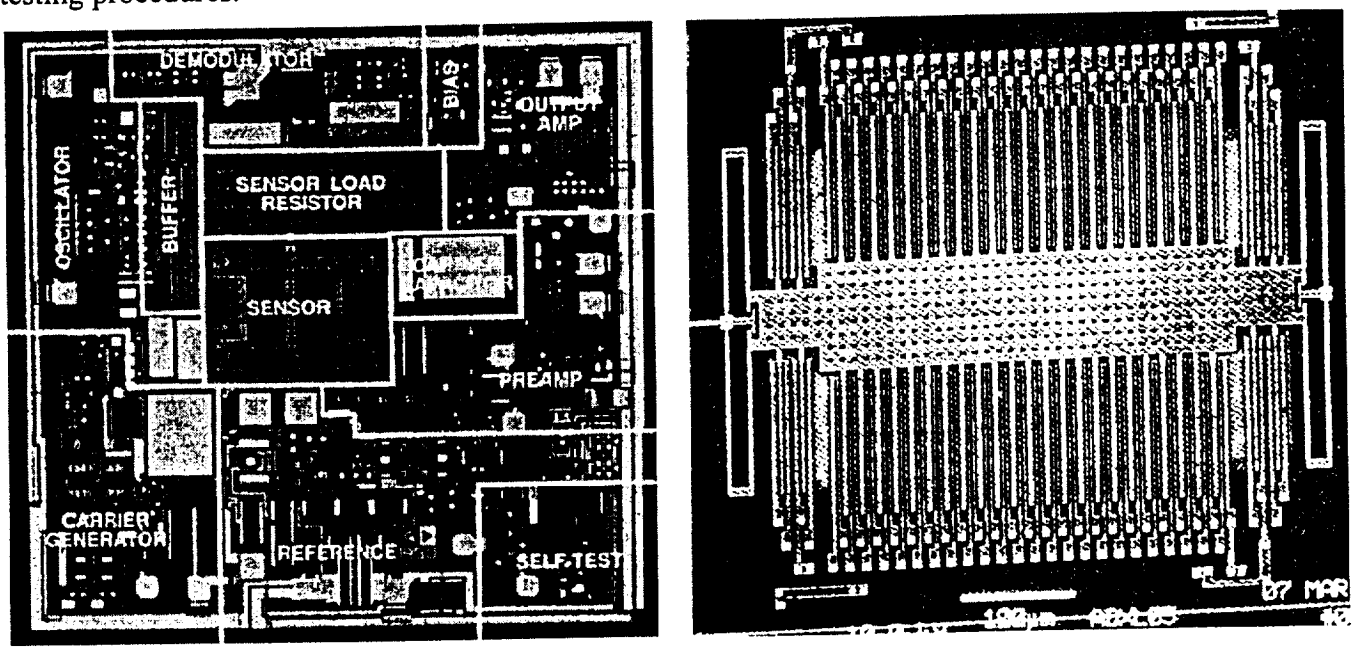


Figure 7. Simplified MAPLE payload.

Presently, most MAPLE configurations consists of four sub-experiments and experiment master controller (Figure 7). Each subexperiment contains a low-power dedicated microcontroller and serial interface to a central master controller. Although MAPLE appears to be a single chassis, the subexperiments are built as separate slices and stacked into one assembly. Alternately, the slices may be physically distributed throughout the satellite structure. In some cases, it is conceivable that two or even three copies of the same MAPLE experiment could be integrated in a given satellite, providing the ability to generate more statistics on selected components.

By establishing an indigenous space experiment payload development program, several test flights have been scheduled over the next few years to establish reliability and operating statistics for MEMS devices in space. The first efforts are the Mighty-Sat launch scheduled for delivery in March 1996 and flight in December 1996, and the Space Test Research Vehicle (STRV-2) which is scheduled for delivery in June 1996 and flight in March 1997. Both will carry MEMS-based accelerometer devices. Originally, it was felt that more devices would be available for these initial missions, but most devices examined were not in a condition where spaceflight would be practical.

The Mighty-Sat test flight will carry Analog Devices' ADXL-02 and ADXL-05 chips, arranged as orthogonal, three axis sets. The ADXL-05 device may well constitute the only MEMS devices that has received any previous exposure to spaceflight.[11] For the sake of expeditiousness, conventional packaging approaches were required, hence the true miniaturized potential of MEMS devices will not be realized in these initial experiments (photomicrographs of the devices are shown in Figure 9). The primary mission of this test flight is to the performance of representative MEMS devices in the space environment. One set of accelerometers will be placed in close proximity to a panel containing explosively released bolts and non-explosive shaped-based memory equivalents. In addition to monitoring these events, a conventional "doorbell"-type solenoid actuator will be attached to one accelerometer cluster to provide additional mechanical stimulation for the purpose of extensive self-testing procedures.



(a) (b)

Figure 9. Analog Device commercial MEMS-based accelerometers. (a) ADXL-50 die[12]. (b) Close-up of sensor portion of ADXL-05 die[13].

Initial low-dose rate ionizing radiation tests were performed on the ADXL-50 in the PL/VTE Cesium chamber, which was biased and calibrated before and after exposure with available vibration tables. The first test was terminated after 5,000 rads dose accumulation, as predicted by the apparatus dosimeter. No changes in performance were found, and more testing is planning to predict and simulate potential on-orbit degradation modes[14].

The STRV-2 test flight will carry the ADXL-05 in a configuration similar to that used in Mighty-Sat. Additionally, the performance of a unique and very sensitive accelerometer developed by the Jet Propulsion Laboratory based on tunneling effects will be examined[15]. In tandem, the two classes of MEMS accelerometers will provide a large dynamic range of response in vibration. Furthermore, the

STRV-2 orbit is highly elliptic and will consequently present a much more severe radiation environment for the devices. If successfully completed and launched, both MAPLE experiments will provide early windows on the performance of at least two types of MEMS devices. Follow-on MAPLE experiments are planned, and the prospects of obtaining a greater variety of MEMS devices for these missions are encouraging.

Conclusions

Microelectromechanical Systems (MEMS) has become the latest technology to expand the range of operation of standard mechanical systems. Boasting such advantages as dramatic reductions in size, weight, cost, and failure rate, MEMS devices have proven their potential to effectively address a wide variety of modern applications. As a means of continuing this trend, MEMS is being considered for use in space systems in which the advantages of such devices are of specific interest. It is of the greatest concern that the cost of launching space systems be minimized and the reliability of such systems be maximized. These inherent abilities of MEMS devices lend themselves favorably to such applications in which the overall weight of a given system can be reduced by several orders of magnitude while simultaneously improving reliability.

In this paper, several current research endeavors were presented which seek to utilize these characteristics of MEMS devices for space systems. Each of these are expanding on the operational limits of standard systems such as accelerometers and inertial sensors. Additionally, other research topics are discussed which are projected for future applications. These topics are currently under consideration and planning for research in the near future.

References

[1]Cosnyka, Edward. *Application of Silicon Micromachinng to Thermal Dissipation Issues in Wafer Scale Integrated Circuits.* MS thesis, Air Force Institute of Technology (AFIT/GE/ENG/91D-12).

[2]Cook, Paul. *Silicon Micromachining Applied to the Management of the Thermal Environment in Wafer Scale Integration Technology.* MS thesis, Air Force Institute of Technology (AFIT/GE/ENG/92D-12).

[3]Walker, Michael. *Shielded High Tempeerature Superconductor Interconnect Technology for Packaging High Speed Cryoelectronic Systems.* Technical report, Phillips Laboratory (USAF) PL-TN--93-1057, March 1994. Distribution limited to Department of Defense except by special permission.

[4]Bowman, Lyn. *Microstructures for Heat Transfer in Stirling Micro-refrigerators.* Technical report, Phillips Laboratory (USAF) PL-TN--93-1032, June 1994. Distribution limited to Department of Defense except by special permission.

[5]Johnson, Barry. *Vertically Integrated multichip Modules.* Technical report, Phillips Laboratory (USAF) PL-TR--95-1066, to be published. Distribution limited to Department of Defense except by special permission.

[6]Sweet, James N. *et. al. Assembly test Chip Version 04 (ATC04) Description and User's Guide.* Publication by Sandia National Laboratory, 20 August 1993. For ordering information, contact James Sweet at (505)845-8242.

[7]Phipps, Mark W. *Design and Development of Microswitches for Microelectromechanical Relay Matrices.* MS thesis, Air Force Institute of Technology (AFIT/GE/ENG/95J-02).

[8]Carderelli, D. and Sappupo, MS. *Flat-Pack Gyroscope.* Technical report, Phillips Laboratory (USAF, to be published. Distribution limited to Department of Defense except by special permission.

[9]Edmans, Daniel. *Micromachined Vibration Sensors for Space.* Technical report, Phillips Laboratory (USAF), to be published. Distribution limited to Department of Defense except by special permission.

[10]Michalicek, M. Adrian, *Design, Fabrication, Modeling, and Testing of Surface-Micromachined Micromirror Devices.* MS thesis, Air Force Institute of Technology (AFIT/GE/ENG/95J-02).

[11]"ADXL05 Accelerometers Fly on Space Shuttle!", *Accelerometer News*, published by Analog Devices, Inc. September 1995, issue 4.

[12]Obtained from Analog Devices World Wide Web site, linked through http::/mems.isi.edu/mems (among other places).

[13]"Low-gravity accelerometer provides DC response," *EDN*, 13 April 1995, pp. 20-22.

[14] Hong, Winn. "Space Worthiness of MEMS Devices: Initial Studies". Draft Phillips Laboratory technical report, prepared in support of Air Force Office of Scientific Research Graduate Summer Research Program, September 1995.

[15]Rockstad, Howard K., *et.al.*, "A Minaiature High-Sensitivity Broad-Band Accelerometer Based on Electron Tunneling Transducers", *Sensors and Actuators A*, **43**(1994) 107-114.

Subminiaturization for ERAST instrumentation (Environmental Research Aircraft and Sensor Technology)

71106
030619
1 P.

Marc Madou

University of California, Berkeley

Max Loewenstein and Steven Wegener

NASA Ames

We are focusing on the Argus as an example to demonstrate our philosophy on miniaturization of airborne analytical instruments for the study of atmospheric chemistry. Argus is a two-channel, tunable-diode laser absorption spectrometer developed at NASA Ames for the measurement of N^2O (4.5 μm) and CH^4 (3.3 μm) at the o.1 ppb level from the Perseus aircraft platform at altitudes up to 30 km. Although Argus' mass is down to 23 kg from the 197 kg Atlas, its predecessor, our goal is to design a next-generation sub-miniaturized instrument weighing less than 1 kg, measuring a few cm^3 and able to eliminate dewars for cooling.

In the new micromachined designs, replacing the multipass resonator Herriott cell (26.1 cm long cell and 72 passes, i.e., an optical pathlength of 18.8 m) is a subminiaturized mirror array to effect a similar optical path length. Replacing the two-channel tunable-diode laser is a self-focusing grating, focusing light from an integrated light source into two narrow bands at 4.5 and 3.3 μm. The optical path, the self-focusing grating, light source, and detectors are all lithographically defined on a common substrate, avoiding separate alignment of optical components and allowing the monolithic structure to minimize thermal effects Under NASA Ames sponsorship, working with CAMD (Center for Advanced Micro Devices) at LSU, we have optimized the design for mirror's arrays on a 1 cm^2 footprint and prototyped such mirrors using LIGA. Separately, at UC Davis, we obtained micromachined broad-band light sources by relying on a micro-heater element on an SOI wafer. In a next step, the LIGA components such as grating and lightpath will be integrated on the same SOI substrate containing the surface micromachined light source as well as detectors.

Current designs enable us to make a small, inexpensive monolithic spectrometer without the required sensitivity range. Further work is on its way to increase sensitivity. We are pursing but not limiting ourselves to LIGA and SOI micromachining approaches. We are continuing to zero-base the technical approach in terms of the specification for the given instrument. We are establishing a check list of questions to hone into the best micromachining approach and superpose on the answers insights in scaling laws and flexible engineering designs to enable more relaxed tolerances for the smallest of the components.

Poster not available at time of publication.

DoD SPACE TEST PROGRAM (STP)

71109
030623
4p

Maj Llwyn Smith
USAF Space and Missile Systems Center

STP Mission

The Space Test Program was chartered by the Office of the Secretary of Defense in 1965 to provide access to space for the DoD-wide space R&D community. The Air Force was directed to serve as the Executive Agent for the program to reduce duplication and achieve efficiencies of scale. To carry out this mission, STP matches a ranked list of sanctioned experiments with available budgets and searches for the most cost effective mechanisms to get the experiments to space. STP has successfully flown over 350 experiments in its long history, using dedicated freeflyer spacecraft, secondary space on the Space Shuttle, and various host satellites. Typical missions (other than those on the Shuttle) provide one year of on-orbit experiment data to the sponsor.

Organization

The Space Test Program belongs to Space and Missile Systems Center's (SMC) Space and Missile Test and Evaluation Directorate (SMC/TE). Early in 1995, SMC/TE was relocated to Kirtland AFB, in Albuquerque NM, from Los Angeles AFB, CA. Consolidation of some of SMC/TE's elements is continuing at Kirtland AFB as of this writing. The Space Test Program Office (SMC/TEL) is managed by Col Peter Young, who oversees the efforts of the following divisions: Mission Design and Management (TELO), Shuttle Payloads (TELH, co-located with NASA at the Johnson Space Center, Houston), the Tri-Service Spacecraft Division (TELS), and the Spacecraft Development Division (TELM, the remainder of the former SMC/CU contingent in Los Angeles). Launch Services, Program Control and Contracting activities are provided through matrixed support.

Space Experiments Review Board (SERB) Process

Each year, the experiment submission process begins with individual research agencies (typically service laboratories or research institutions) ranking their space experiments for submission to the respective services. The services then submit their rankings to the DoD SERB for consolidation, which usually occurs in May of each year. The DoD SERB ranks the experiments based on the their assessment of the experiments' military relevance, along with other factors, such as experiment quality, maturity, and internal service priority. STP is then tasked to fly the optimum number of highly ranked experiments the budget will allow.

Spaceflight Opportunities

Once an experiment makes the priority list there are three general ways which it can gain spaceflight. Mission design and planning effort is initiated to study the most cost effective means for spaceflight, attempting to match up flight opportunities with specific experiments. For experiments with unique orbital requirements that can best be met by free-flying spacecraft, STP contracts for spacecraft development, experiment integration, and launch service. STP also flies experiments as secondary payloads (piggybacks) on spacecraft of various agencies, including NASA and DoD, and various countries, including Russian and French. The third way in which STP gains spaceflight for SERB ranked experiments is through a collaboration with NASA to fly experiments on the Space Shuttle. Some experiments can be located in the payload bay and either retained within the bay or ejected for orbital flight. Other experiments are flown in the mid-deck area of the Space Shuttle crew cabin.

Since the beginning of the program in 1965, 30 piggyback missions have been flown, 27 free-flyer satellites have been launched, and 53 Shuttle flights have carried STP experiments. The total number of individual experiments flown to date is 373. Launch vehicles used so far have included the Space Shuttle, Scout, Ariane,

Delta, Pegasus, Proton, and Taurus. Host vehicles carrying STP experiments have included DSCS, SPOT, the STEP series, and RESURS. Eleven missions have suffered launch failures.

Some examples of how STP spacecraft have played a significant role in the development of military space systems are:

- Spacecraft Charging at High Altitudes (SCATHA) - Measured the charge on spacecraft surfaces and the conditions of the plasma surrounding the spacecraft.
- Combined Release and Radiation Effects Satellite (CRRES) - Collected data on charged particles, electric and magnetic fields, and waves in the near-earth environment.
- STACKSAT - Collected data on communications transceiver and solid-state recorder devices.

Often, technology proven through STP experiments evolves into a new realm of capabilities. The knowledge gained from STP experiments on radiation belts, satellite charging, and rubidium atomic clocks contributed to the creation of the Global Positioning System (GPS), which provides precise navigation information to a variety of military and civilian users.

Operational Missions

Operating Freeflyers

- APEX - Provides spaceflight for three experiments: Photovoltaic Array Space Power Plus Diagnostics (PASP-Plus); Cosmic Ray Upset Experiment (CRUX); Thin Film Ferroelectric Experiment (FERRO).
- RADCAL - Provides a radar calibration target for approximately 70 C-band sites around the world
- STEP Mission 0 - Provides spaceflight for the Technology for Autonomous Operational Survivability (TAOS) experiment.
- STEP Mission 2 - Provides spaceflight for the Signal Identification Experiment (SIDEX) which evaluates multi-integrated small signal detection methods.

Operating Piggybacks

- POAM II - The Polar Ozone and Aerosol Monitor is returning data on the ozone hole in unprecedented detail.
- MAHRSI - Middle Atmosphere High Resolution Spectrograph Investigation measures the concentration of the hydroxyl radical and the nitric oxide in the middle atmosphere and lower thermosphere.
- SWIM - Solar Wind Interplanetary Measurements will be used to develop ways to predict space weather conditions.
- CCGEO - Satellite Charge Control Experiment at Geosynchronous Orbit demonstrates the ability to detect the build-up of charges on spacecraft and then discharge the satellite.

Recent Shuttle Experiments

- RME-III - Radiation Monitoring Equipment III gathered data on the ionizing radiation environment in the Space Shuttle as a function of geographic location, altitude, spacecraft shielding, and spacecraft orientation. Will result in improved risk assessment models for crew and equipment radiation exposures in low earth orbit.
- CREAM - Cosmic Radiation Effects and Activation Monitor measured radiation effects in the crew compartment as a function of time, orbital location and shielding. Data will improve space environment and radiation shielding models used to predict single event upset rates in electronics and background rates in sensors.
- HERCULES/MSI - Provided test and evaluation of a multispectral imager/geolocator system. The geolocator will determine surface location of each image taken by the multispectral imager within 2.5 nautical miles.

- STL - Space Tissue Loss experiment studied the micro-gravity induced tissue loss phenomena of exposed tissue cultures. Results will improve pharmacological agents to extend human activity in the space environment.
- MIS - Microcapsules Production in Space produced space-made microcapsules for performance comparison with similar earth-made microcapsules for improved drug efficacy and decreased drug toxicity.

Planned Missions

Planned Freeflyers

- FORTE - The Fast On-Orbit Recording of Transient Events satellite will address a gap in the space-borne nuclear detonation detection system.
- REX II - The Radiation Experiment II will do further research to overcome and understand the physics of the electron density irregularities that cause disruptive scintillation effects on radio signals.
- ARGOS - The Advanced Research & Global Observation Satellite will house eight high priority space experiments (including 31 sub-experiments) and will flight-qualify key components for operational programs.
- STEP Mission 4 - The Space Test and Experimentation Platform #4 will provide spaceflight for three experiments: Orbiting Ozone and Aerosol Measurement (OOAM); Electro-Magnetic Propagation Experiment (EMPE); and Digital Ion Driftmeter (DIDM).
- TSX-5 - The Tri-Service eXperiments satellite #5 will provide spaceflight for two experiments: The Compact Environmental Anomaly Sensor (CEASE) and the US/UK Space Test Research Vehicle-2 (STRV-2).

Planned Piggybacks

- BINRAD - The Beryllium-7 Induced Radiation Experiment will provide basic knowledge of the low Earth orbit environment explaining unexpected high levels of Beryllium-7.
- POGS II - The Polar Orbiting Geomagnetic Survey II will collect data to update the geomagnetic maps of the Earth.
- POAM III - The third Polar Ozone and Aerosol Monitor will collect upper atmosphere ozone concentration data at lower latitudes as well as at polar regions.

For more information about the Space Test Program, please contact Col Peter Young or Lt Col Joseph Marino at (505) 846-8812, FAX (505) 846-8814. Or write to: SMC/TEL
3550 Aberdeen Ave SE
Kirtland AFB NM 87117-5776

7 1111

Weaves as an Interconnection Fabric for ASIMs and Nanosatellites

Michael M. Gorlick
The Aerospace Corporation
P.O. Box 92957
Los Angeles, California 90009
gorlick@aero.org

Abstract

Many of the micromachines under consideration require computer support, indeed, one of the appeals of this technology is the ability to intermix mechanical, optical, analog, and digital devices on the same substrate. The amount of computer power is rarely an issue, the sticking point is the complexity of the software required to make effective use of these devices.

Micromachines are the nanotechnologist's equivalent of "golden screws," in other words, they will be piece parts in larger assemblages. For example, a nanosatellite may be composed of stacked silicon wafers where each wafer contains hundreds to thousands of micromachines, digital controllers, general-purpose computers, memories, and high-speed bus interconnects. Comparatively few of these devices will be custom designed, most will be stock parts selected from libraries and catalogs. The novelty will lie in the interconnections, for example, a digital accelerometer may be a component part in an adaptive suspension, a monitoring element embedded in the wrapper of a package, or a portion of the smart skin of a launch vehicle. In each case this device must inter-operate with other devices and probes for the purposes of command, control, and communication.

We propose a software technology called *weaves* that will permit large collections of micromachines and their attendant computers to freely intercommunicate while preserving modularity, transparency, and flexibility. Weaves are composed of networks of communicating software components. The network, and the components comprising it, may be changed even while the software, and the devices it controls, is executing. This unusual degree of software plasticity permits micromachines to dynamically adapt the software to changing conditions and allows system engineers to rapidly and inexpensively develop special-purpose software by assembling stock software components in custom configurations.

1 Introduction

Without extensive software support nanomachines, microdevices, and application-specific integreated microinstruments (ASIMs) are just so much fancy dirty glass. Indeed from the perspective of a computer scientist many of the devices being proposed can be regarded as multicomputers with "unusual" peripherals. Another perspective, one which emphasizes their information content, regards these constructions as collections of sensors, actuators, and transmuters, whose purpose is to obtain, produce, and transform information. Even small assemblages may require significant amounts of software. For example, a modern rechargable electric razor contains about 2 kilobytes of software, a digital thermostat about 12 kilobytes, and an automotive emissions control system contains in excess of 300 kilobytes of software. Satellites based on nanotechnology will contain a wide assortment of interconnected digitally mediated subsystems including attitude control, power, navigation, communication, and sensors whose combined software elements may easily exceed tens of megabytes. The economic assembly of such systems will require:

- software that can be assembled component-wise from stock piece-parts;

- software that can be flexibly reorganized to cope with the introduction of novel components or new combinations of common subassemblies; and

- software that can be dynamically reconfigured to compensate for hardware failures or changes in the mission of the satellite.

One medium that addresses these issues is weaves, a component-based approach to software composition and interconnection. Weaves are networks of components in which streams of arbitrary objects flow from one component to another. They occupy a computational niche midway between fine-grain dataflow and large-grain stream processing (as exemplified by Unix pipes and filters). A detailed discussion of the computational and communication semantics of weaves (including the features that distinguish it from data-flow languages like Khoros[2], Show-and-Tell [6], and Prograph [1]) can be found in [5]. We have implemented a visual software composition and integration environment for constructing systems as weaves. The weave visual editor, *Jacquard*, provides users with mechanisms for rapidly assembling weaves from components, executing and observing weaves, and combining and modifying weaves dynamically — all using nothing but point, click, drag, and drop. A more detailed view of the environment for constructing weaves can be found in [3].

Weaves are well suited for systems characterized by processing on continuous or intermittent streams of data, and they have been applied to such tasks as satellite telemetry processing, tracking, and the rapid prototyping of satellite ground stations. Figure 1 shows a portion of a stereo tracker implemented as a weave. Weaves are comprised of sockets (which are either unpopulated or populated), tool fragments, and jumpers. Unpopulated sockets are placeholders for tool fragments that consume objects as inputs and produce objects as outputs. Jumpers between sockets provide transport services for moving objects from one place to another and thus define the topology of the network. Users assemble weaves by interconnecting sockets and populating each socket with a tool fragment (which may itself be a weave). Type information about the objects flowing through a connection is specified by labeling the jumper.

Weaves adhere to the three rules of *blind communication*:

- no tool fragment in the network is aware of the sources of its input objects or the destinations of its output objects, consequently, all tool fragments are independent of their position in the topology of the network;

- no tool fragment is aware of the semantics of the transport services that are used to deliver its input objects or transmit its output objects, consequently, new transport services can be freely substituted for old.

- no tool fragment is aware of the loss of a connection and to the extent that the computation can continue it will, consequently, weaves can by dynamically edited and rewired without risking the integrity of the weave.

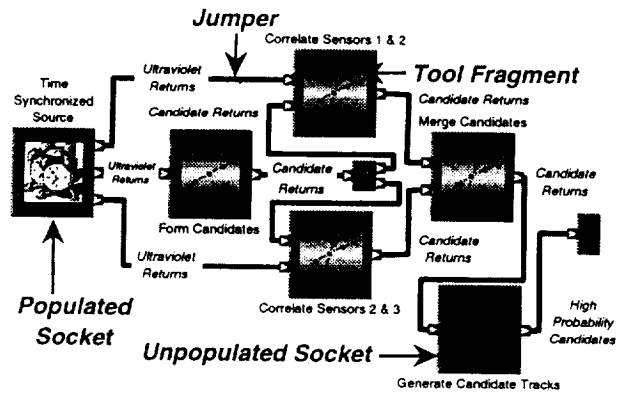


Figure 1: A portion of a weave-based stereo tracker.

Weaves were specifically designed to tackle the problem of constructing large systems by composing components and interconnections — the visual equivalent of a module-interconnection language. A weave that contains one or more unpopulated sockets can be thought of as a framework for an entire family of implementations that are customized by populating empty sockets with components. Figure 2 illustrates such a framework for a sensor and its controller. The sensor is connected to the controller via a feedback loop through which the controller issues commands and control messages to the sensor and receives sensor-specific measurands and status information.

This weave can be used as a tool fragment in some higher level construction since it contains an input pad that passes commands and controls to the controller socket in from the outside and an output pad that transmits measurands and status from the controller socket to the outside. Figure 3 illustrates a particular

instantiation of the framework, an integrated vibration sensor, that contains the tool fragments to command and control a vibration microsensor and deliver its measurements to some higher level device.

In the following sections we illustrate how weaves can be used to integrate large numbers of diverse microinstruments. Two different approaches are shown. In Section 2 we examine a weave framework designed to support a small number of tightly coupled microinstruments bound together as a multiparameter sensor. In Section 3 we outline a weave framework based on message buses that is suitable for large numbers of loosely coupled subsystems such as that found on a nanosatellite. Finally in Section 4 we frame some of the research questions that must be resolved to make this approach viable.

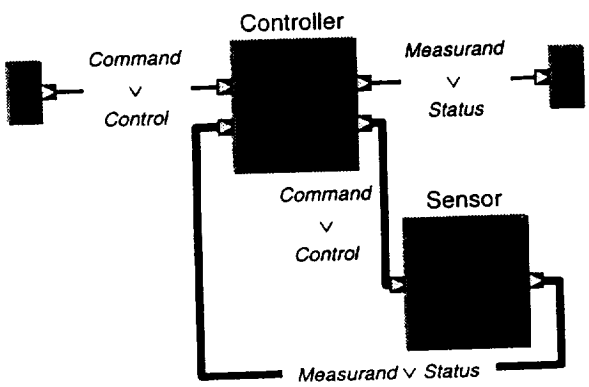


Figure 2: A generic framework for a sensor and its controller.

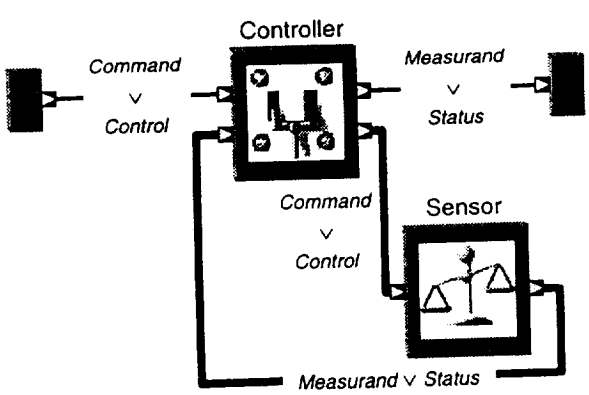


Figure 3: A instantiation of the generic sensor framework as a vibration sensor.

2 Weaves for Multiparameter Sensors

A multiparameter sensor module combines several sensors on a single substrate. A simplified layout of one such hypothetical module is shown in Figure 4. This module provides location, acceleration, sound level, and the detection of one or more chemical species and contains an wireless link for the transmission of telemetry and the receipt of commands. It would be powered by a thin film battery layered on the underside of the substrate and would be so small (certainly no larger than a pack of cigarettes) that it could pasted almost anywhere such telemetry might be required. Possible applications include environmental monitoring, industrial process control, and launch vehicle data acquisition. Specialized versions of such modules could be incorporated into the skins and structures of aircraft or trucks, or strategically placed on bridges or within buildings.

Multiparameter sensors offer three advantages over a comparable collection of independent sensors:

- all of the sensor readings are location correlated, that is, all measurands are being collected at the same location (within a few tens of millimeters);
- the individual instruments can be tightly coupled, for example, their sampling rates can be phased or synchronized;
- the activity or sampling frequency of one instrument can be made dependent on the measurands of another, for example, a multiparameter sensor in a rocket motor compartment could increase the sampling frequency of its temperature sensor when the vibration level crosses a preprogrammed threshold.

It makes sound development and economic sense that it should be as easy to construct control software for a multiparameter module as it is to construct the multiparameter module itself, that is, by combining and interconnecting stock piece parts into a cohesive whole. Logically the integration of an N parameter sensor should be the flat composition of N individual sensors, that is, for the multiparameter device that we are considering all sensors are peers equitably sharing a common substrate and resources such as power bus and specialized services such as analog/digital conversion. In the generic sensor framework sketched in the introduction each individual sensor is comprised of a

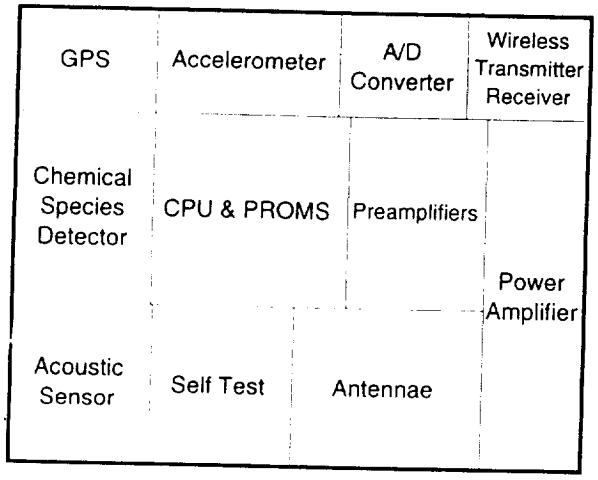


Figure 4: A hypothetical multiparameter sensor module.

controller and the sensor proper. This base organization suggests that a multiparameter sensor be organized hierarchically where each sensor device has its own local controller but reports to, and is commanded by, a higher order controller which is responsible for coordinating the activities of the individual sensors and mediating resource contention.

A sample weave framework for such an organization is shown in Figure 5. The framework is designed to accomodate four independent devices; the changes required for a different number of devices are obvious. The framework supports three different principal classes of tool fragments. Moving from right to left in Figure 5 the first class is represented by the "device" sockets which are intended for tool fragments that are the "embodiment" of the individual sensor devices. These tool fragments are weaves in their own right whose general form is suggested by Figures 2 and 3. Since weaves are indifferent to the composition or form of the tool fragments that populate the sockets of the framework these sensors can be arbitrarily complex.

The second class is represented by the unpopulated "router" socket in the middle of the network. Routers are responsible for the distribution of command and control messages from the higher level controller in the multiparameter sensor to the individual sensor devices. The router helps to insulate the controller from the multiplicities of the sensors, and to a certain extent, from some of the characteristics of the sensors themselves. The router can perform protocol conversions or command translations to supply a uniform unvarying interface to the controller.

The third and final class is represented on the far left by the "controller" which is responsible for accepting higher-level command and control and transforming that into individual sensor commands. Using a feedback loop analogous to that which appears in the individual sensors it also accepts the N-way merged output of the N sensors. Like the lower-order devices that it controls, it has input and output pads thereby permitting the entire multiparameter sensor itself to be embedded in some higher-order device.

A particular instantiation of this framework is illustrated in Figure 6 where a wireless transceiver, an accelerometer, a chemical species sniffer, and an acoustic sensor are combined into a single integrated multiparameter sensor. To the extent that the individual sensors observe a common command and control protocol the high-level controller and its matching router can be generic elements. Note that the same intermixing of custom and stock software components can be applied to both the controller and router which, like any other tool fragment, may be weaves in their own right. The combination of weave frameworks and the hierarchical composition of weaves permit one to construct the software for highly integrated, tightly coupled multidevices in a straightforward and elegant manner.

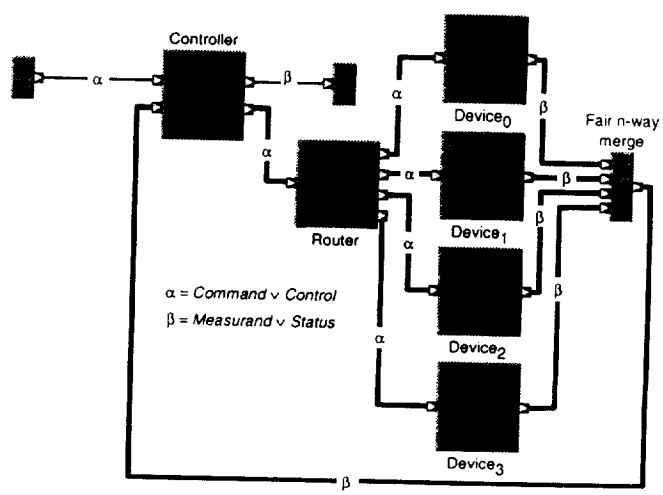


Figure 5: A weave framework for a generic multiparameter sensor.

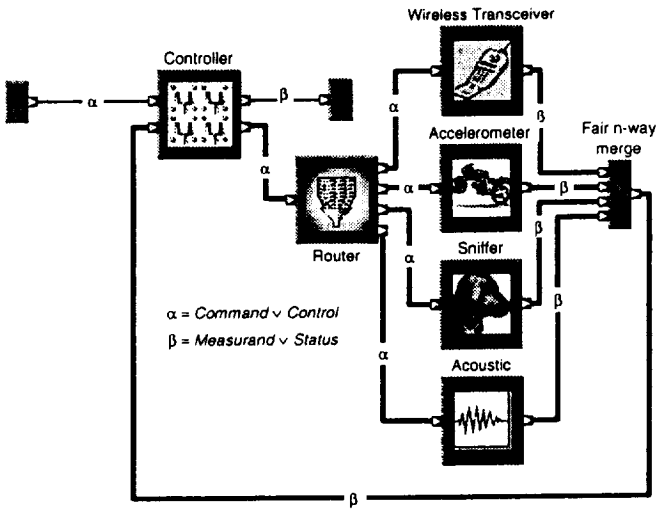


Figure 6: A sample weave for a specific multiparameter sensor.

3 Weaves for Nanosatellites

Unlike a tightly integrated multiparameter sensor a nanosatellite will be composed of a number of loosely integrated subsystems. Those subsystems in turn will encompass a broad degree of integration and coupling ranging from lightly coupled systems that communicate infrequently or irregularly to highly coupled, synchronized systems that require substantial bandwidth. The multiparameter sensor described in Section 2 illustrates some of the techniques required for tight coupling among components. The interconnections among weave components in Figure 6 are all point-to-point which is suitable for components that intercommunicate frequently. However new components can not be added to the weave without rewiring. While weaves fully support rewiring on-the-fly during execution dynamic editing may force the system into an unsafe state due to temporal or behavioral constraints. Furthermore complex systems contain cooperating subsystems that exhibit a variety of communication behaviors ranging from infrequent, low bandwidth communication to regular, frequent, high volume message traffic.

A highly integrated device such as a nanosatellite will be composed from independent subsystems that are mechanically, electrically, or optically integrated with one another. For example, one might construct a nanosatellite by stacking and bonding individual wafers where each wafer is a stock subsystem (guidance, batteries, power management, ...). Ideally the nanosatellite software should be able to recognize the stacking arrangement and arrange its communication paths accordingly. This capability would give system designers the freedom to insert custom subsystems without changing the software base that supports the stock subsystems.

This capability will prove increasingly important as swarms of nanosatellites cooperate to accomplish a task. Hundreds of nanosatellites in close physical proximity could organize themselves as a giant phased array thereby allowing space system architects to assemble on-orbit powerful communications "megasatellites" from individual satellites the size of a tea saucer. It would be advantageous if the software structures of the individual nanosatellites generalized smoothly to the software structures required for the control and management of swarms.

Hierarchical message buses [4] are flexible communication architectures that hold the promise of scaling smoothly from a single nanosatellite to large groups of independent, coordinated nanosatellites. We discuss below a bus-like communication structure that significantly reduces the amount of coupling and therefore may be more appropriate for a collection of semi-independent, cooperating subsystems.

Figure 7 illustrates a framework for a single message bus. Imagine a weave assembled on a sheet of paper where the sheet itself is a broadcast medium for the transmission of objects. Taps into the medium allow sockets to receive messages broadcast on the sheet or to themselves send broadcast messages over the sheet. This arrangement is typified by the socket *Alpha* whose inputs arrive from a tap (an *on-sheet receiver*) and whose outputs are in turn transmitted (via an *on-sheet transmitter*) to any other socket that has a comparable tap into the sheet. As a consequence of blind communication it is impossible for a tool fragment seated in the *Alpha* socket to determine if its inputs are arriving courtesy of a point-to-point connection or are obtained from a bus tap. Similarly, the same tool fragment can not discover that its outputs are being placed on the sheet bus for broadcast.

Sheets (message buses) can be cross-wired using *off-sheet* transmitters and receivers. Each sheet has a unique name. An off-sheet receiver is shown in Figure 7 that is receiving broadcast messages from a sheet named *Source*. All of the messages broadcast on sheet *Source* are being fed as inputs into the socket *Beta*. Likewise all objects output by any tool fragment populating socket *Gamma* will be broadcast on the sheet named *Sink* by the off-sheet transmitter attached to the output pad of *Gamma*.

Finally sheets, like any other weave, are permit-

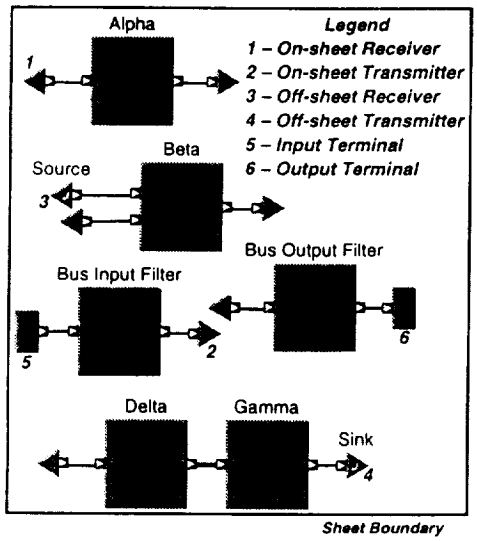


Figure 7: A generic weave bus.

ted input and output terminals. Consequently a sheet can be encapsulated as a tool fragment and may appear as a component in some higher-order weave. This permits message buses to composed hierarchically and dramatically reduces the scope and volume of the object traffic on any one sheet (bus). The combination of hierarchical composition and inter-sheet cross-wiring via off-sheet receivers and transmitters allows system architects to construct complex multicast architectures that scale as the number of sheets (subsystems) increases.

To illustrate some of the possibilities we briefly sketch a high-level nanosatellite software architecture as shown in Figure 8. At the top level the spacecraft is organized as a single sheet (message bus) with a sub-weave responsible for each individual major subsystem. The individual subsystems are each constructed using a combination of point-to-point and bus topologies. A sample framework for the power subsystem is given in Figure 9. It bears a strong resemblance to the framework for the integrated multiparameter sensor shown in Figure 6 with a few important differences. In the power subsystem all of the outputs of the controller are fed into a filter that generates two granularities of monitoring and status information. The monitoring and status information that is fed to the output terminal is of coarser grain than that fed to the on-sheet transmitter. The message traffic appearing on the output terminal is a summary of the activities of the power subsystem that is suitable for processing by a higher-level controller or monitor. The message traffic appearing on the power sheet itself is a finer-grain, more detailed view of those same activities.

The flexibility of weaves and the bus-based architecture outlined above make it possible to add monitoring elements to the spacecraft architecture while the craft is on-orbit without extensive weave "rewiring." Figure 10 illustrates this approach. A specialized monitor has been added to the top-level spacecraft software architecture. The inputs of the monitor are derived from two sources, the spacecraft sheet and the power subsystem sheet using an on-sheet and an off-sheet receiver respectively. Note that this modification is completely transparent to the power subsystem and because no rewiring was required can be performed without endangering the integrity of the craft as a whole. When the monitor is no longer required it can be safely removed and the software restored to its original state. Similar techniques could be applied for fault tolerance, hot sparing, on-orbit testing, or reconfiguration.

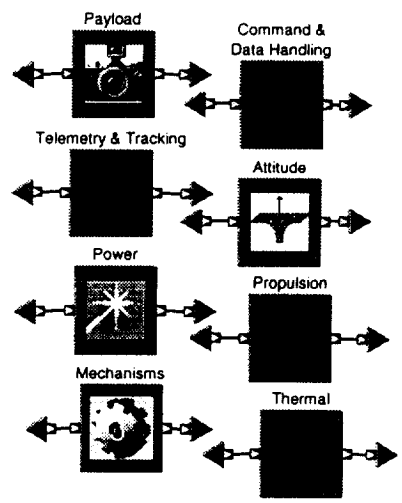


Figure 8: A generic spacecraft architecture.

4 Future Research

One outstanding problem is porting the weave runtime infrastructure to a generic micromachine hardware environment. This must be a cooperative venture between micromachinists and computer scientists. The history of processor architectures is illuminating in this respect. For many years processor architectures were designed by electrical engineers and

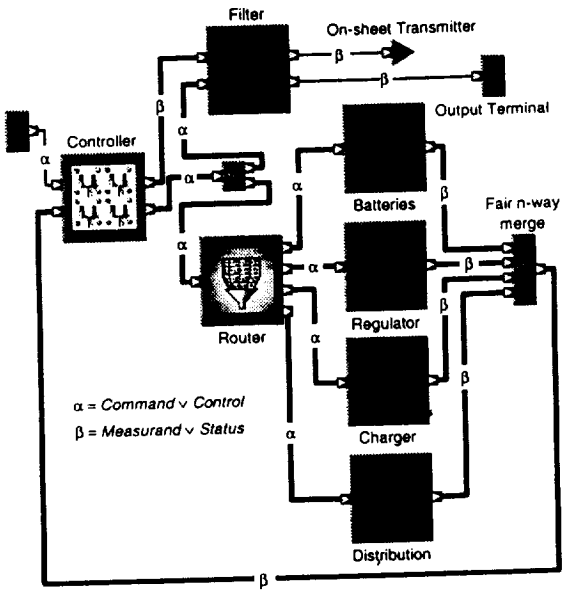


Figure 9: A power system framework.

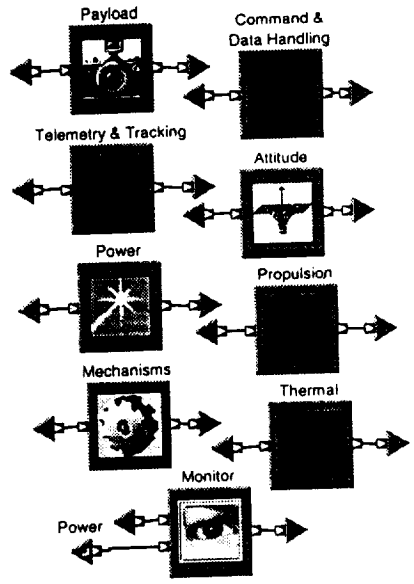


Figure 10: A spacecraft architecture with additional monitoring.

device architects as if software didn't exist and processors were useful in their own right wholly independent of software. We now understand that view to be misguided and the rise in popularity of RISC architectures is the outcome of a joint venture to arrive at architectures that were designed from the outset to support complex software.

The weave execution infrastructure is non-trivial and makes numerous demands on the underlying operating system. It also has strong implications for the bus structure that is used as the communication path among microdevices. The best of all possible worlds would be to design micromachines and their digital controllers with weaves in mind from the very beginning including hooks for atomic weave components (that is weave device drivers) where the bottom of the weave ecology is connected to the particulars of the custom devices provided by the micro-environment.

If one accepts the argument that the missing piece in the weaves-to-micromachines picture is the software/hardware glue, then we must build generic weave components that can talk at the hardware level to a micromachine device, and at the same time, encourage micromachine hardware designers to settle on a generic hardware interface structure that is, at worst, not hostile to weaves. Once this is in place, it should be relatively easy to demonstrate the viability and power of weaves for the software structures of assemblages of micromachines. One attractive, low risk possibility is to simulate a micromachine assemblage using chip-level components and construct a prototype weave infrastructure for such a simulation.

References

[1] P. T. Cox, F. R. Giles, and T. Pietrzykowski. Prograph: a step forward liberating programming from textual conditioning. In *Proceedings of the 1989 IEEE Workshop on Visual Languages*. IEEE Computer Society Press, 1989.

[2] R. A. Earnshaw and N. Wiseman. *An Introductory Guide To Scientific Visualization*, chapter 8. Springer-Verlag, 1992.

[3] Michael Gorlick and Alex Quilici. Visual programming in the large versus visual programming in the small. In *Proceedings of the 1994 IEEE Symposium on Visual Languages*, St. Louis, Missouri, October 1994. IEEE Computer Society Press.

[4] Michael M. Gorlick. Cricket: A domain-based message bus for tool integration. In *Proceedings of the 2nd Irvine Software Symposium*, Department of Information and Computer Science, University of California, Irvine, Irvine, CA 92717, March 1992. Irvine Research Unit in Software.

[5] Michael M. Gorlick and Rami R. Razouk. Using weaves for software construction and analysis. In *Proceedings of the 13th International Conference on Software Engineering*, pages 23–34, Austin, Texas, May 1991. IEEE Computer Society Press.

[6] M. A. Najork and E. Golin. Enhancing Show-and-Tell with a polymorphic type system and higher-order functions. In *Proceedings of the 1990 IEEE Workshop on Visual Languages*, pages 215–220. IEEE Computer Society Press, 1990.

71113

Performance Thresholds for Application of MEMS Inertial Sensors in Space.

030609
6 P.

GEOFFREY N. SMIT
Vehicle and Control Systems Division
The Aerospace Corporation

Abstract

We review types of inertial sensors available and current usage of inertial sensors in space and the performance requirements for these applications. We then assess the performance available from MEMS (Micro-Electro-Mechanical Systems) devices, both in the near and far term. Opportunities for the application of these devices are then identified. A key point is that although the performance available from MEMS inertial sensors is significantly lower than that achieved by existing macroscopic devices (at least in the near term), the low cost, size and power of the MEMS devices opens up a number of applications. In particular, we show that there are substantial benefits to using MEMS devices to provide vibration, and, for some missions, attitude sensing. In addition, augmentation for GPS navigation systems holds much promise.

Introduction

The prospect of the availability of very small, very low cost and low power inertial sensors opens up the possibility of a number of applications for these devices in space. The viability of these applications will clearly depend on the performance of the MEMS devices. The main performance measures we shall discuss are bias stability (drift rate) and radiation hardness. Depending on the application, other performance measures may be more relevant, however these represent key issues for space. In general, the time scales of space vehicle maneuvers are relatively slow compared with those of terrestrial vehicles, justifying the interest in drift rate. The space radiation environment is far more hostile than most terrestrial environments, so that some degree of radiation hardening is required, even for low altitude orbits.

Since the size of the space market is very small compared with many terrestrial markets, most MEMS devices are being developed with an eye toward terrestrial applications (e.g. automotive braking controls). It is of value to assess the extent to which such devices can be adapted to space applications. Alternatively, special devices will have to be developed for space (with a corresponding cost impact).

Current uses of inertial sensors in space

The following table summarizes the main current applications of inertial instruments in space, including typical performance requirements.

Table 1

Current Uses of Gyros in Space

Application	Drift Rate
Launch Vehicles	0.1°/hr
Spacecraft delta-V	0.1°/hr
Spacecraft Pointing	0.01°/hr

Attributes of typical instruments currently in use to meet these requirements are listed in the following table

Table 2

Typical Space Gyros (IMUs)

Type	Weight	Power
SKIRU DII	28 lb	15 - 26 W
Honeywell YG9666	3.6 lb	17.5 W
Delco HRG TNS 311	3.5 lb	10 W

The cost, weight and power requirements of these devices are high enough that their use (particularly in situations where redundancy is required) represents a significant impact on the cost of a program. In many low cost programs, a minimal set of attitude determination instruments is used. Often this does not include a gyro. In many situations this can result in compromised performance, or excessive operational costs later in the mission in the event of component failures or anomalies. Some examples are cited in the following table:

Table 3

Anomalies indicating desirability of MEMS back-up

Vehicle	Anomaly	MEMS could have simplified..
STEP M0	Gyro loss	- delta-V maneuvers - normal mode (reconfiguration from 3-axis to momentum bias was required)
Classified	flat spin	Recovery from flat spin

The desirability of being able to add components to achieve enhanced performance, flexibility or redundancy is clear. We now turn to the question of whether, or to what extent, MEMS devices can fill this need.

Current and near term performance from MEMS devices

The following table summarizes the performance available from experimental units fabricated at C.S.Draper Laboratories.

Table 4

Performance of current MEMS gyros

Attribute	Performance
Angular random walk	0.037 deg/rt.hr.
Scale factor stability	100-150 ppm
Drift rate (60 Hz B/W)	24 deg/hr
Drift rate (0.1 Hz B/W)	1 deg/hr

Mid-term performance (18-24 month delivery) would be of the order:

Angular random walk	0.008 deg/rt.hr.
Scale factor stability	50 ppm
Drift rate (60 Hz B/W)	6 deg/hr
Drift rate (0.1 Hz B/W)	0.25 deg/hr

Long term (3+ years) these numbers may come down to 0.001, 10+, 1 and 0.024 respectively.

The cost of these items depends on the time scale of delivery and on the packaging. Using hybrid electronics (about a 4" x 4" board) and a 6-12 month delivery time, would cost around $300-$500K. With the electronics in an ASIC, the time scale would stretch by about 6 months, and the cost would go up by about $200K. The size of the ASIC system would be less than 1" square. Power and weight would be of the order of 0.25W and 5 grams respectively. In the long term, costs would come down substantially, although the actual numbers would depend on a number of details, including the emphasis placed on the needs of the space community when production units are developed. High production units (e.g. for automotive use) would be very low cost. However, the actual cost to integration into a space vehicle would include whatever modifications would be required.

When we compare the near term performance numbers with the requirements in Table 1, we note that in general the MEMS units cannot be used as direct substitutes for current devices. We must therefore either determine when (or if) the performance of the MEMS devices will reach this level, or we must look into the possibility of finding new or modified applications. We shall concentrate on the latter.

From the standpoint of technological limitations, the MEMS devices have relatively high drift rates, and therefore must be used in "short time scale" applications. The need for radiation hardness, although critical from a practical standpoint, is not driven by a lack of technological capability as much as by a lack of need in the main markets driving the development of MEMS inertial sensors. The vulnerability to radiation occurs in the electronics (FETs etc., used in the preamplifiers and signal processing circuits), not in the MEMS devices themselves.

The following table summarizes a number of applications in which the time scale is short enough to permit the effective use of MEMS sensors.

Table 5

Short time-scale applications in space

Launch vehicles

Augment GPS (esp. for range safety)
Environment monitoring

Spacecraft

Maneuvers
Detumble (e.g., Acquisition, Safehold Modes)
Vibration Control (e.g., Large Structures, Deployables)
Vibration Monitoring (e.g., fault detection on wheel bearings)

In addition to these applications, which would improve the performance and redundancy of current types of satellites, there is the question of future "nanosatellites". In these systems the whole satellite will be built on a "chip" (or at least some wafers). For such applications, the use of MEMS will be mandatory. The necessary performance could be acquired by using the MEMS devices to augment a long time scale sensor (e.g., a miniature star sensor).

Conclusions

There are a number of applications which could benefit from the availability of space hardened MEMS gyros (and accelerometers) - even with the performance limitations currently associated with these devices.

Bibliography

Technical reviews of MEMS inertial sensors can be found in:

1. J. Soderquist, "Microsystems for Navigation," Calibri Pro Development, AB, 1994.
2. B. E. Boser, R. T. Howe, A. P. Pisano, *Course Notes from UCB Short Course on Monolithic Surface Micromachined Inertial Sensors,* May 1995.

An overview of current inertial sensors appears in

3. A. Lawrence, *Modern Inertial Technology* (Springer, 1993).

Information on the Draper MEMS gyros was from J. Gilmore (verbal communication).

ADVANCED MODELING OF MICROMIRROR DEVICES

71114

M. Adrian Michalicek, Darren E. Sene, and Victor M. Bright

230635

Department of Electrical and Computer Engineering
Air Force Institute of Technology
Wright-Patterson AFB, Ohio 45433-7765

16 P.

ABSTRACT

The Flexure-Beam Micromirror Device (FBMD) is a phase-only piston style spatial light modulator demonstrating properties which can be used for phase adaptive-corrective optics. This paper presents a complete study of a square FBMD, from advanced model development through final device testing and model verification. The model relates the electrical and mechanical properties of the device by equating the electrostatic force of a parallel-plate capacitor with the counteracting spring force of the device's support flexures. The capacitor solution is derived via the Schwartz-Christoffel transformation such that the final solution accounts for non-ideal electric fields. This complete model describes the behavior of any piston-style device, given its design geometry and material properties. It includes operational parameters such as drive frequency and temperature, as well as fringing effects, mirror surface deformations, and cross-talk from neighboring devices. The steps taken to develop this model can be applied to other micromirrors, such as the Cantilever and Torsion-Beam designs, to produce an advanced model for any given device.

The micromirror devices studied in this paper were commercially fabricated in a surface micromachining process. A microscope-based laser interferometer is used to test the device in which a beam reflected from the device modulates a fixed reference beam. The mirror displacement is determined from the relative phase which generates a continuous set of data for each selected position on the mirror surface. Plots of this data describe the localized deflection as a function of drive voltage.

INTRODUCTION

A growing trend in optical processing and related fields is the implementation of micromirror-based spatial light modulators (SLMs) for various optical applications [1]. The Flexure-Beam Micromirror Device (FBMD) is a phase-only piston style SLM demonstrating properties which can be used for phase adaptive optics. High optical efficiency and individual micromirror addressability make large arrays of devices well suited to phase-front modulation applications [2,3]. For example, Fig. 1 shows an array of devices used to discretely lengthen or shorten the optical path of incoming light to correct for phase-front aberrations. Other designs of micromirrors, such as the Cantilever or Torsion-Beam devices, have become increasingly favorable for applications in which a redirection of incoming light is desired.

The micromirrors studied in this paper were commercially fabricated using a standard polysilicon micromachining process. In this paper, ideal models are developed for all three designs. Additionally, an advanced model is developed for the FBMD which accounts for surface deformations, fringing losses, and cross-talk from neighboring devices as well as operating conditions such as temperature and drive frequency. The steps taken in this process can be applied to other designs of micromirrors in order to create advanced models for any given device. The model is verified with a microscope-based laser interferometer used to study the behavior of the micromirror devices.

The FBMD, shown in Fig. 2(a), is a 60×60 μm square mirror with flexures attached at the corners spanning two sides of the mirror. The actuation of the device is electrostatic such that a voltage is applied to an address electrode beneath the mirror creating a potential difference between this electrode and the mirror which is grounded. This creates a downward electrostatic force on the mirror which is counteracted by an upward spring force of the flexures. Figure 2(b) represents the actuation and characteristic behavior of the device illustrating these forces.

COORDINATE SYSTEM AND VARIABLES

A convenient characteristic of most micromirror devices is the symmetry designed about the center of the device. Most micromirror devices are designed in the shape of squares or other polygons that share similar symmetric traits. Therefore, a simple Cartesian coordinate system can be assigned to analyze the behavior of micromirror devices which makes use of this symmetry. As shown in Fig. 3(a), the x and y axes lie in the plane of the top of the address electrode and intersect at the center of the device. The z axis defines the vertical dimension within the device. The mirror widths along the x and y axes, w_x and w_y respectively, are shown such that the coordinates used to describe a position along the mirror surface range from negative to positive values of half the width. This coordinate system will help simplify the solutions of symmetric physical properties such as the electric field intensity which is uniform only at the center of the device.

In order to describe the mechanical behavior of micromirror devices, a set of variables must be defined that fully accounts for the physical geometry and motion of the mirrors and flexures. These variables are graphically defined using a simple micromirror device consisting of two flexures supporting the device at opposite ends of the mirror. The flexures and support posts are shown separated from the mirror for the purpose of clarification between the resting and actuated positions of the device.

The flexure variables shown in Fig. 3(b) are comprised of the initial deflection due to gravity, d_g, the actuated deflection at the end of the flexures, d_f, the resting separation distance between the mirror and address electrode, z_o, the actuated separation distance at the end of the flexures, z_f, and the spacer thickness, t_s, used in the fabrication of the device. The mirror variables shown in Fig. 3(a) are a function of position along the surface of the mirror and include the vertical separation distance between the mirror and address electrode, $z_m(x,y)$, and the surface distribution of mirror position relative to the ideal uniform deflection, $\Delta z(x,y)$. This includes mirror surface deformations and tilting of the mirror due to cross-talk or variances in the spring constants of each flexure.

The initial deflection due to gravity can be found using the combined mass of the mirror, M, and the characteristic spring constant of the device, k, such that

$$d_g = \frac{Mg}{k} \tag{1}$$

where g is the acceleration constant due to gravity. If the spring constants of each flexure are known to be unequal, this deflection can be found for each flexure using its portion of the total weight of the mirror. For the purpose of simplicity, however, it is assumed that each flexure of a given device is identical.

As shown in Fig. 3(b), the resting position of the device at the end of the flexures, z_O, is given by:

$$z_o = t_s - d_g = d_f + z_f \tag{2}$$

which describes the vertical separation distance between the address electrode and the mirror at the end of the flexures when no address potential is applied. Likewise, as shown in Fig. 3(a), the separation of the mirror and address electrode is given as

$$\begin{aligned} z_m(x,y) &= z_o - d_f - \Delta z(x,y) \\ &= z_f - \Delta z(x,y) \end{aligned} \tag{3}$$

The most important relationship defines the relative deflection as a function of position along the mirror surface, $d(x,y)$, such that

$$d(x,y) = d_f + \Delta z(x,y) \tag{4}$$

which describes the deflection observed for a given voltage at any point (x,y) along the surface of the mirror. Ultimately, this is the independent variable used to characterize a micromirror device such that this deflection is plotted against the address voltage.

ELECTROSTATIC ACTUATION

In order to compute the electrostatic force on the mirror, it must first be determined by which means this force will be calculated. More specifically, it must be decided whether the charge distribution, which is not uniform over the mirror surface, will be considered. The charge distribution will change with the position of the mirror surface and will also be altered by any mirror surface deformations or discontinuities such as etch holes. This leads to a complicated solution when integrating across the mirror. As an alternative, since both the charge distribution of the mirror and the applied electrode voltage are related to the electric field within the device, it is possible to express the potential energy, ξ, of the electric charge distribution solely in terms of this field such that

$$\xi = \frac{1}{2}\int_A \sigma V dA = \frac{1}{2}\int_v \varepsilon_o E^2 dv \tag{5}$$

where σ is the surface charge distribution on the mirror, V is the voltage between the mirror and electrode, A is the area of the mirror, ε_O is the free space dielectric constant and E is the electric field intensity at any point in the volume v within the device [4]. By assigning a relative electric energy density of $\frac{1}{2}\varepsilon_O E^2$ to each point in space within the device, the physical effect of the charge distribution on the mirror surface is preserved. From this approach it is easy to see that the non-uniform charge distribution on the mirror surface and the fringing effects of electric fields around the edges of the mirror are complementary descriptions of the same electrical phenomenon.

With the ability to express the energy of the device in terms of the electric field, the electrostatic force on the mirror surface is determined by a method known as virtual work [4]. This theory states that the change in the electrical energy of a capacitor is equal to the sum of the mechanical work done by displacing the plates and the change in the electrical energy of the source. The total electrostatic force of an ideal capacitor was determined to be

$$F = \frac{\varepsilon_o E^2}{2} A \tag{6}$$

which represents the total force on the surface of the mirror as a function of electric field. It also demonstrates that the force per unit area on the mirror surface is equal to the electrical energy density per unit volume within the micromirror device [4].

This relationship holds for non-uniform electric fields as well. The fringing electric field around the perimeter of the device alters the force per area on the mirror as a function of position on the mirror surface. The total electrostatic force acting on the mirror is

$$F = \iint f(x,y)dxdy = \frac{\varepsilon_o}{2}\iint E^2(x,y)dxdy \tag{7}$$

The fringing electric field will diminish the force per unit area around the edges of the mirror and will produce a total electrostatic force that is slightly less than the ideal force calculated by neglecting fringing effects.

IDEAL FLEXURE-BEAM MODEL

Since the electric field is symmetric about the center of the device and the mirror and electrode are assumed to be rigid, the electric field lines along the outer edges of the cell shall be assumed uniform as well. Therefore, the induced electric field is initially assumed to be uniform and orthogonal to both the mirror and electrode at all points along both surfaces. This neglects deformations of the mirror surface during operation as well as fringing effects of the electric field around the edges of the device. The total electrostatic force of the Flexure-Beam device, F_{FB}, is found by the method of virtual work and reduces to:

$$F_{FB} = \frac{\varepsilon_o}{2} E_{FB}^2 A \tag{8}$$

where E_{FB} is the ideal electric field, ε_O is the free space dielectric constant, and A is the surface area of the mirror. The uniform separation distance between the address electrode and mirror, z_m, is given in Eq. (3) in which $\Delta z(x,y) = 0$ such that:

$$z_m = z_o - d_f \tag{9}$$

where z_O is the resting separation when no electrode voltage is applied and d_f is the vertical displacement of the mirror at any point along the surface.

The total force is found by substituting Eq. (9) into the expression for the ideal electric field, E_{FB}, in Eq. (8) which yields the magnitude of the downward force applied on the mirror:

$$F_{FB} = \frac{\varepsilon_o}{2}\left(\frac{V}{z_0 - d_f}\right)^2 A \tag{10}$$

The restoring force produced by a spring displaced a distance, d_f, from its equilibrium position is given by Hooke's Law:

$$F_S = kd_f \tag{11}$$

where k is the characteristic spring constant distinct to a particular spring system. This constant is distinct to each spring and can be

measured experimentally or determined using mechanical analysis. It is obvious that the linear response of the restoring force is valid only for a limited range of displacement distances. Forces greater than some critical force applied to the mirror must be avoided to ensure that the flexures do not deform and that the restoring force exhibits a linear response.

It is expected that the flexures will deform linearly. Therefore, balancing the upward restoring force of the micromirror flexures against the downward force of the parallel plate capacitor:

$$F_{FB} = F_S, \qquad \frac{\varepsilon_o}{2}\left(\frac{V}{z_0 - d_f}\right)^2 A = kd_f \tag{12}$$

produces an equality that can be solved to determine the necessary voltage, V, to vertically displace the mirror a desired distance, d_f, from the resting position:

$$V = \left(z_0 - d_f\right)\sqrt{\frac{2kd_f}{\varepsilon_o A}} \tag{13}$$

In this ideal model, the deflection along the mirror surface is assumed to be uniform. In a more realistic model, surface deformations invalidate this assumption and other non-ideal effects of geometry and device operation must be included as well.

As described above, the characteristic spring constant, k, can be experimentally determined for a specific micromirror device. However, mechanical analysis of the geometry and material properties comprising the flexures can approximate this value. As a result, the behavior of a Flexure-Beam Micromirror Device can be obtained without the need for experimental observations.

IDEAL CANTILEVER MODEL

The Cantilever micromirror device can be modeled using the same ideal conditions assumed for the Flexure-Beam micromirror device. Unlike the FBMD, however, the deflection is not uniform along the surface of the mirror, but a function of position along one dimension since the device tilts away from the support post. Assuming no surface deformations, the deflection becomes a linear function of position. Figure 4(a) illustrates the motion of the device and defines the dimension variables. It is known that the flexures will deflect according to Hooke's Law given in Eq. (11), but another aspect of the Cantilever device is the additional bending of the flexure which determines the angle of deflection, θ, at which the mirror is tilted.

As the mirror deflects downward, the force distribution along the surface of the mirror is no longer uniform since the end of the mirror is closer to the address electrode than elsewhere along the mirror. As a result, the total electrostatic force applied to the device will change according to the vertical deflection of the flexure, d_f, and the angle of deflection, θ. To account for this behavior, two spring constants are introduced such that

$$F_S = k_1 d_f, \qquad \theta = k_2 d_f \tag{14}$$

where k_1 describes the vertical deflection at the end of the flexure and k_2 describes the angle of mirror deflection. Both constants are directly related to the amount of electrostatic force acting on the device since they determine the position of the mirror.

The electrostatic force acting on the mirror is found by integrating the linear force distribution across the surface of the mirror. Since this force distribution is uniform in the y dimension, the force is only dependent on the integral over the x domain. Likewise, the separation distance between the mirror and address electrode, z_m, and the vertical deflection distance of the mirror, d, are functions of x and are defined as

$$z_m(x) = z_o - d_f - x\sin(\theta) \tag{15}$$

$$d(x) = d_f + x\sin(\theta) \tag{16}$$

The total electrostatic force for a Cantilever micromirror device, F_C, is found to be [5]

$$F_C = \frac{\varepsilon_o}{2} w_y V^2 \int_0^{w_x} \frac{dx}{z_m^2(x)} = \frac{\varepsilon_o A V^2}{2 z_f z_t} \tag{17}$$

where the vertical separation distances at the flexure end of the mirror and tip of the device, z_f and z_t respectively, are shown in Fig. 4(a) and are defined as

$$z_f = z_0 - d_f, \qquad z_t = z_f - w_x \sin(\theta) \tag{18}$$

Using the deflection relationships of Eqs. (14) and (16), the angle of deflection, θ, becomes

$$\theta = k_2 d_f = k_2\left[d - x\sin(\theta)\right] \tag{19}$$

Since the length of the micromirror device is significantly larger than the separation distance between the mirror and address electrode, the angle produced by the actuation of the device is sufficiently small to allow for an approximation such that

$$\theta \approx \sin(\theta) = \frac{k_2 d}{\left(k_2 x + 1\right)} \tag{20}$$

Equating the electrostatic force in Eq. (17) with the restoring spring force of Eq. (14) yields the ideal characteristic model of the Cantilever micromirror device:

$$V = \sqrt{\frac{2k_1\left[d - x\sin(\theta)\right] z_f z_t}{\varepsilon_o A}} \tag{21}$$

where the address potential, V, is required to deflect a device some distance, d, at some position, x, along the surface of the mirror. Similar to the FBMD model, the spring constants of the Cantilever device can be found from mechanical analysis of the deflection and bending properties of the material comprising the flexure.

IDEAL TORSION-BEAM MODEL

The Torsion-Beam model is similar to the Cantilever model with the exception that only the rotational constant need be considered. The operation of the device is shown in Fig. 4(b) which illustrates that the ideal motion of the mirror does not include a deflection at the flexures. Therefore, the torque produced by an electrode on one side of the device, τ, is directly related to the angle of rotation of the mirror surface, θ, such that:

$$\tau = k\theta = F\bar{x} \tag{22}$$

where F is the total electrostatic force produced by the electrode and $\bar{x}$ is the centroid position at which it is located given by [5]

$$F_T = \int_{x_A}^{w_x/2} f(x)\,dx, \qquad \bar{x} = \frac{1}{F_T}\int_{x_A}^{w_x/2} x f(x)\,dx \tag{23}$$

where x_A is the lateral position at which the edge of the address electrode is located and the ideal force distribution, $f(x)$, is given as:

$$f(x) = \frac{\varepsilon_o W_y V^2}{2z_m^2(x)}, \qquad z_m(x) = z_o - x\sin(\theta) \tag{24}$$

Using the following angle approximation for rotation

$$\theta \approx \sin(\theta) = \frac{d}{x} \tag{25}$$

where d is the desired deflection at some position x and solving for the address potential, V, in Eq. (24) yields the ideal model:

$$V = \sqrt{\frac{2kd^3 / \varepsilon_o W_y x^3}{\ln\left(\frac{z_o - \frac{dW_x}{2x}}{z_o - \frac{dx_A}{x}}\right) + \left(\frac{z_o}{z_o - \frac{dW_x}{2x}}\right) - \left(\frac{z_o}{z_o - \frac{dx_A}{x}}\right)}} \tag{26}$$

which produces singularities at the center of the mirror, $x = 0$, since ideally no deflection can occur at that position [5]. Likewise, a limiting factor must be used so that the model does not predict a desired deflection past the point where the tip of the mirror would touch the substrate and prevent further rotation. The counterweight of the opposite side of the mirror is incorporated into the model by fitting the curve, via the spring constant, k , to the empirical data.

SPRING CONSTANTS

The flexures are modeled as simple springs in which the restoring force in the upward direction is linearly related to the vertical deflection of the mirror by a spring constant that can be determined from the geometry and material properties of the flexures. Furthermore, the mirror and flexures of the device comprise an undamped harmonic oscillator when the device is actuated with a periodic voltage at low frequencies. As a result, the restoring force of the flexures is not only a function of geometry and material properties, but also of temperature and driving frequency. At higher frequencies, however, squeeze film damping may become increasingly significant as the mirror must force air out of the volume of the device during operation.

To analyze the behavior of the flexures, another beam is rigidly supported on one end and free-floating on the other. A force, F, acts in the downward direction at the end of the beam where the maximum deflection, d, from the horizontal is known. The relation between force and deflection produces the cross sectional spring constant, k_{cs}, such that

$$d = \frac{FL^3}{3EI}, \quad I = \frac{1}{3}wt^3, \quad k_{cs} = \frac{Ewt^3}{L^3} \tag{27}$$

where L, w, t, and E are the length, width, thickness, and modulus of elasticity for the beam, respectively [6].

In addition to standard beam theory, the spring constant of the flexures must account for their layout such that a corner will produce a flexure that is more resistant to deflection than one of the same length that is straight. Therefore, a torsional spring constant must be added. As shown in Fig. 2(a), the square FBMD has flexures which span half the perimeter of the device and have several turns in their layout. The torsional spring constant must be evaluated for each corner of the flexure where L_1 and L_2 are the lengths of the primary and secondary portions of the flexure under consideration respectively. The torsional angle through which the primary flexure is rotated by the secondary flexure, ϕ, is given as:

$$\phi = \frac{F_2 L_1 L_2 \sin(\theta)}{KG} \approx \frac{L_1}{(L_1 + L_2)}\frac{d_2}{L_2}, \qquad G = \frac{E}{2(1+\nu)} \tag{28}$$

where

$$K = wt^3\left[\frac{1}{3} - \left(\frac{0.21t}{w}\right)\left(1 - \frac{t^4}{12w^4}\right)\right] \tag{29}$$

and F_2 is the force observed at the end of the secondary portion of the flexure, d_2 is the deflection at the same position, θ is the planar angle between the two portions, and G is the shear modulus of the flexure material. The approximation of ϕ is valid since the deflection observed by the primary portion of the flexure, d_1, is much smaller than the lengths of both portions of the flexures [7]. Solving for the relationship between force and deflection yields the torsional spring constant for a given portion of the flexure:

$$k_t = \frac{KE}{2(L_1 + L_2)L_2^2(1+\nu)\sin(\theta)} \tag{30}$$

There is also a stress term that can be added, k_s, given as:

$$k_s = \frac{\sigma(1-\nu)wt}{2L} \tag{31}$$

where σ and ν are the stress and Poisson ratio of the flexure material, respectively [8]. The system spring constant, k, is found by summing these constants per flexure, k_f, and multiplying by N, the number of flexures for a given device:

$$k = N(k_f) = 4[k_{cs} + k_t + k_s] \tag{32}$$

This constant is a function of temperature since the elastic modulus decreases as temperature increases and the thermal expansion of the flexures will slightly alter their geometry. This constant will be used to extract the elastic modulus as a function of temperature.

SCHWARTZ-CHRISTOFFEL TRANSFORMATION

The electrostatic force of the device is developed using a conformal mapping technique known as the Schwartz-Christoffel transformation. In any map of an electric field, the electric flux and equipotential lines are orthogonal to each other and form curvilinear squares between points of intersection. The sides of these squares will be perfectly linear for uniform electric fields and curved for any non-uniform field. As shown in Fig. 5, the electric field is taken from an original complex plane $\gamma = x + iz$ which describes some polygon and transformed to a complex plane W, where W is an analytic function of γ. This transformation preserves the orthogonal nature of the flux and equipotential lines and alters the sides of the curvilinear squares thus mapping the electric field to the W plane. It provides the means to determine the functional relationship between the two planes such that any electric field can be mapped about any geometry given the initial polygon [9].

The fringing electric field is analyzed using a parallel plate capacitor whose plates extend to infinity along the y axis and for negative x values. This symmetry approach is valid since the fringing effects of the device are localized at the outer edges of the mirror. Transforming a finite plate capacitor results in a solution with several elliptic integrals which is virtually unusable for further calculations [9]. The Schwartz-Christoffel transformation is a widely-accepted tool for such analysis which describes the initial polygon in terms of the exterior angles about which its perimeter

traverses and the points at which the angle is located. The conformal mapping equation is given in as:

$$\gamma = \gamma_o + A\int (w-b_o)^{\alpha_o}(w-b_1)^{\alpha_1}\ldots(w-b_n)^{\alpha_n}\,dw \qquad (33)$$

where γ_O and A are constants determined by boundary conditions, b is the value of each point mapped into the W plane, and n is the number of points mapped into finite values. The quantity exponents, $\alpha = \theta/\pi - 1$, are functions of the external angle, θ, of the transformed polygon at each mapped point in the γ plane [9,10].

The electric field of a parallel plate capacitor originally drawn in the γ plane is shown in Fig. 5(a) where the points being mapped into the W plane are labeled A through D and are enclosed by the polygon drawn around the upper and lower plates of the capacitor. Figure 5(b) represents the mapping of these points in the W plane showing the finite values of points B, C, and D. The constant electric flux lines are mapped into W circularly about point C which produces the relationship:

$$W = \Psi + i\Phi = \left(\frac{V}{i\pi}\right)\ln(w), \quad w = \exp\left(\frac{i\pi W}{V}\right) \qquad (34)$$

where Ψ and Φ represent electric flux and potential respectively, V is the potential applied to the capacitor and w represents the W plane in polar form. Evaluating the exponents at each mapped point, $\alpha_B = \alpha_D = 1$ and $\alpha_C = -1$, the transformation becomes:

$$\gamma = \gamma_o + A\int \frac{w^2-1}{w}\,dw = \gamma_o + A\left[\tfrac{1}{2}w^2 - \ln(w)\right] \qquad (35)$$

Applying the boundary conditions at points B and D in both planes, the constants of the transformation are $\gamma_O = -\tfrac{1}{2}A$ and $A = -(z_m/\pi)$ which produces the final relationship:

$$\gamma = x + iz = \frac{z_m}{\pi}\left[\frac{i\pi}{V}W + \frac{1}{2}\left(1-\exp\left\{\frac{2\pi i W}{V_o}\right\}\right)\right] \qquad (36)$$

This can be solved for the real and imaginary parts to produce the parameterized solution in two dimensions for the edge of a parallel plate capacitor. Doing so yields

$$x = \frac{z_m}{2\pi}\left[\psi + 1 - e^{\psi}\cos(\varphi)\right] \qquad (37)$$

$$z = \frac{z_m}{2\pi}\left[\varphi - e^{\psi}\sin(\varphi)\right] \qquad (38)$$

where z_m is the vertical position of the mirror above the address electrode. The index parameters ψ and φ are normalized functions of flux and potential respectively, such that

$$\psi = -\frac{2\pi\Psi}{V}, \qquad -\infty \le \psi \le \infty \qquad (39)$$

and

$$\varphi = \frac{2\pi\Phi}{V}, \qquad 0 \le \varphi \le 2\pi \qquad (40)$$

where Φ is the potential variable, Ψ is the electric flux variable, and V is the potential applied between the mirror and electrode. The result of Eqs. (37) and (38) is plotted in Fig. 6(a) which demonstrates that the fringing effects are only present at the edges of the mirror. Moving toward the center of the device, away from the edges of the mirror, the electric field and equipotential lines approach ideal uniformity. The fringing effects are only considered for field lines on the underside of the mirror ($\psi \le 0$) since neighboring micromirror devices prevent the extended fringing that would produce field lines emanating from the top of the mirror and underneath the electrode. Micromirror devices standing alone may experience a larger fringing loss than devices positioned within an array due to the existence of these extending electric field lines.

For devices standing alone, the electrostatic force along the electric field lines outside the device acts in the opposite direction as those within the device. Although the arc lengths of these lines are much larger and thus the electric intensity much weaker, the net electrostatic force of these lines should not be neglected. Integrating this solution along the top of the mirror produces some non-zero force in the upward direction that counters the actuation of the device. The net electrostatic force acting on devices standing alone is somewhat less than that on devices within an array.

ELECTRIC FIELD INTENSITY

To find the electric field intensity as a function of position along the mirror surface, the length of the arc traced along a constant electric flux value must be determined. Recognizing that differential change in the potential function, $d\varphi$, will result in differential change in position, dx and dz, the relation is found to be

$$dl = \sqrt{dx^2 + dz^2} = \sqrt{\left(\frac{dx}{d\varphi}\right)^2 + \left(\frac{dz}{d\varphi}\right)^2}\,d\varphi \qquad (41)$$

where dl is the differential change in the arc length. Using the parameterized solution of x and z to find the derivatives with respect to the potential function, φ, and integrating yields

$$l = \frac{2z_m}{\pi}\left(1+e^{\psi}\right)\int_0^{\frac{\pi}{2}}\left[1-m\sin^2\theta\right]^{\frac{1}{2}}\,d\theta, \quad m = \frac{4e^{\psi}}{\left(1+e^{\psi}\right)^2} \qquad (42)$$

which is simply an elliptic integral of the second kind where the need for $m \le 1$ is valid for all values of ψ. Therefore, the elliptic integral series solution is

$$l = z_m\left(1+e^{\psi}\right)\left[1-\sum_{n=1}^{\infty}\left[\left(\prod_{k=1}^{n}\frac{2k-1}{2k}\right)^2\frac{m^n}{2n-1}\right]\right] \qquad (43)$$

which is somewhat difficult to use for real-time modeling due to the recursive multiplication and the need for a large number of terms to converge to a solution [5,11].

As an alternative to the elliptic integral series solution, two approximations were developed that are far more efficient and simpler to employ. First, the numerical integration of Eq. (43) produces a series solution that converges much more quickly and requires significantly fewer terms to maintain a certain degree of accuracy. Another approximation is a curve-fitting approach which produces a closed-form solution in the form of an exponentially increasing function. For all calculations, however, the arc lengths were evaluated by finding the converging limit of Eq. (43) with at least $N = 500$ terms in order to minimize error propagation.

With the address electrode at some potential, V, and the mirror grounded, the field intensity at a position, x, along the mirror is

$$E = \frac{V}{l}, \qquad x = \frac{z_m}{2\pi}\left[\psi + 1 - e^{\psi}\right] + \frac{w_x}{2} \qquad (44)$$

which parametrically represents the electric field intensity as a function of position over half the mirror ($0 \le x \le \tfrac{1}{2}w_x$) where w_x is the width of the mirror in the x direction.

To project this solution into the y domain as well, algebraic averaging was used such that the net electric field intensity at some position along the surface of the mirror is the average of that given by the x and y coordinates.

$$E_{xy} = \frac{1}{2}\left[E_x + E_y\right] \tag{45}$$

where the x and y coordinates are evaluated as:

$$x = \frac{z_m}{2\pi}\left[\psi_x + 1 - e^{\psi_x}\right] + \frac{w_x}{2} \tag{46}$$

$$y = \frac{z_m}{2\pi}\left[\psi_y + 1 - e^{\psi_y}\right] + \frac{w_y}{2} \tag{47}$$

The normalized magnitude of the electric field along the surface of the mirror is shown in Fig. 6(b) as a function of x and y over one quarter of the mirror surface. At the center of the mirror, no fringing effects exist and the ideal uniform electric field is shown. At the edges, however, the fringing effects are quite significant. At the corners of the mirror and address electrode, the electric field intensity is reduced to 78.7% of the ideal magnitude. The solution in one dimension, given in Eq. (44), is the cross section along the diagonal of the solution in two dimensions.

To determine the total electrostatic force acting on the mirror in the downward direction, as given in Eq. (7), the square of the electric field intensity in Eq. (45) must be numerically integrated across the mirror surface using the flux parameter, ψ. It is obvious that the total force will be less than the ideal force calculated using an ideal uniform electric field. The total force becomes

$$\begin{aligned} F_{FB} &= \frac{\varepsilon_o}{2}\int_{-\frac{w_x}{2}}^{\frac{w_x}{2}}\int_{-\frac{w_y}{2}}^{\frac{w_y}{2}}\left[E_{xy}\right]^2 dy\,dx \\ &= \frac{\varepsilon_o w_y}{4}\int_0^{\frac{w_x}{2}} E_x^2\,dx + \varepsilon_o\int_0^{\frac{w_x}{2}}\int_0^{\frac{w_y}{2}} E_x E_y\,dy\,dx \\ &\quad + \frac{\varepsilon_o w_x}{4}\int_0^{\frac{w_y}{2}} E_y^2\,dy \end{aligned} \tag{48}$$

Each of these integrals must be numerically integrated individually due to their distinct integrands. In order to do so, the corresponding parameter, ψ, is divided into N segments which are used to evaluate discrete samples of the electric field intensity and position. The definite integral is evaluated as the sum of the area of rectangles formed in this process. For example, the following integral is numerically integrated such that

$$\int_0^{\frac{w_x}{2}} E_x\,dx = \frac{1}{2}\sum_{i=1}^{N}\left[(x_i - x_{i-1})(E_i + E_{i-1})\right] \tag{49}$$

where the height of the rectangle is defined as the average of the values of the electric field intensity at each side of the rectangle. The remaining integrals are evaluated using the converging limit of the series solution generated by this technique.

The range of parameter values must be chosen to correspond to the range of integration over position. In order to do so, a relationship must be developed between the index parameter, ψ, and the center of the mirror, $x = 0$. Moving away from the edges of the device, the index parameter becomes increasingly negative. Therefore, Eq. (37) can be reduced and solved for the index parameter at the center of the device, ψ_o, such that

$$x = 0 = \frac{z_m}{2\pi}\left[\psi_o + 1\right] + \frac{w_x}{2}, \quad \psi_o = -\left[\frac{w_x \pi}{z_m} + 1\right] \tag{50}$$

The index parameter at $y = 0$ is determined using the same technique. Since it is known that the value of the index parameter at the edge of the device is zero, the resulting range in the index parameter can be used to describe the desired range of integration with respect to position over the surface of the mirror.

ELECTRIC FIELD FRINGING LOSSES

The parametric numerical integration was performed for numerous values of device dimensions, w_x and w_y, and mirror separation distance, z_m, such that an analytic equivalent of this approach could be determined. It was found that the fringing losses are best described as a fractional loss in the ideal force:

$$\Delta f_{FL} \approx \frac{z_m}{165.4}\left[\frac{2(w_x + w_y)}{w_x w_y}\right], \qquad 0 < \Delta f_{FL} < 1 \tag{51}$$

This approximation function is shown in Fig. 7 along with the results of numerical integration. It is obvious that as the mirror area increases, the effects of fringing decrease, thus smaller devices are more affected by such losses to the extent that the ideal solution can not be used. It should be noted that Eq. (51) is valid for other device geometries such that the quantity in brackets is the ratio of the length of the perimeter to the area of the mirror.

Another reduction in the magnitude of the ideal force of a capacitor is the unused area in the surface of the mirror devoted to etch holes. The fractional loss is simply a ratio of the total etch hole area to the ideal area of the device. When this is added to the fringing loss, Δf_{FL}, the total loss, Δf, describes the reduction of the ideal electrostatic force of the device due to such non-ideal characteristics. The net electrostatic force acting on the surface of the mirror in the downward direction becomes

$$F = \frac{\varepsilon_o}{2}\left[1 - \Delta f\right] \iint\left(\frac{V}{z_m(x,y)}\right)^2 dxdy \tag{52}$$

where $z_m(x,y)$ represents the vertical separation distance between the electrode and mirror at any given position within the device and will not be uniform due to mirror surface deformations.

CROSS-TALK INTERFERENCE

Another characteristic of the electric field within a device is the interference produced by the electric field lines of neighboring devices. This could alter the electrostatic force on the mirror in two ways. First, the fringing field lines of one device can be distorted by partially conforming to those of another which would change the amount of fringing losses as calculated above. However, since the flexures and support posts between each device are grounded with the mirrors and a gap exists between these geometric features, the electric field fringing loss at the edge of an individual mirror is still dominated by the fringing effects within the device itself.

The second cross-talk effect would be the added force on the mirror supplied along additional field lines emanating from the electrode of a neighboring device. This interference is only present when the primary device is not actuated since the creation of an much stronger electric field within the primary device would

prevent the interference field. As shown in Fig. 8, the mirror of a primary device experiences a small force along the electric field lines from the first of four neighboring device that are actuated. If the address potential of the primary device, V_P, is approximately zero, the net cross-talk force supplied along the electric field lines is simply the integral of the linear force distribution along the surface of the mirror. This distribution is determined by the address potential of the neighboring device, V_1, the length of each electric field line, L, and the angle of the force vector, θ. The length of the electric field lines is given by

$$L(x) = \sqrt{\left(\Delta x^2 + z_o^2\right)} = \sqrt{\left(x + x_s + \frac{w_x}{2}\right)^2 + z_o^2} \qquad (53)$$

where x_s is the separation distance between each device as shown in Fig. 8 which also shows Δx as the horizontal distance between the neighboring address electrode and any point along the surface of the primary mirror. The linear force distribution is found to be

$$f_1(x) = \frac{\varepsilon_o}{2} w_y \left(\frac{V_1}{L(x)}\right)^2 \cos(\theta) = \frac{\varepsilon_o}{2} \frac{w_y V_1^2 z_o}{\left[L(x)\right]^3} \qquad (54)$$

which is not a function of position in the y direction. Since this distribution is not symmetric about the center of the device, the side of the mirror nearest the neighboring device will experience a greater force than the opposite side of the mirror. In order to determine the amount of force at both ends of the device, the centroid position, $\bar{x}_1$, and the total force due to cross-talk, F_1, must be found and are defined as [5]

$$F_1 = \int_{-w_x/2}^{w_x/2} f_1(x)\,dx, \quad \bar{x}_1 = \frac{1}{F_1} \int_{-w_x/2}^{w_x/2} x f_1(x)\,dx \qquad (55)$$

It is important to note that the centroid position, $\bar{x}_1$, is not a function of address potential of the neighboring device, V_1, due to the common symmetrical design of the devices within the array. Figure 9 illustrates the linear force distribution of the cross-talk interference for a single neighboring device and illustrates the total force at the centroid position. In one dimension, shown in Fig. 9, the resulting force observed by the flexures supporting each end of the device, F_a and F_b, determines the deflection at each end which will not be equal. The end of the device nearest the actuated neighbor will deflect more than the other. The force at each flexure is proportional to F_1 such that

$$F_a = F_1\left(\frac{a_x}{w_x}\right), \qquad F_b = F_1\left(\frac{b_x}{w_x}\right) \qquad (56)$$

where

$$a_x = \frac{w_x}{2} + \bar{x}_1, \qquad b_x = \frac{w_x}{2} - \bar{x}_1 \qquad (57)$$

These forces are directly related to the deflection at each end by the spring constant of the flexure.

Expanding this analysis into two dimensions, it is known that the y centroid falls on the x axis due to the device symmetry. The total force due to cross talk from the first device, F_1, is localized at $(x,y) = (\bar{x}_1, 0)$ and produces a net downward force at each of the four corners of the device. For a square device, the other three neighbors produce similar forces located at the same position, given in Eq. (55), relative to each mirror. The centroid positions of all four neighbors are shown in Fig. 10(a) as circles numbered according to the corresponding device. The total force due to cross talk, F_{CT}, is centered at the final centroid position, $(\bar{x}_{CT}, \bar{y}_{CT})$, which is determined by the forces of the surrounding devices:

$$\bar{x}_{CT} = \frac{1}{F_{CT}} \sum_{n=1}^{4} \bar{x}_n F_n, \qquad \bar{y}_{CT} = \frac{1}{F_{CT}} \sum_{n=1}^{4} \bar{y}_n F_n \qquad (58)$$

where n is the index of the neighboring devices and F_{CT} is simply the sum of their forces. Similar to the analysis in one dimension, given in Eq. (56), the force observed at each corner of the device is proportional to the total force, F_{CT}, as a function of position relative to the centroid, $(\bar{x}_{CT}, \bar{y}_{CT})$. Figure 10(b) illustrates the final effect of cross-talk which shows the uneven tilting of the mirror in response to the location of the final centroid. In this example, the first and fourth devices are actuated more than the second and third devices which determines the position of the centroid and results in a tilting of the mirror.

The deflection of the mirror due to cross talk is a function of position across the mirror surface and can be obtained by developing an equation of the plane formed by joining the four corners. The function Δz_{CT} represents this deflection such that

$$\Delta z_{CT}(x,y) = \frac{D_{sum}}{4} + \frac{D_x}{2}\left(\frac{x}{w_x}\right) + \frac{D_y}{2}\left(\frac{y}{w_y}\right) + D_{xy}\left(\frac{xy}{w_x w_y}\right) \qquad (59)$$

and the deflection coefficients are given as

$$\begin{aligned} D_{sum} &= d_A + d_B + d_C + d_D, \\ D_x &= d_B + d_C - d_A - d_D, \\ D_y &= d_A + d_B - d_C - d_D, \\ D_{xy} &= d_B + d_D - d_A - d_C \end{aligned} \qquad (60)$$

where d_A, d_B, d_C, and d_D are the deflections at corners A, B, C, and D respectively. The amount of the cross-talk deflection, Δz_{CT}, increases as the distance between devices, x_s, decreases. Therefore, arrays containing micromirror devices in close proximity to each other may be significantly affected by neighboring devices.

To determine the effect of proximity, the maximum deflection of a primary device due to cross-talk, d_{max}, was found by fully actuating all neighboring devices. This analysis was completed for a variety of device separation distances, x_s, and primary mirror surface areas of a square mirror and was found to be:

$$d_{\max} = \frac{4 d_{2\pi}\left(z_o - d_{2\pi}\right)^2}{z_o w}\left[\frac{(x_s + w)L_{\min} - x_s L_{\max}}{L_{\min} L_{\max}}\right] \qquad (61)$$

where $d_{2\pi}$ is the 2π modulation deflection for any arbitrary wavelength, w is the width of the square mirror, and L_{min} and L_{max} are the minimum and maximum arc lengths between devices, respectively, shown in Fig. 8 and defined as:

$$L_{\min} = \sqrt{x_s^2 + z_o^2} \qquad (62)$$

$$L_{\max} = \sqrt{\left(x_s + w_x\right)^2 + z_o^2} \qquad (63)$$

This result is shown if Fig. 11 which illustrates that the effect of cross-talk is dramatically reduced as devices are placed further apart. However, devices in close proximity to each other were found to be susceptible to this interference. Since the actuation of the primary device dominates over the cross-talk interference from neighboring devices, the effects of cross-talk can be removed by setting a resting bias for the micromirrors so that their resting position is at some small deflection.

MIRROR SURFACE DEFORMATION

Another major factor in the behavior of the device is the deformation of the mirror surface during actuation. This behavior is compared to the deformation of a rigid beam supported on each end by ball supports such that the free-floating flexures allow the edges of the mirror to angle upwards as the center of the mirror deflects downward. The maximum deflection, δ, of the beam under a uniform force per unit length, q, is given by

$$\delta = \frac{5qL^4}{384EI}, \quad q = \frac{F}{L}, \quad I = \frac{1}{12}wt^3 \qquad (64)$$

where L, w, t, and E are the length, width, thickness, and modulus of elasticity of the beam, respectively [6].

Although the edges of the beam are allowed to angle upward, the angles produced by very small deflections at the center of the beam compared to its length, $\delta << L$, are negligible. Therefore, the deformation is modeled as a beam rigidly supported at the ends and is represented as one period of a cosine wave having an amplitude equal to half the maximum deflection at the center of the beam, δ. Figure 12(a) represents this beam deflection. For a micromirror device of area A, the maximum surface deformation including an initial deformation due to gravity reduces to

$$\delta = \frac{FA}{(6.4)Et^3} = \left[\frac{\varepsilon_o}{2}\left(\frac{V}{z_m}\right)^2 A + Mg\right]\frac{A}{(6.4)Et^3} \qquad (65)$$

where M is the combined mass of the mirror and g is the acceleration constant due to gravity. Using the above beam analysis, the deformation of the mirror surface becomes

$$z_m(x,y) = z_f - \delta\left[1 + \frac{1}{2}\left(\cos\left(\frac{2\pi x}{w_x}\right) + \cos\left(\frac{2\pi y}{w_y}\right)\right)\right] \qquad (66)$$

where z_f is the vertical position of the flexures at the corners of the mirror. Figure 12(b) shows a surface plot of this function which depicts the maximum deflection along the surface (z_f - 2δ) to be at the center of the mirror ($x = y = 0$). It should be noted that the elastic modulus for the mirror surface will be difficult to predict for devices with several layers of structural, adhesive, and reflective material. Likewise, the peak deflection coefficient, δ, does not include the effects of stress which can significantly alter the deformation behavior of larger devices.

For micromirror devices with the flexures attached at some point along the edge of the mirror, the solution in Eq. (66) is simply rotated and scaled down to fit within the dimensions of the mirror. The rotated coordinates of the solution are found to be

$$x' = s_x\left[x\cos(\theta_x) - y\sin(\theta_y)\right] \qquad (67)$$

$$y' = s_y\left[x\sin(\theta_x) - y\cos(\theta_y)\right] \qquad (68)$$

where the scale factors, s_x and s_y, and rotation angles, θ_x and θ_y, are determined by the geometry of the device and the position at which the flexures are attached. The scale factors must be included in order to generate a solution with areas of zero deformation at the flexures. If neglected, these areas would appear outside the geometry of the device and the model becomes discontinuous at the position of the flexures along the mirror.

The rotated coordinates given in Eqs. (67) and (68) are used in Eq. (66) to produce the contour plot shown in Fig. 13(b) in which the original solution is shown within the dashed lines. The surface of the rectangular mirror where the flexures are attached has no deformation and is shown as white while deeper deformations are shown darker relative to their depth. This contour illustrates the effects of deformations at the corners of the mirror which are free to deform without rigid support by the flexures. In both solutions, the peak deformation is given as z_f -2δ although the peak deformation of the rotated solution will be slightly less than the original solution shown in Fig. 13(a) since the center of the mirror is much closer to the flexures. Therefore, the deflection coefficient, δ, must be reduced. The surface deformation of any rectangular Flexure-Beam device can be represented with this solution.

FREQUENCY RESPONSE

Since the mirror is an oscillator, the spring constant directly determines the resonant frequency of the mirror given its mass. The time response of any harmonic oscillator can be found by solving a differential equation relating Newton's second law and Hooke's law to a sinusoidal driving force [12]. The solution is

$$z(t) = \frac{F_o\cos(\omega t)}{M\sqrt{\left(\omega_o^2 - \omega^2\right)^2 - 4\omega^2\beta^2}}, \quad \omega_o = \sqrt{\frac{k}{M}} \qquad (69)$$

where $z(t)$ is the deflection of the oscillator in time, F_o and ω are the amplitude and frequency of the driving force, respectively, ω_o is the resonant frequency of the oscillator, k is the spring constant in Eq. (32), M is the combined mass of the mirror as determined from the densities and geometries of the materials comprising it, and β is the damping parameter of the device. The device experiences a squeeze-film damping effect by displacing the air within the device as it deflects. The peak deflection response of an oscillator is found by obtaining the maximum deflection of Eq. (69) as a function of frequency. The combined restoring force of the flexures simplifies to a frequency-dependent spring force given by

$$F_s = d_f\sqrt{\left[k - M(2\pi f)^2\right]^2 - 4M^2\beta^2(2\pi f)^2} \qquad (70)$$

where d_f is the vertical deflection of the mirror at the flexures and f is the operating frequency of the device. For low operating frequencies, ($2\pi f << \omega_o$), the force reduces to the static spring force of $F_s = k\,d_f$ given by Hooke's Law.

TEMPERATURE DEPENDENCE

The temperature effects are analyzed by considering the coefficients of thermal expansion for the materials comprising the flexures and mirrors. The length, width and thickness of the device components will increase with temperature which alters such factors as the spring constant of the flexures or the total electrostatic force on the mirror. Consider the length of the flexures as a function of temperature, T, in which

$$L = l_o\left[1 + \alpha(T - T_o)\right] \qquad (71)$$

where l_o is any length at temperature T_o and α is the coefficient of thermal expansion for the flexure material [13]. The temperature dependence of the entire device can then be predicted by applying this analysis to all dimensions of length in the final model.

Additionally, the elastic modulus is a function of temperature where the device becomes more flexible as temperature rises. To find this relationship, the resonant frequency is obtained at various temperatures and Eqs. (32) and (69) are used to extract the spring constant and the elastic modulus as a function of temperature.

ADVANCED FLEXURE-BEAM MODEL

To develop the characteristic model for the device, the electrostatic force given in Eq. (52) is set equal to the spring force in Eq. (70) and solved for the address potential, V, such that

$$V = \sqrt{\left[\frac{2F_s}{\varepsilon_o[1-\Delta f]}\right]\left[\frac{1}{\iint z_m^{-2}(x,y)\,dxdy}\right]} \qquad (72)$$

Recognizing that δ is a function of V as given in Eq. (62), this creates a circular reference when calculating the voltage required to deflect the device a desired distance. Therefore, the spring force is used to replace the electrostatic force given in this equation since these forces are ideally equal. The temperature and surface deformation effects then can be added such that:

$$V = \sqrt{\frac{2F_s}{\varepsilon_o[1-\Delta f]}\tan\left[\frac{\left(z_o - d - \Delta z(x,y) - \delta\right)^2}{w_x w_y\left[1+\alpha_M(T-T_o)\right]^2}\right]} \qquad (73)$$

where

$$F_s = (d - \Delta z(x,y))\sqrt{\left[k_o - M(2\pi f)^2\right]^2 - 4M^2\beta^2(2\pi f)^2}\,, \qquad (74)$$

$$k_o = \frac{k}{\left[1+\alpha_F(T-T_o)\right]}, \qquad (75)$$

$$\delta = \left[\frac{\left[k_o(d-\Delta z(x,y)) + Mg\right]w_x w_y}{(6.4)Et^3}\right]\left[1+\alpha_M(T-T_o)\right] \qquad (76)$$

and where α_F and α_M are the coefficients of linear expansion for the flexures and mirror respectively, d is the desired deflection distance at some location (x,y) on the mirror, and z_o is the resting height of the flexures. This height is related to the initial spacer thickness such that the initial deflection due to gravity is a result of the weight of the mirror related by the spring constant of the flexures. This model is valid as long the desired mirror deflection is greater than the surface deformation at that point.

FABRICATION

The mirror arrays were commercially fabricated by the Microelectronics Corporation of North Carolina (MCNC) using the ARPA-sponsored Multi-User MEMS Process (MUMPS). This fabrication process has three structural layers of polysilicon and silicon dioxide as the sacrificial material. The first polysilicon layer, Poly-0, is non-releasable and is used for address electrodes and local wiring while the second and third layers, Poly-1 and Poly-2 respectively, can be released to form mechanical devices. The MUMPS process allows a layer of metal to be deposited only on the top of the Poly-2 layer. The metal is deposited as the last layer of the fabrication process since the metal is non-refractory and the polysilicon layers are annealed at 1100°C to reduce stress. These active layers are built up over a silicon nitride layer which insulates them from the conductive silicon substrate.

This process is illustrated using a simple device consisting of a metallized mirror, one flexure, and one support post. Note that this design does not use Poly-1. Figure 14(a) shows a cross-section of this design prior to metallization. After fabrication, the sacrificial layers must be etched away to release the mechanical layers. Figure 14(b) shows the released structure after the metal has been deposited and the sacrificial material has been removed.

The unreleased die are delivered from MCNC in a protective photoresist which is stripped off in a three minute acetone bath. The die are then rinsed in deionized water for two minutes. The actual release etch is a two minute dip in concentrated (49%) hydrofluoric acid. The die are then rinsed for five minutes in gently stirred deionized water. After the rinse, they are soaked for five minutes in 2-propanol, then baked dry in a 150°F oven for five minutes. The propanol displaces the water, and when it evaporates its lower surface tension prevents the pull-down and destruction of the released polysilicon structures.

EXPERIMENTAL SETUP

The experimental setup is shown in Fig. 15 in which a microscope-based laser interferometer is used to modulate a fixed reference beam with the beam reflected from the device. An incident laser beam is split into a reference and object beam and each is allowed to travel some distance before they are joined together at an aperture to create an interference pattern. A photo-detector placed behind this aperture produces a current which is linearly related to the intensity of the interference pattern. Along the path of the object beam, the path length increases by twice the vertical displacement of the device under test. Therefore, by using a periodic drive signal and knowing the exact wavelength of the incident laser beam, a continuous sample of the detector current yields an accurate measurement of the displacement of the micromirror surface. Comparing this displacement with the input signal yields the response characteristics of the device [14].

The microscope allows the object beam to be finely focused onto the surface of the mirror such that the spot size is approximately 4 μm in diameter. Since the translation stage supporting the device can be moved in increments of 0.1 μm, the displacement at any location on the mirror can be measured and compared to measurements taken elsewhere throughout the mirror surface. The result is a mapping of the surface deformations or tilting of the mirror as a function of applied potential. A system precision of 2 nm was measured using multiple characterization curves for a single location on one micromirror device.

An additional setup was used to measure the frequency response of the devices studied. A device under test is placed in a temperature-controlled evacuation chamber at 20 mTorr of pressure. A spectrum analyzer is used to measure the mechanical energy of the device using the principle of virtual work. A peak in the mechanical energy is observed at the resonant frequency of the device. The output, however, is relative only to the mechanical energy of the device and does not represent deflection. This procedure can produce accurate evaluations of the spring constant of a device given its resonant frequency and mirror mass.

PROCEDURES AND RESULTS

The ideal models were verified by characterizing the devices and fitting the curve to the data using the spring constant. Data was taken for the FBMD and the Cantilever devices, but not for the Torsion-Beam device because the model was developed after the test chip was sent to fabrication with no Torsion-Beam devices. The model and experimental data for the Cantilever device is shown in Fig. 16 and the Torsion-Beam model is shown in Fig. 17 which illustrates the behavior at several positions along the surface of the mirror. The slight error shown at the center of the Cantilever device in Fig. 16 can be partially attributed to the uncertainty in positioning the 4 μm laser spot. The ideal model of the FBMD is not shown. It was determined that a spring constant of 2.6 N/m accurately fit the curve to the experimental data.

The frequency response of the square FBMD was analyzed using a complex transfer function derived from the Fourier transform of Eq. (69) such that the phase response of the device is preserved. As shown in Fig. 18, which shows this theoretical behavior and mechanical data of the device in arbitrary units, there is a slight miscalculation of 40 Hz between the predicted and observed resonant frequency. This is due to the value used for the mass of the mirror in which the mass of the flexures was neglected.

Theoretically, the damping coefficient, β, can be found as a function of device area, A, mirror separation distance, z_m, and atmospheric pressure by finding the resonant frequency of several devices at various pressures. However, for this particular FBMD, the resonant frequency could not be achieved above 100 mTorr of pressure which indicates that the squeeze-film damping effects on Flexure-Beam devices is quite significant. Devices with lower resonant frequencies and other geometries may not be as affected.

The resonant frequency of the square FBMD was found at various temperatures at 20 mTorr of pressure. The resulting spring constant of each sample, from Eq. (69), was then used to extract the elastic modulus, from Eq. (32), as a function of temperature. This function is shown in Fig. 19 which demonstrates a linear behavior. This function for thin-film polysilicon was found to be

$$E = (-0.03225)T + (172.3931)\ GPa \qquad (77)$$

where T is the Kelvin temperature. The range of temperature could not be expanded due to the limits of the experimental setup. At colder temperatures, condensation from the humidity in the air prevented an accurate characterization of the device.

Tests were conducted to verify the cross-talk and mirror surface deformations. The cross-talk testing involved generating a behavior curve for a device at normal operation and then another curve while its surrounding devices were fully actuated. No significant changes in the behavior were observed which stands to verify the assumption that such cross-talk effects are negligible for this device due to the 18 μm separation distance within the array.

The peak surface deformation in the center of the device was predicted to be 5 nm and measured to be 7 nm. The predicted value is based on the modulus of elasticity for polysilicon extracted from other devices (168 GPa) and is also affected by the stress of the mirror which is comprised of three material layers. As a result, this exact value of deformation is somewhat difficult to predict.

In order to verify the advanced Flexure-Beam model, the square FBMD was driven by a 250 Hz signal ranging from zero to approximately 16 volts while the laser spot was positioned at the corner of the mirror. Comparing the input signal with the resulting phase curve, the device behavior is plotted in Fig. 20 which shows that the device created a 2π phase change in a $\lambda = 632.8$ nm HeNe laser, $d_f = 316.4$ nm, with an address potential of 15.25 volts.

The theoretical behavior of the device, shown as a dashed line, is calculated using design dimensions and the modulus of elasticity, E=168 GPa, determined from a separately fabricated device. The actual modulus of elasticity of a thin film material depends on the fabrication process, and the modulus can vary significantly. Unless the modulus is determined exactly for the device being modeled, the value for bulk silicon, or a value determined from another thin film polysilicon device, must be used as a starting point in the model. Given this uncertainty in the value of the modulus of elasticity, the model will produce a representative behavior for the device. However, by altering only the modulus of elasticity, the representative curve can be shifted to match the observed data.

CONCLUSION

As Fig. 20 illustrates, the characteristic model in Eq. (73) closely predicts the actual behavior of the device presented in this paper. It has also been found to model other devices of various geometries and materials with similar accuracy. The ideal models were found to closely describe the behavior of a large portion of the devices tested once the model was fit to the data by the spring constant. The material analysis performed in the advanced model seems to remove the need to empirically determine this constant.

Micromirror devices can be commercially fabricated in a variety of surface micromachining processes due to their simple, robust design. The ideal models presented in this paper can be used to describe the behavior of a large majority of devices based on their design and motion. For very large or very small devices, the advanced model may be required to characterize the device. These thresholds at which the advanced model should be used is defined by the size of the device and its fabrication process which incorporates other variables such as stress and thickness of various layers. An advanced model was developed for any rectangular piston-style Flexure-Beam Micromirror Device. An advanced model for other micromirror designs can be developed by following the same steps for a particular geometry and fabrication process.

REFERENCES

[1] R. M. Boysel, J. M. Florence, and W. R. Wu, "*Deformable mirror light modulators for image processing.*" in Proc. SPIE, vol. 1151, pp. 183-194, 1989.

[2] R. H. Freeman and J. E. Pearson, "*Deformable mirrors for all seasons and reasons,*" Applied Optics, vol. 21, pp. 580-588, Feb. 15, 1982.

[3] L. J. Hornbeck, "*Deformable mirror spatial light modulators,*" in Proc. SPIE , vol. 1150, pp. 86-102, 1990.

[4] P. Lorrain, D. R. Corson, and F. Lorrain. Electromagnetic Fields and Waves. 3rd ed. W. H. Freeman and Company, New York, 1988.

[5] D. R. Lide, ed. CRC Handbook of Chemistry and Physics. 71st ed. CRC Press, Boston, 1990.

[6] K. K. Stevens. Statics & Strength of Materials. 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[7] M. R. Lindeburg, EIT Reference Manual, 7th ed., Professional Publications, Inc., CA, 1990.

[8] T. H. Lin, "*Implementation and characterization of a flexure-beam micromirror spatial light modulator,*" in Optical Engineering, vol. 33, pp. 3643-3648, Nov. 1994.

[9] H. B. Palmer, "*The Capacitance of a parallel-plate capacitor by the Schwartz-Christoffel transformation,*" in Electrical Engineering, pp. 363-366, March 1937.

[10] P. M. Morse and H. Feshbach. Methods of Theoretical Physics. McGraw-Hill, New York, 1953.

[11] M. Abramowitz and I. Stegun. Handbook of Mathematical Functions. Dover Publications, New York, 1972.

[12] J. B. Marion and S. T. Thornton. Classical Dynamics of Particles & Systems. 3rd ed. Harcourt Brace Jovanovich, San Diego, 1988.

[13] R. A. Serway. Physics for Scientists and Engineers. 2nd ed. Suanders College Publishing, Philadelphia, 1986.

[14] T. A. Rhoadarmer, V. M. Bright, B. M. Welsh, S. C. Gustafson, and T. H. Lin, "*Interferometric characterization of the flexure beam micromirror device,*" in Proc. SPIE, Vol. 2291, 1994, pp. 13-23.

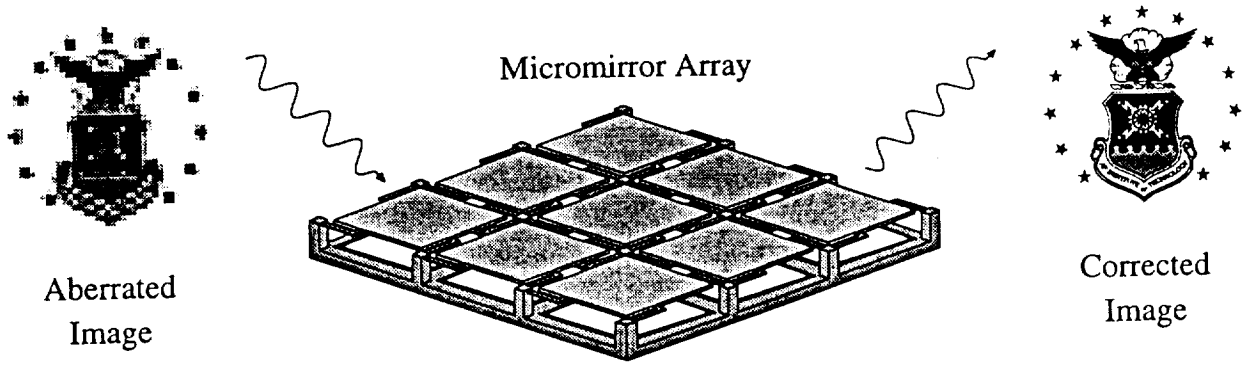
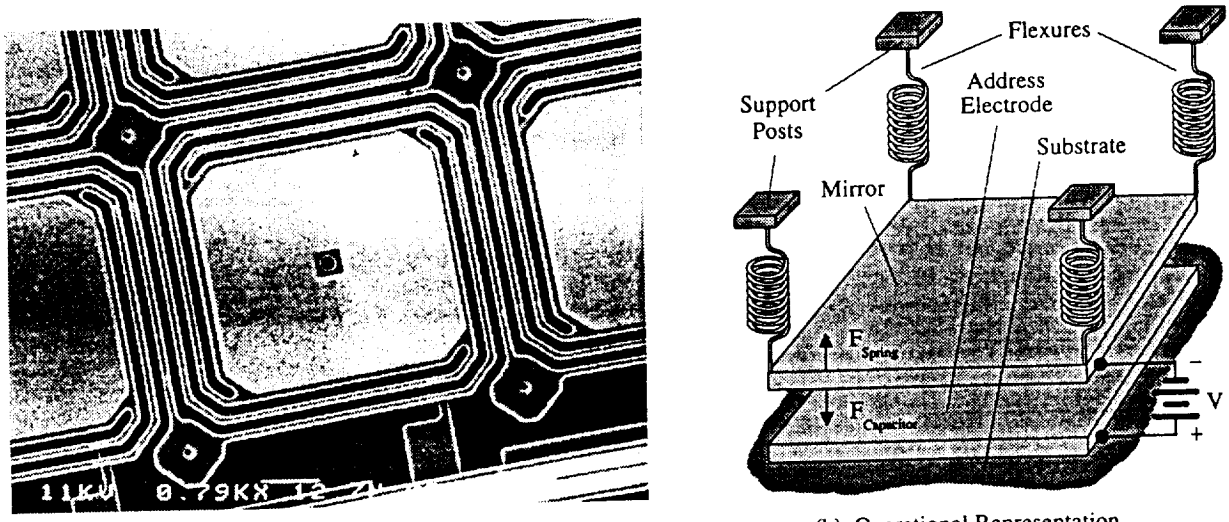
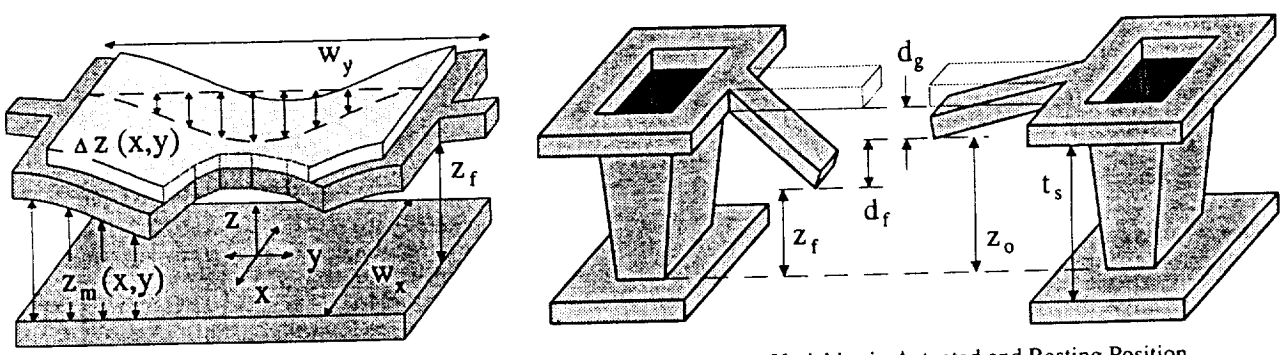


Figure 1. Use of Flexure-Beam Micromirror Devices in phase corrective optics.



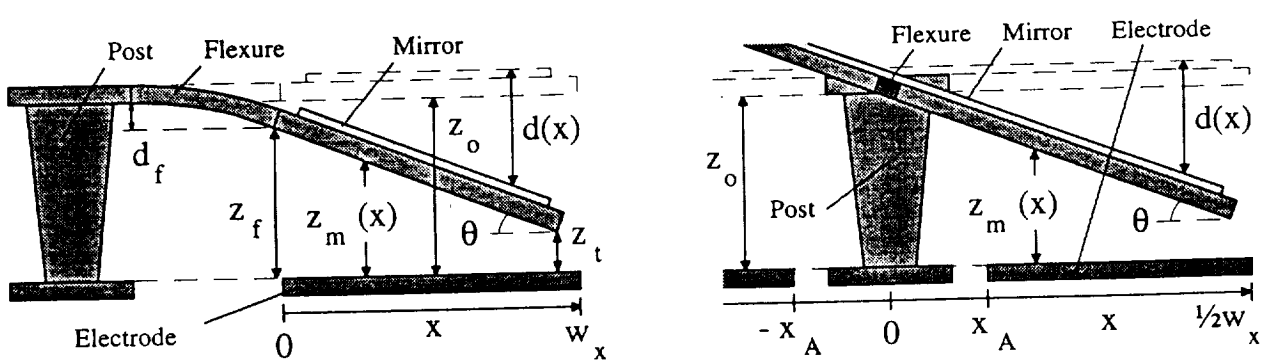
(a) SEM Micrograph (b) Operational Representation

Figure 2. Micrograph and representation of the square Flexure-Beam Micromirror Device.



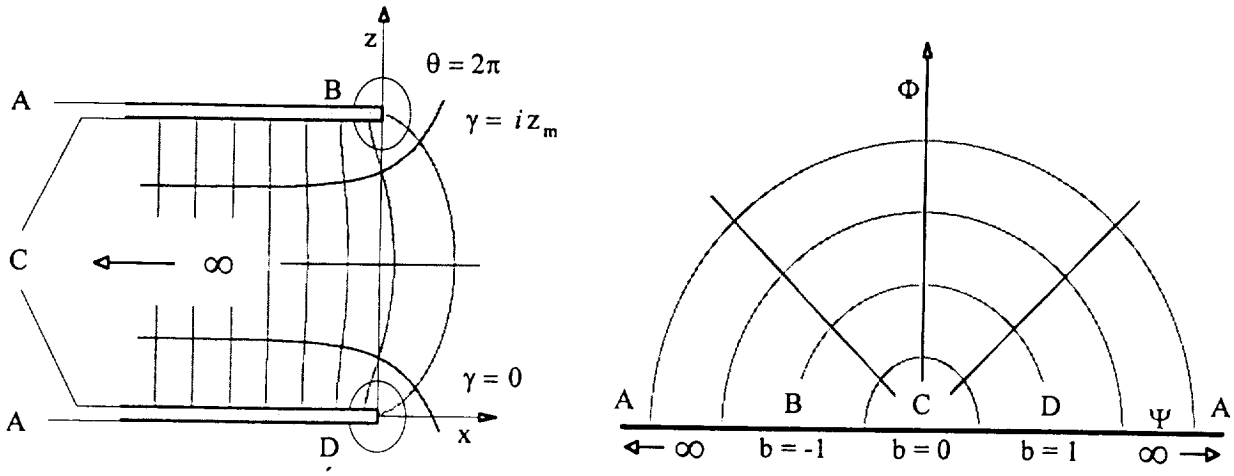
(a) Mirror Variables and Coordinate System (b) Flexure Variables in Actuated and Resting Position

Figure 3. Graphical identification of micromirror device dimension variables and coordinate system.



(a) Cantilever Micromirror Device (b) Torsion-Beam Micromirror Device

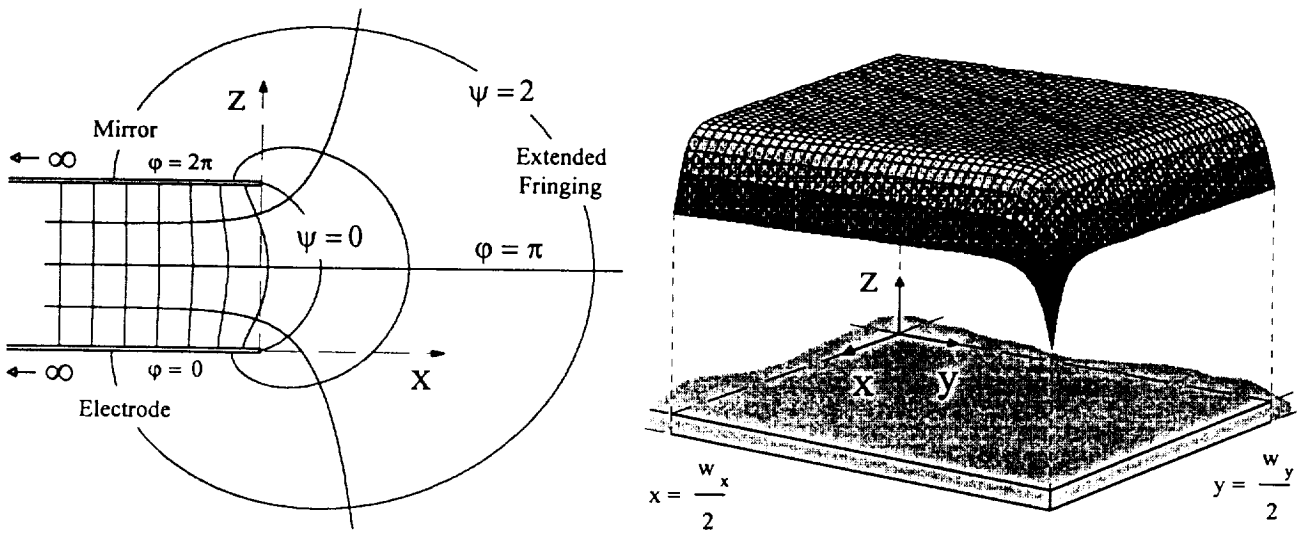
Figure 4. Side view of the Cantilever and Torsion-Beam micromirror deflection with assigned variables.



(a) Original $\gamma = x + iz$ Plane

(b) Transformed $W = \Psi + i\Phi$ Plane

Figure 5. Origianl and transformed planes used in the Schwartz-Christoffel transformation of a parallel plate capacitor [10].



(a) Schwartz-Christoffel transformation solution

(b) Normalized fringing electric field magnitude

Figure 6. Electric field fringing analysis of a parallel-plate capacitor using the Schwartz-Christoffel transformation.

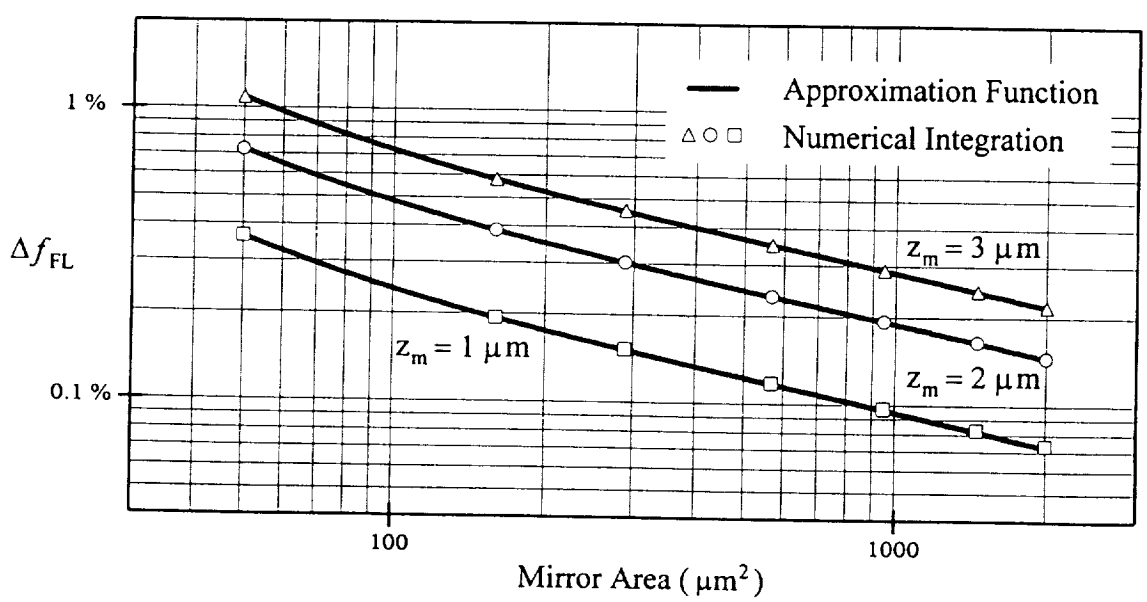


Figure 7. Plot of fringing loss approximation function with respect to mirror area along with numerical integration results.

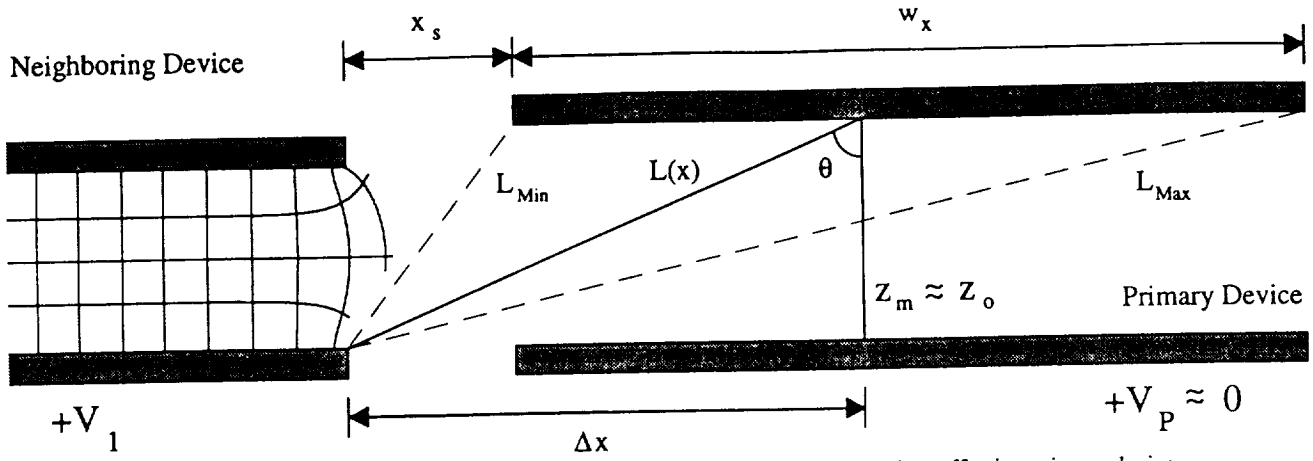


Figure 8. Range of cross-talk electric field lines of neighboring micromirror devices affecting primary device.

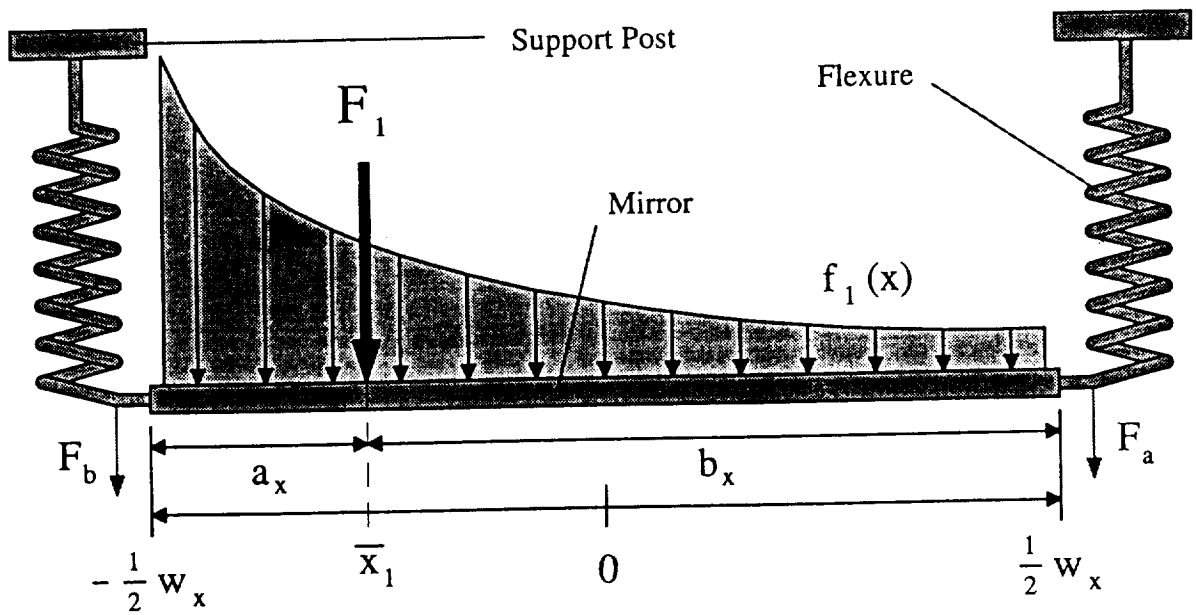
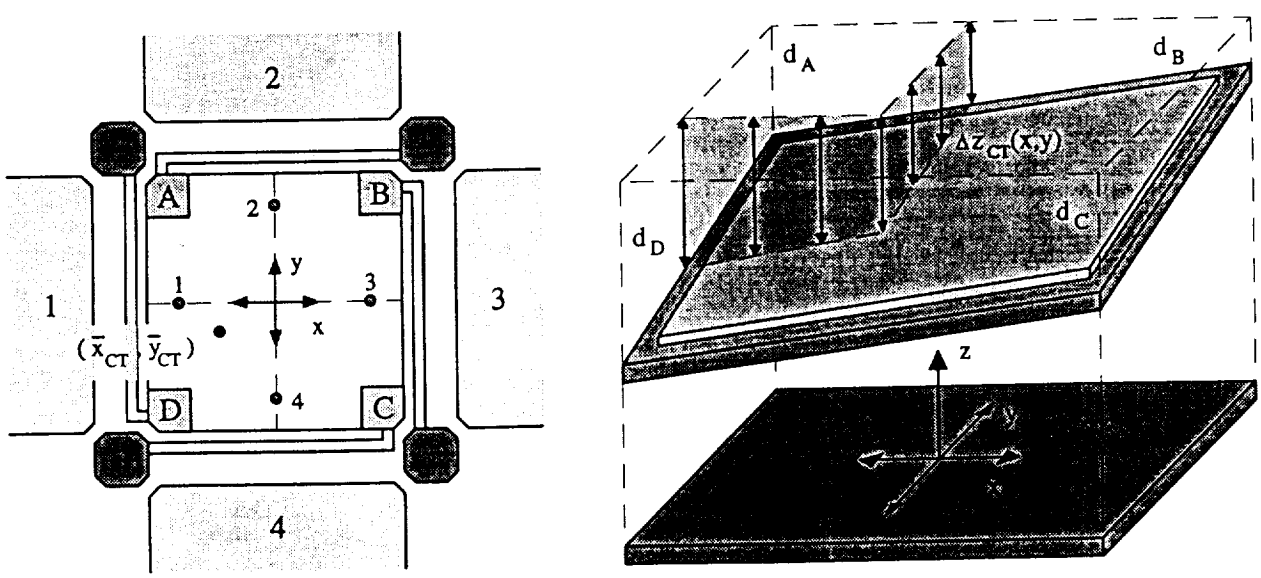


Figure 9. Cross talk linear force distribution along primary micromirror device and resulting forces at each end.



(a) Adjacent Micromirror Devices

(b) Tilting Of Primary Mirror Surface

Figure 10. Cross-talk interference of adjacent devices and resulting mirror surface tilt of the primary mirror surface.

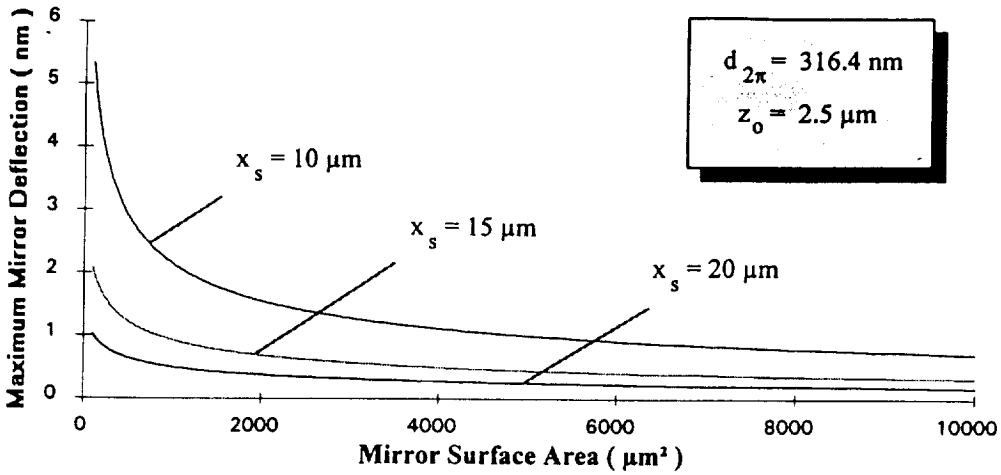
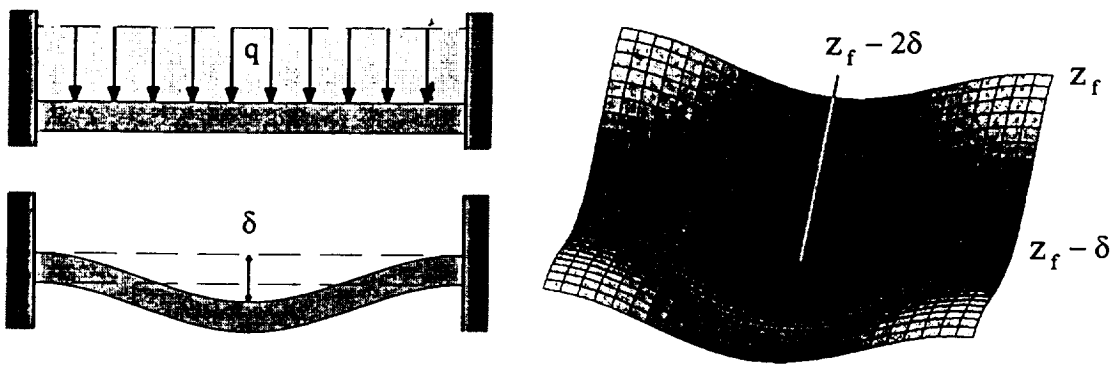
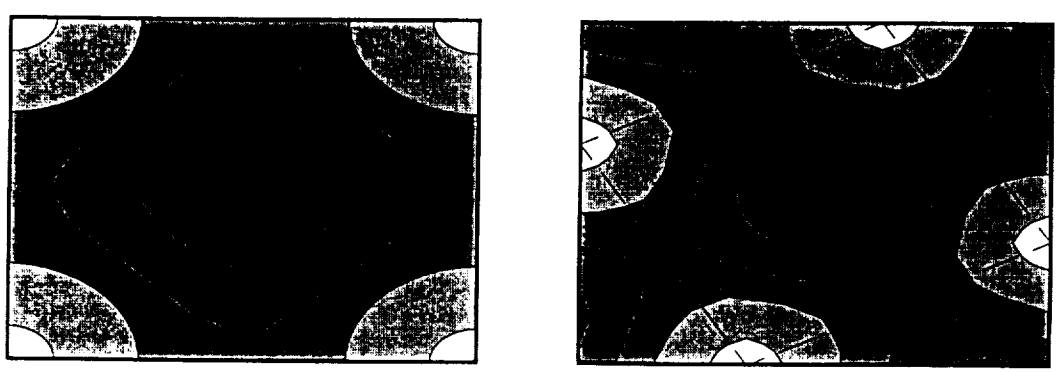


Figure 11. Plot of maximum deflection of primary mirror surface due to cross-talk versus micromirror area.



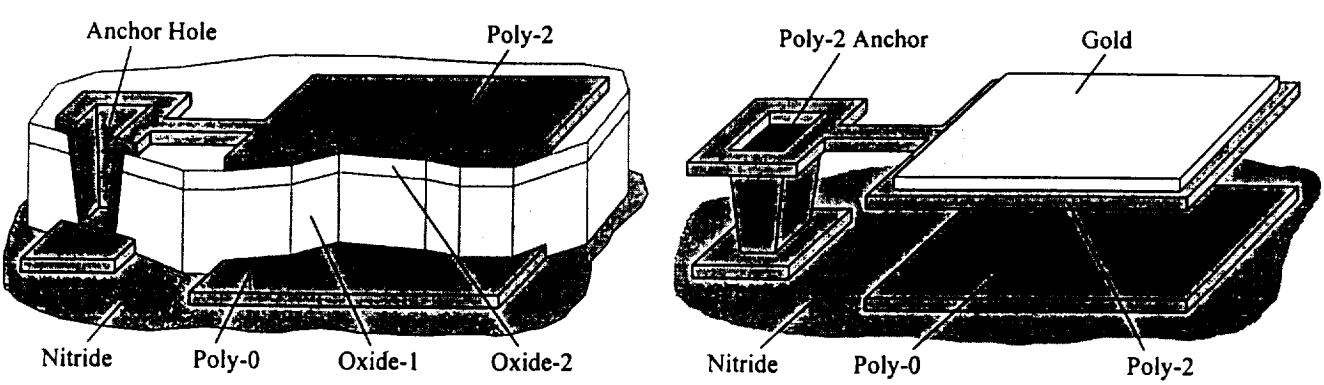
(a) Beam Deformation (b) Mirror Surface Deformation

Figure 12. Use of beam deflection analysis along one dimension to represent mirror surface deformation along two dimensions.



(a) Flexures Attached at Corners (b) Flexures Attached Along Sides

Figure 13. Plot of surface deformation function for a rectangular Flexure-Beam device with two flexure locations.



(a) After Poly-2 Deposition (b) Released Device

Figure 14. Graphical illustration of the MUMPS fabrication process using a simple Cantilever micromirror device.

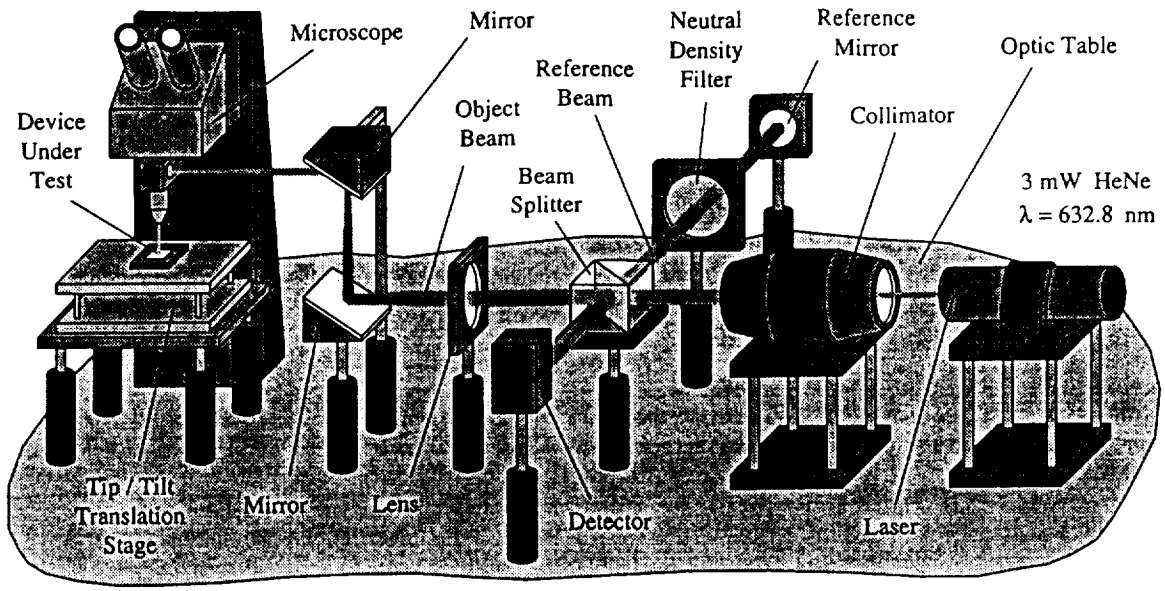


Figure 15. Experimental setup of the microscope-based laser interferometer.

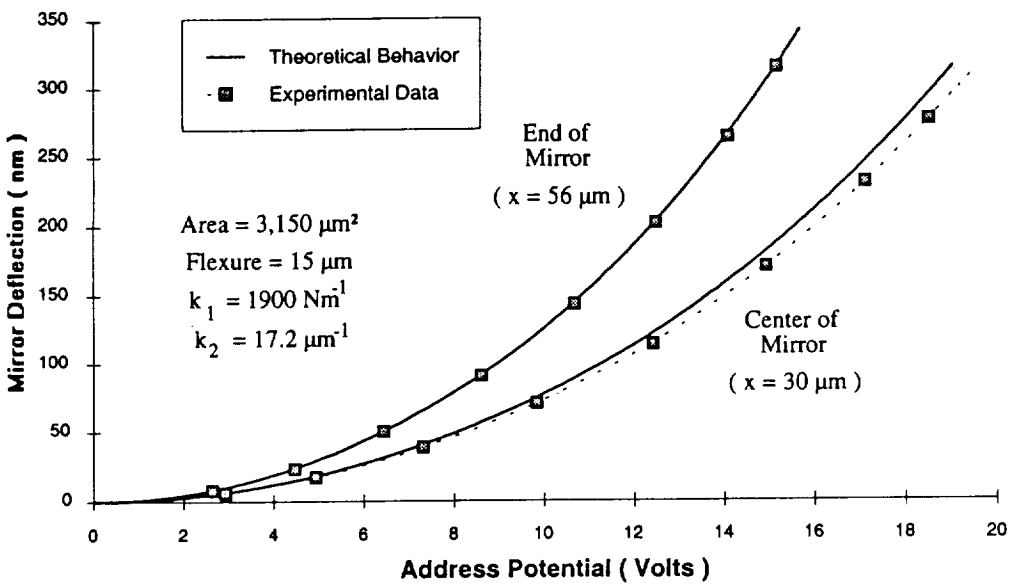


Figure 16. Characteristic behavior curves for two locations on the Cantilever micromirror device.

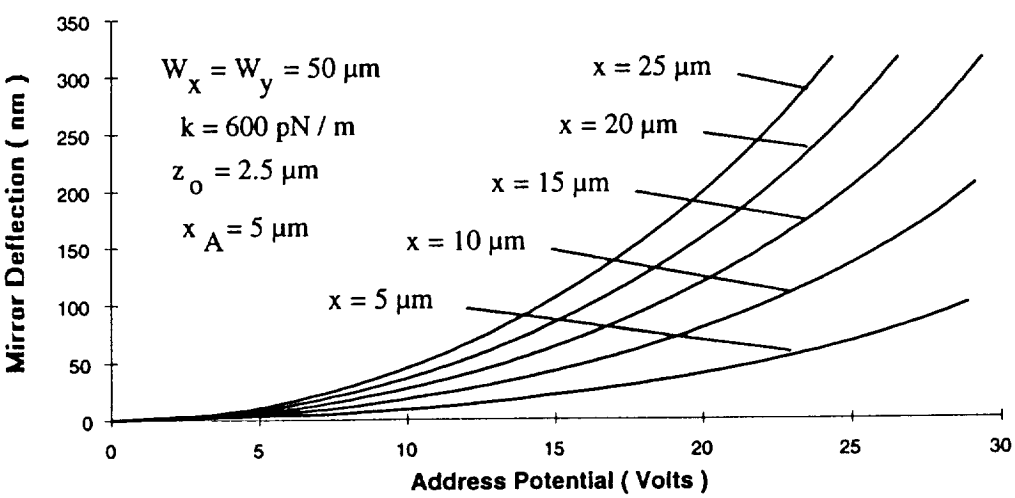


Figure 17. Characteristic behavior of a Torsion-Beam Micromirror Device at various positions along the surface.

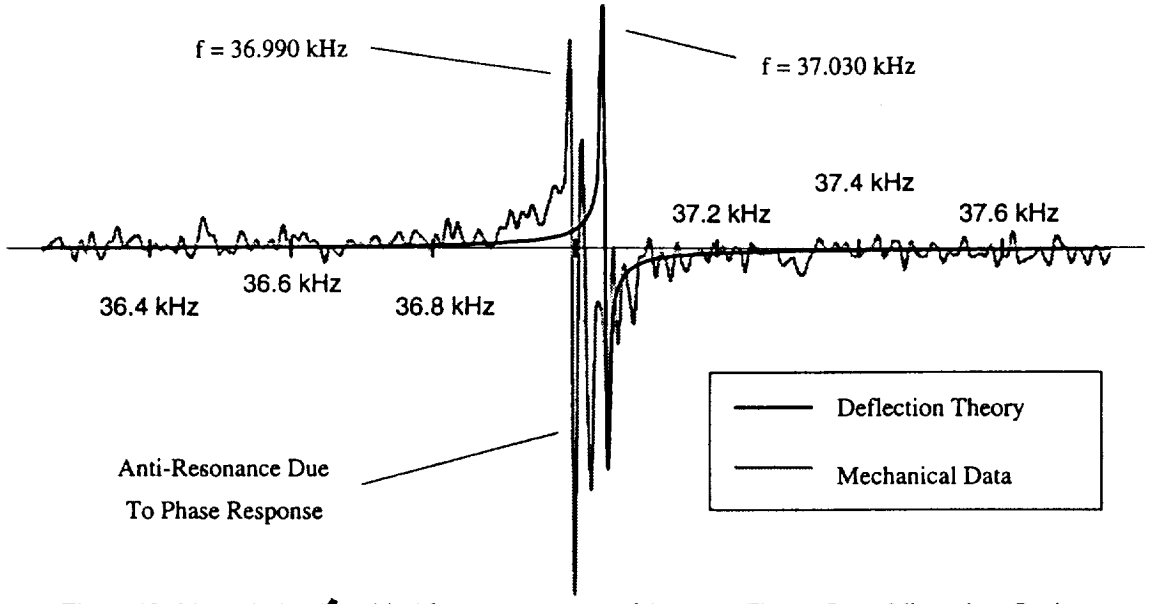


Figure 18. Theoretical and empirical frequency response of the square Flexure-Beam Micromirror Device.

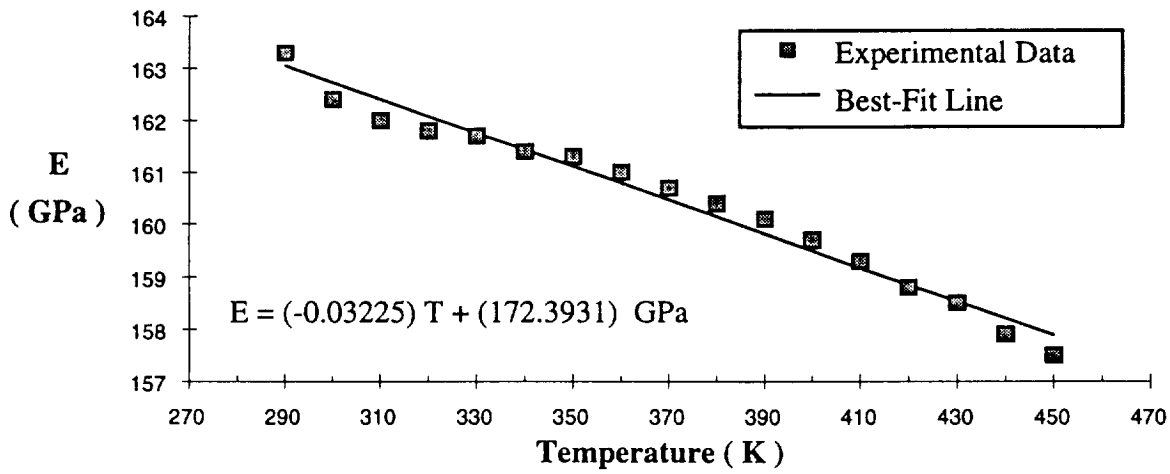


Figure 19. Elastic modulus as a function of temperature extracted from the square Flexure-Beam Micromirror Device.

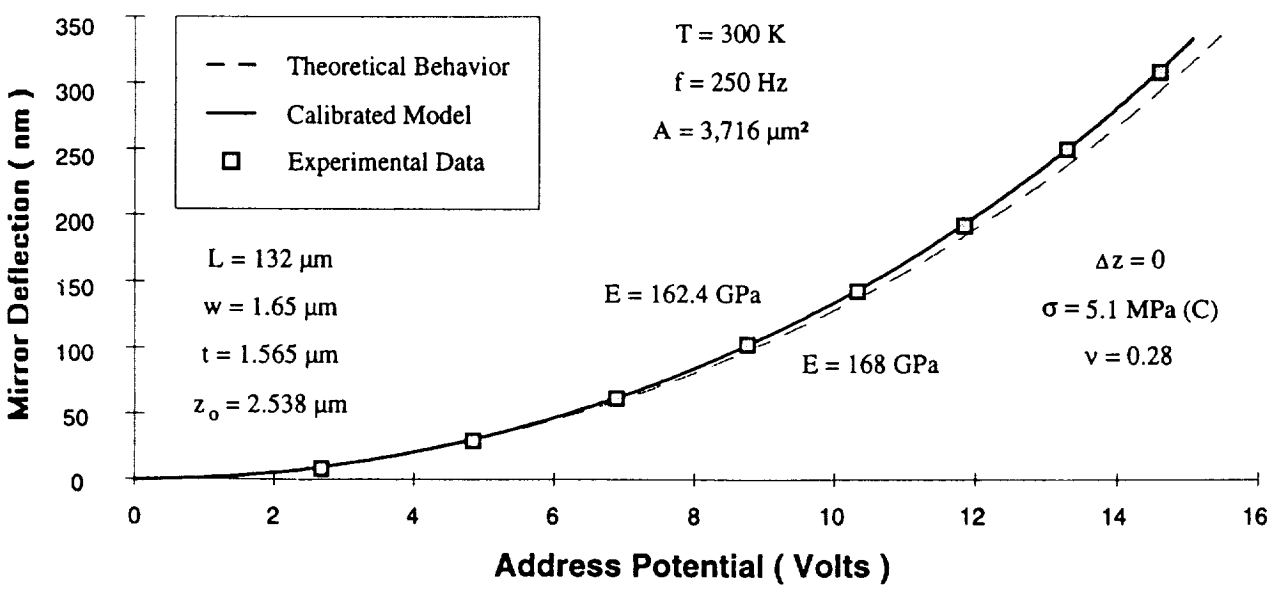


Figure 20. Theoretical and empirical characteristic behavior of the square Flexure-Beam Micromirror Device.

MONITORING COMPOSITE MATERIAL PRESSURE VESSELS WITH A FIBER-OPTIC/MICROELECTRONIC SENSOR SYSTEM

C. Klimcak, and B. Jaduszliwer
Electronics Technology Center
The Aerospace Corporation
Los Angeles, CA 90009-2957

71116
230640
6 P

ABSTRACT

We discuss the concept of an integrated, fiber-optic/microelectronic distributed sensor system that can monitor composite material pressure vessels for Air Force space systems to provide assessments of the overall health and integrity of the vessel throughout its entire operating history from birth to end-of-life. The fiber optic component would include either a semiconductor light emitting diode or diode laser and a multiplexed fiber optic sensing network incorporating Bragg grating sensors capable of detecting internal temperature and strain. The microelectronic components include a power source, a pulsed laser driver, time-domain data acquisition hardware, a microprocessor, a data storage device, and a communication interface. The sensing system would be incorporated within the composite during its manufacture. The microelectronic data acquisition and logging system would record the environmental conditions to which the vessel has been subjected during its storage and transit, e.g., the history of thermal excursions, pressure loading data, the occurrence of mechanical impacts, the presence of changing internal strain due to aging, delamination, material decomposition, etc. Data would be maintained in non-volatile memory for subsequent readout through a microcomputer interface.

INTRODUCTION

Fiber reinforced composites are the structural material of choice for fabricating high pressure vessels suitable for spacecraft applications. The superior specific properties of these materials greatly benefit systems for which weight minimization is a paramount concern. However, composite materials possess poor impact performance, a consequence of their limited ability to undergo plastic deformation. Much of the kinetic energy of an impacting mass is dissipated through the creation of large areas of fracture, resulting in a significant reduction in both the strength and stiffness of the composite (1). Even relatively low velocity impacts ($\cong 10\ ms^{-1}$) can cause delamination and matrix cracking with little or no evidence of exterior damage. In a documented instance a 50 % loss in the strength of a composite rocket motor casing resulted from an impact that produced damage at only the threshold of visibility (2). Rough handling at the factory or during transport are potential sources of damage. Furthermore, exposure of the vessel to excessive temperatures or internal pressure during testing and storage can also produce invisible damage that may result in a significant reduction in the burst strength of the vessel. To prevent failures, one could employ non-destructive testing methods to evaluate structural integrity by inspecting for damage just prior to deployment. Alternatively, one could maintain strict control and surveillance of the vessel at all times after manufacture, ensuring that all thermal, pressurization, and impact events are diligently logged for a subsequent assessment of structural integrity. Both of these options are costly, requiring access to either elaborate off-line instrumentation or the implementation of painstaking control and documentation procedures.

This paper discusses the concept of a fiber-optic/microelectronic sensor system for composite pressure vessels that will automatically detect and record the occurrence and magnitude of damaging environmental stresses suffered by the vessel. Fiber optic sensors would acquire data on both accidental

and intentional stresses experienced by the vessel during storage, transit, and operation. This data would be logged and evaluated with an on-board microprocessor that incorporated non-volatile storage for subsequent readout. This fiberoptic/microelectronic sensing system constitutes an application-specific integrated microinstrument (ASIM) that would enable the manufacture of smart composite pressure vessels that permit damage assessment via an on-demand capability of reporting critical life-stress history.

CONCEPT OVERVIEW

The environmental stress events that must be recorded are the occurrences of impacts, overpressurization, and exposure to high temperatures. These phenomena can be measured with a multiplexed array of optically-interrogated fiber optic strain and temperature sensors. A schematic illustration of the overall system is shown in Figure 1. It includes several fiber optic sensors attached to the external wall of the pressure vessel. The sensors are bonded at appropriate locations onto the surface of a composite vessel prior to adding the final protective overwrap. External attachment avoids perturbation of the composite matrix by the fiber optic waveguide and ensures that the sensors themselves do not compromise the intrinsic strength of the material. If necessary, the network can be mounted after the completion of the high temperature cure to preclude the possibility of thermal damage to the sensing network. The sensors are interconnected via standard fiber optic waveguide which is coupled to an electrooptic control unit via a fiber optic connector. The control unit contains an interrogating near infrared light source light source, a photodetector, and other optical and microelectronic components that are required for detecting and demultiplexing the signals arising from each of the sensors. A microprocessor manages data acquisition, demultiplexing, and the conversion of the individual sensor responses to localized strain and temperature data.

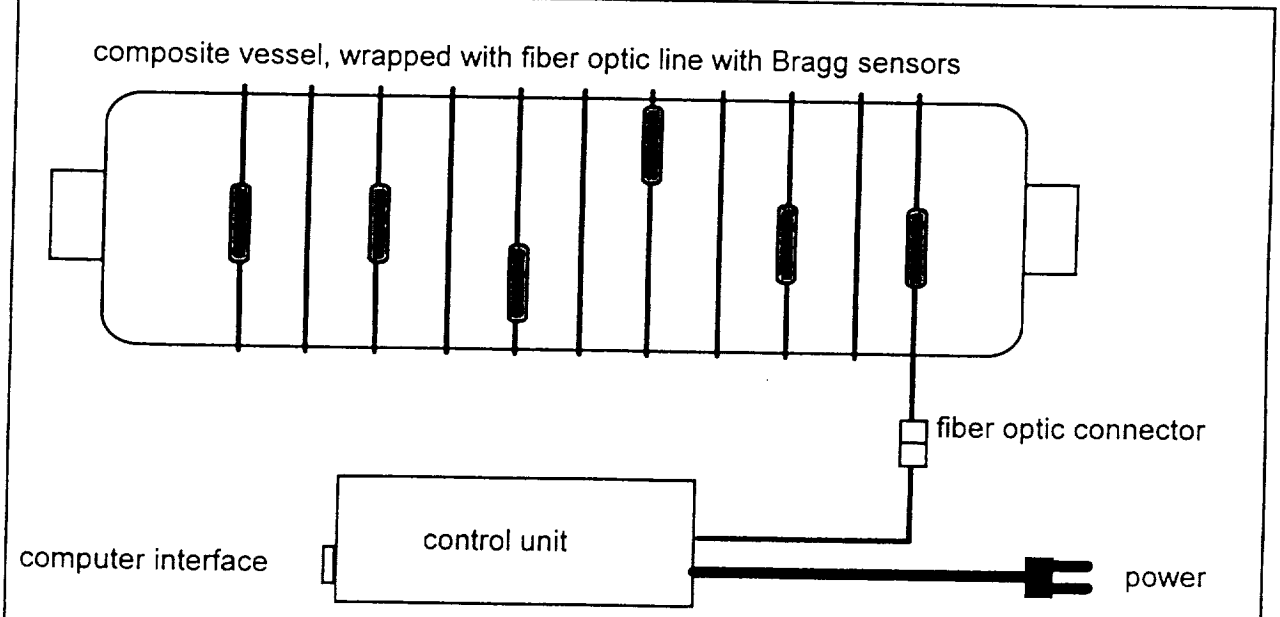


Figure 1: Conceptual Design of "Smart Composite Vessel"

This integrated sensor/microelectronic system would operate continuously and record data in non-volatile memory if preset thresholds of temperature or strain were exceeded. Data processing would be used to differentiate between static and dynamic strain effects. The DC response of a fiber optic strain sensor would be used to determine the steady-state strain resulting from pressurization while its AC

response would be used to detect vibration following a mechanical impact. In the latter case the transient ringing produced by the normal vibrations of the vessel would be detected by strain sensors positioned at the mechanical antinodes of the vessel. Wall temperature would be determined with a DC-coupled fiber optic temperature sensor.

Fiber Optic Sensing System:

Strain and temperature will be detected with fiber optic Bragg grating sensors (FOBGS). A FOBGS is a moderately short gauge length sensor (≅ 1 cm) fabricated within the core of photosensitive optical fiber by producing a periodic spatial modulation of the core index of refraction with a UV laser. Permanent modulation can be produced by either lithographic (3) or holographic exposure techniques (4). The resulting grating acts as a narrow-band reflector, with a central wavelength that is determined by the grating spacing. Since light transmitted by the grating will be missing those wavelength components that have been reflected, then the grating also behaves as a narrow bandwidth notch filter. Changes of strain or temperature modify that spacing to produce a detectable shift in the center wavelength of the reflection (notch). Transmitted signals arising from a FOBGS that is interrogated by a broadband LED light source are illustrated in Figure 2. The notch wavelength is directly proportional to the grating spacing and hence the absolute strain and temperature at the site of the grating. Sensing is accomplished by measuring the wavelength of the retroreflected light with a suitable spectral technique. Resolutions of 1°C (temperature) and 5 με (strain) are easily achievable. Differentiation between the effects of temperature and strain can be performed by encapsulating some sensors in strain-free manner within a housing so that these sensors respond only to temperature changes and are unaffected by strain. Alternatively, a dual-wavelength techniques employing superimposed grating sensors could be used (5).

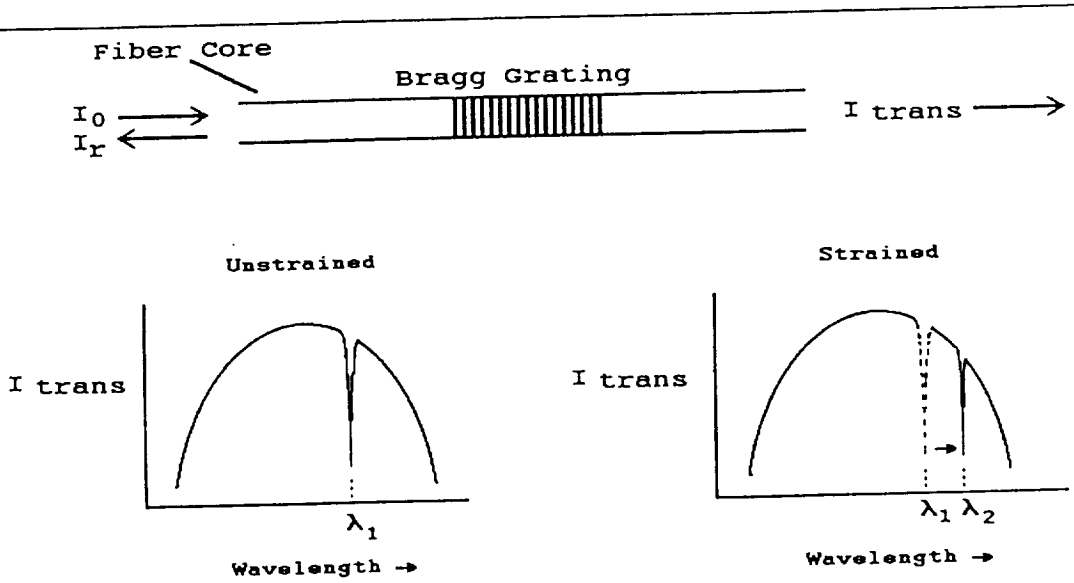


Figure 2: Depiction of the effect of strain on broadband light transmitted by a Bragg grating

Control Unit Subsytem:

Several Bragg gratings will be multiplexed on a fiber optic network that will be interrogated with a broadband NIR LED light source. The LED will be contained within an externally mounted electrooptic module that includes signal acquisition, demultiplexing, and all data processing, storage, and readout electronics. This module will interrogate the strain and temperature sensors and demultiplex the signals reflected by the Bragg sensor array. Numerous demultiplexing techniques exist that are based upon wavelength or combined time and wavelength decoding that permit extraction of the temperature and strain data from these reflected signals. In time-division demultiplexing, one utilizes pulsed interrogation of the sensor array and performs time-domain signal processing to infer sensor location from the arrival order of the reflected pulses. In wavelength-division demultiplexing, one utilizes spectral decoding of the returns from sensors that have been fabricated with unique central wavelengths that serve to identify the sensor location. Numerous wavelength demodulation options are available, including a fiber coupled CCD spectrometer, a fiber Fabry-Perot etalon, an acousto-optic tunable filter, and a matched Bragg grating demultiplexer array (6-8) . Combined time and wavelength division methods may be employed if it is necessary to increase the number of sensors beyond that which can be demultiplexed by wavelength alone.

The selection of a specific demultiplexing scheme is dependent upon many factors including: strain and temperature resolution, the number of sensors, response time, the mechanical and thermal behavior of the composite, cost, and tolerable system complexity. A detailed requirements analysis that considers all these factors must be performed prior to selecting the optimum scheme.

A signal processing module will separate and process the temperature, steady state strain, and vibration components of the electrooptic module output signal, determine if preset alarm thresholds have been exceeded and generate digital output for storage in non-volatile memory. A non-interruptible power supply will provide continuous power service during momentary power failure. A self-check module will be incorporated that generates system diagnostics for storage at a predetermined periodicity.

Development Roadmap:

Structural Analysis

As noted above, a thorough evaluation of the mechanical and thermal behavior of the composite is necessary for designing a suitable demultiplexing system. The normal modes of the vibrating vessel must be determined and the points of maximum vibration amplitude must be identified for several different types of excitation. A measurement of the relationship between the deposited shock energy and the resulting vibrational amplitude is necessary. If not already available, an alarm threshold for shock energy must be determined from a measurement of the damage induced at different shock energy levels. This will allow us to establish alarm thresholds for vibrational amplitude. Likewise, alarm thresholds for steady state strain and temperature must be determined from previously known data or new measurements.

Sensing and Control Unit Subsystems

The principle parameters of the sensing subsystem that must be determined are the number of strain and temperature sensors required to achieve our objectives, the interrogating light source, the design wavelengths for each of the multiplexed sensors, and the spacing between sensors. Several issues related to sensor attachment must also be addressed including determination of the manufacturing point at which the sensors will be incorporated within the vessel, the identification of bonding methods that assure good thermal and compliant mechanical coupling, selection of sensor protective overwrap, and the design of fiber ingress and egress ports.

The demultiplexing technique incorporated within the electrooptic module must be designed, fabricated and tested assuring that sensitivity, dynamic operating range, signal to noise ratio, and response bandwidths match system requirements. The signal processing module that receives analog input from the electrooptic module and provides digital output for storage must also be designed, fabricated, and tested. The diagnostic parameters of the self-check module must be defined and the module designed, fabricated, and tested. Finally, commercially available power supplies and non-volatile memories suitable for this application must be identified.

CONCLUSION

A conceptual solution for automatically monitoring and logging environmentally-induced overstress on composite high pressure vessels has been described. A roadmap for determining the system requirements has been developed that summarizes the mechanical properties that must be experimentally determined and the design issues that must be met to achieve this objective.

REFERENCES

1. W. J. Cantwell and J. Morton, "The impact resistance of composite materials - a review," Composites, 22, 347 (1991).

2. T. A. Solars, "Centaur G' Composite Adapter Stress Report: Damage Tolerance/Repair Programs," GDSS Report No. CG' CA-86-001, Jan 1986.

3. K. O. Hill, B. Malo, F. Bilodeau, D. C. Johnson, and J. Albert, "Bragg gratings fabricated in monomode photosensitive optical fiber by UV exposure through a phase mask," Appl. Phys. Lett., 62, 1035 (1993).

4. G. A. Ball, G. Meltz, and W. W. Morey,"Formation of Bragg gratings in optical fibers by a transverse holographic method," Opt. Lett., 14, 823 (1989).

5. M. G. Xu, J. L. Archambault, L. Reekie, and J. P. Dakin, "Discrimination between strain and temperature effects using dual wavelength fiber grating sensors," Electron. Lett. 30, 1085 (1994).

6. C. G. Askins, M. A. Putnam, and E. J. Friebele, "Instrumentation for interrogating many-element fiber Bragg grating arrays," Proc. SPIE Volume 2444, 257 (1995).

7. M. A. Davis, D. G. Bellemore, T. A. Berkoff, and A. D. Kersey, "Design and Performance of a Fiber Bragg Grating Distributed Strain Sensor System," Proc. SPIE Volume 2446, 227 (1995).

8. M. A. Davis and A. D. Kersey, "Serially Configured Matched Filter Interrogation Technique for Bragg Grating Arrays," Proc. SPIE Volume 2444, 295 (1995).

Micromechanical Switches on GaAs for Microwave Applications

71118

John N. Randall, Chuck Goldsmith, David Denniston, Tsen-Hwang Lin

Texas Instruments Incorporated
P.O. Box 655936
Mail Station 134
Dallas, Texas 75265
TEL (214) 995-2723
FAX (214) 995-2836
email: randall@resbld.csc.ti.com

Over the past few years, developments in microelectromechanical devices (MEMS) have enabled the exciting development of numerous microminiature sensors. These include accelerometers for automotive crash sensors as well as pressure sensors for medical blood pressure monitors. MEMS have also been used to build a variety of control device such as motors, spatial light modulators, and high-definition television displays. To date, there have been few efforts to exploit this technology for use in the design of microwave components. In 1979, Peterson [1] described the fabrication of low-frequency electrical switches using SiO_2 cantilevers. In 1991, Larson [2] described the operation of rotary mechanical microwave switches.

The switches discussed in this presentation are related to micromechanical membrane structures used to perform switching of optical signals on silicon substrates [3]. These switches use a thin metal membrane which is actuated by an electrostatic potential. Actuation of the membrane causes the switch to make or break contact. Selecting the proper materials and dimensions for the membrane and electrodes allows these devices to switch microwave signals with a low loss and reasonable switching voltages. Currently, the design of these devices centers on building capacitive microwave switches where the movement of the membrane can alter the capacitance by a factor of 100 or more. These devices exhibit high on-off impedance ratios with low parasitics at microwave frequencies. These devices are fabricated on gallium arsenide to be integrable with existing microwave amplifiers. The switch elements require only 4 mask levels with an additional one or two masks to form air bridge crossovers and gold plating to lower the RF impedance. These devices are expected to exhibit better than 100:1 impedance ratios, and switch microwave signals up to 18 GHz.

In this presentation, we will describe the fabrication process and could show video of operating switches. We will present RF data on simple switches and discuss the design of high performance multi-element switches. Advantages include superior isolation, high power handling capabilities, high radiation hardness, very low power operation, and the ability to integrate onto GaAs MMIC chips. Various applications of these devices will be discussed.

[1] K.E. Peterson, "Micromechanical Membrane Switches on Silicon," IBM J. Res. Develop., vol. 23, no. 4, pp. 376-385, July 1979.

[2] L.E. Larson, R.H. Hackett, M.A. Melendes, and R.F. Lohr, "Micromachined Microwave Actuator (MIMAC) Technology - A New Tuning Approach for Microwave Integrated Circuits," IEEE Microwave and Millimeter-Wave Monolithic Circuit Symp., pp. 27-30, 1991.

[3] G.M. Magel, T.H. Lin, L.Y. Pang, and W.R. Wu, "Integrated optic switches for phased-array applications based on electrostatic actuation of metallic membranes," SPIE Conf. on Optoelectronic Signal Processing for Phased-Array Antennas IV, vol. 2155, pp. 107-113, January 1994.

Poster not available at time of publication.

HIGH DENSITY MEMORY BASED ON QUANTUM DEVICE TECHNOLOGY

Paul van der Wagt, Gary Frazier, and Hao Tang

Texas Instruments Incorporated
P. O. Box 655936, MS 134
Corporate Research and Development
Dallas, TX 75265

tel (214) 995-6968, fax (214) 995-2836
wagt@resbld.csc.ti.com

71119
10P.
030614

ABSTRACT

We explore the feasibility of ultra-high density memory based on quantum devices. Starting from overall constraints on chip area, power consumption, access speed, noise margin, we deduce boundaries on single cell parameters such as required operating voltage and standby current. Next, the possible role of quantum devices is examined. Since the most mature quantum device, the resonant-tunneling diode (RTD), can easily be integrated vertically, it naturally leads to the issue of 3D integrated memory. We propose a novel method of addressing vertically integrated bistable two-terminal devices, such as resonant-tunneling diodes (RTDs) and Esaki diodes, that avoids individual physical contacts. The new concept has been demonstrated experimentally in memory cells of FETs and stacked RTDs.

1 Introduction

Progress in computer memory design has largely been evolutionary instead of revolutionary, because of the large number of constraints involved.[1] For 64 Gbit DRAM using planar technology the projected minimum feature size is 0.07 μm and a first shipment date of 2010.[2] To do better than this, it will be necessary to consider new approaches such as using quantum effect devices and 3D integration.

Memory is unique among computer building blocks in the sense that the basic memory cell does not necessarily require the use of transistors. Cells consisting of only resonant-tunneling diodes (RTDs) and single tunnel barrier diodes have already been demonstrated.[3] However, since a complementary "quantum device" logic does not presently exist (there is no fast pull-up device), surrounding logic should use CMOS technology to satisfy low power constraints.

In what follows, we concentrate on the basic memory cell and consider the application of tunneling-based devices to memory. Our experimental results have been obtained with III/V material RTDs, but the same ideas apply to Esaki diodes (and any other two-terminal bistable device), opening up a wide range of possible material systems, including Silicon.

2 Constraints

We separate constraints on a conceptually new memory system into four groups: area, power, speed, and noise margin. From these we infer limits on the design options.

The area constraint comes from considerations of wiring delays and yield problems. The 0.07 μm design rule of 2010, results in a 64 Gbit chip area of about 12.5 cm^2. For a 1 terabit chip the corresponding area would be around 200 cm^2, making it unwieldy. Achieving higher bit densities is possible using a 3D memory architecture where bit cells are stacked vertically on chip. However, full 3D processing, where each bit cell is physically contacted, must be avoided to minimize wiring complexity.[4] If we assume fabrication by epitaxy and standard processing, without regrowth steps, it follows that the basic memory cells must be vertically stacked two-terminal devices. We thus have the following logical chain:

3D => no wires to each device => stacked two-terminal devices.

Thus far, the planar (2D) approach has worked well enough that a solution of this 3D "wireless" addressing problem has not been required.

Power dissipation of large memories is entirely dominated by standby power, i.e. the additional energy expended during writing to or reading from the memory is negligible. The power per bit cell is given by the expression:

$$P = V_{cell} I_{st} , \qquad (1)$$

where V_{cell} is the (average) voltage drop over one bit cell and I_{st} is the standby current through the cell. Let C be the capacitance that is charged/discharged when the bistable cell switches between its two states

and let I_{sw} be the average current available from the cell during switching. Then a rough estimate for the switching time, t, of the cell is given by:

$$t\, I_{sw} = V_{cell}\, C = \text{charge moved during switching.} \tag{2}$$

Combining Eqs. (1, 2):

$$P\, t = C\, V_{cell}^2\, I_{st}/I_{sw}\ , \tag{3}$$

which is the "static power switching delay product" per bit cell.

It is clear from Eq. (3) that, for given C, V_{cell}, and I_{sw}/I_{st} ratio, standby power and switching speed can be traded off against each other. If we specialize to a bi-stable current device, such as an RTD or Esaki diode, then we can associate I_{sw}/I_{st} with the peak-to-valley current ratio (PVR),[5] and take C about 4 fF/μm^2 x A, where A is the device area (assuming about 250 Å separation between the cathode/anode charges). This leads to a single equation relating power, speed, and cell area with the voltage drop over the cell and PVR:

$$P\, t\, /\, A = 4 \cdot 10^{-15}\, V_{cell}^2\, /\, PVR\ . \tag{4}$$

A is the conducting cell area in μm^2. Figure 1 shows power per cell versus cell switching time for various values of V_{cell} and A and a conservative PVR of 20.

We did not include in C of Eq. (3) the data line capacitance. This capacitance may be larger than the device capacitance by orders of magnitude (few 100 fF versus < 1 fF). If an extremely high PVR (10^4) can be achieved for some two-terminal device, it may be possible to drive this capacitance directly with the cell switching current. However, just as is expected for future generations of DRAM cells,[6,7] a gain stage at the cell will likely be needed.

In addition to some of the noise sources present in high-density DRAMs,[1,8] we have to consider the probability of resetting a device with two current minima separated by a voltage DV. For thermal noise we find:

$$V_{noise} = (kT/C)^{1/2} = 2\, A^{-1/2}\ (\text{in mV})\ , \tag{5}$$

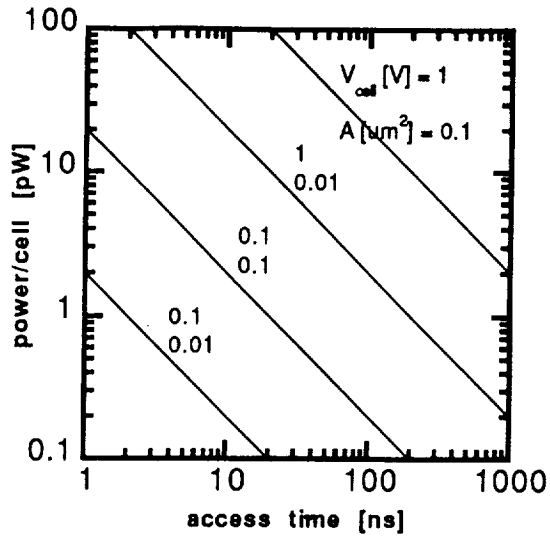


Figure 1. Static power dissipation vs. access speed for various memory cell voltages and cross sectional areas.

(room temperature) and for shot noise induced current fluctuations:

$$I_{shot} = (2eI_{st}B)^{1/2} = 0.06\ A^{-1/2}\ I_{st}\ , \tag{6}$$

where B is the noise frequency bandwidth. The effects of each of these sources is negligible as long as A $\geq 0.01\ \mu m^2$ and $\Delta V \geq 0.1$ V, consistent with earlier analyses.[9]

3 Design options

Quantum mechanical tunneling is responsible for the nontrivial current-voltage characteristics of both the RTD and the Esaki diode. As pointed out above, a practical 3D integrated memory implies the use of stacked two-terminal devices (although addressing these has been left open for the moment). Another big advantage of using two-terminal devices as bit cells is their potential for very low voltage operation. It is well-known that reduction of the CMOS circuit operating voltage below 1 V poses serious problems because any lowering of the threshold voltage is accompanied by an exponential rise in subthreshold current (at $V_{GS} = 0$ V).[10] However, for RTDs and Esaki diodes, two current minima can occur much closer than 1 V and a separation as low as 0.1 V should be realizable. Although surrounding circuitry would still use standard voltage levels, the product of cell power and access time decreases by a factor 100 if the cell voltage drops from 1 to 0.1 V.

Let us assume then that $V_{cell} = 0.1$ V, $A = 0.01\ \mu m^2$ and PVR = 20 in Eq. (4). This yields:

$$P\ [pW/cell]\ \tau[ns] = 20 \tag{7}$$

This equation shows that a terabit memory would be feasible with total power in the 1 W range and access times below 100 ns.

If the power per bit cell is indeed 1 pW, then it follows from Eq. (1) that the standby current density should be about 0.1 A/cm^2. If using bistable diodes, their valley current density will have to be of the same order of magnitude (or below). These numbers are about 4 orders of magnitude lower than standard values. Figure 2 shows the I-V characteristic of an ultra-low current RTD that we designed and fabricated. The valley current density is 1 A/cm^2 and the PVR is about 7. There is no fundamental problem to lower this current density even further and we fully expect to improve the PVR with future designs.

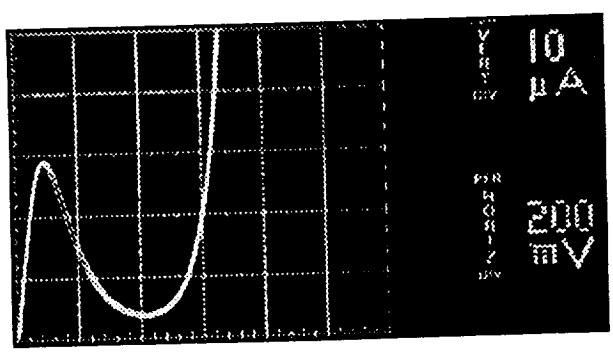


Figure 2. I-V curve of 20x20 mm^2 low-current RTD.

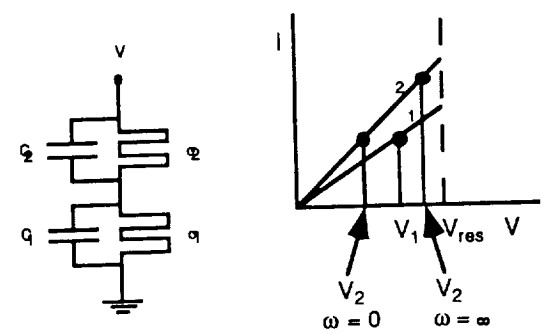


Figure 3. Lumped element model of RTDs and V_2 vs. V_1 relation for low and high frequencies.

4 Slew Rate Addressing

The outstanding issue is now whether it is possible to access bits stored in vertically stacked bi-stable diodes without physically contacting these devices. If a voltage ramp is applied over two RTDs (or Esaki diodes) in series, one RTD will switch prior to the other. This event can be analyzed with a lumped element RTD model[11] taking only intrinsic parallel capacitance into account, as is shown in Fig. 3. Let s_n be the conductance associated with the pre-peak I-V slope and C_n the capacitance of RTD n.

A sinusoidal applied voltage with angular frequency w produces voltages V_1 and V_2 over each RTDs 1 and 2 with a ratio

$$V_1 / V_2 = (s_2 + jw\, C_2) / (s_1 + jw\, C_1) \ . \tag{8}$$

When the real part of one of these voltages reaches the corresponding RTD peak voltage, a switching event takes place (and the current linear analysis breaks down). Now let the peak currents and capacitances satisfy $I_1 < I_2$ *and* $C_1 > C_2$. Then for low frequencies RTD_1 switches first, while for high frequencies RTD_2 switches first. Another way of understanding this behavior is to think of the parallel RTD capacitance as shunting away current that would otherwise have been available for conduction through the RTD, and that this happens more for RTD_1 than for RTD_2 at higher frequencies.

Figure 4 shows a simulation of this switching order reversal. In this case, the relevant slew rate lies below the intrinsic RTD slew rate, which is just the RTD speed index (I/C). Therefore, when the first switching RTD goes beyond its peak, the voltage over it increases faster than the overall ramp voltage and the voltage over the other RTD must decrease. If the relevant slew rate would lie much above this, both RTDs would "switch" (and one of them would "switch back" if the applied voltage was suddenly held constant).

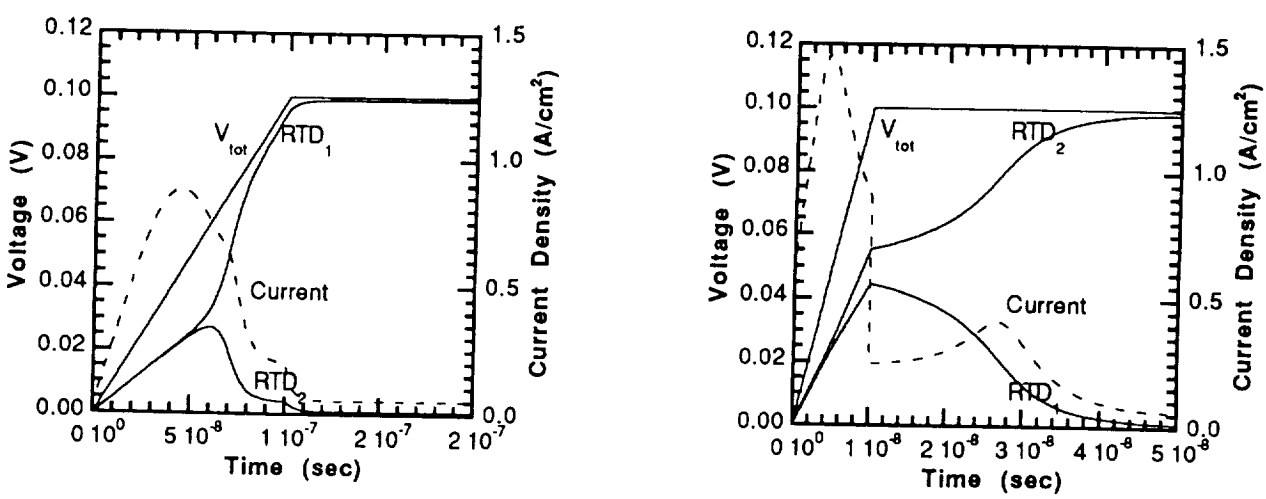


Figure 4. Simulated RTD stack switching under applied voltages with different slew rates. The slew rate is 10x higher on the righthand side (note different time scales).

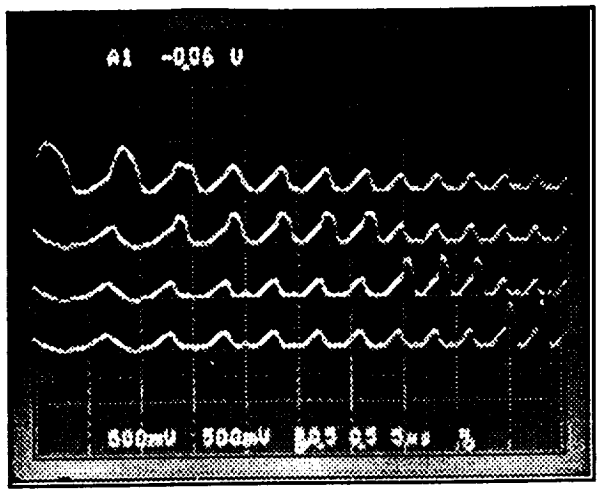


Figure 5. Voltages of four series connected RTDs under an applied frequency sweep from 100 to 340 kHz.

These ideas generalize to multiple RTDs connected in series. Figure 5 shows the voltages over four RTDs connected in series. We used standard high-current RTDs and slowed these down by adding external capacitors, so that slew rate dependent "addressing" of the individual RTDs can be observed in the few hundred kHz range. Peak currents form an ascending and capacitances a descending series as a function of RTD number. Using TI SPICE augmented with RTD models, we have been able to accurately reproduce these experimental results. These results show that by using more information of the applied signal than just its final levels, one can "address" individual stacked RTDs without extra contacts.

Since propagating slew rate information over word lines to a memory cell may pose problems, we have looked at a model stacked memory cell for which the desired slew rates are locally generated from a multiple-valued word line level. The schematic circuit is shown in Figure 6. The pass FET acts as a variable resistor, and together with the capacitances of the RTDs forms an adjustable low frequency filter (the pass capacitor merely provides DC isolation between the top of the RTD stack and the bit line). After the word line voltage "opens" the pass FET to some degree, a positive step voltage step on the bit line generates a corresponding upward ramp on the "storage node," which in turn switches the desired RTD from a low voltage to a high voltage state. SPICE simulations confirm the proper operation of this circuit over full write cycles without unwanted "resets" when the word and bit line return to their original values.

Fig. 7 shows experimental results obtained for the case of two series connected RTDs. As in the case of Fig. 5, external capacitances were added to standard current density RTDs. Word level 1 (- 0.8 V) selects RTD1 (when bit line goes high), while word level 2 (0.0 V) selects RTD2. In order to be able to

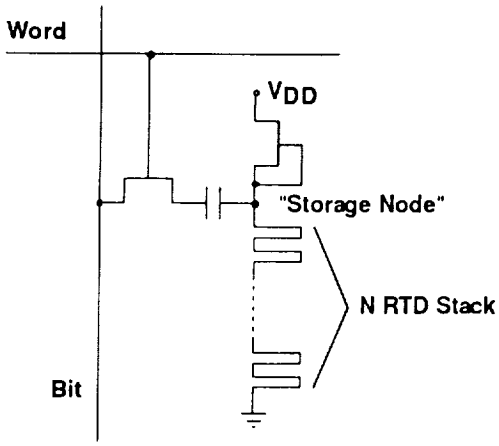


Figure 6. Multiple-valued Word Line Memory Cell.

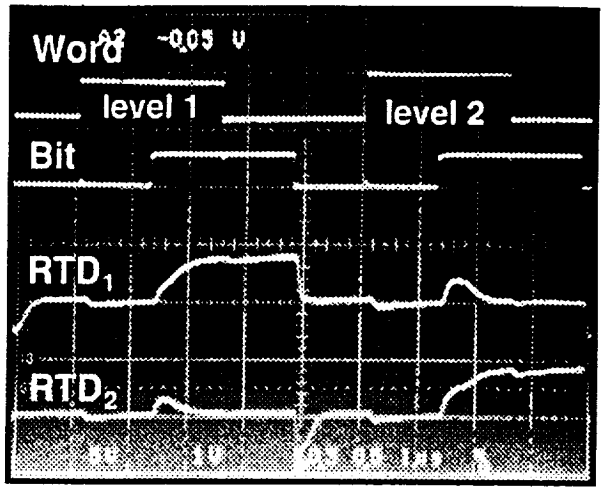


Figure 7. Write Operation for 2 RTD Stack. Vertical: 5, 1, 0.5, and 0.5 V/div, Horizontal: 1 ms/div.

display this process on an oscilloscope, the entire RTD stack was reset at the negative edge of the bit pulse.

The above results were obtained by adding external capacitances to standard RTDs. However, we have also observed the switching order reversal with ultra-low current RTDs *without* any added capacitance, justifying our more convenient testing of the slew rate based addressing concept with externally added capacitances.

Finally, we should remark that reading the data from an RTD stack is more complex than writing to it. We are currently investigating circuit topologies for fast readout of all bits stored in an RTD stack.

5 Conclusion

We have discussed a possible rôle for quantum effect devices in high density memory. Two-terminal bistable diodes, such as the RTD and the Esaki diode, may enable a practical form of 3D integration. However, this would require a "contactless" method of selecting vertically stacked bit cells. We have developed and experimentally verified such a contactless scheme utilizing the slew rate of an addressing signal.

Acknowledgments

The authors would like to thank A. Seabaugh for supplying many of the devices for this research and P. Stickney, R. Aldert for excellent technical assistance.

References

[1] K. Itoh, IEEE J. Solid-State Circuits **25** (1990) 778

[2] The National Technology Roadmap for Semiconductors (Semiconductor Industry Association, San Jose, CA, 1994) p. 11, and Sematech Communiqué Vol. 6, Number 1, Jan/Feb 1995

[3] T. Mori et al., Ext. Abs. Intl. Conf. SSDM (1993) 1074; N. Yokoyama, S. Muto, K. Imamura, M. Takatsu, T. Mori, Y. Sugiyama, Y. Sakuma, H. Nakao, and T. Adachihara, Ext. Abs. 2nd Int. Workshop on Quantum Functional Devices (1995) 40

[4] Very High Speed MOS Devices, Ed. S. Kohyama (Oxford University Press, 1990) Sec. 9.4

[5] For RTD physics overview, see H. C. Liu and T. C. L. G. Sollner, High-Frequency Resonant-Tunneling Devices, in High Speed Heterostructure Devices, Eds. R. A. Kiehl and T. C. L. G. Sollner (Academic Press, Boston, 1994) p. 359; or E. R. Brown, High-Speed Resonant-Tunneling Diodes, in Heterostructures and Quantum Devices, Eds. N. G. Einspruch and W. R. Frensley (Academic Press, San Diego, 1994) p. 305

[6] M. Terauchi, A. Nitayama, F. Horiguchi, and F. Masuoka, VLSI Symp. Tech. Dig. (1993) 21

[7] W. Kim, IEEE J. Solid-State Circuits **29** (1994) 978

[8] Y. Nakagome , M. Aoki, S. Ikenaga, M. Horiguchi, S. Kimura, Y. Kawamoto, and K. Itoh, IEEE J. Solid-State Circuits **23** (1988) 1120; H. Masuda, R. Hori, Y. Kamigaki, K. Itoh, H. Kawamoto, and H. Katto, IEEE J. Solid-State Circuits **15** (1980) 846

[9] R. Landauer, J. Appl. Phys. **33** (1962) 2209

[10] T. Sakata, IEEE J. Solid-State Circuits **29** (1994) 887

[11] B. G. Park, E. Wolak, K. L. Lear, and J. S. Harris, Jr. , IEEE IEDM Tech. Dig. (1989) 563

7123

InGaAlAsPN: A MATERIALS SYSTEM FOR SILICON BASED OPTOELECTRONICS AND HETEROSTRUCTURE DEVICE TECHNOLOGIES

T. P. E. Broekaert*, S. Tang, R. M. Wallace, E. A. Beam III, W. M. Duncan, Y.-C. Kao, H.-Y. Liu
*System Components Laboratory
and
Materials Science Laboratory
IS&S Corporate R&D, Texas Instruments, Dallas, TX 75265

8 p.

ABSTRACT

A new material system is proposed for silicon based optoelectronic and heterostructure devices: the silicon lattice matched compositions of the (In,Ga,Al)-(As,P)N III-V compounds. In this nitride alloy material system the bandgap is expected to be direct at the silicon lattice-matched compositions with a bandgap range most likely to be in the infrared to visible. At lattice constants ranging between those of silicon carbide and silicon, a wider bandgap range is expected to be available and the high quality material obtained through lattice matching could enable applications such as monolithic color displays, high efficiency multi-junction solar cells, optoelectronic integrated circuits for fiber communications, and transfer of existing III-V technology to silicon.

1. Introduction

Recently, the growth of the $GaAs_xN_{1-x}$[1,2,3,4,5] compound semiconductors has been gaining interest. This increased interest in nitride based semiconductors is primarily related to the potential for short wavelength (blue and violet) emitting devices[6,7] and the capability of the arsenide nitride compound semiconductors of being lattice-matched to silicon[3]. This combination of direct bandgap materials and the potential for high quality lattice-matched crystalline materials on a silicon substrate is heretofore unrealized and is ideal for silicon-based optoelectronics[8]. However, the applications for this material system are potentially much broader, depending on the range of bandgaps that will be available. The ternaries of interest that lattice match to silicon are, assuming that Vegard's law holds for the III-V N alloys, $AlAs_{0.82}N_{0.18}$, $GaAs_{0.81}N_{0.19}$, $InAs_{0.41}N_{0.59}$, $InP_{0.5}N_{0.5}$. Thus, the silicon based ternaries all require large amounts of nitrogen to be incorporated. The binaries GaP and AlP nearly lattice match to silicon but are of limited interest for electro-optic devices owing to their indirect bandgap. However, strained pseudomorphic layers of GaP_xN_{1-x} and AlP_xN_{1-x} may be of use in devices with nanometer scale layer thicknesses.

2. Challenges and opportunities

The challenge in growing the group III - V nitride alloy compounds is due to the large difference in the atomic radii of nitrogen and the other group V elements. Such material systems with large differences in atomic radii are known to have large miscibility gaps and cannot be grown by equilibrium techniques[9]. However, immiscible material systems can be grown by non-equilibrium techniques

such as MBE. Indeed, immiscible materials with large microscopic strain such as $GaP_{1-x}Sb_x$[10] and $(GaAs)_{1-x}Si_x$[11] have already been grown successfully. Two questions need to be answered to determine the potential of these heterojunction materials : (1) Are any of the large-nitrogen-content alloys (meta-) stable? and (2) what bandgap range is possible?

Concerning question (1), the high microscopic strain in the III-V N alloys is expected to lead to ordering[12] which is known to stabilize the epitaxial layers[13] and to change the bandgap as compared to the bandgap of the random alloy[14]. For x=0.5 alloys, the chalcopyrite ordering typically results in an increased bandgap, while CuAu or CuPt type ordering typically lowers the bandgap as compared to that of the random alloy[14]. Epitaxial stabilization in ordered compounds is strongest for the stoichiometric compound[13], x=0.5, with decreasing stability towards the constituents. Thus, $InP_{0.5}N_{0.5}$ may form a stable ordered chalcopyrite compound, while decreasing epitaxial stability against decomposition is expected for $InAs_{0.41}N_{0.59}$, $GaAs_{0.81}N_{0.19}$ and $AlAs_{0.82}N_{0.18}$. As an example, in the case of $InAs_{0.41}N_{0.59}$ the ordering may result in $InAs_{0.41}N_{0.59}$ being an alloy of the chalcopyrite-like In_2AsN and the famatinite-like In_4AsN_3, or as an $In_5As_2N_3$ ordered compound, rather than a random alloy of InAs and InN. For quaternary alloys such as the AlGaAsN compounds on silicon an ordered group V sublattice would be expected, while the group III sublattice can be random owing to the similar radii of gallium and aluminum.

The second question is that of the bandgap range available in the III-V N compounds. Munich and Pierret first calculated the bandgaps for AlGaAsN compounds[15], in particular, for the silicon matched compounds they find a direct-bandgap range of about 1.7 eV to 3.0 eV. However, they did not take into account the extrinsic bowing as described by Van Vechten and Bergstresser[16]. On the other hand, Sakai[6] *et al.* by taking into account the extrinsic bowing predict a semimetallic nature (zero bandgap) over a wide range of the (InAlGa)(SbAsP)N compounds. The "negative" bandgaps predicted correspond to a structure that has zero bandgap between the main conduction and valence bands but that has negative E_0, *i.e.* an inverted band structure as in $Hg_xCd_{1-x}Te$[17]. However, the extrapolations used in Ref. 6 to calculate the bandgap at high alloy concentrations, *i.e.* $x \approx 0.5$, may be beyond the validity of the model used, as described in Ref. 16, since the perturbation potential caused by the alloying has become sufficiently large to affect the character of the valence and conduction bands. An *ab initio* calculation is needed to obtain a more reliable theoretical estimate. Recently Rubio and Cohen[18] calculated a finite bandgap of 0.7 eV for $GaAs_{0.75}N_{0.25}$ using the local density approximation (LDA) without, however, internal relaxation of the crystal structure. Neugebauer and Van de Walle[19] calculated the bandstructure for a (111) superlattice of $GaAs_{0.5}N_{0.5}$ with internal relaxation and found a zero bandgap mettallic structure. Recent results obtained by Tang[20] *et. al.* at Texas Instruments using LDA with internal relaxation have shown that some of the ordered structures of $GaAs_{0.5}N_{0.5}$ and $GaAs_{0.75}N_{0.25}$ have a finite bandgap.

Preliminary experimental bandgap results at large nitrogen content are so far inconclusive[3]. The bandgap range available in the nitride alloy system is thus still open to debate. An estimate can be made for the random alloys using the recently obtained experimental values for the bowing coefficient in $GaAs_xN_{1-x}$[1,5] (18 eV) and GaP_xN_{1-x}[21] (14 eV). As pointed out in the previous section the bandgap values obtained at high alloy concentrations using this method may not be accurate. Figure

1 shows the estimated bandgap versus lattice constant for the ternaries of the InGaAlAsPN material system, using a bowing coefficient of 14 eV for the phosphide nitrides and 18 eV for the arsenide nitrides. The solid lines represent the Γ-point bandgap while the dashed lines represent the X-point bandgap. Also shown in figure 1 is the estimated bandgap for the Famatinite ordered $GaAs_{0.75}N_{0.25}$. Although the bandgap range that would be available on silicon is still uncertain, the early indications are that the range will be in the infrared to red part of the spectrum, enabling such applications as 1.55 μm detectors and lasers, and devices such as FETs and HBTs using the low bandgap AlGaAsN quaternaries. At lattice constants smaller than that of silicon, alloys covering the entire visible bandgap range might be feasible, enabling applications such as color displays and high efficiency multijunction solar cells.

As a result of the strong bowing in the III-V N alloys which shifts the bandgap towards the infrared, the applications requiring a red to blue wavelength bandgap range such as color displays and high efficiency multijunction solar cells may need to be grown on smaller lattice constant materials, *e.g.* Si_xC_{1-x} with x = 0.75 to x = 0.60 as most likely.

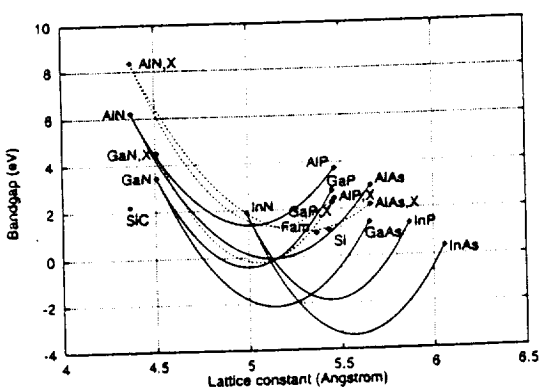


Figure 1. Bandgap vs. lattice constant in the InGaAlAsPN material system. The solid lines represent empirical extrapolation from low nitrogen content alloys for the direct bandgap, while the dashed lines represent the X-point. The point labeled Fam. is the ab-initio theoretical prediction for the bandgap of Famatinite $GaAs_{0.75}N_{0.25}$.

Therefore, the growth of high quality lattice-matched materials covering a bandgap range as wide as the entire visible range on a single substrate may be achieved, making possible the monolithic integration of full color displays and detectors. As with other material systems quaternaries can be used for bandgap engineering and device optimization. An example of the versatility possible with a wide available bandgap range is shown in figure 2: by making use of the optical low pass nature of semiconductors, a true multi-color pixel can be achieved by vertically stacking the red, green and blue pixel of a color display. The vertical stacking results in a pixel with a true color appearance irrespective of viewing angle and distance. Though the bandgap range on silicon is possibly more limited, several heterostructure devices are possible using AlGaAsN quaternaries. For example a resonant tunneling diode could be made using AlAsN as the barrier material and GaAsN as the well and contact material. Similarly, FETs and HBTs can be made using the AlGaAsN quaternaries. In addition, the piezoelectric properties of III-V semiconductors in general and of the nitride alloys in specific would

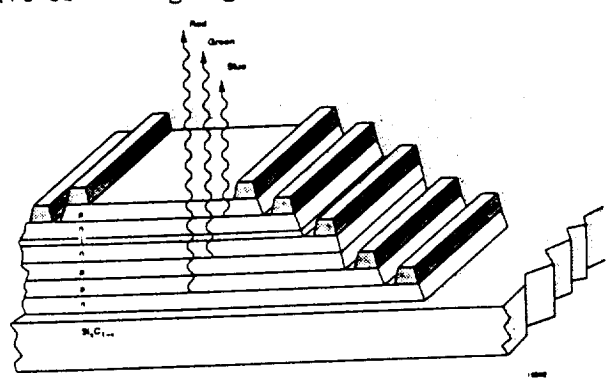


Figure 2. A true color pixel as an example of one of the new applications possible with the InGaAlAsPN material system.

allow the integration of surface acoustic wave devices on silicon.

Finally, it should be noted that due to the uncertainty of theoretical predictions of bandgaps that the actual bandgap range of the InGaAlAsPN compounds will not be accurately known until it is experimentally determined.

3. Experimental results

Several methods have been used in this study for the growth of $GaAs_{1-x}N_x$ alloys. The first method used in this study is MOMBE: a metalorganic source is used for the arsenic (TBA or TDMAAs) and a solid source is used for the group III elements; an ECR nitrogen source is used for the nitrogen supply. A number of varying nitrogen flow rates, ECR power and substrate temperature were used for the growth. The initial choice of substrates in this study is GaAs to avoid the antiphase domain boundary defects that would be associated with growth on silicon. All films had a rough surface with a hazy to orange peel look. At high substrate temperatures, 600 °C, the epitaxial material formed is a polycrystalline cubic GaN film. For one of the growths, at low substrate temperature, 500 °C, a polycrystalline film is formed which has been identified as having two major components: a $GaAs_{1-x}N_x$ compound with a nitrogen content of about 6.5% and a second compound Ga_3As_2N , i.e. with 33 % nitrogen. Figure 3 shows the thin film X-ray diffraction pattern for the polycrystalline film containing Ga_3As_2N [22]. Three of the intensity X-ray peaks are identified as (111), (220), and (400) reflections of $GaAs_{1-x}N_x$ with x = 6.54%. Two of the other peaks are identified as (111) and (224) reflections of $GaAs_{1-x}N_x$ with x = 33.75%. This last result of a nitrogen fraction of 33.75% is interesting since it indicates that high nitrogen fraction compounds can be formed. Moreover, the close to 2 to 1 ratio of As to N suggests that this material system may be likely to form ordered compounds at high nitrogen fractions.

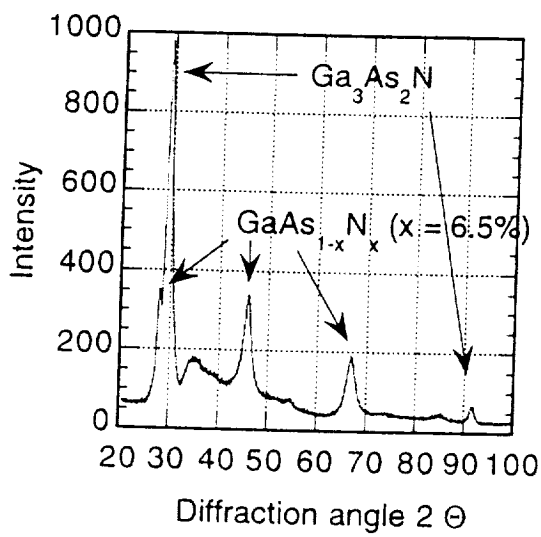


Figure 3. Thin film X-ray diffraction of a polycrystalline sample grown by MOMBE.

The second method used to grow the $GaAs_{1-x}N_x$ alloys in this study is conventional MBE: solid sources are used for both group III and group V elements, except for the nitrogen which is again provided for by an ECR nitrogen source. At small nitrogen fractions up to 1.5% it is possible to obtain a single crystal epitaxial film and the bandgap of the material has a red shift with increasing nitrogen fraction with a bowing coefficient of about 18 eV as has been previously obtained in other studies[1,2]. Figure 4 shows the bandgap dependence of $GaAs_xN_{1-x}$ at small nitrogen fractions[23]. Also shown in the same figure is the empirical bandgap dependence assuming an 18 eV bowing coefficient.

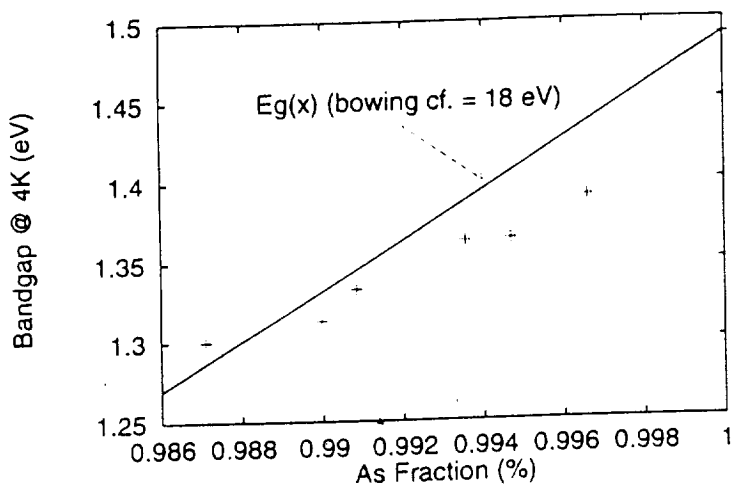


Figure 4. Bandgap of $GaAs_xN_{1-x}$ as a function of As fraction as measured by PL at 4 K.

For films with greater than 1.5% nitrogen no single crystal epitaxial film has been obtained to date. Again, a polycrystalline film is obtained for the higher nitrogen fraction films. Figure 5 shows the thin film diffraction pattern of such a polycrystalline film, one grown on silicon and the other on GaAs. The large peak at 34.5 degree is associated with (111) cubic GaN or $GaAs_xN_{1-x}$ with a small As fraction. Three peaks are identified as $GaAs_{1-x}N_x$: the peaks near 28, 46, and 54 degree are associated with (111), (220), and (311) crystal plane reflections. Assuming Vegard's law, and without correcting for strain deformation, the nitrogen composition associated with each peak is 25.5%, 2.8%, and 5.8% for the GaAs substrate based sample and 11.7%, 9.7%, and 9.7% for the silicon substrate based sample respectively. Due to the polycrystalline nature of the film it is not possible to determine the amount of strain deformation, and therefore also the true nitrogen content, in each of the differently oriented grains. From the silicon substrate based sample one can conlude that the nitrogen content is at least 10%.

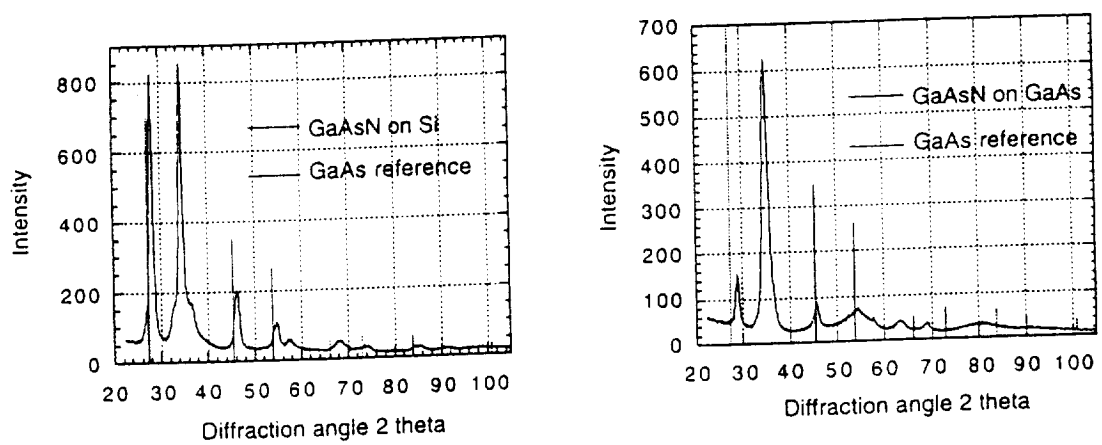


Figure 5. Thin film X-ray diffraction of GaAsN grown on silicon (shown on the left) and on GaAs (shown on the right) in comparison to a GaAs reference diffraction pattern.

Thus far we have not been successful at detecting any photoluminescence from the high nitrogen fraction compounds observed here, which leaves open the question of the electronic properties of the high nitrogen fraction arsenic nitride alloys.

4. Conclusions

In conclusion, InGaAlAsPN technology may impact all of the existing III-V technology. Silicon-based epitaxy will enable III-V technology to keep in step with Si technology and wafer size, resulting in a significant reduction in cost of III-V technology. Reliable and high yield monolithic cointegration of lasers, optical detectors, surface acoustic wave filters, microwave circuits and VLSI circuits on a single chip may be possible for the first time, thereby drastically reducing system size and increasing system reliability. To achieve the full potential of this material system further growth studies of the nitride and arsenide nitride materials are needed. The authors would like to acknowledge stimulating discussions with Drs. G. Frazier, T. Moise, J. Randall and A. Seabaugh. The outstanding technical assistance of D. Chasse, A. Fowler, F. Stovall, R. Thomason, and K. Vargason is much appreciated.

References

[1] M. Weyers, M. Sato, H. Ando, Jpn. J. Appl. Phys. **31**, L854 (1992).
[2] M. Weyers and M. Sato, Appl. Phys. Lett. **62**, 1396 (1993).
[3] C. T. Foxon, T. S. Cheng, D. E. Lacklison, S. V. Novikov, D. Johnson, N. Baba-Ali, T. L. Tansley, S. Hooper, and L. J. Challis, 1994 Electronic Mat. Conf. Proc. A35.
[4] S. Bharatan, K. S. Jones, C. R. Abernathy, S. J. Pearton, F. Ren, P. W. Wisk and J. R. Lothian, J. Vac. Sci. and Tech. **A 12**, 1094 (1994).
[5] M. Kondow, K. Uomi, K. Hosomi, T. Mozume, Jpn. J. Appl. Phys. **33**, L1056 (1994).
[6] S. Sakai, Y. Ueta, and Y. Terauchi, Jpn. J. Appl. Phys. **32**, 4413 (1994).
[7] For a review see H. Morkoc, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, and M. Burns, J. Appl. Phys. **76**, 1363 (1994).
[8] R. A. Soref, Proc. of the IEEE **81**, 1687 (1993).
[9] G. B. Stringfellow, J. of Cryst. Growth **58**, 194 (1982).
[10] M. J. Jou, Y. T. Cherng, H. R. Jen, G. B. Stringfellow, Appl. Phys. Lett. **52**, 549 (1988).
[11] J. E. Greene, J. of Vac. Sci. and Tech. **B 1**, 229 (1983).
[12] A. A. Mbaye, A. Zunger, D. M. Wood, Appl. Phys. Lett. **49**, 782 (1986).
[13] C. P. Flynn, Phys. Rev. Lett. **57**, 599 (1986).
[14] S.-H. Wei and A. Zunger, Appl. Phys. Lett. **56**, 662, 1990.
[15] D. P. Munich and R. F. Pierret, Solid St. Electron. **30**, 901 (1987).
[16] J. A. Van Vechten and T. K. Bergstresser, Phys. Rev. **B 1**, 3351 (1970).
[17] R. Dornhaus, G. Nimtz, and B. Schlicht, ``Narrow gap semiconductors", Springer-Verlag, New York (1985).
[18] A. Rubio and M. L. Cohen, Phys. Rev. **B 51**, 4343 (1995).
[19] J. Neugebauer and C. G. Van de Walle, Phys. Rev. **B 51**, 10568 (1995).
[20] S. Tang, T. P. E. Broekaert, and R. M. Wallace, to be published.
[21] J. N. Baillargeon, K. Y. Cheng, G. E. Holfer, P. J. Pearah, K. C. Hsieh, Appl. Phys. Lett. **60**, 2540 (1992).
[22] E. A. Beam III, H.-Y. Liu, and T. P. E. Broekaert, unpublished results.
[23] Y.-C. Kao, W. M. Duncan, H.-Y. Liu, and T. P. E. Broekaert, unpublished results.

Session 3: Posters and Demonstrations (continued)

The Fundamentals of Using the Digital Micromirror Device (DMD™) for Projection Display
by Lars A. Yoder, Texas Instruments

Aercam Demonstration
by Charles Price, NASA/JSC

MEMS Spaceborne Testbed
by Robert H. Stroud, The Aerospace Corporation, and
Mark Holderman, NASA/JSC

Low Volume Packaging for a Microinstrumentation System
by Andrew Masonand Ken Wise, University of Michigan

Pyroelectric Applications of the VDF-TrFE Copolymer
by J.J. Simonne, Laboratoire d'Analyse et d'Architecture des Systèmes,
Ph. Bauer, SOFRADIR, L. Audaire, CEA/DTA/LETI/DOPTS/S-LIR,
and F. Bauer, Institut de Recherche Franco-Allemand de St Louis

Stereo-Based Region-Growing using String Matching
by Robert Mandelbaum and Max Mintz, University of Pennsylvania

The Fundamentals of Using the Digital Micromirror Device (DMD™) for Projection Display

(Revised 10/20/95)

Presented at:

The International Conference on Integrated Micro/Nanotechnology for Space Applications
Houston, TX October, 1995

Lars A. Yoder
Texas Instruments Inc.
Digital Video Products
P.O. Box 655474, MS 429
Dallas, Texas 75265
214-995-2377

ABSTRACT

Developed by Texas Instruments, the Digital Micromirror Device (DMD™) is a quickly emerging and highly useful Micro-Electro-Mechanical Structures (MEMS) device (See Figure 1). Using standard semiconductor fabrication technology, the DMD's simplicity in concept and design will provide advantageous solutions for many different applications. At the rudimentary level, the DMD is a precision, semiconductor light switch.

In the initial commercial development of DMD technology, Texas Instruments has concentrated on projection display and hardcopy. The paper will focus on how the DMD is used for projection display. Other applicational areas are being explored and evaluated to find appropriate and beneficial uses for the DMD.

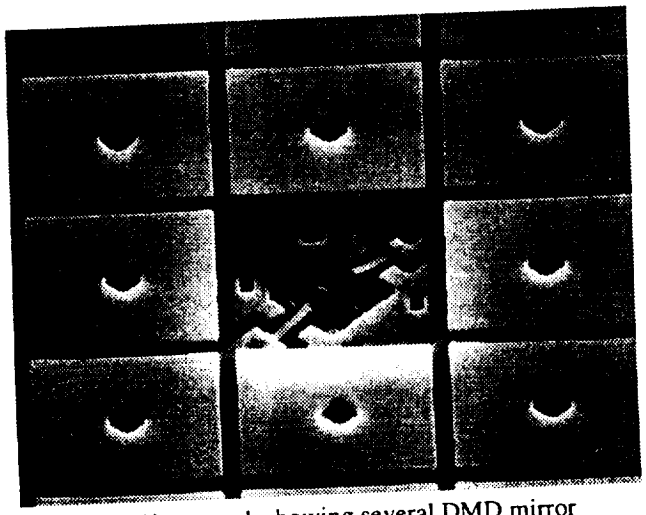


Figure 1. Photograph showing several DMD mirror elements and the mirror substructure (center).

Introduction

Starting in 1977, Texas Instruments began working on an analog light modulating technology called the Deformable Mirror Device, or DMD. This technology had limited performance and yield characteristics. By 1987, a bistable (digital) DMD was created that offered enhanced performance and showed no fundamental yield limitations. The DMD acronym was maintained but it now abbreviates the Digital Micromirror Device.

A DMD is a Micro-Electro-Mechanical Structures (MEMS) device composed of a two dimensional array of thousands of small, tilting, mirrors mounted atop of complementary metal-oxide-semiconductor (CMOS) Static RAM (S-RAM). In addition to having MEMS properties, the DMD also has an optical component, the tilting micromirrors. These mirrors, when combined with the proper optical projection system, create a truly digital process for displaying images.

Since the DMD is built over a standard CMOS circuit using conventional semiconductor processes, fabrication costs are expected to drop in line with a CMOS-like learning curve. These anticipated cost reductions have created a lot of excitement at Texas Instruments as it opens up the possibility of exploring many new markets. In investigating other applicational areas and markets for the DMD, the complete interrelationship of the DMD's MEMS and optical properties must be considered. Proper understanding of all aspects of DMD technology will give better insight as to how a DMD might become the solution to yet another technological challenge.

Markets

Texas Instruments is bringing DMD technology to the marketplace through its Digital Light Processing (DLP™) subsystem. At the core of a DLP subsystem is the DMD. Other DLP components are: memory, electronics, a power supply, a light source, a color filter system and projection optics. The goal for DLP is to compete in the projection display market; a market that is expected to have world wide sales of $4.6 billion in 1995[1].

Three projection markets have been targeted in which to sell DLP subsystems: consumer, business, and professional. The consumer market consists of front and rear-screen projection televisions. Projection television sales have been steadily increasing. This year alone, U.S. sales are up 29.1% over 1994[2]. DLP will enter the business market in the form of a conference room business projector. High brightness, 3-DMD DLP subsystems will be sold into the professional market for large screen display applications. Ultimately, these high brightness systems would be the cornerstone of future, digital cinemas.

DMD Structure

Each DMD consists of thousands of tilting, microscopic, aluminum alloy mirrors. The mirrors are 16 μm square and are separated by 1 μm gaps. These mirrors are mounted on a yoke and hidden, torsion-hinge structure which connects to support posts. The torsion hinges permit mirror rotation of +/- 10 degrees. The support posts are connected to an underlying bias/reset bus. The

[1] Source: Stanford Resources, Inc. "Projection Displays," Second Edition, 1994-95, pg. 3
[2] Source: Television Digest, Vol. 35, No. 38, pg. 12

bias/reset bus is connected such that both the bias and reset voltage can be supplied to each mirror. The mirror, hinge structure, and support posts are all formed over an underlying CMOS address circuit and a pair of address electrodes (See Figure 2).

Applying voltage to one of the address electrodes in conjunction with a bias/reset voltage to the mirror structure, creates an electrostatic attraction between the mirror and the addressed side. The mirror tilts until it touches the landing electrode that is held at the same potential. At this point, the mirror is electro-mechanically latched in place. Placing a binary "1" in the memory cell causes the mirror to tilt +10 degrees while a "0" causes the mirror to tilt -10 degrees. Each mirror on a DMD array has the ability to modulate incident light digitally as a semiconductor light switch. Full "on" to full "off" switching time is less than 20 µs.

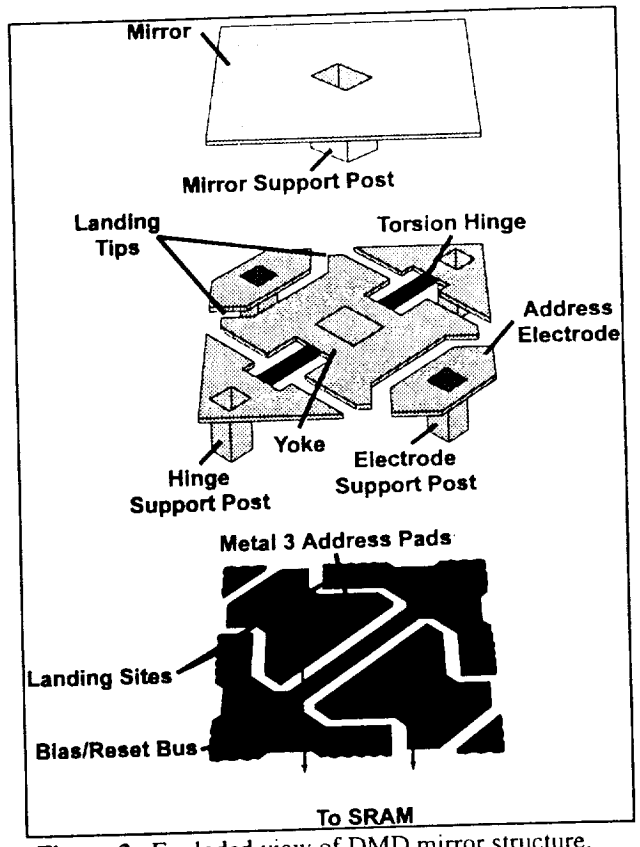


Figure 2. Exploded view of DMD mirror structure.

Mirror Assembly

The DMD chip is fabricated using 0.8 µm CMOS technology. The mirror system is built in CMOS wafer form on top of a chem-mechanical protective layer that separates the mirror system from the CMOS SRAM layer.

Construction of the mirror systems begins after contact openings for the address electrodes and bias/reset bus have been formed in the circuit's protective chem-mechanical oxide layer. Five photolithographically defined layers are surface micromachined to form the DMD mirror structure. From bottom to top, these five layers include: the metal layer that is connected to the CMOS structure via address electrodes and the bias/reset bus, a sacrificial layer, a coplanar hinge and address electrode layer, another sacrificial layer, and the reflective, aluminum alloy mirror layer. The sacrificial layers are made of an organic material that is plasma-ashed to form the two air gaps, one between the bottom metal layer and the coplanar hinge and address electrode layer, the other between the mirror and the hinge/address electrode layer. The other layers are formed from dry plasma etched, sputtered aluminum. The spacing between the mirror and bottom metal layer is 2 µm, enough room for each mirror to tilt +/- 10 degrees (See Figure 3).

After fabrication, the wafer on which the DMDs have been formed undergoes a wafer saw step. During this process, it is extremely important that the DMDs not be contaminated with particles. Micron size particles can interfere with the mechanical operation of the DMD as well as reduce its optical performance.

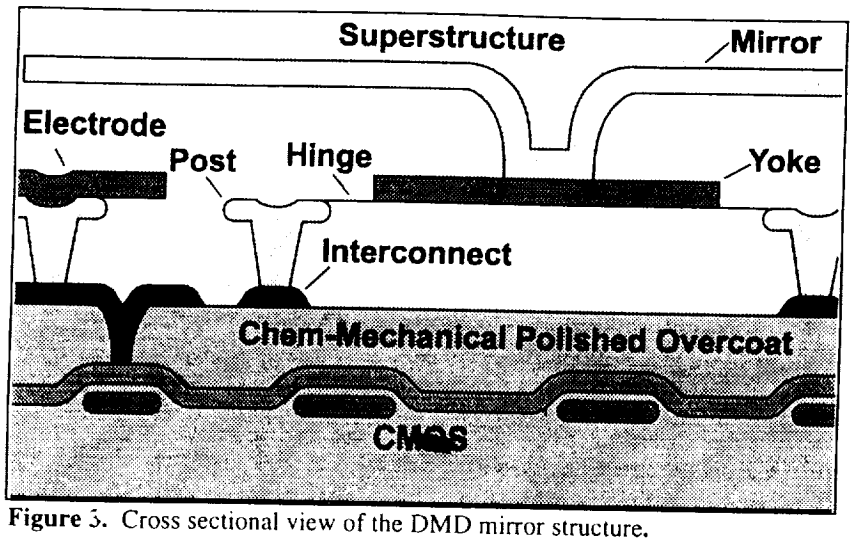


Figure 3. Cross sectional view of the DMD mirror structure.

Upon completion of the wafer saw, the chips are placed in a plasma etching chamber, where isotropic etching removes the sacrificial layers. Once these sacrificial layers are removed, the DMD finally becomes functional. Die attach, wire bond, window seal, and final test complete the sequence of assembly operations.

Video and Graphic Sources

To fully comprehend how a DMD functions in a display system, it is important to understand what needs to be displayed. The two dominant projection display sources are video and graphics. Standards for these sources are outlined below in Table 1.

NTSC: National Television System Committee-the interlaced, 60 Hz, 525 line color TV standard adopted by the United States in 1953. NTSC is also used in North and South America and Japan. Of the 525 lines, only 480 lines are active or visible to the viewer. In the interlaced mode, two 1/60 of a second TV fields make up one TV frame. Each field contains 240 alternating lines of information. As the two fields are interlaced, viewers see 480 active lines per TV frame. In the horizontal direction, NTSC TV can display about 330 vertically drawn black and white lines. This is often referred to as the horizontal lines of resolution, meaning the number of lines viewers can resolve when counting vertically drawn, black and white lines, across the picture. NTSC specifies a 4:3 aspect ratio.
PAL: Phase Alternation Line-the 625 line color TV standard used in Europe. PAL has 576 active lines per TV frame, the same 4:3 aspect ratio, operates in an interlaced mode but is specified for 50 Hz operation. PAL TV has about 420 horizontal lines of resolution.
HDTV: High Definition TV-a higher resolution TV standard that will have a 16:9 aspect ratio. Since all TV applications do not have the same requirements, multiple formats have been proposed. The highest proposed resolution is a 1,920 x 1,080 non-interlaced format.
VGA/SVGA/XGA/SXGA: Computer industry resolution standards with 4:3 aspect ratios, video, super video, extended, and super extended graphics adapter. VGA resolution specifies 640 x 480 pixels, SVGA specifies 800 x 600 pixels, XGA specifies 1,024 x 768 pixels, and SXGA specifies 1,280 x 1,024 pixels. Graphics modes are usually shown in a non-interlaced mode.

Table 1: Video and Graphics Signals standards

DMD Array Size

Depending on the requirement, DMD arrays can be configured in various formats. For projection display, arrays are built according to the resolution of the information intended to be displayed.

Initial Array

The first DMD chip was an array designed to display PAL broadcast signal, the European television standard. With square pixels, a 4:3 aspect ratio, and 576 visible lines in the vertical dimension, a DMD PAL chip is 4/3 x 576 = 768 mirrors wide. A DMD with an array of 768 x 576, containing 442,368 micromirrors was designed specifically for the purpose of displaying this PAL resolution TV signal. This chip also has the capability of displaying NTSC broadcast signal since NTSC has slightly lower resolution than PAL. When displaying the lower resolution, mirrors that are not addressed with information are simply tilted to the "off" position.

Other Arrays

Other array sizes of 864 x 576, 848 x 600, 1,280 x 1,024, and 2,048 x 1,152 have been built to support different aspect ratios and multiple resolutions of video and graphics input sources. A long, linear array of 64 x 7,056 mirrors has also been fabricated. Its intended use is to provide 600 dots per inch printing resolution across a 297 mm wide page for hardcopy applications.

HDTV Array

Demonstrating the DMD's ability to display future video sources, Texas Instruments with financial support from the Advanced Research Projects Agency (ARPA), developed a HDTV prototype projection display system. This system was designed to display the highest proposed 16:9 HDTV resolution, 1,920 x 1,080. The projector is based on three DMD chips, each containing 2.3 million, 16 μm square, micromirrors in 2,048 x 1,152 arrays. The project was successfully completed at the end of 1993, solidifying DMD display technology as a competitive solution for displaying current and future, video and graphics information.

DMD Display and Systems

Every time a binary "1" or "0" is delivered to the memory cell directly below each mirror, the proper address electrode is activated, and the mirror tilts to the "on" or "off" position. By using the DMD as a spatial light modulator, light incident on each of the thousands of mirrors on the chip can be precisely reflected to, or away from, a lens configuration. The light transmitted through the lens system is imaged onto a screen. The light that is reflected away from the lens system hits a black, light absorbing material. Through this process, a digitally projected image is possible (See Figure 4).

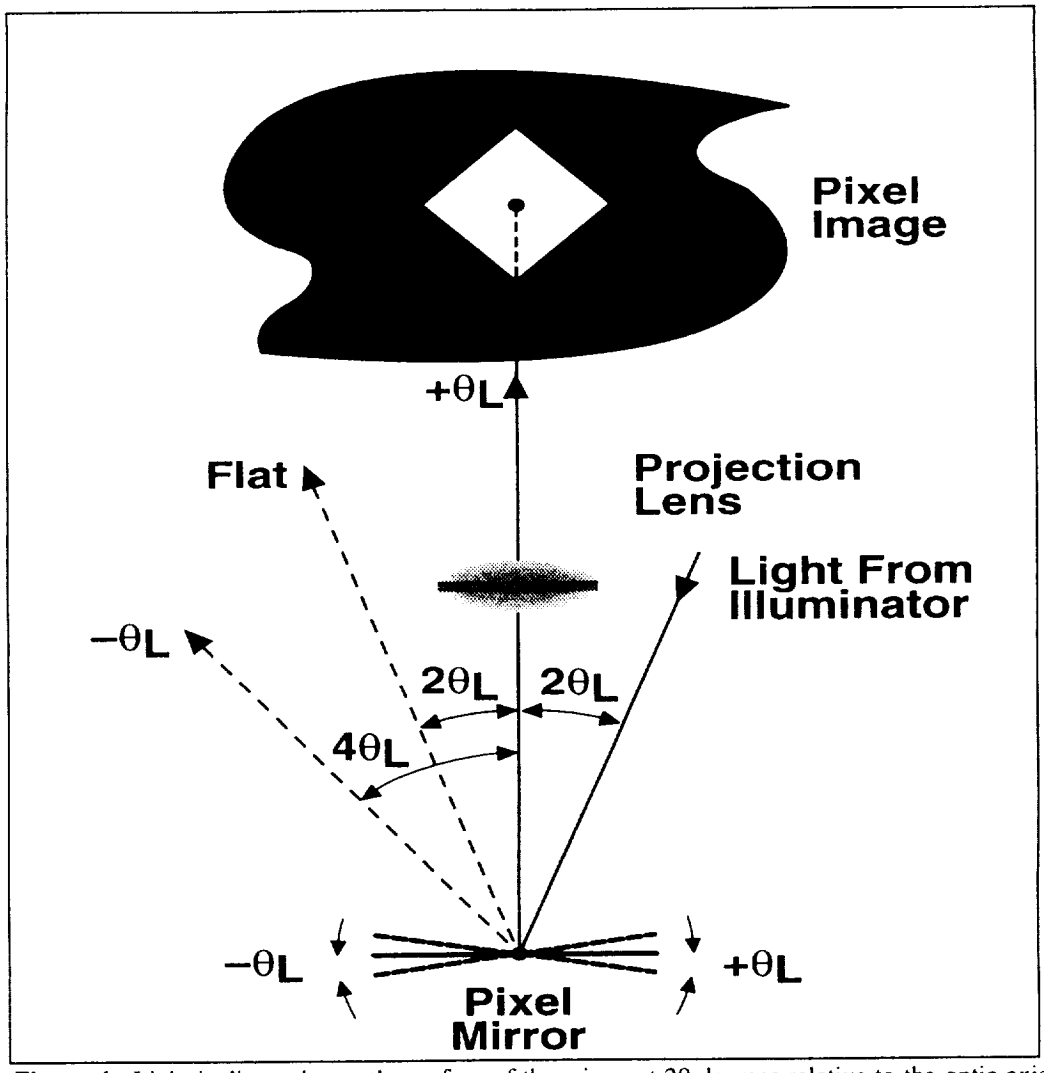


Figure 4. Light is directed onto the surface of the mirror at 20 degrees relative to the optic axis (which is perpendicular to the surface of the chip). A mirror that is tilted +10 degrees to the "On" position will direct the incident light directly up the optic axis, through a projection lens. This light then forms a pixel image on a screen. A mirror that is tilted -10 degrees will reflect light -40 degrees from the optic axis, away from the projection lens. This light is directed towards a black, light absorber.

Grayscale is achieved through a digital technique called pulse width modulation (PWM). With PWM, the amount of time that a mirror is "on" for a given TV field is controlled by the binary code sent to the memory cell. By varying the time the mirror is on, grayscale levels are generated.

For example, if 128 grayscale levels are desired, a 7 bit binary signal would be used (See Figure 5). The most significant bit is assigned to half of the TV field while. Proceeding bits are assigned to half of the remaining TV field with the least significant bit being assigned to 1/128th of the TV field. The DMD mirrors will tilt "on" or "off" for the given bit and the amount of reflected light is integrated by the human visual system to create perceived intensities or grayscale levels.

Color is added through a color filter or prism system, depending on the application and performance requirements.

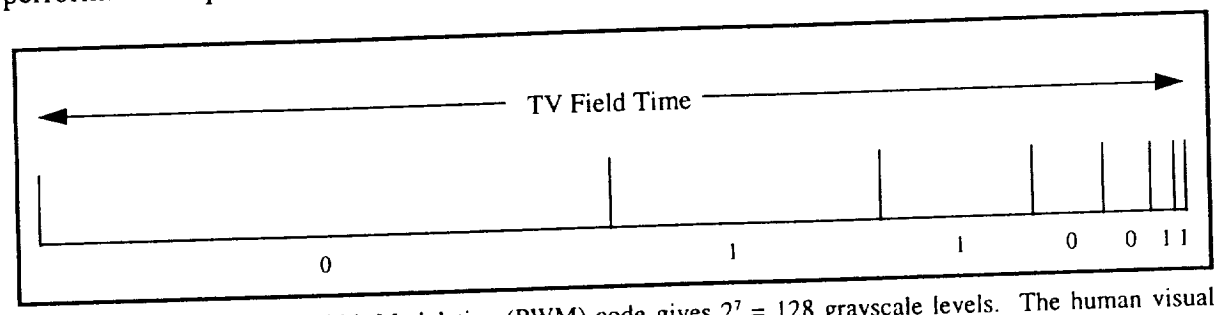


Figure 5. A 7 bit Pulse Width Modulation (PWM) code gives $2^7 = 128$ grayscale levels. The human visual system integrates the light that is reflected or "on" during the TV field and grayscale levels are realized. The 7 bit signal of 0110011 results in an intensity encoding of 40%, or a grayscale level of 51.

Video Projection Display

Currently, a video TV frame consists of two interlaced fields. One field writes every other line and the subsequent field writes the alternating lines. Interlacing works well for cathode ray tube (CRT) display because the phosphor glow persists long enough to make the image visible. DMD projection display is inherently a non-interlaced or progressive display technology; the entire array of mirrors is addressed for each TV field. It is therefore necessary to convert conventional interlaced TV signals into a non-interlaced format.

Several techniques can be adopted to convert interlaced signals into non-interlaced ones: line doubling, field jamming, and adaptive motion interpolation. Line doubling simply displays each TV line twice during the TV field, filling in the space between alternating lines. With field jamming, a memory buffer stores one of the interlaced fields and displays it with the other field, superimposing the two fields. Adaptive motion interpolation looks at the alternating lines in real-time, on a pixel-to-pixel basis, and "interpolates" what the picture should look like in between the two lines. The interpolated information is then scanned in between the alternating lines. This seems to be the best approach to compensate for motion artifacts that are more noticeable when the other two techniques are used. Proper conversion of the interlaced signal into a progressively scanned signal also has the advantage of increasing the apparent perceived resolution of the projected image.

Graphics Projection Display

The majority of computer monitors today operate in a non-interlaced mode. They scan every line and write the entire frame in just one field. Without non-interlaced scanning, the fading of the phosphors between each alternating line can be more noticeable, causing the screen to flicker.

Since the DMD shows a full, progressively scanned image, it couples well with graphics inputs and displays razor sharp graphics images. Ideally, a DMD used for graphics projection would map each pixel of information to its own mirror. This way, exact digital pixel control can be achieved. It is also possible to display higher resolution graphics sources on a DMD having less mirrors than pixels through the use of scaling techniques.

1-Chip System

In a single DMD projection system, a color wheel is used to create a full color, projected image. The color wheel is a red, green, and blue, filter system which spins at 60 Hz to give 180 color fields per second. In this configuration, the DMD operates in a color field sequential mode.

The input signal is broken down into red, green, and blue (RGB) components. The signal goes through the DMD formatter and picture buffer electronics. It is then written to the DMD's SRAM. A white light source is focused onto the color wheel through the use of condensing optics. The light that passes through the color wheel is then imaged onto the surface of the DMD. As the wheel spins, sequential red, green, and blue light hits the DMD. The color wheel and video signal are in sequence such that when red light is incident on the DMD, the mirrors tilt "on" according to where and how much red information is intended to be displayed. The same is done for the green and blue light and video signal. The human visual system integrates the red, green, and blue information and a full color image is seen. Using a projection lens, the image formed on the surface of the DMD can be projected onto a large screen (See Figure 6).

Since a NTSC TV field is 16.7 ms (1/60 of a second), each of the primary colors must be displayed in about 5.6 mili-seconds. Given that the DMD has <20 μm switching time, 8-bit grayscale per color (256 shades) is possible with a single DMD system. This gives 256 shades for each of the primary colors or, $256^3 = 16.7$ million possible colors that can be generated.

When a color wheel is used, 2/3 of the light is blocked at any given time. As white light hits the red filter, the red light is transmitted and blue and green light is absorbed. The same holds true for the blue and green filters; the blue filter transmits blue and absorbs red and green, the green filter transmits green and absorbs red and blue. The result is that a 1-chip system is inefficient in its use of light, however, for certain applications, this approach offers an excellent price/performance match.

3-Chip System

Another approach is to add color by splitting white light into the three primary colors by using a prism system. In this approach, three DMDs would be used, one for each of the primary colors. The main reason for using a 3-DMD projection system is for improved brightness. With 3 DMDs, light from each of the primary colors is directed continuously at its own DMD for the entire 16.7 mili-second TV field. The result is that more light gets to the screen, giving a brighter projected image. In addition to increased brightness, higher bit color can be realized. Since light is directed to each DMD for the whole TV field, 10 or 11-bit grayscale per color is possible. This highly efficient, 3-chip projection system would be used for large screen and high brightness applications.

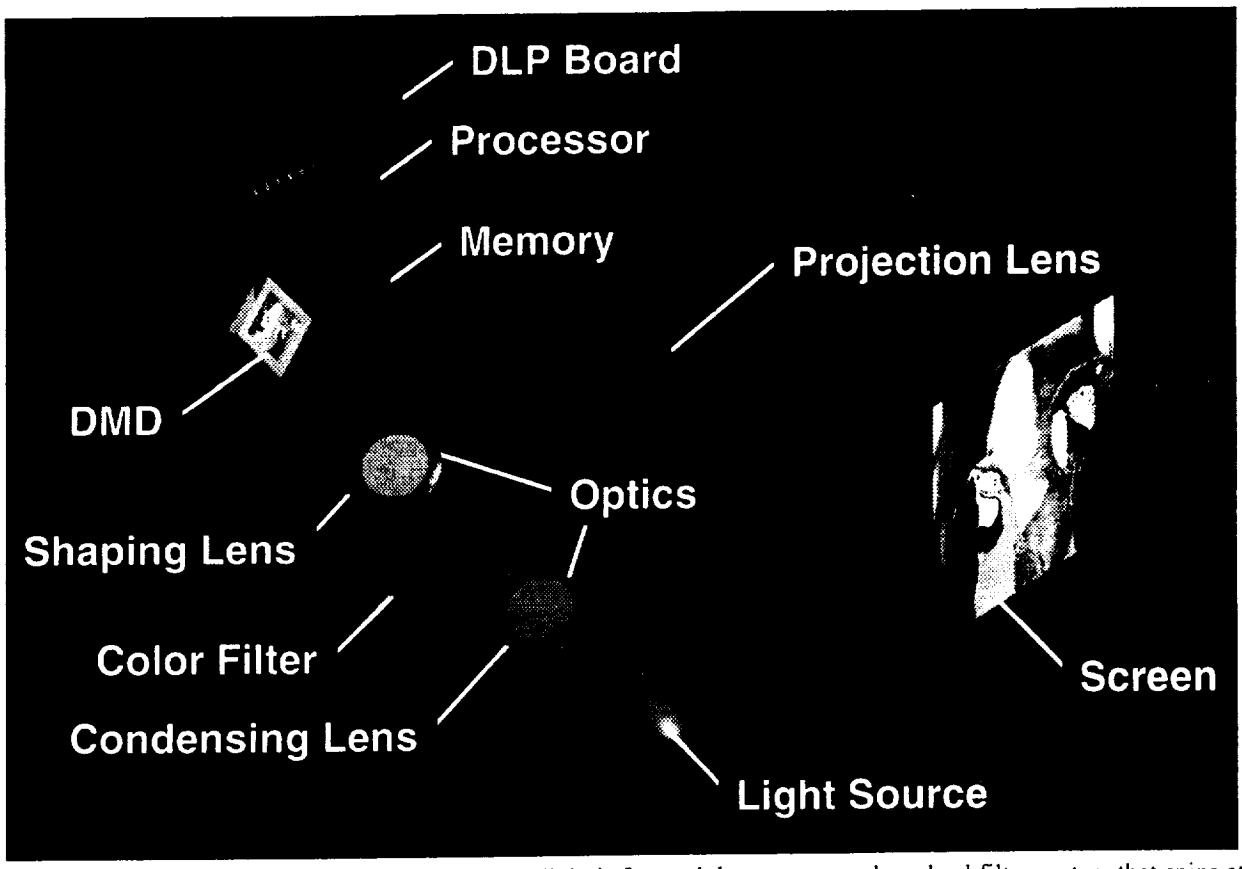


Figure 6. 1-chip, DMD projection system: White light is focused down onto a color wheel filter system that spins at 60 Hz. This wheel spins in sequence with the red, green, and blue video signal being sent to the DMD. Mirrors are turned "on" depending on where and how much of each color is needed for each TV field. The human visual system integrates the sequential color, and a full color image is seen.

Advantages

DMD based projection display systems have many advantages over the existing Cathode Ray Tube (CRT) and Liquid Crystal Display (LCD) projection technologies.

Pixel Fill Factor and Uniformity

DMD chips with 16 μm square mirrors spaced on 17 μm centers have a fill factor of up to 90%. In other words, 90% of the pixel/mirror area can actively reflect light to create a projected image. Pixel size and gap uniformity is maintained over the entire array and is independent of resolution. LCD's have at best, a 60% fill factor. CRT's are not capable of producing square pixels since they rely on an electron beam scan, not a pixelated array. The DMD's higher fill factor gives a higher perceived resolution, and this, combined with the progressive scanning, creates a projected image that is much more pleasing to the eye than conventional projection display.

Light Efficiency

The DMD is capable of having an overall light efficiency of over 60%. The definition of light efficiency here is simply the percentage of output light as compared to the amount of input light.

For a DMD, there are four multiplying components that make up its light efficiency: a temporal component, the reflectivity of surface, the fill factor, and diffraction efficiency.

DMD Light Efficiency = (Actual time "on") x (Re. of Surface) x (Fill Factor) x (D. Efficiency)

= (92%) x (88%) x (90%) x (85%) = 61.9%

LCD projection displays are inherently light inefficient. First, they are polarization dependent so one of the polarized light components, or half of the lamp light is not used. Other light is blocked by the transistors, gate, and source lines in the LCD cell. In addition to these light losses, the liquid crystal material itself absorbs a portion of the light. The result is that only a small amount of the incident light gets transmitted through the LCD panel and onto the screen.

Higher Resolution and Brightness

Increasing the input resolution simply means that more mirrors on the DMD have to be activated. Higher resolution can be achieved independent of brightness. As brightness is increased with CRT projectors, the phosphors "bloom" causing resolution to drop. In a DMD projection system, increasing brightness simply means that more light is reflected off of the DMD and onto the screen, with no loss in resolution. The amount of brightness for a given DMD system is a function of the light source used and the number of DMDs in the system. Several hundred to thousands of lumens of brightness can be displayed using DMD technology. This provides flexibility to both the manufacturers and markets that DMD projection display will serve. The DMD's higher resolution and brightness capabilities makes it more desirable for current and future projection display applications.

Digital Control

Each pixel of information displayed on a screen is precisely and digitally controlled independent of surrounding pixels. Spatial repeatability is achieved and through the use of PWM, grayscale and color levels can be accurately repeated, time after time.

Reliability

The DMD has passed all standard, semiconductor qualification tests. In addition to these tests, Texas Instruments has evaluated the DMD's performance reliability for the DLP subsytems that will be sold to the three previously mentioned markets: consumer, business, and professional. The DMD has passed a barrage of tests meant to simulate actual DMD environmental operation including: thermal shock, temperature cycling, moisture resistance, mechanical shock, vibration, and acceleration testing.

Because the DMD relies on a moving hinge structure, most of the reliability concerns are focused on the hinge life. To test hinge failure, approximately 100 different DMDs were subjected to a simulated one year operational period. Some devices have been tested for over 1 trillion cycles,

equivalent to 20 years of operation. Inspection of the devices after these tests showed no broken hinges on any of the devices. Hinge failure is not a factor in DMD reliability.

Conclusion

Produced using standard semiconductor processes, the DMD is becoming a highly useful optical MEMS device for projection display. Device reliability has been validated leaving no reason to inhibit marketplace acceptance. Other applicational areas for DMD technology are being evaluated but the current thrust for DMD deployment into the marketplace is in the area of projection display. DMD and DLP technology will offer many performance advantages over current projection display technologies. High speed operation, brightness, resolution, fill factor, high optical efficiency, and 16.7 + million color reproductions are all advantages of DMD projection display. Combining these DMD advantages with complete digital control makes DLP an exciting new technology. Texas Instruments' Digital Light Processing, based on the DMD, offers an excellent solution for projection display as the world begins to enter into a true, digital, multimedia age.

References

1. Jack Younse, IEEE Spectrum, "Mirrors on a Chip," November, 1993

2. Michael A. Mignardi, Solid State Technology, "Digital micromirror array for projection TV," July, 1994

3. Jeffrey Sampsell, "An Overview of the Performance Envelope of Digital Micromirror Device (DMD) Based Projection Display Systems," SID International Symposium, Digest of Technical Papers, June, 1994

4. Robert J. Gove, "DMD Display Systems: The Impact of an All-Digital Display," SID International Symposium, Digest of Technical Papers, June, 1994

5. Edison H. Chiu, Can Tran, Takeshi Honzawa, and Shigeki Numaga, "Design and Implementation of a 525 mm^2 CMOS Digital Micromirror Device (DMD) Display Chip."

6. Nelson and Rohit L. Bhuva, "Digital Micromirror Imaging Bar for Hardcopy," Color Hard Copy and Graphics Arts IV, Proceedings Preprint from SPIE-The International Society for Optical Engineering, February, 1995

7. Jack Younse, "Projection Display Systems Based on the Digital Micromirror Device (DMD)," Micromachining and Microfabrication '95 Conference, October, 1995

8. Feather, G.A., "Digital Light Processing: Projection Display Advantages of the Digital Micromirror Device," Montreux '95-The International Television Symposium and Technical Exhibition.

9. M.R.Douglass, D.M. Kozuch, "DMD Reliability Assessment for Large-Area Displays," SID International Symposium, Digest of Technical Papers, June, 1995

10. Brian Evans, "Understanding Digital TV-The Route to HDTV," IEEE Press, 1993

Aercam Demonstration

Charles Price, NASA/JSC

Hardware display and video; no text presently available.

MEMS Spaceborne Testbed

Robert H. Stroud
The Aerospace Corporation

Mark Holderman
NASA/JSC

7/128
030653
1P

Microelectromechanical systems (MEMS) technology can significantly reduce the cost to design, build, launch and operate space systems. This will require new enabling technologies for the next generation of spacecraft. The MEMS research community is also actively looking for applications. Collaboration can bring great benefits to both groups.

MEMS technology offers low cost, small, light weight and reliable devices. These can significantly reduce the size, weight and cost of spacecraft while increasing their reliability through durability and redundancy. A standard spaceborne interface/cabinet that is readily available for experimenters will enable inexpensive testing of MEMS devices in space. Such a testbed is being proposed for use on the Space Shuttle, and an initial design and set of specifications is now being developed. This will allow the MEMS community to explore new applications that have thus far been cost prohibitive.

There are many advantages to a predefined and readily available testbed: leverage of the technology and innovations from universities and industry, rapid and inexpensive flight testing of devices in the space environment and incorporation into space programs, a vehicle for technology insertion studies, built in environmental monitoring by the host platform, and experiment recovery for post mission analysis. It will also provide a means for incremental development and testing of spacecraft systems from individual MEMS devices through completed, flight tested MEMS components and subsystems, possibly stimulating new technologies and devices in the process.

The MEMS testbed will provide fixtures for holding MEMS devices, ambient environment monitoring, an assortment of available operating voltages, RF and low frequency signal inputs, sources of various gases and pressure monitors, a command and control interface computer, and a data interface which will record or transmit the data and performance parameters of these small systems. Since the testbed itself will have been space qualified, the burden of space qualification for the experimenter will be limited to the experiment itself, not the ancillary operating systems.

Individual experiments have, to date, either provided these capabilities themselves or relied on the intervention of the shuttle crew. Both are expensive alternatives. Providing them on a single, predefined MEMS testbed alleviates much of the expense and logistics burden from the experimenter and shuttle crew, and encourages participation in the space community by MEMS researchers. This readily available spaceborne testbed will be available to experimenters worldwide at a minimal cost.

Low volume packaging for a microinstrumentation system

71130

A. Mason and K. Wise

University of Michigan

2 30655

1P.

A folding, multi-platform assembly has been developed for the packaging of a multi-chip microinstrumentation system. The assembly includes three solid platforms connected by flexible micromachined ribbon cables and can be populated by sensors and electronics from a variety of technologies including surface mount and IC processes. The entire structure is fabricated from a four-inch silicon wafer using a simple four mask process and a post-process EDP etch. The micromachined ribbon cables allow the platform assembly to be folded into a three level structure with control electronics on the bottom level, microsensors and interface electronics on the second level, and sensors that need environmental access on the top level. Utilizing the silicon multi-platform assembly, a prototype microinstrumentation system has been developed that includes a microcontroller unit and sensors for measuring temperature, barometric pressure, humidity, altitude, and acceleration as well as a telemetry device for wireless communication. When folded into the three level structure, this microsystem occupies only 5cc and can be placed in an outer case the size of a wristwatch.

Poster not available at time of publication.

PYROELECTRIC APPLICATIONS OF THE VDF-TrFE COPOLYMER

J.J.Simonne[1], Ph.Bauer[2], L.Audaire[3], F.Bauer[4]

[1] Laboratoire d'Analyse et d'Architecture des Systèmes LAAS-CNRS, 7, Avenue du Colonel Roche, F-31077 Toulouse Cedex, France

[2] SOFRADIR, 43/47 rue Camille Pelletan, F-92290 Chatenay-Malabry, France

[3] CEA/DTA/LETI/DOPT/S-LIR, 17 rue des Martyrs, F-38054 Grenoble Cedex, France

[4] Institut de Recherche Franco-Allemand de St Louis, ISL, BP 34, F-68301 Saint Louis, France

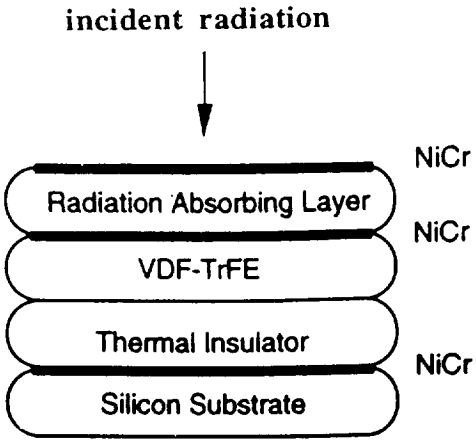
ABSTRACT

VDF/TrFE pyroelectric sensors have now definitely reached the level of a product. Based on a bidimensional staring array, it can be considered as a whole system with a monolithic technology processed on a silicon substrate provided with the integrated read out circuit.

The paper will describe the main procedures dealing with the elaboration of a 32x32 Focal Plane Array developed, in the context of the PROMETHEUS PROCHIP European Program (EUREKA), as a passive infrared obstacle detection (1) applied to automotive. Additional experimental data suggest that this microsystem could operate in space environment .

INTRODUCTION

A few years after the emergence of copolymers as pyroelectric elements, the high grade material currently commercially available (2) and its capability to be processed on a silicon substrate in a way similar to any microelectronic system, offer a real chance of industrial development to these sensors. Large sectors of industry dedicated to mass production like automotive, but also, medecine, agronomy, space... are concerned, and new developments of the Microelectromechanical system (MEMS) which can be adapted to the technology of these sensors deserve to pay attention to a material, which appears as a true challenger of 'high tech - low cost' infrared detectors already present or arriving on the market.



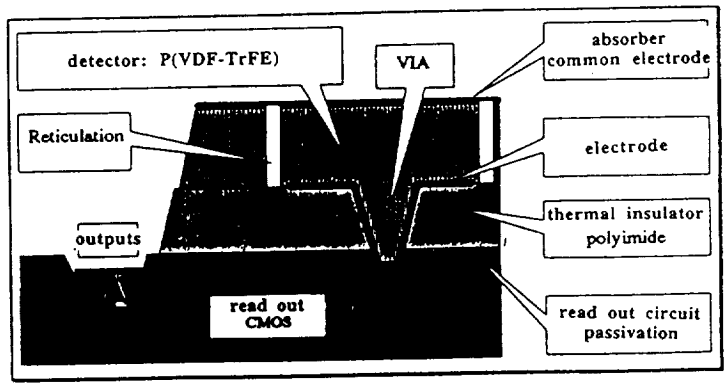
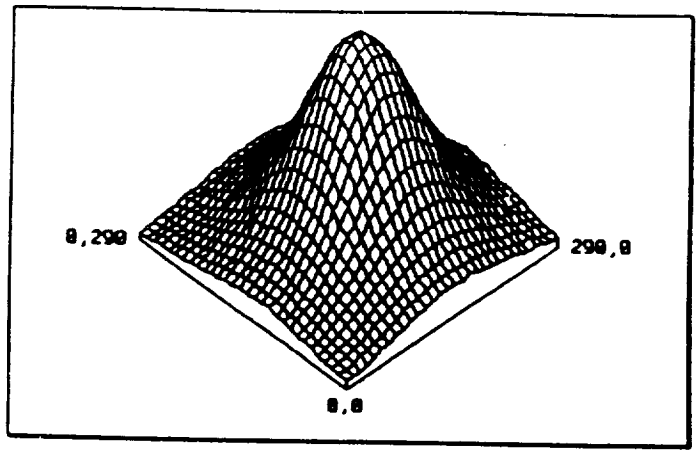


Fig.1 Basic Sensor Structure **Fig.2 Detector array - Pixel cross-section**

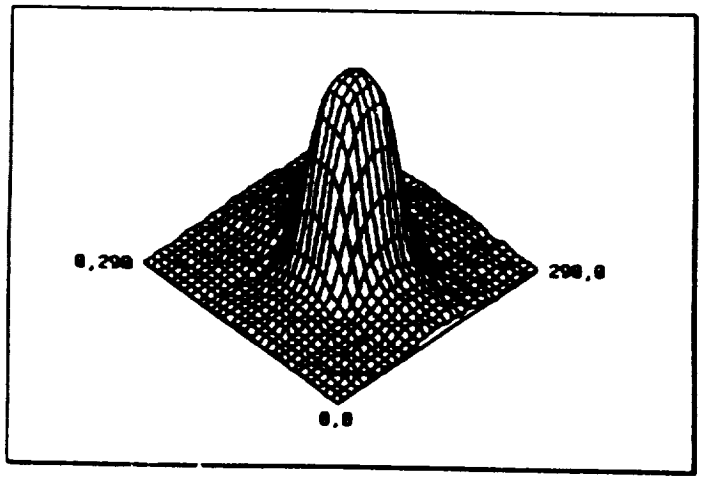
This paper intends to describe the latest development achieved in the design and the technology of a 32x32 copolymer VDF-TrFE staring array, associated with a focal plane, and coupled with a signal processing circuit integrated in a silicon substrate.
The heterostructure (Fig.1) basically includes:
- a copolymer layer (VDF,0.7 - TrFE, 0.3) allowing the pyroelectric detection after absorption of the chopped incident infrared radiation; converted into a temperature variation, it will induce a pyroelectric signal between the electrodes deposited above and underneath the copolymer. This material is provided on top with a thin absorber film.
- a silicon substrate in which the read out circuit is processed in a 1.2μm C.MOS technology.
- a thermal insulator (polyimide), sandwiched in between copolymer and silicon, which improves the thermal sensitivity of the pixels.

ELABORATION OF THE STARING ARRAY

A schematic partial cross-section of the array is displayed in Fig.2, showing the different layers of the heterostructure. The 'via' allows to connect electrically copolymer and signal processing circuits.



a)



b)

Fig.3 Cross-talk 3D map of a pixel (pitch 100μm)
a) when the whole area is covered by the copolymer layer
b) when the copolymer is reticulated

Three main steps were successfully completed to validate this approach:
- the copolymer, which cristallizes directly in a beta polar phase, is available under a liquid form and deposited by spin coating upon its support (polyimide+silicon). The real difficulty is to perform the poling procedure across the whole heterostructure already provided with the read out circuit. This has been achieved without any damage of the electronic circuit. The method used, first developed by F.BAUER (3), consists in the application of a very low frequency electric field (0.1Hz) at room temperature, allowing to generate many times the hysteresis loop while the magnitude of the field is increased continuously up to a limit slightly under the breakdown voltage of the material ($3x10^8$V/m). Reproducible values for the remanent polarization, equal or higher than $8\mu C/cm^2$ are commonly reached.

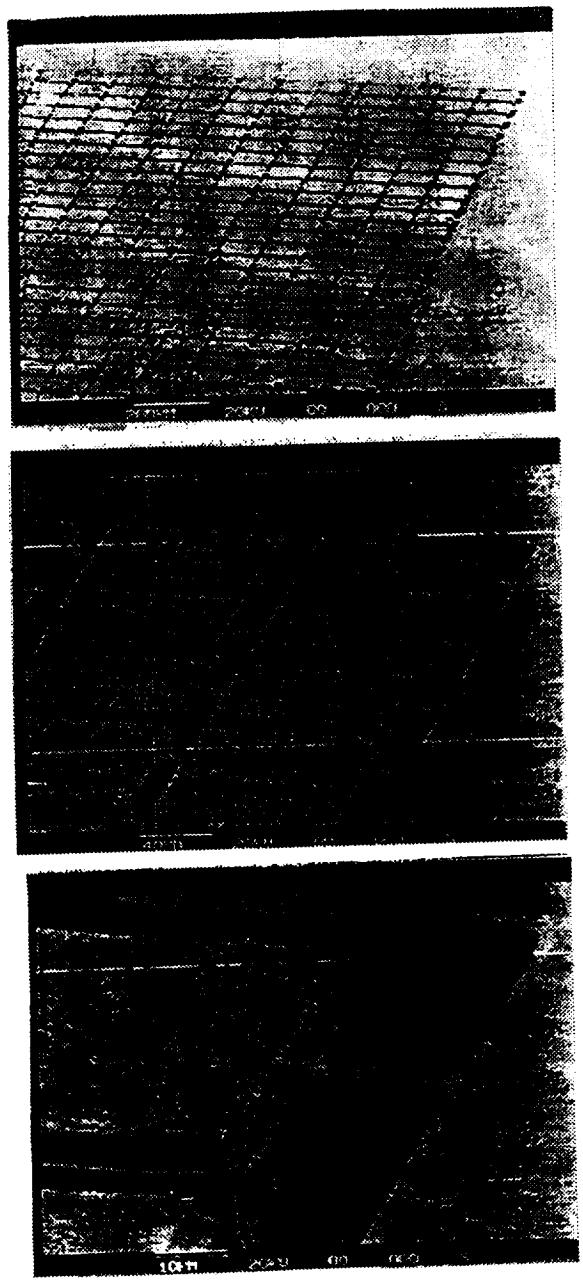


Fig.4 Dry etching for reticulation of the sensor array

- The low thermal conductivity of (VDF-TrFE) should allow to avoid any reticulation of the copolymer layer in the matrix array; However, according to Broadhurst et al (4) and confirmed by our experiments, a large part of pyroelectricity in this material is mostly due to a coupling between piezoelectricity and thermal dilatation of the material rather than to the variation of the spontaneous polarisation in the crystalline phase at constant volume. For this reason and for its beneficial impact on the thermal cross-talk as well, the reticulation by oxygen plasma etching of the different layers constituing the active part of the detector - metal, absorber, copolymer - has been successfully tested; this procedure allows a maximum of deformation under the effect of a temperature variation, improves the pyroelectric coefficient, and a minimum thermal cross-talk has been also checked after this procedure (fig.3). A SEM picture of the reticulation is shown in figure 4.
- The signal processing silicon circuit has a direct impact on the sensitivity of the sensor. A first system (IRPY1) elaborated has demonstrated the necessity to reduce the fixed pattern noise by a self calibration integrated at the level of each unit pixel. Therefore, a second generation (IRPY2) has been designed, including a reduction of the read out capacitance and of the access resistances, allowing a significant decrease of the whole area by a factor 2 (see the picture of both sensors on their headers).
In order to optimize the performance of the sensor, the theoretical analysis of the response of the array has been investigated through the thermal diffusion model adapted to a general heterostructure by Shu-Yau Wu (5), and through a second model based on a network of distributed thermal resistances and capacitances associated to each layer, to include the effect of the 'via' on the thermal behaviour of the sensor.

As a result, a thickness of 17µm for the copolymer, and of 10µm for the polyimide have been selected in the project.(4). The thermal cut off frequency has been optimized below 50Hz for the frame rate application of 10Hz.

The read out circuit is a 32 lines x 32 columns, 100µm pitch staring array. The pixel signals are addressed by lines and the internal data bus buffers are multiplexed to the video amplifier. Fig.5 gives a schematic view of the array architecture whose characteristics are detailed in Ref.(6).

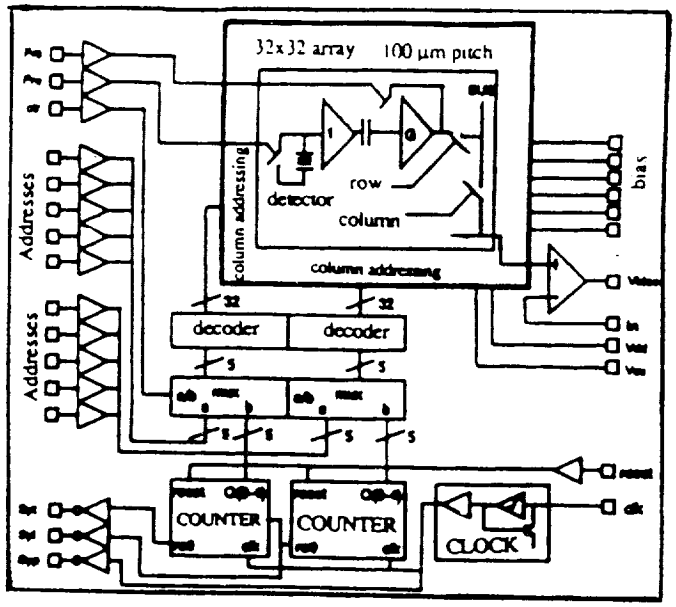


Fig.5 Copolymer Focal Plane Array: Architecture of the read out circuit

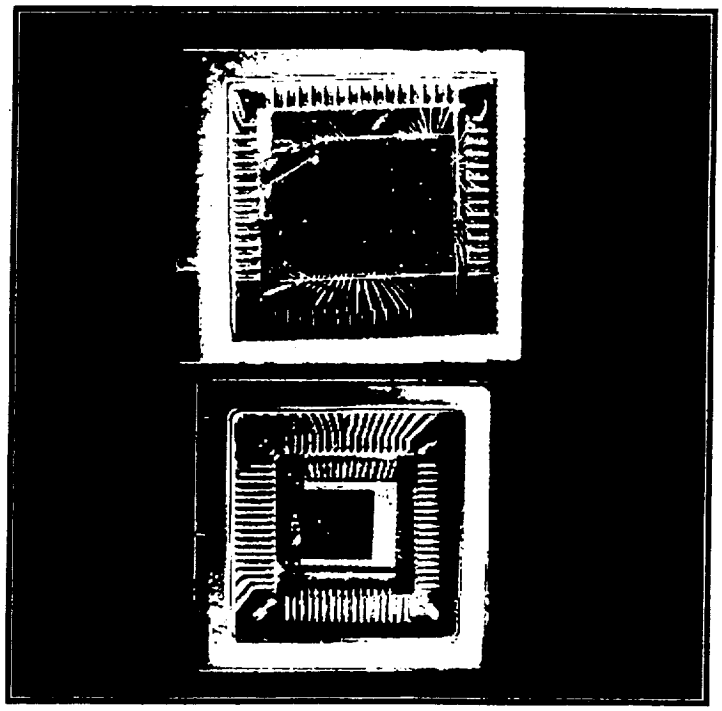
SENSOR PERFORMANCE

Interest of (VDF-TrFE) in the present application does not rely on its pyroelectric coefficient, fairly low when compared to other traditional pyroelectric materials (see Table 1). Other parameters like dielectric constant (fig.6) and loss tangent (fig.7) have hopefully a direct impact on the sensor performance and allow the copolymer to exhibit a good voltage merit factor. In addition, the material does not require costly preparation, is insensitive to humidity, is chemically inert, stable, and, as already said, available under a liquid form, properties which make it attractive in IR imagery.
Improvement on two particular performances has been emphasized: sensitivity and noise properties.

Material	Form	Curie point °C	Pyroelectric coefficient nCb/cm²/K	Permittivity F/m	Thermal diffusion cm/°C	Voltage readout merit factor	Charge readout merit factor
TGS	Crystal	49	35	50	72	7	4.9
$LiTaO_3$	Crystal	620	17	43	-	4	2.6
$PbTiO_3$	Ceramic	490	30	200	140	1.5	2.1
PZT	Ceramic	> 300	35	300	110	1.2	2
$P(VF_2\text{-}TrFE)$	Copolymer	125	4	7	20	5.7	1.5

Pyro-electric materials : properties, performances and ability to readout.

Table 1



32x32 pyroelectric copolymer arrays

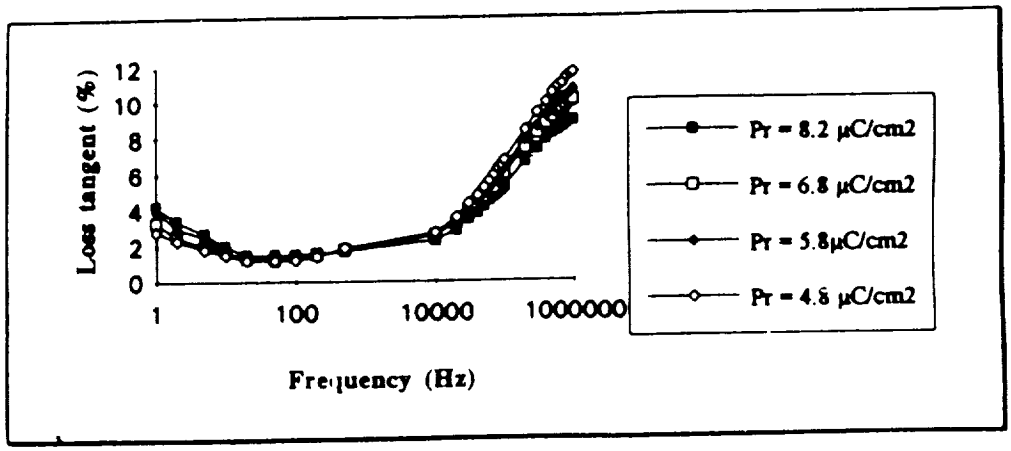


Fig.6 tgδ versus f for several remanent polarisation P_r
(tgδ is minimum between 10 and 200Hz)

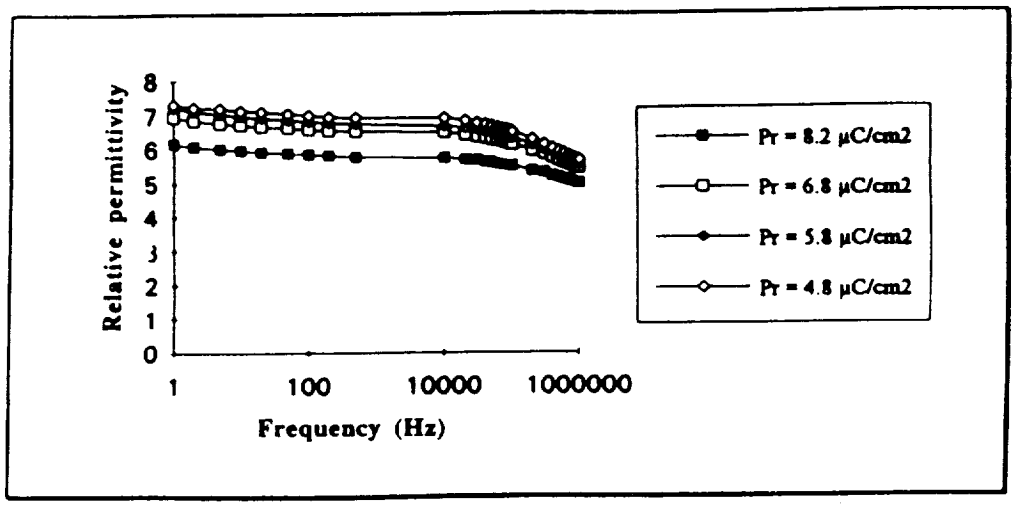


Fig.7 ε_r versus f for several remanent polarisation P_r

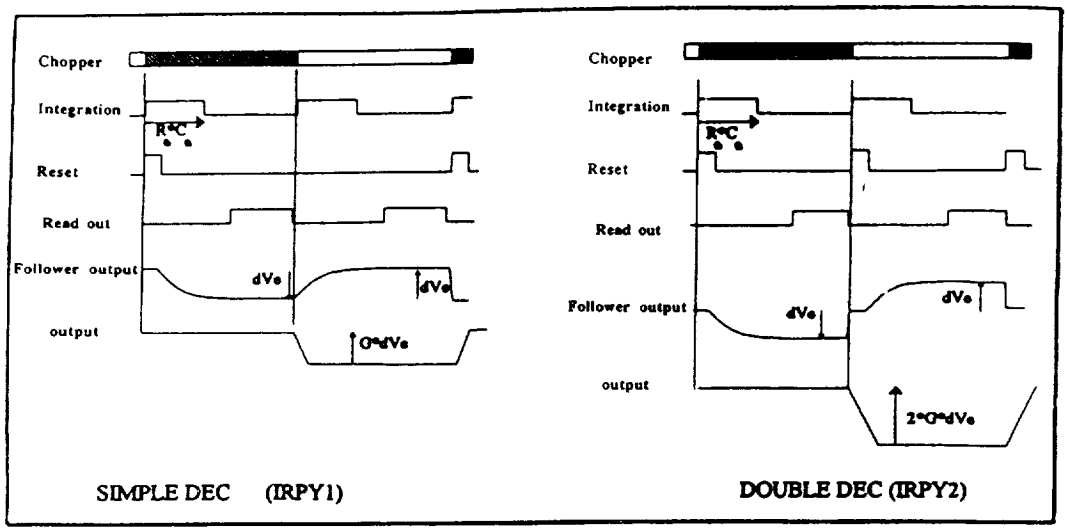


Fig.8

- concerning the sensitivity, a double correlated sampling (double DEC), using the subtraction of the signals generated in both initial positions of the chopper inducing in a pixel +ΔT (aperture), then -ΔT (shutting), allows to double the response (fig.8).
- concerning noise properties, the fixed pattern noise is ruled out by the double DEC procedure, while other noise sources have also been shrunk by optimization of the connexion lengths, leading to the circuit lay out of the second generation. Figure 9 and figure 10 illustrate the improvements of the responsivity and of the Noise Equivalent Temperature Difference (NETD) with both technologies IRPY1 and IRPY2, including different upper electrodes.

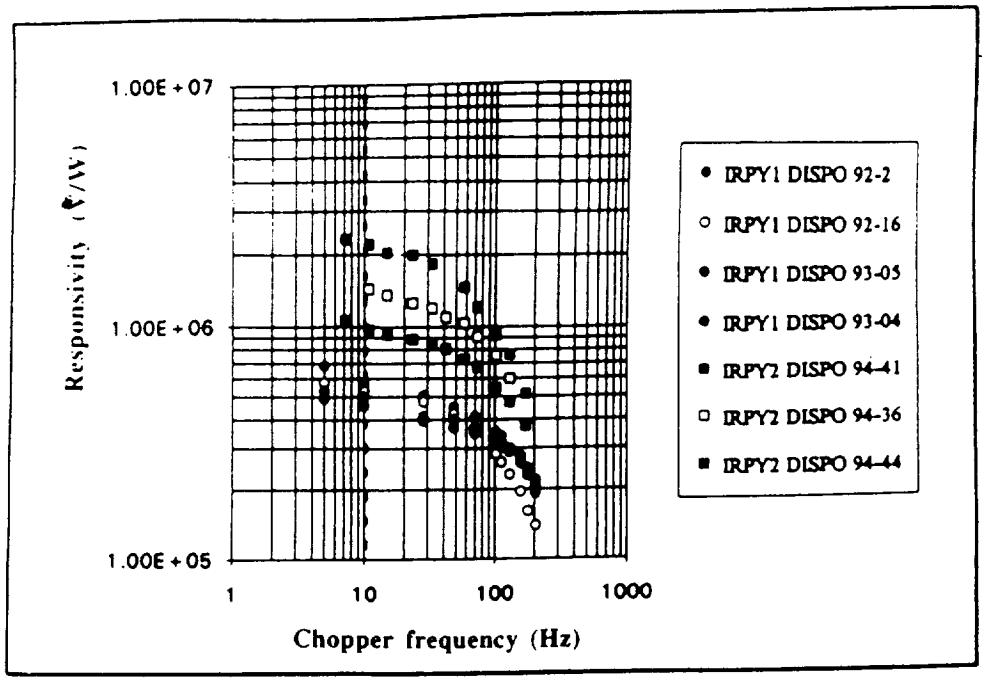


Fig.9 **Evolution of the responsivity with the improvement of the technology.**
nature of the upper electrode: 92-2 thin Cr layer; 92-16 graphite spray; 93-05 thin Cr layer; 93-04 thin Cr layer; 94-41 absorber; 94-36 thin Cr layer; 94-44 reticulated absorber.

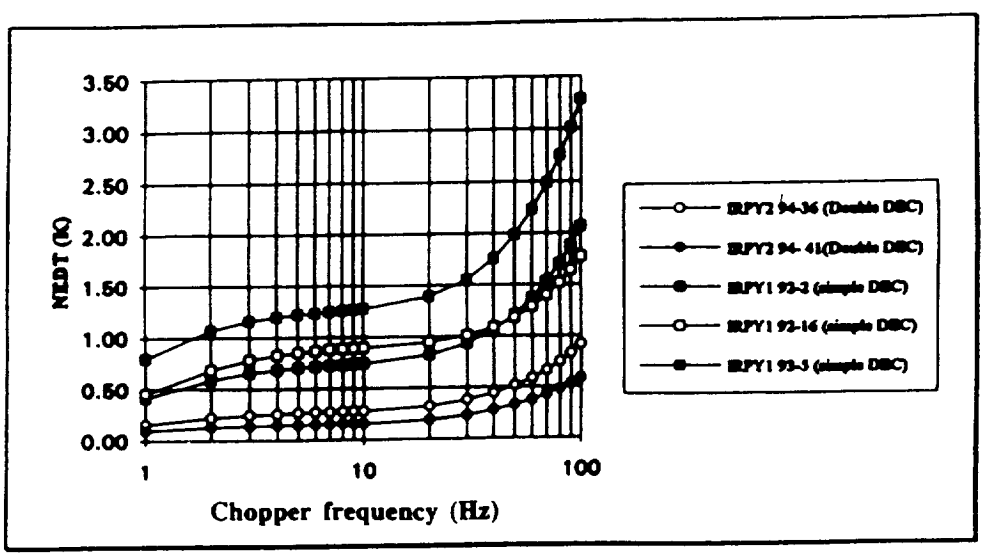
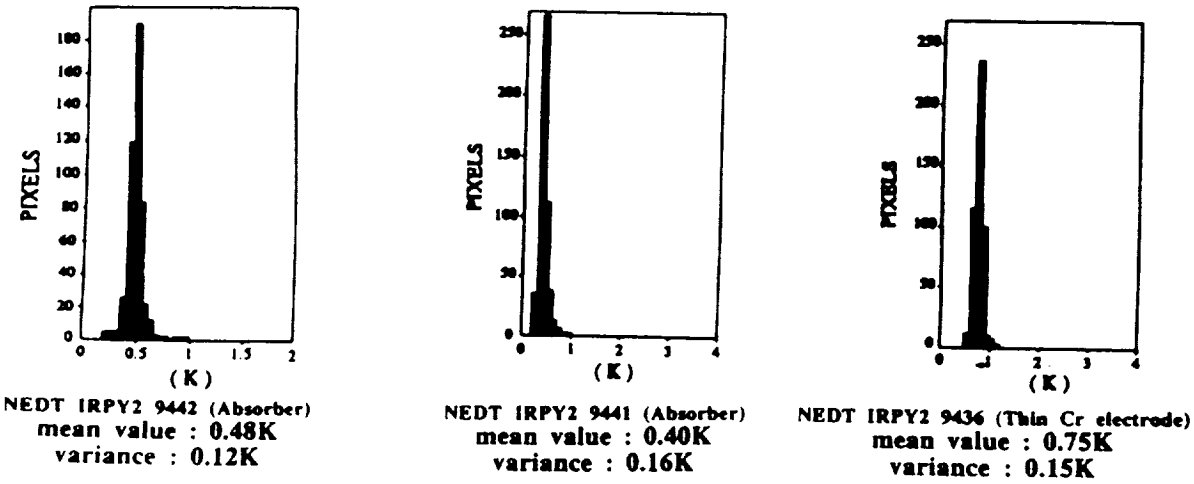


Fig.10 **Noise properties for different technologies (se fig.9 for the nature of the upper layer)**

The NETD histograms presented in figure 11, with an ideal optic transmission and a F/1 numerical aperture, give a mean measured value of 0.40K performed on all pixels of the IRPY2 array, which could be even decreased to 0.16K as noticed through simulation on the same figure.



Optimum simulation NEDT : 0.16K
bandwidth : 8-14μm, optics f/1, transmission : 1

Fig.11 NETD Histograms of the IRPY2 devices

COMMENTS

This project has demonstrated the capability of VDF-TrFE to be processed as a pyroelectric Focal Plane Array dedicated to infrared imagery, with latest performances able to reach those presented in recent alternative solutions, namely the microbridge resistive bolometer array developed by Honeywell (7), and the hybrid dielectric bolometer processed by Texas Instruments (8). This copolymer thin film heterostructure and a micromachined linear array also implemented elsewhere (9) make clear that the Microelectromechanical systems (MEMS) will be the most predictable approach selected for 'high tech-low cost' IR imagers.
If automotive has been the driving force of this application, this Conference is a good opportunity to examine how space can be also considered as a potential field for this material. Tests experimented at different temperatures have demonstrated a working stability between -50°C and +50°C before unfolding, and in the range -150°C/+150°C, when operating outside. Stability with time in another useful factor in favour of copolymer, property which emphasizes the role which could be played by this system in the payload of a satellite: a 10 years storage does not disturb the response of the active material. Furthermore, we must keep in mind the tremendous characteristics of (VDF-TrFE) in piezoelectricity, demonstrated in shock compression (10) and accelerometer applications (11). Some early successful tests have been made to examine the behaviour in a heavy ion ambient and more investigation is likely needed to check its properties versus other kinds of charged particles; nevertheless, in a period where the space technology is going to know a huge transformation, capabilities of the pyroelectric and piezoelectric copolymer cannot be ignored, even if additional data must be provided to state whether they appear as a useful approach to the space field.

ACKNOWLEDGEMENTS

We would like to acknowledge the technology teams of LETI/LIR, LAAS and ISL for their contribution to this work, and PROMETHEUS-PROCHIP for financial support of this program.

REFERENCES

(1) EUREKA/PROMETHEUS/PROCHIP, Project 2211
see also D.Esteve, P.A.Rolland, J.J.Simonne and G.Vialaret, SPIE, 2470, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing VI, 386 (1995)
(2) Supplier: Piezotech/ISL, BP34, F-68301, Saint louis, France
(3) F.Bauer, French patent 8221025; US patent 4611260
(4) M.G.Broadhurst, G.T.Davis, J.E.McKinney, R.E.Collins, Journal of Applied Physics, 49, N°10, 4992, Oct.1978
and J.J.Simonne, Ph.Bauer, L.Audaire, F.Bauer, PROC. Workshop on the Technology of Ferroelectric Polymers, in press, Albuquerque NM, USA, Oct.17-19 (1995)
(5) Shu-Yau Wu, IEEE Trans. on ED, 1, 88 (1980)
(6) Ph.Bauer, L.Audaire, Ph.Rambaud, M.Pirot, F.Bauer and J.J.Simonne, PROC. of the ESSCIRC'95 Conf., Lille, F, 330 (1995)
(7) R.A.Wood, SPIE, 2020 Infrared Technology XIX, 34 (1993)
(8) Ch.Hanson, H.Beratan, R.Owen, M.Corbin and S.McKenney, SPIE, 1735, *Infrared Detector: State-of-the-Art*, 17 (1992)
(9) N.Neumann, R.Kohler and G.Hofmann, PROC.Ferroelectric Polymer Shock and Dynamic Sensor workshop, St Louis, F, R126/94, 23.1 (1994)
(10) F.Bauer, R.A.Graham, M.U.Anderson, H.LeFebvre, L.M.Lee, and R.P.Reed, PROC 8th IEEE Int.Symp.on Application of Ferroelectrics ISAF'92, Greenville SC, USA, 273 (1992)
(11) B.Andre, J.Clot, E.Partouche, J.J.Simonne and F.Bauer, *Sensors and Actuators A*, 33, 111 (1992)

7134

Stereo-Based Region-Growing using String Matching *

230665

Robert Mandelbaum and Max Mintz

21

General Robotics and Active Sensory Perception (GRASP) Laboratory
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

October 7, 1995

Abstract

We present a novel stereo algorithm based on a coarse texture segmentation preprocessing phase. Matching is performed using string comparison. Matching substrings correspond to matching sequences of textures. Inter-scanline clustering of matching substrings yields regions of matching texture. The shapes of these regions yield information concerning objects' height, width and azimuthal position relative to the camera pair. Hence, rather than the standard dense depth map, the output of this algorithm is a segmentation of objects in the scene. Such a format is useful for integration of stereo with other sensor modalities on a mobile robotic platform. It is also useful for *localization*: height and width of a detected object may be used for *landmark recognition*, while depth and relative azimuthal location determine *pose*.

The algorithm does not rely on the monotonicity of order of image primitives. Occlusions, exposures, and foreshortening effects are not problematic. The algorithm can deal with certain types of transparencies. It is computationally efficient and very amenable to parallel implementation. Further, the epipolar constraints may be relaxed to some small but significant degree. A version of the algorithm has been implemented and tested on various types of images. It performs best on random dot stereograms, on images with easily filtered backgrounds (as in synthetic images), and on real scenes with uncontrived backgrounds.

1 Introduction

1.1 Motivation

A common deficiency among standard stereo algorithms is that they do not provide a *segmented* representation of the scene, with each region corresponding to a distinct object in the scene. This reduces the potential usefulness of such algorithms within the context of a mobile robotic system. A common assumption is that if a mobile robot is to use a stereo system as a sensing modality, it

*Portions of this research were supported by the following grants and contracts: ARPA Contracts N00014-92-J-1647, and DAAH04-93-G-0419; ARO Contracts DAAL03-89-C-0031PRI, and DAAL03-92-G0153; NSF Grants CISE/CDA-88-22719, IRI92-10030, IRI92-09880, IRI93-03980, and IRI93-07126.

is the task of other modules further along in the dataflow pipeline to segment the output (usually a dense depth-map) and extract from it whatever information is required by the system.

In contrast, we designed the stereo algorithm described in this paper with the *specific intention* of using the output on a mobile robot to aid in localization of the agent. Furthermore, we wished to use stereo in conjunction with other sensor modalities. In essence, we addressed the design of a stereo processing technique by (a) deciding what type of output would be most useful for the task of localization, and (b) taking into consideration how the modality would be integrated into the system as a whole.

For these reasons, we stressed the following attributes:

1. *Computational efficiency.*

2. *Predictive power:* If the data representation allows prediction of expected sensor measurements, then correspondence matching between extracted and stored features is facilitated. This supports localization. The dense depth-map output by standard stereo algorithms does *not* facilitate this prediction.

3. *Robustness:* The stereo algorithm should work well on real, uncontrived indoor images with ordinary high-textured backgrounds.

This paper describes a stereo algorithm which is computationally highly efficient, performs well on real scenes with highly textured backgrounds, and produces output in a form which is very compatible with other sensor modalities. In particular, the output of height, width, azimuthal location and range of an object in the scene is very useful for landmark recognition and localization.

1.2 Overview

In [5] it is pointed out that there are three major components of any stereo algorithm: preprocessing, matching and 3-D structure determination. This is based on the assumption that the data-flow consists of (i) an input of two or more images of the same scene from different vantage points, and (ii) an output of a dense depth-map of the scene. Algorithms differ in the type of preprocessing performed, and hence in the nature of the primitives upon with the matching is executed, in the type of matching, as well as in the post-matching 3-D reconstruction.

Many trade-offs are made. Several algorithms extract *features* from the images. Examples of typical features are single-edge points (often extracted by finding the zero-crossings of the convolution of the image with the $L \circ G$ operator)[1] [2, 7, 15], and linear edge segments extracted using some edge detector [1, 13]. Feature extraction reduces the number and ambiguity of the primitives to be matched, thus reducing the complexity of the matching problem. Feature-based methods also lead potentially to great accuracy since features can be located in each image to sub-pixel precision. On the other hand, feature extraction can itself be computationally expensive. Furthermore, the product of matching features is a sparse depth-map, which must then be interpolated. Not only can the interpolation be a difficult and computationally expensive task, but it can also lead to ambiguities in reconstruction and the blurring of the very edges which were used for matching.

Area-based correlation methods compare windows surrounding points in both images. Various metrics are used to evaluate how well the windows are correlated [6, 8, 14]; the windows around points on epipolar lines with greatest correlation values are deemed to match. While such an

[1] where L is some discretized form of the Laplacian (second derivative) and G is a Gaussian smoothing operator.

approach leads to a dense disparity map (in theory, *all* non-occluded pixels have associated disparities), the correlation process is very computationally intensive. In effect, the set of primitives is the set of all windows. A large set of primitives (i) necessitates a large amount of matching, and (ii) leads to frequent occurrences of ambiguities. A trade-off exists in selecting the window size: the window size must be large enough to include enough intensity variation for reliable matching, but small enough to avoid the effects of projective distortion [9]. Also, the larger the window size, the greater the computational expense. For $n \times n$ images and a window size of $w \times w$, the naive correlation-based stereo algorithm has complexity $O(n^3w^2)$. An effective adaptive windowing technique is discussed in [9]. Being pixel-based, the precision of area-based correlation methods is limited to pixel-sized discretization [3].

There are several algorithms which are pixel-based and yet do not rely on windowing techniques [4]. This allows them both to yield a dense disparity map and to be faster since no features need be extracted, nor are window comparisons necessary. Indeed, the algorithm described in this paper is of this type.

In general, the greater the distinctiveness of the features extracted, the less matching is required, but also the more time has to be spent in feature extraction, and the sparser the resultant disparity map.

In many streo algorithms, assumptions are made regarding the nature of the scene:

1. Many algorithms rely on the *monotonicity* assumption, i.e. that the ordering of primitives along epipolar lines is the same in both images. In fact, not only does this assumption not always hold, but it is violated in cases where the effects of taking two different views of a scene are greatest; it would seem that an algorithm exploiting stereo effects should *utilize* rather than *avoid* these cases.

2. In the interests of computational efficiency, many stereo algorithms limit the search for matches to small disparities. This is, in effect, assuming that all objects of interest are far enough away that disparities will not be large. Once again, it would seem such algorithms are avoiding the very effects upon which stereo is based. By limiting themselves to small disparities, such algorithms constrain the area of interest to relatively great depths, where errors are greatest [12, 3].

3. Most stereo algorithms have difficulty dealing with *occlusions* and *exposures*. In actual fact, *any* object which stands out from the background and thus differs from its background in disparity will cause part of that background to be occluded in one of the images. Moreover, any occlusion in one image corresponds to an exposure in the other. Once again, it would seem that occlusion and exposures are necessary and expected artifacts of stereopsis, and should be manageable, if not exploited, by stereo algorithms.

4. Many stereo algorithms assume a frontal planar nature to detected surfaces. This arises from the assumption of orthogonal rather than perspective projection, combined with a parallel-axis stereo geometry: under these conditions, the projections of a surface onto the two image planes are identical, regardless of the orientation of the surface; therefore, upon reconstruction, nothing more complex than frontal planar surfaces is justified. In reality, some information may be gleaned from the effects of foreshortening and the use of perspective projection in stereo.

5. During 3D reconstruction, some algorithms interpolate a dense depth-map from a sparse disparity map. Many such algorithms assume a continuous underlying "rubber" surface,

which has been stretched to pass through the detected depth points. Such a reconstruction model blurs and smooths the edges between objects in the scene; rather than facilitating the segmentation of the scene, clustering of points of similar depth into "objects" is made more difficult.

While these assumptions may indeed be valid in certain environments, the invocation of these assumptions detracts from the versatility and applicability of an algorithm. Moreover, it seems the previous assumptions all arise from a *single* underlying requirement: that the product of a stereo system should be a dense depth-map. It is assumed that if a system is to use a stereo system as a sensing modality, it is the task of other modules further along in the dataflow pipeline to segment this dense depth-map and extract from it whatever information is required by the system.

In this paper we present a stereo algorithm whose output is *not* a dense depth-map of the scene. In fact, the stereo modality is not used to yield depth information at all. In this work, we move in the opposite direction to the standard data-flow. Our stereo algorithm makes use of depth information acquired from ultrasound sensors to guide the search for correspondences. The output of our algorithm is a *partial segmentation* of the scene: at the very least, the algorithm yields information regarding the extents and azimuthal position in the scene of the particular "segment" (i.e. object) which reflected the ultrasonic energy. We believe the output of clustered, segmented data to be a novel aspect of stereo algorithm design. Such a format has proved to be very useful for the integration of stereo with other sensor modalities, especially for the purposes of landmark recognition and localization.

Several other attributes of the stereo algorithm presented in this paper are:

1. The *monotonicity* constraint is not inherently necessary: a string-matching algorithm which handles transposition can be used. However, in our current implementation, we do not make use of this capability.

2. Highly textured, or easily filtered backgrounds (as in synthetic images) are preferred.

3. Since the output of the algorithm is *not* a dense depth-map, occlusions and exposures do not, in general, hinder the operation of the algorithm.

4. Foreshortening effects can be detected by the algorithm, and may, therefore, be exploited for the reconstruction of surfaces at a non-zero angle to the image plane. We do not look for forshortening effects in our implementation.

5. The algorithm can deal with certain types of transparencies.

6. The algorithm is computationally efficient.

7. The algorithm is very amenable to parallel implementation.

8. Preliminary experimentation has shown the approach to be qualitatively robust to slight relaxation of the epipolar constraints.

1.3 Basic operation

The basic operation of the algorithm is described in detail in Section 2. Like most stereo algorithms, it is divided into three stages:

1. In the pre-processing phase, each image is segmented very coarsely according to texture. Texture-segmentation here includes color-segmentation and intensity-based segmentation, among others. Each texture region is then labeled with a letter from the alphabet of possible textures $\mathcal{T}$.

2. In the matching phase, the two *strings* of texture labels associated with epipolar scanlines in the two images are compared. Matching substrings are extracted. Each substring corresponds to a sequence of texture labels, regardless of whether members of that sequence have been foreshortened or not. In fact, once matches have been established, the pixel widths of corresponding texture regions may be compared; a change in region width indicates either foreshortening or occlusion. Some higher-level reasoning system may be used to disambiguate the two cases.

   A string matching approach has several advantages: By the nature of string matching, the uniqueness constraint[2] is propagated automatically. If long substrings are searched for first, cohesivity is stressed, possibly at the expense of the number of total matches. Occlusions and exposures in the scene correspond to string deletions and insertions respectively, and cause no problems. Non-monotonicity of order of image primitives corresponds to substring transposition, and is manageable by most sophisticated string matching algorithms. Any of a host of new string matching algorithms developed for use with genetic data may be called upon for the efficient execution of this phase. Finally, since each scanline is processed independently, this phase is amenable to parallel implementation.

3. In the 3D reconstruction phase, we cluster matching substrings over multiple adjacent scanlines. Substrings beginning or ending at approximately the same horizontal location in the image plane over multiple scanlines are clustered together.In this way, objects consisting of similar texture patterns are segmented. Since we are interested in properties of the multi-scanline *segment* and not each *scanline*, the algorithm allows for a certain amount of relaxation of the epipolar constraint. Misalligned scanlines will simply result in the top or bottom of a region being in error.

1.4 Domain of applicability

The algorithm described in this paper has been implemented and tested on various types of images including:

- random dot stereograms,
- synthetic "blocks world" images with controlled backgrounds,
- real images of indoor scenes with controlled (low texture) backgrounds, and
- real images of indoor scenes with uncontrived (highly textured) backgrounds.

Several of the image pairs and the resultant output of the algorithm are shown in Section 3. In general, since the algorithm treats the background in exactly the same way as the foreground objects, it performs better on more highly textured backgrounds. For cases where the background is less textured than the foreground objects of concern, and where foreground objects are sparse,

[2]The uniqueness constraint states that each primitive in the left image can be matched to only one primitive in the right image.

most of the matching is performed on the background, and the foreground objects do not stand out in the resulting set of matches. Matches on a low-texture background do not generally yield accurate disparity estimates since there are no texture changes on which to fixate; background pixels match background pixels with many different disparities. Note that low texture causes problems only for the *background*; low-textured foreground objects will still result in reliable string matches.

For this reason, the algorithm seems to perform best on random dot stereograms, on synthetic images for which the background can be filtered out, and on real scenes with uncontrived backgrounds. See Section 3 for examples. We are currently investigating the application of this algorithm to mobile robot localization.

1.5 Related work in stereo

Reference [5] presents a comprehensive survey of recent developments in establishing stereo correspondence for the extraction of the 3D structure of a scene. In particular, we mention here the work of Lim and Binford [10], Ohta and Kanade [15], and Cox et al [4], with which our work shares some similarities.

In [10], preprocessing of each image consists of edgel (edge elements) detection. Edges are linked into connected edges and curves. Surfaces are identified by boundary-tracing, and bodies are identified as groups of surfaces that share edges. Matching is attempted at the highest level. Results of matching are propagated to each successive lower level (surfaces, curves, edgels) [5]. As is pointed out in [5], "the advantage of this hierarchical stereo system is that the depth-map obtained is already segmented and ready for surface interpolation." The algorithm described in this paper shares this desireable property. One of the differences, however, is that in this work, stereo matching is used in the region-growing phase.

In [15], the search for matches "is formulated as a path-finding problem in a 2D search space in which vertical and horizontal axes are the right and left scanlines, respectively" [5]. Dynamic programming is used to find the path which minimizes a cost function "based upon variances of gray-level intensities of the scanline intervals being matched." The results of the intrascanline search "are used to establish global consistency among matches achieved in neighboring scanlines using an *interscanline* search." Edge connectivity is used to impose a consistency constraint. Dynamic programming is also used in [4], though in that work a *maximum likelihood* cost function is optimized. Certain assumptions are made about underlying probability distributions. Several *cohesivity constraints* are imposed to guarantee "that solutions minimize the original cost function and preserve discontinuities" [4]. "The constraints are based on minimizing the total number of horizontal and/or vertical discontinuities along and/or between adjacent epipolar lines, and local smoothing is avoided."

There are several correspondences between the work described in this paper and those in [4] and [15]:

1. String matching is similar to the dynamic programming approach. A fundamental difference, however, is that for string matching, no cost or regularization function need be defined; rather, specific behavior can be guaranteed by the appropriate selection of string matching criteria. As an example, in [4], it is shown that more accurate results are obtained if the sum of horizontal discontinuities is minimized. The cost function is then modified to incorporate this criterion. Using string matching, this cohesivity constraint may be satisfied more directly by simply searching for long substring matches first; in effect, a set of matches

involving m_a *adjacent* pixels is deemed superior to a set of m_{nc} non-contiguous matches, even if $m_{\mathrm{nc}} > m_a$.

2. Differences exist in the approach taken for *inter-scanline* clustering. In [15] the problem is posed as that of finding the least-cost path in a 3-D search space [5]. In [4], the problem is formulated as that of minimizing the sum of horizontal *and vertical* discontinuities. Since "minimizing vertical discontinuities between epipolar lines cannot be performed by dynamic programming" [4], an approximation is obtained by using a relaxation technique to "minimize the *local* discontinuities between adjacent epipolar lines." In our implementation, matching substrings over multiple scanlines are compared; those with similar starting or ending locations in the image plane are deemed to belong to the same cluster.

Though the algorithm described in this paper is pixel-based, it differs from area-based correlation or sum of squared differeces (SSD) approaches such as those in [6, 8, 9, 14] in that no windowing is used. Similarly, though the preprocessing phase of coarse texture segmentation may be seen as a form of feature detection, this approach does not really fall within the genre of *feature*-based stereo algorithms such as those in [1, 2, 7, 13, 15], since we are using *strings* of these features for matching.

2 Stereo based on string matching

Our approach for extracting segmentation information using stereopsis comprises three phases: texture segmentation, string matching and inter-scanline clustering.

2.1 Texture segmentation

In the pre-processing phase, each image is segmented very coarsely according to texture. The purpose of the stereo algorithm is to "grow" these coarse segments into regions of similar texture *patterns* in both images. Possible types of texture-segmentation here include color-segmentation and intensity-based segmentation, among others. Each texture patch is then classified and labeled with a letter from the alphabet of possible textures $\mathcal{T}$. Ideally, each texture segment would correspond to a part of an object in a scene. Since foreshortening of texture patches is expected, ideally texture classification should also be invariant under scaling in the horizontal direction. Similarly, the texture classification should be invariant under small changes in illumination. See Figure 1 for an example image and the desired type of segmentation.

Let P be a texture-based segmentation operator which partitions an image into patches according to texture. Let T denote a texture classifier assigning one of $|\mathcal{T}|$ labels to each patch. Let R be a binary relation over elements of $\mathcal{T}$, $R \subseteq \mathcal{T} \times \mathcal{T}$, such that for any textures t_i, $t_j \in \mathcal{T}$, $(t_i, t_j) \in R$ implies that texture t_i is similar to texture t_j. In other words, t_i, $t_j \in \mathcal{T}$, $(t_i, t_j) \in R$ implies that a patch of texture t_i in one image may be considered to correspond to a patch of texture t_j in the other image.

Hence, an implementation of this phase of the algorithm consists of

- a texture segmentation algorithm $T \circ P$ capable of reliably segmenting a scene into texture patches and classifying each patch into one of $\mathcal{T}$ textures, and
- a similarity relation R defined below.

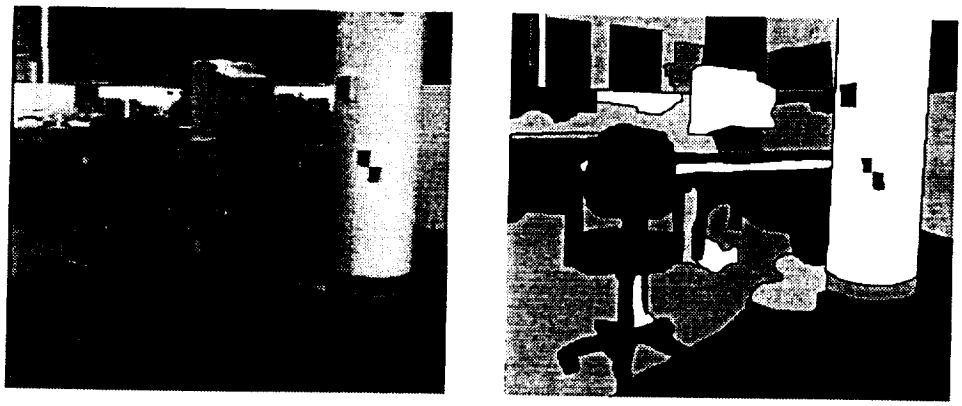


Figure 1: (Left) Example of scene to be analyzed using stereopsis. (Right) Idealized example of the type of texture segmentation suitable for stereo analysis by string matching (as described in this paper). This segmentation was obtained manually.

Since efficient region-growing based on texture is still an open problem, in our implementation we simplify the texture-segmentation phase and rely on the string-matching phase to "grow" the texture segments. We consider each pixel to be its own texture patch, and we discretize the only information we have about the pixel's "texture", i.e. its intensity, into $|\mathcal{T}|$ levels. Each texture patch (i.e. each pixel in our implementation) is therefore assigned one of $|\mathcal{T}|$ labels. Represent these labels with the first $|\mathcal{T}|$ letters of the alphabet. In our implementation, $|\mathcal{T}| = 16$. The coarseness of the discretization makes the texture classification robust in the face of small foreshortening and illumination effects. Though this is a simple "segmentation" requiring very little computation, the trade-off is that, for $n \times n$ images, each scanline consists of long strings of n texture patches, making the string matching phase more computationally intensive. The stereo algorithm is not inherently pixel-based, and the use of any reliable texture-segmentation algorithm yielding texture patches larger than a single pixel would enhance the performance of the string matching greatly. The specification of such a texture-segmenter is beyond the scope of this work.

Due to the simplicity of the texture space we selected for our implementation, the similarity relation R we chose is also simple: each texture is considered to match its immediate neighboring textures in the alphabet, and to be dissimilar to all other textures. A binary matrix representing R is shown in Figure 2.

If color images are available, a possible "texture" classification would consist of the discretization of 3-dimensional (Red×Green×Blue) color space into appropriately sized bins. A possible similarity relation R would then be

$$R = \{(t_i, t_j) \ : \ t_i \text{ is 26-adjacent to } t_j, \ t_i, t_j \in \mathcal{T}\} \ \cup \ \{(t_i, t_i) \ : \ t_i \in \mathcal{T}\}$$

The crucial attribute of any such $(T \circ P) - R$ combination is that T discretizes the space of textures sufficiently finely to distinguish textures, and yet R is sufficiently rich to allow matches among non-identical texture patches. This is necessary since noise in each image, foreshortening and illumination effects often result in the same surface in the scene being classified differently in the two images. In other words, if p_l and p_r are corresponding texture patches in the left and right images respectively, we do not insist that $T(p_l) = T(p_r)$, but merely that $(T(p_l), T(p_r)) \in R$. This is a much easier criterion to meet since R may be designed with knowledge of the behavior of $T \circ P$ and the types of scenes involved. Indeed, R may be changed for various scene and illumination types.

R	a	b	c	d	e	f	$\cdots$
a	1	1	0	0	0	0	
b	1	1	1	0	0	0	
c	0	1	1	1	0	0	
d	0	0	1	1	1	0	
e	0	0	0	1	1	1	
f	0	0	0	0	1	1	
$\vdots$							$\ddots$

Figure 2: Simple similarity relation R between elements of the texture alphabet $\mathcal{T}$. A "1" in position (i, j) indicates that the texture designated by the ith character in the texture alphabet $\mathcal{T}$ is deemed "similar" to the texture designated by the jth character. In this instantiation of R, all texture are considered similar only to themselves and to their immediate neighbors in the alphabet.

Area-based correlation approaches to stereo may be thought of as consisting of a classifier T which assigns a texture label to each pixel based on the surrounding window. Matches are determined by finding the *closest* point in texture space on the corresponding epipolar scanline. This relies on the underlying assumption that texture is numerically quantifiable, and that the distance metric used over texture space imposes a topology which adequately reflects reality. In our approach, the use of the binary relation R as the metric results in a very different topology. In effect the metric is of the $0-1$ type, and two textures are considered either to match or not. The relation R gives us much greater control over the topological neighborhoods of texture space, and hence over the model of reality it reflects.

2.2 String matching for each scan-line

In the matching phase, the two *strings* of texture labels associated with epipolar scan-lines in the two images are compared. Matching sequences of texture patches are extracted, even if some of those patches have undergone foreshortening; a difference in patch widths of corresponding textures indicates either occlusion or foreshortening.

The field of efficient string-matching alogrithms has received extensive attention in recent years in the context of genetics and the human genome project. Any of a host of new algorithms may be utilized for this phase of the stereo algorithm. Desireable properties of a selected string-matching algorithm are:

- Efficient handling of deletions and insertions. These correspond to occlusions and exposures in the image scanlines respectively.
- Efficient handling of substring transpositions. This corresponds to non-monotonicity in the order of matching primitives in the scene.

In our implementation, we use an algorithm which does not explicitly look for transpositions, but does produce substring matches in a form which is easily amenable to a search for transpositions. The string matching algorithm begins by looking for long substring matches, and if no such matches are found, progressively shorter and shorter substring matches are searched for. In this way, cohesivity rather than high pixel-match count is stressed.

Let `left` and `right` represent corresponding epipolar strings of texture labels in the left and right images respectively. Let the length of `left` and `right` be n. Let $\texttt{string}_i^m$ denote the

substring of `string` beginning at the ith location and ending at the mth. Denote the kth *character* of `string` by `string`$[k]$. Note that $\texttt{string}_i^n[k]$ denotes the same character as $\texttt{string}_1^n[i+k-1]$.

If a matching substring is found, say between the first p characters of $\texttt{left}_i^n$ and $\texttt{right}_j^n$, then the algorithm recursively searches for matches of length smaller than p between the substrings $\texttt{left}_1^{i-1}$ and $\texttt{right}_1^{j-1}$, as well as for matches of length p and smaller between substrings $\texttt{left}_{i+p}^n$ and $\texttt{right}_{j+p}^n$. The algorithm used to compare two strings `left` and `right` in our implementation is shown in Figure 3.

The output of the algorithm consists of the two strings with various substrings designated as matches. Transpositions may be taken into account in a second pass: matches may searched for among unmatched portions of `left` and `right` which have not yet been compared.

If a naive approach is used for the string comparison function `string_match`, the complexity of algorithm in Figure 3, per scanline, ranges from $O(n)$ in the best case (complete match) to $O(n^3)$ in the worst case (no matches found). In our implementation, we employ a dynamic programming approach in order to expedite the `string_match` function in the worst case scenario.

Let M be a $n \times n$ matrix such that $M_{i,j}$ contains the length of the longest common prefix of $\texttt{left}_i^n$ and $\texttt{right}_j^n$. In other words, for R as in Figure 2,

$$\begin{aligned} \forall_{1 \le k \le M_{i,j}} \quad & \left(\texttt{left}_i^n[k],\, \texttt{right}_j^n[k]\right) \in R \\ \text{and} \quad & \left(\texttt{left}_i^n[M_{i,j}+1],\, \texttt{right}_j^n[M_{i,j}+1]\right) \notin R \end{aligned}$$

As an example, consider the strings `left = 'aaddbbe'` and `right = 'gcdbbdd'`. Then $\texttt{left}_3^7 =$ `'ddbbe'`, and $\texttt{right}_2^7 =$ `'cdbbdd'`. Since $(d, c) \in R$ and $(e, d) \in R$, the longest matching prefix of these substrings has length 5. Hence $M_{3,2} = 5$. For this example,

$$M = \begin{bmatrix} 0 & 0 & 0 & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 4 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Note that if $M_{i,j} > 0$, then $M_{i+1,j+1} = M_{i,j} - 1$. It is this property of M that we exploit to expedite the search for substring matches. If the matrix M for two strings `left` and `right` was known *a priori*, then the comparison of two p-length substrings of `left` and `right` would consist of a simple table look-up rather than the $O(p)$ character-by-character comparison naive approach; the overall complexity of a scanline match would drop to $O(n^2)$ in the worst case, a significant improvement. Unfortunately, the matrix M is *not* known a priori. In our implementation, however, we partially construct M as string matching progresses.

We begin by setting all values of M to -1. When searching for a substring match of length p between $\texttt{left}_i^{i+q}$ and $\texttt{right}_j^{j+r}$, where $q, r \ge p$, we consult the current value of $M_{i,j}$. If this is still -1, then these substrings have not been compared before. We therefore compare the substrings and find the true value of $M_{i,j}$. $M_{i,j}$ cannot be larger than p since if it was, then we would have found a match while searching for longer substring matches. If $M_{i,j} = p$, then we have located a match. If $M_{i,j} < p$, we update M as follows: for each value of $0 \le k \le p$, set $M_{i+k,j+k} = M_{i,j} - k$.

If upon consulting the current value of $M_{i,j}$, we find that it is *not* equal to -1, then these substrings have been compared before; there is no need to compare them again. Hence, though the update of M requires $O(M_{i,j})$ operations, it potentially saves $O(M_{i,j}^2)$ character comparisons

```
match_strings(left, right, left_length, right_length)
   max_length = min(left_length, right_length);
   if (max_length == 0) {
      /* Base case */
      record_match(left, right, 0);
      return (0);
      /* 0 matches found */ }
   else {
      p = max_length;
      while (p > minimum_match_size) {
         i = 0;
         while (i <= left_length - p) {
            j = 0;
            while (j <= right_length - p) {
               if (string_match(left_i^{i+p}, right_j^{j+p}, p) == TRUE) {

                  /* Recursively search for matches between left_1^{i-1} and right_1^{j-1} */
                  pre_match = match_strings(left_1^{i-1}, right_1^{j-1}, i-1, j-1);

                  record_match(left_i^{i+p}, right_j^{j+p}, p);

                  /* Recursively search for matches between left_{i+p}^n and right_{j+p}^n */
                  post_match = match_strings(left_{i+p}^n, right_{j+p}^n, n-i-p+1, n-j-p+1);

                  return (pre_match + 1 + post_match);
                  /* Number of matches found */ }

               j = j+1; }
            i = i+1; }
         p = p-1; } }
```

Figure 3: The algorithm used to compare two strings `left` and `right` in our implementation. The ranges over which i, j and p vary may be changed to take into account minimum possible disparities between the left and right images.

```
string_match(left_i^{i+q}, right_j^{j+r}, length)
   if (M_{i,j} > -1) {
      /* We have already tried a match starting at these points before */
      if (M_{i,j} ≥ length) return (TRUE);
      else                  return (FALSE); }
   else {
      /* Must do the comparison. If successful, there is no need to update
         M since we will not be accessing these portions of the strings again */
      count = 0;
      while (count < length AND (left_i^{i+q}[count+1], right_j^{j+r}[count+1]) ∈ R) {
         count = count + 1; }
      if (count == length) return (TRUE);
      else {
         /* Update M */
         m = 0;
         while (m ≤ count) {
            M_{i+m,j+m} = count - m;
            m = m+1; }
         return (FALSE); } }
```

Figure 4: An efficient algorithm for the `string_match` function which checks whether strings $\texttt{left}_i^{i+q}$ and $\texttt{right}_j^{j+r}$ share a common prefix of length `length`.

in subsequent (shorter) substring comparisons. Efficiency is improved significantly. Hence, an efficient algorithm for the `string_match` function is shown in Figure 4.

2.3 Inter-scanline clustering

Once matching substrings, and their associated disparities, have been found, a 3D reconstruction of parts of the scene may be performed: each substring match corresponds to a horizontal strip of an object in space. By analyzing the distortion in pixel width of each "character" (i.e. texture patch) in the match, foreshortening or partial occlusion may be inferred. In our implementation, however, since each primitive texture patch is a single pixel, foreshortening is more difficult to detect.

In our implementation, we choose to cluster matching substrings over multiple scanlines by examining their disparities, as well as their beginning and ending points: substrings with similar disparities and beginning or ending at approximately the same horizontal location in the image plane over multiple scanlines are clustered together. The output is a polygon in the image plane circumscribing all contributing substrings. The polygon is assigned a disparity corresponding to the average disparity of the contributing substrings. Thus, each polygon corresponds to a region of corresponding texture *patterns* of similar disparity in the left and right images. For calibrated cameras, the polygon may be projected back into 3D space, and corresponds to a section of a frontal planar surface. This approach blurs information concerning disparity variation *within* each region. However, for our purposes — the extraction of objects' extents and azimuthal position

within the scene — such information is not necessary.

One advantage of this approach to inter-scanline clustering is that it can handle certain types of transparencies. Consider, for example, a scene comprising an object behind a Venetian blind. Each scanline either contains the object, or one of the blades of the Venetian blind. In either case, matching substrings are extracted, though the set of disparities will be easily partitioned into two subsets: those corresponding to the depth of the object, and those corresponding to the depth of the Venetian blind. The output of the algorithm in this case will consist of two polygons: one circumscribing the object, the other outlining the Venetian blind. Rotation of the Venetian blind by 90 degrees will cause a completely different effect: Each blade of the Venetian blind will now be segemented into its own region, and the object will similarly be dissected into "vertical strips" See Section 3.2 and Figure 7 for experiments involving synthetic images of Venetian blind scenes.

3 Experiments

The algorithm described in this paper has been implemented and tested on various types of images including:

- random dot stereograms,
- synthetic "blocks world type" images with controlled backgrounds,
- real images of indoor scenes with controlled (low texture) backgrounds, and
- real images of indoor scenes with non-contrived (highly textured) backgrounds.

In general, since the algorithm treats the background in exactly the same way as the foreground objects, it performs better on more highly textured backgrounds. For cases where the background is less textured than the foreground objects of concern, and where foreground objects are sparse, most of the matching is performed on the background, and the foreground objects do not stand out in the resulting set of matches. Matches on a low-texture background do not generally yield accurate disparity estimates since there are no texture changes on which to fixate ; background matches background with many different disparities. For this reason, random dot stereograms, synthetic images for which the background can be filtered out, and real scenes with uncontrived backgrounds are most suitable for application of our algorithm.

3.1 Random dot stereograms

In order to test the correctness of the stereo matching algorithm, the algorithm was tested on a random dot stereogram pair. Figure 5 shows the output generated. The output is in the form of a polygon circumscribing the region deemed to match. Note the jagged vertical edges on both sides of the polygon. Since the algorithm detects matches by way of string comparison, "extra" pixels on either end of the "genuine" shifted string will often be included in the match: there is a 50% chance of a single random pixel matching, a 25% chance of two consecutive pixels matching, etc. Once these random artifacts of random dot stereograms are taken into account, the algorithm is seen to perform flawlessly, correctly matching 100% of the shifted pixels.

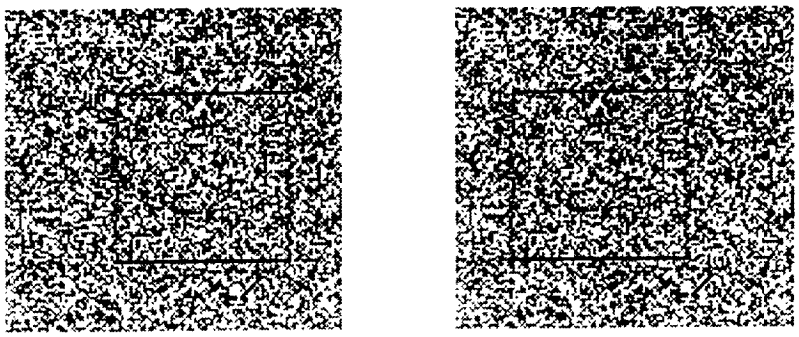


Figure 5: A random dot stereogram pair and the output of the algorithm in the form of a polygon circumscribing the matching region. The jagged vertical edges are a result of random "extra" matching pixels adjacent to the shifted region. 100% of actual shifted pixels are identified. Processing time for these 125 × 120 images, including the clustering into the polygon, is approximately 20 seconds on a Sparc 10.

3.2 Synthetic images

Figure 6 shows a pair of blocks world images of an assortment of objects. Directly below the images is the output of the algorithm where no distinction has been drawn between background and objects; the longest matching strings therefore comprise mostly background, and swamp the output. The next image illustrates the output once the algorithm has been instructed to filter out the dark background and black "floor". Regions correspond very well to the objects. The last pair of images shows the polygons thus found superimposed on the original images to show how the regions match the underlying images.

Figure 7 illustrates the ability of the stereo algorithm to handle certain restricted types of transparency. The top figure shows a rectangular block behind a horizontal Venetian blind. The output of the algorithm is shown below it: two regions are found, each corresponding to one of the objects in the scene. The numbers next to each polygon are the disparities associated with the region: The Venetian blind has a disparity of 18 pixels, whereas the more distant block has been successfully segmented at a lower disparity of 8 pixels, despite the foreground occlusion by the Venetian blind.

When the Venetian blind is placed vertically, the algorithm fails to cluster all portions of the block into a single region. The block has been "dissected" into three vertical strips. Due to large perspective differences between the two images, the blades of the Venetian blind are not found.

3.3 Real images

3.3.1 Non-textured background

Figure 8 shows two sets of real images involving a chair against a relatively low-texture background. In the first set of images, the illumination was from above; in the second set of images, an additional illumination source was placed facing the chair so that a shadow of the chair was cast on the background wall. In both sets of images, the output polygons of the algorithm are superimposed on the images.

In both sets of images, the non-textured background is seen to swamp the output: the largest polygons correspond to the background wall or curtain, or to the floor. Since these are real images,

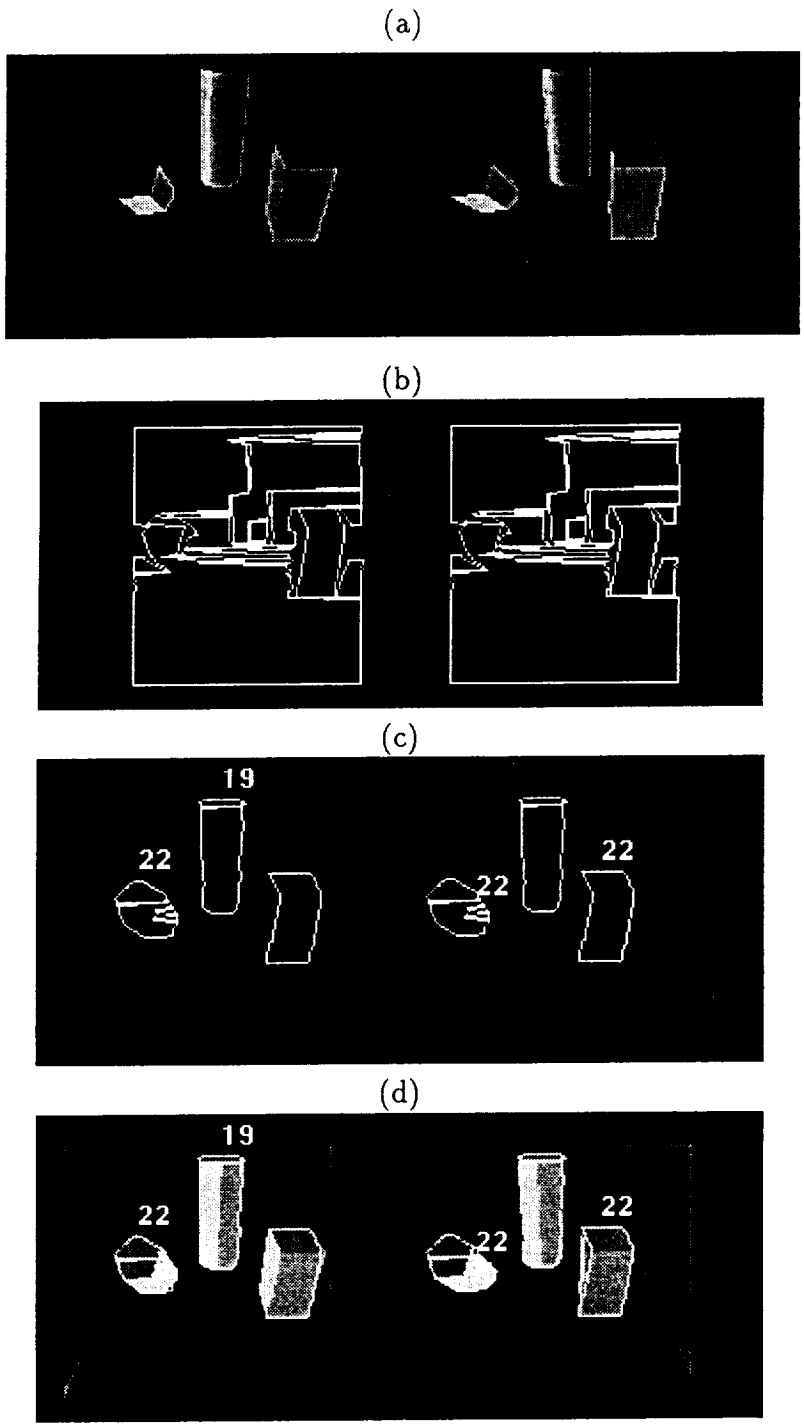
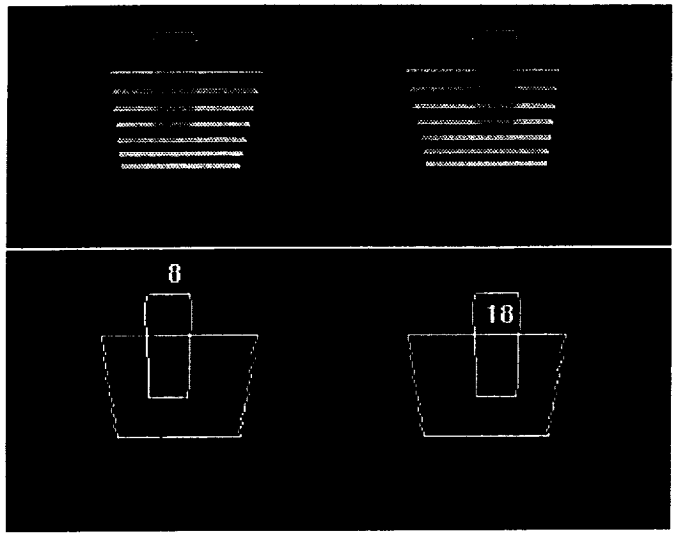
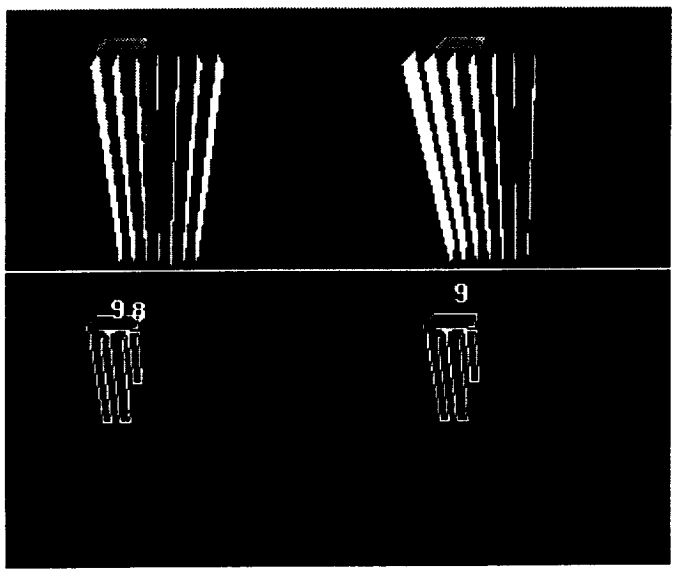


Figure 6: (a) Left and right blocks world images of an assortment of objects. (b) Output of algorithm on images including background and floor. (c) Output of algorithm on objects only (background and floor filtered out). Numbers represent average disparity for each region: more distant cylinder has lower disparity than front two objects. (d) Corresponding regions superimposed on the images.



(a)



(b)

Figure 7: Example illustrating ability to handle certain types of transparency. (a) Left and right images of block behind *horizontal* Venetian blind, and corresponding output: two regions are found, one corresponding to the Venetian blind (average disparity 18), and the other to the more distant block (average disparity 8). (b) Left and right images of block behind *vertical* Venetian blind, and corresponding output: the block has been "dissected" into three vertical strips of average disparities 8, 9 and 9. Large perspective differences between the two images precludes the matching of the blades of the Venetian blind.

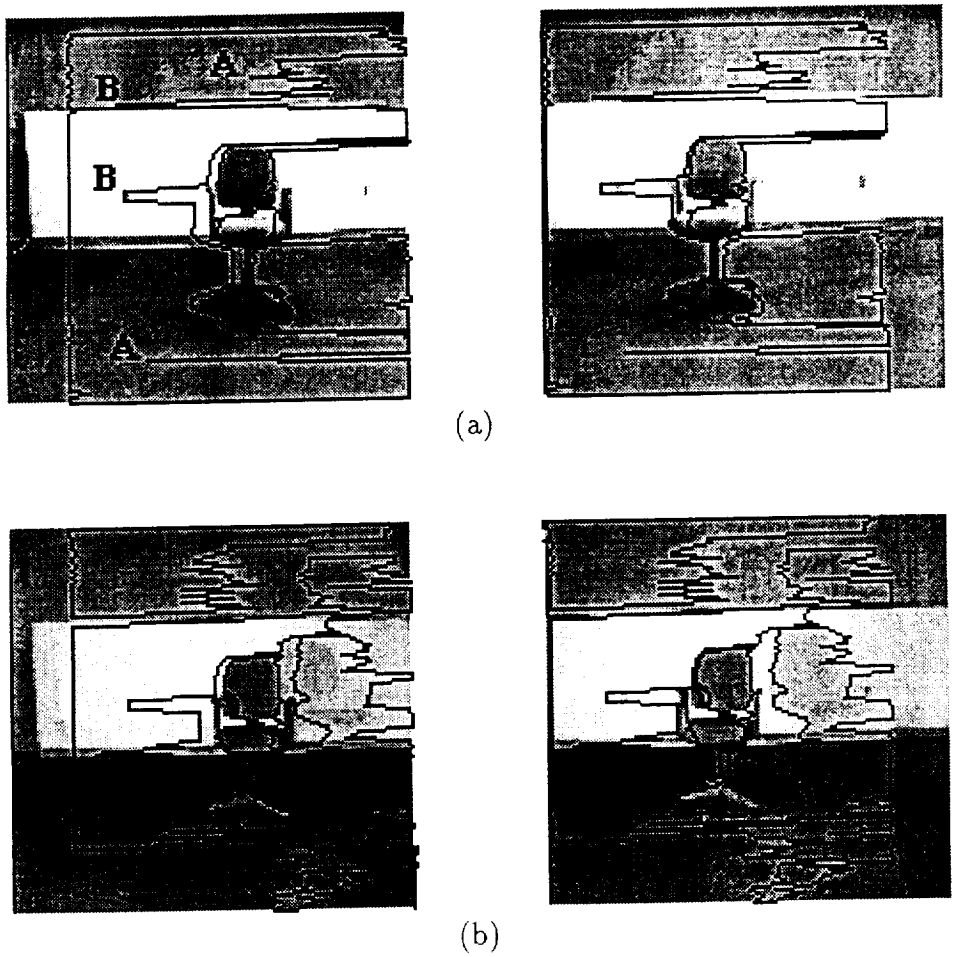


Figure 8: (a) Real left and right images of a chair against a controlled (low-texture) background, and the associated output. Though the shape of the chair is segmented, the largest regions correspond to the background. Since the intensities of objects and background are similar in real images, it is much more difficult to filter out background before processing than in synthetic images. (b) Real left and right images of a chair against a controlled (low-texture) background, and the associated output. Additional illumination from the front casts a shadow of the chair on the wall. Both chair and shadow are segmented, though the output is swamped by matches of the background.

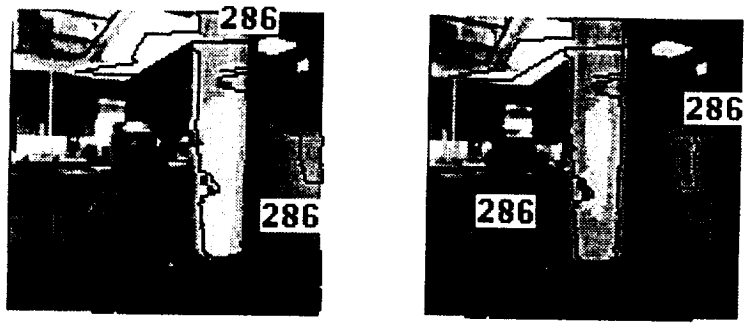


Figure 9: Output for a pair of real images with a non-contrived background. All regions are found to be portions of a frontal planar surface at distance 286 centimeters. Most notable region corresponds to the column. Examine the distances of the extracted region from the portion around the checker-board pattern at the lower left of the column: the algorithm has matched corresponding portions of the column, despite changes in perspective between the two images.

it is much more difficult to filter out the background and floor without also removing parts of the foreground objects of interest. Nevertheless, in the first set of images, the back and seat of the chair are segmented. Similarly, in the second set of images, both the chair and its shadow are clustered well into regions.

The errors in matching which are labelled "A" in the first set of images are due to noise in the images and differences in illumination between the left and right image. Those labelled "B" are due either to image noise, illumination effects, or to a misalignment of epipolar lines. The upper error occurs at a horizontal junction between dark and light areas, while the lower error is caused by a a small spot (electric socket) on the rear wall. The additional illumination in the second set of images creates greater contrast, and hence exacerbates these effects.

3.3.2 Textured background

Finally, Figure 9 shows the algorithm operating on a real set of images involving an non-contrived, textured background. Most of the regions corresponding to background objects do not exceed the minimum size requirement to be considered interesting. In fact, only five regions are displayed: two correspond to carpet, one to a patch of the side wall, one to the light fixture (matched with incorrect disparity), and one to the foreground column. The scene is similar to that of Figure 1, except that no chair is present.

The quality of the region corresponding to the column is significant: at first glance, it may appear that the algorithm has matched *different* portions of the column: in the left image, the region is aligned with the right-hand edge of the column, whereas in the right image, the region is aligned with the left edge of the column. However, the algorithm has, in fact, matched *corresponding* portions of the column very well. This can most easily be checked by examining the portion around the checker-board pattern on the lower left of the column: the polygon weaves around the pattern at approximately the same distance in both images. The reason for the *apparent* misalignment is that the two images show different perspectives of the column. The numbers in the image represent approximate depths; all regions were found with the same approximate disparity, and hence are shown with the same approximate depth of 286 centimeters. As mentioned in Section 2.3, each region is deemed to correspond to a portion of a frontal planar surface in

space.

4 Conclusions

We have presented a novel stereo algorithm based on a coarse texture segmentation preprocessing phase. We selected a very simple texture segmentation preprocessing phase for our implementation: each pixel defines its own texture patch, and its intensity level is used to infer a texture label. More sophisticated texture segmentation would enhance the performance of the algorithm considerably. The use of texture *labels* rather than *values* allows greater versatility in the choice of the similarity relation between textures, and hence in the choice of topology in texture space: instead of the relation imposed by the usual metric on real numbers, any similarity relation may be specified.

Matching is performed using string comparison. A string matching approach has several advantages: By the nature of string matching, the *uniqueness* constraint is propagated automatically. If long substrings are searched for first, cohesivity is stressed, possibly at the expense of the number of total matches. Occlusions and exposures in the scene correspond to string deletions and insertions respectively, and cause no problems. The *monotonicity* constraint is not inherently necessary: a string-matching algorithm which handles transposition can be used. Any of a host of new string matching algorithms developed for use with genetic data may be called upon for the efficient execution of this phase. Finally, since each scanline is processed independently, this phase is amenable to parallel implementation.

Matching substrings correspond to matching sequences of textures, even in the presence of foreshortening. By analyzing the widths in the image plane of matching members of a sequence, foreshortening effects can be detected. In theory, these effects may be exploited for the reconstruction of surfaces at a non-zero angle to the image plane. In practice, however, it is difficult to disambiguate foreshortening from occlusion.

Inter-scanline clustering of matching substrings yields *regions* of matching texture. The shapes of these regions yield information concerning objects' heights, widths and azimuthal positions relative to the camera pair, while the average disparity of pixels in these regions may be used to estimate the objects' distances. Hence, rather than the standard dense depth map which must still be segmented and further processed, the output of this algorithm is a partial description of *objects* in the scene. We believe the output of clustered, partially segmented data by a stereo algorithm to be novel. Such a format has proved to be very useful for the integration of stereo with other sensor modalities. In particular, we are currently investigating the utility of this approach, in conjunction with other modalities, for pose estimation and mobile robot localization [11].

The algorithm can deal with certain types of transparencies. Further, the nature of the approach permits a slight realxation of the epipolar constraints. Furthermore, the algorithm is computationally efficient: the particular version of the algorithm which was implemented for the experiments in this paper uses a dynamic programming approach to keep the complexity somewhere between $O(n^2)$ (best case) and $O(n^3)$ (worst case) for an $n \times n$ image, *including the time required for intra-scanline clustering.* This compares favorably with the complexity for window-based correlation or SSD methods ($O(n^3w^2)$ for w-sized windows and for *unclustered* output). Comparison of efficiency with feature-based approaches is difficult, since the complexity of a feature-based approach depends on the type of feature detection performed.

The speed of the algorithm may be greatly enhanced by integration with other sensing modalities. In our implementation, for example, the ultrasound modality is used to measure depth of an object of interest. This measurement is used as to infer approximate disparity, and hence to

guide the search for substring matches. The output of the stereo algorithm is *not* the depth of the object, which has already been measured accurately by the ultrasound modality. The output is, instead, the physical extents (width and height) and azimuthal location (relative to the camera pair) of that object of interest. This information, in conjuction with data from other sensor modalities, is very useful for landmark recognition and localization of a mobile robot. We are currently investigating this area of application.

The version algorithm has been tested on random dot stereograms, synthetic "blocks world" images with controlled backgrounds, indoor real images with controlled (low texture) backgrounds, and indoor real images with uncontrived (highly textured) backgrounds. The results are encouraging, with 100% successful matching on shifted pixels in the random dot stereogram, and good qualitative segmentation of images with easily filtered backgrounds (such as in synthetic images). Perhaps most significant, however, is the successful segmentation of foreground objects against a highly textured backgound in real uncontrived scenes.

References

[1] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *International Journal of Computer Vision*, pages 107–131, 1987.

[2] H. Baker and T. Binford. Depth from edge and intensity based stereo. *Proc. of 7th Int. Joint Conf. Artificial Intell.*, pages 631–636, August 1981.

[3] S. Blostein and T. Huang. Error analysis in stereo determination of 3-d point positions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(6):752–765, November 1987.

[4] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. A maximum likelihood stereo algorithm. *Submitted to CVGIP*, 1994.

[5] U. R. Dhond and J. K. Aggarwal. Structure from stereo — a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(16), November/December 1989.

[6] D. Gennery. Object detection and measurement using stereo vision. In *Proceedings of the 1980 Image Understanding Workshop*, pages 161–167, April 1980.

[7] W. Grimson. Computational experiments with a feature-based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, January 1985.

[8] M. Hannah. Bootstrap stereo. In *Proceedings of the 1980 Image Understanding Workshop*, pages 201–208, April 1980.

[9] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, September 1994.

[10] H. S. Lim and T. O. Binford. Stereo correspondence: A hierarchical approach. In *Proceedings of the 1987 Image Understanding Workshop*, pages 234–241, February 1987.

[11] R. Mandelbaum. *Sensor Processing for Mobile Robot Localization, Exploration and Navigation.* PhD thesis, University of Pennsylvania, 1995.

[12] L. Matthies and S. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239–248, June 1987.

[13] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, Image Processing*, 31:2–18, 1985.

[14] H. Moravec. Towards automatic visual obstacle avoidance. *Proc. of 5th Int. Joint Conf. Artificial Intell.*, page 584, 1977.

[15] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, March 1985.

Session 4: Microinstruments (Tuesday, AM)

Session Chairs: Bart van der Schoot (U. Neuchatel, Swiss) and William Trimmer (Belle Mead Research)

Little Spaces for Outer Space (Invited)
by William Trimmer, Belle Mead Research, Inc.

Micro-Electromechanical Instrument and Systems Development at the Charles Stark Draper Laboratory
by J. H. Connelly, J. P. Gilmore, and M. S. Weinberg, C. S. Draper Laboratory, Inc.

PZT Thin Film Piezoelectric Travelling Wave Motor
by Shen Dexin, Zhang Baoan, Yang Genqing, Jiao Jiwei, Lu Jianguo, and Wang Weiyuan, State Key Laboratories of Transducer Technology, Shanghai Institute of Metallurgy

Demonstration of a Monolithic Micro-Spectrometer System
by S. Rajic and C. M. Egert, Oak Ridge National Laboratory

Silicon Micromachined Sensor for Broadband Vibration Analysis
by Adolfo Gutierrez, Daniel Edmans, Chris Cormeau, and Gernot Seidler, InterScience Inc.

A Microinstrumentation System for Remote Environmental Monitoring
by Andrew Mason, Wayne G. Baer, and Kensall D. Wise, University of Michigan

Spacecraft Applications of Compact Optical and Mass Spectrometers
by N. M. Davinic and D. J. Nagel, Naval Research Laboratory

High-performance Electronics at Ultra-Low Power Consumption for Space Applications: From Superconductor to Nanoscale Semiconductor Technology
by Robert V. Duncan, Sandia Corp.

Microrefrigeration by a Pair of Normal Metal/Insulator/Superconductor Junctions
by M. M. Leivo and J. P. Pekola, University of Jyväskylä,
and D. V. Averin, State University of New York (SUNY) at Stony Brook

Little Spaces for Outer Space

William Trimmer
Belle Mead Research, Inc.

It is difficult and expensive to work above the earth's atmosphere. Launch costs are huge, and once launched, the pay loads are difficult to access. Micromechanical devices have some inherent advantages in satellites.

Consider what happens if the size of a satellite system is decreased by a factor of a hundred. The mass and volume scale as the size to the third power, s^3, and the mass volume of this system decreases by a factor of a million. This is a substantial decrease in launch weight and volume requirements. Small objects also tend to be more robust and less easily damaged. Ants demonstrate this as they carry loads many times their weight, and can fall from heights without damage.

Because they are light, many micromechanical (or MEMS or MST) devices can be flown on a mission for less than the cost of one big satellite. These multiple systems allow redundancy and can improve the reliability of a mission.

Often there is excess launch capacity; the launch vehicle can orbit more than the weight of the planned satellite. Small satellites can tuck into this excess launch capacity, and perform auxiliary missions for smaller costs. Conversely, smaller and less expensive launch vehicles can be used for smaller satellites.

There are a range of new fabrication technologies that can inexpensively make large numbers of micro systems.

Some things scale less favorably into the micro domain. Power for example, small batteries have diminishing amounts of energy. Communications is a second example. One still needs antennas of sufficient apertures and sufficient power to communicate. Designing micromechanical space systems requires a total system approach to maximize the advantages of these small, light weight systems in space.

Key to all micromechanical applications in space is the greatly reduced size and mass of the micro systems. It is possible to build systems whose incremental costs to orbit are quire small. Using this concept to advantage, however, requires rethinking system philosophies.

In the conventional launch vehicles and systems, the failure of one small part can abort an entire mission. Hence the philosophy of "Zero Defects." Things are different for micro systems. Large numbers of micro systems can be included with a single launch to accomplish a myriad of interesting, but not system critical missions. There is not a major penalty if a few of these micro systems fail. A better philosophy for the plethora of micro hitch hikers is "Zero Cost."

To take advantage of the micro devices requires the proper support systems and infrastructure. Designers of interesting micro systems need easy access to power, communication, computing capabilities, and special requirements such as pressurized gasses and optical systems. To design this infrastructure for each micro device, reduces the cost advantage. What is needed is "standard services" that these micro systems can easily access.

Not only will the integration of micromechanical devices enhance our space capabilities, reduce costs, and speed the development, the incorporation of inexpensive microsystems and appropriate support systems will help bring other new technologies into space.

The intriguing job of designing micromechanical systems to help our space program is before us.

MICRO-ELECTROMECHANICAL INSTRUMENT AND SYSTEMS DEVELOPMENT AT THE CHARLES STARK DRAPER LABORATORY

7136

J.H. Connelly, J.P. Gilmore, M.S. Weinberg

The Charles Stark Draper Laboratory, Inc.
Cambridge, Massachusetts 02139

8p

Abstract

Draper Laboratory has been developing miniature micromechanical instruments for over 10 years, using and maturing silicon microfabrication techniques to achieve high yields in this batch processing environment. During this time, we have made considerable progress in the development and fabrication of micromechanical gyroscopes, accelerometers, and acoustic sensors. We have fabricated gyroscopes and accelerometers with dynamic ranges from 50 to 500 deg/s and 10 to 100,000 g, respectively. Bias stability of 33 deg/h and 0.55 mg has been demonstrated over a wide range of thermal and environmental conditions. In recent room temperature tests, 1°/h performance over a 0.1 Hz bandwidth, corresponding to 24°/h performance over 60 Hz, has been achieved. Our continuing development activities are expected to yield over an order of magnitude in performance enhancement. Draper builds its micromechanical instruments using a silicon wafer process that results in crystal silicon structures that are anodically bonded on a Pyrex (glass) substrate that contains sensing and control electrodes. This silicon-on-glass configuration has low stray capacitance, and is ideally suited for hybrid or flip-chip bonding technology.

Several generations of micromechanical gyros and accelerometers have been developed at Draper. Current design effort centers on tuning-fork gyro design and pendulous accelerometer configuration. Over 200 gyros of different generations have been packaged and tested. These units have successfully performed across a temperature range of -40 to 85°C, and have survived 30,000-g shock tests along all axes. Draper is currently under contract to develop an integrated Micromechanical Inertial Sensor Assembly (MMISA) and Global Positioning System (GPS) receiver configuration. Ultimate projections of size, weight, and power for an MMISA (after electronic design of the application-specific integrated circuit (ASIC) is completed) are 2 x 2 x 0.5 cm, 5 gm, and less than 1 W, respectively. This paper describes Draper's fabrication process, the current gyro and accelerometer designs, and system configurations.

Introduction

Draper has been developing miniature micromachined instruments for over 10 years, using and contributing to the maturation of silicon microfabrication technology. During this time, we have made extensive progress in the development and fabrication of micromechanical gyros, accelerometers, microphones, and hydrophones. In this context, we have fabricated gyros and accelerometers with dynamic ranges from 50 to 500 deg/s and 10 to 100,000 g, respectively. Performance resolution of 33 deg/h and 0.55 mg have been demonstrated over a wide range of thermal and environmental tests. Our continuing development activities are expected to yield an order of magnitude in performance improvement. Our acoustic sensors have demonstrated sensitivities that are significantly higher than comparably sized commercial units.

Consistent with the goal of transitioning technology to industry, Draper has entered into an alliance with the Rockwell International Corporation (RI). In 1993, the Draper/Rockwell alliance was consummated for the purpose of transitioning Draper-developed inertial micromechanical technology for production and commercialization. The proven performance and high-volume,

low-cost manufacturability achievements have demonstrated commercial viability, and Rockwell has established a manufacturing capability.

Initial RI products are targeted for automotive applications. Extensive investments have already been made in the development of this technology and its preparation for production. While RI's current primary emphasis is oriented toward large-volume applications, Draper remains committed to enhancing the performance of these designs and extending our micromachining capabilities to other instruments and applications. We currently provide, and will continue to seek, additional opportunities to apply our micromechanical development expertise to meet the unique needs of DoD and NASA. Toward this end, we believe our current instruments can be tailored to address miniature spacecraft objectives for GN&C and robotics. These capabilities can also be applied to a host of integrated payload instrumentation suites, as well as vehicle health monitoring systems.

This paper specifically details Draper's micromechanical silicon dissolved wafer fabrication process and describes and illustrates the current gyro and accelerometer development and fabrication status. Draper's achievements in micromechanical inertial systems technology has demonstrated a level of performance and manufacturability that warrants its consideration in Space and DoD applications. We are prepared to provide gyros, accelerometers, and inertial system development and fabrication activities that could result in near-term deliveries of components and systems for ground test and space flight demonstration. For example, under U.S. NAVY-NAVSEA sponsorship, Draper is developing an MMISA of three gyros and accelerometers, and is integrating it with a miniature GPS receiver for an Extended-Range Guided Munitions Demonstration Program (ERGM). We are interfacing these units with a TMS320C31 processor, and implementing software to perform attitude determination, instrument calibration, receiver aiding, and navigation, with an initial delivery scheduled for February 1996.

Fabrication Process

Draper builds its micromechanical instruments using a dissolved silicon wafer process that results in crystal silicon structures anodically bonded on a Pyrex (glass) substrate that contains the electrodes. Compared to conducting substrates, this silicon-on-glass configuration has low stray capacitance. Although on-chip electronics are not yet possible with P+ silicon on glass, this configuration is ideal for hybrid or flip-chip bonding technology.

The micromechanical instrument dissolved wafer fabrication process is illustrated in Figure 1. Reaction ion etching (RIE) and boron diffusion are used to define the final structure. As shown, the process starts with a silicon wafer of moderate doping. In the Mask 1 step, recesses are etched into the silicon using KOH. These recesses define the height of the silicon above the glass to allow gap spacing for the capacitive sensing plates and metal runs. A boron diffusion step follows, which defines the thickness of the structure. The structure's features (patterns) are then defined (Mask 2) and micromachined using RIE by etching past the diffused boron layers. This etching process results in straight side walls and high aspect ratios.

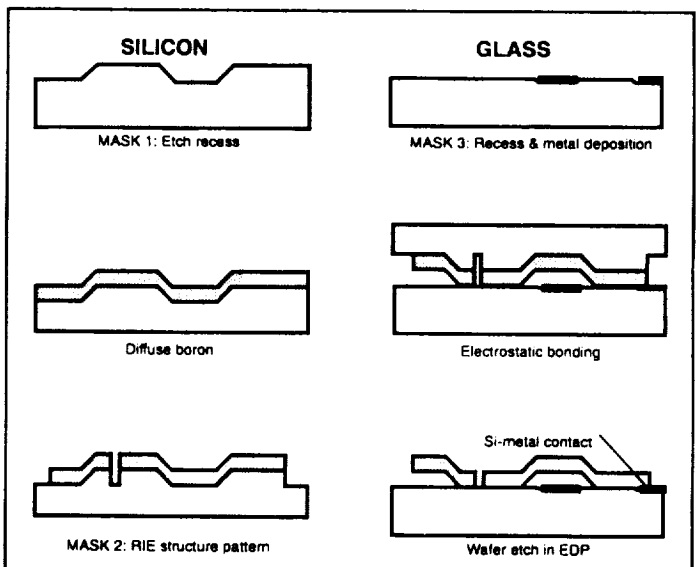


Figure 1. Dissolved wafer, silicon-on-glass process.

The glass processing (Mask 3) recesses the glass and then deposits and subsequently lifts off a multimetal system on a glass wafer. This results in a planar structure with metal electrodes and runs protruding only slightly above the glass surface. The metal forms the sense and drive plates of capacitor transducers and their output leads. The silicon and glass are then electrostatically bonded together. This electrostatic bonding process draws the silicon and glass tightly together to ensure a low-resistance contact. The final step corresponds to a selective etch in ethylene diamine pyrocatechol (EDP), which dissolves the undoped silicon and stops at the heavily boron diffused layers. This overall fabrication sequence requires only single-sided processing with 3 masking steps. Figure 2 illustrates the quality of the straight wall etching process.

Draper has achieved high yields using this batch processing technique. Hundreds of gyros are made on single wafer. Figure 3 (courtesy of RI) illustrates a tuning-fork gyro (TFG) wafer under probe test. Under Internal Research and Development (IR&D), Draper continues efforts to perfect these processing techniques to enable fabrication of higher performance instruments. Our fabrication capabilities will address and assess the future potential of micromechanical fabrication for a large variety of instrument applications.

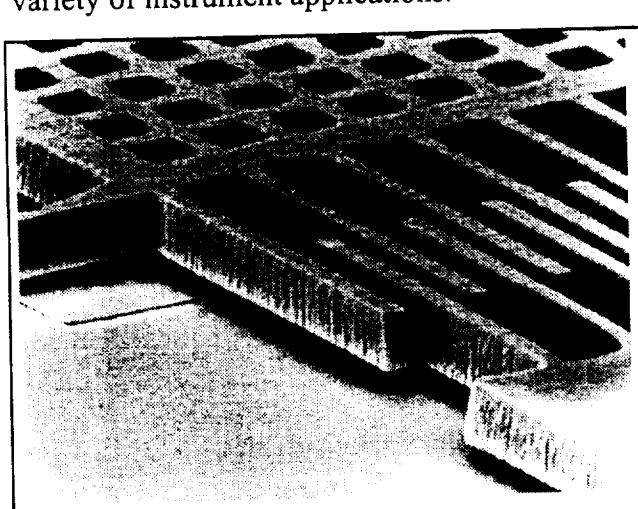


Figure 2. The comb structure on a TFG.

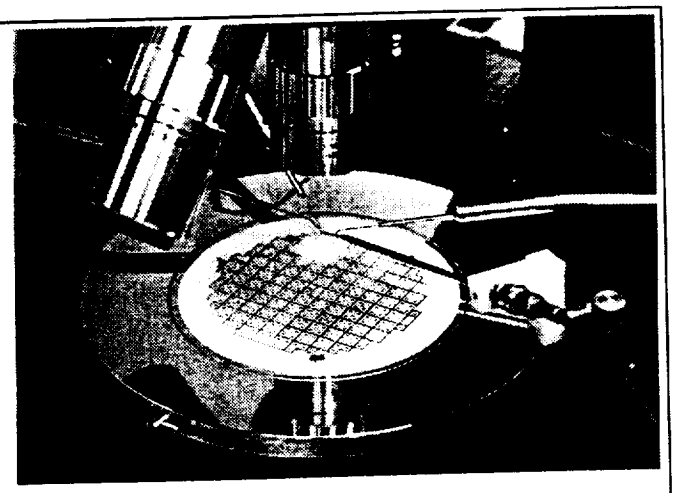


Figure 3. TFG wafer under test probe.

Micromechanical Inertial Sensors

Draper has developed several generations of micromechanical gyros and accelerometers. The current build process is described in the Fabrications section of this paper. This section describes our current micromechanical TFG and pendulous mass accelerometer designs.

Tuning Fork Gyro

The TFG's principle of operation and construction features are illustrated in the mechanical schematic shown in Figure 4. Both an in-plane (top view) and a cross-sectional (side view) are shown. The top view shows two vibrating mass members suspended by sets of flexural (struts) supports. The proof masses are vibrated in the plane of the structure by the operation of electrostatic forces applied through the interaction of the motor drive comb structures. An AC excitation is applied to the comb drives to sustain a lateral oscillation of the proof mass members. The resulting oscillation yields an in-plane peak velocity, V, that is a function of the drive frequency (f) and the peak amplitude of the vibration displacement (Y_A).

$$V \doteq 2\pi f Y_A \tag{1}$$

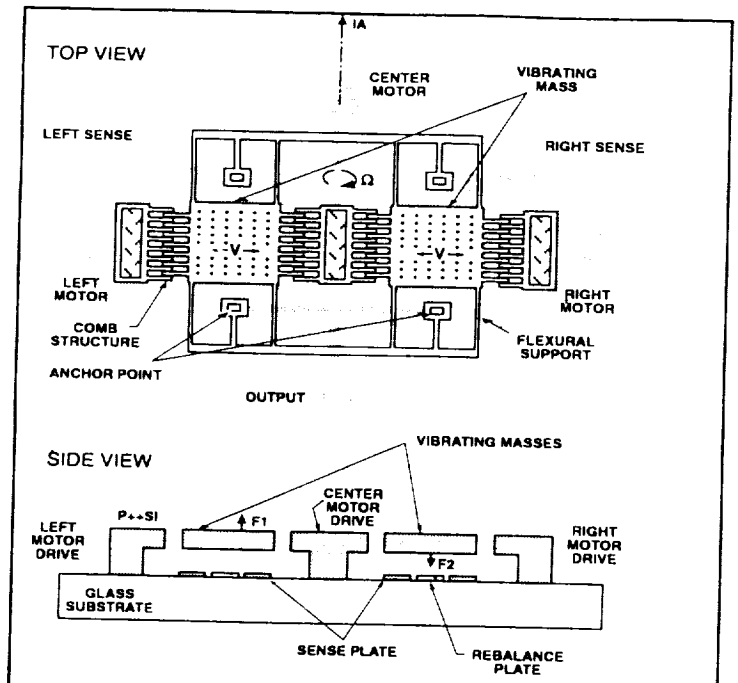


Figure 4. Micromechanical yaw rate sensor configuration.

The mass members are excited so that their velocities are 180 deg out of phase with respect to each other.

When an angular rate (Ω) is applied about the input axis (IA), one of the vibrating masses will lift up out of the plane and the other will move down due to Coriolis forces F_1 and F_2 (side view), respectively.

$$F = 2\Omega V m \qquad (2)$$

where m is the integrated mass of the vibrating member.

The capacitor electrodes (sense plates) below these proof masses sense this motion. Feedback through a control loop can apply voltages to the rebalance plate electrodes to provide nulling electrostatic forces. For lower cost and performance applications, open-loop operation is adequate.

A scanning electron microscope (SEM) photo of the TFG is shown in Figure 5. Figure 6 shows the device next to an ant for size comparison. The unit shown corresponds to the product of many design and test iterations. Over 200 gyros of different generations have been packaged and tested. Steady performance improvement has been achieved. These units have survived 30,000-g shock and centrifuge tests along all axes.

In open-loop, low-bandwidth tests, compensated bias stability of 33 deg/h has been demonstrated over a temperature range of -40 to +85°C. Figure 7 illustrates the open-loop voltage output of the TFG across a ±100 deg/s input angular rate. Scale-factor (SF) repeatability of better than 0.1% has

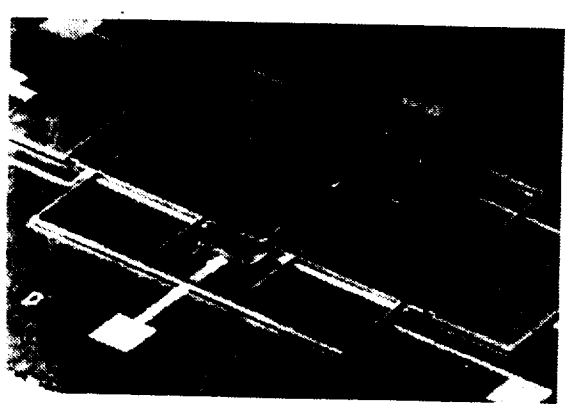


Figure 5. Micromachined comb drive TFG.

Figure 6. Silicon micromechanical gyro size comparison

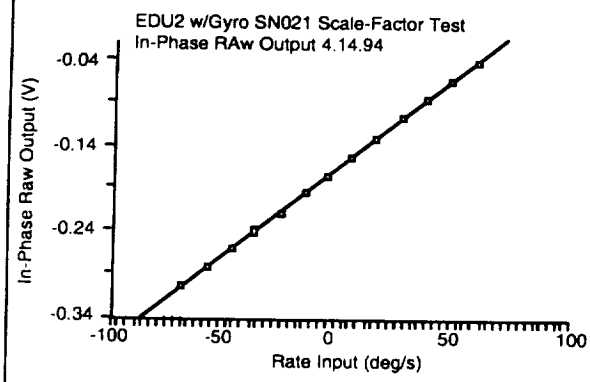


Figure 7. Gyro SF test.

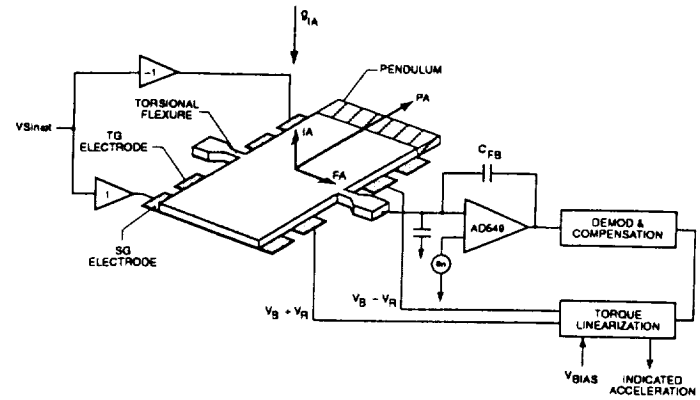


Figure 8. 100-g accelerometer schematic.

been achieved in testing across numerous temperature cycles and shutdowns. Continued TFG development and design enhancement is planned, and a long-term performance goal of 1 to 10 deg/h over -40 to +85°C is projected. Draper has made steady progress in the evolution of the micromechanical gyro with continuing improvements in performance. In recent room temperature tests, 1°/h performance over a 0.1 Hz bandwidth, corresponding to 24°/h performance over 60 Hz, has been achieved. Table 1 illustrates the performance progress that has been achieved from 1993 to date. A projected 1996 performance target of 0.25°/h is shown. Fabrication of a larger unit and improved sense electrode preamplifier is expected to provide significant performance enhancements.

Table 1. Micromechanical gyro performance history*.

DATE/STATUS		RMS DRIFT (deg/h, 0.1-60 Hz)	ARW (deg/√h)	DRIFT (0.1 Hz)	SF STABILITY (ppm)
5/93	Measured	1000	1.52	40.8	~150
6/94	Measured	500	0.76	20.4	<100
8/95	Measured	200	0.3	8.17	—
10/95	Measured	24	0.037	0.98	~50
11/96	Projected	6	0.009	0.25	~10

* Room temperature tests

Pendulous Accelerometer

A schematic of the accelerometer is shown in Figure 8. A simple mechanical design is employed. The acceleration-sensitive element corresponds to a single thickness proof mass that is suspended on a pair of micromachined torsional flexures. The pendulous effect is achieved by mounting the structure off center (one side is longer than the other). A pair of electrodes lies below the proof mass structure. Capacitive changes occur when the proof mass rotates about the flexure axis (FA) in response to acceleration inputs along the unit's input axis (IA). The gaps between the proof mass

structure and the SG electrodes change (one gap opens while the other closes). The resultant capacitive changes are proportional to the input acceleration, and a net current flows out of the proof mass flexure into the low-noise preamplifier. The flexure and proof mass structures are scaled to accommodate the desired g range . Units may be operated in an open- or closed-loop torque-to-balance mode. Closed-loop operation is achieved by electrostatic forces resulting from feedback voltages applied to the torque generator (TG) electrodes. Selection of open- or closed-loop operation is a function of cost considerations and accuracy requirements. Draper has configured units with three design ranges: 100,000, 100, and 10 g. For the 100,000-g unit, open-loop operation has been more than adequate. The high-g range development activity has been under U.S. Army sponsorship for kinetic energy projectile tests.

A SEM photo of the 100-g accelerometer is shown in Figure 9. As in the case of the gyro, the proof mass structure is perforated. These devices have undergone vibration and shock tests, and centrifuge testing has been performed to set the unit's SF and determine its linearity. Figure 10 shows a centrifuge test run on a 100-g accelerometer stack (with electronics) with a 100-Hz control loop. Due to centrifuge limitations, the SF test was constrained to a 22-g input. A 0.57% SF linearity was measured. Bias stability on the order of 0.55 mg, have been demonstrated in tests on the 100-g accelerometer.

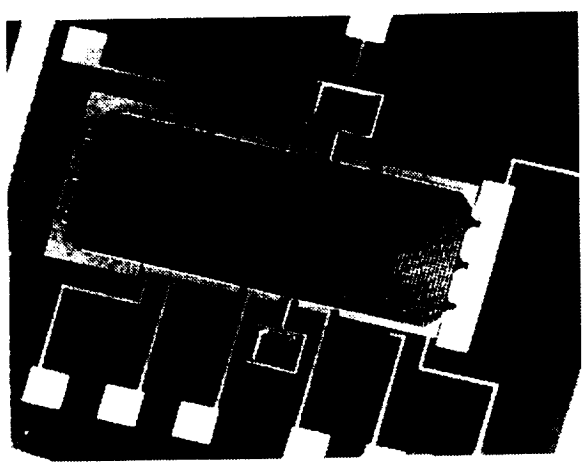


Figure 9. 100-g accelerometer.

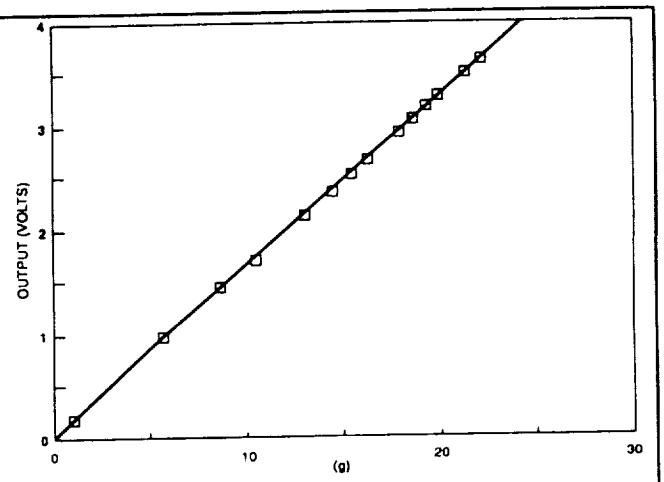


Figure 10. 100-g accelerometer stack output voltage vs acceleration input.

System and Device Development

As noted in the introductory section, Draper is currently developing an integrated MMISA/GPS for the ERGM Demonstration Program. Draper has also configured a number of single-axis gyro packages with hybrid electronics, and configured TFG units with RI-fabricated ASICs for test evaluation, and has delivered accelerometer units for high-g applications.

Representative sensor packages are shown in Figures 11 and 12. Figure 11 is a magnified photo of a TFG installed in a 1-in x 1-in hybrid single-axis package, and Figure 12 is a photo of a triad of microaccelerometers in a 1-in x 1-in hybrid package with preamplifiers. Device development could be oriented to provide further device performance enhancements or to tailor devices for relevant payload applications. For example, TFG devices could complement camera imaging sensors by providing wide-band rate stabilization signals or be used in a structural array to provide sensing for flexural mode sensing and damping.

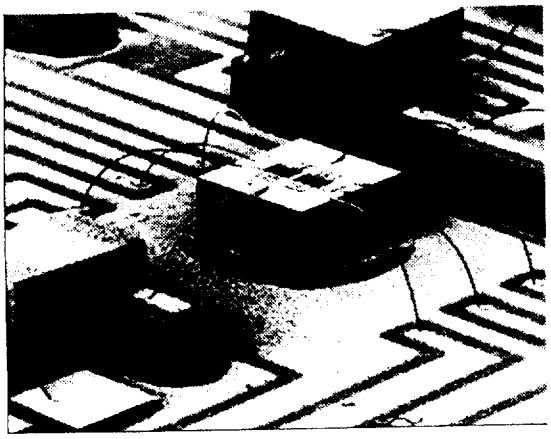


Figure 11. TFG in hybrid package.

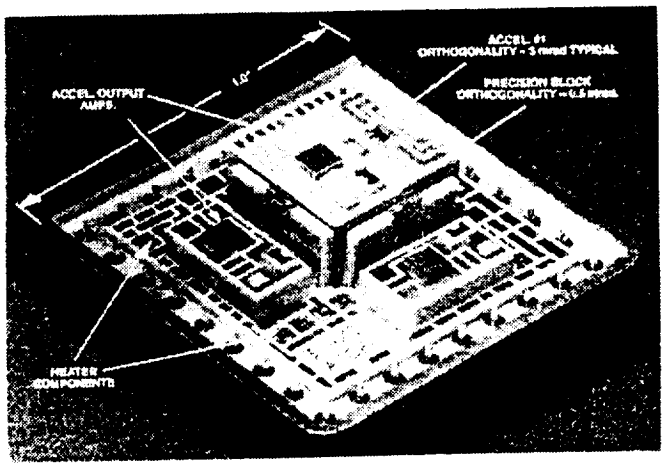


Figure 12. Three-axis 100-g accelerometer hybrid.

Conclusion

Draper Laboratory has made significant progress in the development of micromechanical instruments. Further design enhancements and high density packaging development are in process.

A gyro package with compensated performance on the order of 10 deg/h over -40 to +85°C with a ±100 deg/s dynamic range will be realized by early 1997.

ASIC technology design iterations and validations will be completed in mid-1996, and the goal of a fully miniaturized micromechanical inertial sensing system will be realized:

- Three axes of gyros and accelerometers.
- 2 x 2 x 0.5 cm package; 5-gm weight.
- Less than 1 W power.

Performance enhancements are expected as continued IR&D design improvements are realized. For example, currently, IR&D efforts have doubled the proof mass thickness and the drive amplitude. Improved fabrication techniques are also under evaluation.

- Gyro TFG performance goals:
  - ± 100 deg/s rate range.
  - 0.1% SF stability.
  - <1 deg/h drift stability.
- Accelerometer performance goals:
  - 100-g measurement range.
  - 0.05% SF stability.
  - 100-μg bias stability.

Micromechanical Development Facilities

Draper has a well-equipped 1200-ft^2 laboratory dedicated to micromechanical device fabrication. Diffusion, oxidation, photolithography, metallization, chemical vapor deposition (CVD), and plasma and wet etching are performed within the laboratory. Internal fabrication allows process

optimization for our sensors and maintains proper control over the development. The processing staff are experienced in both bulk and surface micromachining.

Facilities now include: 11 diffusion/oxidation tubes for oxidation, including low-pressure chemical vapor deposition (LPCVD) tubes for the deposition of polysilicon, silicon nitride, silicon dioxide, and phosphoro silicate glass; and two contact mask aligners, including a Karl Suss with infrared capability, allowing front-to-back side alignment of silicon wafers. Reactive ion etch and plasma CVD equipment were recently added to the microfabrication resources. A plasma oxygen asher is used to remove trace organic residues and photoresist. A dc magnetron sputtering machine allows up to three metals to be deposited in one pump-down. In-house computer-aided design (CAD) facilities are used to design photomasks and to perform finite-element analysis on all designs prior to fabrication.

System and component packaging is performed at Draper. Conventional die mounting and wire bonding facilities, and a hybrid assembly facility are available. Custom ceramic substrates are fabricated, and chip components can be assembled, along with the sensors, preamplifier, and electronics in the sensor package.

Electronics packaging facilities at Draper include CAD design for printed circuit boards, ceramic wiring boards, and multichip module wiring. Fabrication facilities include an extensive printed circuit prototype facility and a ceramic wiring board facility that can produce screened multilayer ceramic boards and multilayer boards using green tape processing. Draper also maintains a multichip module rapid prototyping facility that is compatible with industry manufacturing "foundries." Draper ASIC electronic design capabilities are mature, and many Draper designs have been manufactured in quantities for fielded systems using qualified high-reliability foundries. These facilities will be available for development activities in support of small spacecraft initiatives.

Draper References:

[1] *A Micromechanical INS/GPS System for Guided Projectiles,* D. Gustafson, R. Hopkins, N. Barbour, and J. Dowdle, Draper Laboratory, ION Proceedings 51st Annual Meeting, Institute of Navigation, Colorado Springs, CO, June 2, 1995.

[2] *Micromachined Silicon Inertial Sensors, A Defense Conversion Technology,* G. Brand, J. Connelly, Rockwell, Draper Laboratory, 1994 Joint Service Data Exchange, October 31-November 3, 1994, Scottsdale, AZ.

[3] *A Micromechanical Comb Drive Tuning Fork Gyroscope for Commercial Applications,* M. Weinberg, J. Bernstein, S. Cho, A.T. King, A. Kourepenis, P. Ward, and J. Sohn, Draper Laboratory, Sensors Expo, September 20-22, 1994.

[4] *A Micromachined Comb Drive Tuning Fork Rate Gyroscope,* M. Weinberg, J. Bernstein, S. Cho, A.T. King, A. Kourepenis, and P. Maciel, Draper Laboratory, Proceedings of the ION 49th Annual Meeting. "Future Global Navigation and Guidance," June 21-23, 1993.

[5] *High Dynamic-Range Microdynamic Accelerometer Technology,* James L. Sitomer, Draper Laboratory, Proceedings of the ION 49th Annual Meeting, "Future Global Navigation and Guidance," June 21-23, 1993.

[6] *Micromechanical Sensors for Kinetic Energy Projectile Test and Evaluation,* A. Hooper, U.S. Army Yuma Proving Grounds, and R. Hopkins, and P. Greiff, Draper Laboratory, U.S. Army TECOM Test Technology Symposium VII, March 29-31, 1994.

[7] *Transducers '91, Digest of Technical Papers,* pp. 966-969, 1991 International Conference on Solid-State Sensors and Actuators, P. Greiff, B. Boxenhorn, T. King, and L. Niles, Draper Laboratory, June 24, 1991, IEEE Cat. #91CH2817-5.

PZT Thin Film Piezoelectric Traveling Wave Motor

71138

Shen Dexin, Zhang Baoan, Yang Genqing*,Jiao Jiwei, Lu Jianguo, Wang Weiyuan
State Key Laboratories of Transducer Technology,
*Ion Beam Laboratory
Shanghai Institute of Metallurgy, Chinese Academy of Sciences
865 Changning Road, Shanghai 200050, P.R. China

Abstract

With the development of MEMS, its applications in many fields attract more and more attention in the world. Among kinds of MEMS, micro motors which include electrostatic, electromagnetic type are the typical and important ones. As an alternative approach, the piezoelectric traveling wave micro motor, based on thin film material technology and IC technologies, circumvents many of the drawbacks of the above two type of micro motors. It displays its distinct advantages. In this paper, we report a PZT piezoelectric thin film traveling wave motor. The PZT film with thickness of 150μm and diameter of 8mm was first deposited onto a metal substrate as the stator material. Then, eight sections were patterned to form the stator electrodes. The rotor was 8kHz frequency power supply. The rotation speed of the motor is 100rpm. Besides, the influences of the friction between stator and rotor and the structure of rotor on the rotation have been studied.

Introduction

With the development of microelectromechanical systems(MEMS), the application attracts the attention of the world. Among kinds of MEMS, micro motors that include electrostatic, electromagnetic type are the typical and important ones. As an alternative approach, the piezoelectric traveling wave micro motor, based on thin film material technology and IC technologies, circumvents many of the drawbacks of the above two type of micro motors, and can be possibly applied in the research of piezoelectric vibratory gyroscope.

In this paper, we report a new structure for PZT thin film piezoelectric traveling wave motor, and discuss the relationship between rotation speed of rotor and friction, and that between excitation frequency and rotation speed.

Design and fabrication process of the piezoelectric motor

1.Traditional design of the piezoelectric motor structure

The structure of the traditional piezoelectric mini-motor[1] or micro motor[2] is shown in Fig. 1. In this structure, the public electrode is fabricated on the substrate. Then, the piezoelectric thin film is deposited, sputtered or coated on the substrate. The piezoelectric thin film is patterned to form fans, which support the rotor. The piezoelectric thin film is excited by AC source to produce acoustic vibration that leads to the rotation of the rotor. The design principle for the piezoelectric motor is that the mechanical deformation of the piezoelectric material electrically excited produces the traveling wave that drives the rotor. In the traditional structure, a short possible happens between the metal rotor and the fans. Therefore, another insulator layer on the fans is necessary. However, this extra layer causes a series of disadvantages and increases the complexity of process, even though suitable material has been selected to improve the rotation condition by increasing the friction between the rotor and the stator.

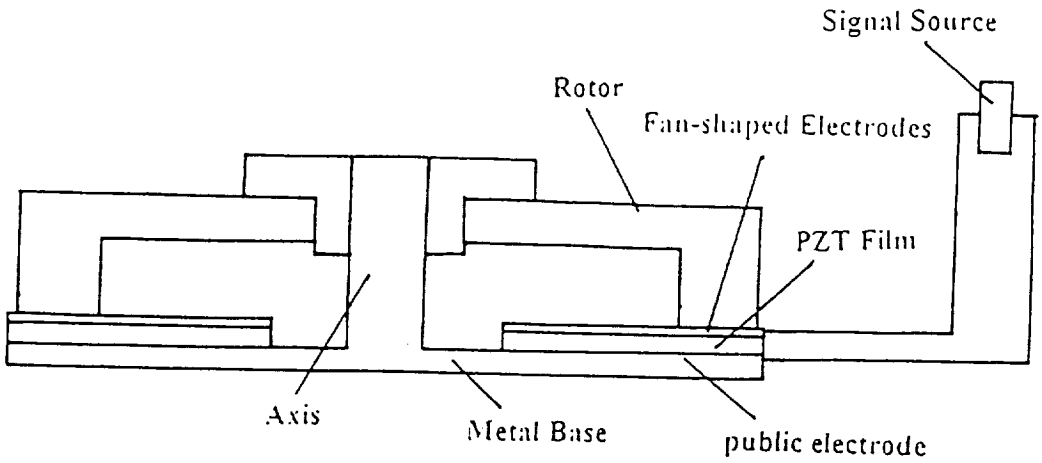


Figure 1 . Schematic diagram of the traditional motors

2. New design of the piezoelectric motor structure

After analyzing the vibration mode of PZT material for the rotor of the piezoelectric motor, the mechanical deformation of PZT thin film under AC excitation is determined to be an entirety deformation of the material. Meanwhile, the piezoelectric motor with new structure has some advantages compared with other motors fabricated with IC processes. The new designed structure is shown in Fig.2. In this structure, the public electrode, or the stator, of the piezoelectric motor faces upwards to support the rotor as shown in Fig.3, while the fans face downwards for linking leads easily as shown in Fig.4.

3.Fabrication processes for the piezoelectric motor

The stator material is commercially available. The substrate is copper with a thickness of less 200μm, which supports the PZT thin film with a thickness of less 150μm. Both faces of the PZT thin film are fused with silver layers as the electrodes. The outer silver face is also patterned to form eight fans for the connection with AC source. The inner silver face is a ring-shape public electrode and adhered to the copper substrate. This stator is then assembled on an axle with the public electrode upwards and fans downwards. When AC source is applied, the stator with above new structure can drive the rotor of the piezoelectric motor on the public electrode. In our experiment, the piezoelectric motor with a 8mm diameter rotates in a rotation speed of above 100rpm under 10V AC excitation.

There is another fabrication process for the rotor of the thin film piezoelectric motor. The copper sheet with a thickness of less 200μm is the substrate, on which SiO_2 layer is deposited. The PZT thin film with etched fans and leads is adhered to the substrate with epoxy. The public stator faces upwards to support the rotor. The final structure of the piezoelectric motor is shown in Fig.5.

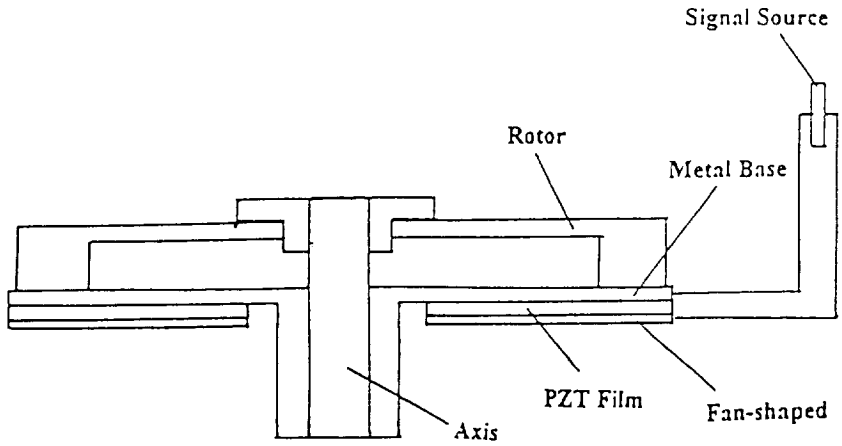


Figure 2. Schematic diagram of the motor

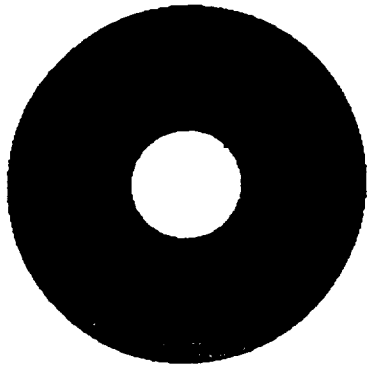


Figure3. Schematic diagram of the public electrode

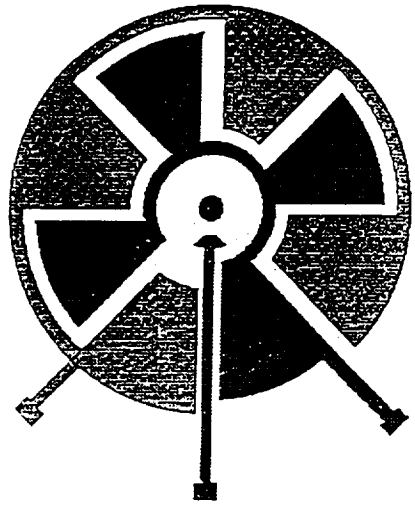


Figure 4. Schematic diagram of the fan-shaped electrode

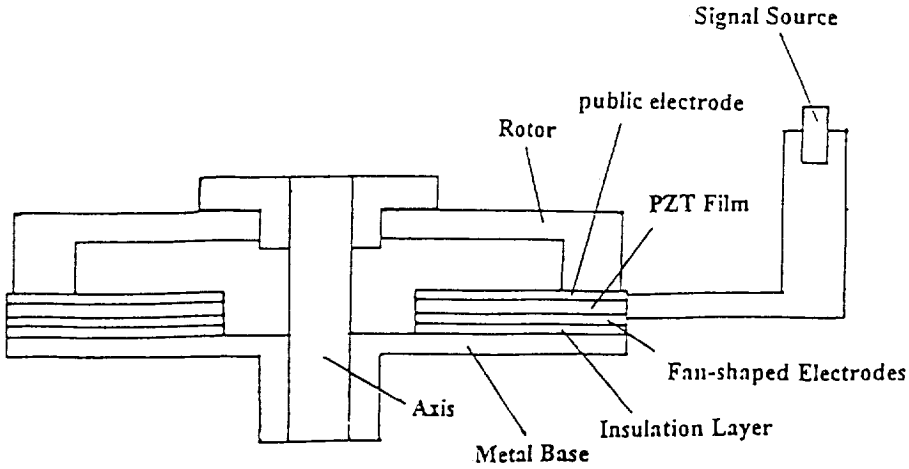


Figure 5. Detailed diagram of the motor

Results

1. Weight of rotor and rotation speed

The rotor on the public stator is even supported by the public stator. It is important to determine the weight of the rotor. The relationship of the weight of the rotor and the rotation speed is shown in Fig.6. When the rotor is very light, the friction force between the rotor and the stator is too small to drive the rotor. On the other hand, when the rotor is very heavy, the friction force hinders the rotation on the contrary. Only in the case of the rotor with a proper weight, it can rotate stably and continuously. In addition, the structure of the rotor is another important factor that influences the rotation mode. Therefore, the piezoelectric motor with a proper rotor, i.e., that is of proper weight and well-designed structure, can be in the best rotation state.

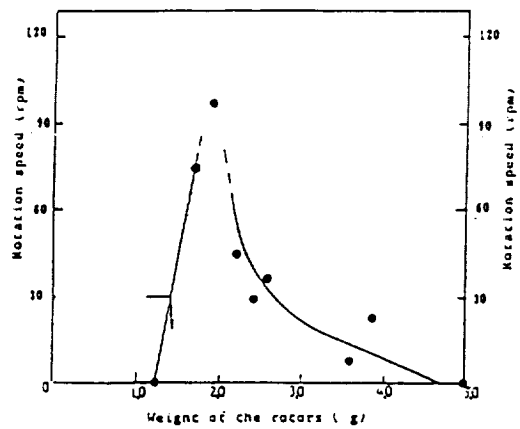


Figure 6. Relationship between the rotation speed and the weight of the rotors

2. Comparison of the rotation speed for the rotor with the two structures

There is no sensible difference of the rotation speed wherever the rotor is supported, by the public electrode or the fans as shown in Fig.7 and Fig.8.

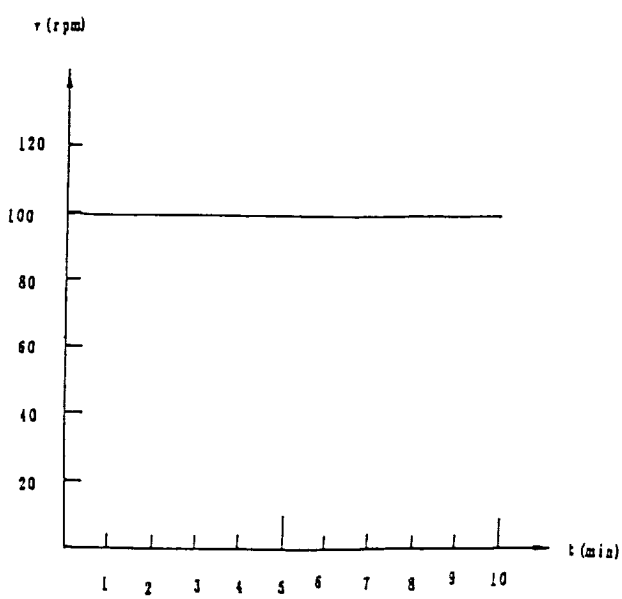


Figure 7. Rotation characteristics of the new motor

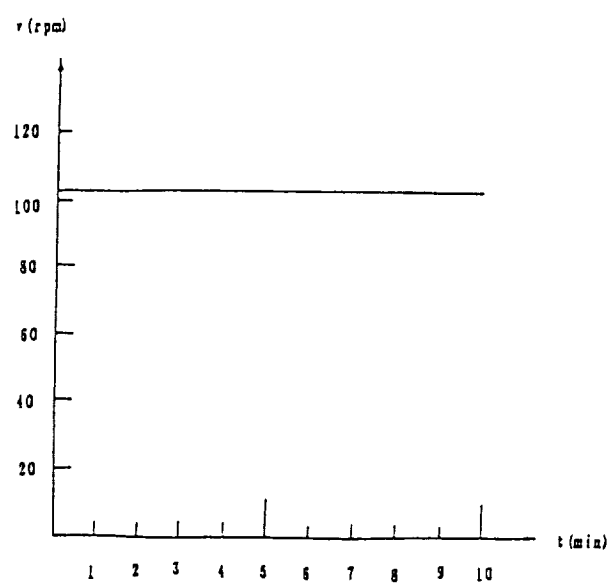


Figure 8. Rotation characteristics of the traditional motor

3.Relationship of the resonant frequency and rotation speed

The rotor of the piezoelectric motor is driven by the vibration of the piezoelectric thin film under AC excitation. The experiment showed that the rotation speed is sensitive to the excitation frequency, i.e., unless the piezoelectric thin film is resonantly excited, it produces large mechanical output to drive the rotor. The rotor of the new structure motor can rotate only near at the resonant and anti-resonant frequency, though with a great difference of mechanical output. With a small frequency shift from the resonant and anti-resonant value, the rotation speed decreases sharply and soon reaches zero. In Fig.9, the relationship of the excitation frequency and rotation speed for the rotor with proper weight is shown.

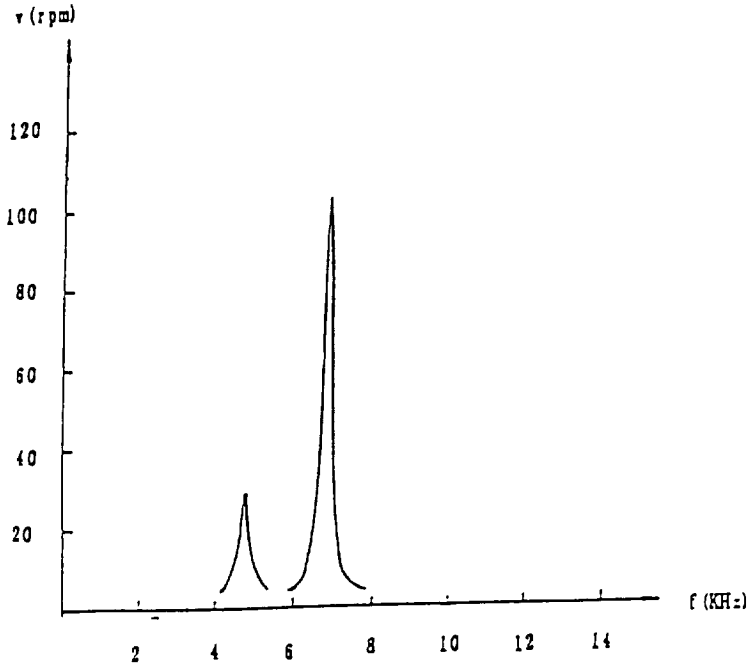


Figure 9. Relationship between the rotation speed and the resonance frequency

Discussion

In this piezoelectric motor with such a new and simple structure, its rotor of micro piezoelectric motor can be fabricated with IC process. Combined with electro-moulding process, it is feasible to fabricate piezoelectric micro motor. The problem of the short between fans can be solved in our new structure piezoelectric motor though the rotor is fabricate with metal materials. The new structure of the piezoelectric micro motor with 2mm diameter is shown in Fig. 10. In the structure, the transition layer of TiN, which is deposited on the silicon substrate and patterned, helps adhere the electrodes and leads on PZT thin film to the substrate. The fans and leads are all made of platinum. The other face of the PZT thin film is deposited with metal and patterned to form the public electrode. All above construct the stator of the new structure piezoelectric micro motor. Assembled with the axle and rotor fabricated with electro-moulding process, a piezoelectric micro motor can be developed completely.

References

[1]Patent CH Appl. 2282/92 20 July 1992.

[2]A.M. Flynn, L.S. Tavrow, S.F. Bart, R.A. Brooks, D.J. Ehrlich, K.R. Udayakuar and L.E. Cross, Journal of Microelectromechanical Systems, 1(1) (1992),44.

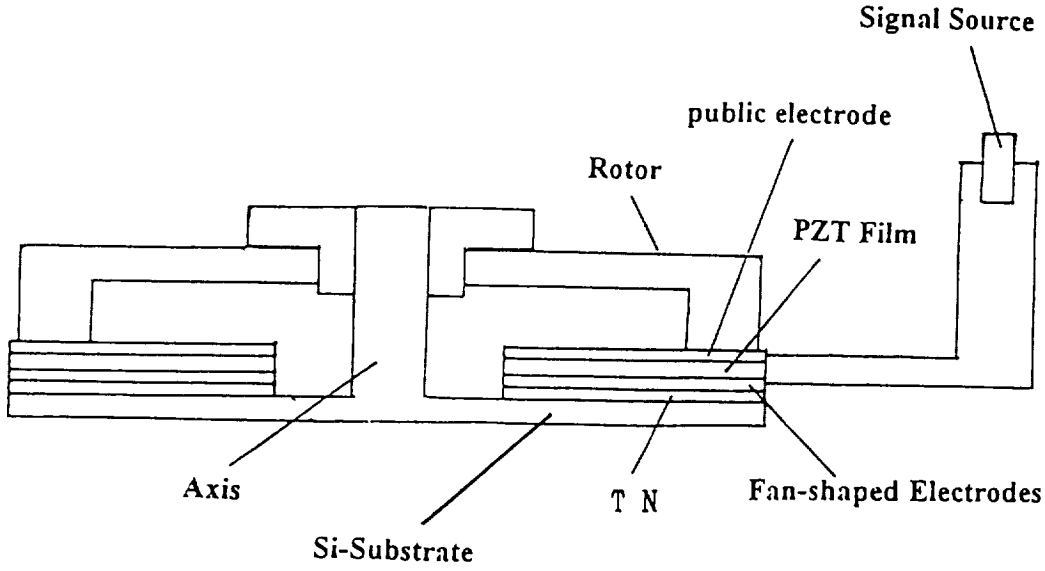


Figure 10. Schematic diagram of the Si-based micromotor

DEMONSTRATION OF A MONOLITHIC MICRO-SPECTROMETER SYSTEM

S. Rajic and C.M. Egert

Oak Ridge National Laboratory
P.O. Box 2009, MS-8039
Oak Ridge , TN 37831

71139 6P
030673

INTRODUCTION

The starting design of our spectrometer was based on a modified Czerny-Turner configuration containing five precision surfaces encapsulated in a monolithic structure. We were interested in exploring novel system designs, fabrication technologies, and examining numerous material systems, in the development of the micro-spectrometer. Our purpose at the early stages was to demonstrate the feasibility of the technology and not an attempt to address a specific sensing problem. Thus, we had great liberty to select the first prototype material, which is usually application specific. The first substrate material chosen was optical quality polymethyl methacrylate (PMMA).

DESIGN

Many starting designs were initially examined for the demonstration of this micro-sensor.[1] The final system design decision was eventually narrowed down to two possible configurations containing five and six precision surfaces which are shown respectively in figures 1a and 1b. The design shown in figure 1a was chosen for development. This five surface design was chosen since it contained one less precision optical surface, yet included multiple off-axis aspheres. Although the alternate design had greater linear dispersion, due to a longer focal length, the difficulty of fabricating an additional surface could not be justified in this technology demonstrator. This is especially true since the six surface design contained a folded optical path such that scatter / optical homogeneity became much more important. Among the many and often conflicting design goals, the physical parameters were assigned particular importance. The overall volume was targeted at less than 6 cm^3. This requirement was met with the extreme dimensions consisting of 2.4 cm X 1.9 cm X 1.2 cm. In this particular design, and material system, the mass was kept below 7 g. The wavelength range (bandpass) design goal was 1 μm (0.6 μm - 1.6 μm). The PMMA is particularly transparent in this wavelength region and there are interesting effects to monitor within this band. The optical system was designed and optimized using the ZEMAX Optical Design software program to be entirely alignment free (self-aligning).[2]

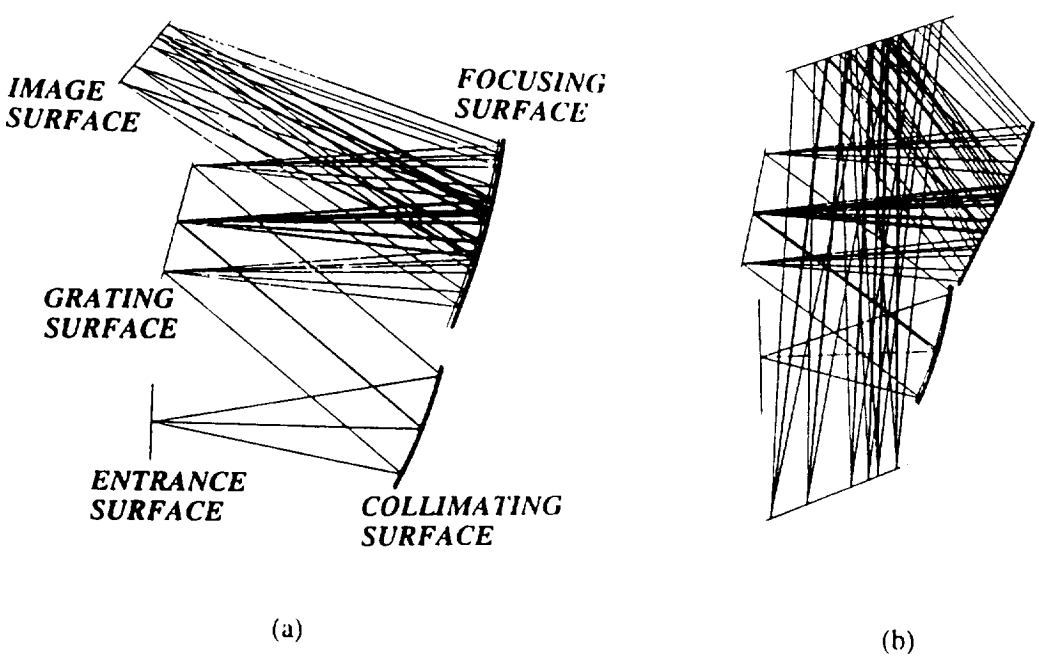


Figure 1. (a) Micro-spectrometer system layout of prototype. (b) Alternate design with higher resolution.

The entrance aperture consists of an optical fiber input. This fiber is positioned directly on the entrance surface at a specific location on that plane relative to the other four precision surfaces. The numerical aperture (NA) of the input fiber must be large enough to fill the useful portion of the collimating surface (first off-axis parabola). In this design the NA of this 62.5 μm core diameter fiber was 0.275 (~16^0 half angle). This core diameter represents the spectrometer entrance slit width and is not variable as in most laboratory sized instruments. The collimating surface redirects the light energy toward the grating with a focus at infinity. The light impinging on the grating arrives at a specific angle and with a collimated diameter of approximately 6mm. The grating in this design is a 300 lines / mm flat surface operating in the -1 order. The grating surface disperses the incident light toward the focusing surface (second off-axis parabola). The light at this point is diverging from the approximately 6mm diameter beam that impinged onto the grating. Thus the useful area of the focusing surface is the largest of the five precision surfaces at approximately a diameter of 10 mm. The focusing surface intercepts the diverging cone of light and focuses it onto the image surface. A linear detector array is then attached directly to this image surface.

The tolerances on the angles and spaces between these five precision surfaces were very challenging to maintain. However since a state-of-the-art diamond turning machine was used for the fabrication process, coordinates locations were achieved that met the required tolerances. The specification on surface figure was < 0.1 waves @ 632.8 nm. Although this requirement appears very stringent for an off-axis asphere, the full parent surfaces were generated in an uninterrupted single cut. The parent surfaces and useful clear apertures are shown in figure 2. The surface finish requirement was < 50 å rms. The mechanical tolerances of intersurface tilt and despace were +/- 0.3 mRad and +/-5 μm, respectively. This sensor was designed around readily available high power light sources. Since this was a prototype demonstration, the freedom to choose the wavelength range simplified that aspect of the optical design. The laser diodes chosen for calibration were 0.635 μm, 0.780 μm, 0.850 μm, 0.980 μm, 1.310 μm, and 1.550 μm. Miniature tungsten-halogen broadband, and mercury-argon calibration, sources were also used in the testing phase.

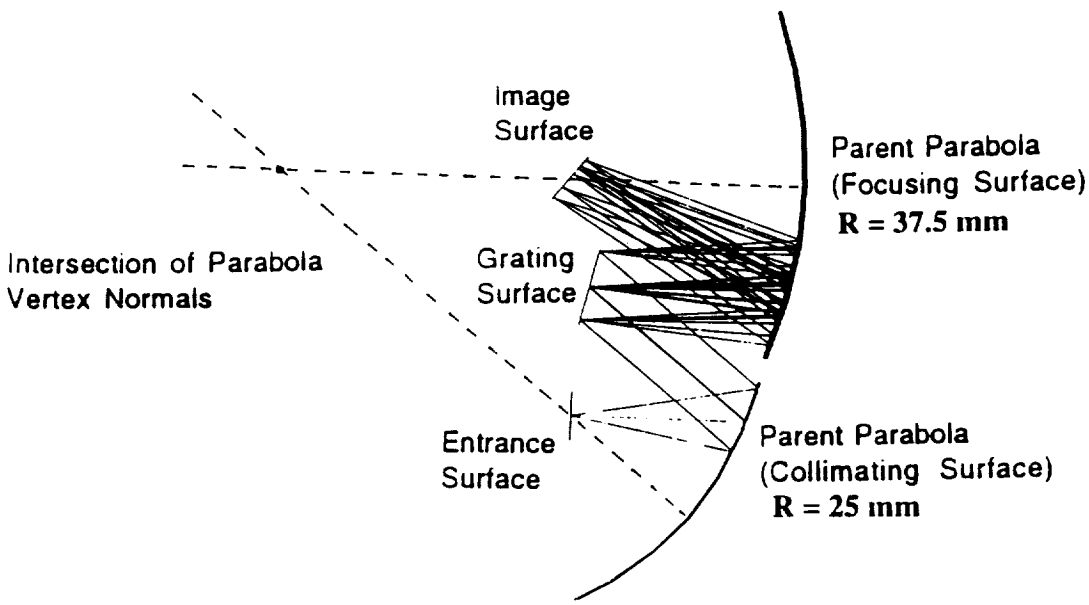


Figure 2. Optical layout with full parent surfaces included.

The bulk of the optical design effort centered on minimizing the focused spot size across the entire wavelength bandpass. Figure 3 illustrates the spot diagram for the final design for all six laser diode wavelengths.

Since the linear detector array lays along a vertical line connecting the spot diagrams for all six wavelengths, the horizontal spectral spreading is of no significance. A compromise focus was chosen for performance optimization across the detector array. This is evident from figure 4 which shows an expanded view of the individual spot diagram for the 0.980 μm laser. However, to maximize performance across the spectrum this particular prescription was selected. Since typical detector arrays have pixel sizes

ranging from ~ 15 μm to 50 um in length, we see that at most one to two pixels will be illuminated in this case. Thus achieving diffraction limited performance may not be required in all applications.

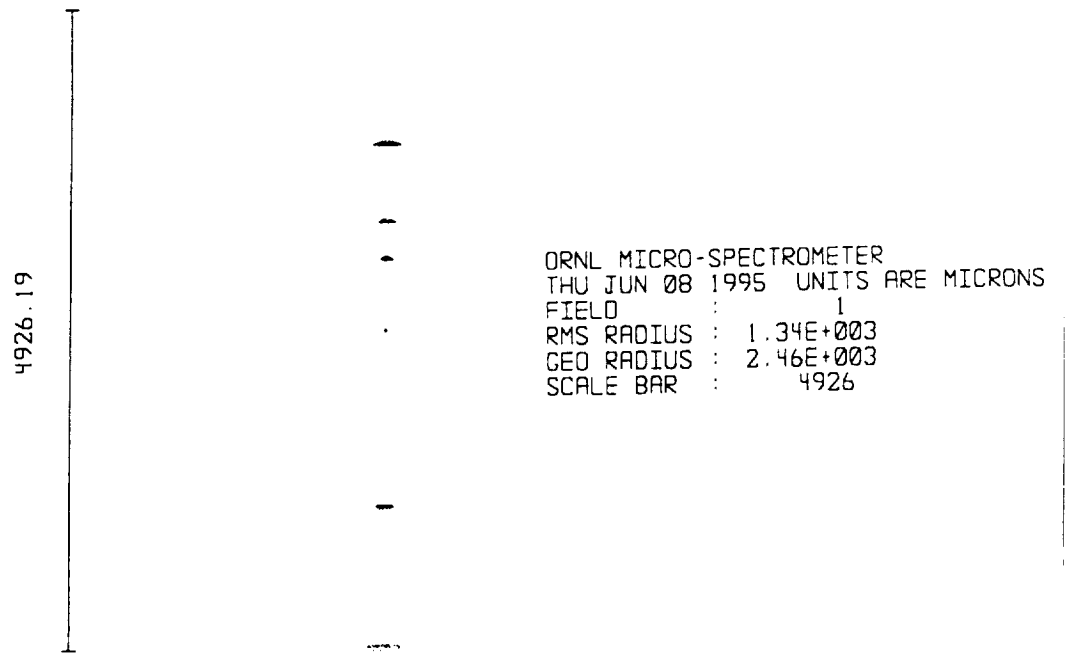


Figure 3. Spot diagram for all six laser wavelengths.

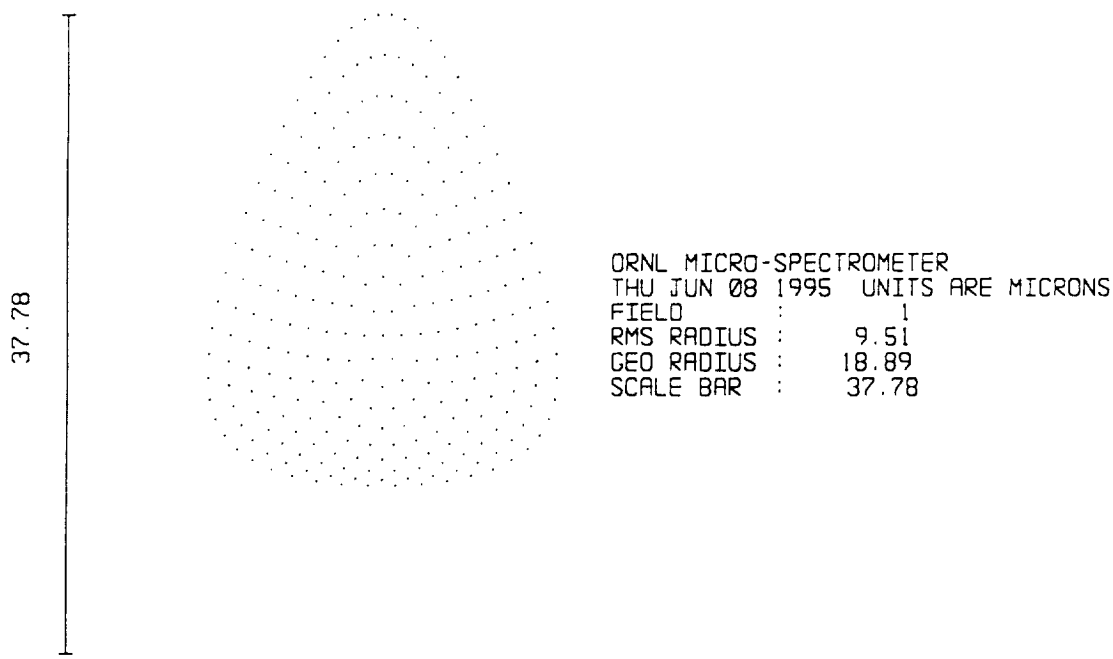


Figure 4. Spot diagram at 0.980 μm shows essentially single pixel illumination.

Once the optical prescription was finalized, a monolithic mechanical design was needed to encompass the internal light path. The design was optimized for minimum size. The relationship between the optical and mechanical designs is evident from figure 5. The five precision surfaces are labeled as in figure 1.

The mechanical design was accomplished using the Pro Engineer software package. This design tool was a great aid in linking the optical design to the fabrication process. This was particularly important since the diamond turning machine required global coordinates.

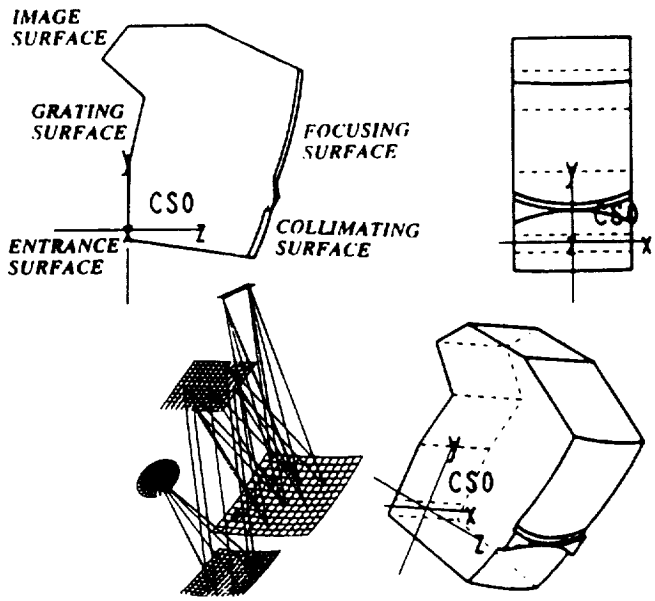


Figure 5. Micro-spectrometer mechanical design.

3. FABRICATION

The fabrication technique that was developed is particularly applicable to the manufacture of monolithic optical components and systems. The entire fabrication process including; fabrication engineering, rough machining, diamond turning, fixture fabrication, assembly, and coating, was performed at the Oak Ridge National Laboratory (ORNL). All of the diamond-turning was performed on the Nanoform 600 machine. The six major diamond-turning operations are shown in figure 6.

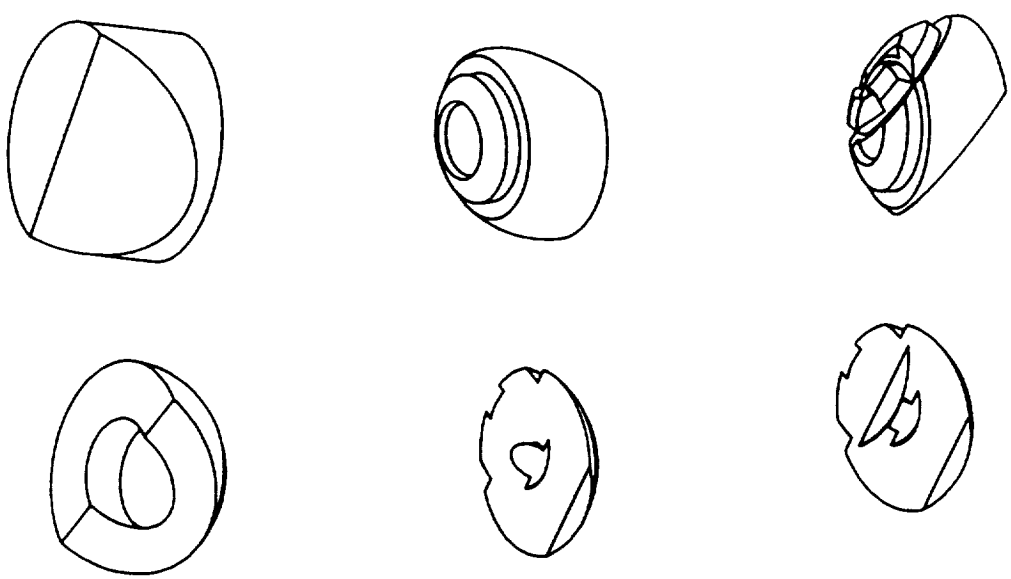


Figure 6. Fabrication steps.

TESTING

The micro-spectrometer was immediately subjected to qualitative tests after the initial fabrication process was complete. Light was injected into the entrance plane by means of a optical fiber. A visible wavelength was used to aid in the alignment process. However since this was a monolithic structure, with all optical surfaces predefined, only the input fiber and the detector array required alignment. A simple translucent white card was used to identify the image surface as the focal plane of the system. The different orders of the injected visible source were brought to a very sharp focus at the image surface as required. This quick test verified that the system contained no gross errors in either the design or fabrication phases. Thus the system was performing as a spectrometer although it had not yet been cut out of its surround.

Some of the system requirements were difficult to measure directly. For this reason system signal / noise was not specified explicitly. Since scattered light energy from all the internal surfaces contributes to the noise of the system the surface finish, which is directly related to scatter, was specified instead. This parameter, as well as surface figure, were readily accessible to physical measurement with both interferometers and profilometers.[3] Since all of the precision diamond turning was preformed on the Nanoform 600 machine, the measured values of ~ 50 å rms met our specification as expected.[4]

Since this was a totally new spectrometric device, calibration was pursued soon after reasonable signals were detected. The calibration was accomplished with multiple laser diodes. The diodes were first analyzed through a more conventional spectrometer with greater resolution. Then the same signals were introduced into the micro-spectrometer and compared. Data was first taken before the device was separated from its fabrication surround. The results showed a higher than expected noise floor due to the presence of fabrication "dummy surfaces" that were not in the optical design. However, reasonable signals were achieved. After the device was separated from its surround the noise level was reduced substantially. Finally the device was blackened on all its rough machined non-optical surfaces and the desired results were achieved. The wavelengths used thus far for calibration were; 635nm, 780nm, 850nm, and 980nm.

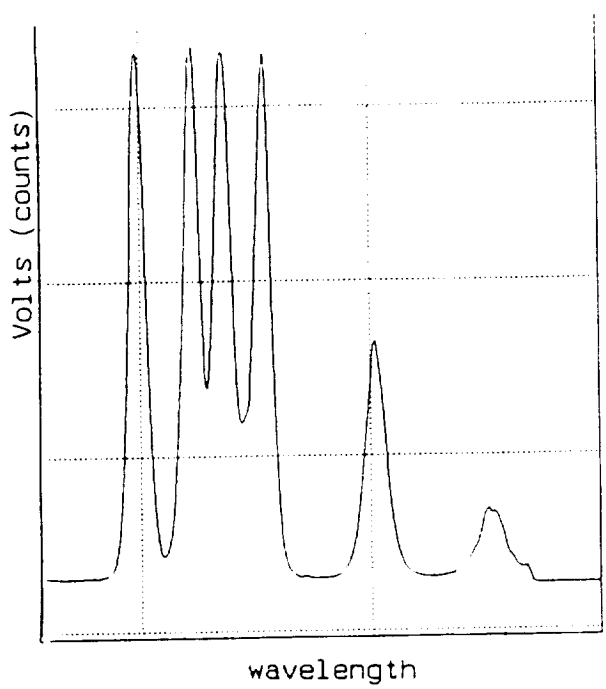


Figure 7. Calibration laser diodes 635nm to 980nm.

Evident from figure 7 are the second order spectra associated the 635 nm and 780 nm calibrating laser signals. This was due to the use of an oversized linear detector array containing 1024 pixels. Although for calibration purposes this can be advantageous, if more than a single wavelength is used the results can be confusing. The 850 nm laser in figure 16 also shows the beginnings of a second order signal. However the detector runs out of useful pixels before anything more than a small tail is seen. The 980 nm signal shows no signs of second order signals as expected.

CONCLUSIONS

It now appears clear that ultra-precision monolithic sensors, such as a micro-spectrometer, can be successfully fabricated. The prototype system that has been demonstrated can now serve as the basis for low cost production techniques involving mold fabrication. This would include both the Sol-Gel technique, which can produce silica systems, as well as traditional plastic injection molding. The system was designed for a linear dispersion of ~200 nm/mm. However a design has been shown that, with the addition of a single flat surface, will produce a system with linear dispersion between 75-100 nm/mm. This is typically sufficient for low / medium resolution sensor applications and would be tailored to the specific applications as required. The detailed analysis of the micro-spectrometer performance will be ongoing. The already identified applications include, emission mode laser warning receiver, transmission mode chemical / environmental detector, and a reflection mode corrosion monitor.

ACKNOWLEDGEMENTS

This work was sponsored by the laboratory directors research and development (LDRD) fund within the Oak Ridge National Laboratory (ORNL). ORNL is managed by Lockheed Martin Energy Systems for the U.S. Department of Energy under contract no. DE-AC05-84OR21400.

REFERENCES

1. Instruments SA, JOBIN YVON/SPEX Division, Guide for Spectroscopy, Edison NJ, 1994.

2. S. Rajic, "Snap-Together Directed Energy Threat Protection System", OSA, Optical Fabrication & Testing Workshop, 1992.

3. L.C. Maxey, "Novel Technique for Aligning Paraboloids", SPIE Vol.1532, Advanced Optical Manufacturing and Testing II, 1991.

4. M.C. Gerchman, "Specifications and Manufacturing Considerations of Diamond-Turned Optical Components",SPIE, O-E Lase, 1986.

Silicon Micromachined Sensor for Broadband Vibration Analysis

Adolfo Gutiérrez, Daniel Edmans, Chris Cormeau, Gernot Seidler
InterScience, Inc.
Troy, NY 12180

and

Dave Deangelis†, Edward Maby††
†Department of Mechanical Engineering
††Department of Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

SBIR Sponsors: DoD AF, DoD ARMY-ARPA

ABSTRACT

The development of a family of silicon based integrated vibration sensors capable of sensing mechanical resonances over a broad range of frequencies with minimal signal processing requirements is presented. Two basic general embodiments of the concept were designed and fabricated. The first design was structured around an array of cantilever beams and fabricated using the ARPA sponsored MUMPS process at MCNC. As part of the design process for this first sensor, a comprehensive finite elements analysis of the resonant modes and stress distribution was performed using PATRAN. Dependence of strain distribution and resonant frequency response as a function of Young's Modulus in the Poly-Si structural material was studied. Analytical models were also studied. In-house experimental characterization using optical interferometry techniques were performed under controlled low pressure conditions. A second design, intended to operate in non-resonant mode and capable of broadband frequency response, was proposed and developed around the concept of a cantilever beam integrated with a feedback control loop to produce a null mode vibration sensor. A proprietary process was used to integrate a MOS sensing device, with actuators and a cantilever beam, as part of a compatible process. Both devices, once incorporated as part of mutifunction data acquisition and telemetry systems will constitute a useful system for NASA launch vibration monitoring operations. Satellites and other space structures can benefit from the sensor for mechanical condition monitoring functions.

INTRODUCTION

Over the past 10 years the research and development efforts aimed towards producing very small electromechanical components using microelectronics fabrication techniques have received ever increasing attention. As a result of these efforts, various competing technologies have emerged as leading contenders. In Europe, high aspect ratio microstructures fabricated using the LIGA (Lithographie, Galvanoformung und Abformung) process based on the use of thick photoresist, synchrotron based x-ray lithography and injection molding has become dominant. Japan, driven by MITI coordinating efforts, has chosen to miniaturize components through mechanical microfabrication using high precision machinery. In the United States the most developed technology for the fabrication of micromachined devices relies on using silicon bulk or surface micromachining.

A typical silicon surface micromachining process begins with a silicon wafer upon which alternating layers of structural (Poly-Si) and sacrificial material (SiO_2) are deposited and selectively etched to produce structures that have relatively small aspect ratios. The thickness of the depositions is limited to a few microns because of photoresist thickness and etch rate limitations, and because of built-

in or intrinsic residual stresses that tend to increase as the thickness of the layers increases. These process created stresses can only be partially relieved through thermal annealing and the remaining residual stress can determine severe structural deformations after final chemical release.

Although silicon micromachining of Micro-Electromechanical Systems (MEMS) has rapidly become a major research area, only a handful of successful commercial devices have been produced. The consolidation of MEMS devices as a segment of the sensors and microelectronics industry ultimately will rely on proper product definition, skillful market identification, and competitive cost-benefit ratios. InterScience, Inc. has decided to focus its research and development efforts in a group of relatively simple integrated mechanical vibration sensors suitable for real time vibration analysis of mechanical structures. These types of sensors would have applications in transportation and other mechanical systems where reliability if of utmost concern. Systems such as launch and orbital vehicles, permanent space structures, aircrafts and missiles are examples of such systems of direct interest to NASA, FAA and DoD. Also, high reliability mechanical systems are found throughout the nuclear cycle. Some potentially mass producible applications, such as engine analysis in the ground and sea transportation industries, are also readily conceivable. Cost-benefit tradeoffs will ultimately be the determining factor in the commercial success of these types of sensors.

Earlier work by Benecke and Csepregi[1] demonstrated the feasibility of building resonant vibration sensors using a piezoelectric sensor located at the base of the cantilever beams, where the strain is maximum. Additional work by Motamedi[2] from Rockwell demonstrated cantilever based accelerometers fabricated in silicon using piezoelectric capacitive sensors and a PMOS amplifier, in a process fully VLSI compatible. Both sensors exhibited narrow band responses that could potentially be improved through the use of feedback control schemes. Inspired by these initial investigations, our team decided to pursue the development of a family of vibration sensors, both resonant and non-resonant, using silicon surface micromachining[3]. The goal of the first effort was to develop a resonant vibration sensor capable of real time spectral analysis using an array of cantilever beams. The goal of the second related effort was to produce a wideband vibration sensor incorporating feedback control and integrated capacitive position sensing.

This paper describes the various aspects that we have encountered in the development of our vibration sensors. Critical proprietary information has been left out of the discussion. The development is still an ongoing effort implying that some critical information for performance assessment is missing. The paper starts with a review of the effort of producing an array of resonant piezoresistive vibration sensors. It follows with a review of the fabrication process which was used by MCNC in our first design. Numerical analysis of our sensors, based on finite element methods is then surveyed. It follows a discussion of the rationale for incorporating feedback forced balance schemes. Finally a discussion of our in-house mechanical characterization capabilities based on interferometry is presented.

SILICON MICROMACHINED RESONANT VIBRATION SENSOR

A number of functional prototypes were fabricated using the MCNC (formerly Microelectronics Center of North Carolina) foundry as part of the MUMPS4 (Multi-User MEMS Process, Run # 4). This multiproject process is supported by ARPA as part of a multi-year infrastructure program for MEMS development. Figure 1 is a general cross section of the two-layer polysilicon surface micromachining MUMPs process[4]. This process has the general features of a standard surface micromachining process: (1) polysilicon is used as the structural material, (2) deposited oxide is used as the sacrificial layer, and (3) silicon nitride is used as electrical isolation between the polysilicon and the substrate.

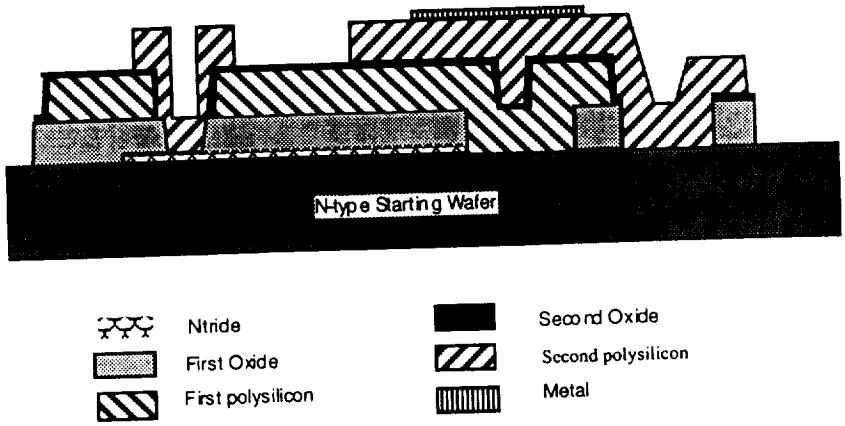


Fig. 1 Double-layer Poly-Silicon MUMPS process (not to scale)

The MUMPS fabrication process begins with n-type <100> 100 mm silicon wafers. Next, a 500nm thick low stress silicon nitride layer is deposited on the wafers as an electrical isolation layer, followed by a thin layer of electrical polysilicon deposited to allow the user to design an electrical sensing or electrode plate, isolated from the wafer, below the first mechanical polysilicon layer. It is followed by a 2.0 μm thick first oxide sacrificial layer. This layer of oxide is will be removed at the end of the process to free the first mechanical layer of polysilicon. Next, the first structural layer of polysilicon is deposited. The thickness of the polysilicon layer is 2.0 μm. A thin masking layer of second oxide is deposited. After the first layer of polysilicon is patterned, a second sacrificial thin oxide layer is deposited. This oxide layer will be removed at the end of the process to facilitate the release of the second mechanical polysilicon layer. Next, the second 1.5 μm thick polysilicon layer is deposited. As with the first polysilicon layer, a thin sacrificial oxide layer is deposited as an etch mask and dopant source. The final deposited layer in the MUMPs process is a 1.0 μm thick aluminum layer. The front side of the wafer is protected and the polysilicon and oxide layer on the back of the wafer are removed by wet chemical etching. The nitride layer is left on the back of the wafer in case post-processing of the die is desired at the MUMPS participants own facility.

Each of the dies fabricated using MUMPS4 contained an array of resonators. The typical resonator consisted of a cantilever beam that has an additional loading mass of deposited aluminum placed on the free end to reduce the resonant frequency. It could also have a piezoelectric material deposited at the base that acts as a strain gauge. A depiction of the 1 cm^2 die fabricated is shown in Figure 2. An schematic representing the basic concept is shown on Figure 3.

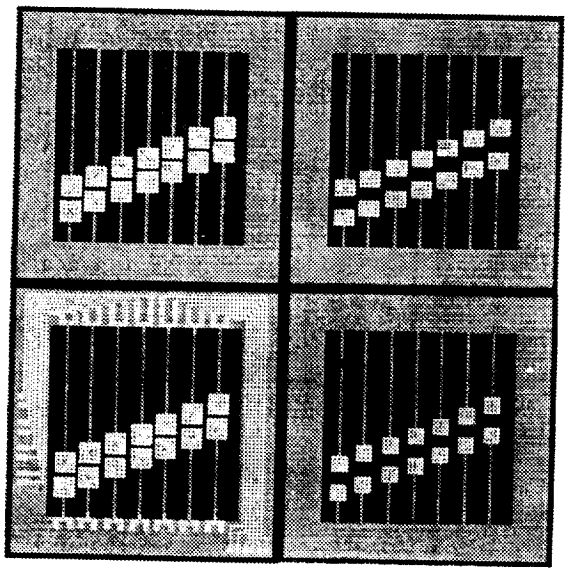


Figure 2. A 1 cm^2 die containing the first four cantilever test structures fabricated using MUMPS 4.

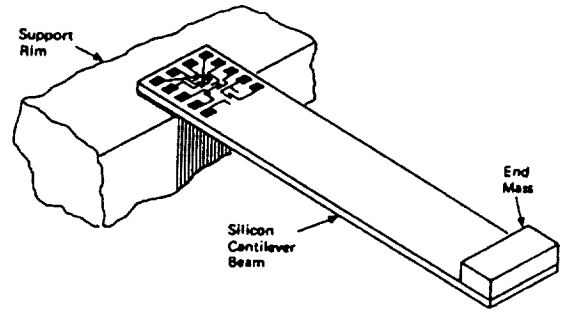


Figure 3. Cantilevered beam showing placement of piezoresistor and loading mass.

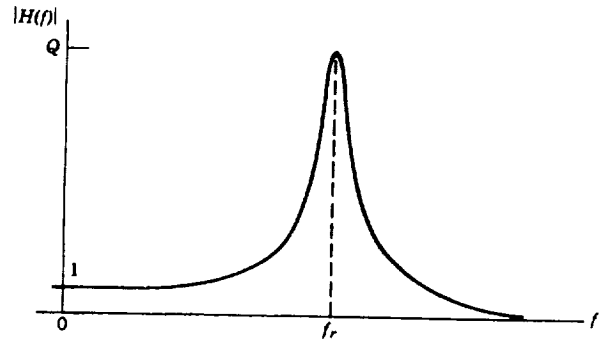


Figure 4. Transfer function of a single cantilever beam oscillator.

The resonant frequency of the beam is a function of several parameters including Young's Modulus of elasticity, the beam length, the thickness of the beam, the density of the polysilicon, and the density of the surrounding media. A typical transfer function is shown in Figure 4. The selectivity is represented by the peak at the resonant frequency, f_r, and it is characterized by the width at half maximum power (FWHM), which is strongly dependent on the geometry of the end mass and the viscosity of the surrounding gas or fluid. For the particular designs implemented, the width of the resonant peak is about 1 Hz in vacuum and 800 Hz in air at STP. The selectivity and sensitivity are trade-offs that have to be made to provide the right combination for the particular application. The strain is related to the driving acceleration and the physical parameters of the beam. The maximum acceleration that the beam can sustain is given by substituting the maximum yield strength for silicon of $10^{-3} N/m^2$ into the strain equation. The resulting value is in the thousands of g's range for our design. A variable resistance directly proportional to the strain is developed in the piezoelectric crystal. It is also proportional to the piezoelectric material's coupling coefficient, and the area of the piezoelectric deposition, and inversely proportional to the combination of the capacitance of the piezoelectric material and parasitic capacitances.

FINITE ELEMENT ANALYSIS

The critical aspect of the design is the resonant frequency and the mechanical quality factor, Q, of each cantilever beam. An advantage of a micro-machined vibration sensor, in addition to its small size and potential for low cost fabrication, is the ability to provide high sensitivity to a discrete set of vibrational frequencies.

In order to accurately predict the response of the beams to the driving forces, it is necessary to perform finite element analysis calculations. The use of finite element analysis also provides estimates for the strain induced at the anchored end of each beam. The strain distribution thus calculated is used

to predict the optimum location for the piezoelectric material in order to maximize the charge that will accumulate in the piezoelectric material, which is directly related to the output piezo-voltage to be measured.

Initially, the numerical analysis was carried out using the commercial software package for finite element analysis, PATRAN. The analysis of the shortest cantilever beam in the array, measuring 380 μm to the inside edge of a 300 μm square paddle. Based on the finite element analysis performed, the resonant frequency of the beams vary from 220 Hz to 2,245 Hz. It is important to be aware that the FEA models rely on *a priori* knowledge of the Young's Modulus, a parameter that for the case of Poly-Si, varies significantly for each specific fabrication process. More recently we have centered our numerical analysis work around the FEA package WECAN.

The fundamental frequency of the beams depicted in this section is much higher than will be designed and is only being used for illustrative purposes. Figure 5 shows the stress distribution in a beam that has a paddle on the free end. The paddle reduces the resonant frequency and broadens the resonant peak. This shows the use of the FEA, to identify areas of high stress during the design process in order to alleviate any problems by careful early planning.

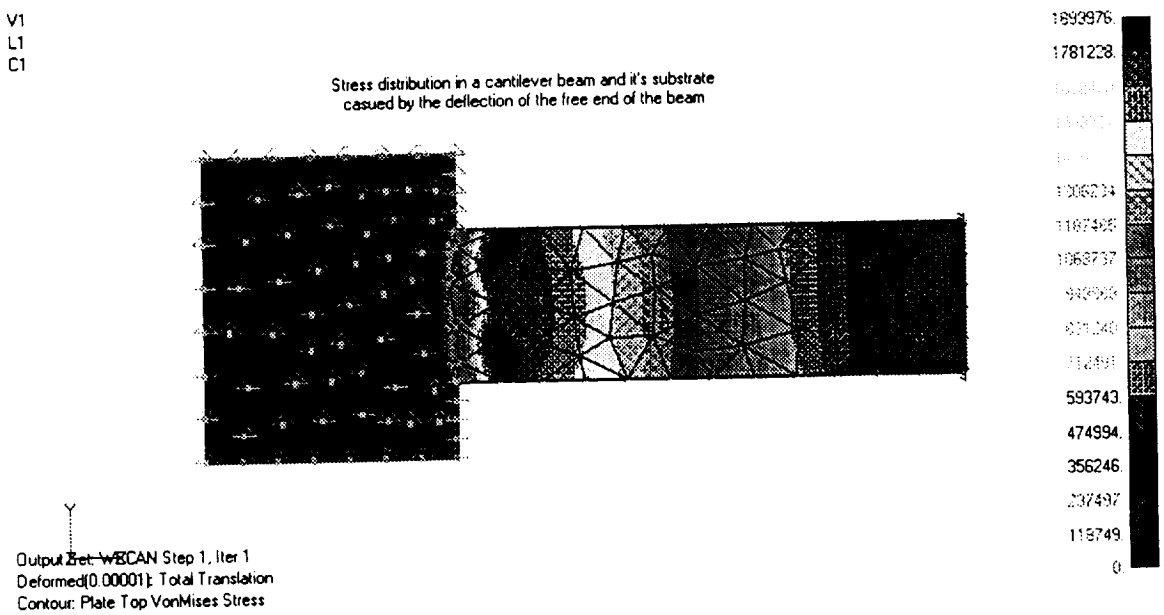


Figure 5. Stress distribution in a typical beam showing the highest stress concentration at the constrained end.

For a structure which is 1.5 μm thick, 60 μm wide and 200 μm long, the fundamental mode, shown in Figure 7, occurs at 50,859 Hz and the typical deflection is 144 μm when it is not constrained. The stress distribution is represented by the shading. The shape of the deflection is important because the capacitance will vary as the beam is deflected. At higher modes 3 and 5 the beam twists. The next four frames in Figure 8 show the shape the beam takes in each of the first four higher order modes, numbers 2 through 5.

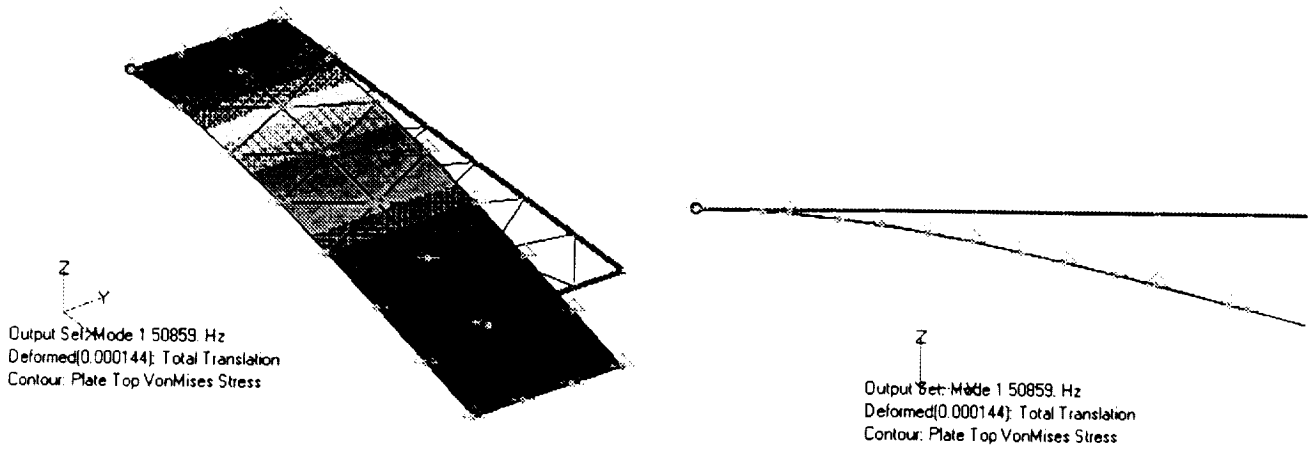


Figure 7. The shape of a simple beam in the fundamental mode.

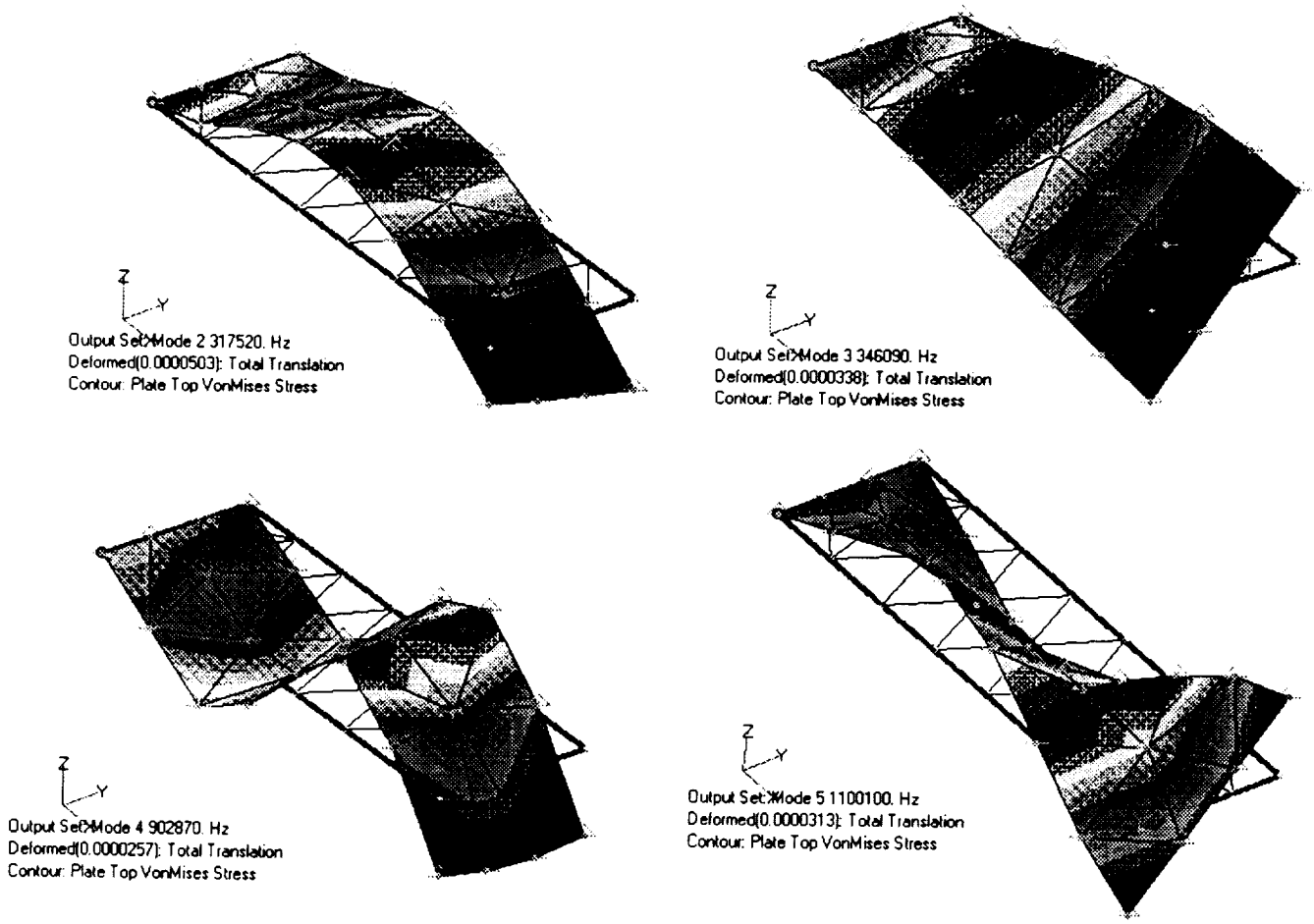


Figure 8 The first four higher order modes, 2-5, of a typical beam showing the shape and the scaled deflection. The strain distribution is represented by the shading.

The comparison of the normalized deflections shown in the preceding pictures shows that the maximum displacement associated with the higher order modes is always less than 30% of the displacement of the beam in its fundamental mode. Figure 9 below shows the relative displacement of the beam in each of the first five modes relative to the sixth mode. It shows that there is very little deflection associated with the higher modes, although it cannot be ignored.

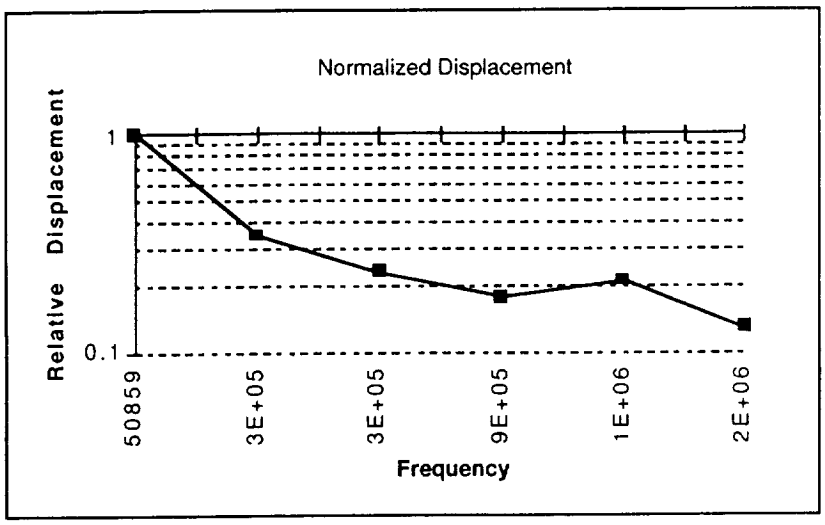


Figure 9 Relative deflection of the beam in each of the first six modes.

The resonant vibration sensor was determined to be adequate for applications in which *a priori* knowledge of the mechanical frequencies is available. Although in principle it is possible to cover a large spectral range through the use of a very large number of cantilever beams, in practice however, this approach will result in large and highly redundant devices, implying high unit cost. Furthermore, resonant devices impose severe packaging restrictions since they require vacuum sealed operation to reduce viscous damping.

A concept that has been used for several decades in the field of seismology is the null mode forced balance feedback operation of an inertial vibration sensor which improves all the mechanical performance figures. It enables wider broadband, higher sensitivity and simplified packaging requirements.

SILICON MICROMACHINED WIDEBAND NON-RESONANT VIBRATION SENSOR

The force-balance vibration sensor derives its name from the forces applied electrostatically between its suspended mass and frame through the use of a feedback loop. Two forces are involved, one variable force proportional to the displacement of the mass relative to the frame that tends to balance this displacement and so behaves as a variable spring, a second force is proportional to the relative velocity and provides lossless dynamic damping.

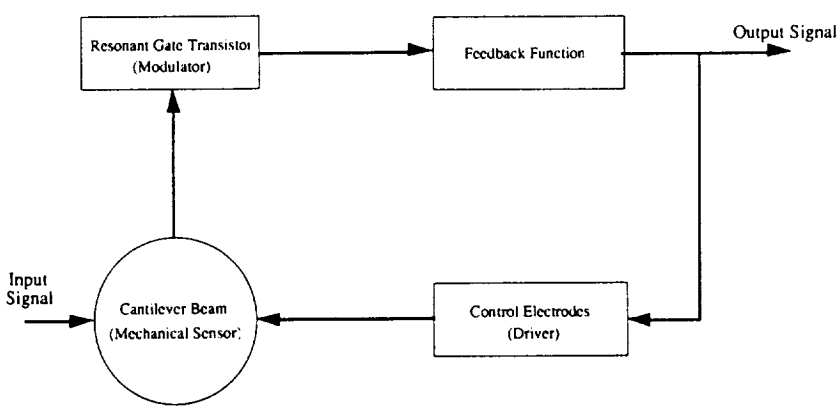


Figure 10. Force-balance feedback control block diagram.

The fundamental equation describing any inertial vibration sensor is[5]:

$$\frac{d^2x_r}{dt^2}+\frac{R}{M}\frac{dx_r}{dt}+\frac{1}{MC}x_r=\frac{d^2x_i}{dt^2}$$

where M is the suspended mass, x_i is the external mechanical excitation, R is the viscous damping resistance and C is the compliance of the supporting spring.

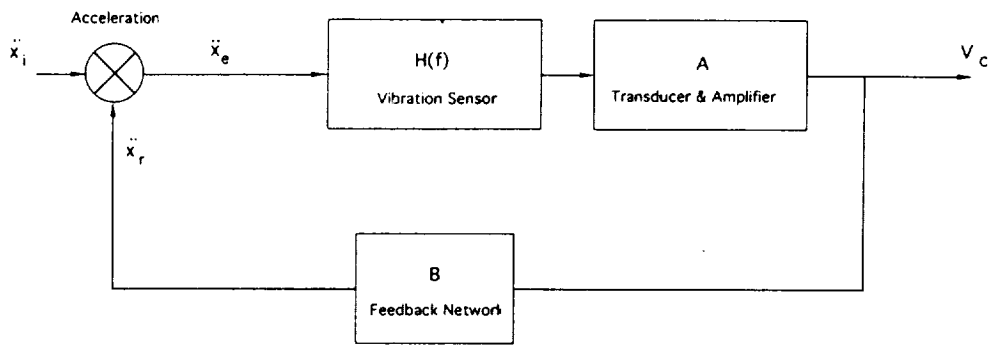


Figure 11. Force-balance feedback control diagram

The application of negative force-feedback to a seismometer produces a number of advantages and is in fact essential when a small mass is suspended with a high Q, in order to achieve a satisfactory transient response. The most useful form of feedback is negative displacement, which tends to keep the mass fixed in a position with respect to its support, making the suspension appear more stiff and increasing the natural frequency.

The transfer function of the vibration sensor represented in Fig. 11 is :

$$\frac{x_r}{\ddot{x}_i}=\frac{1}{s^2+2\xi\omega_0 s+\omega_0^2}$$

$s=jw$ is the Laplace operator and ξ is the damping ratio. The closed-loop transfer function is

$$\frac{v_0}{\ddot{x}_i}=\frac{A}{s^2+2\varepsilon\omega_0 s+\omega_0^2}$$

where v_0 is the output voltage, A is the gain in the forward path and B is the transfer function of the feedback path, and becomes $1/B$ when AB is the dominant term. If it is assumed that AB is independent of frequency, the DC loop gain L is $(1/\omega_0^2)AB$ and the natural frequency is increased by a factor $L^{1/2}$, the damping being reduced by the same factor. The response is essentially flat (to acceleration) from DC to the new natural frequency and the transient response can be controlled by a compensation network in the feedback path.

The fundamental limit of detection of any vibration sensor is determined by the Brownian motion of the mass. It has been shown that the noise equivalent acceleration (d^2x_i/dt^2) for a bandwidth f is given by

$$(x)_{ne}^2=\frac{4RkT\Delta f}{M^2}=\frac{4kT}{M}\frac{\omega_o}{Q}$$

where kT is the equi-partition energy and Q is the quality factor of the suspension. It can be observed that a small mass may be used provided that the damping is low. A typical micromachined device, such as the one we have developed has a mass of about 0.3 μg; Q of about 1000, and a natural frequency of a 1kHz which determines that the noise equivalent acceleration is about 3.5 x 10^{-13} m^2s^{-3}. In fact this

device is extremely sensitive, thus Brownian noise will not be important and the electronic noise, mainly thermal, will ultimately determine the sensitivity of the micro seismometer.

Mechanical design requirements are eased and the desired wideband response is simply determined by the feedback parameters. The signal-to-noise is unaffected by feedback. The response is controlled by applying forces to the mass; this does not affect the Brownian motion, whereas adding damping to control the response in an open loop-system increases the dissipation and therefore increases the Brownian motion. Furthermore, it is easy to verify that a very small mass can be suspended with a high Q and employed in a feedback system of suitable loop gain, can thus provide a flat response and adequate detectivity over the whole range of interest in seismology. Additionally advantages, over conventional open-loop instruments are obtained in linearity, dynamic range and calibration.

In the control loop shown in Figure 10, two electronic transducers are attached to the sensor, one being the modulator, in our case our proprietary capacitive sensing device, and the other being a driver or control electrodes to provide mechanical force opposing the displacement. The cross section of a device currently being fabricated that implements this closed feedback control loop is shown in Figure 12. It can be observed in the cross-sectional view that two control electrodes have been inserted. The sensing element of the loop is a proprietary capacitance measuring device.

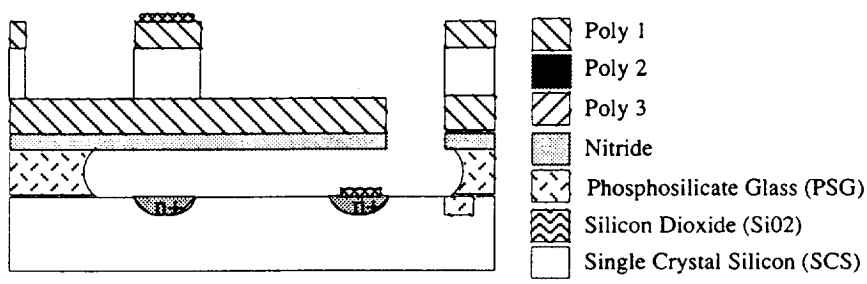


Figure 12 Cross-sectional view of the Silicon Micromachined Wideband Vibration Sensor.

EXPERIMENTAL CHARACTERIZATION USING OPTICAL INTERFEROMETRY TECHNIQUES

Following fabrication of a vibration sensor die, they must be tested and characterized. This is accomplished by using an interferometric technique developed in our own laboratories. A diagram of the optical test set-up used to characterize the frequency response of the resonators is shown in Figure 13. A Fabry-Perot (FP) type interferometer is to be used to measure very small deflections of the cantilever beam. In general an FP interferometer contains two plates separated by a gap. In the case of the FP interferometer used here, the first plane is a partial reflector, formed by the end surface of the fiber. The total reflector is the silicon target. Provided that the resonator is moving, the standoff distance will change producing a changing interference pattern. The fringe pattern can be resolved using our system to an accuracy of $\lambda/30$ or up to ± 27 nm for an 820 nm wavelength source.

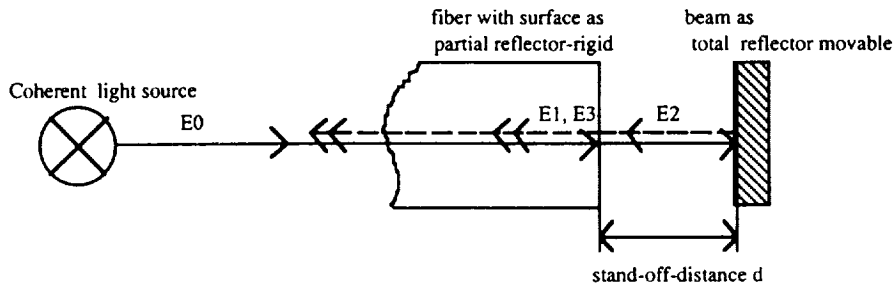


Figure 13. Fabry-Perot-Interferometer based on single mode fiber interface.

The fringe visibility is given as

$$\alpha = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

where I_{max} and I_{min} are the relative maxima and minima of the intensity in the range of $d \pm \Delta d$. The deflection of the beam, which is variation of the stand-off distance, will change with the applied voltage. The resulting optical phase difference varies linearly. The optical phase difference $\phi(t)$ corresponds to

$$\phi(t) = \frac{4\pi}{\lambda} d(t)$$

where $d(t)$ is deflection of the beam and λ is the wavelength of the light source.

The beam is deflected with the application of a modulation current through the electrode. A variety of deflection waveforms can be used to characterize the material parameters and the response of the beam. A moving beam will modulate the frequency of the interferometric signal. In addition, a reference signal is needed to differentiate the motion of the vibrating beam from any residual motion of the surrounding support structure. The reference signal is achieved by a second interferometer that is pointed to the die substrate. Figure 14 shows the setup of the measurement system.

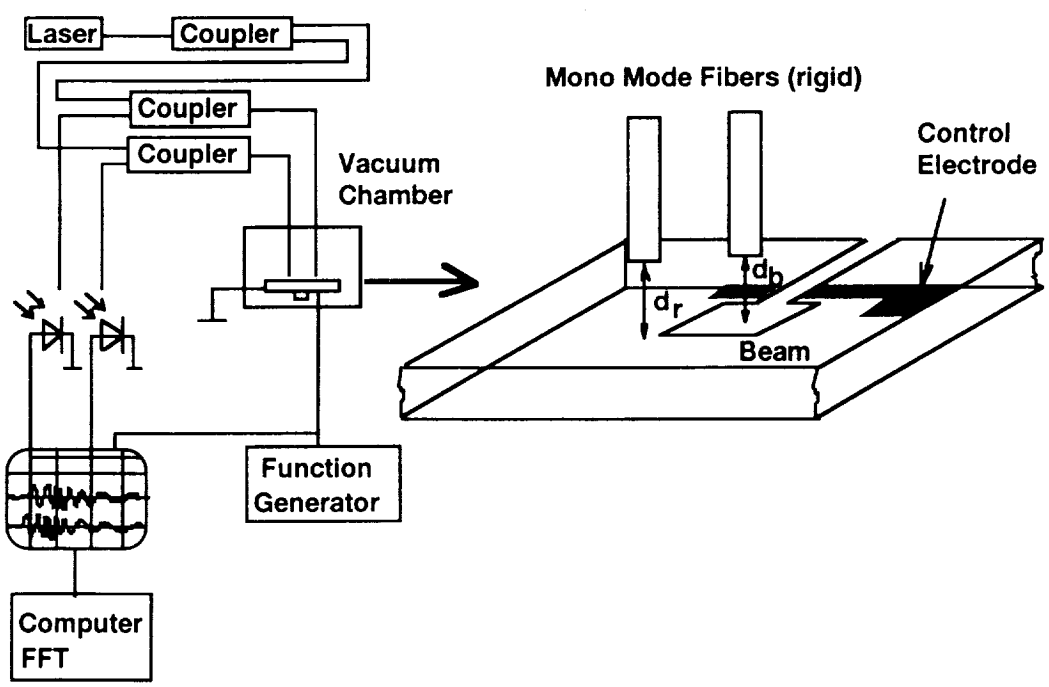


Figure 14. Setup for measuring the movement of the beams.

As mentioned above, the data resulting from this interferometric system is acquired using LabView. This software package is used to both control laboratory equipment and to analyze the resulting measurements through the use of virtual instruments. However, an illustrative example is included below. First data is collected and processed to show the magnitude and phase of the signal. This data is consequently processed to determine the spectral content. The information in the frequency domain will include the fundamental frequency as well as a beat frequency generated from the mixing that occurs due to the nature of the interferometric system. This beat frequency comes from the convolution of the two signals, one is the reference and the other is the beam itself. Figure 15 shows the results of the analysis with the spectral content of the beam response clearly shown.

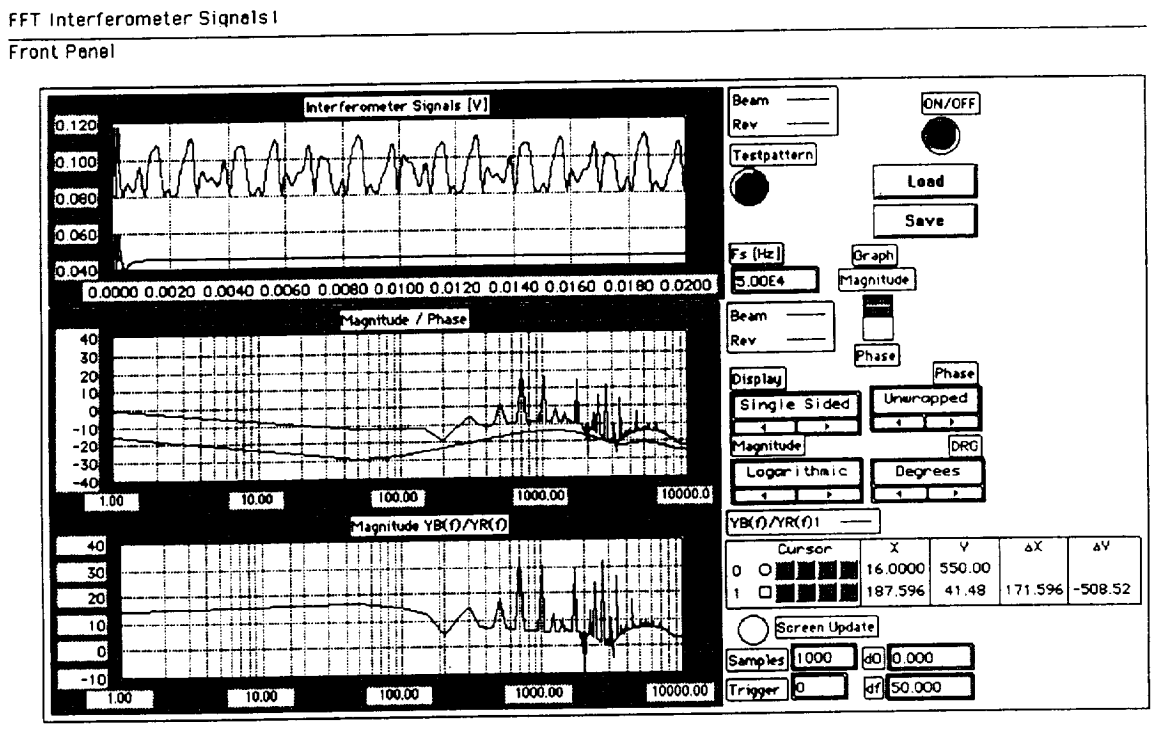


Figure 15. Results screen of virtual instrument showing the spectrum producedfrom the Fourier Transform and the impulse response of the beams.

CONCLUSIONS

The evolutionary development of a micromechanical vibration sensor has been described Starting from a preliminary concept implemented using the ARPA sponsored MUMPS at MCNC, the fabrication and characterization of a number of resonant vibration sensor was carried out. During characterization of this type of sensor, it was determined that viscous damping imposes severe limitations on the performance of the simple resonators indicating the need for vacuum sealed encapsulation of the devices.

Finite element modeling efforts were able to provide qualitative information. The fact that the Young's Modulus for Poly-Si can assume a wide range of values requiring experimental determination for each particular process, reduces the quantitative potential of the FEA in earlier design stages.

The use of a closed feedback control loop in a configuration known as force balanced vibration sensor, has the advantage of enabling broadband operation thus simplifying the number of cantilever sensors required at expense of higher processing complexities. At present we are still carrying out the various fabrication steps and currently this second set of devices is awaiting final processing.

Metrological characterization of the devices has been implemented using a fiber optic interferometric technique developed at InterScience that has been succesfully applied for directly measuring deflection and frequency response of cantilever type vibration sensors.

---

1 W. Benecke, L. Csepregui et al, "A frequency selective, piezoresistive silicon vibration sensor", 3rd International Conference on Solid-State Sensors and Actuators, 1985, pp. 105-108.

2 M.E. Motamedi, et al., " Application of electrochemical etch-stop in processing silicon accelerometer", Proc. 1982 Electrochemical Society Meeting, May 9-14, p 188.

3 DoD sponsored both efforts under USAF contract #F29601-94-C-0095 and under ARPA/BMDO contract number DASG60-94-C-0081.

4 D. Koester, R. Mahadevan and K. Markus, "MUMPs: Introduction and Design Rules", Rev. 3, MCNC, Oct. 1994

5 M.J. Usher et al., "A miniature wideband horizontal-component feedback seismometer", Journal of Physics E: Scientific Instruments 1977 Vol. 10

A MICROINSTRUMENTATION SYSTEM FOR REMOTE ENVIRONMENTAL MONITORING

Andrew Mason, Wayne G. Baer and Kensall D. Wise

Center for Integrated Sensors and Circuits
Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109-2122, USA

71144
030675
6P.

ABSTRACT

This paper reports a hybrid micro-instrumentation system that includes an embedded microcontroller, transducers for monitoring environmental parameters, interface/readout electronics for linking the controller and the transducers, and custom circuitry for system power management. Sensors for measuring temperature, pressure, humidity, and acceleration are included in the initial system, which operates for more than 180 days and dissipates less than 700μW from a 6V battery supply. The sensor scan rate is adaptive and can be event triggered. The system communicates internally over a 1MHz, nine-line intramodule sensor bus and outputs data over a hardwired serial interface or a 315MHz wireless link. The use of folding platform packaging allows an internal system volume as small as 5cc.

INTRODUCTION

The development of highly-integrated "smart" microsystems merging sensors, microactuators, low-power signal-processing electronics, and wireless communication promises to have a significant and pervasive impact during the coming decade [1], finding applications in such diverse areas as industrial process automation, health care, automotive systems, and environmental monitoring. These systems will require the high-density integration of state-of-the-art circuitry, sensors formed using a variety of process technologies, and intelligent system management. This paper describes a generic multiparameter sensing system for environmental monitoring that could serve as a prototype for such devices. The microsystem is implemented in a hybrid fashion as shown in Figure 1. This architecture allows the process technologies for each of the sensing elements and circuit components to be individually optimized and allows high precision sensors to be fabricated without unduly complicating the production of the integrated circuits. Additionally, a hybrid implementation allows commercial components to be used and individual elements to be updated without modifying the entire system.

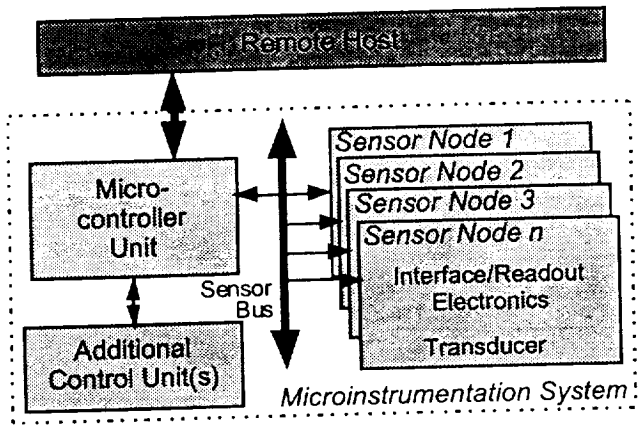


Figure 1: Hybrid system architecture for a microinstrumentation system.

SYSTEM DEFINITION

Based on the generic microinstrumentation system architecture shown in Figure 1, a multi-element sensing system has been developed. A block diagram of the system is shown in Figure 2, and specifications for the prototype system implementation are listed in Table 1. The microsystem is built around an embedded Motorola 68HC11 microcontroller unit (MCU) having on-chip memory, an 8b ADC, a timer, and serial communications hardware. The MCU communicates with the front-end transducers via a nine-line intramodule sensor bus and custom interface circuitry integrated on the transducer chips or on a separate hybrid. Sensor data collected by the MCU is calibrated in-module, stored, and sent out either through a hardwired RS-232 I/O port or via an on-board telemetry device. A custom power management chip performs several functions for minimizing power consumption in the battery powered system. The system employs an open architecture that permits it to be populated as desired by transducers using a mix of technologies. Transducers for measuring barometric pressure, altitude, humidity, temperature, and acceleration are included in the initial system.

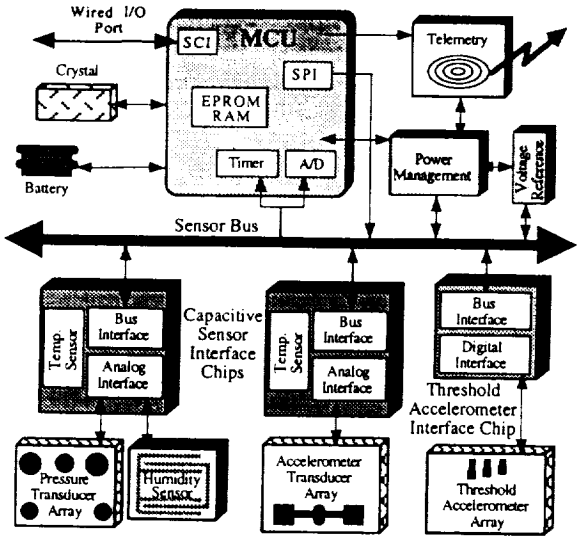


Figure 2: Block diagram of the microinstrumentation system.

Table 1. Specifications for the initial version of the microinstrumentation system.

Parameter	Typical Value
Power Supply	6V, External or Battery
Ave. Power Dissipation*	400-700μW
Lifetime (with battery)*	330-180 Days
Package Configuration	Card or Wristwatch
System Volume	5cc (wristwatch structure)
Packaged Volume	25cc (card structure)
Sensor Bus Frequency	1MHz
Wireless Range	≥100 feet
Temperature	-20 to +60°C ±0.2°C
Barometric Pressure	600-800 Torr ±25mTorr
Humidity	30-90%RH ±3%RH
Vibration/Acceleration	-2 to +2g ±0.1g

* scan rate dependent

During normal operation the system runs in a periodic scan mode in which data is collected from each sensing element and stored in the MCU's RAM. Each transducer has a block of code in the MCU which contains information regarding the sensor characteristics, self-test procedures, etc. Floating point software routines allow the sensors to be calibrated and digitally compensated for cross-parameter sensitivities (e.g., temperature dependence) in-module before the data is delivered to a remote host for further analysis. Digital compensation algorithms vary depending on the sensor response but have been chosen to minimize processing time, power consumption, and MCU memory. The compensation methods currently employed are look-up tables and polynomial evaluation [2].

Each sensor scan is followed by a sleep period in which the system enters a low power mode. The duration of this sleep period is determined adaptively based on the variation in the sensed parameters with previous measurements. At the end of the sleep period the system is awakened and the cycle starts again with another sensor scan.

INTERFACING TO THE SENSOR BUS

During each sensor scan the MCU communicates with the front-end devices over an intramodule sensor bus described in Figure 3. The sensor bus [3] contains three power leads (ground, a continuous 6V line from the system supply, and a switched 5V reference), four outputs to the front-end circuitry (chip enable and strobe handshaking signals, a 1MHz clock, and a serial data line), and two inputs from the front-end electronics (data out and data valid). The data valid line signals the MCU when valid data is present on the data out line and is also used to initiate interrupts triggered by the front-end devices. Figure 3 also shows the bus protocol for serial data delivered by the MCU, which includes a 4b chip address followed by a 5b sensor/actuator address, 3 command bits, and up to 12b of input data. This format allows each microsystem to access up to 16 interface chips which can, in turn, service up to 32 sensors and 32 actuators.

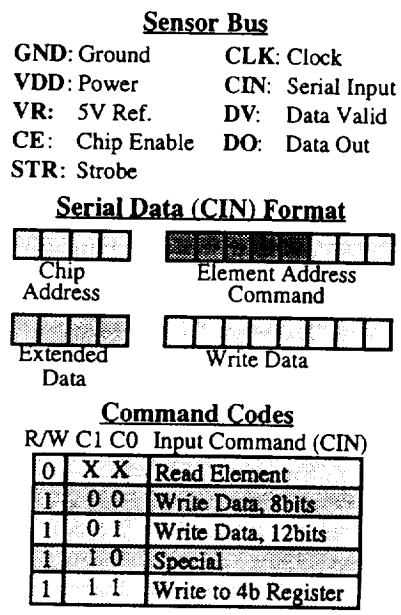


Figure 3: The intramodule sensor bus, serial data format, and command format used in the microsystem.

To interface between the front-end transducers and the bus, a standardized interface chip, shown in Figure 4, has been designed. This interface chip contains switched-capacitor readout circuitry [4] for up to six capacitive sensors with digitally programmable gain and three digitally selected reference capacitors that can be laser trimmed. The chip also contains all bus interface circuitry, a 3b

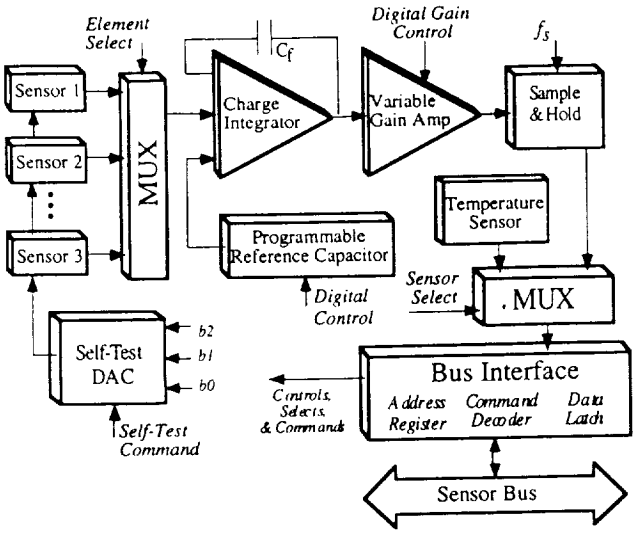


Figure 4: Block diagram of the capacitive sensor interface chip.

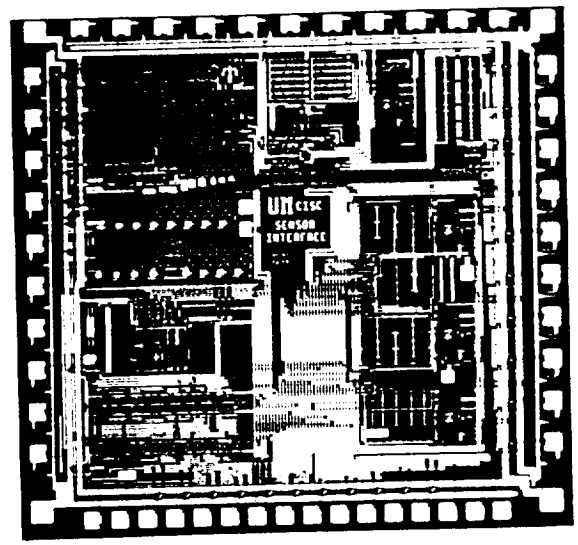


Figure 5: Integrated CMOS capacitive sensor interface chip for a standardized intramodule sensor bus.

DAC for sensor self-test, and an on-chip temperature sensor. A photograph of the 3.2mm x 3.2mm 3μm p-well 1M/2P CMOS chip is shown in Figure 5. The sensor interface chip operates from a 5V supply at about 400μA.

SENSORS FOR ENVIRONMENTAL MONITORING

The sensors in the initial system and their specifications are listed in Table 1. The temperature sensor, integrated on the sensor bus interface chip, is a simple temperature-dependent oscillator, although a bandgap sensor is being developed for future versions of the system. For barometric pressure, a multi-element, thin-diaphragm, micromachined, capacitive pressure transducer [5] is used. While one diaphragm measures pressure over the entire measurement range, the other elements measure segments of that range with much greater resolution (equivalent to changes in altitude of approximately one foot at sea level). Humidity is measured by a high aspect ratio inter-digitated hygrometer. This capacitive transducer utilizes micromolding [6] and electroplating to form electrodes that are separated by a polymer (e.g., DuPont PI2723) having a moisture-sensitive dielectric constant. Acceleration is measured by a bulk-silicon capacitive microaccelerometer with overrange protection and force feedback electrodes [7]. This device has a bandwidth of 75Hz using a bridge structure with four folded beam supports. Many of these transducers have utilized optimal design of experiments [8] for calibration and testing. Future versions of the system will add sensors for gas type, gas purity, and acoustic inputs, as well as a link for a global positioning system (GPS). Thus, the device will be able to monitor many aspects of its environment and its position within that environment.

POWER MANAGEMENT

The microinstrumentation system can be powered though the wired I/O port or, as many applications demand, can operate as a wireless, battery-powered system using two 3V, 270mA-hr lithium batteries. To control power consumption and maximize the life of the battery-powered microsystem, a power management (PM) chip has been designed. Figure 6 shows a block diagram of this 2.2mm x 2.2mm chip fabricated using the same technology as the interface chip above. The primary function of the PM chip is to control and monitor system functions between sensor scans when the MCU goes into a low-power sleep mode and power is removed from many system components. The duration of this interval is programmed by a 4b coded input from the MCU and can range from 40 seconds to 8 minutes. The MCU adaptively adjusts this interval based on the amount of variation in the sensed parameters. An on-chip clock generator and counter provide the PM timing functions that wake the system from its sleep mode. This chip also includes integrated switches for shutting off the voltage reference and the telemetry device. The PM chip also contains circuitry for converting the telemetry data to the necessary 3V level and provides pull-ups at the appropriate MCU inputs.

During the sleep period, only the MCU (in sleep mode), the power management chip, and a threshold accelerometer interface chip remain powered, keeping the continuous power

dissipation under 400μW. The threshold accelerometer is a micromachined device with three cantilever beams that act as switches measuring accelerationat three different thresholds (1.5g, 10g, and 100g). These switches feed to a 2.2mm x 2.2mm threshold accelerometer interface chip formed in 3μm p-well 1M/2P CMOS technology . This chip is programmed by the MCU through the sensor bus to monitor one of the threshold values, latch results, and wake ·the system from its sleep mode for an event-triggered response. The addition of this device to thesystem allows slowly-changing environmental variables to be scanned at a wide time interval while continuously (and autonomously) monitoring higher-frequency vibration activity. The combination of these functions enables the microsystem to operate using 700μW at the maximum scan frequency, providing a battery-powered lifetime of about 180 days.

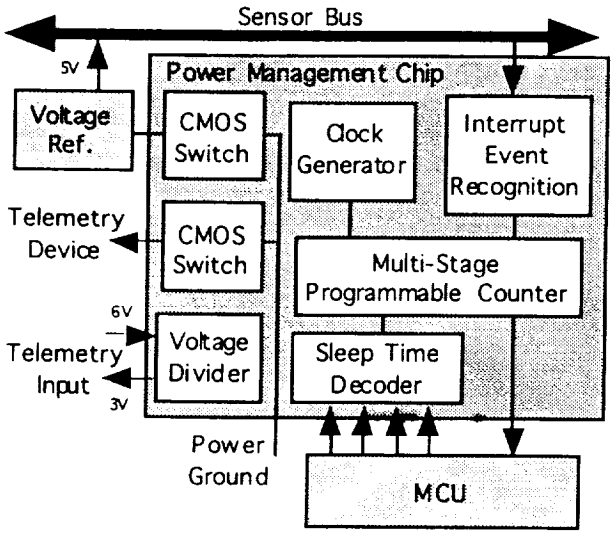


Figure 6: Block diagram of the power management chip and its connections to the system.

EXTERNAL I/O

The system offers both a hardwired external port and a wireless link. Included in the eight-line hardwired bus are power and ground, two lines for asynchronous serial RS-232 communications, and four advanced feature selects (e.g., programming toggles). For normal data transfer operations this bus can be reduced to four lines. The telemetry device on the microsystem is a commercially available 315MHz, amplitude-modulating transmitter (HX1005; RFM, Inc.) driven at a 3kHz bit rate. This device operates from a 3V supply and requires an average current of 4mA. The transmitter is active only in battery powered operation where its 3V supply can be tapped off the center contact of thetwo series 3V batteries. With a superheterodyne receiver and a one inch loop antenna, a range of over 100 feet has been observed. The message format for wireless transmission consists of eight data bits precededby a start bit and followed by a parity bit and a stop bit. Additional error checking is handled by software at the receiver end, and some data compression is performed in-moduleby the MCU. Future additions to the microsystem will include a 2.4GHz link using a micromachined antenna structure expected to produce a range of several thousand feet.

PACKAGING

A major requirement of this system is that it be very small and still compatible with a variety of sensor technologies. The packaging scheme must permit selective environmental access for those sensors requiring it and should permit repair or replacement of defective chips during test. In the initial version of the microinstrumentation system, a three-level printed circuit board (PCB) has been used to integrate the hybrid system components. Figure 7 shows a microinstrumentation cluster populating the 65mm x 35mm PCB which provides access to all system components during testing. This "card-like" packaging structure with two coin cell batteries attached below the board is placed into the cavity of an anodized aluminum outer case. The case has an external volume of approximately 25cc and provides battery access from the back. An O-ring, seated between the PCB and the outer case, seals the system components from the external environment while providing access to the humidity and pressure transducers. The transmitter's loop antenna is wrapped into a groove on the perimeter of the case.

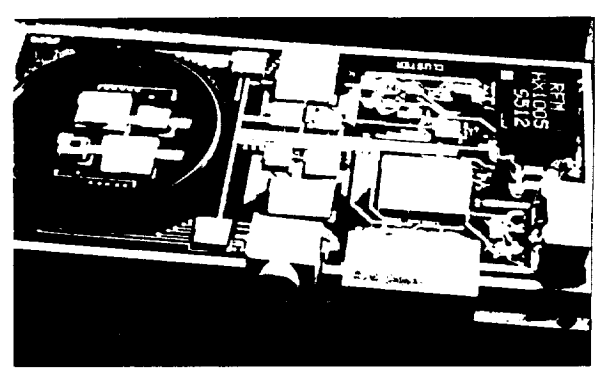


Figure 7: Complete microinstrumentation system mounted on the printed circuit board version of the system packaging.

An additional package that has been developed to further reduce the system volume is shown in Figure 8 and consists of several micromachined silicon platforms connected by gold-plated silicon ribbon cables [9]. The platforms include sites fora

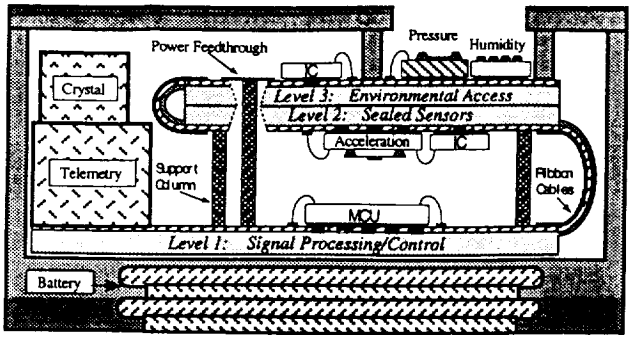


Figure 8: Three-level folding-platform packaging using micromachined silicon platforms and ribbon cables.

number of sensing chips and, when folded, form three levels: one for the MCU and other system electronics, one for sensors that do not require environmental access, and a final level for sensors that do. The system is populated as desired, tested, and then folded into the required package shape which provides room for large components such as the transmitter and crystal. The use of silicon platforms allows high-density interconnects, a low-resistance ground plane under the components, the possibility of in-platform switching/buffering, and compatibility with micromachined antenna structures. Figure 9 shows a prototype platform assembly before it is folded into a 5cc wristwatch-size cavity.

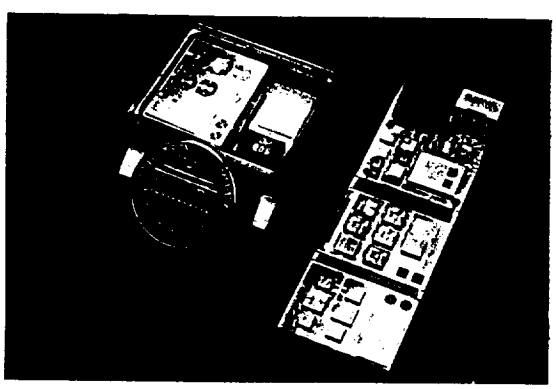


Figure 9: Prototype microinstrumentation system using a silicon micromachined platform assembly.

CONCLUSIONS

A generic structure for a multi-element sensing system has been developed and applied to a specific microinstrumentation system for environmental monitoring. Transducers for pressure, humidity, temperature, and acceleration are present in this system along with a microcontroller for in-module sensor calibration, digital compensation, and system control. A standardized intramodule sensor bus is used along with interface/readout circuitry that links the transducers to the embedded microcontroller. A custom circuit provides additional power management control permitting the system to operate up to 330 days at 400μW (average) from a 6V battery supply while communicating with a remote host through a wireless transmitter that provides a range greater than 100 ft. Both printed circuit board and micromachined silicon platform structures have been developed for system packaging.

ACKNOWLEDGMENT

The authors would like to thank the faculty, staff, and students of the Center for Integrated Sensors and Circuits, University of Michigan for their assistance in this research. This work was supported by the Advanced Research Projects Agency under contract JFBI 92-149.

REFERENCES

1. K. D. Wise, "Integrated Microinstrumentation Systems: Smart Peripherals for Distributed Sensing and Control," *Digest IEEE Int. Solid-State Circuits Conf.*, pp. 126-127, February 1993.
2. S. B. Crary, W. G. Baer, J. C. Cowles, and K. D. Wise, "Digital Compensation of High-performance Silicon Pressure Transducers," *Sensors and Actuators*, A21-A23, 1990, pp. 70-72.
3. N. Najafi and K. D. Wise, "An Organization and Interface for Sensor-Driven Semiconductor Process Control Systems," *IEEE Journal of Semiconductor Manufacturing*, pp. 230-238, November 1990.
4. Y. Park and K. Wise, "An MOS Switched-Capacitor Readout Circuit for Capacitive Pressure Sensors," *Proc. IEEE Custom Circuit Conf.*, May 1983, pp. 380-384.
5. Y. Zhang and K. D. Wise, "A High-Accuracy Multi-Element Silicon Barometric Pressure Sensor," *Digest Transducers '95*, Stockholm, V. 1, pp. 608-611, June 1995.
6. A. B. Frazier and M. G. Allen, "Metallic Microstructures Fabricated Using Photosensitive Polyimide Electroplating Molds," *IEEE J. Microelectromech. Systems*, *2*, pp. 87-94, 1993.
7. K. J. Ma, N. Yazdi and K. Najafi, "A Bulk-Silicon Capacitive Microaccelerometer with Built-in Overrange and Force Feedback Electrodes," *Digest, IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, pp. 160-163, June 1994 .
8. S. B. Crary, "Reducing Sensor Calibration Expense using Optimal Design of Experiments," *Proceedings of the Ninth Annual Conference and Exposition of Sensors and Systems*, Cleveland, OH, pp. 475-484, September 1994.
9. J.F. Hetke, J. L. Lund, K. Najafi, K. D. Wise, D. J. Anderson, "Silicon Ribbon Cables for Chronically Implantable Microelectrode Arrays," *IEEE Trans. On Biomedical Engineering*, pp. 314-321, April 1994.

7145
230676
22 p

SPACECRAFT APPLICATIONS OF COMPACT OPTICAL AND MASS SPECTROMETERS

N. M. Davinic and D. J. Nagel
Naval Research Laboratory
Washington, DC 20375-5345

ABSTRACT

Optical spectrometers, and mass spectrometers to a lesser extent, have a long and rich history of use aboard spacecraft. Space mission applications include deep space science spacecraft, earth orbiting satellites, atmospheric probes, and surface landers, rovers, and penetrators. The large size of capable instruments limited their use to large, expensive spacecraft. Because of the novel application of microfabrication technologies, compact optical and mass spectrometers are now available. The new compact devices are especially attractive for spacecraft because of their small mass and volume, as well as their low power consumption. Dispersive optical multi-channel analyzers which cover the 0.4-1.1 μm wavelength range are now commercially available in packages as small as 3 x 6 x 18 mm, exclusive of drive and recording electronics. Mass spectrometers as small as 3 x 3 x 3 mm, again without electronics, are under development. A variety of compact optical and mass spectrometers are reviewed in this paper. A number of past space applications are described, along with some upcoming opportunities that are likely candidate missions to fly this new class of compact spectrometers.

I. INTRODUCTION

This paper focuses on two classes of compact sensors, both spectrometers, which are already commercially available or now under development. They are optical and mass spectrometers, both of which have clear promise for use in spacecraft applications. Their smaller weight and lower power requirements, as well as their potentially greater reliability, will clearly have a major impact on new spacecraft. This is true for two reasons: (a) they permit putting greater functionality on conventional spacecraft, for the same weight and power, or reducing the launch and housekeeping needs for the same functionality, and (b) they enable the design of radically new miniature spacecraft which have not been possible before the emergence of microsystems (1).

Microsystems, for example, microelectromechanical systems (MEMS), are particularly attractive for spacecraft. Specifically, the motivations for designing microsystems into spacecraft include the following. Smaller instruments can

enable the use of smaller launch vehicles. Smaller launch vehicles translate to lower program costs, more flight opportunities, and possibly less program risk. Microsystems can similarly reduce secondary program costs such as facilities, integration and test, and transportation costs by reducing the physical scale of these operations. Another motivation for using microcomponents is that they permit the collocation of sensors with their electronics, which can yield better signal to noise performance and more efficient packaging. Components and subsystems which are lighter and smaller can relieve a variety of mission requirements. If they require less power, spacecraft can be designed with smaller solar arrays (often the largest single contributor to spacecraft surface area) and smaller batteries (often the largest single contributor to spacecraft mass). Microsystems can improve the capability and performance of a spacecraft by making it possible to fly more sensors, either of the same variety, for redundancy, or of different types, for improved functionality.

Practical factors and limits to miniaturization have to be evaluated. It is only somewhat beneficial if one or more of the smaller or less power hungry subsystems is miniaturized, because the weight and power budget of the entire craft is little reduced. Also, reduction in the size of a component has limits, because of the need for a package which can be handled, fixed in place and connected appropriately. Other limits on microsystems concern their performance, rather than merely their physical characteristics. Several examples of performance limits have been widely discussed. Among them are the aperture limits for small optical systems, the need to match antenna sizes to radio-frequency wavelengths and the fact that the very small proof masses in MEMS accelerometers are limited by thermal (Brownian) noise.

The hardware for any system can be subdivided into a number of subsystems. For spacecraft, these include the structure, power, thermal control, computer, communications and the payload (2). For example, the payload in many cases includes sensors, along with additional control electronics and data storage components. Each of the subsystems on a spacecraft or payload, is a candidate for miniaturization, either in the near term or in the future. Sensors for either housekeeping or as part of the payload are very attractive targets for miniaturization.

II. SPECTROMETERS

Like imagers, spectrometers provide an abundance of data. The spectral lines each have at least a position (energy or mass) and an intensity. For both optical and mass spectra, the position of lines gives qualitative analyses (what is it?) while the intensities yields quantitative analysis (how much?). The continuum

in many optical spectra also provides a great deal of information. Both the lines and continuum, properly interpreted, indicate the mechanisms which are active in the production of the observed spectra. Optical spectra can also be used for the characterization of hot or energetic sources. Mass spectra provide data on the makeup of natural plasmas, for example, ionization in the upper atmosphere, and plasmas produced by hypervelocity impact of man-made objects.

The next section deals with compact optical spectrometers, many of which are already on the market. The third section focuses on miniature mass spectrometers which are under development, and may be available in small numbers in the foreseeable future. In both these sections, some background is provided before a review of specific instruments. Then the fourth section reviews some earlier space applications of optical and mass spectrometers, before considering new opportunities for utilization of available and emerging compact spectrometers aboard spacecraft in the fifth section. Special emphasis is given to the coming Clementine II program which might include both optical and mass spectrometers for observation of the impact of microsat interceptors with a sequence of three asteroids.

III. COMPACT OPTICAL SPECTROMETERS

The utilization of optical spectrometers on earth and in space has a long and rich history. There are two major trends now in the evolution of optical spectral instruments. The first is continued reductions in size, which is the focus this paper. The other is full integration of spectrometers with imagers to provide a "cube" of information, namely two spatial dimensions (say, X and Y) and the spectral distribution at each pixel (the wavelength dimension). Such hyper-spectral imagers are instruments which work in a way analogous to our vision, which provides an image with colors. At this time, hyper-spectral imagers have been integrated into aircraft, but not into spacecraft. Multi-spectral imagers, such as on Landsat, which provide images in a few wavelength bands, are really not spectrometer systems in the usual sense. That is, they provide intensities in only a few wavelength bands and not the entire spectrum in a more or less continuous manner.

Many good references on the basics of optical spectrometers are available (3,4). There are two main types of optical devices, each with a pair of major variants. The first is dispersive instruments based on prisms or gratings. The second is interferometric spectrometers, such as the Fabry-Perot or Michelson types. Compact optical spectrometers now available commercially employ diffraction grating dispersion elements in reflection geometries. Their description constitutes the body of this section. MEMS spectrometers, with moving

components providing dispersion in Fabry-Perot and other designs, are under development. They will be mentioned briefly at the end of the section.

Micro-optical spectrometers behave very much like their larger counterparts. The wavelength of light (from the ultraviolet near 0.2 μm, through the visible 0.4-0.7 μm, to the near infrared around 1.0 μm) is small compared to geometries in even compact spectrometers. The light collection efficiency depends on optics external to the spectrometer, the size of the entrance slit, and the angular acceptance (called the f number). The ability of the spectrometer to register photons, namely its efficiency, varies across the spectrum and depends on the individual efficiencies of the grating and the detector. The resolution is determined by the size of the entrance slit, the geometry of the grating and spectrometer, and the size of the detector elements (pixels).

The range over which a given spectrometer works depends on its design, the grating ruling density, and the size of the instrument. When only point electronic detectors were available, grating spectrometers had to be scanned to record an entire spectrum sequentially (serial readout). With the availability of solid-state detector arrays, such as charge-couple devices (CCD) and photo-diode arrays (PDA) in the 1980s, it is possible to capture an entire spectrum simultaneously (parallel readout). This mode, the so-called optical multi-channel analyzer (OMA), is similar to the way in which photographic film was used to record spectra before there were any electronic detectors.

Many standard, laboratory-size OMA spectrometers and spectro-photometers are on the market. Three commercial grating optical multi-channel analyzers, which are compact and relatively inexpensive, will now be described. To our knowledge, none of these has been flown on a spacecraft, but all are candidates for missions to be described in the fifth section.

A spectrometer which literally fits in the palm of a hand is shown in Figure 1. It, and variants of it, are manufactured by Ocean Optics, Inc. (5). Light arrives at the instrument in an optical fiber, the end of which serves as the entrance slit. Hence, one can trade resolution (linear in the fiber core diameter) for intensity (quadratically dependent on the diameter) in a straightforward manner. The instrument is assembled from separate components, a folding mirror, grating and detector array, and then sealed. That is, switching of internal components is not possible in the current versions. The company offers several different input fibers and gratings so that the buyer can tailor the intensity, resolution and range to the application. It is possible to cover a 500 nm range, say, from 250 to 750 nm, with 2.5 nm resolution, or a 125 nm region with 0.7 nm resolution. The detector is a CCD containing 1024 elements. Spectra can be collected in as little as 8.2 msec, that is, it is possible to record over 100 spectra per second. As

illustrated, the instrument is mounted on a half-board which fits a standard PC. The computer performs both control and data recording functions, with subsequent spectral analysis being readily performed with either the vendor supplied, or third party, software. The cost of the spectrometer on the board, input fiber, and software is about $2000.

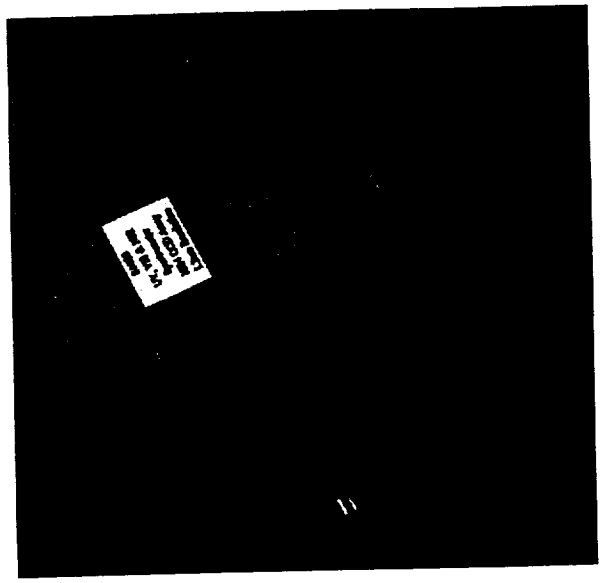


Figure 1. Optical grating spectrometer, fed by the optical fiber shown at the bottom, covers half a short board for a PC.

Another compact grating OMA is made by Carl Zeiss (6) and pictured in Figure 2. This system, which contains a grating and a diode array, is also fixed in construction. It is offered in two versions. One covers the 190-735 nm region with 2.2 nm resolution, while the second is good for the 320-1100 nm range with 2.2 nm resolution. The detector is a PDA with 256 elements. It can provide 12 bit information in 10.5 msec or, with different electronics, 14 bit data in 9 msec. While the spectrometer itself is quite compact (24 mm diameter body 28 mm long), it is mounted in a housing so that the entire unit is 40 mm x 50 mm x 65 mm in size (for the longer wavelength version). In addition, a separate external electronics box 60 mm x 100 mm x 100 mm is needed between the spectrometer and a PC plug-in board. The entire basic system including the spectrometer with the housing, 12-bit electronics box and PC board, and connecting cables costs $3000. The 14-bit electronics are an additional $2850.

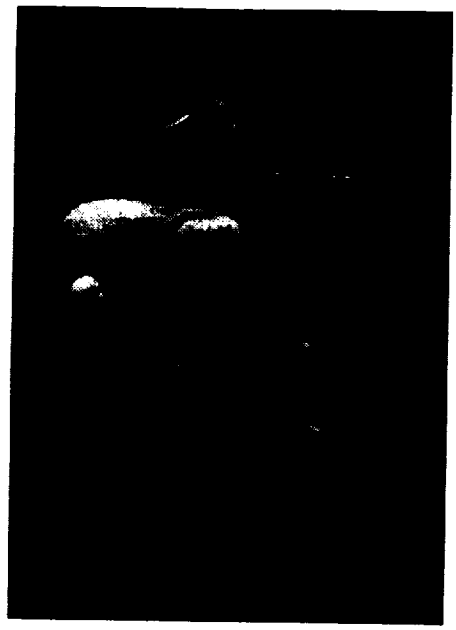


Figure 2. Compact optical spectrometer, with the input optical fiber, a grating under the curved top and detector array near the output pins.

The most compact commercial spectrometer is made by a German company called MicroParts Gesellschaft fur Mikrostrukturtechnik mbH (7) using the LIGA technique (8). This extraordinary device is sketched in Figure 3. The holder for the input optical fiber, the grating, and a 45 deg. mirror which deflects the light down into the horizontally mounted array are produced as a unit. It provides 7 nm optical resolution over the wavelength range from 370 to 850 nm. The spectrometer itself is only a few mm thick, 6 mm wide and 18 mm long. It costs about $600, including the 256-element PDA, without electronics. The electronics board permits spectral acquisition times variable from 40 to 1256 msec, with 16-bit resolution. The entire system, spectrometer with PDA, electronics, software, and power supply costs about $4100.

The focus in this section so far has been on available instruments. However, since tomorrow's product is commonly a recent or current research prototype, we will briefly survey some other reports of microsystem optical spectrometers in the remainder of the section.

An infrared spectrometer on a chip was fabricated with integral light guides and grating (9). It consisted of an InP substrate, an intermediate layer of InGaAs and a top layer of InP through which the radiation traveled. That is, the layout is generally similar to the commercial spectrometer shown in Figure 3, in which the optical radiation travels in air. The infrared transmissivity within a

solid semiconductor was exploited in this developmental device. Radiation in the 1.48 to 1.59 μm region was dispersed with 3.75 nm resolution.

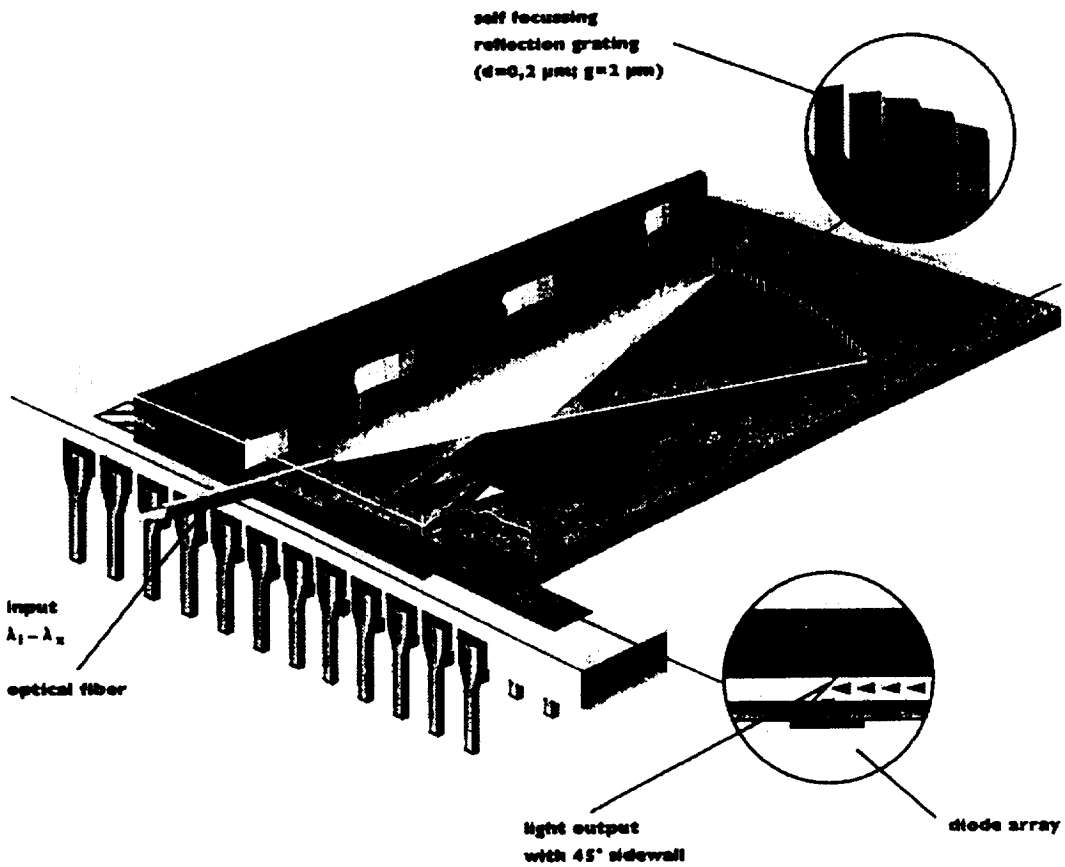


Figure 3. Micro-optical-spectrometer, 18 mm in length and 6 mm wide, with fiber optic input and a silicon diode detector array.

A silicon spectrometer was fabricated by deep etching of two wafers and their subsequent bonding to form an integrated device with a 4 mm total optical path length (10). Dispersion was obtained with a 32-slit transmission diffraction grating on the same wafer as a 200-element PDA.

There have been at least four reports of MEMS Fabry-Perot interferometric components. The first two came from the same group, which used electrostatic tuning of bulk micromachined and bonded structures. They achieved tunability in the 1.30-1.38 μm range by applying voltages up to 70V. The optical resolution was 0.9 nm (11,12). A similar approach in the visible region yielded 1.6 nm resolution at 450 nm (13). A process and design study for

surface micromachined silicon tunable interferometric arrays is also available (14).

A tunable, high-pass infrared MEMS filter based on deformable metallic structures was produced by the LIGA technique (15). Spectral resolution can be attained with this device by scanning the cutoff. Variation in the cutoff, which is similar to the waveguide cutoff at radio frequencies, from about 5 to 25 μm has been demonstrated.

There are many other papers on optical MEMS, where the components do not act as spectrometers. Functions such as light modulation, coupling of fiber optics and formation of beams for displays are being aggressively pursued. A review of optical MEMS is in preparation (16).

IV. COMPACT MASS SPECTROMETERS

Mass spectrometry has a long and rich terrestrial history, which can be traced back to around 1910. However, mass spectrometers have not been used in space anywhere near as much as optical spectrometers. Because they require contact with the material to be analyzed, mass spectrometers are limited to direct rather then remote sensing, both of which are possible with optical spectrometers. Mass-resolving instruments are evolving in the direction of smaller devices without much loss of capability. This is making them increasingly attractive for a wide variety of space missions, as will be discussed in the fifth section.

Good references on the fundamentals and applications of mass spectrometers are available (17,18). As with optical spectrometers, there are a few basic types of instruments which can be used to distinguish the masses, and therefore type of atoms and molecules. A conceptually simple approach to mass separation is taken in time-of-flight devices. In these, diverse ions are given the same energy by pulsed electron or photon (laser) excitation. They then move through a drift region in the instrument at different speeds because of their different masses. The drift region can be one directional, or can include an ion mirror, which serves to shorten the overall length of he instrument. The varying arrival times at the ion detector provides the mass spectrum.

The second class of mass spectrometers involves the use of time-varying, often radio-frequency, fields. Three major versions of these systems exist. In one variant, called fourier-transform ion cyclotron resonance mass spectrometry, a radio-frequency field is applied to a vacuum trap containing the ions to be

analyzed. Image charges in nearby plates provide a signal, which is fourier transformed to obtain the mass spectrum. In another, termed a quadrapole mass filter, four electrodes are biased with both steady and varying electric fields in such a way that specific masses can transit the filter. Recording a mass spectrum requires scanning of the pass band of the quadrapole analyzer. The third RF variant has multiple deflection plates along the track taken by he ions to be analyzed. The plate geometry and RF frequency are chosen so that the mass of interest will pass to the detector, but all other material will be shunted aside. Hence, this is also a sequential method of mass spectrometry.

The final class of mass spectrometer employs magnetic fields, either separately or in combination with electric fields, to disperse charged species with missing or added electrons. This approach can be used in either a scanning (series) mode, if the fields are varied appropriately, or for simultaneous (parallel) acquisition of a spectrum, if an array of detectors is available. The latter is amenable to either pulsed operation, as in a time-of-flight instrument, but it can conveniently be operated in a DC mode to accumulate a spectrum.

Development of smaller mass spectrometers requires shrinking all of the needed components, notably the source of ions, the means of dispersal in space or time and the detector. The method of ionization varies widely, depending on the material to be analyzed (gas, liquid or solid, and the chemical makeup) and the source of energy for the production of ions (optical, electrical or other means). The collection efficiency of the system is set by the external optics (if any), the size of the entrance slit and the geometry. The spectrometer efficiency depends on the quality of the internal vacuum and the efficiency of the detector. Mass resolution, the ability to separate neighboring isotopes different by only one atomic mass unit, is determined by the slit widths, dispersion element(s) and the detector pixel size. The mass range over which a particular device is useful depends on the design, the dispersion element, and the size of the key components. All of the different classes of mass spectrometers described above, except of the fourier-transform ion cyclotron resonance approach, are being made more compact with conventional technology or being designed anew with micro-machining technology. Some commercially-available, relatively-compact mass spectrometers, and small instruments under development, will be described in the remainder of this section.

A few companies offer time-of-flight mass spectrometers for laboratory applications. Comstock, Inc. is an example (19). They sell both linear and reflection models in which the drift region is 1 m. Now, they are offering more

compact 450 mm drift region devices in field-portable configurations. Adaptation of the new designs to space missions is more attractive, because of the lower mass in the smaller instruments.

A compact time-of-flight spectrometer is under development at the Johns Hopkins Applied Physics Laboratory (20). The mass analyzer is 200 mm long, 6000 mm^2 in cross section and weighs 500 grams, with the system weight being greater due to electronics, vacuum pumps and computer. Spectra from the prototype instrument have been measured in the mass-to-charge range of 0-500. This device is also being aimed at field use and is a candidate for space missions.

Quadrapole mass analyzers are widely available commercially because of their use in leak detectors for vacuum systems. Miniature quadrapole mass spectrometers are under development at the Jet Propulsion Laboratory (21). Unit mass resolution has been demonstrated from 4 to 86 amu. Large arrays of small, lithographically-produced mass filters are being envisioned for planetary, near comet and space plasma measurements.

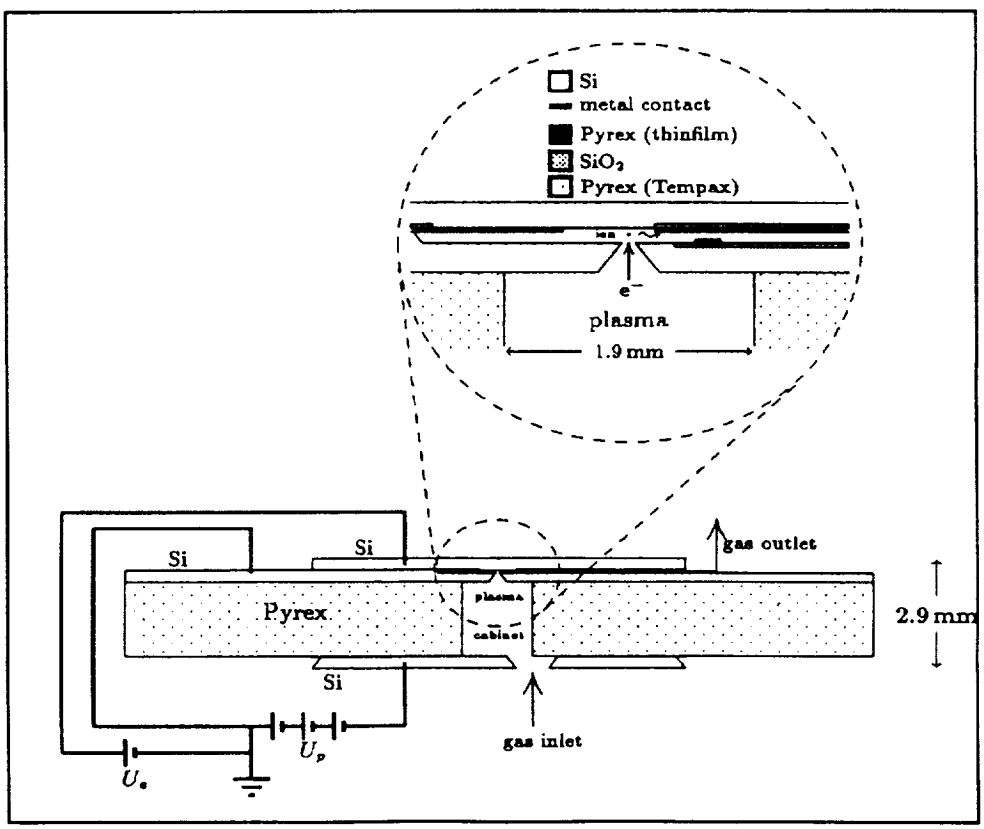


Figure 4. Cross section of a compact RF mass spectrometer.

Two current projects are seeking to build mass spectrometers "on a chip". In one of them, a combination of lithography, anisotropic etching of silicon, thin film deposition and anodic bonding of wafers is being used (22). The spectrometer, shown in Figure 4, included a plasma electron source, an ionization chamber and a mass separator in a $(3\ mm)^3$ volume. The separator consists of a series of plates to which an RF is applied (around 50 MHz). Data published one year ago showed a modest mass resolution of 18 (mass/delta mass) in a separator length of 2 mm.

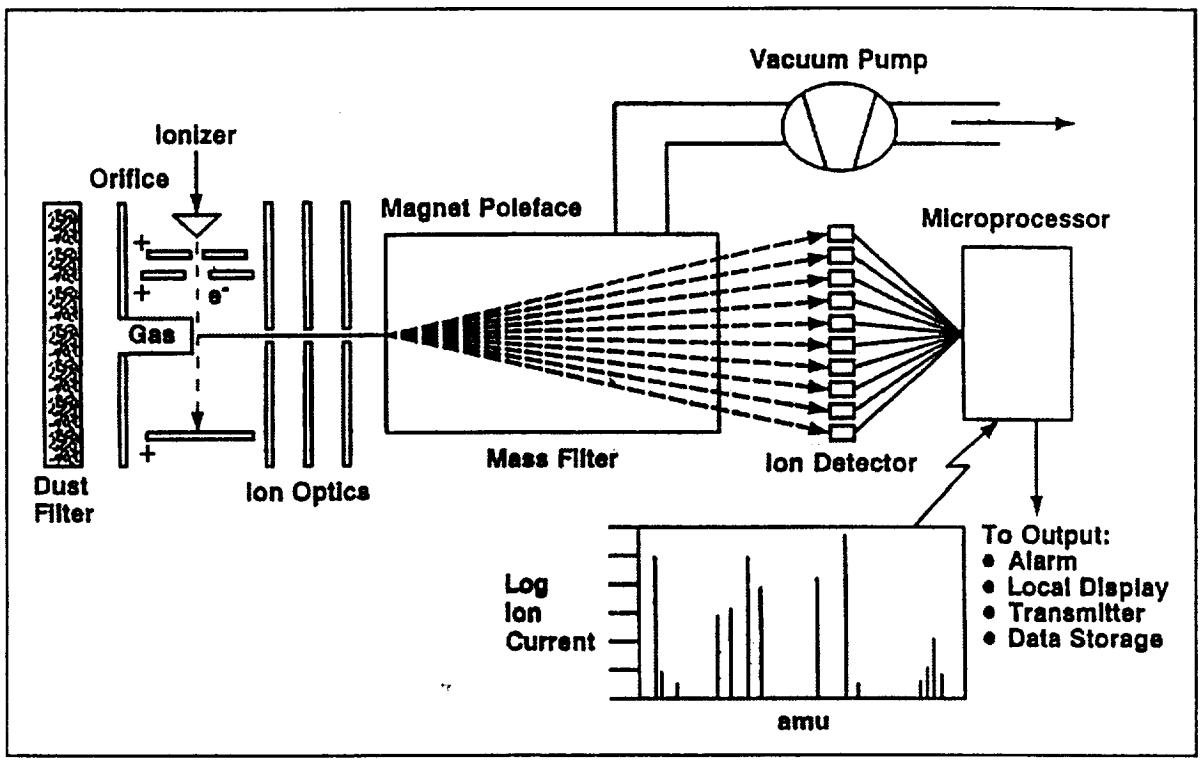


Figure 5. Plan view of a mass spectrometer on a chip.

An ambitious development project for a very small mass spectrometer is underway at Westinghouse, Inc (23). It involves producing four key components by silicon processing technology: (a) the electron ionization source, (b) a mass analyzer with crossed magnetic and electric fields, (c) an array of charge collection devices, and (d) micro-machined vacuum pumps. Production of pumps in series which are capable of achieving vacuum conditions in the atmosphere are a stressing aspect of this project.

One serious difference between optical and mass spectrometers for

ordinary uses deserves emphasis. Optical devices do not require a vacuum, as do mass spectrometers. For terrestrial uses, the need for a vacuum system can be the factor limiting the overall system size, weight and power. Micromechanical pumps are under development (24). However, these small pumps, and associated micromechanical valves, are mainly for the movement of gases against modest pressure differentials or for transporting liquids at pressures near one atmosphere. The potential application of compact mass spectrometers in the natural vacuum of space could reduce or obviate the need for the vacuum system normally associated with a mass spectrometer. This is especially attractive for the spectrometers "on a chip".

This completes our review of the compact optical and mass spectrometer technologies. Examples of past use of conventional spectrometers in space will be examined in the next section. Future possibilities for the exploitation of the emerging technologies discussed above will be presented in the following section.

IV. PAST AND CURRENT SPECTRAL MEASUREMENTS IN SPACE

Remote sensing is used to provide information about an object without being in physical contact with it. One way that information can be gathered is by measuring changes in the electromagnetic field associated with the object. Different parts of the electromagnetic spectrum are utilized for remote sensing (i.e. visible, ultraviolet, infrared). Spectrometers are used to remotely measure the spectral content in and near the visible region of the electromagnetic field of interest. In addition to spectral data, spatial or intensity information may also be required. Hybrid systems that provide both spectral and spatial information (imaging spectrometers) or spectral and intensity information (spectroradiometers) are in use today (25).

A variety of space related applications exist where spectrometry has, and will continue to play a significant role, see Table 1. Spectral imaging of the earth provides data on numerous environmental factors such as ozone depletion, vegetation characterization, ocean color (chlorophyll levels), and soil moisture content. Spectral characterizations are made of the earth's surface, both dry land and oceans, as well as the atmosphere at various elevations. Similar types of measurements can be made on other celestial objects including planets, comets, and asteroids. Lunar and planetary measurements can be made from orbiting platforms or surface landers.

	Optical Spectrometry	Mass Spectrometry
Earth Sensing		
• Earth Observation (Imaging Spectrometers)	X	
• Remote Sensing (Surface)	X	
• Atmospheric/ Ionospheric Mapping	X	
Planetary Sensing		
• Orbital Survey	X	
• Atmospheric Probes	X	X
• Landers	X	X
• Rovers	X	X
• Surface Penetrators		X
Asteroid / Comet Sensing		
• Orbital Survey	X	
• Landers	X	X
• Rovers	X	X
• Surface Penetrators		X
Solar Sensing	X	
Stellar Sensing	X	
Deep Space Sensing		
• Space Chemistry	X	X
• Plasma Characterization	X	X

Table 1. Summary of space applications for optical and mass spectrometers.

Earth observation from space offers a number of advantages compared to airborne and ground based techniques, including larger field of regard, synoptic coverage over that field of regard, some immunity to local disturbance effects, and the potential for continuous coverage (geostationary systems). There are several significant disadvantages to space-based earth sensing including severe weight and volume restrictions for payloads due to launch vehicle limitations, no access to sensors for maintenance or repair, and harsh operating environments (radiation dose, atomic oxygen, thermal cycling). The fact that several hundred earth sensing satellites have been launched by more than 20 countries indicates that often the benefits outweigh the disadvantages.

Life cycle cost is another important factor in evaluating space-based systems. Launch vehicle costs are extremely high, $15M - $400M, so program life cycle costs are high. Payload development costs are high due, in part, to very high reliability requirements for space systems that, once launched, cannot be repaired. The aerospace industry has become very sensitive to these cost concerns and the current trend is towards smaller, cheaper, and in some cases less capable

satellite systems (away from larger, expensive, infrequently flown, but very capable systems). This trend is an ideal opportunity for the introduction and use of compact optical and mass spectrometers for remote and direct sensing.

Space-based platforms have been used for remote sensing for years. One of the first remote sensing platforms, Tiros 1, was launched by the United States in 1960. Tiros 1 was a meteorological satellite which returned more than 22,000 images to earth. The Tiros satellites were the precursors to the National Oceanic and Atmospheric Administration (NOAA) Geostationary and Polar Operational Environmental satellites (GOES and POES) (26) .

The European Space Agency (ESA) is currently flying a number of environmental sensing platforms. The Envisat 1, scheduled for launch in 1998, will be flying a medium resolution imaging spectrometer. The instrument will work in the visible and near IR ranges and will be used for water and land quality measurements such as plankton content, water depth, bottom type classification, and pollution monitoring (27).

The Total Ozone Mapping Spectrometer Earth Probe (TOMS-EP) will be launched by the US in 1997 to continue monitoring of atmospheric ozone levels. The instrument is a 308.6-360 nm Fastie-Ebert Monochrometer used for ozone content characterization in six wavelength bands. This program uses small satellite technologies throughout the spacecraft and is further indication of the current trend in the aerospace industry towards smaller, cheaper satellites (28).

Various spectrometric systems have also been used on planetary and lunar probes. The latest lunar mission, Clementine, was launched in 1994. The mission profile included a lunar mapping phase and a near earth asteroid fly-by. The program was designed to be a flight demonstration for a set of light weight space based sensing technologies developed by the Ballistic Missile Defense Office (BMDO).The satellite was successfully launched on schedule in January 1994 and completed the lunar mapping phase in May 1994. During the mapping phase Clementine produced 1.8 million multi-spectral digital images (visible, ultraviolet, and infrared) of the moons surface (29).

Optical spectroscopy has long contributed to solar and stellar astronomy. Solar spectral studies at wavelengths below 200 nm, necessarily done from space because of atmospheric absorption, probe the appropriate range of temperatures (30). Stellar and other deep space spectral measurements are epitomized by current results from the Hubble Space Telescope (HST).

The HST, launched in April 1990, is the largest optical telescope and the largest civilian payload ever flown on the shuttle. Of the five primary scientific

instruments integrated on the HST two were spectrometers. The Faint Object Spectrograph uses the full HST resolving capability and operates in the 115 - 850 nm range allowing for measurements of stellar objects up to 14 billion light years away. The High Resolution Spectrometer is optimized for higher resolution observations in the UV (110 - 320 nm). This device has experienced operational difficulties due to power supply fluctuations on orbit. The Near-IR Camera & Multi-Object Spectrometer (NICMOS) will be integrated onto the HST during a later shuttle servicing mission scheduled for 1997. NICMOS will use three separate grating spectrometers operating in the 1.0 - 3.0 μm range to simultaneously observe different portions of the instrument field of view (31).

Direct measurement of materials using mass spectroscopy has not been employed in space to the same extent as optical spectrometry. However, in-situ measurements on planets and other celestial bodies are possible. A current example is the Galileo mission, now on its final leg to Jupiter. The Heavy Ion Counter on the Galileo spacecraft uses direct sensing; it registers the characteristics of ions in the spacecraft's vicinity which enter the instrument. It does not form an image of the ions' source. In addition, the Galileo mission will be the first mission to make in-situ measurements of an outer planet's atmosphere. A separate atmospheric probe will be deployed from the Galileo spacecraft into the Jovian atmosphere to evaluate chemical composition at various altitudes. The probe uses an 11 kg neutral-ion mass spectrometer (1-150 amu) (32).

V FUTURE OPPORTUNITIES

A Clementine II mission has been proposed and is currently being studied by the Naval Research Laboratory and Phillips Laboratory. The Clementine II mission, see Figure 6, would involve three asteroid encounters and the use of "micro-satellites" as asteroid interceptors. The plan is to develop and launch an upgraded Clementine bus, sketched in Figure 7, that can support multiple interceptor micro-satellites. When the primary spacecraft is in close proximity to a selected asteroid a micro-satellite will be deployed. The micro-satellite will navigate to the asteroid using an onboard suite of imaging sensors. Any devices on the micro-satellites will have to be very compact since the weight of the entire micro-satellite will be on the order of 20 kg. The micro-satellite will impact the asteroid, generating a cloud of debris through which the primary spacecraft will fly using a suite of sensors to make various measurements.

Near each asteroid optical spectrometers would be used to characterize the composition of the asteroid surface by measuring the asteroid in various spectral bands. Once an impact occurs on the asteroid surface, spectroscopic

measurements of the internal asteroid structure can also be made. Mass spectrometers could be used to characterize some of the debris created by the impact if a means of capturing the material safely can be developed.

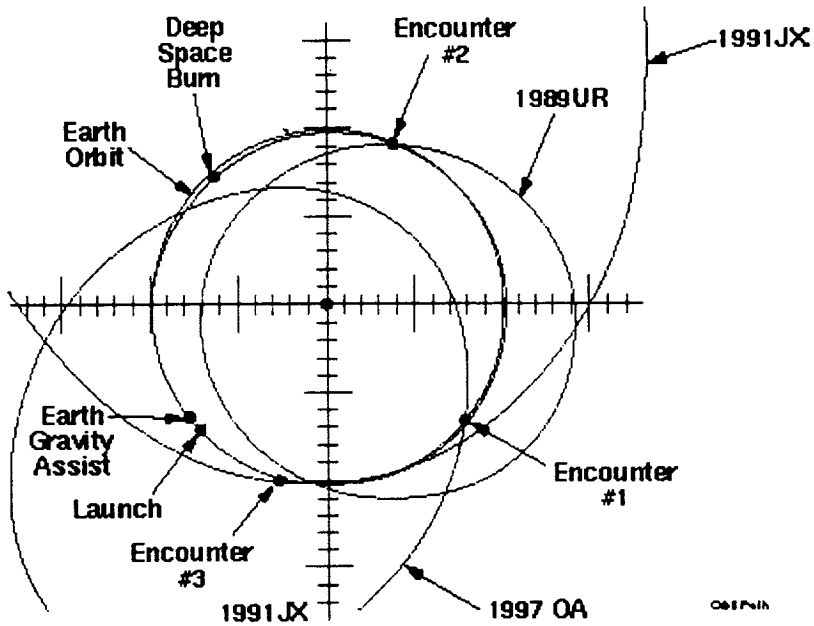


Figure 6. Clementine II Mission Profile

A presidential mandate requires that the current NOAA satellite programs be combined with the Defense Meteorological Satellite Program (DMSP). The data collection requirements of these two individual systems have significant overlap. Many of the observational parameters, such as ocean chlorophyll content and moisture profiles, require spectroscopic measurements. The combined program, the National Polar-orbiting Operational Environmental Satellite System (NPOESS), is currently evaluating instrument technologies for the future flight systems (33). While the present performance of compact optical spectrometers is not adequate for the primary systems, potential secondary applications are being discussed.

The NPOESS primary satellite constellation will consist of several large multi-instrument platforms. One possible application for compact spectrometers is as spares, either resident on the large primary host vehicles or on subsequently launched small satellites. As a secondary instrument on the host the compact devices could be bore-sighted with the primary systems and used in case of primary instrument fails to provide some data at degraded levels. A proposal is

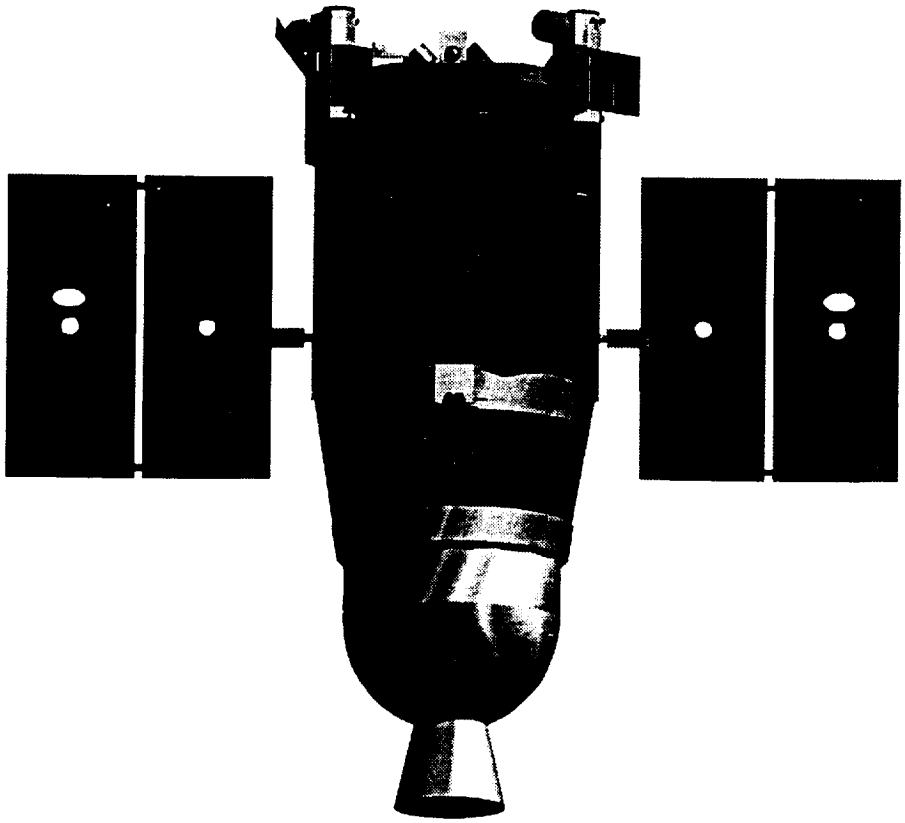


Figure 7. Clementine Spacecraft (2 Micro-Satellites Shown)

being investigated for using small satellites launched only after a primary instrument failure in order to supplement the larger vehicle system. This approach would allow instrument replacements to be delivered without replacing the entire primary satellite. The concept would require constellation maintenance to ensure that the original vehicle and the small satellite carrying the spare instrument remain in close proximity to each other (requirements exist for measurements from several instruments at the same time and location).

In general, the use of smaller satellites is a trend in industry, especially in remote sensing platforms. In the last two years several small imaging (< 500 kg) systems have been proposed: the CTA World View venture which uses a 235 kg satellite planned for launch in 1995 will provide 3 m resolution; the Lockheed Commercial Remote Sensing Satellite (CRSS) venture which delivers 4 m multi-spectral resolution and will be launched on an LLV2; and the Orbital Sciences Corporation (OSC) Eyeglass venture, which will provide 1 m panchromatic resolution, is slated for a Taurus vehicle. All three systems could benefit from adding on a small spectrometer payload to provide additional wide area (lower spatial resolution) coverage in other spectral bands.

The New Millennium Program (NMP) is an effort by NASA to provide frequent flight opportunities for new enabling technologies. The program started early in 1995 and is managed by the Jet Propulsion Laboratory, will create collaborative partnerships among NASA centers, government agencies, industrial firms, nonprofit research organizations, and academic institutions (34). Partners will participate in NMP teams from proposal, development and implementation through design, launch, and operation of validation flights. Five technology areas were selected as the focus of the project, one of which was Microinstruments and MEMS. All the technologies enable smaller, more capable spacecraft

NMP will fly several technology qualification flights per year and also transition newly flight proven technologies to operational systems in a timely manner. One of the first operational systems to take advantage of this process will be the Pluto-Express mission. The Pluto mission plan calls for launch of two spacecraft early in the next decade toward encounters with Pluto around 2010 or later. The mission objective is a low cost, initial reconnaissance of the Pluto system which would include: geology and morphology - high resolution global maps of the planet; surface composition - global composition maps of Pluto; and atmosphere composition - composition with vertical temperature and pressure profiles of Pluto's atmosphere. Optical and mass spectrometers are potentially applicable instruments, keeping the mission relatively low cost by assuring a small payload weight.

VII. CONCLUSION

Spacecraft applications for recently available and emerging compact spectrometers are abundant. Such instruments can yield a great deal of quantitative information in support of diverse space missions.

Optical spectrometers have a history in space spanning almost half a century. Compact devices now available sacrifice some performance, mainly resolution, because of their small scale. However, they are light-weight, small, require relatively little power, and are easily integrated into a spacecraft (because of their common integration with microcomputers for control and data acquisition). Current small commercial optical systems cover only a few of the design options; that is, it is straightforward to tailor the designs of compact spectrometers to the needs of a given mission.

Mass spectrometers have neither so long nor rich a history in space as do optical spectrometers. Nor are compact mass spectrometers as available

commercially today. However, their increasing availability, capabilities, and the design flexibility they enable, all bode well for further utilization. This particularly true for planetary exploration missions such as Galileo.

This paper has focused on applications of compact spectrometers for spacecraft sensors applications. There is also a potential for uses in support of space missions. For example, the condition of exterior materials, such as reflectivity, can be tracked with optical spectrometers. The Long Duration Exposure Facility could have provided dynamic degradation data if it had been equipped with small optical spectrometers. Diagnosis of in-flight ascent, or on station, conditions are also possible when the monitoring instruments are small and light weight Mass spectrometers could be used to monitor environments within spacecraft, either prior to launch or on station, and might be used as a quality assurance tool. Potential applications include the Space Station Freedom. In short, it seems reasonable to expect that truly compact spectrometers will be an enabling technology for a number of new spacecraft operational and supporting applications.

Acknowledgments

We thank U. Feldman and W Hunter for discussions and providing references.

References

1. "Micro- and Nanotechnology for Space systems: An Initial Evaluation," edited by H. Helvajian and E.Y. Robinson, The Aerospace Corporation (31 March 1993); "Microengineering Technology for Space Systems," edited by H. Helvajian, The Aerospace Corporation (30 September 1995); "Microelectromechanical Systems (MEMS) Technology Integration into Microspacecraft," Lilac Muller, Ninth Annual Conference on Small Satellites, Logan UT (8-12 September 1995).

2. "Fundamentals of Space Systems," edited by V. L. Pisacane and R. C. Moore, Oxford University Press (1994).

3. "Experimental Spectroscopy," R. A. Sawyer, Dover Publications, Inc. (1963).

4. "The Design of Optical Spectrometers," J.F. James and R.S. Sternberg, Chapman and Hall LTD (1969)

5. Ocean Optics, Inc., 1104 Pinehurst Road, Dunedin FL 34698.

6. Sold in the U.S. by Hellma Cells, Inc., 118-21 Queens Boulevard, Forest Hills NY 11375.

7. Sold in the U.S. by American Laubscher Corporation, 80 Finn Court, Farmingdale NY 11735.

8. "A Microspectrometer Fabricated by the LIGA Process," C. Müller, J. Mohr, et al., KFK-Nachr., 23(2-3), 93-9 (Ger) 1991.

9. "Grating Spectrograph in InGaAsP/InP for Dense Wavelength Division Multiplexing," C. Cremer, et al., Appl. Phys. Lett. 59, 627 (1991)

10 "Integrated Monochromator Fabricated in Silicon Using Micromachining Techniques," R.F. Wolffenbuttel and T.A. Kwa, International Conference on Solid-State Sensors and Actuators, IEEE CH2817-5/91/00000-0832 (1991).

11. "Miniature Micromachined Fabry-Perot Interferometers in Silicon," S.R. Mallinson and J.H. Jerman, Elec. Let., V. 23, N. 20, Sept. 1987, pp 1041-1043.

12. "Miniature Fabry-Perot Interferometers Micromachined in Silicon for use in Optical Fiber WDM Systems," J.H. Jerman and D.J. Clift, IEEE, CH2817-5/91/00000-0372 (1991).

13. "A Fabry-Perot Microinterferometer for Visible Wavelengths," N.F. Raley, et al., Solid State Sensor and Actuator Workshop, IEEE, 0-7803-0456-X(1992).

14. "Process and Design Considerations for Surface Micromachined Beams for a Tuneable Interferometer Array in Silicon," K. Aratani, et al., Micro Electro Mechanical Systems, IEEE CH3265-6 (1993).

15. "Tunable IR Filters Using Flexible Metallic Microstructures," T.R. Ohnstein, et al., Micro Electro Mechanical Systems, Amsterdam (1995).

16. "MEMS and Optics," S. Walker, et al., NRL Memorandum Report (to be published).

17. "Mass Spectroscopy" 2nd Edition," H.E. Duckworth, et al., Cambridge University Press (1986).

18. "Mass Spectrometry Applications in Science and Engineering," Frederick A. White and G.M. Wood, John Wiley & Sons (1986).

19. Comstock, Inc., 1005 Alvin Weinberg Drive, Oak Ridge TN 37830.

20. "The Tiny-TOF Mass Spectrometer for Chemical and Biological Sensing", W.A. Bryden et al., Johns Hopkins APL Technical Digest, Vol. 16(3), 296-310 (1995).

21. "Miniature and Micromachined Mass Spectrometer Arrays," A. Chutjian, M. Hecht and O. Orient, Jet Propulsion Laboratory, Pasadena, CA (unpublished).

22. "A Microsystem Mass Spectrometer," A. Feustel et al., in "Micro Total Analysis Systems" edited by A. v.d. Berg and P. Bergveld, Kluwer Academic Publishers (1995) p. 299-304

23. "A Miniature Mass Spec Chip Concept," C. Fredhoff et al., ARPA Workshop Paper, Pittsburg PA (July 1993) and "Novel Functionality Using Micro-Gaseous Devices" by H.C. Nathanson, et al., IEEE 95CH35754 (1995) p 72-6.

24. "Bonding and Assembling Methods for Realising μTAS," S. Shoji and M. Esashi in "Micro Total Analysis Systems," edited by A.v.d. Berg and P. Bergveld, Kluwer Academic Publishers (1995) p. 165-179

25. "Introduction to the Physics and Techniques of Remote Sensing", Charles Elachi, John Wiley & Sons (1987) p. 2-5

26. "Space Log: 1957 - 1991", T. D. Thompson, Editor, TRW Corporation,

(1991) p. 53

27. "Jane's Space Directory", A. Wilson, Editor, Jane's Information Group Ltd., (1994) p. 380-382

28. *ibid*, p. 406

29. "The Clementine Collection", Naval Research Laboratory, (1994), NRL/PU/1230-94-261, p. 2-4

30. "The Use of Spectral Emmision Lines in the Diagnostics of Hot Solar Plasmas", U. Feldman, Physica Scrypta 24, 681-711 (1981) and "Solar Spectroscopy in the Far Ultraviolet - Xray Wavelength Regions: Status and Prospects". U. Feldman, et al, Journal of the Optical Society of America, 5, 2237-51 (1988)

31. "Jane's Space Directory", A. Wilson, Editor, Jane's Information Group Ltd., (1994) p. 380-382

32. Internet site: http://www.jpl.nasa.gov/galileo/

33. "Small Satellites and NOAA: A Technology Study", The Johns Hopkins University Applied Physics Laboratory, SDO-10559 (1995) p. 2-4

34. "NASA Workshop On Miniature Spacecraft Technology", Kane Cassani, Oral presentation (1995)

71147
030680
2P.

Abstract for presentation at the International Conference on Integrated Micro/Nanotechnology for Space Applications

Houston, Texas
30 October to 3 November, 1995

High-performance electronics at ultra-low power consumption for space applications : from superconductor to nanoscale semiconductor technology

Briefed by : Robert V. Duncan, Ph.D.
Distinguished Member of the Technical Staff
Leader, Cryogenic and Superconductive Technologies Team 5725-1
Sandia National Laboratories, Albuquerque, NM 87185-0980
and Adj. Assoc. Prof. of Physics and Astronomy, Univ. of New Mexico

We present a detailed review of Sandia's work in ultralow power dissipation electronics for space flight applications, including Sandia's superconductive electronics, new advances in quantum well structures and ultra-high purity III-V materials, and finally our recent advances in MEMS technology. Sandia has developed superconductive Josephson-effect systems for use in electrical standardization of manufacturing throughout the Nuclear Weapons Complex, and for use in seismic verification activities. [1] This technology is now being miniaturized for field and space applications through DOE and NASA sponsorship. Other superconductive measurement systems are under development by Sandia, the Jet Propulsion Laboratory (JPL), and the University of New Mexico (UNM) for use in a fundamental science mission tentatively scheduled for deployment in the cargo bay of the Space Shuttle in the year 2000 [2]. These systems have demonstrated revolutionary performance with virtually no intrinsic electrical power dissipation, however their need for cooling to 2 K drives large associated engineering costs. In an attempt to overcome this limitation, Sandia has teamed with UNM and the Air Force Phillips Laboratory to explore the use of ultra-pure nanoscale heterostructure and quantum-well materials to obtain near-superconductive performance at higher temperature and at sufficiently high frequencies. [3] The fundamental limit of the accuracy which may be obtained using certain sub-classes of these Josephson devices has been explored theoretically. [4] Sandia has recently developed quantum-well materials and heterostructures with mobility exceeding 2×10^6 cm^2/Vs [5], and associated new phase-coherent electronic devices [6]. The development of advanced three-level lithographic techniques has permitted the construction of new MEMS structures, including a micro-steam engine, micro-motors which have operated at over 200,000 RPM, and electrostrictively actuated micro-tweasers. These micromechanical devices are well suited for applications in micro-robotics, micro-rocket engines, and advanced sensors [7]. Work reported here has been conducted by the author, Jerry Simmons, Ph.D., Stuart Kupferman, Ph.D., and Paul McWhorter, Ph.D. of Sandia, Prof. David Dunlap, Ph.D. of UNM, and V. Kovanis, Ph.D. of the Phillips Laboratory, and sponsored by the US Department of Energy under contract number DE-AC04-94AL85000, and by the Microgravity Science and Applications Division of NASA and the NASA Metrology Working Group.

References

1. R. Duncan, "Thermal Effects on the Josephson Series-Array Voltage Standard", *Physica B* **165&166**, 101 (1990); and "A Refrigerated Dewar for the Josephson Array Voltage Calibration System" *IEEE Transactions on Instrumentation and Measurement* **40**, 326 (1991).

2. W. A. Moeur, M. J. Adriaans, S.T.P. Boyd, C. Weidensteiner, D. M. Strayer, and R. Duncan, "Cryogenic Design of the Liquid Helium Experiment : Critical Dynamics in Microgravity", to appear in *Cryogenics;* and S. Boyd, W. Moeur, S. Robinson, R, Akau, S. Gianoulakis, and R. Duncan, "Critical Dynamics in Microgravity", to appear in the *International Journal of Thermophysical Properties* (1996).

3. D. H. Dunlap, V. Kovanis, J. Simmons, and R. Duncan, "A frequency to voltage converter based on Bloch oscillations in a capacitively-coupled GaAs-GaAlAs quantum well", *Phys. Rev. B* **48**, 7975 (1993).

4. D. H. Dunlap and R. Duncan, "Measurement accuracy of macroscopic quantum circuits with RF-biased Josephson junction arrays", *Superconductor Science and Technology* **4**, 413 (1991); R. Duncan, "Superconducting Instrumentation for Precision Measurement and Control", in Superconducting Devices and Their Applications, page 446, H. Koch and H. Lubbig, editors (Springer-Verlag, Berlin, 1992); and D. H. Dunlap and R. Duncan, "Fundamental Measurement Accuracy of RF-Biased Josephson Device Comparisons", *J. Appl. Phys.* **71**, 6177 (1992).

5. H. C. Chui, B. E. Hammons, N. E. Harff, J. A. Simmons, and M. E. Sherwin, "Two Millon cm^2/V-s electron mobility by metalorganic chemical vapor deposition with tetriarybutlarsine", to appear *in Applied Physics Letters.*

6. J. A. Simmons, S. K. Lyo, N. E. Harff, and J. F. Klem, "Conductance modulation in double quantum wells due to magnetic field-induced anticrossings", *Phys. Rev. Lett.* **73**, 2256 (1994); M. E. Sherwin, J. A. Simmons, T. E. Eiles, N. E. Harff, and J. F. Klem, "Parallel quantum point contacts fabricated with independently biased gates and a submicrometer airbridge post", *Appl. Phys. Lett.* **65**, 2326 (1994).

Microrefrigeration by a pair of normal metal/ insulator/superconductor junctions

M.M. Leivo and J.P. Pekola

Department of Physics, University of Jyväskylä, P.O. Box 35, 40351 Jyväskylä, Finland

D.V. Averin

Department of Physics, SUNY Stony Brook, NY 11794

Abstract

We suggest and demonstrate in experiment that two normal metal /insulator/ superconductor (NIS) tunnel junctions combined in series to form a symmetric SINIS structure can operate as an efficient Peltier refrigerator. Specifically, it is shown that the SINIS structure with normal-state junction resistances 1.0 and 1.1 kΩ is capable of reaching a temperature of about 100 mK starting from 300 mK. We estimate the corresponding cooling power to be 1.5 pW per total junction area of 0.8 μm^2 at $T = 300$ mK. This cooling power density implies that scaling of junction area up to about 1 mm^2 should bring the cooling power in the μW range.

Introduction

Recently it was shown [1] that in the sub-Kelvin temperature range the Peltier effect in normal metal /insulator/ superconductor (NIS) junctions can be used to cool electrons in the normal electrode of the junction below lattice temperature. The mechanism of the cooling in the NIS contacts is the same as that of the well-known Peltier effect in metal/semiconductor contacts – see, e.g., [2]. Due to the energy gap in the superconductor, electrons with higher energies (above the gap) are removed from the normal metal more effectively than those with lower energies. This makes the electron energy distribution sharper, thus decreasing the effective temperature of electron gas in the normal metal. The decrease in electron temperature demonstrated in this first experiment was, however, limited to about 10% of the starting temperature. There are two possible reasons for this. The first and most obvious, is that the resistance of the refrigerator junction was relatively large leading to low cooling power. The second possible reason is heat leakage through the SN contact used to bias the refrigerator. Nominally, an ideal SN contact with large electron transparency should be able to provide electric conductance without thermal conductance at low temperatures. However both finite temperature and finite subgap density of states in a superconductor lead to non-zero thermal conductance of the biasing contact which

degrades the refrigerator performance. This poses the problem of how to bias the refrigerator without compromising its thermal insulation.

The aim of the present work was to address these two problems and to show that once they are solved the refrigerator performance is improved dramatically. In particular, we show that refrigerator with two NIS junctions in series is capable of reaching temperatures of about 100 mK starting from 300 mK. This brings the NIS refrigerators quite close to practical applications, for instance, in cooling space-based infrared detectors [3].

The problem of large specific refrigerator junction resistance can be alleviated to some degree in a straightforward way by making the insulator barrier thinner. Although it is a challenging technological problem to push this process to its limit, we could conveniently reduce the specific resistance of the junctions to about 0.3 kΩ $\times$ μm^2.

It is less obvious how to solve the second problem of heat leakage through the biasing junction. The solution we suggest here is to combine two NIS junctions in series to form a symmetric SINIS structure. Since the heat current in the NIS junction is a symmetric function of the bias voltage V, the heat flows out of the normal electrode regardless of the direction of the electric current if the junction is biased near the tunneling threshold, $V \simeq \pm\Delta/e$. This means that in a symmetric SINIS structure we can realize the conditions when the electric current flows into the normal electrode through one junction and out through the other one, while the heat flows out of the normal electrode through both junctions. In this way the heat leakage into the normal electrode of the structure is minimized. The experiment with the SINIS structures described below supports this idea.

Basic concepts and refrigerator structure

We begin by briefly outlining the basic theoretical concepts concerning the heat flow in the NIS junctions. Under typical conditions when the transparency of the insulator barrier is small, the heat current P out of the normal electrode (cooling power) of an individual NIS junction is:

$$P(V) = \frac{1}{e^2 R_T} \int_{-\infty}^{+\infty} d\epsilon N(\epsilon)(\epsilon - eV)[f_1(\epsilon - eV) - f_2(\epsilon)]\,, \tag{1}$$

where R_T is the normal-state tunneling resistance of the barrier, f_j is an equilibrium distribution of electrons in the jth electrode, and $N(\epsilon) = \Theta(\epsilon^2 - \Delta^2)\,|\epsilon|\,/\sqrt{\epsilon^2 - \Delta^2}$ is the density of states in the superconductor. From eq. (1) we can deduce several properties of the cooling power P. First of all, by changing the integration variable $\epsilon \to -\epsilon$ we prove that for equal temperatures of the two electrodes P is indeed a symmetric function of the bias voltage, $P(-V) = P(V)$. Plotting

eq. (1) numerically one can see that P is maximum at the optimal bias points $V \simeq \pm\Delta/e$. The optimal value of P depends on temperature and is maximum at $k_BT \simeq 0.3\Delta$, when it reaches $0.06\Delta^2/e^2R_T$, and decreases at lower temperatures as $(k_BT/\Delta)^{3/2}$ [4]. Specifically, at $V = \Delta/e$ (i.e., quite close to the optimal bias voltage) one can get from eq. (1):

$$P(\Delta/e) = \frac{\sqrt{\pi}(\sqrt{2}-1)}{4}\zeta(3/2)\frac{\Delta^2}{e^2R_T}(\frac{k_BT}{\Delta})^{3/2} \simeq 0.48\,\frac{\Delta^2}{e^2R_T}(\frac{k_BT}{\Delta})^{3/2}. \tag{2}$$

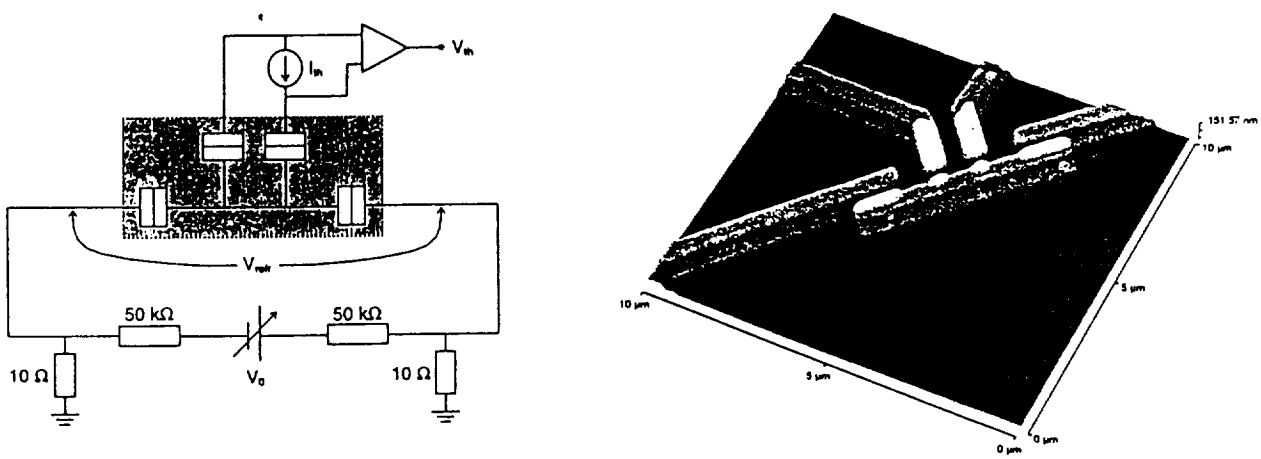


Figure 1. (a) The schematics of the SINIS refrigerator used in the measurements, and (b) an AFM image of the actual structure.

Figures 1a and 1b show, respectively, a schematic diagram of the SINIS structures studied in our experiments and the corresponding AFM image of the structure. Four tunnel junctions were fabricated around a normal metal (Cu) central electrode and four superconducting (Al) external electrodes. The electrodes were made with electron beam lithography using the shadow mask evaporation technique. The tunnel junctions were formed by oxidation in pure oxygen between the two metallization steps. Two junctions at the edges with larger areas were used for refrigeration, while the pair of smaller junctions in the middle was used as a thermometer. A floating measurement of voltage across the two thermometer junctions at a constant bias current was used to measure temperature in the same way as in the simpler one junction case [5].

Single-junction refrigerator

Prior to our experiments with the SINIS refrigerator we repeated measurements in the geometry of Nahum et al. [1]. In our case the refrigerating junction had a resistance of $R_T = 7.8$ kΩ, the copper island was 10 μm long, 0.3 μm wide and 35 nm thick. Results of the measurements of the electron temperature in the island as a function of refrigerator voltage V_{refr} for several starting

temperatures at $V_{refr} = 0$ are shown in Fig. 2. We see that only a few per cent refrigeration can be obtained, as expected.

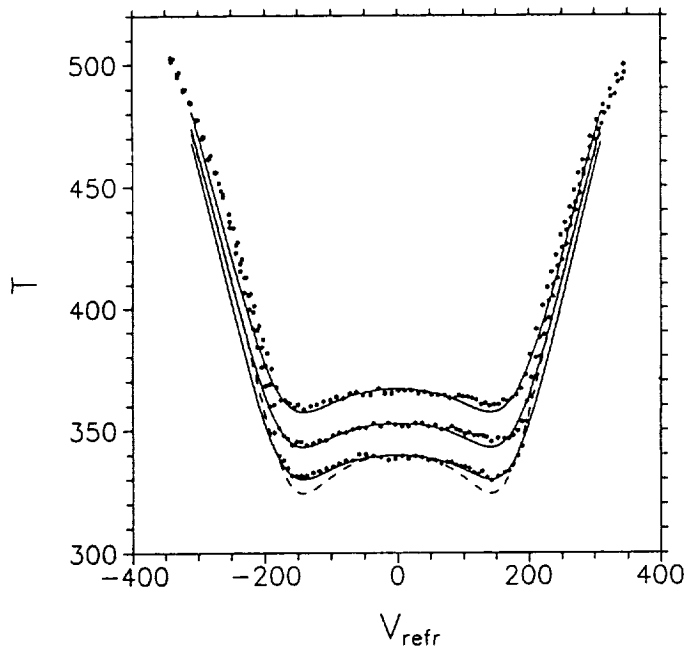


Figure 2. Results of the NIS single junction experiment: temperature T [mK] of the N-electrode versus refrigerator voltage V_{refr} [μV]. Dots and solid lines show, respectively, experimental data and the theoretical fit including heat conductance of the biasing contact. For comparison, the dashed line shows the fit without the heat conductance.

Solid lines in Fig. 2 represent a theoretical fit obtained within the standard model of electron energy relaxation [7-9]. Within this model, we assume that the electron-electron collision rate is large so that the electrons maintain an equilibrium distribution characterized by the temperature T, which is in general different from the lattice temperature T_l. The rate of energy transfer from electrons to phonons is then [8]: $P_l = \Sigma U(T^5 - T_l^5)$, where Σ is a constant which depends on the strength of electron-phonon coupling, and U is the volume of the island. Another element of the fitting process is a heat conductance κ of the biasing SN contact. (For simplicity, we neglect temperature dependence of κ, since the temperature range of interest in Fig. 2 is not very large.) The value of the superconductor gap Δ is almost fixed by the position of the temperature dips in the refrigeration curves (Fig. 2) and is taken to be 155 μeV for this sample. Solving numerically the equation $P = P_l + \kappa(T - T_l)$, where P is the cooling power (1) we can calculate T as a function of V_{refr}. The fit in Fig. 2 is obtained in this way with $\Sigma = 0.9$ nW/K^5 μm^3 and $\kappa = 8$ pW/K. For comparison, the dashed line shows the fit obtained for the lowest-temperature curve without κ; in this case $\Sigma = 1.4$ nW/K^5 μm^3. Although there is no drastic disagreement with the data even in

this case, we see that κ improves the fit considerably.

To see whether the value of the heat conductance κ deduced from the fit in Fig. 2 is reasonable, we calculated the heat conductance of the SN contact with perfect electron transparency at low temperatures, using the method developed in [9,4]. In this approach, the heat current J in the contact can be written as follows:

$$J = \frac{1}{e^2 R_N} \int d\epsilon (\epsilon - eV)[f_1(\epsilon - eV) - A(\epsilon) f_1(\epsilon + eV) - (1 - A(\epsilon)) f_2(\epsilon)] , \tag{3}$$

where R_N is the normal-state contact resistance, functions $f_j(\epsilon)$ are introduced in eq. (1), and $A(\epsilon)$ is the probability of Andreev reflection from the ideal NS interface:

$$A(\epsilon) = \begin{cases} (|\epsilon| - (\epsilon^2 - \Delta^2)^{1/2})^2 / \Delta^2 , & |\epsilon| > \Delta , \\ 1 , & |\epsilon| < \Delta . \end{cases}$$

At low temperatures $k_B T \ll \Delta$ and vanishing bias voltage V eq. (3) gives for the linear heat conductance κ:

$$\kappa = \frac{2\Delta}{e^2 R_N} \left(\frac{2\pi\Delta}{k_B T}\right)^{1/2} \exp\left(-\frac{\Delta}{k_B T}\right) . \tag{4}$$

Making use of the gap value and temperature corresponding to Fig. 2 we get from this equation that $\kappa = 8$ pW/K corresponds to R_N on the order of 100 Ohm, which is the same order of magnitude as in the experiment.

Double-junction refrigerator

Figure 3 shows our main results with the SINIS refrigerator of Fig. 1. The two refrigerating junctions had resistances $R_T = 1.0$ and 1.1 kΩ, respectively, and the island was 5 μm long, 0.3 μm wide, and 35 nm thick. In Fig. 3 we see several refrigeration curves starting at various ambient temperatures at $V_{refr} = 0$. Note that maximum cooling power is now obtained at $V_{refr} \simeq 2\Delta/e$ because of two junctions involved. From the position of the temperature dips we get $\Delta = 180$ μeV. We see that the drop in temperature is immensely improved over that of the single junction configuration. Solid lines in Fig. 3 show the theoretical fit which was obtained within the same model as for the single-junction configuration with the two modifications. We do not have heat conductance κ this time, and we need to solve the balance equations simultaneously for the energy and electric current in order to determine the electric potential of the island. The best fit shown in Fig. 3 corresponds to $\Sigma = 4$ nW/K^5 μm^3. The fit can be classified as reasonable, although there are some obvious discrepancies between the data and the theory. Possible origins of these discrepancies include an oversimplified model of electron-phonon heat transfer on the theory side,

and poor calibration of the thermometer toward higher temperatures on the experimental side. The inset in Fig. 3 shows the maximum cooling power as a function of temperature deduced from this fit, together with the analytical dependence obtained by summing eq. (2) over the two junctions. We see that the simple analytical expression (2) gives a very accurate description of the lower-temperature cooling power.

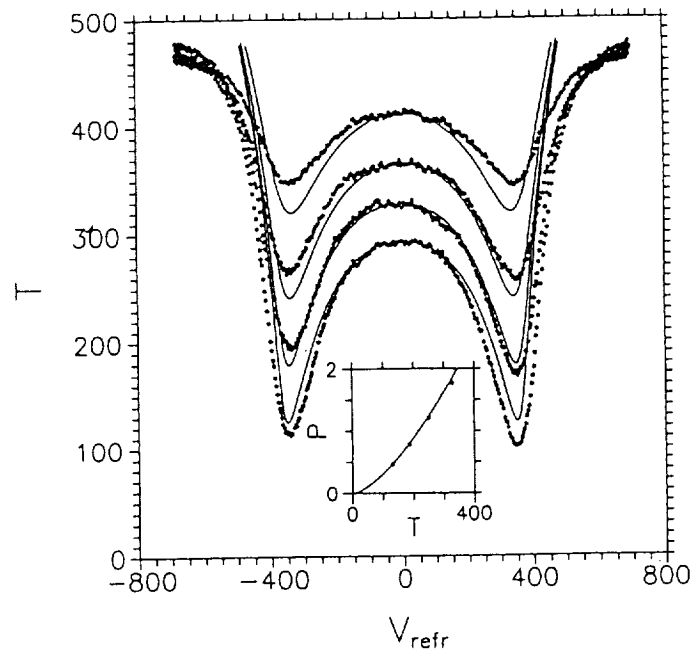


Figure 3. SINIS refrigerator performance at various starting temperatures. Notations are the same as in Fig. 2. Dots are the experimental data, while the solid lines show the theoretical fit with one fitting parameter Σ for all curves. The inset shows the cooling power P [pW] from the fits (dots), together with the analytical result from eq. (2).

We must stress that Fig. 3 shows electron, not lattice temperature. By cooling only electron gas we avoid the problem of parasitic phonon heat conductance through refrigerating junctions. (Electron heat conduction due to tunneling plays a minor role and was accounted for in our model.) If we would try to cool down the normal island as a whole, the phonon heat conduction through the insulator barrier would become important limiting factor also in our refrigerator. In order to estimate the magnitude P_{ph} of the parasitic phonon heat flow through the tunnel junctions we note that neither the thin insulator layer of the junction nor the difference between parameters of the junction electrodes (aluminum and copper in our experiment) give rise to substantial reflection coefficient of the large-wavelength phonons relevant at low temperatures. Therefore we can

estimate P_{ph} as the flow of energy associated with ballistic propagation of acoustic phonons:

$$P_{ph} = \frac{\pi^2 (k_B T)^4}{120\hbar^3}\left(\frac{1}{c_{\parallel}^2} + \frac{2}{c_{\perp}^2}\right),$$

where $c_{\parallel}$ and $c_{\perp}$ are longitudinal and traversal sound velocities. For aluminum at T=1 K this equation gives the power density $P_{ph} \simeq 0.6$ nW/μm^2. Comparing this with the cooling power of our refrigerator we see that the refrigerator can cool the normal island as a whole only for starting temperatures not greater than 0.3 K. Decrease in tunnel resistance of the refrigerating junctions could shift this temperature limit to few Kelvins.

Conclusion

In conclusion, we have shown that the nominally-symmetric SINIS structure can be used as an efficient Peltier refrigerator. One of the advantages of the symmetric structure is that it is easier to fabricate than the asymmetric single-junction configuration. Besides this, SINIS structure provides more efficient thermal insulation of the central electrode, which allowed us to demonstrate a temperature drop of about 200 mK starting from 300 mK. The achieved cooling power density was approximately 2 pW/μm^2, with the total power being 1.5 pW at $T = 300$ mK.

The next step in the development of a practical NIS refrigerator could be further optimization of the refrigerator performance with respect to the resistance R_T of the insulator barrier. As we saw above, the cooling power of the refrigerator increases with decreasing R_T. For a fixed junction area this trend would continue only up to an optimal resistance, at which point the transport starts to be dominated by Andreev reflection [4]. The theoretical limit for the maximum cooling power density is on the order of 10^{-8} W/μm^2 for aluminum junctions and should be reached in the junctions with unrealistically low specific resistances on the order of 10^{-2} $\Omega \times \mu$m^2. In practice the limiting factors will be the technological ability to fabricate uniform tunnel barriers with high transparency.

The authors gratefully acknowledge useful discussions with A. Korotkov, K. Likharev, J. Luckens and M. Paalanen. This work was supported in part by the Academy of Finland and ONR grant # N00014-95-1-0762.

References

1. M. Nahum, T.M. Eiles, and J.M. Martinis, Appl. Phys. Lett. **65**, 3123 (1994).

2. R.R. Heikes and R.W. Ure, Jr., *"Thermoelectricity: science and engineering"* (Interscience publishers, NY, 1961).

3. For space cryogenic applications, see, e.g., C. Hagmann, D.J. Benford, and P.L. Richards, Cryogenics **34**, 213 (1994); and a special issue of Cryogenics **34**, No. 5 (1994).

4. A. Bardas and D. Averin, Phys. Rev. B **52** (1995), to be published.

5. M. Nahum and J.M. Martinis, Appl. Phys. Lett. **63**, 3075 (1993).

6. M.L. Roukes, M.R. Freeman, R.S. Germain, R.C. Richardson, and M.B. Ketchen, Phys. Rev. Lett. **55**, 422 (1985).

7. F.C. Wellstood, C. Urbina, and J. Clarke, Appl. Phys. Lett. **54**, 2599 (1989).

8. R.L. Kautz, G. Zimmerli, and J.M. Martinis, J. Appl. Phys. **73**, 2386 (1993).

9. G.E. Blonder, M. Tinkham, and T.M. Klapwijk, Phys. Rev. B **25**, 4515 (1982).

Session 5: Materials Processing, Packaging and Architectures (Tuesday, PM)

Session Chairs: James C. Lyke (Air Force Phillips Laboratory) and Henry Helvajian (The Aerospace Corporation)

Laser Material Processing for Microengineering Applications
by Henry Helvajian, The Aerospace Corporation

Recent Developments in Microsystems Fabricated by the LIGA-Technique
by J. Schulz, K. Bade, A. El-Kholi, H. Hein and J. Mohr,
Institute für Mikrostrukturtechnik, Karlsruhe, Germany

Micromechanical Machining Processes and Their Application to Aerospace Structures, Devices, and Systems
by Craig Friedrich and Robert O. Warrington, Louisiana Technical University

Advancing MEMS Technology Usage through the MUMPs Program
by David A. Koester, K. W. Markus, V. Dhuler, R. Mahadevan, and A. Cowen,
MCNC MEMS Technology Applications Center

Method of Making Large Area Microstructures
by Alvin M. Marks, Advanced Research Development, Inc.

Surface Micro-Machining: Progress Towards Micro-Electro Mechanical Systems
by James Doscher, Analog Devices Inc.

Advanced Packaging for Integrated Micro-Instruments
by James C. Lyke, U.S. Air Force Phillips Laboratory

The Impact of Space Radiation Requirements and Effects on ASIMs
by C. Barnes, A. Johnston and G. Swift, Jet Propulsion Laboratory (NASA)

3D Thin Film Microstructures for Space Microrobots
by Isao Shimoyama, University of Tokyo

Laser Material Processing for Microengineering Applications

H. Helvajian
Mechanics and Materials Technology Center
The Aerospace Corporation
Los Angeles, CA 90009
USA

71151
030685
10P

Abstract

The processing of materials via laser irradiation is presented in a brief survey. Various techniques currently used in laser processing are outlined and the significance to the development of space qualified microinstruments are identified. In general the laser processing technique permits the transferring of patterns (i.e. lithography), machining (i.e. with nanometer precision), material deposition (e.g. metals, dielectrics), the removal of contaminants/debris/passivation layers and the ability for providing process control through spectroscopy.

Introduction

The microelectronics processing technique has been successfully used to cofabricate microstructures (i.e. transducers) alongside electronics (i.e. "intelligence"). The result is a "smart" microinstrument comprising of materials compatible with microelectronics processing. The use of materials outside the fabrication recipe list is typically not recommended as it may jeopardize the product yield or contaminate a process line. As a consequence, a limited set of materials are available for most microinstrument development. This limitation is dictated by the processing approach and would be less restrictive for a processing tool which could alter materials with "site-specific" precision and without the need for raising the sample bulk temperature. Laser processing is one such technique which complements the microelectronics processing technique and permits the site-selective processing of materials at low bulk temperatures. For example, the laser based technique has been used as a "direct-write" processing tool for circuit repair operations. Still, more is possible if the knowledge in laser material interaction physics/chemistry is used to advantage. A laser "tool" can provide capabilities which are beyond the post-assembly simple-repair operations. First, as a processing tool the laser is non intrusive and can easily be integrated into a microelectronics fabrication assembly line. Second, the laser tool is amenable to automation. Third, a laser tool can be configured for multitasking in situ operations (i.e. it can put down and/or remove material, serve as process diagnostic). Finally, a laser based tool offers new capabilities, such as processing at an imbedded interface. A technique currently unavailable in the microelectronics processing tool chest. There is sufficient experimental evidence that a laser based material processing tool will serve to accelerate the development of microinstruments by allowing materials with unique chemical and physical properties (e.g. ceramics, diamond, polymers) to be processed. This capability will be specially important for space applications where materials are designed to meet stringent standards.

Space qualified hardware must meet a diverse set of environmental conditions. Prior to launch, space qualified hardware is typically placed in storage, sometimes for years. The storage environment may include both high temperature and humidity. During the launch the hardware must survive the high acoustic and acceleration forces and subsequent to the launch, they must operate in vacuum, without the benefit of cooling air. Furthermore, the hardware is exposed to cosmic radiation and must survive thousands of heating-cooling cycles. Therefore,

materials are specifically chosen or "engineered" to withstand both the launch and the space environments. Likewise, component packaging for space applications is also non trivial. The package or an enclosure may serve two purposes. First, to mitigate the effects of the space environment and second, to provide an exit pathway for outgasing of materials contained. Consequently, some components may have to be locally sealed while others may require pathways for outgassing.

Microinstruments designed for space applications must also meet the diverse set of environmental conditions a described. As a consequence space qualified microinstruments (and application specific integrated microinstruments- ASIMs) may incorporate materials not commonly used in terrestrial applications. Likewise, the integrated packaging approaches used must employ localized hermetic seals to protect contaminant sensitive components. A hypothetical example of such a package is given in Figure 1 where various devices and components, comprising of a variety of materials, are assembled on a common substrate. To fabricate, test and repair such a module requires a material processing tool with "direct-write" capability. The laser is a "direct-write" processing tool which does not require vacuum, can deposit, remove or alter materials with both site and material specificity and can deposit energy at an imbedded interface. The result is the ability for localized material processing without raising the temperature of adjoining structures or components. Figure 1 shows a few of the viable laser processing tasks for the hypothetical package example and Table 1 lists a few of the current laser applications in material processing.

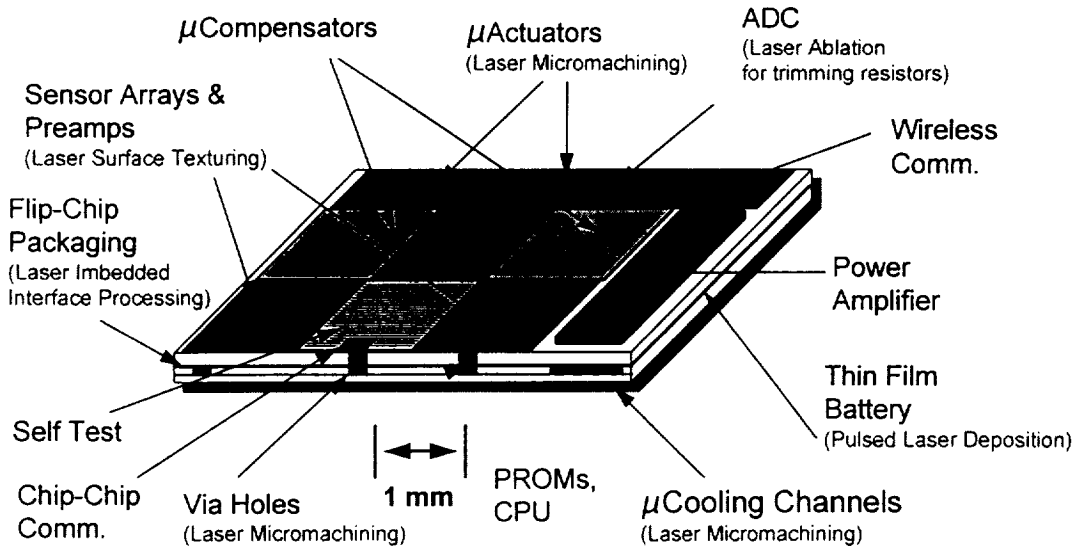


Figure 1: Hypothetical ASIM in a hybrid multichip module (MCM) package employing flip-chip technology.

Table 1: Examples of Laser Applications in Material Processing

1	Pulsed Laser Deposition (via Ablation) - scale-up to 200mm wafers shown. Bibliography of deposited thin films[1] Bandgap Engineering[2] Biocompatible Materials[3]	8	Micrometer scale "Machining" Etching: Metals, Insulators, Semiconductors Soft Ablation: Polymers Scribing and Contouring
2	Nanometer Scale "Machining" (via nonthermal ablation) --- Crystals	9	Spectroscopy Surface, Above surface, Process Control
3	Photochemical Deposition	10	Photolithography
4	Surface Annealing	11	Localized Oxidation
5	Annealing	12	Surface texturing/contouring
6	Semiconductor Doping and Drive-in.	13	Surface Debris Removal (i.e. cleaning)
7	Laser-LIGA	14	Embedded Interface Diffusion

Fundamentals and Selected Application Examples

To refine the laser material processing technique requires understanding the laser material interaction phenomenon. Many textbooks and articles cover this material at depth[4]. This report only summarizes the salient features and without the benefit of supporting equations. Under most laser processing conditions the criteria for optimum laser material interaction is described by a handful of equations[5]. These equations describe the propagation of the laser beam energy through the beam delivery optics, the photophysical interaction with the surface (i.e. absorption, surface chemical interactions) and the subsequent surface modification. The equations are simplified for a Gaussian laser beam propagating in a diffraction-limited optical system, though for most material processing a top-hat or flat top homogenized beam is used[6]. In general, the fundamental physics is driven by the incident laser fluence (J/cm^2), the responsivity of the material and the photochemistry of the ablated material. In vacuum and for very low laser fluences the particle or "processing" yield is governed via non thermal photoactivated mechanisms. In this processing regime the yield is non linear with the laser fluence and the surface is "processed" with nanometer precision. At somewhat higher laser fluences the irradiated surface is abruptly heated and material processing is via thermal initiated mechanisms. In this regime, processing yield is typically linear and the processing precision is micrometers. At much higher laser fluences, where an above surface plasma is ignited, the interaction of the laser with the surface is reduced as a result of absorption in the plasma. In this regime, the surface morphology is altered not by the laser but by the above surface plasma. The processing precision is not easily controllable but is nominally on the millimeter scale. Table 2 shows the process criteria and related parameters for laser "direct-write" processing and Table 3 gives a summary of common laser parameters that can be controlled and the induced effect in processing. Application examples of processing materials in these different regimes are also given below.

Table 2: Process considerations in direct write processing (from Ref [7])

Process Considerations	Related Mechanisms
Resolution	Wavelength Beam spot size Thermal diffusivity Nonreciprocity
Writing Speed	Beam size Film growth rates
Resistivity	Composition Impurity Physical structures
Morphology	Coherence effects Nonuniform heating Instabilities
Adhesion	Interfaces Thermal mismatch

Table 3: Laser parameters and related processing parameters (from Ref [8])

Laser Parameters	Effect on Material Processing
Power (average) (W)	Temperature (steady state) Process throughput
Wavelength (μm)	Optical absorption, reflection, and Transmission, Resolution Photochemical effects
Spectral linewidth (nm)	Temporal coherence Chromatic aberration
Beam size (mm)	Focal spot size, Depth of focus Beam transformation characteristics Intensity
Lasing modes	Intensity distribution Spatial uniformity Speckle, Spatial coherence Modulation transfer function
Peak power (W)	Peak temperature Damage/induced stress Nonlinear effects
Pulsewidth (sec)	Interaction time Transient processes
Stability (%)	Process latitude
Efficiency (%)	Cost
Reliability	Cost

For some materials, with UV laser irradiation and at very low laser fluences (<100mJ/cm^2) the laser material interaction is driven by electronic rather than thermal excitation. Under these circumstances nanometer scale precision surface processing becomes possible. By controlling the laser fluence during the irradiation, a surface can be "peeled" atomic layer-by-layer. Furthermore, by tuning the laser wavelength selective removal of species from the surface is possible. Figures 2 and 3 present both of these phenomenon for the perovskite ceramic $Bi_2Sr_2CaCu_2O_8$. Figure 2 shows the calcium atom photodesorption yield as a function of laser shot number. The oscillation in the calcium signal as measured by the mass spectrometer corroborates with the layer-by-layer removal concept. Figure 3 shows two mass spectra of the photoejected species for two laser irradiation wavelengths. For 351 nm irradiation the mass spectrum is representative of the all the species while for 248 nm irradiation a single mass peak is measured. Most laser material processing is conducted at higher laser fluences than that used in acquiring the data shown in Figures 2 and 3. Laser material processing via non thermal excitation becomes economical for laser repetition rates greater than 1 MHz. At the higher fluences, microscribing or surface texturing become economically feasible for most laser systems commercially available. Laser microscribing is a value added process because laser scribing is generally faster than most techniques and leaves the surface with less overall damage. Even in comparison diamond scribed surfaces. An example is in the development of large area (1m x 1m) photovoltaics where laser scribed surfaces result in cleaner cuts. Microsribed surfaces also have properties which may be important in space applications. Surfaces with complex topologies are known to enhance chemical reactivity (e.g. catalysis), optical reflectivity (e.g. light trapping), mechanical action (e.g. microVelcro, precision abrasion) and load-handling capability in lubricated journal bearing systems. Surfaces can be laser microscribed either via a "direct-write" method, through the formation of an interference pattern in a chemical etching environment or via a rapid heating and recrystallization process (temperature rise rate >10^7 K). Figure 4 shows a scanning electron micrograph of a microtextured crystalline aluminum surface prepared in vacuum at The Aerospace Corporation using the latter method. The inset shows the scale where the measured corrugation period is approximately 0.15 μm. Most surface texturing or scribing by laser is typically done by the two other methods using an etching solution. Table 4 lists typical etching/ablation rates for a laser based material processing system. For the etching of Si and SiO_2 the table also includes, for comparison, the predicted etching rates used in microelectronics processing ---- for etching without a laser and using a relatively vigorous KOH/water recipe.

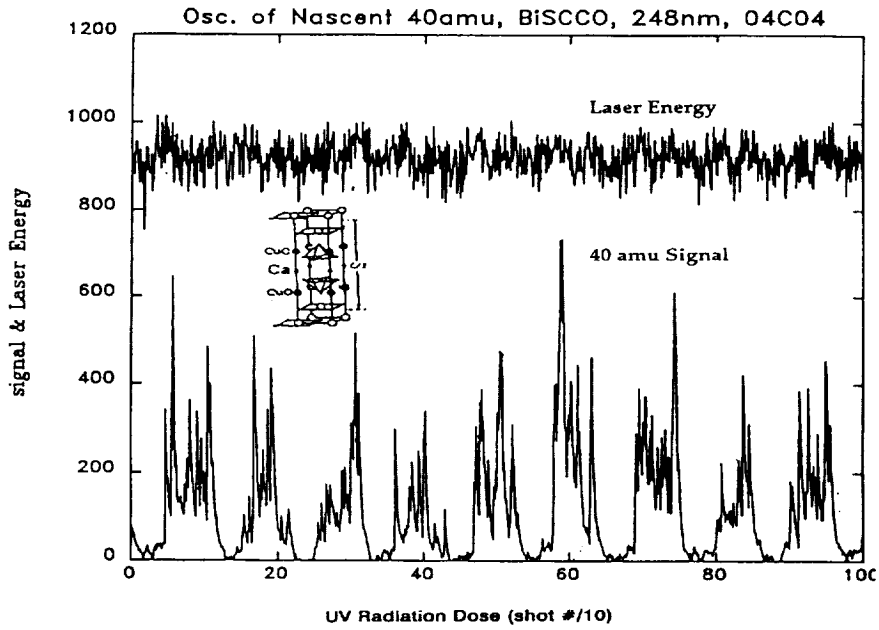


Figure 2: Measurement of the photoejected atomic calcium yield as a function of laser pulse number. The irradiation is via a UV laser nonthermal photosputtering from an aligned single crystal, high T_c perovskite ceramic ($Bi_2Sr_2Cu_2CaO_8$)[9].

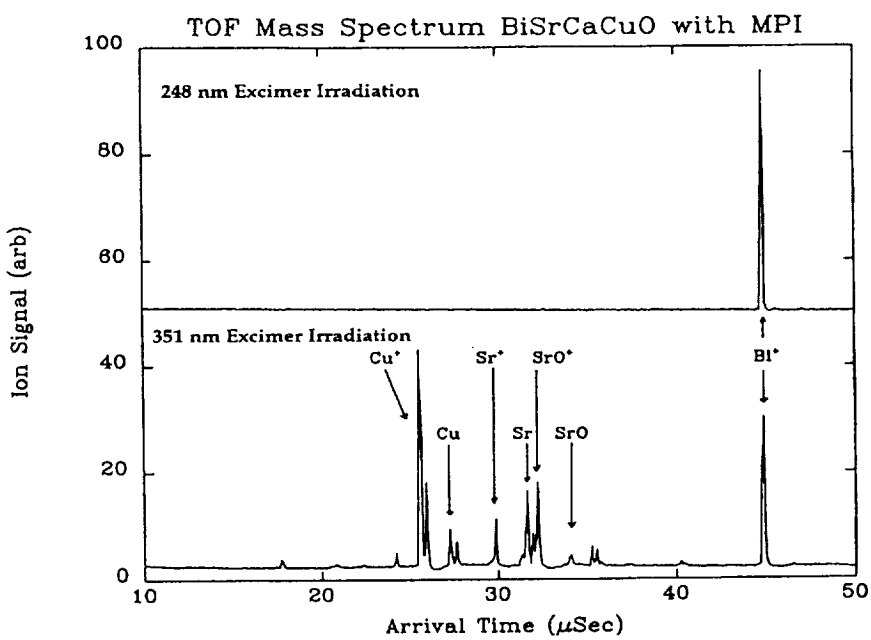


Figure 3: Time-of-flight mass spectrum of photoejected species following pulsed UV laser irradiation of $Bi_2Sr_2Cu_2CaO_8$[10].

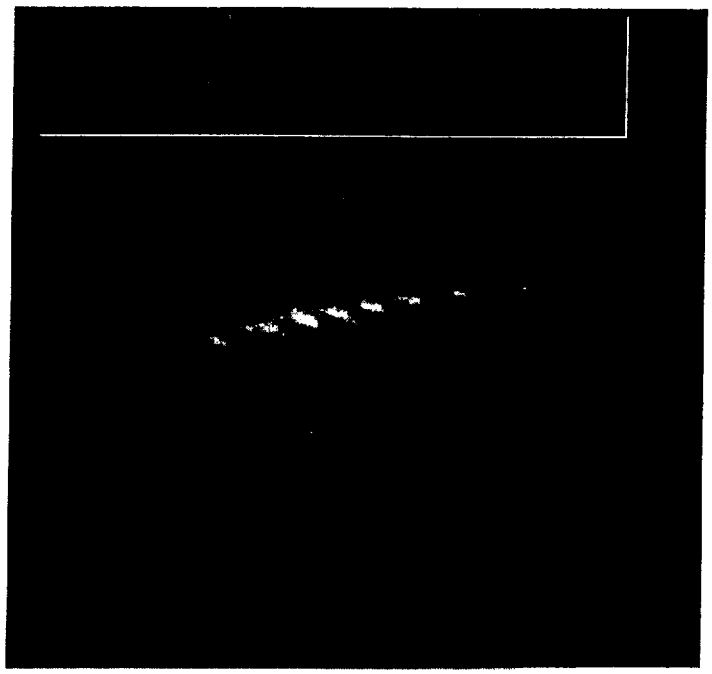


Figure 4: UV laser microtexturing of an aluminum (111) surface prepared in UHV.

Table 4: Etching/ablation rates for typical materials used in semiconductor processing (data from tables in Ref.[11])

Si:: ~ 7-15 μ/sec	SiO_2:: ~ 0.5 -3 nm/sec
(KOH etch rate Si<100>, 35%, 100C is 55 nm/sec)	(KOH etch rate 40%, 100C is 0.5 nm/sec)
-αSi:: ~ 50 nm/pulse	Ge:: ~ 36 μ/sec
GaAs:: ~ 0.5 - 2 μ/sec	InP:: ~ 140 μ/sec
GaP:: ~ 60 nm/sec	Diamond:: ~ 140 nm/pulse
Polimide:: ~ 0.1 - 1.0 μ/pulse	PET:: ~ 0.1-1.0 μ/pulse
PMMA:: ~ 0.3 μ/pulse	Ag: ~ 0.1 μ/pulse
Al:: ~ 0.1 - 1.0 μ/sec	Au:: ~ 50-80 nm/pulse
Cr:: ~ 80 nm/pulse	Cu:: ~ 0.1 μ/pulse
Mo:: ~ 20 μ/pulse	W:: ~ 3 nm/sec

Recent interest has focused on the ability of UV lasers to micromachine diamond. Diamond is a material which offers significant advantages in applications for space systems. It is both a good electrical insulator and a chemically inert material. Moreover, it has low thermal expansion when compared to silicon, aluminum nitride, Kovar and alumina. It could serve as an ideal material for packaging. Lasers can not only machine diamond but can also grow a variant of true natural diamond. Amorphous diamond (~75% true diamond by volume) is one form which is similar to natural diamond in hardness (~78GPa) and has a low coefficient of friction (~0.1).

Using high intensity ablation, amorphous diamond can be deposited without catalysts[12]. Various steels have been coated (304, 316L, 440C)[13] long with silicon, titanium and IR optics (Ge, ZnS)[14]. Growth rates of 0.5 μm/hr over 100 cm^2 have been realized. Figure 5 shows a free standing laser micromachined diamond microgear while Figure 6 shows a laser micromachined valve in a diamond substrate.

Figure 5: a) Freestanding laser micromachined diamond microgear. b) Engraved laser micromachined gear in type 1b diamond substrate (Courtesy of Potomac Corp.[15])

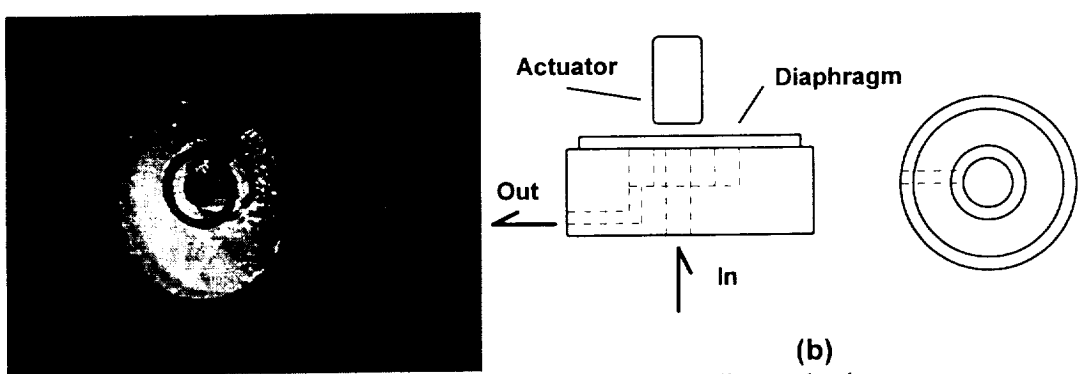


Figure 6: a) Laser micromachined valve in diamond substrate without the actuator b) Schematic of the valve design. (Courtesy of Potomac Corp.)

Experiments now in preparation at The Aerospace Corporation will utilize the laser processing techniques as described to develop microthrusters for space applications. The intent is to develop a miniature (15 cm diameter) space micropropulsion platform. Several concepts are under consideration, they include the traditional monopropellant thrusters and a miniaturized ion propulsion thruster. Key to developing this subsystem is the fabrication of a very low-leak rate microvalve which is resistant to caustic gases (e.g. hydrazine) and the integrating of the valving, propellant tank and the thruster nozzles.

A general survey of current space systems finds that there are applications where the laser material processing technique would be of value. A few examples are given. Under the general area of fabrication, laser processing could be used to machine microchannels in sensor arrays for integrated cooling systems, lasers could also be used to machine vias for multilevel interconnects and for developing media for permanent high density information storage (WORM). Laser processing tools are certainly needed in the development of large area photovoltaics, diamond coating of surfaces and in advanced packaging schemes were localized hermetic seals are required. In the area of repair or modification, laser tools can be used to trim resistor and capacitor values, can be used for fixing interconnects in DRAM chips and the alloying of metals for increased corrosion resistance. Laser processing tools have the greatest impact in the area of prototyping. Whether for prototyping complex multitiered microoptics in various materials or in fabricating microstructures in diamond or ceramics --- the laser tool is unsurpassed in its versatility. The only barrier to implementing laser processing in microengineering is the lack of prototyping service-centers and in the lack of process development.

Acknowledgments
The author acknowledges the support received from the Aerospace Sponsored Research Program which sponsored this review on laser material processing technology. The author also acknowledges the support received from the Space Research & Development Office.

[1]K. L. Saenger, " Bibliography of Films Deposited by Pulsed Laser Deposition", in Pulsed Laser Depostion of Thin Films, D. B. Chrisey, G. K. Hubler Eds. (John Wiley & Sons, New York, 1994) pg. 582.

[2] J. T. Cheung, and H. Sankur, "Growth of Thin Films by Laser-Induced Evaporation", Vol. **15** CRC Critical Rev. Solid State Mat. Sci. (1988), pg.63

[3]C. M. Cotell, "Pulsed Laser Deposition of Biocompatible Thin Films" in Pulsed Laser Deposition of Thin Films, D.B. Chrisey, G. K. Hubler Eds. (John Wiley & Sons, New York) 1994, pg. 549.

[4]J. Y. Tsao and D. J. Ehrlich Eds. "Laser Microfabrication--Thin Film Processes and Lithography", 1989(Academic Press, NY); C. I. H. Ashby, "Laser Driven Etching" in Thin Film Processes II, 1991(Academic Press, NY) pg. 783; J. H. Davies, A. R. Long Eds. "Physics of Nanostructures", 1992 (IOP Publishing Ltd, London); R. F. Haglund, Jr. and R. Kelly "Electronic Processes in Sputtering by Laser Beams" in Fundamental Processes in Sputtering of Atoms and Molecules (SPUT92) P. Sigmund Ed. 1993 (Munksgaard, Copenhagen); L. D. Laude, D. Bäuerle, M. Wautelet, "Interfaces Under Laser Irradiation", NATO ASI Series E- **134**, 1987 (Martinus, Nijhoff Publishers, Boston); D. B. Chrisey, G. K. Hubler Eds. "Pulsed Laser Deposition of Thin Films", 1994 (John Wiley & Sons, New York); R. R. Kunz, M.W. Horn, T. M. Bloomstein, D. J. Ehrlich, " Applications of Lasers in Microelectronics and Micromechanics", Appl. Surf. Sci. , **79** (1994) 12.

[5] H. Helvajian, "Laser Material Processing: A Multifunctional In-Situ Processing Tool for Microinstrument Development", in Micrengineering Technology for Space Systems H. Helvajain Ed. Aerospace Report ATR-95(8168)-2, 1995, pg. 87, and references therein.

[6]A. G. Cullis, H. C. Webber, and P. Bailey, "A Device for Laser Beam Diffusion and Homogenization" J. Phys. E. Sci. Instrum. **12**, (1979) 688.

[7]Y.S. Liu " Sources, Optics and Laser Microfabrication Systems for Direct Write and Projection Lithography" in Laser Microfabrication--Thin Film Processes and Lithography, J. Y. Tsao and D. J. Ehrlich Eds. 1989(Academic Press, NY) pg. 3.

[8]Y.S. Liu, "Sources, Optics and Laser Microfabrication Systems for Direct Write and Projection Lithography" in Laser Microfabrication--Thin Film Processes and Lithography, J. Y. Tsao and D. J. Ehrlich Eds. 1989(Academic Press, NY) pg. 3.

[9] L. Wiedeman, and H. Helvajian, The Aerospace Corporation, unpublished work.

[10]H. Helvajian, L. Wiedeman, and H.-S Kim, " Photophysical Processes in Low-Fluence UV Laser-Material Interaction and the Relevance to Atomic Layer Processing," Adv. Mat. for Optics and Elect. **2**, (1993) pg 31.

[11]H. Helvajian, "Laser Material Processing: A Multifunctional In-Situ Processing Tool for Microinstrument Development", in Micrengineering Technology for Space Systems H. Helvajain Ed. Aerospace Report ATR-95(8168)-2, 1995, pg. 87, and references therein.

[12] C. B. Collins, F. Davanloo, T. J. Lee, J. H. You, and H. Park, " The Production and Use of Amorphic Diamond", Amer. Ceramic Soc. Bull. **71**, (1992) 1535.

[13]C. B. Collins, F. Davanloo, J. H. You and H. Park, "Tribology of Amorphic Diamond on Stainless Steel", Diamond Films and Tech. **4**, (1994) 15.

[14] F. Davanloo, T. J. Lee, H. Park, J. H. You, and C. B. Collins, " Protective Films of Nanophase Diamond Deposited Directly on Zinc Sulfide Infrared Optics", J. Mater. Res. **8**, (1993) 3090.

[15]MRS Bulletin, " Microgear produced from Single-Crystal Diamond", October (1994) pg. 5; C. P. Christensen, "Fine Diamonds with Laser Machining", Photonics Spectra, November, 1993; C. Paul Christensen "UV Lasers: Key Tools for Micromachining", Medical Device & Diagnostic Industry, January 1995; C. P. Christensen "Laser Processing Works on a Micro Scale", Industrial Laser Review, June 1994.

RECENT DEVELOPMENTS IN MICROSYSTEMS FABRICATED BY THE LIGA-TECHNIQUE

J.Schulz, K.Bade, A.El-Kholi, H.Hein and J.Mohr
Forschungszentrum Karlsruhe, Institut für Mikrostrukturtechnik
Postfach 3640
76021 Karlsruhe
Germany

Abstract

As an example of microsystems fabricated by the LIGA-technique three systems are described and characterized: a triaxial acceleration sensor system, a micro-optical switch and a microsystem for the analysis of pollutants. The fabrication technologies are reviewed with respect to the key components of the three systems: an acceleration sensor, an electrostatic actuator and a spectrometer are made by the LIGA-technique, a micropump and microvalve are made using micromachined tools for molding and optical fiber imaging is made possible by combining LIGA and anisotropic etching of silicon in a batch process. These examples show that the combination of technologies and components is the key to complex microsystems. The design of such microsystems will be very much facilitated if standardized interfaces are available.

Introduction

At the Research Center in Karlsruhe (FZK), the LIGA technology (Bec86) as one of the dominant micro patterning technologies has been developed over the last ten years. Three years ago, the promising economic perspective of microsystems has led to an increase of FZK's activities in this field by coordinating the work from areas such as patterning technologies, chemical sensor fabrication, assembly, material science, electronics and computer science.

In this paper, the results of our efforts on three microsystems will be reviewed as an example of the activities of FZK:

- a triaxial acceleration sensor system in a planar setup consists of two LIGA-acceleration sensors for the detection of acceleration within the substrate plane and a commercially available silicon acceleration sensor for the direction normal to the substrate; this setup simplifies assembly tremendously;

- a switch for monomode fiberoptical communications uses an electrostatic LIGA-actuator on a silicon substrate with etched grooves for vertical placement of lenses and with LIGA lateral fixing elements for lenses and fibers;

- a microsystem for the optochemical analysis uses a LIGA-spectrometer with especially designed electronics and a micropump for fluid control to and from a micro cuvette which carries optochemical sensors..

In the next section we will give a brief review of the involved technologies and we will describe the key components of the microsystems. Subsequently, the microsystems and their performance will be described.

Microstructure Technology and Components

The fabrication procedure of microcomponents usually involves all of the three technologies: patterning, material deposition or etching and microassembly. Full batch fabrication is the goal since it promises high repeatability, good quality control and high output per processed substrate. All the technological steps of the LIGA-technique (x-ray lithography, electroplating and molding) are batch processes. Though molding of these structures has been successfully implemented (Bot 94, Mül 95) it will not be covered here since mass fabrication is hardly an issue for space applications and for small numbers of structures x-ray lithography and electroforming is more economic than to setup LIGA molding facilities.

To fabricate acceleration sensors as well as electrostatic actuators the LIGA-process has been improved by a sacrificial layer technique. The process scheme is shown in *Figure 1*. Before the PMMA x-ray resist is cast onto the substrate, an electrical layer of Cr/Ag and a sacrificial layer of Ti are patterned by optical lithography and wet etching. After x-ray exposure and development, the PMMA pattern is filled by electroplating, Ni is used for the acceleration sensor but Ni/Fe permalloy, Cu or Au are also available as a standard. When the resist has been stripped, the sacrificial layer is etched away and finally the substrates are diced and the individual sensors are mounted and bonded onto the circuit board (Moh 90) using the Cr/Ag bond pads on the substrate.

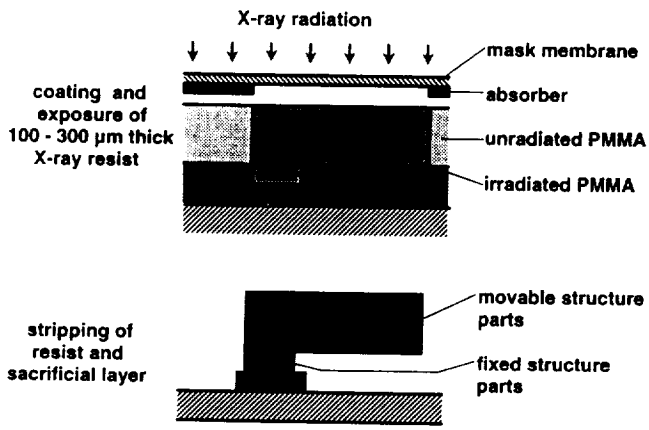


Figure 1: Process Sequence for the Fabrication of Moveable LIGA Micro-structures (sacrificial layer technique)

Figure 2 shows the schematic of the LIGA acceleration sensor where the moveable seismic mass forms two capacitors with the adjacent fixed electrodes. The final design is shown in an SEM picture in Figure 3. The smallest lateral dimension is the width of the capacitor gaps which is merely 4µm and which results in a high zero acceleration capacitance of 4.5pF. The height of the structure of 200µm requires special attention with respect to the development process (Elk 94). The complete sensor is 3mm long and 1mm wide. The mass is suspended with two parallel cantilever beams to two stationary blocks (on the right side). They assure a parallel deflection of the seismic mass which is favorable for high linearity. The width of the beams determines the sensors' maximum acceleration range which may be between 1 and 20g, depending on the design. Four fixed blocks in the corners of the structure prevent a short between the seismic mass and the fixed electrodes. The fixed electrodes are interrupted every 100µm for two reasons: first, during processing, the corresponding PMMA bar stabilizes the 4µm thin PMMA wall which forms the capacitor separation and, second, the width of the interruptions may be used to tune the damping coefficient of the structure depending on the atmosphere to 0.7 as desired for maximum bandwidth. The complex forked geometry of the seismic mass ensures that the capacitance of parts for which the gap width decreases with temperature is identical to parts for which the gap width increases with temperature (Str 93). Figure *4* shows that the

linear term of the temperature dependence is completely suppressed by the forked design. Table 1 summarizes the experimental data of a 1g LIGA-sensor element (Str 94).

Table 1: Experimental Data of the LIGA-Acceleration Sensor (Str 94)

sensitivity	resonance frequency	damping coefficient	thermal zero shift	thermal shift of sensitivity
20%/g	557Hz	0.45	$6*10^{-5}$g/K	$2.3*10^{-4}$/K

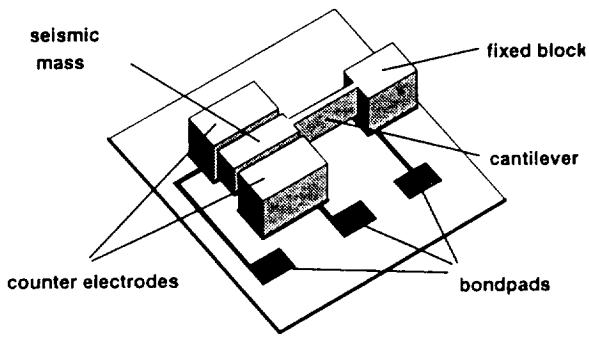


Figure 2: Schematic of a LIGA Acceleration Sensor.

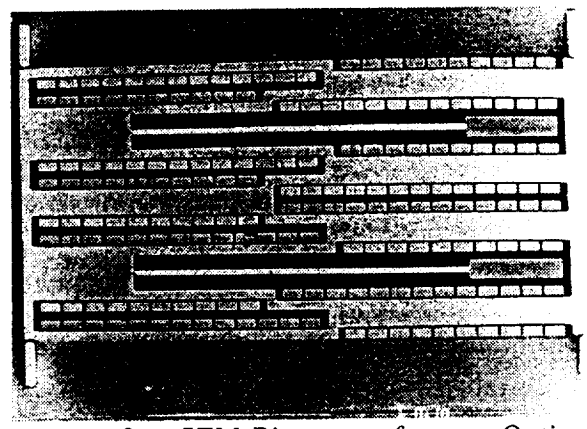


*Figure 3: SEM-Picture of an Optimally Designed Sensor Element. Overall dimensions are 3mm*1mm. For details see the text.*

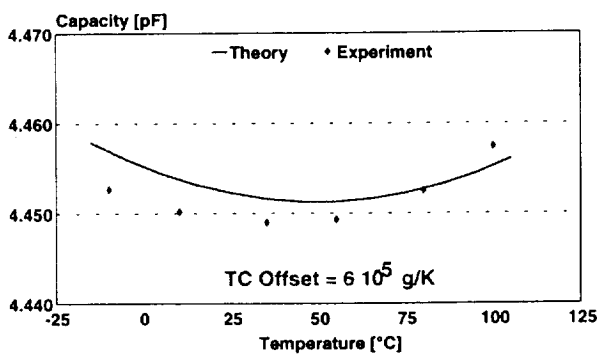


Figure 4: Temperature Dependence of one of the Sensor Element's Capacitances.
A strong linear term which may result from different expansion coefficients of the substrate and the electroplated nickel is completely compensated for by the forked design (Str 93)..

The electrostatic linear actuator shown in Figure *5* has been especially developed with respect to large displacements (Moh 93). *Figure 6* shows the individual teeth in an enlarged way. From Figure *7* it can be seen that a design with parallel plates has a limited displacement of 70µm because for this design the electrostatic force remains constant while the spring force increases. For large displacements, the conical design uses also the force component that rises as $1/d^2$ where d is the separation of the plates. Fabrication tolerances result in an unsymmetric placement of the slider by 1µm. The $1/d^2$ component of the force not only pulls the slider in the desired direction but also onto the sides. For small displacements, this component can be compensated by springs that need to be designed stiffly in this direction but at higher displacements, guiding elements are required that prevent the slider from contacting the electrodes. When the guiding elements are touched (≈ 190µm), friction leads to a drastic decrease of the net force.

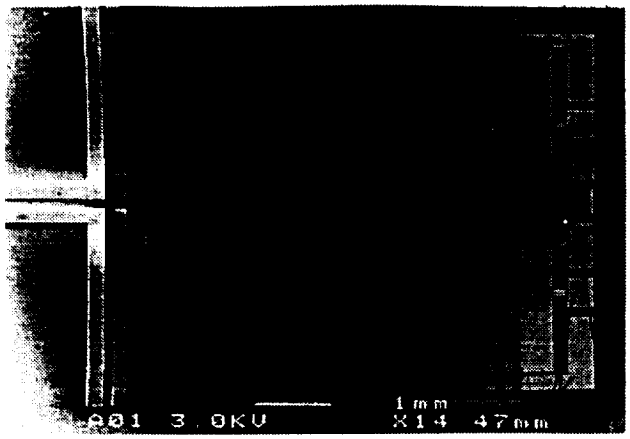


Figure 5: Electrostatic Linear LIGA-Actuator Optimized for Large Displacement.
On the left side, the fixing elements for fibers can be seen. Parallel cantilever springs on the left and right ends of the actuator are used to hold and guide the slider. They need to be stiff with respect to a side motion of the slider (vertical in the picture)(Moh 93).

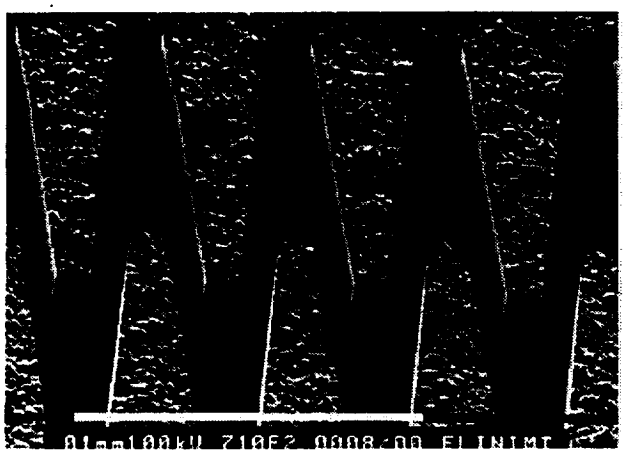


Figure 6: Close-Up View of Individual Capacitor Elements.
The cone angle is 15°. A parallel design referred to in Figure 7 has a cone angle of 0°.

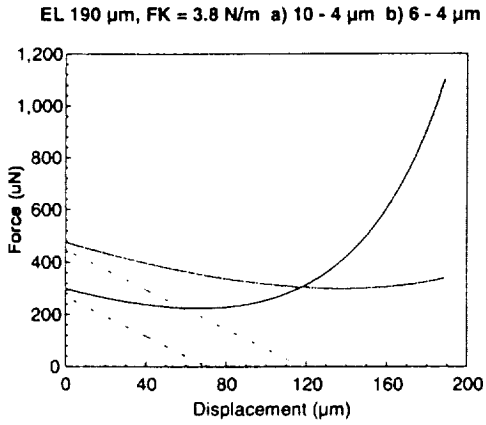


Figure 7: Calculated Force-Displacement Characteristic of the Conical Actuator (70V).
The dashed lines correspond to a parallel plate design with (constant) plate separation of 10µm and 6µm for the lower and higher forces, respectively. The full lines correspond to the conical design with the same initial plate separations.

To fabricate a blazed spectrometer, we use a 3 layer x-ray resist consisting of PMMA as a core-layer and PMMA and a copolymer as cladding layers which are bonded to each other by high pressure welding at temperatures slightly above the glass transition temperature. The transition zone between the layers is approximately 10µm thick (Göt 91). The core layer has a higher index of refraction and guides the light coupled into the structure by a multimode fiber (Figure *8*). By a fiber connector, a clear optical interface is defined. Figure *9* shows part of the grating whose step height is 0.2µm at a step length of 3µm and a vertical height of 125µm which corresponds to the thickness of the fiber. The fabrication of this structure is a technological challenge in all process steps involved: mask

making, x-ray exposure and development and, for mass fabrication, molding. Further details can be found in (Mül 94).

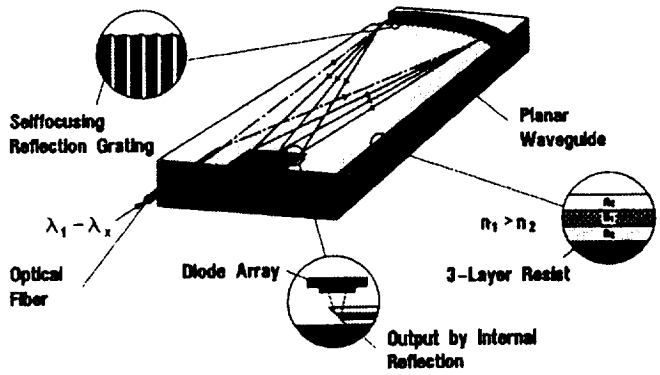


Figure 8: Principle of the LIGA-Microspectrometer. The light is transferred to the device by an optical fiber and is guided by the core layer. A slanted sidewall is used to reflect the light out of the device onto a linear CCD diode array.

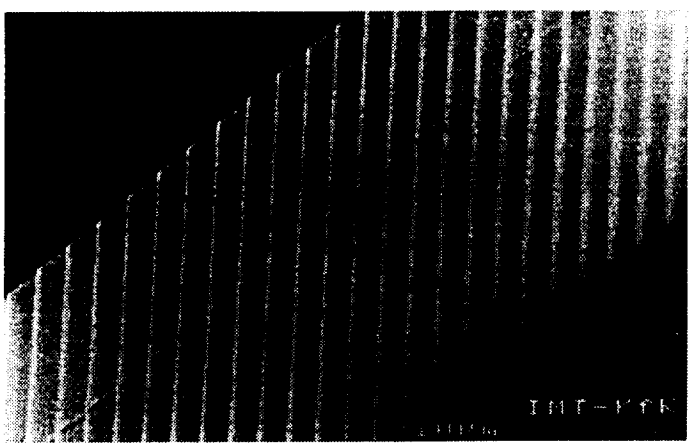


Figure 9: Close-Up View of the Spectrometer's Teeth.
The interface between the core and cladding layer is clearly visible. The step height is 0.25µm, their length is 3µm. The sharpness of the teeth extends over the whole resist height of 175µm (Mül 93).

In order to measure the intensity of the diffracted light, a slanted sidewall of the PMMA structure is used to direct the light out of the spectrometer plane (Figure *8*). This makes the alignment of a linear CCD-diode array with a pixel size of 25*500µm very easy. An evaluation board which includes ADCs, a microcontroller and a serial port has been developed and has been fabricated in SMD multilayer technique. The complete set-up fits into a box of 70mm*60mm*15mm. The experimental data of the spectrometer are listed in Table 2.

Table 2: Characteristic Data of LIGA-Microspectrometer

spectral range	transmission at λ_{blaze}	spectral resolution	attenuation to scattered light	optical fiber	diode array	dynamic range	measuring time
400-1100nm	25%	7nm	25dB	50/125µm gradient index	Hamamatsu S5464-512F	10000 - 20000	40ms - 2560ms

Besides x-ray lithography, micro milling and drilling has evolved as an important patterning technology to fabricate molding tools (Bie 93). As shown in Figure *10*, several levels may be fabricated with smallest dimensions being limited by the available tools: 50 µm diameter for a drill and 300 µm for a micro-end mill. To fabricate pneumatic devices for example a pump (Büs 94) or an active valve system (Fah 94) such tools have been extensively used to mold the casings from PSU. Their fabrication also requires thin-film deposition, optical lithography and adhesive bonding. Particularly adhesive bonding had to be developed and the main contribution has been the

development of aligned adhesive bonding for batch processing using capillary forces in order to deposit the adhesive in the desired places (Maa 94).

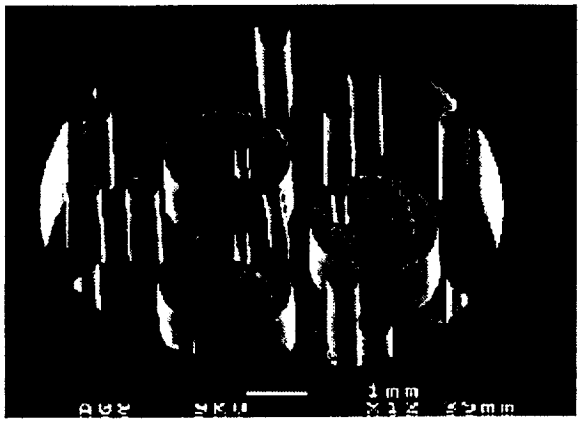


Figure 10: SEM-Picture of a 4 Level Molding Tool Fabricated by Micro Milling and Micro Drilling.
The smallest dimensions are holes of 50µm and the width of areas where material has been removed can not be smaller than 300µm. The limitations result from the dimensions of the smallest commercially available tools (Fah 94).

The micropump shown in Figure *11* is made of two PSU casings between which a polyimid membrane is placed. The polyimid membrane seals the actuator chamber and by pulsed resistive heating, the air volume underneath the actuator chamber is displaced. Two passive valves are used to obtain a directed flow of the displaced gas. With these pumps, flow rates in the range of 220µl/min can be achieved at a pulse rate of 30Hz (Table 3).

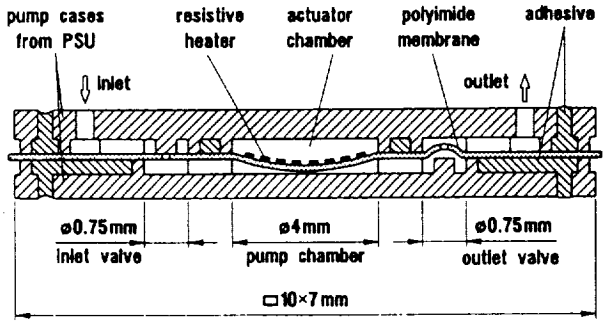


Figure 11: Schematic of a Micropump for Gases.
The fabrication makes use of thin-film technology, optical lithography, molding and adhesive bonding (Büs 94).

Table 3: Characteristic Data of the Micropump for Gases

drive voltage	max. outlet pressure	max. flow rate	pulse length	pulse frequency	dimensions
15V	130 hPa	220µl/min	2ms	30Hz	10*7*1mm³

Triaxial Acceleration Sensor System

Since acceleration is a vector, it needs to be measured in all three directions of space. This can be accomplished by using two LIGA-sensor elements for the directions perpendicular to the substrate normal (Str 94) in combination with a silicon sensor element for the normal direction. With this planar set-up little or no alignment of the sensor elements is necessary since the LIGA-structures may be processed orthogonally by design on a single substrate and the third direction is the natural direction of sensitivity of the silicon sensor. Figure *12* shows the completed sensor system including hybrid electronics and digital signal processing.

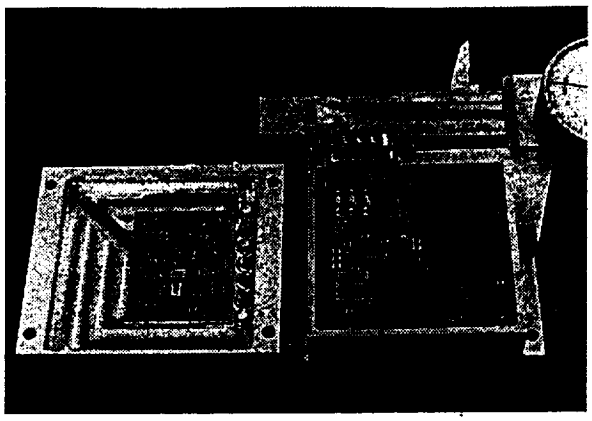


Figure 12: Complete Triaxial Acceleration Sensor System.
Two LIGA-Sensors measure within the substrate direction, one silicon sensor measures the normal direction. On the left side, the sensor elements and the analogue board, on the right side ADCs and digital electronics for preprocessing and external communications are shown.

The key to low-cost and efficient sensor systems are sensor elements with high linearity and low thermal drift because for such an excellent sensor element, expensive corrections of the data are not necessary. In order to preserve the good characteristics of the LIGA-acceleration sensor element, a hybrid feedback read-out circuit has been developed. It uses controlled electrostatic forces to keep the seismic mass centered between the fixed electrodes. It has been optimized with respect to a high 3dB bandwidth and low noise. Table 4 summarizes the experimental results obtained with this circuit. It should be noted that the current resolution limit (1.2μg/√Hz) is due to the mechanical noise of the equipment and that the bandwidth includes zero frequency (DC) as well. Table 4 summarizes the characteristic data of the sensors. A system that has been extended to obtain redundancy and that includes data preprocessing has been described in (Stro 94).

Table 4: Characteristic Data of 2g Acceleration Sensor Element with Feedback Readout

range	sensitivity	linearity	thermal zero shift	resolution	3dB frequency
2g	1971mV/g	0.8%	90μg/K	28μg (1.2μg/√Hz)	570Hz

Microoptical Bypass Switch

In optical communications, a device is needed to bypass a faulty user or amplifier. The electrostatic actuator shown in the previous section fulfills three major requirements: the LIGA sidewall has a very small roughness so that it can be used as a mirror, the displacements are somewhat larger than the diameter of a collimated fiber beam and no energy is required to keep the actuator at any position. For monomode applications, an optical bench is required that images the fiber ends onto each other. The schematic of Figure *13* illustrates the principle. Since imaging with ball lenses of the diameter of the fibers (125μm) is physically prohibitive, commercially available ball lenses of 900μm diameter have been used. This makes the fabrication of the system more complex because a major requirement is to have the optical axis parallel to the substrate surface within very narrow tolerances (Mül 93). By anisotropic etching of (100) silicon wafers, the lenses may be positioned vertically with the required level control of less than 1μm. For the lateral fixing of the lenses a LIGA resist pattern is used (Figure *14*). The precision of the lateral position of the fixing elements is strongly dependent on the stresses that bend the whole substrate during processing. For the most influential process steps, sputter deposition and electroplating, special parameters had to be found that minimize this stress. The coupling losses of the optical part has been measured in a separate set-up. They vary between 1.6 and 2.4dB of which almost 1dB may be attributed to reflection (Mül 95). The measured values are well within the requirements of 3dB for this system. The complete system that includes the actuator is currently being investigated experimentally.

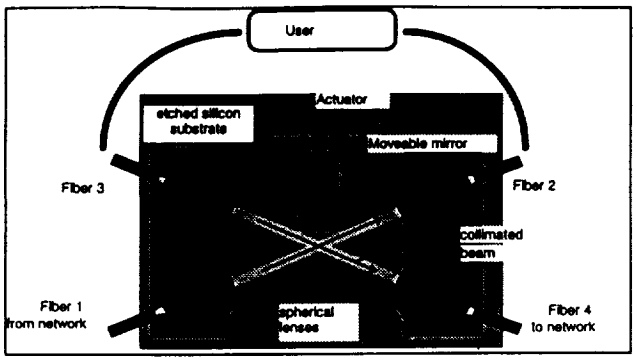


Figure 13: Schematic of the Bypass Switch. In the active mode which is shown, the signal proceeds from fiber 1 to fiber 4 via the user and an amplifier. In the inactive mode, the signal is coupled directly from fiber 1 via the mirror to fiber 4 (Mül 95).

Figure 14: SEM-Picture of a Lens Leveled by an Etched Silicon Substrate and Fixed by LIGA-Structures.

Microsystem for Optochemical Analysis of Pollutants

Figure *15* shows the schematic layout of a microsystem for optochemical analysis of heavy metal ions in water. It consists of a module of 4 micropumps for fluid handling, a microcuvette with chemical sensors on its sides, the LIGA-spectrometer to measure the extinction over the whole wavelength range and a 16 bit µ-controller to control the micropump and the data transfer to a PC. The 4 pumps are working in a special sequence so that a reference liquid as well as the sample liquid are pumped through the microcuvette. To avoid contamination of the pumps with liquids, they are working as pressure or vacuum pump.

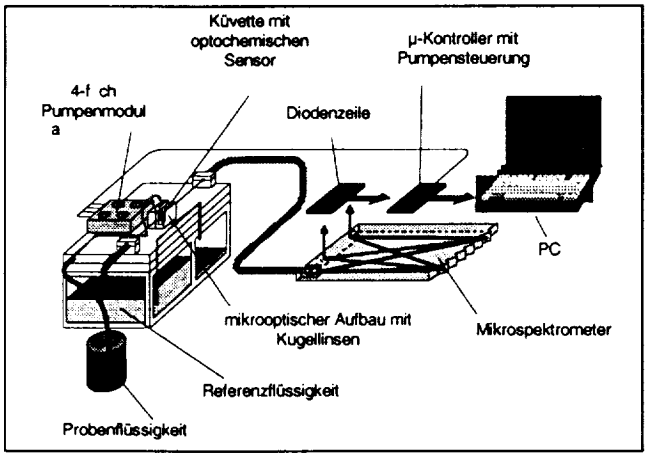
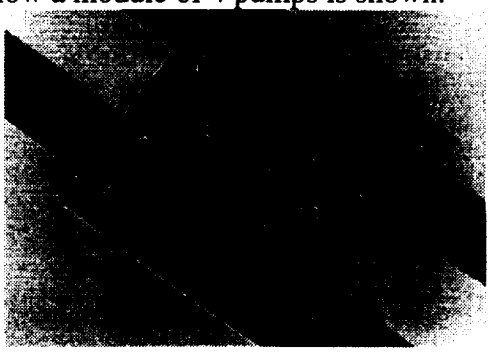


Figure 15: Principle of the Optochemical Analysis System.
Below a module of 4 pumps is shown.



The optochemical sensors change their transmission behavior depending on the concentration of Hg, Pb, and Cd. A multicomponent analysis has been realized by using derivatives of porphyrin so that different metal ions show different absorption coefficients in different spectral ranges (Rec 93). The following limits of detectability have been measured:

Hg(II)	30µg/l
Cd(II)	1µg/l
Pb(II)	20µg/l

Long time stability has been considerably improved by binding the porphyrin to a macromolecular carrier with a molar weight of 60000. At constant concentration, the sensor signal decays by merely 15% over a time period of 40 days (Rec 93).

Conclusions

In the past, the efforts in microtechnology have concentrated on different fabrication technologies and on components compatible with these technologies. The LIGA-technique is most important because the large structural height is advantageous with respect to actuator forces, the extreme sidewall smoothness makes optical applications possible and the small structural detail may be exploited in diffraction optics.

The microsystems presented here clearly indicate that the technologies and components are sufficiently advanced in order to design real systems. It has been shown that future improvement is expected by combining different technologies, for example LIGA and silicon-etching or LIGA and micromachining. To facilitate this, a major prerequisite for the near future is the definition of standardized interfaces between components fabricated by different technologies.

Acknowledgments

The microsystems presented here have been mainly developed during the last three years. Many people have contributed to the design and to their successful fabrication. The authors would like to thank the coworkers of the Karlsruhe Research Center who are named in the references. Their cooperation in submitting figures and photographs is gratefully appreciated. The authors would also like to express their deep gratitude to the technical personnel who have contributed to the fabrication of the actuators through all these years.

Literature

(Bec 86) E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Münchmeyer, Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic molding (LIGA process).Microelectronic Engineering 4, (1986) pp. 35-56.

(Bie 93) W. Bier, A. Guber, G. Linder, Th. Schaller, K. Schubert:
Mechanische Mikrofertigung - Verfahren und Anwendungen
KfK-Bericht 5238 (1993) pp.132-137.

(Büs 94) B. Büstgens, W. Bacher. W. Bier, R. Ehnes, D. Maas, R. Ruprecht, W.K. Schomburg
Micromembrane Pump Manufactured by Molding

Proceedings Actuator '94, Bremen,15. -17. Juni 1994, S. 86 - 90

(Elk 94) A.El-Kholi, J.Mohr, R.Stransky, Ultrasonic Supported Development of Irradiated Microstructures, Microelectronic Engineering 23 (1994), 219-222

(Fah 94) J. Fahrenberg, D. Maas, W. Menz, W.K. Schomburg
Active Microvalve System Manufactured by the LIGA Process
Proceedin s Actuator 94, Bremen,15. -17. Juni 1994, S. 71- 74

(Göt 91) J.Göttert, J.Mohr, C.Müller, Mikrooptische Komponenten aus PMMA, hergestellt durch Röntgentiefenlithographie, Werkstoffe der Mikrotechnik -Basis für neue Produkte-, VDI/VDE Gesellschaft Feinwerktechnik, Karlsruhe, Oktober 1991.Tagungsbericht VDI 933, (1991) S. 249-263.

(Maa 94) D. Maas, B. Büstgens, J. Fahrenberg, W. Keller, D. Seidel
Application of Adhesive Bonding for Integration of Microfluidic Components,
Proceedings Actuator 94, Bremen,15. -17. Juni 1994, S. 75 - 78

(Moh 90) J. Mohr, C. Burbaum, P. Bley, W. Menz, U. Wallrabe:
Movable Microstructures Manufactured by the LIGA Process as Basic Elements for Microsystems.
H.Reichl (ed.): MICRO SYSTEM TECHNOLOGIES 90, Springer Verlag (1990),
pp. 529-537.

(Moh 93) J.Mohr, M.Kohl, W.Menz, Micro Optical Switching by Electrostatic Linear Actuators with Large Dsiplacements, 7th Int. Conf. Solid-State Sensors and Acutators (Transducers '93) Yokohama, June 7-10 (1993) 120.

(Mül 94) C.Müller, Miniaturisiertes Spektrometersystem in LIGA-Technik, Disseration, Fakultät für Maschinenbau, Universität Karlsruhe, (1994)

(Mül 93) A Müller, J. Göttert, J. Mohr, LIGA microstructures on top of micromachined silicon wafers used to fabricate a microoptical switch, J. Micromech. Microeng. 3 (1993) 158 - 160

(Mül 95) A.Müller, J.Göttert, M.Kohl, J.Mohr, K.Schorb, W.Voit, "Aufbau und Verbindungstechnik als Basistechnologie für elektrische und optische Mikrosysteme", 4. Zwischenbericht zum 6. Statusseminar, 23. März 1995, BMBF-Verbundprojekt AVT-KEO

(Rec 93) J.Reichert, R.Czolk, S.C.Kraus, G.Heinzmann, A.Morales-Bahnik, J.Reinhardt, W.Sellien, H.J.Ache, Development of Optochemical Microsensors for Environmental Analysis, Proceedings Sensor 93, Oct. 11-14, 1993, Nürnberg, Vol 2, 55

(Str 93) M.Strohrmann, P.Bley, O.Fromhein, J.Mohr, Acceleration Sensor with Integrated Compensation of Temperature Effects Fabricated by the LIGA Process, Proc. Eurosensors VII, Budapest, September 26-29, 1993, 426-429

(Str 94) M.Strohrmann, F.Eberle, O.Fromhein, W.Keller, O.Krömer, T.Kühner, K.Lindemann, J.Mohr, J.Schulz, *Smart Acceleration Systems Based on LIGA Micromechanics,* Microsystem Technologies '94, 4th Int. Conf. Micro-, Opto-, Mechanical Systems and Components, Oct. 19-21, Berlin, 753 (1994)

MICROMECHANICAL MACHINING PROCESSES AND THEIR APPLICATION TO AEROSPACE STRUCTURES, DEVICES, AND SYSTEMS

Craig R. Friedrich and Robert O. Warrington
Institute for Micromanufacturing (*IfM*)
Louisiana Tech University

/
71156
030692
6 P.

Abstract

Micromechanical machining processes are those micro fabrication techniques which directly remove work piece material by either a physical cutting tool or an energy process. These processes are direct and therefore they can help reduce the cost and time for prototype development of micro mechanical components and systems. This is especially true for aerospace applications where size and weight are critical, and reliability and the operating environment are an integral part of the design and development process. The micromechanical machining processes are rapidly being recognized as a complementary set of tools to traditional lithographic processes (such as LIGA) for the fabrication of micromechanical components. Worldwide efforts in the U.S., Germany, and Japan are leading to results which sometimes rival lithography at a fraction of the time and cost. Efforts to develop processes and systems specific to aerospace applications are well underway.

Micromechanical Machining Processes

The micromechanical machining processes are divided into two distinct categories based on the method of material removal. The force processes include single point diamond micromachining (SPDM), micromilling, microdrilling, and sawing, grinding and polishing. These processes are spinoffs of traditional mechanical processes but have been specifically adapted to machine features as small as several micrometers with sub-micrometer tolerances. Merely shrinking the traditional processes has not proven entirely successful. The micromechanical processes have undergone specific changes for adaptation at the microscale. The second category is the forceless processes which include laser ablative micromachining, micro electrical discharge micromachining (mEDM), and focused ion beam machining. Although it is more akin to microlithographic processes, laser-based photopolymerization (micro stereo lithography) is also included in this category. All of these processes are undergoing various levels of integration to provide much more flexibility to the microsystems designer. A summary comparison of the various micromanufacturing processes in common use are shown in **Table 1**.

The micromechanical processes directly remove material in a single step rather than many serial steps found in lithography. To create a microstructure by the LIGA method, for example, an x-ray mask must be first fabricated. This requires CAD layout of the pattern, an electron beam writing step to transfer the pattern to a thin photoresist. The resist is then developed and electroplated with chromium to form an optical lithography mask. The mask is then used to transfer the pattern to a photoresist which is typically 2-3 micrometers thick. The developed pattern is then electroplated with gold to form an intermediate x-ray mask. This mask is used with an x-ray source, a synchrotron for example, to form a final x-ray mask with a gold absorber 10-15 micrometers thick. The final mask is then used to transfer the pattern to a resist up to centimeters in thickness. The exposed pattern is developed and electroplated with nickel or other metals to form the final microstructure. A portion of a mold for a micro fluidic device, produced by x-ray lithography, is shown in **Fig. 1.**

It is readily seen that the walls are vertical, straight, and smooth. In addition the top edge is sharp and without burrs, and the ratio of the vertical to lateral dimensions (aspect ratio) is large. These are typical advantages of molds and structures made by x-ray lithography and are attributes which a final product may need to have. The disadvantage is that

Table 1. Process Comparison

	Micromechanical Machining	X-ray Micromachining	Bulk Si Processing Techniques	Surface Processing Techniques
Feature Height	um's - mm	1000 um - cm's	500 um - mm's (using bonding techniques)	10 - 20 um
Cross-Sectional Shape	Very Good	Very Good	Fixed by Crystal	Very Good
Cross-Sectional Shape Variation with Depth	Good	Limited	Limited	Very Limited
Materials	Wide Range Possible	New Materials Being Developed Continuously	Fixed	Fixed
IC Compatibility	Good	Good	Fair	Excellent
Process Maturity	Developing	First Products	Fairly Mature	First Products
Aspect Ratio	Large	Large	Small	Medium
Low Volume	Excellent	Poor	Fair	Poor
High Volume	Possible	Good	Excellent	Excellent

an x-ray mask made with a gold absorber typically costs $10K to $30K or more depending on the dimensional tolerances required. The cost for using a 100keV electron beam writer to create an x-ray mask in one step, thus eliminating the need for the optical and intermediate masks, has been quoted at $1K per hour. Gold electroplating must still be performed and this process uses toxic and environmentally hazardous materials such as cyanides. The cost of a synchrotron is tens-of-millions of dollars and the annual operating costs are over a million dollars. For component and system prototyping where low cost and rapid turn around are important, the lithographic processes have distinct disadvantages.

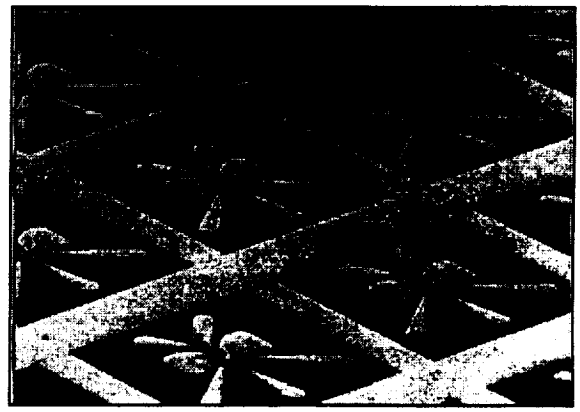


Figure 1 Lithography Mold

To help reduce the cost and time for mask and mold fabrication at the prototype stage, or at the final stage if the component attributes meet the design requirements, micromilling has been developed [1-4] at the *IfM*. This is presented as an example of the use of the micromechanical machining techniques. Many of the other processes are also being adapted in a similar manner. At present, the process uses micromilling tools which are 22 micrometers in diameter and fabricated with focused ion beam micromachining. The tool material is M42 cobalt-high speed steel (Rc 65-67). A typical tool is shown

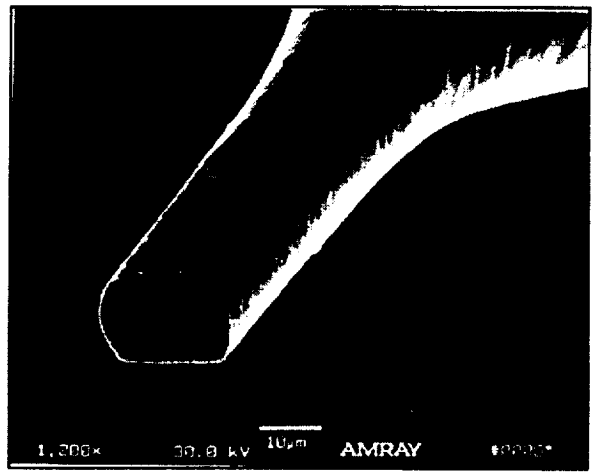


Figure 2 Micromilling Tool

in **Fig..2**. The tools are used to directly mill polymethyl methacrylate (PMMA) mold and mask features. The deepest features to date are 62 micrometers and the thinnest walls are 8 micrometers wide. Sidewalls are very vertical (89.5 degrees or better) and the rms roughness is 0.1 micrometers. Complex shapes can be milled simply by programming the high precision micromilling machine used for this process. An example of this complex geometry is shown in **Fig. 3** which is just a small portion of the overall pattern. The pattern is over 2mm in diameter and over 35 million cubic micrometers of material was removed in just three hours. A specially designed micromilling/micromilling/mEDM machine was developed which has nanometer-level positional resolution, sub-micrometer positional accuracy and repeatability, a massive granite structure for thermal and vibrational stability, and air bearing slides and spindle. Although the machine cost nearly $300K to develop, the fact that it allows very rapid production of molds and masks gives this process considerable potential to become a mainstream micromanufacturing technique.

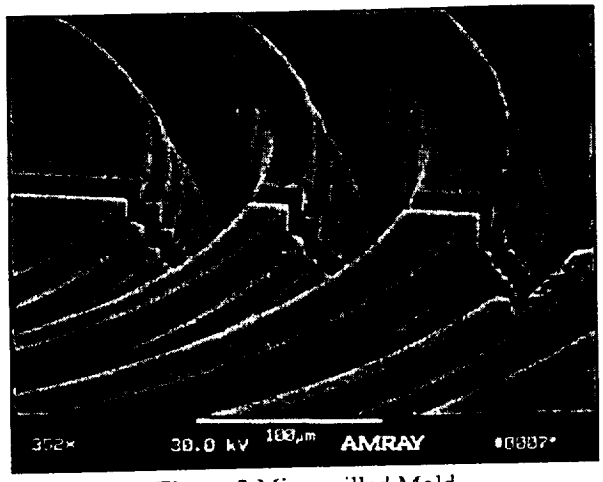


Figure 3 Micromilled Mold

Worldwide Programs in Micromechanical Machining

Because the micromechanical processes are evolving and being applied very rapidly, it is difficult to provide a complete overview of efforts. The programs shown for the U.S., Germany, and Japan are somewhat representative of those also taking place in England, Switzerland, The Netherlands, and Taiwan. Historically, Germany is known for high precision manufacturing at traditional scales. The LIGA process was commercialized for micromechanical structures in Germany and since that time (the 1980's) there has been a primary focus on that technique. As the technology has matured, the precision micromechanical processes are seeing increased use to support LIGA[5].

In the U.S., deep x-ray lithography for micromechanical applications has primarily resided at the University of Wisconsin. Recently, LSU-Baton (Center for Advanced Microstructures and Devices/CAMD) and Louisiana Tech University/IfM, are applying deep x-ray lithography for production of microstructures and systems. Because the LIGA process technology had been widely published prior to the establishment of the IfM, it was known that the micromechanical machining processes would have widespread utility and were therefore included as an integral part of the Institute. The IfM operates all of the micromechanical processes previously identified.

Japan still lags behind in the area of deep x-ray lithography primarily due to an apparent lack of interest or economic driver[6]. Because of this, Japan has many successes in the micromechanical techniques, especially in micro EDM, silicon-based technologies, and microsystem concepts and designs. The Japanese government, through MITI, continues to aggressively fund the microtechnologies and there is considerable support and involvement from the leading technology companies based in Japan. This support is also evident from the establishment of the Micro Machine Center (MMC) several years ago. This center has served as a clearinghouse for R&D information from around the world.

Aerospace Applications of Micromechanical and Lithographic Fabrication

Smaller, lighter weight, more reliable, and less expensive are adjectives for components and systems used in aerospace applications. These are also phrases often used to describe micro-manufactured components and systems. Such devices are inherently smaller and therefore lighter weight. Often, these devices are designed to have very few independent parts because of the difficulty in assembly. With fewer parts, greater reliability can be achieved. With smaller parts also comes reduced

statistical incidence of material defects or non-homogeneities. Using mass fabrication techniques pioneered by the semiconductor industry, or direct fabrication techniques during prototype development, the unit cost of micromanufactured parts can be less than by traditional processes. The following examples represent a summary of applications worldwide and is not all inclusive. The components, systems, and processes are generally for both aeronautical and space applications.

The cost of placing payloads in Earth orbit or for planetary missions is continually increasing. Therefore smaller analytical instruments are the obvious next step. Microspectrometers have been designed and built to the commercial stage in Germany[7,8]. The device operates in the 2.75-3.35 micrometer spectrum by using a reflection grating patterned in a PMMA layer by x-ray lithography. After development, the PMMA is given a reflective coating. Light is introduced onto the grating by a 150 micrometer diameter multi-mode fiber. The resolution of the device is 31 nanometers. A total of 2000 of the devices have been placed into production.

An electrochemical microanalytical system has been develop[9] where potentiometric microsensors are combined with solid state ion sensitive membranes, plastic molded micropumps, a microchannel system for connecting the components, and microelectronic components for signal processing and system control. At present, the system has been developed to measure pNa and pH.

In Japan, reactive ion etching has been used to fabricate an electromagnetically driven resonant angular rate sensor[10]. A silicon mass, 60 micrometers wide by 190 micrometers thick, is caused to resonate along one axis by electromagnets. Rotation about a second orthogonal axis causes a vibration along the third axis by the Coriolis effect. This vibration is sensed by a capacitance change. The device had a reported sensitivity of 6 fF sec./deg. and rates in the range of 240 deg./sec. to 360 deg./sec. were measured.

Applications which are being studied or have been developed to the prototype stage at the IfM include electrostatically driven rotary and linear actuators, self diagnostic bearing elements which can report several factors concerning the bearing operating environment, micro scale surface modifications to increase heat transfer in micro heat exchangers and to reduce drag on free and enclosed surfaces, small "smart" projectiles on the scale of a 0.50-caliber bullet and the control surfaces required at this scale, micro antennae for near-infrared communication systems, and large L/D micro channel plate detectors for night vision applications.

Summary

As with many process technologies, there are a variety of ways in which they can be applied to solve development problems. The micromechanical machining processes should be viewed as an additional set of tools available to the microsystems designer and fabricator. The processes can be used as standalone techniques or integrated with more traditional lithographic techniques by pre- or post-processing. These techniques include substrate preparation such as polishing for smoothness and dimensional control, for planarizing thick resist layers by polishing or direct diamond machining, for planarizing multiple layers of resist and electroformed structure materials in the so-called "step LIGA" process for creating variable vertical geometries, to using microdrilling to remove taper left in deep holes by electron beam lithography. The current challenge is to develop standard procedures, fixtures, and reference location schemes so the integration of these processes will not detract from productivity or dimensional quality of the finished parts and systems.

References

1. Friedrich,CR and Vasile,MJ, "Development of the Micromilling Process for High Aspect Ratio Microstructures," ***Journal of Microelectromechanical Systems*** (in press).
2. Friedrich,CR and Vasile,MJ, "The Micromilling Process for High Aspect Ratio Microstructures," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).

3. Friedrich,CR and Kikkeri,B, "Rapid Fabrication of Molds by Mechanical Micromilling:Process Development," ***Proceedings of SPIE Conference on Micromachining and Microfabrication '95***, Austin, TX (in press).
4. Friedrich,CR and Kikkeri,B, "Mechanical Micromilling of PMMA," ***Proceedings of 1995 ASPE Annual Conference***, Austin, TX (in press).
5. Fahrenberg,J, Schaller,T, Bacher,W, El-Kholi,A, and Schomburg,W, "High Aspect Ratio Multi-Level Mold Inserts Fabricated by Mechanical Micro Machining and Deep Etch X-Ray Lithography," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).
6. Umeda,A, "Present State of LIGA Technology in Japan," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).
7. Muller,C, Mohr,J, and van der Sel,C, "Microspectrometer for the IR Range," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).
8. Bacher,W, Hagena,O, Heckele,M, and Muller,C, "Small-Scale Series Production of LIGA Microspectrometer Components," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).
9. Suss,W, Eggert,H, Georges-Schleuter,M, Jakob,W, and Lindemann,K, "Step by Step Development of an Electrochemical Microanalytical System," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).
10. Choi,J, Minami,K, and Esashi,M, "Application of Deep Reactive Ion Etching for Silicon Angular Rate Sensor," ***Proceedings of HARMST'95***, Karlsruhe, Germany (in press).

47 71157

ADVANCING MEMS TECHNOLOGY USAGE THROUGH THE MUMPS (MULTI-USER MEMS PROCESSES) PROGRAM

D. A. Koester, K. W. Markus, V. Dhuler, R. Mahadevan, A. Cowen
MCNC MEMS Technology Applications Center
Research Triangle Park, NC 27709-2889
mems@mcnc.org
http://www.mcnc.org/HTML/ETD/MEMS/memshome.html

D30695

8P

Abstract

In order to help provide access to advanced MEMS technologies and lower the barriers for both industry and academia, MCNC and ARPA have developed a program which provides users with access to both MEMS processes and advanced electronic integration techniques. The four distinct aspects of this program, the Multi-User MEMS Processes (MUMPs), the Consolidated Micromechanical Element Library, Smart MEMS and the MEMS Technology Network will be described in this paper. MUMPs is an ARPA-supported program created to provide inexpensive access to MEMS technology in a multi-user environment. It is both a proof-of-concept and educational tool that aids the development of MEMS in the domestic community. MUMPs technologies currently include a 3-layer polysilicon surface micromachining process and LIGA processes that provide reasonable design flexibility within set guidelines. The Consolidated Micromechanical Element Library or CaMEL, is a library of active and passive MEMS structures that can be downloaded by the MEMS community via the internet. Smart MEMS is the development of advanced electronics integration techniques for MEMS through the applications of flip chip technology. The MEMS Technology Network or TechNet is a menu of standard substrates and MEMS fabrication processes that can be purchased and combined to create unique process flows. TechNet provides the MEMS community greater flexibility and enhanced technology accessibility.

Introduction

Over the last decade silicon process technology, synonymous with integrated circuit processing, has been increasingly applied to the field of micromechanics, leading to the emerging field of MEMS (microelectromechanical systems). The extensive characterization of silicon processes by the IC industry, integrated with silicon's high Young's modulus and yield-strength, high thermal conductivity and low thermal expansion coefficient makes it one of the best understood and well suited materials available for coupled electronic and mechanical applications.

MEMS is an enabling technology, which partially accounts for the projections of 10 - 20% annual growth and the potential of a greater than $8 billion market by the year 2000[1]. Current market estimates of approximately $1 billion are possibly low since MEMS are already being incorporated into much more complex systems. Due to the enabling nature of MEMS and because of the significant impact they can play on both the commercial and defense markets, the federal government has taken special interest in nurturing growth in this field. One of the many ways this is being accomplished is through the MCNC MEMS Infrastructure program, supported by the Advanced Research Projects Agency (ARPA).

The MEMS Infrastructure program was established in 1993 to provide low-cost, easy access to advanced MEMS technologies. By lowering the barriers to the technology, it is hoped that the cost (both time and dollars) of developing and incorporating MEMS into new applications will be significantly reduced. There are three key components to the Infrastructure program; the Multi-user MEMS processes (MUMPs), a publicly-accessible standard element MEMS library and the generation of Smart MEMS through the use of flip chip technology. This Infrastructure base has been expanded through the MEMS Technology Network or TechNet. TechNet consists of two components, a wafer inventory of standard substrates common to MEMS fabrication and a menu of common process modules that allow users to piece together unique MEMS processes for more overall fabrication flexibility than offered in MUMPs. These activities are described below.

The Multi-User MEMS Processes (MUMPs)

As with integrated circuits, the cost of MEMS devices benefits from the leveraging of batch fabrication technology. Even so, the costs of micron-scale silicon processing is enormous. Building, maintaining and operating a cleanroom with knowledgeable workers can out price most universities, small and even medium-sized companies with interest in developing and/or commercializing MEMS technology. Commercial prototyping services are available but the cost of developing a full prototype can easily exceed $250K. As is often the case, MEMS fabrication is too costly for individuals interested in experimenting with this technology. As a result, the broad dissemination of MEMS technology is inhibited and the pool of creativity contributing to new and potentially lucrative products is restricted. It is these access and cost barriers to MEMS development that prompted ARPA and MCNC to establish a program of multi-project micromachining processes.

By providing an inexpensive route to MEMS fabrication technologies, the MUMPs program has filled two important niches. First, MUMPs provides cost effective proof-of-concept fabrication. This is particularly important for small (and even large) companies where R&D funds are limited and providing some kind of physical evidence beyond the "idea stage" is a prerequisite to further project funding. Once ideas have been displayed through the MUMPs process, users may wish to take their ideas further to the prototyping stage. Second, MUMPs provides a excellent, low-cost, educational tool. Since MEMS is, in many ways, still in its infancy, there is plenty of room for growth. Many universities are beginning to develop programs in MEMS. Small companies and government agencies interested in MEMS may find themselves relegated to reading journals but never testing their knowledge because of fabrication costs. The low cost of the MUMPs program gives these individuals an opportunity to learn more about MEMS technology and a means of testing their ideas.

The MUMPs program currently offers access to two distinct MEMS technologies: polysilicon surface micromachining and LIGA (hereafter referred to as LIGAMUMPs), which is provided in conjunction with the University of Wisconsin.

MUMPs

Polysilicon surface micromachining encompasses many of the same fabrication techniques as traditional silicon IC fabrication where layers of CVD films are deposited and subsequently patterned using photolithography and plasma etch techniques. Alternating layers of silicon dioxide and polysilicon are deposited so when all the layers have been patterned, the silicon dioxide can be etched away with hydrofluoric acid leaving only the polysilicon structures behind. The chief benefit to this process is the ability to fabricate moving parts on a silicon substrate. The MUMPs process provides two layers of structural polysilicon and a third layer of thin polysilicon that serves as an electrical ground plane or electrode. Devices fabricated in this way are typically several microns in thickness. With some devices having in-plane dimensions of greater than 500 μm, surface micromachined devices can take on a two-dimensional appearance. Nevertheless, the mechanical strength of polysilicon films makes such structures surprisingly robust.

Table 1 MUMPs (polysilicon) process specifications

Mask Level	Nom. Geometry	Material	Comments
Poly 0	3.0 μm	0.5 μm poly	Ground plane, low stress poly
Dimple	3.0 μm	-	0.75 μm dimple into first oxide
Anchor 1	3.0 μm	2.0 μm PSG (1st oxide)	Creates poly 1 anchors
Poly 1	3.0 μm	2.0 μm poly	First structural layer
Poly1_Poly2_Via	3.0 μm	0.75 μm PSG (2nd oxide)	Creates vias between poly 1 and poly 2
Anchor 2	3.0 μm	2.75 μm PSG (1st & 2nd oxide)	Creates anchor for poly 2
Poly 2	3.0 μm	1.5 μm poly	Second structural layer
Metal	3.0 μm	0.5 μm Cr/Au	Evaporated metal

The MUMPs process provides seven different films (layers) with which to build devices. These films are silicon nitride, poly 0, first oxide, poly 1, second oxide, poly 2 and metal. Table 1 describes these layers, their purpose and their associated lithography levels. The purpose of the

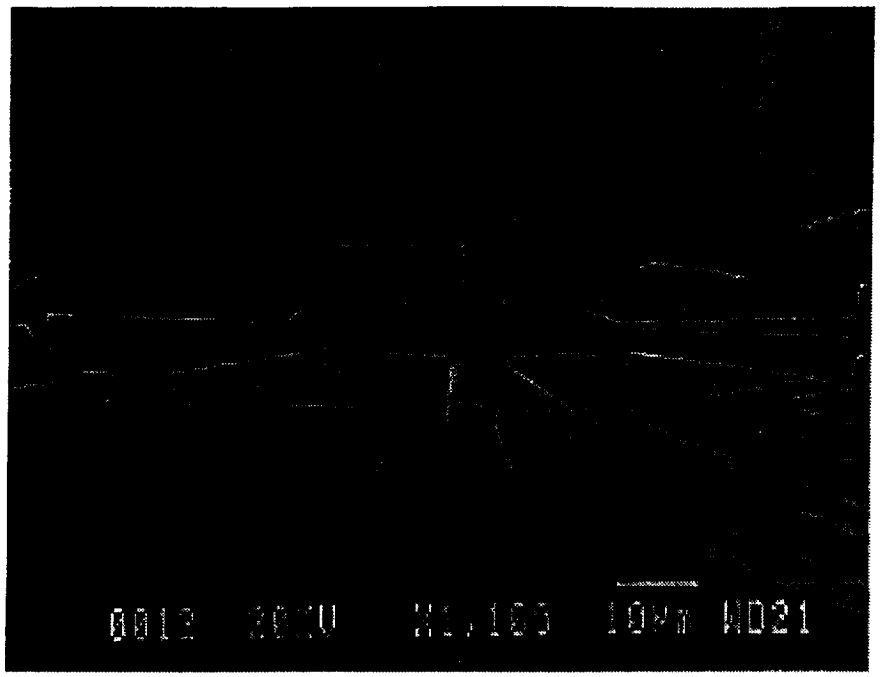


Figure 1 SEM of a micromotor fabricated using the MUMPs process. The rotor is approximately 80 μm in diameter and the rotor-stator gap is 2.0 μm.

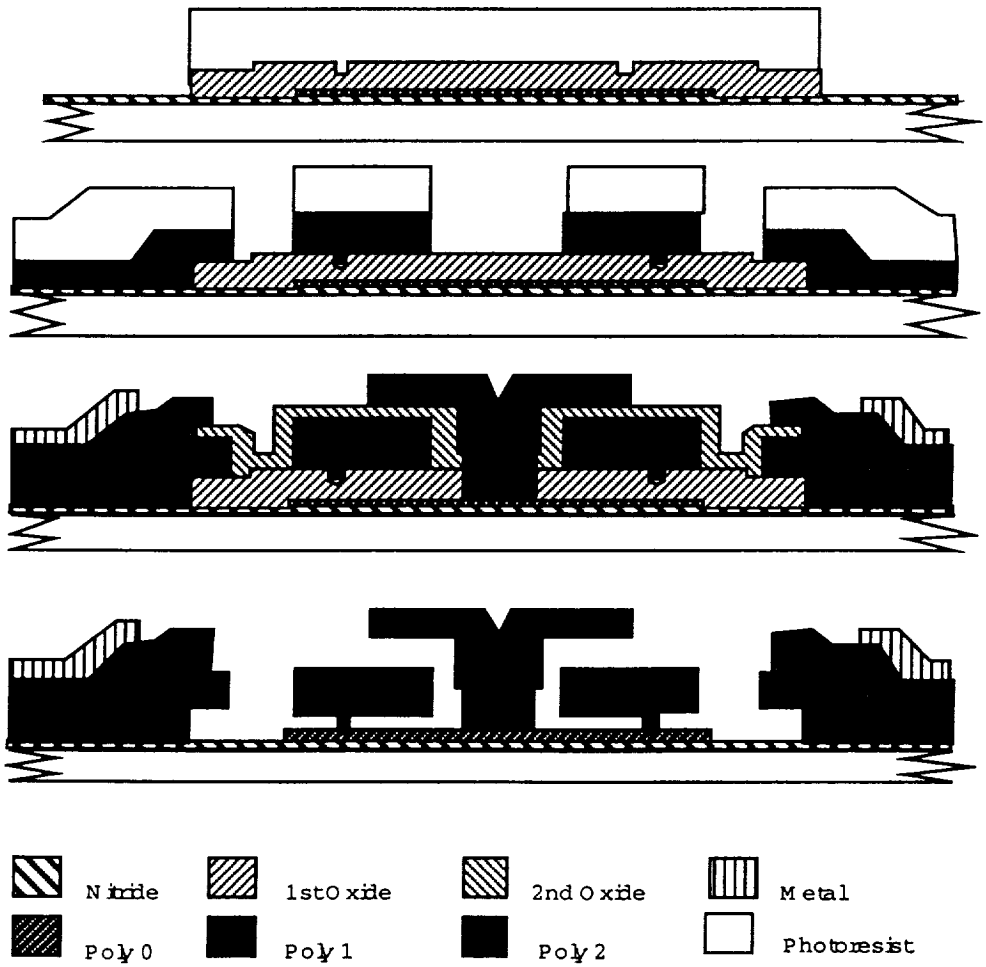


Figure 2 MUMPs 3-layer polysilicon process flow for fabricating a micromotor.

oxide and polysilicon films has been described above. The silicon nitride film is a blanket layer that serves to isolate all structures electrically from the substrate. The metal layer serves two purposes, it provides electrical contact to the polysilicon simplifying wire bonding and second, it provides a reflective surface for optical applications. Figure 1 shows a scanning electron micrograph of a salient pole micromotor[2] fabricated using the MUMPs process. Figure 2 illustrates the progression of layers used to build the micromotor.

LIGAMUMPs

LIGA, a German acronym which translates to Lithography, Electroforming and Injection Molding, was developed in Germany in the 1980's[3] and has slowly gained wide-spread interest. There are two key factors to LIGA's attractiveness–the ability to mass-replicate high aspect ratio structures out of metals, polymers and ceramics, and the ability to fabricate structures which can be assembled with a high degree of precision. To produce LIGA requires the use of an extremely energetic, highly collimated light source, which restricts its practice to those facilities with access to x-ray synchrotrons. The synchrotron generated x-rays act to expose a thick layer of PMMA (polymethylmethacrylate) through an x-ray mask of high Z material. The x-rays provide deep exposure of the PMMA as well as excellent edge acuity (eg. 0.1 μm in a 400 μm tall structure). Once exposed, the PMMA becomes soluble in certain solvents. By "developing" in special solvent mixtures, the unexposed PMMA is left behind forming structures that can be used as-is or as an electroforming template for metals. After plating the PMMA template is removed leaving behind stand-alone metal structures. These structures can then be released from the plating base or used as injection molds for mass replication. Figure 3 shows a stator housing of nickel formed using the LIGAMUMPs process performed at the University of Wisconsin.

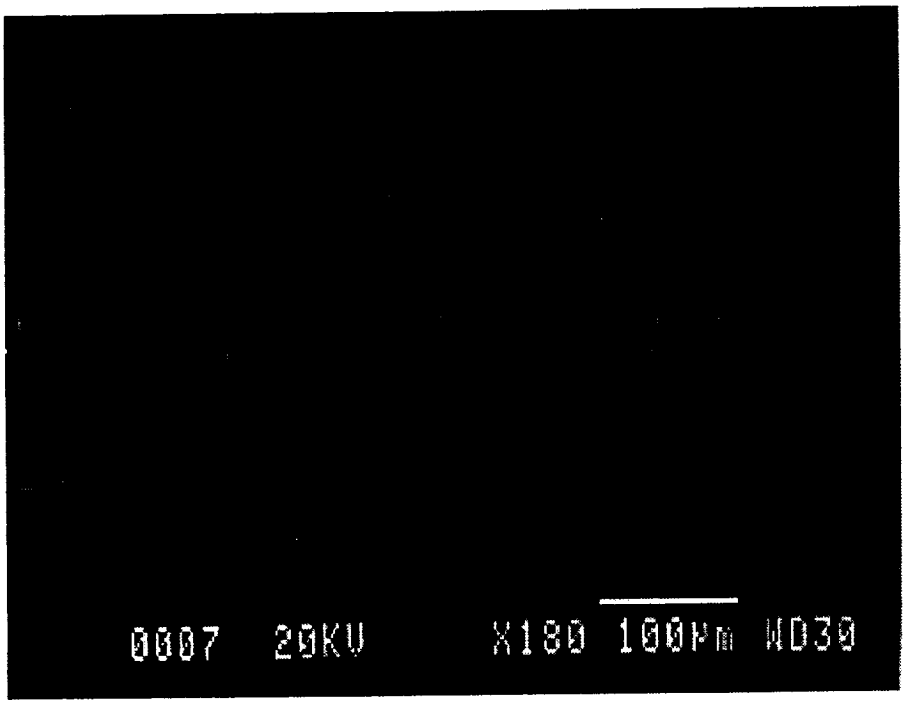


Figure 3 Stator housing fabricated using LIGAMUMPs process. Structures are electroplated nickel, 150 μm in height.

Consolidated Micromechanical Element Library

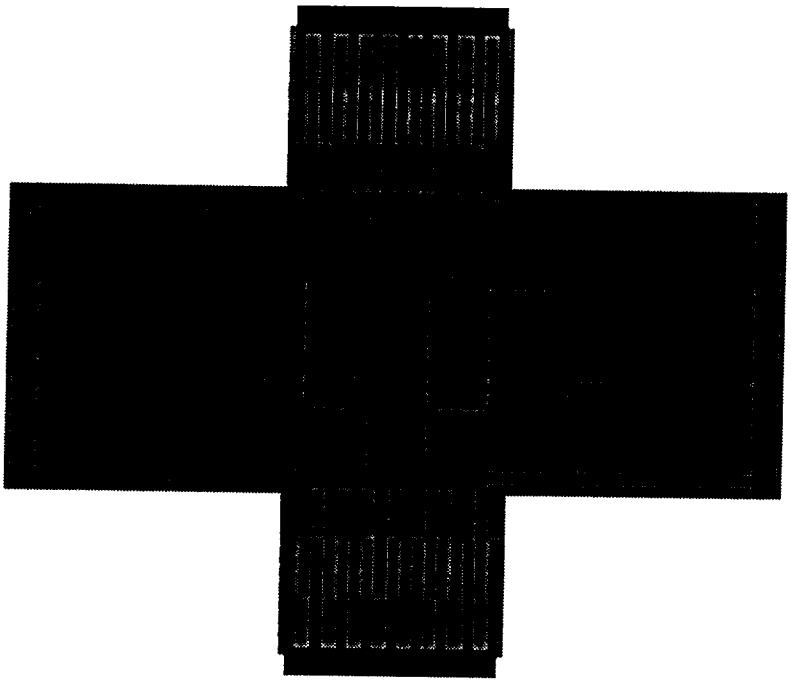
The Consolidated Micromechanical Element Library (CaMEL) is a library of MEMS cells and is similar to standard cell libraries that proliferate in VLSI design. The CaMEL library consists of two independent parts, the nonparameterized cell database and the parameterized micromechanical element library. The aim is to provide MEMS cell libraries that are useful for novice MEMS designers, as well as experienced ones. Both libraries are intended to assist the user in the design and layout of MEMS devices and it is assumed that the user will modify and customize these elements using a suitable mask layout editor.

The nonparameterized cell library is a database of MEMS designs in various process technologies contributed by different sources. It is a resource of working MEMS devices and structures. The library browser, DBRead, permits the user to peruse brief descriptions of the cells and select desired ones. The selected cells can then be retrieved from the database in either Caltech Intermediate Form (CIF) or

CALMA GDS II format. A companion program, DBSubmit, allows designers to submit MEMS designs for inclusion in the database along with the accompanying process information. Both programs are written in the Practical Extraction and Report Language, PERL. Cells currently available in the library include designs for MUMPs, UCB 2 poly, and LIGA processes.

The parameterized micromechanical element (PME) library is a set of generators that allow users to create customized versions of commonly used elements in a quick and easy manner. The PME library also provides a framework for writing cell generators. It enables the generators to be relatively process independent and allows limited cell hierarchy. Designs can be generated in CIF, GDS II, or PostScript output formats. Technology dependent design rules are read in from an environment specified technology file. The library provides various geometric primitives and a set of available generators. Various types of elements are available, including active micromechanical elements, passive micromechanical elements, test mechanical structures, and electrical elements.

The PME input is provided in an ASCII text file and defines cells by calling generators with desired parameter values and then placing instances of defined cells at chosen locations within the top level cell. Instances of defined cells can be reflected, translated, or rotated through arbitrary rotation angles.



```
/* Linear comb resonator */
PME p1,p2;
/* Create comb and suspension */
p1=lcomb1(98,12,14,60,4,3,30,
      comb);
p2=lfbs1(150,4,50,12,30,25,12,
      98,suspension);
/* Place cells for resonator */
instance(p1,'*',0,0,75);
instance(p1,'*',0,0,-75);
instance(p2,'*',0,0,0);
```

Figure 4 Example of linear comb resonator generated using the PME library.

Figure 4 is an example of a linear comb resonator generated by the PME library and its associated input file. It is generated using the first structural layer (poly 1 in the MUMPs process) using the **lcomb1** linear comb drive and **lfbs1** linear folded beam suspension. The fingers in the comb drive are 4 μm wide with an airgap of 3 μm, and beams in the suspensions are chosen to 150 μm long and 4 μm wide.

Smart MEMS: Electronics integration for MEMS

There has been a growing interest in placing electronics closer to sensors and actuators to improve their performance. Among the several approaches that have been demonstrated are hybrids and embedded electronics. The hybrid approach, in which the different chips including electronics, sensors and actuators are placed in a single package and connected by wire bonding, has long been the industry standard. This approach is flexible with few restrictions on the types of usable electronics and substrates. However, hybrids are not batch fabricated, can suffer from system performance degradation due to stray or large capacitance and also result in increased size of the integrated system[4]. A recent approach[5,6] has been to build MEMS sensors and actuators on top of underlying electronics on the same substrate. This embedded approach is suitable for batch processing and results in significant improvement of

performance. However, it involves a large number of processing steps that can increase processing complexity and reduce yield, driving up the costs.

Flip Chip MEMS

Flip chip MEMS combines the advantages of the hybrid and embedded approach. The electronics and the MEMS are batch fabricated on different substrates and are then connected using solder bumping. Flip chip has been successfully used to connect IC chips to printed circuits or substrate carriers for almost 20 years. In conventional flip chip attach, the IC chip is turned upside-down (i.e. *flip chipped*) and an array of solder bumps on the chip are joined to a matching array of solder wettable pads on the substrate. This conventional approach has been modified to facilitate the connection of the MEMS chip to the electronics chip, taking into account the mechanical or 'released' nature of MEMS chips (fig. 5).

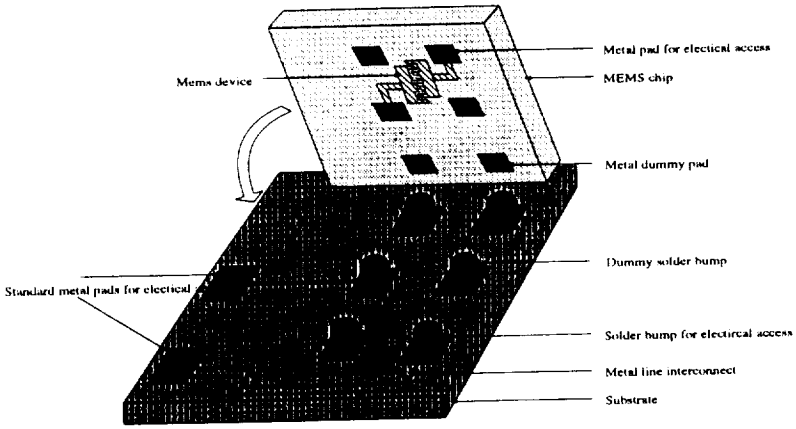


Figure 5 Schematic of flip chipped MEMS and substrate die

MCNC is a world leader in the development of flip chip process technology for MCM's, and is the site of the ARPA-supported Flip Chip Technology Center. As part of the MEMS Infrastructure program we are investigating the various issues involved in bump attaching both surface and bulk micromachined devices to different types of substrates, including silicon, quartz, Pyrex and GaAs. Methodologies for the handling of released MEMS structures and design rules are being developed to allow the MEMS community to access this advanced integration technique for the production of Smart MEMS systems.

Electromechanical Control System

The Electromechanical Control System (ECOSYS) is a program to facilitate the fabrication of MEMS systems including the electronics. A standardized IC controller with various functional blocks is currently under production. The aim is to implement a system incorporating sensing, feedback, and control by allowing users to connect blocks and attach them to a MEMS device via flip chip. A limited degree of user programmability will be provided via external RC components to tailor the response of the blocks for the application. Interconnections to the MEMS elements will be via solder bumps and the parasitics introduced by the solder bump and pad will be accounted for in the design of the blocks.

Expanded Infrastructure–MEMS TechNet

Once proof-of-concept has been established using MUMPs, users often require larger numbers of prototypes for testing and evaluation than MUMPs can provide. An expansion of the current Infrastructure program is underway that is designed to provide users more flexibility to design and fabricate unique devices and also produce larger quantities as the need arises. The added flexibility is beneficial in cases were the uniqueness of a MEMS device may necessitate custom fabrication processes that are not provided by MUMPs.

The Expanded Infrastructure, known as the MEMS Technology Network or MEMS TechNet, will consist of two parts, the Wafer Inventory and the Process Modules. The Wafer Inventory will provide the MEMS community with a variety of standard films (and combinations of films) including low stress nitride, polysilicon and phosphosilicate glass (PSG). These wafers can be purchased and subsequently

processed at MCNC or elsewhere, giving the user the greater flexibility and capacity needed by more mature programs. The Process Modules will provide users with a menu of MEMS fabrication processes and sequences such as anisotropic wet etching (KOH or EDP), deep boron diffusion, wafer bonding and laser assisted etching. The focus of the modules is to give users a wide variety of fabrication technologies thereby allowing them to fine-tune their device processing. Many of these processes will be supplied by providers other than MCNC who have considerable experience with the process. MCNC will operate as a clearinghouse for the modules thereby simplifying the complex task of locating and purchasing such services.

Acknowledgments

The authors would like to thank the fabrication facility staff at MCNC for their hard work and dedication in helping make the MUMPs program a success. We would also like to thank the MCNC Advanced Packaging Group for their support and technical advice in implementing the Smart MEMS. We would also like to thank Kaigham Gabriel, ARPA MEMS Program Manager, for his faith and advocacy in making MEMS infrastructure a reality. This work is supported under ARPA contract DABT63-93-C-0051.

References

1. "Micro-machines Hold Promise for Aerospace", Aviations Weekly & Space Technology, March 1, 1993.

2. M. Mehregany, S.D. Senturia, J.H. Lang and P. Nagarkar, "Micromotor Fabricaton," *IEEE Trans. Electron Devices*, vol. 39, pp.566-575, Apr. 1992.

3. E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Münchmeyer,, "Fabrication of Microstructures with High Aspect Ratios and Great Structural Heights by Synchrotron Radiation, Lithography, Galvanoforming and Plastic Moulding (LIGA Process)," *Microelectron Eng.* [4] 1986, pp. 35-36.

4. N. Najafi, K. D. Wise, R. Merchant, and J. W. Schwank, "An Integrated Multi-Element Ultra-Thin-Film Gas Analyzer", Technical digest, IEEE Solid-State Sensor and Actuator Workshop, (Hilton Head, SC 1992), pp. 19-22.

5. T. Core, R. S. Pyne, D. Quinn, S. Sherman, and W. K. Wong, "Integrated, Complete, Affordable Accelerometer for Air-Bag Applications", Proceedings of Sensors Expo, Chicago IL, pp. 204B-1 (1991).

6. W. Yun, R. T. Howe, and P. R. Gray, "Surface Micromachined, Digitally force-balanced Accelerometer with Integrated Detection Circuitry", Technical digest, IEEE Solid-State Sensor and Actuator Workshop, (Hilton Head, SC 1992), pp. 126-131.

METHOD OF MAKING LARGE AREA NANOSTRUCTURES

Alvin M. Marks
Advance Research Development, Incorporated

71159
030706

ABSTRACT

Present technology appears to be limited to incremental improvements in the slow speed formation of nanostructures on small areas.

20 P.

A new method is described which enables the high speed (0.1 m^2 /sec) formation of nanostructures on large area surfaces (1 m^2). The method uses a "Supersebter", an acronym for super sub-micron electron beam writer.

The Supersebter utilizes a large area multi-electrode (Spindt type emitter source) to produce multiple electron beams simultaneously scanned to form a pattern on a surface in an electron beam writer. Spindt electrodes are well known and available commercially.

Proposed is a 100,000 x 100,000 array of electron point sources, demagnified in a long electron beam writer to simultaneously produce 10 billion nanopatterns on a 1 m^2 surface by multi-electron beam impact on a 1 cm^2 surface of an insulating material. The surface is coated with a monomolecular or monoatomic layer. The monomolecular layer is altered when the electron beam impacts the surface to form a +, -, or 0 charge pattern of adjacent charges. A multiple charge pattern is thus produced on a large area. The 1 cm^2 charge pattern is then stepped over the surface to form a 1 m^2 nanopattern in 10 seconds.

Metal is deposited at atmospheric pressure and temperature onto the negatively charged pattern area from an electroless coating solution. The pattern is then rinsed dried and protected in any manner. Successful implementation of this process will result in major advances in light/electric power conversion, HDTV, lasers, computers and telecommunications.

BACKGROUND

1. Field of the Use

This describes a novel monoatomic or monomolecular resist for use with a beam writer for the production of high resolution submicron circuit patterns on an insulating substrate.

2. Description of the Prior Art

The State of the Art of "Nanometer-Scale Fabrication" has been given with an excellent bibliography, as of 1982 [1]; wherein, it is stated on p. 3:

"...electron, ion and X-ray exposure....limitations of the resist...not those of the exposure system...set the ultimate limit on...resolution", and: "The most commonly used resist for high resolution (<100 nm or 1000Å) is (PMMA) polymerthylmethacrylate. A resolution of <50 nm or 500Å may be obtained with other resists...not well studied. "; and: "exposure of PMMA with a high intensity 50 KV field-emission electron beam source with a 20 nm. beam of 10^{-7} amps...takes 1 day (86,400 sec.) to expose a dense pattern on a 4" square (100 cm^2 or 10^{-2} m^2), with additional time for stage motion and alignment."

This is a speed of about 10^{-7} m^2/sec; and a resolution of only 500Å.

The use of reactive ion etching to produce localized probes 1000 Å apart is reported [2] but this resolution is also small enough, and there is no increase in speed.

For the manufacture of Lepcon™, Elcon™ and such devices this speed is too small; and the resolution not small enough. A speed of about 0.1 m^2sec and a resolution of 10Å is required, not obtainable with these prior art devices.

A 10 Å resolution is reported [3]:

> "Using a 1/2nm (5Å) diameter beam of 100 keV electron, we have etched lines, holes and patterns in NaC1 crystals at the 2 nm. (20Å) scale size. Troughs about 1.5 nm. wide on 4.5 nm centers and 2 nm dia holes have been etched completely through NaC1 crystals more the 30 nm thick." and "The scanning transmission electron microscope (VG Microscopes, Ltd., Model HB5) in operation in the National Research and Resource Facility for Submicron Structures at Cornell University can produce up to 1 nÅ of 100 kV electron in a beam dia as small as 1/2nm (5Å). This beam current density of 1/2 x 10^6 A/cm^2 means that it takes only 10μs to deposit a dose of 5 coulombs/cm^2 in the sample." and "Two types of materials...alkali halides and aliphatic amino acids...can easily be vacuum sublimated or evaporated as uniform thin films...readily vaporized by electron beams. Using 100 kV electrons a dose of about 10^3 C/cm^2 is sufficient to etch through 30 nm of L-glycine, while a dose of 10^2C/cm^2 is needed to etch through a similar thickness of NaC1."

In the latter reference the resolution is satisfactory but the speed is too slow. Recently (1986) there has been a report on a new X-Ray lithography device [4]. This article stated:

> Submicron lithography "using storage ring XRay sources may be closer...volume production...1990'2. A compact synchroton storage ring will be mated with a vertical stepper...will produce 12Å wavelength at 630 MeV energy level. When mated to a storage ring, the XRS should have a resolution of 0.2μm (2000Å)...alignment accuracy to within 0.1μm (1000Å). The stepper will expose wafers up to 8 in. (0.2 m) dia.

A Field-Emission Scanning Transmission Microscope (STEM) has been described [5.1]. A field emitter is employed to produce an emission area having a diameter of 30-300Å. One magnetic lens with a short focal length and low spherical aberration, is used to demagnify this source to a resolution of 2 to 5Å on the specimen surface. The field emission gun and lens is mounted in an ultrahigh vacuum vessel that operates, at 10^{-8} to 10^{-7} Pa. When a short focal length lens is utilized to keep the system compact, aberration may be compensated by a "stigmator".

In these Prior Art Devices the speed is too slow for the rapid production of devices for the Lepcon™ - Elcon™ devices.

DEFINITIONS

ELCON™ A trademark for a submicron array of dipole antennae on a sheet which emits light photon power by transducing the equivalent input electrical power from a direct electric current input; thus directly converting input electric power per unit area to output light, power per unit area (watts/m^2). The light is emitted from the sheet as a parallel laser beam of a particular color or frequency. With its electric vector parallel to the long axis of the antenna. [39]

LEPCON™ A trademark for a submicron array of dipole antennae capable of receiving and transducing light photon energy $\varepsilon = hv$ into its equivalent electron energy $\varepsilon = Ve$ as direct current; thus directly converting light power per unit area (photons/sec-m^2) to electric power per unit area (watts/m^2). Randomly polarized light photons are resolved. Two orthogonal antennae arrays totally absorb and transduce the randomly oriented incident photons. In bright sunlight the electric power output is about 500 watts at 80% efficiency. [37]

SUPERSEBTER™ An acronym for Super Submicron Electron Beam Writer for the rapid fabrication of submicron arrays for LEPCON™ or ELCON™ devices, or other submicron circuits.

Strip: A metal electrical conductor coated or deposited on an insulating surface in the shape of a long narrow parallelpiped.

OBJECTS

Objects of this work are to provide a process and an apparatus for the manufacture of submicron circuits which have:

1. a high speed (about 0.1 m^2/s)
2. a large area (1 m^2)
3. a high resolution (less than about 5Å)
4. a low cost (less than \$250.00/m^2-1986 prices).

DISCUSSION

The present device may be employed for the rapid manufacture of submicron circuits as hereinabove set forth.

These devices comprise single crystal metal strips of Copper, Aluminum, or the like; in which, an energetic electron provided by direct conversion from photon energy travels freely in the metal for distances of about 10,000Å without collision with an impurity atom; and, hence without energy loss.

The strips may be for example 600Å long, separated by a "tunnel-junction" gap of 28-35Å, the width of the strip may be, for example, 30Å.

In the formation of this structure a single scan with a 30Å dia. electron beam may be used. The electron beam is preferably shaped with a square or rectangular section to provide a constant gap of about 30Å between successive in-line strips.

1. A small diameter image is formed from the electron emitter array by a demagnifying lens.
2. One or two long focus electron lenses of the magnetic or electrostatic type are used resulting in negligible aberration.
3. A large aperture lens may be used; to 30mm.
4. The focal length of the lens is 2.5m to 20m. compared to about 2.5cm in a standard SEM.
5. A plurality of images of the electron emitter array is simultaneously imaged onto the work surface.
6. Writing speed is increased by the simultaneous scanning with a plurality of electron beams. For example: 2 x 10^9 electron beams are scanned simultaneously to imprint the same number of identical patterns.
7. The pattern is imprinted by an electron beam impinging on a surface coated with a monoatomic or monomolecular layer, which may comprise an electric double layer. The electron beam breaks the chemical bonds, changes the chemical or electrical characteristics, or ablates the layer. Prior art masking layers were usually about 300Å thick. The layer used herein is only 1/2% to 10% of the thickness of prior art coatings. Consequently the present method is more efficient than prior art methods, requiring considerably less electron beam energy per unit area, and will accurately imprint nanostructures.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a cross section through a plate surface having no free electric surface charges.

FIG. 2 shows a cross section through a glass plate having a surface electrical double layer, with the negative charges on the outside of the layer, with no pattern impressed.

FIG. 3 shows a cross section through a glass plate having a surface electrical double layer, with the positive charges on the outside of the layer, with no pattern impressed.

FIG. 4 shows a cross section through a glass plate having a surface electrical double layer, with a pattern impressed by a reversal of the sign of the charge layer.

FIG. 5 shows a cross section through a glass plate having a surface electrical double layer, with the negative charges on the outside of the layer, a pattern being impressed by an electron beam onto adjacent areas of the surface which are thereby charged or not charged.

FIG. 6 shows a cross section through a glass plate having a surface electrical double layer, with the positive charges on the outside of the layer, a pattern being impressed by an electron beam onto adjacent areas of the surface which are thereby charged or not charged.

FIG. 7 shows a cross section through the glass plate and a back electrode, located in a Supersebter during pattern inscription with an incident electron beam.

FIG. 8 diagrammatically shows the generalized process steps of submicron pattern deposition along a production line on the OX axis of the Supersebter™, according to Process A.

FIG. 9 diagrammatically shows the generalized process of submicron metal pattern deposition along a production line on the OX axis of the Supersebter™, according to Process B.

FIG. 10 diagrammatically shows an Ion Beam Coater.

DESCRIPTION OF THE PRODUCTION PROCESS

Generalized production processes for the manufacture of submicron circuits on a plurality of substrate sheets are shown diagrammatically in FIGS.8 and 9. The main or first vacuum tube 24 of the Supersebter™ is along the OZ axis. One or more vacuum tubes 53, 54 and atmospheric pressure processing stations 51, 52 are located along the OX axis. The Production line 61, driven by the stepping motor 62 enters from the left at Station 1, and leaves at 59, Station n, to finished sheet inventory, circulating continuously. In the two production processes described herein, Process A and Process B, substrate sheets, usually glass plates, are supplied from stock 50 to a production line 61. Both Processes include the step of electron beam writing a pattern on the surface of the sheet using the Supersebter™ electron beam writer described herein.

In the generalized process, Station x refers to a processing step at position x. Referring to FIG. 8, the substrate sheet is loaded from stock 50 onto the production line 61 at Station 1. Several processes, hereinafter described, may be employed at atmospheric pressure from Stations 2 to h. The sheet enters the second tube 53 through the first air locks at Stations h + 1, h + 2, h + 3; respectively, the first low, medium and high vacuum airlocks 55. Vacuum treatment steps are located from Stations h + 4 to Station j - 1. The electron beam writer step, located at Station j, creates a pattern on the surface of the sheet. This pattern may include a gap for example 15-30Å, used for an asymmetric tunnel junction. Further treatment steps may occur at Stations j + 1 to k - 3. Stations k - 2 are second airlocks 56, which restore the sheets to atmospheric pressure. In Process B Stations k + 1 to m - 1 at 52 are at atmospheric pressure for several wet process steps such as the application of electroless metal sensitizing and metal deposition solutions, rinsing and drying.

Next, the sheets may enter airlocks 57 at m, m + 1 and m + 2. At Station m + 3 the deposited metal may be crystallized to single crystal areas. (As shown in Tables I, II, III). Then, using the ion-coating device shown in FIG. 10, the metal pattern is coated with high and low work function materials respectively at opposite faces of a gap in the deposited metal pattern. At Station m + 5 the entire pattern is coated with an insulator coating as described in Section 08.7. After passing through the fourth airlocks 58 at Stations n - 3, n - 2, n - 1, n the sheets pass from the third tube into the atmosphere and are removed from the production line to finished sheet inventory 60. The production line 61 circulates back to the start, and the process is intermittently continuous.

Two basic Processes A and B are described below:

FIG. 1

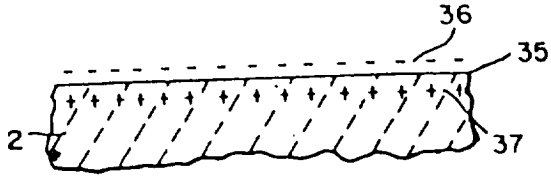


FIG. 2

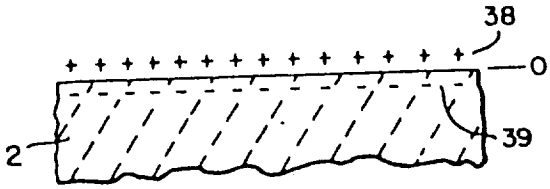


FIG. 3

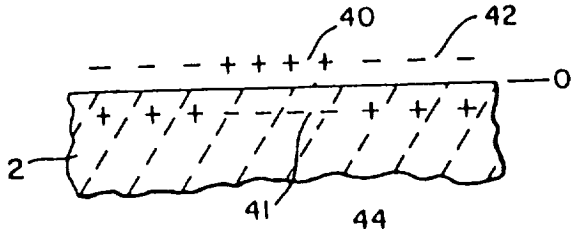


FIG. 4

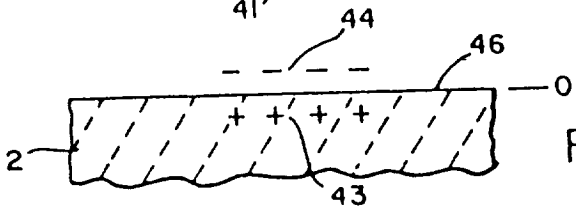


FIG. 5

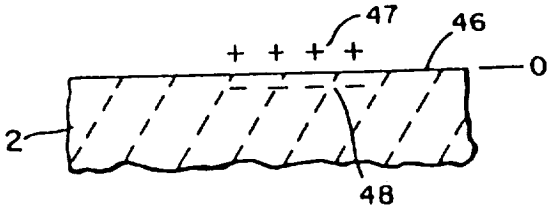


FIG. 6

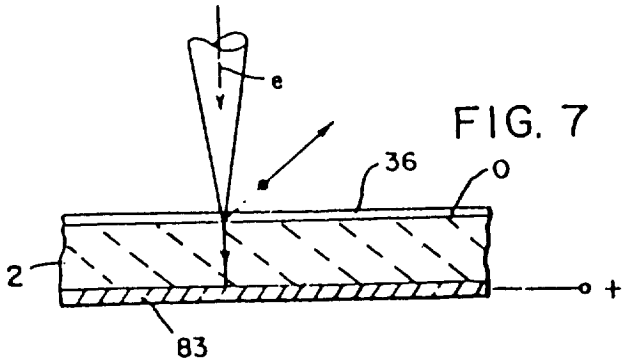
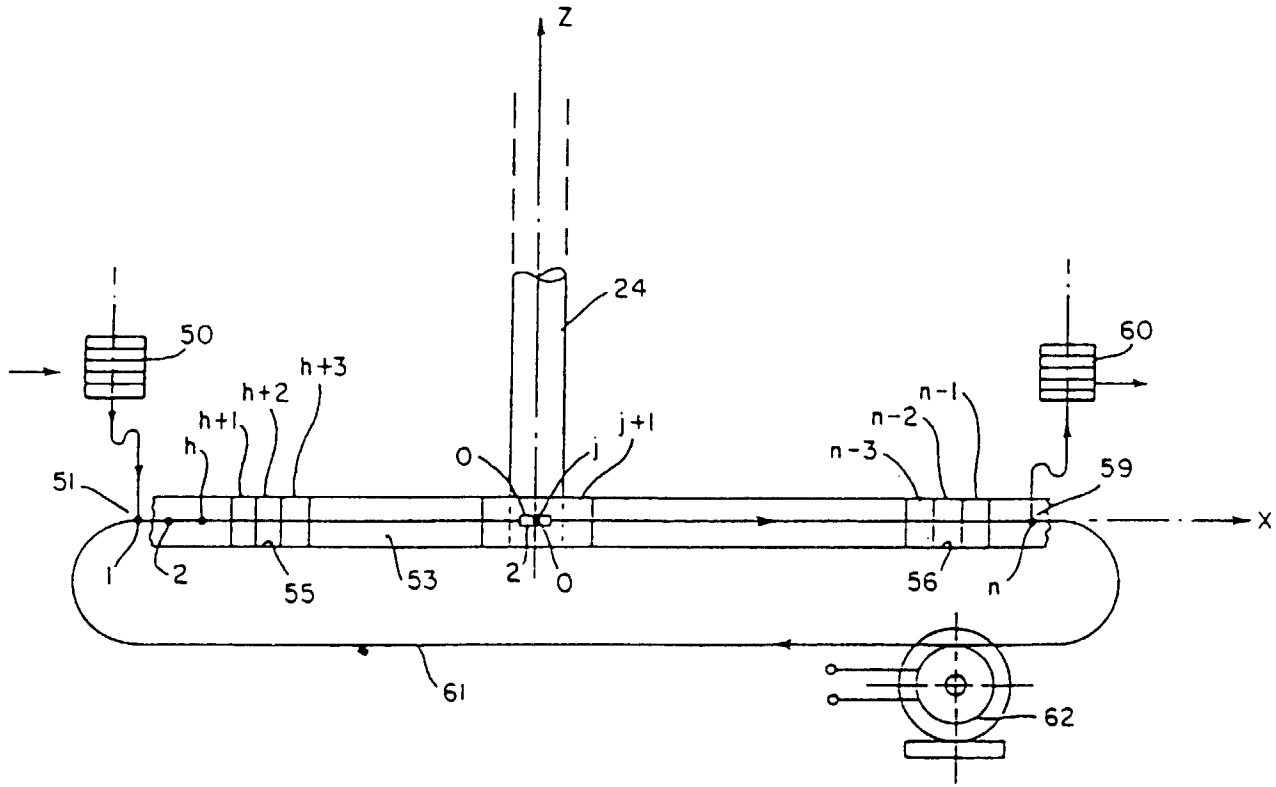
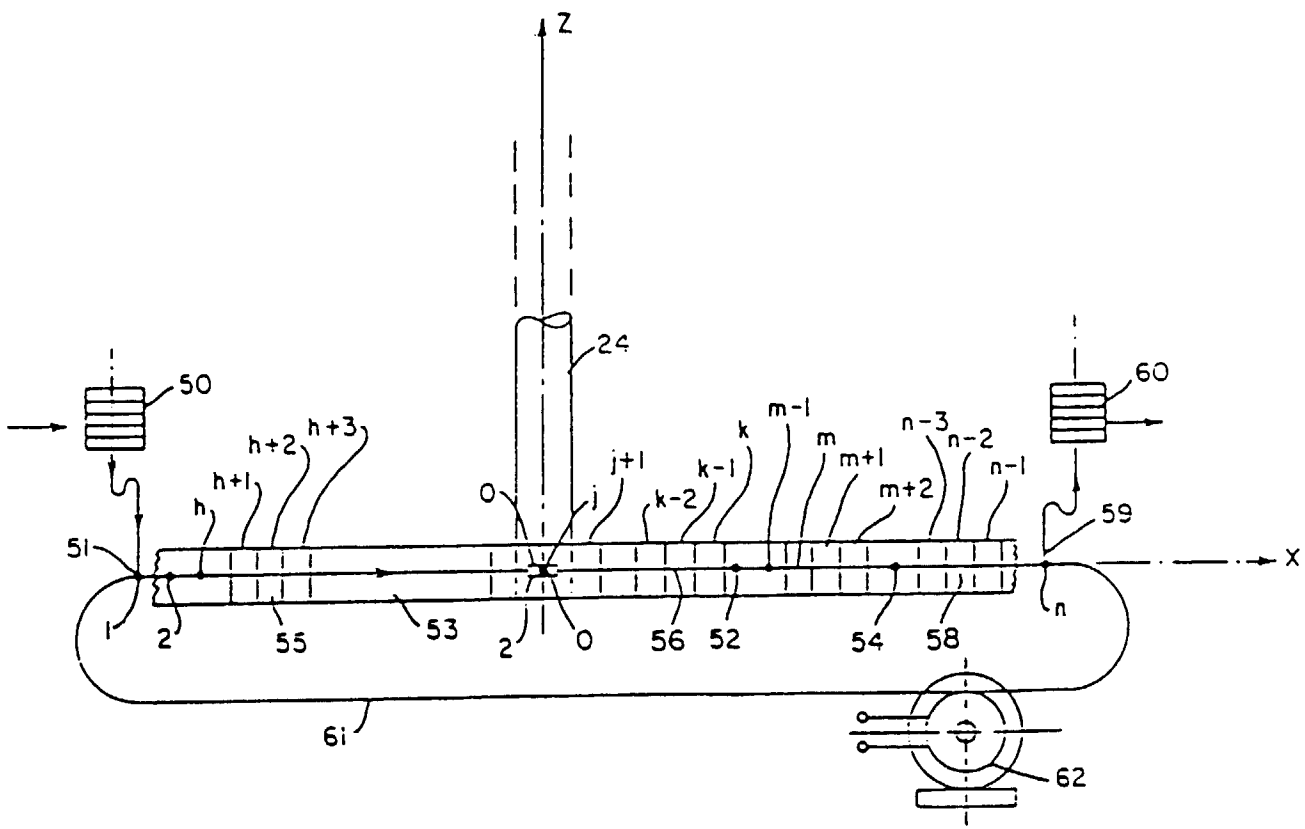


FIG. 7



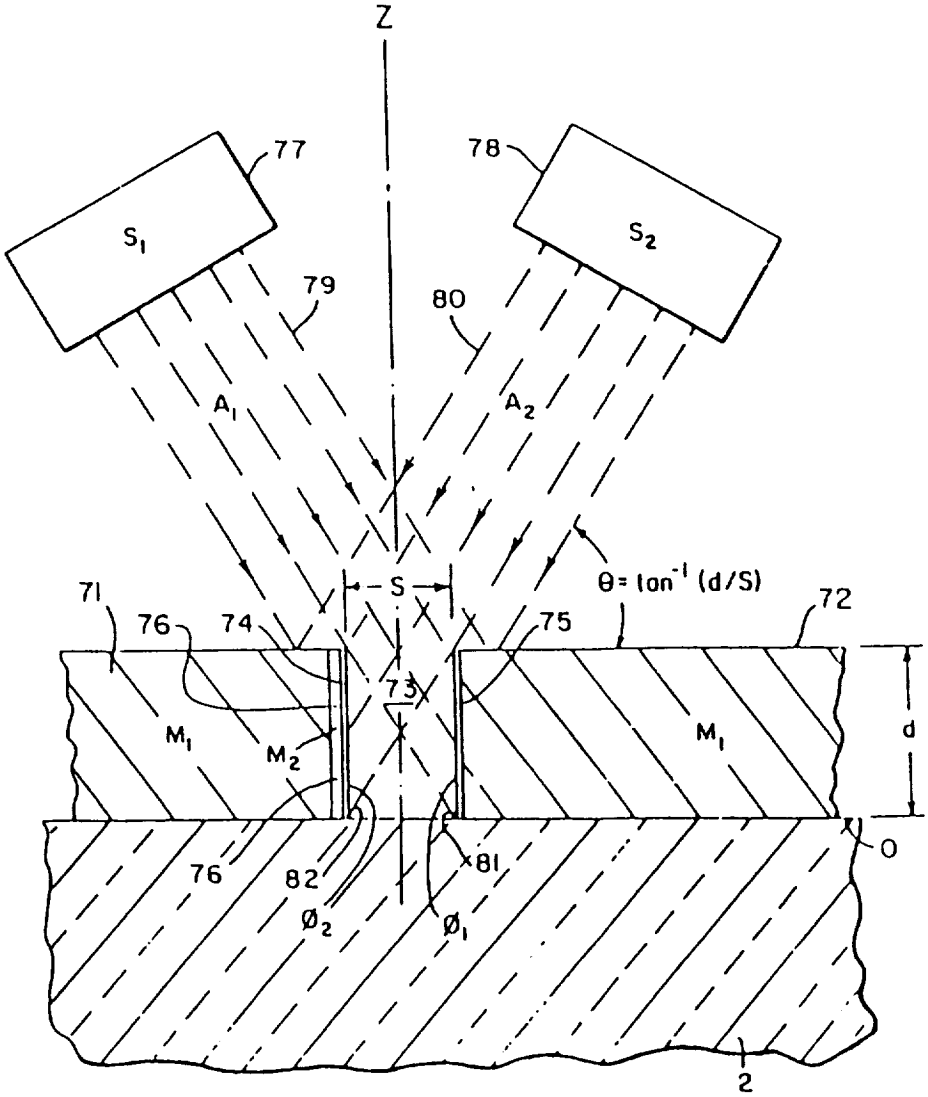
PROCESS A

FIG. 8



PROCESS B
FIG. 9

FIG. 10
ION COATER



PROCESS A

FIG. 8 shows a device employing Process A: the first vacuum tube 24 of the electron beam writer joins the production-line second vacuum tube 52 in a T section. Process A takes place entirely within the second vacuum tube 53. The substrate sheets enter at Station 2, being processed before and after the electron beam writing step at Station j, and leaving at Station n as finished sheets. An example of the processing steps which may occur at the various Stations follows:

1. The sheets enter the airlocks at stations h + 1 to h - 3.
2. At Station h + 4 the sheets are heated to 700° C. and Argon-plasma cleaned to de-gas their surface.
3. The surface chemistry may be changed by ions implanted into the surface by known means.
4. At Station j the pattern is imprinted onto the surface of the glass panel by the electron beam. The electron beam "sensitizes" the surface atoms by directly altering their electrical and/or chemical properties.
5. The sensitized surface is exposed to positive ions such as Sn^{++}, Pd^{++}. The positively charged ions are attracted to negatively charged pattern areas where they deposit on and, adhere to the surface; but are repelled by positively charged areas. The positive ions do not adhere to surface areas having no charge, but may reflect elastically.
6. Tile surface is exposed to a metal vapor such as Cu or Al which may comprise positive ions of these metals. These ions are attracted to the previously metallized areas by an induced negative image charge. The thickness of the metal deposit is controlled by the metal vapor concentration and exposure time.
7. The patterns are formed with a gap S = 15-30Å at locations where an asymmetric tunnel junction is to be placed. The pattern includes gaps 73, 15 - 30Å wide in the metal deposit.
8. The deposited metal is crystallized to single crystal by locally heating with an electron beam or laser, and then cooling to ambient temperature.
9. At station m + 3 the gaps 73 in the metal pattern are ion coated in an ion coating device such as shown in FIG. 10. This provides the facing surfaces 81 and 82 of the gap 73 with two materials having different work functions ϕ_1/ε and ϕ_2/ε.
10. The entire surface is then coated with an insulator layer; for example, silicon dioxide, titanium dioxide, silicon nitride, and the like. The insulating material is chosen from one having a dielectric constants, such that the effective work functions are ϕ_1/ε and ϕ_2/ε. Conventional vapor coating techniques and apparatus may be used.
11. The sheet is again removed to atmospheric pressure through the airlocks 58, exiting at 56, Station n, as a finished product.

PROCESS B

FIG. 9 shows a device utilizing Process B. Process B takes place in two vacuum chambers 53 and 54, with a space 55 between them for stations for process steps at atmospheric pressure.

1. The process steps are the same as in Process A, to and including the electron beam writing step at Station j.
2. The sheet is removed from the second vacuum tube 53 into the atmosphere through the airlocks 56, and wet-processed with solutions which provide an electroless deposit of metal on the sensitized pattern areas [8].1].
3. The sheet is then dried, and passed into a third vacuum tube 54 through the airlocks at Stations m to m + 2.
4. From Stations m + 2 to n - 1, the steps are the same as in Process A, from 0.18 to 0.21 inclusive.

SURFACE CHEMISTRY OF GLASS

There is an electric double layer surface charge naturally existing on the surface of glass, which varies with glass composition [14-17]. The silica network is the most important in determining the surface charge. Most silicas have a population of silanol SiOH groups on the surfaces which may disassociate:

$$(\equiv SiOH) + OH \;<---->\; (\equiv Si - O^{-}) + H_2O$$

This forms negatively charged $\equiv Si - O^{-}$ groups in water, the greater the pH, the greater the number of $\equiv Si - O^{-}$ groups formed. The pH for which $\zeta = O$ is similar to that for other forms of silica pH = 2 to 2.5).

The charge on the surface of the glass depends on the physical and chemical treatments to which it is subjected, as discussed in connection with the following examples:

.1 At Station 2 the glass surface is polished with Cerium Oxide.
.2 At station 3 the surface is rinsed with distilled water.
.3 At station 4 the sheet is dried at about 60° C. for about 10 minutes.

This process results in hydroxyl groups (–OH) attached to the glass surface (=Si–OH) making the surface electronegative, and capable of reacting with positive ions in the vapor or liquid phases; such as Sn^{++}.

An alternate method is:

.4 Expose the surface of glass to fuming sulphuric acid H_2SO_4 for a few minutes [25, 26].

This process attaches a sulphonic group (SO_3=) to the glass forming an electronegative surface charge on the surface (=Si=$S0_3$–).

In a similar manner the glass surface may be made electropositive:

.5 At Station 4 the glass surface is exposed to a solution or vapor containing bifunctional electropositive groups such as:($-NH_2^{+}$) which results in a glass surface with a positive charge having the composition $\equiv Si-O-NH_2^{+}$.

A positive surface charge tends to repel positively charged ions and attract negatively charged ions from the vapor or solution. A third method is to render the glass surface charge neutral by utilizing reactive monofunctional atoms; such as a halogen, fluorine –F.

An H atom in 0.3 above may be removed by an electron beam to form patterns on the glass surface, as shown in FIGS. 5 and 6, forming positively and negatively charged patterned areas. FIG. 3 shows a uniformly charged positive layer 38, and a negatively charged lower layer 39, on the glass surface. FIG. 4 shows the same surface after a pattern has been created by the electron beam. The pattern appears as adjacent areas of charge reversal, 40 and 42; with the charges reversed at 41, 43 in the lower layer.

The surface chemistry of glass may be altered by for example heating.

Glass surfaces prepared in contact with water as in 0.1 to 0.3 above have Silanol groups; ≡ SiOH.

Such surfaces heated to >400° C. give off water, adjacent group forming siloxane groups:

$$\equiv SiOH$$

Such surfaces heated to >400° C. give off water adjacent group forming siloxane groups:

$$2(\equiv Si - OH) \;<- - ->\; (\equiv Si - O - Si \equiv) + H_2O$$

At 800° C. only a few silanol groups remain. At 1200° C. no silanol groups remain. On cooling surface rehydration is slow. Aging in water restores the hydrated state.

An initial heating stage at Station 3 in which the temperature is maintained for a time at sufficiently high temperature may be utilized to fully dehydrate the surface.

METAL PATTERN DEPOSITION

The pattern is inscribed by an electron beam writer such as a Supersebter. Metal is deposited on the pattern using Process A or Process B described above.

In Process A, metal pattern deposition occurs from the ionic vapor in a vacuum onto charged surface areas that attract the metal ions and cause them to deposit. For example, as shown in FIG. 5, the electronegative portions of the pattern attract Sn^{++} ions which sensitize only those areas. These ions may be produced by known methods [29-31 include.]. A second ion source deposits Copper, for example, Cu^{++} from an ionic vapor, onto the sensitized areas of the pattern.

In Process B, metal pattern deposition occurs at atmospheric pressure from solution, for example by electroless coating of Copper or other metals using known methods [18-24].

SINGLE CRYSTAL METAL GROWTH

The pattern may comprise a deposit of long, narrow thin metal strips, for example having dimensions 1000Å x 100Å x 50Å. The metal is preferably Copper, Aluminum or other metal having a long mean free path for energetic electrons in a single crystal of the metal. A single metal crystal is essential to the best functioning of the device. At these dimensions metal particles may form single crystals spontaneously; but especially when heated and cooled. An electron beam may be employed to locally heat a plurality of the metal strips to induce their conversion to single crystals [27].

ION DEPOSITION

An Ion Beam Device for producing an asymmetric Tunnel Junction is shown in FIG. 10. The cross section of an asymmetric tunnel junction on a substrate surface O is magnified to a scale of 10,000,000 to 1. Thus, in the drawing, the gap S = 20Å measures 2 cm.

The asymmetric tunnel junction comprises a pattern of deposited metal layers 71, 72 of a metal M_1 of substrate surface O. There is a gap 72 between adjacent faces 74, 75 of the metal layers; for example, 20 to 100Å. Two or more ion sources 77 and 78 are provided to coat the gap faces with various materials 76, 81, 82 having different work functions ϕ_1 to ϕ_2. One of the metal faces of 71 may first be coated by the ion source 78 with a thin metal layer 76 of a second metal M_2 a few atoms thick; then coated with said materials.

INSULATOR COATINGS

As a last step in the vacuum tube at Station n - 1, insulator coatings are applied to the sheet, utilizing standard vapor coating or sputtering techniques and apparatus.

The insulator coating material is selected for its dielectric constant, ε, to produce effective work functions ϕ_1/ε, and ϕ_2/ε, as described in the following section entitled Work Function.

At station n, electric connections may be made and the surface of the sheet further protected by lamination to a second sheet of glass using known methods.

FIGS. 1 to 3 show various continuous surface layers of noncharge and electric charge on which the pattern may be inscribed by the electron beam writer.

FIG. 1 shows a noncharged surface 35 comprising, for example, siloxane groups with no exposed reactive chemical radical. Such condition may be reached by heating the glass.

FIG. 2 shows a glass surface having an electronegative charge layer 36 over an electropositive layer 37, forming a monomolecular electric double layer. The layer 36 may comprise, for example, sulphonic radicals, previously described herein.

FIG. 3 shows a glass surface with an electropositive surface layer 38, over an electronegative layer 39, forming a monomolecular electric double layer on the surfaced. The electropositive radical may be, for example an amino radical, previously described herein.

FIG. 4 shows a charge pattern inscribed by the electron beam, for example, in the charge structures shown in FIG. 3. In this case, the electron beam neutralizes an electropositive area and adds negative charge to reverse the charge pattern.

FIG. 5 shows a pattern inscribed by the electron beam which comprises adjacent negative charge - noncharge areas 44, 45 respectively, produced, for example by the electron beam breaking chemical bonds of the noncharged siloxane chains on the glass surface, exposing the negative free bond of an oxygen atom.

FIG. 6 shows a noncharge - positive pattern 46, 67 respectively, produced by an electron beam on the electropositive surface layer 38 of FIG. 2. In this case the amino radicals, are neutralized or ablated and free oxygen bonds and form noncharged surface areas of siloxane.

FIG. 7 shows a cross section through a substrate sheet O mounted on a base electrode 83 in an electron beam writer of this invention. The base electrode 83 may be connected to a positive voltage source to attract and discharge electrons driven through the sheet by the high velocity electrons in the beam.

Reverse current through the junction is constant, being limited by the junction structure and electric voltage. The forward current in the quantum regime is limited only by the input rate of energy quanta. Hence the forward/reverse current ratio of the Femto Diode may be very large. [37]

The tunnel junction used in the Femto Diode of this invention comprises an asymmetric Metal 1 - Insulator - Metal - 2 configuration. The first metal and its insulating interface has a work function ϕ_1 in the range of 1.1 to 1.9 eV; and the second metal and its insulating interface has a work function ϕ_2 in a range from 1.8 to 3.2 eV; A result of the analysis is that a maximum forward/reverse current occurs when $\phi_1/\phi_2 \approx 0.6$. The insulating barrier has a thickness which depends on the selected current density; for example from 28 to 38Å for a current density in the range 0.1 to 10 amps/cm^2. The dimensions of the facing metal surfaces are submicron, < 100Å × 100Å.

The tunnel junction used in the Femto Diode facilitates the tunneling transmission of an electron in the forward direction through the insulating barrier; and impedes the tunneling transmission of an electron through the insulating barrier in the reverse direction. The total absorption of a light photon accelerates a single electron to velocity determined by the energy of the photon.

The tunnel junction is based upon particular values of the work functions ϕ_1 and ϕ_2 of metal 1 and 2, and their insulator interfaces, respectively; the barrier thickness s; and the cross-section A, for which the forward and reverse tunnel currents have a maximum ratio of about 14; and in which forward average current through the diode area of 50Å x 50Å is about 2.2 x 10^{-13} amps; and for which the average forward current density is about 0.88 amps/cm^2.

The Femto Diode has an efficiency of about 80%, and is useful for many applications, particularly in light/electric power converters.

WORK FUNCTION

The work function of a metal is defined as the difference between the electric potential of an electron outside the surface (-eV) and the electron potential of an electron inside the same metal.

$$\phi = -eV - \mu$$

The work function ϕ is also the energy difference separating the top of the valence band (the Fermi energy) from the bottom of the conduction band at the surface of the metal.

The work function of a metal can be changed by the adsorption of one or more monolayers of positive or negative ions at the metal surface to change the electric potential distribution. The change in the work function $\Delta\phi$ depends upon the crystal orientation of the metal surface, the chemical structure of the adsorbate ions, and the number of monolayers. The work function can increase or decrease depending on the nature of the adsorbate. The change in the work function also depends on the order in which the interface ions are deposited.

The dielectric constant ε of the insulator layer changes the work functions ϕ_1 and ϕ_2 on the adjacent metal faces to (ϕ_1/ε) and (ϕ_2/ε), respectively; whereby, the work function (ϕ_2/ε) is adjusted to be approximately equal to the electron and photon energy; that is:

$$(\phi_2/\varepsilon) \approx eV$$

electron energy = hv photon energy enabling electron tunneling and conversion of photon energy to electrical energy to occur at the corresponding wave-length.

Table I shows experimental data for the decrease in work function ϕ on various metal substrates for various absorbates.

TABLE I

Decrease of Work Functions for Metals on Adsorbates

Experimental Data

METAL	SYMBOL	CRYSTAL FACE	WORK FUNCTION	ADSORBATE		RESULTING WORK FUNCTION	DECREASE IN WORK FUNCTION
			Φ eV	Material	Formula/ Order of Deposition	Φ_1 eV	$-\Delta\Phi$ eV
Aluminum	Al		4.28	Aluminum Oxide	Al_2O_3 on Al	1.64	2.64
	Al		4.28		Al on Al_2O_3	2.40	1.88
Iridium	Ir		5.27	Barium Oxide	BaO on Ir	1.4	3.87
Nickel	Ni	100	5.15	Sodium	Na on Ni	2.15	3.00
Nickel	Ni	112	5.15	Oxygen	O_2 on Ni	4.15	1.0

For specific wavelength μ, expressed eV, the work function $e\phi_2$ = eV; or ϕ_2 = V. A peak ratio of (j_2/j_1) occurs across the junction when $\phi_1/\phi_2 = 0.6$. Table II illustrates this relationship for 3 wavelengths.

TABLE II

Wavelength versus work function $\Phi_1\Phi_2$

COLOR	WAVELENGTH	$V = \Phi_2$	Φ_1
	Å	eV	$0.6\Phi_2$
RED	7000	1.77	1.06
GREEN	6300	1.97	1.18
BLUE	4000	3.10	1.86

The following Table III illustrates the method of selecting metals to match the work functions listed in Table III for specific wavelengths.

TABLE III

METHOD OF SELECTING INTERFACE MATERIALS

2-1 Work Functions Required $\phi_1 = 1.06$			
ADSORBATE/ METAL	$-\Delta\Phi$ OBSERVED	REQUIRED WORK FUNCTION OF METAL $\Delta\Phi - \Phi_1 = \Phi_{M1}$	METAL SELECTED FROM TABLE OF WORK FUNCTIONS F_{M1}
BaO/Ir	3.87	4.93	Mo(111),Be,Co,Ni(110)
Na/Ni	3.00	4.06	Al(110),Hf,In,Mn,Z
Al_2O_3/Al	2.64	3.70	Mg,U
Al/Al_2O_3	1.88	2.94	Tb,Y,Li,Gd

2-1 Work Functions Required $\phi_2 = 1.77$			
ADSORBATE/ METAL	$-\Delta\Phi$ OBSERVED	REQUIRED WORK FUNCTION OF METAL $\Delta\Phi - \Phi_2 = \Phi_{M2}$	METAL SELECTED FROM TABLE OF WORK FUNCTIONS F_{M2}
BaO/Ir	3.8~	5.64	Pt
Na/Ni	3.00	4.77	Ag(111),Fe(111),M
Al_2O_3/Al	2.64	4.41	Sn,Ta,W,Zr
Al/Al_2O_3	1.88	65	Sc,Mg

Tables I, II and III illustrate selection principles which may be utilized to obtain the required values of ϕ_2 and ϕ_1 for a tunnel junction matched to a particular energy $eV = \phi_2$.

(1) Metals, absorbates, and the observed change in work function $\Delta\phi$ are listed in Table I.

(2) The work function of the metal surface is added to the decrease in work function $\Delta\phi$ produced by the adsorbate/interface to obtain the work function required for the metal.

(3) From the Table of Work Functions of the Elements, Candidate metals are selected which have a work function close to the required work function calculated in Table II.

(4) The order of deposition must be taken into account; for example: Ni/BaO or NaO/Ni For example: In Table (III) line 1, the metal selected Mo, Be, Co, Ni is tested with the adsorbate BaO such as BaO/Ni; BaO/Co; etc.

The selection of materials for a junction suited to each wavelength range has been illustrated with examples. Other combinations of materials may be employed for the metal surface alloys, surfaces with ion implantation, semimetals such as bismuth and the like, and various crystal orientations.

Other materials may be employed for the adsorbate metal oxides, alkali metals, the number of monolayers may be varied; and mixtures of absorbates may be used. The order of deposition may be varied to change $\Delta\phi$.

The selection of materials for tunnel junctions having the requisite work functions ϕ_2 and ϕ_1 for each wavelength range may be made to meet these requirements using the available materials and techniques described above; or modifications.

The device is simple and inexpensive. It utilizes a readily available amorphous substrate such as glass, although it is not limited thereto. It does however require precision fabrication. In a submicron facility Electron beams may be employed to produce the extremely small structures required, using Process A or Process B described above. Ion beams or molecular beam epitaxy may be used to lay down the appropriate metal and insulating areas in producing submicron electron devices. The insulating layer may be silicon dioxide, aluminum oxide, or other insulating layers. The metal strip may be a single crystal which may form spontaneously in such small dimensions or which may be induced to crystallize by suitably heating and cooling the coating, and/or by the momentary application of electric or magnetic fields, and/or by epitaxial growth on a crystalline substrate. The metal strip may have any cross-section but is preferably a square or rectangular high purity single crystal having a long mean free time, such as tungsten, for which $\tau = 1.6 \times 10^{-13}$ sec.

The laws of physics which apply to large-scale electrical circuits in the macro regime are different from the quantum electrodynamic laws of physics in the quantum regime. Because of the small current and time intervals concerned, individual electrons are utilized one at a time. In a Femto Diode, a single electron approaches the barrier traveling over submicron distances - with an energy $\varepsilon = hv = Ve$.

The penetration of a barrier by an electron possessing an energy eV slightly less than the barrier potential $e\phi_2$ occurs according to quantum mechanics by an effect known as "electron tunneling through a barrier". According to the quantum theory of tunneling, an electron moving in a metal approaching an insulating barrier, either passes through the barrier to the metal on the other side by "tunneling"; or is totally repelled by the electric field potential at the barrier/metal interface, and reverses its direction of motion. The waves are transmission/reflection probability waves, not actual particles.

According to the well-established theory, the effect occurs because an electron has a probability wave function which extends a considerable distance and penetrates a thin barrier ($s \approx 30$Å). The probability of transmission of an electron passing through the barrier depends on various parameters, which are defined in mathematical equations derived from fundamental considerations.

The passage of the electron through the barrier occurs because the location of an electron in space is indeterminate, expressed as a probability wave function of the electron being in a given position. This wave function extends over a distance of at least 100Å which is greater than the thickness s of the insulating barrier; usually 28 to 38Å for a Femto Diode.

The electron penetrates the insulating barrier because its position along the axis normal to the plane of the barrier is described by a probability law derived from the Schrodinger equation, from which the Tunnel Transmittance Equation was derived.

The electron may penetrate the barrier and appear on the other side, with its kinetic energy now converted to an equal quantity of potential energy; or, the electron may be reflected, and reverse its direction without loss of energy. A single electron oscillates back and forth in the well without loss of energy until it passes through the barrier.

These phenomena occur without loss because there are no electron collisions, and in the reversal of direction, the initial and final velocities of the electron are equal and opposite.

In this region the effective mass m* of an electron may be about 0.01 m_e the rest mass of an electron. Hence, in a metal, the electron velocity increase from a given quanta of energy is greater than that of an electron in free space.

As an example, the current through a Femto Diode may be about 1.6 x 10^{-13} amps; for which the number of single electrons/sec is: N = 10^6 electron/sec; or 1 electron per μ sec. These are single electron events.

RECENT WORK

Our early work on the manufacture of nanostructures was to enable the large scale rapid manufacture of photovoltaic sheets which are 80% efficient, and low cost, <50 cents/watt. Now, however, many other uses are foreseen; efficient, high definition (1 pixel/μm_2) 2D and 3D HD TV [37], low cost laser polarized lighting sources for general illumination [39], and ultraminiaturized nanometer computer circuits of high density at low cost, all on glass surfaces. The produce and processes describe herein obviate the present high cost of semiconductor crystal substrates.

The Super Submicron Electron Beam Writer, or "Supersebter" [38, 40] is the device used in the method of making large area nanostructures. A key component of the supersebter is a large area field emitter array. Such arrays with packing densities of 1.5×10^7 tips/cm^2 have been fabricated for flat displays and other uses [46]. These emitter arrasy are commercially available and will facilitate the construction of a Supersebter.

A recent paper [41] describes the desorption of NH_3 from a TiO_2 surface using a 400 eV electron beam and 2.5 x 10^{15} electrons/cm^2; that is, an incidence of about one electron onto an area of $4Å^2$. Thus one 400 eV electron is able to desorb one NH_3 positive ion from a surface. Consequently, a low voltage, low current electron beam should be capable of writing a charge pattern on a surface.

Recent work [42-45] has shown that a submicron antenna-diode structure will polarize incident visible light, and convert incident light power to electric power at a load. Thus, large area photovoltaic panels using a submicron antenna-diode structure was shown to be feasible, and further work recommended.

REFERENCES

[1] Nanometer-Scale Fabrication Techniques; R. E. Howard and D. E. Prober VLSI Electronics-Microstructures: Science Vol, V. 1982, Academic Press, Norman G. Einspruch, Ed. pp. 1-24 include., 118 References 8 Tables, 15 FIG.

[2] R. E. Howard et al. Science 221, 117 (1983)

[3] In situ vaporization of very low molecular weight resists using 1/2 nm diameter electron beams: M. Isaacson and A. Murray; J. Vac. Sci. Technol., Vol. 19, No. 4, Nov. - Dec. 1981, 1117-1120

[4] X-Ray Lithography Gets Closer; Jerry Lyman, Electronics, May 26, 1986, p. 15-16

[5] Scanning Electron Microscope S-800; Catalogue EX-E591P 1986; Hitachi, Ltd., Tokyo, Japan; FIGS. 1 and 2; page 4; FIG. 7; Page 6

[5].1] "The Field Emission Electron Source" and Table: page 4

[6] Evaluation of the Emission Capabilities of Spindt-Type Field Emitting Cathodes, Forman, Ralph, NASA Lewis Research Center, Cleveland, Ohio, 44135, Applications of Surface Science 16(1983)277-99 FIG. 3: Page 284, FIGS. 4 and 5: Page 278, FIG. 6: Page 285

[7] DIELECTRIC CONSTANTS OF CERAMICS, Table includes: Dielectric Strength v/mil, Volume Resistivity ohm-cm @23° C., CRC Handbook of Chemistry and Physica, Page E-55, 65th Edition, 1985, CRC Press, Inc., Boca Raton, FL

[8] Pat. No. 4,720,642 issued Jan. 19, 1988 to Alvin M. Marks
[8].1] Cols. 5-8 include.

[9] The Principles and Practice of Electron Microscopy, Ian M. Watt, Cambridge University Press; 1985, Pages 10-19 include.; particularly page 13.

[10] Transmission Electron Microscopy Physics of Image Formation and Micro Analysis, Reimer, Ludwig, Springer-Verlag, Berlin Heidelberg New York Tokyo 1984, Page 88 Table 4.1 Field Emission cathode, Page 88 Table 4.1 Field Emission Cathode Data pages 104-105 Spherical Aberration and Error Disc, Pages 122-123 Field Emission Scanning Transmission Electron Microscope, Pages 272-273 Electron Spread in Thin Specimens, Pages 284-285 Grid and Anode Electrodes with Field Emitting Electrode

[11] Experimental High Resolution Electron Microscopy, John C. H. Spence, Clarenden Press Oxford 1981, Chapter 2, Electron Optics, pages 28-59

[12] Atomic Resolution of TEM Images of Surfaces, The newest Electron Microscopes that crack the 2 Angstrom barrier can resolve atom positions on metal and semiconductor surfaces. Robinson, Arthur L., Science, Vol. 230 pages 304-306

[13] Dynamic Atomic-Level Rearrangements in Small Gold Particles, Smith, Petford-Long, Wallenberg and Bovin, Science Volume 233, Pages 872-874,

[14] Discharges and Mechanoelectrical Phenomena in Charged Glasses, Vorob'ev, Zavadovskaya, Starodubstev, and Federov, S. M. Korov Tomsk Polytechic Institute, Izvestiya Vishhikh Uchebnykh Zavedenii Fizika, No. 2, pp 40-46 February 1977

[15] Electret Effect in Silica Glass containing Two Types of Alkaline Ions, E. Rsiakiewicz-Pasek Institute of Physics, Technical University of Wroclaw, Poland, Journal of Electrostatics, 16(1984)123-125, Elseview Science Publishers B

[16] Surface Charges in a Zinc Borosilicate Glass/Silicon System, Misawa, Hachino, Haa, Ogawa and Yagi, Hitachi, Ibaraki, Japan, J. Electrochem Soc., SOLID STATE SCIENCE AND TECHNOLOGY pp. 614-616, March 1981

[17] Surface Charge of Vitreous and Silicate Glasses in Aqueous Electrolyte Solutions, Horn and Onoda, Journal of The American Ceramic Society, pages 523-527, Vol. 61 No. 11-12, Nov. - Dec. 1978

[18] Electroless Deposit of Cuprous Oxide, Greenberg and Greininger, J. Electrchem. Soc.: electrochemical Science and Technology, pages 1681-1785, Vol. 124, No. 11, Nov. 1977

[19] A Copper Mirror: Electroless Plating of Copper, Hill, Foss and Scott, University of Wisconsin, River Falls, WI 54022, Journal of Chemical Education, Page 752, Vol, 19

[20] Comparison of Stannous and Stannic Chloride a Sensitizing Agents in the Electroless of Silver on Glass Using X-Ray Photoelectron Spectroscopy, Pederson, L. R., Solar Energy Materials, Pages 221-232, 6, 1982, North Holland Publishing Company

[21] Electroless Gold Plating - Current Status, Yutaka Okinaka, Abstract No. 441, Bell Communications Research, Inc., 600 Mountain Avenue, Murray Hill, NJ 07974

[22] AutoCatalytic Deposition of Gold and Tin, A. Molenaar, Abstract No. 442, Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands

[23] Anodic Oxidation of Reductants in Electroless Plating, Ono and Haruyama, Abstract No. 444 Tokyo Institute of Technology, Graduate School at Nagatsuta, Midori ku, Yokahama 27, Japan

[24] Application of Mixed Potential Theory and the Interdependence of Partial Reactions in Electroless Copper Deposition, Bindra David and Roldan, Abstract No. 445, IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598

[25] Transparent Plastic Articles having Ammonium Suphonic Acid Groups on the Surface thereof and Methods for their Production, U.S. Pat. No. 3,625,751, Issued Dec. 7, 1971, Wilhelm E. Walles

[26] Barrier Plastic Articles, U.S. Pat. No. 3,916,048, Issued Oct. 28, 1975, Wilhelm E. Walles

[27] Laser and Electron Beam Processing of Materials, Edited by C. W. White, Solid State Division, Oak Ridge National Laboratory, Oak Ridge, TN, P.S. Peercy, Division 5112, Sandia Laboratories

Academic Press. New York, N.Y., 1980, Pages 18-19, Fundamental Mechanisms, Pages 65-69, Melting by Pulsed Laser Irradiation, Pages 671-677, Laser Induced Photochemical Reaction for Electronic Device Fabrication, Pages 691-699, Metastable Surface Alloys, Page 697, Crystallization of Copper from the melt using electron beam

[28] Fine Focused Ion Beam System using Liquid Metal Alloy Ion Sources and Maskless Fabrication Gamo, Inomoto, Ochiai and Namba, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan, Pages 422-433

[29] Microscopy and Lithography with Liquid Metal Ion Sources, Cleaver, J. R. A., Department of Engineering, Cambridge University, England, Pages 461-466, Inst. Phys. Conf. Ser. No. 68: Chapter 12 EMAG, Guildford, Aug. 30 - Sep. 2, 1983

[30] Electrohydrodynamic Ion Source, Mahoney, Yahiku, Daley, Moore and Perel, Pages 5105-5106 Journal of Applied Physics, Vol. 40, No. 13, Dec. 1969

[31] Stable and Metastable Metal Surfaces, Spencer M.S., Pages 685-687, Nature Vol. 323, 23, October 1986

[32] Electrons in Silicon Microstructures, Howard, Jacket, Mankiewich, and Skocpol, Pages 346-349, Science Vol. 231, 24 Jan. 1986

[33] Atomic-Resolution of TEM Images of Surfaces, Robinson, Arthur L., Pages 304-306, Science Vol. 230, 18 Oct. 1984

[34] Time-Resolved Electron Energy Loss Spectroscopy, Ellis, Dubois, Kevin and Cardillo, Pages 256-26, Science, Vol, 230, 18 Oct. 1985

[35] Optical Holography, Collier, Burchkhardt and Lin, Bell Telephone Laboratories, Pages 300-301 Academic Press, New York, N.Y. 1971

[36] Patent No. 4,445,050, issued Apr. 24, 1984 to Alvin M. Marks, entitled "Device for the Conversion of Light Power to Electric Power"

[37] Patent No. 4,720,642, issued Jan. 19, 1988 to Alvin M. Marks, entitled "Femto Diode and Applications"

[38] Patent No. 4,798,959, issued Jan. 17, 1989 to Alvin M. Marks, entitled "Super Submicron Electron Beamwriter"

[39] Patent No. 4,972,094, issued Nov. 20, 1990 to Alvin M. Marks, entitled "Lighting Devices with Quantum Electric/Light Power Converters"

[40] Patent No. 5,268,258, issued Dec. 7, 1993 to Alvin M. Marks entitled "Monomolecular Resist and Process for Beamwriter"

[41] Adsorption and electron stimulated desorption of $NH_3/TiO_2(100)$ U. Diebold and T.E. Madey, Department of Physics and Astonomy and Laboratory for Surface Modification, Rutgers, The State University of New Jersey, Piscataway, NJ, 08855, J. Vac. Sci. Technol.A 10(4) Jul/Aug 1992, 0734-2101/92/042327, 1992, American Vacuum Society

[42] Antenna Absorption of Visible Light and the Production of Electricity, Guang H. Lin and John O'M Bockris, Chemistry Department, Texas A&M University, College Station, TX 77843, Oct. 1994, funded by U.S. Gov't; (not published)

[43] Letter dated October 31, 1994 from John O'M Bockris, Distinguished Professor of Chemistry at Texas A&M University, to U.S. Gov't. Project Officer states:

"I must support Dr. Lin's contention that a significant contribution has been made here. It seems to me that a new principal has been proved: one can indeed create electricity from antenna absorption of solar light.

What the efficiency of this electricty is will eventually be is difficult to say at the moment because we have some way to go in understanding rectification. On the other hand, there are certain features which would seem to make the eventual development of such a device more efficient than those of photovoltaics.

But of course, we must always remember that the basic idease from Alvin Marks, and I believe that he should be given some kind of award - along. perhaps, with Dr. Lin - for the verification of this remarkable suggestion."

[44] Fabrication and Optical Properties of Fine Metal Grids, Final Report, May 31, 1994, W. Wright, R. C. Tiberio and H. G. Craighead, School of Applied and Engineering Physics and the National Nanofabrication Facility, Cornell University, funded by U.S. Gov't. (not published)

[45] Letter dated Sept. 7, 1994 to Dr. Alvin M. Marks from U.S. Gov't. Project Officer on Final Report (Cornell) [37], states:

"1. Enclosed please find final report summarizing the work accomplished at Cornell University... over the last year in regards to assessing the feasibility of the collection of solar energy with antenna structures (idea based on your patent #4,445, 050).

2. We judged the effort a success, but there are still many unknowns that would effect the overall efficiency of such a device. Based on the funding availability, we plan to pursue the concept further..."

[46] Field Emitter Arrays for Vacuum Microelectronics, C. A. Spindt, C. E. Holland, A. Rosengreen, Ivor Brodie, IEEE Transactions on Electron Devices, Vol. 38, No. 10, October 1991. The authors are with SRI International, Menlo Park, CA 94025

SURFACE MICRO-MACHINING: PROGRESS TOWARDS MICRO-ELECTRO MECHANICAL SYSTEMS

James Doscher
Analog Devices Inc.
831 Woburn St
Wilmington, MA 01887 USA
Phone: 617-937-1534, FAX 617-937-1607
Email: jim.doscher@analog.com

Surface micromachining is a technique for building electromechanical systems in silicon. Combined with on-board signal conditioning circuits, complete electromechanical systems can be economically built on a single piece of silicon using a standard integrated circuit fabrication line. The first commercially successful applications for surface micromachined sensors were accelerometers for automotive airbags.

Since then, a number of electromechanical systems have been implemented in miniature using the fundamental structural building blocks of tensile and non-tensile springs, differential capacitance sensing cells, and electrostatic drive. Work, supported by an ARPA contract, has demonstrated X, Y and Z axis accelerometers, rate gyros, flow meters, electromechanical filters, resonant accelerometers, and electromechanical relays. An interesting thing about these sensors is that they are not just laboratory curiosities, but have been fabricated on a proven wafer fabrication process and could be built in high volume at reasonable prices.

The integration of complicated mechanical structures and electrical circuits onto a single chip is expected to improve reliability and testability of systems. For example, MEMs accelerometers available today have built in self test functions that are able to physically move the beam structures in order to test the entire electromechanical control loop. Reductions in interconnect wiring, increased use of automation, and the inherent reliability of integrated circuit processes will all contribute to increased reliability of systems. Accelerometers available today demonstrate a failure rate of 10 FITS, (10 failures in 10^9 device hours.

One of the obvious aerospace applications for the technology is miniature, low cost guidance and control systems. The basic technology is in place to show a path to the design of a single chip inertial navigation system. Increases in circuit density and process yield will make this a reality in a few years. If the past is any indicator, however, we may not yet perceive of the best and most useful applications of this technology in space.

Advanced packaging for Integrated Micro-Instruments

71163

230913

10P

James C. Lyke
Phillips Laboratory (USAF)
(505)846-5812 lyke@plk.af.mil

Abstract

The relationship between packaging, microelectronics, and microelectromechanical systems (MEMS) is an important one, particularly when the edges of performance boundaries are pressed, as in the case of miniaturized systems. Packaging is a sort of physical backbone that enables the maximum performance of these systems to be realized, and the penalties imposed by conventional packaging approaches is particularly limiting for MEMS devices. As such, advanced packaging approaches, such as multi-chip modules (MCMs) have been touted as a true means of electronic "enablement" for a variety of application domains.

Realizing an optimum system application of packaging, however, is not as simple as replacing a set of single chip packages with a substrate of interconnections. Research at Phillips Laboratory have turned up a number of interesting options in the two- and three-dimensional rendering of miniature systems with physical interconnection structures with intrinsically high performance. Not only do these structures motivate the redesign of integrated circuits (ICs) for lower power, but they possess interesting features that provide a framework for the direct integration of MEMS devices. Cost remains a barrier to the application of MEMS devices, even in space systems. As such, several innovations are suggested that will result in lower cost and more rapid cycle time. First, the novelty of a "constant floor plan" multichip module (MCM) which encapsulates a variety of commonly used components into a stockable, easily customized assembly is discussed. Next, the use of low-cost substrates is examined. The anticipated advent of ultra-high density interconnect (UHDI) is suggested as the limit argument of advanced packaging. Finally, the concept of a heterogeneous 3-D MCM system is outlined, which allows the combination of different compatible packaging approaches into a uniformly dense structure that could also include MEMS-based sensors.

Introduction

Electronics comprise 30 - 40 % of space systems, and their presence clearly has significant logistics implications and overhead. New approaches and processes have been developed for the packaging and design of electronic systems. The combination of these technologies imply tremendous reductions in size, weight, and power and improvements in performance, which will allow and enable the systematic reductions in spacecraft weight for a given function, or greatly improved functionality for a given weight. New technologies are furthermore under research and development that may reach the theoretical limit for density in a planar packaging approach, where even the semiconductor layers are thinned to a pliable regime. The possibilities of creating conformable electronic building blocks suggests that there is a new paradigm in the missile and satellite construction. Clearly, these payoffs can expand past space applications, which will clearly realize the most significant and immediate advantage.

In the whole of microelectronics and packaging research, two trends are clear: (1) increased functionality per unit volume, and (2) increased mixtures of functionality. The first trend is accelerated through advances in IC processes and the advent of two- and three-dimensional packaging. The second trend refers to the increased tendency to consider the mixture of disparate types of electronics functions (e.g., analog instrument, digital processing, communications, and power) within the same electronic module, also known as "mixed signal technology". The implications of these trends, extended in time, are that eventually most electronic systems of a given capability can be hosted in increasingly smaller volumes. The

degree to which this will be possible in general, depends as much upon packaging as it currently depends on IC feature size reduction, simply because no single monolithic IC process could be made to address disparate functional realms profitably at the feature sizes of interest.

Packaging technologies deal with the problem of physically supporting and interconnecting system components. In the specific case of an all-electronic system, one is concerned with at least four functions: (1) structural support and environmental protection of thin semiconductor slivers or "chips"; (2) thermal management; (3) information-bearing signal distribution; and (4) electrical power and grounding distribution. But a generalization of packaging requires that components which convert information-bearing electrical signals into another form of energy (or vice versa) be dealt with, and that class of components include conventional and MEMS-based sensors and actuators. Such a generalization would also deal with various non-electrical permutations of interfaces between components, such as the capability to deal with aggregations of actuator forces from individual components or routing complex fluidic manifolds throughout a system.

Conventional packaging schemes fall along the package-board-box-system (PBBS) paradigm. The paradigm implies a hierarchy in electronics packaging, as suggested in Figure 1. Chips are placed into packages, which are placed onto printed wiring boards (PWBs), that are in turn grouped into a "card cage" or chassis, an ensemble of which, when combined with the associated harnesses and connectors, constitute an electrical system. Most system designers are enslaved by (as opposed to enabled by) the convenience of conventional packaging, in stressing cases, however. One example of a stressing case is when order-of-magnitude size and weight reductions are clearly indicated, as they are in, for example, micro-spacecraft. If the PBBS cycle could be modified or broken, then tremendous opportunities exist for further optimization. If the normal traversing distances of feet between boxes can be reduced to centimeters or millimeters, it is conceivable that the associated functions could be further re-engineered to exploit the reduced drive required, saving power consumption and reducing latency.

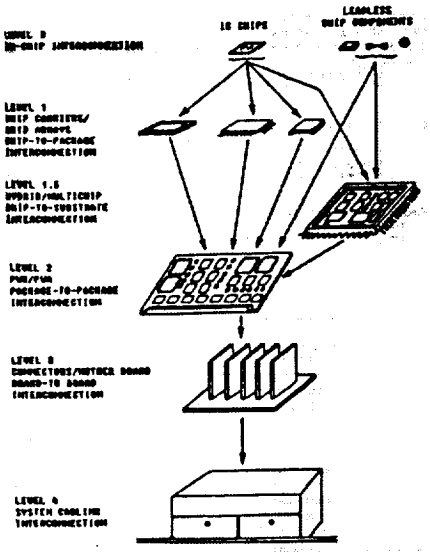


Figure 1. The hierarchical nature of packaging.[1]

It would seem that this hierarchical approach would also apply to electronics systems that contain MEMS components, while the paradigm for many non-electronic systems is more diverse and modular, reflecting a unified integration of structure and functionality, such as the case of a bicycle, combustion engine, or jet fighter aircraft. An integrated microsystems may combine the best features of both hierarchical and modular approaches.

Advanced packaging is very much about optimizing the way that components are contained and interconnected within an electronics system. For space-constrained systems, conventional packaging is particularly poor in efficiency -- typically less than one percent[2]. Through advanced packaging, the volume packing efficiency can be improved upon considerably, perhaps as much as 40 times better than conventional approaches. For performance-intensive systems, conventional packaging systems become increasingly problemmatic, due to the continuous demands for increased throughput and bandwidth. The conventional approaches introduce significant series loss and capacitive parasitic components, and, as frequencies are increased, the time-of-flight delays in packaging become so significant as to warrant treating interconnections as transmission lines.

Planar (2D) Advanced Packaging Approaches

Advanced packaging concepts can be applied at each level at the packaging hierarchy. The most well-known concept in advanced packaging is the multi-chip module (MCM), which combines a number of ICs and discrete components that would normally be individually packaged in a much denser form within a common package, as suggested in Figure 2. MCMs offer advantages similar to monolithic wafer scale integration (WSI), but without many of the intrinsic problems. MCMs are realized in one of two configurations, patterned substrate and patterned overlay, and are categorized by the industry using three types: MCM-C, MCM-D, and MCM-L.

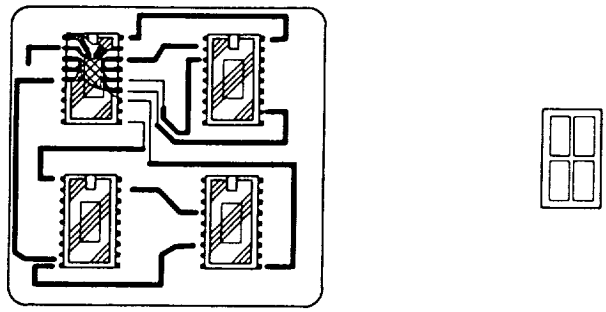


Figure 2. Impact of advanced packaging on size reduction.

Defining the configuration of MCMs requires an understanding of the construction of a typical MCM substrate (Figure 3a). The mechanical substrate provides the essential physical support for components on an MCM. The interconnecting substrate provides the wiring media for signal and power distribution. In the patterned substrate configuration (Figure 3b), the components are placed onto both substrates. In the patterned overlay process (Figure 3c), however, components are contained in the mechanical substrate, but signal distribution is accomplished with an interconnecting substrate which is applied over the components and mechanical substrate.

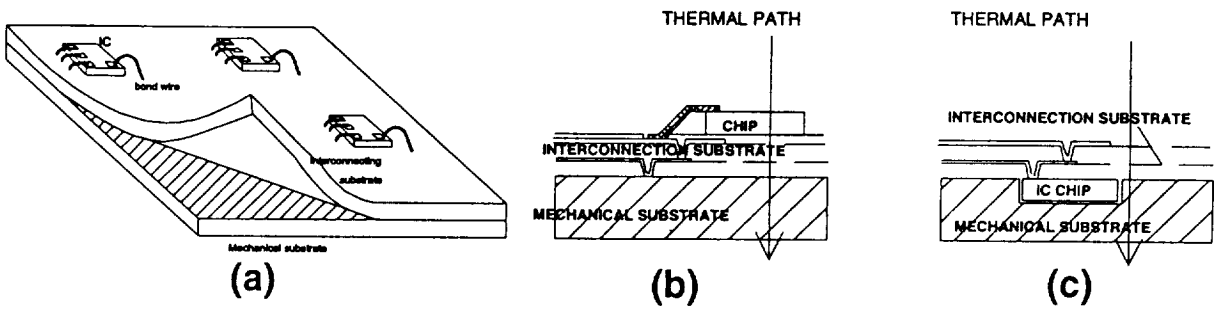


Figure 3. MCM configurations. (a) General construction. (b) Patterned substrate. (c) Patterned overlay.

MCM types are often designated as MCM-C, MCM-D, and MCM-L, which represent ceramic, thin-film, and laminate MCM technologies, respectively. These designations refer to the type of interconnecting substrate material, and in the MCM-C and MCM-L cases, the mechanical and interconnecting substrates coincide. MCM-C approaches are contemporary versions of the "chip and wire" hybrid. MCM-C modules have and continue to be the most prevalent form of advanced packaging in use in military and space systems, due to maturity and lack of organic materials, or, in other words, due to tradition. The thin-film MCM-D approaches are among the most recent developments in MCM technology, often boasting higher connection density and electrical performance (and unfortunately cost). The forms of MCM-D in research vary widely, ranging from nChip's silicon/SiO_2 patterned substrate approach to Lockheed Martin's High Density Interconnect Cu/polyimide/ceramic patterned overlay approach. MCM-L technology is sometimes referred to as "chip-on-board", and is usually thought of as the least capable and expensive of the MCM types, although modules of increasing sophistication have been achieved with this approach. Acceptance of MCM-L in military and space applications has been controversial due to concerns over the high content of organic materials. The significant research interest in these technologies have led to continued advancemeents in each MCM type, making it increasingly difficult to make any absolute assertions about which type is better for a given metric. It is also possible to mix the types, leading to sometimes confusing nomenclature. Another distinguishing characteristic of MCMs is the method by which components are

attached to the substrates. Prevalent forms of chip to substrate attach include wirebonding, tape automated bonding, and flip-chip attachment, with wirebonding remaining by far the most prevalent attachment method.

MCM Technology Benefits and Challenges. The most compelling case for MCMs, particularly for complex and mixed-signal applications, is the ability to greatly compress the size and weight of an otherwise conventionally packaged assembly. For these assemblies, it is not uncommon to observe ten-fold improvements in packaged configurations. For highly regular structures, such as memory components, the benefit is considerably reduced, usually 20-50%, in packaged configurations. MCM packaging enables designers to enhance electrical perfomance. In early ARPA/Phillips Laboratory research, it was not uncommon to operate complex digital systems at twice their normal operating frequency.[3] MCMs heuristically afford a higher reliability, as the number of interfaces are greatly reduced, lowering the actual number of failure modes.

While the benefits are great, MCMs are not without issues. The most significant problem facing theacceptance of MCMs is the so-called "known-good-die" problem. Conventionally, single chip assemblies are usually tested in-line in their final package, and non-functional components are discarded. In MCM assemblies, however, the die are included in substrates prior to packaging, and are usually not fully tested to the confidence levels of single chip packages. Hence, yield loss in MCMs are magnified tremendously as a function of the number of components and the yield of each component. The MCM developer can at this point only choose to develop more complex test fixturing to actually perform at-speed tests on bare die or to simply test a finished MCM assembly and "repair-to-yield" or discard the assembly. Each of these options add cost to the MCM, which is already high. The second most significant problem with MCM technology is cost of the packaging medium itself. The reason most often attributed to the high cost of MCM packaging aside from component issues is the relatively low production volume, particularly for MCM-D processes. In the aerospace communicty, another barrier to MCM use is the lack of adequate acceptance criteria for MCMs, and the general reluctance to use new technologies. The most advanced MCM approaches employ polymeric materials in their construction, which has historically been problemmatic. The most cost effective MCM assemblies feature non-hermetic encapsulation, which is furthermore considered taboo to many systems development programs. As such, plastic encapsulated microcircuits, whether single or multi-chip packages, have faced an uphill battle for use in aerospace systems.

Despite the challenges, considerable progress is being made in the development and insertion of advanced packaging for use in space systems. Most significant and representative forms of MCM technology are in-orbit slated for spaceflight in the near future. Increased awareness and understanding will be key to continuuing the trend.

Three-dimensional (3D) packaging

While MCMs provide significant improvements in planar packaging density, it is not always enough. First, MCMs are not always applied in an intelligent manner. If board designs, for example, are not carefully planned, MCMs may be used in a space inefficient manner, mitigating much of the benefit that an MCM approach could have provided. As such, some designs have realized little if any real improvement, due to lack of balance in considering advanced packaging at all levels in the Figure 1 hierarchy. Still, after all considerations have been effectively dealt with, 2-D is fundamentally limited, and the highest efficiency, which occurs when the sidewalls of ICs contact one another, is still not the best achievable. The situation begs the question that: if one would go to such lengths in planar packaging, then why ignore the third dimension?

Three-dimensional packaging refers to the consideration of non-planar techniques to improve overall system packaging efficiency on a volumetric basis. It is a natural consequencce and extension of planar MCM and board packaging methodologies. Far less mature than MCM approaches, a great diversity of 3-D approaches exist, range from 3-D integrated circuits (monolithic ICs that are "grown" one atop another) to stacked IC and MCM approaches. Examples of representative 3-D approaches are shown in FigureXXX.

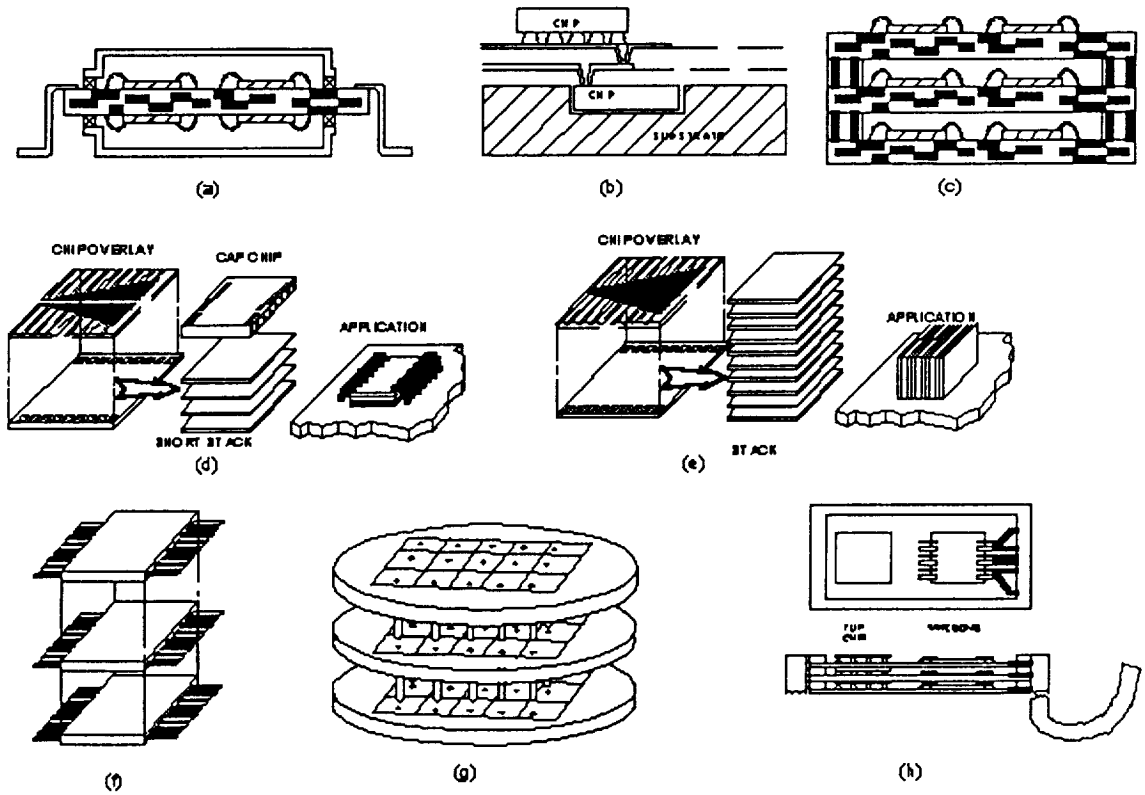


Figure 4. Representative 3-D Packaging Approaches.[3] (a) Double-sided cofired ceramic substrate. (b) Patterned overlay with surface-mounted components. (c) Cofired ceramic or thin-film, patterned substrates, stacked with interposer/spacers. (e) Thin-film patterned chip stack, short form. (f) Thin-film patterned chip stack, cube form. (g) 3-D monolithic wafer scale integration with areal connection technique. (h) Sensor electronics packaging of very thin (0.004") layers with ortogonal mounted imaging sensor detector array.

The benefits and challenges of 3-D packaging are similar to 2-D packaging, and differ largely only by degree. With 3-D packaging, greater reductions in size and weight are possible, and the potential to accelerate performance is improved. On the other hand, the pre-test and cost issues are also enhanced. One must now consider "known good assemblies", in addition to the "known good die" problem of MCMs. Cost for achieving the connections between layers is also high, for much the same reasons that 2-D MCM substrates are presently expensive. If MCMs have a problem in terms of economy of scale, it is only that much worse for 3-D MCMs and packages. It should be noted that sometimes, however, that a simpler form of 3-D packaging can sometimes be employed in an application with greater effectiveness than even the most sophisticated 2-D approach, especially for regular, simple electronic structures, such as mass memory. On the other hand, 3-D packaging is subject to the same abuses of misinterpretation in application as 2-D MCM approaches.

In fact, 3-D packaging promotes new thinking in design to some degree. For space-constrained systems, an increasingly larger fraction of an entire electronics system could be contained within the boundary of a single 3-D assembly. While much of the research in 2-D MCMs today is focussed on digital-only applications, it is the case, particularly for submunitions, missiles, and space systems that the 3-D MCM will have to deal with the packaging of analog instrumentation, microwave, power electronics, and even sensors and actuators. The high performance system, another driver for 3-D packaging, requires shorter electrical paths and many more of them. Thermal management, important in both cases, is a particularly acute concern for high-performance systems. Also of increasing importance is the notion that integrated circuits could and should be re-engineered in 3-D packaging to most advantageously support the

tremendously lower capacitance and series loss potential of that packaging environment. Doing this effectively suggests that it is highly desirable to establish a standard physical packaging architecture for heterogeneous systems. This architecture would include features such as:

- highly compact form factor, as a small monolithic block of functional electronics, with a density one to two orders of magnitude for digital and mixed-signal systems, even those that currently employ hybrid and MCM technology;
- open standards to accommodate a great number of existing and emergent hybrid/MCM technologies;
- flexible interface provisions and serviceable (demountable) layer and computer-aided-design (CAD) protocols;
- high signal integrity potential (based on adherence to specified design rules), allowing the merger of digital, analog instrument, microwave, and power circuitry within a single, functional block;
- adequate thermal management facilitation and comprehensive thermal characterization, allowing cookbook design methodologies (vs PhD-operated sophisticated numerical analyses)

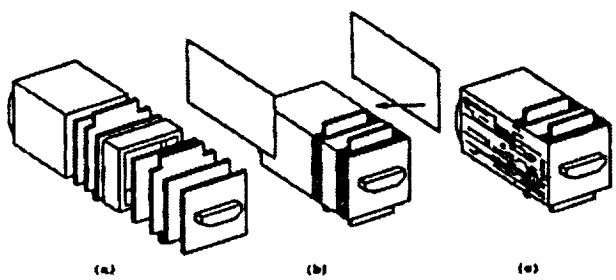


Figure 5. Three-dimensional packaging system developed for the Monolithic Interceptor Processor (MiP) project. Removes several levels of packaging hierarchy, integrates many forms of functionality, includes mechanical components, in a coffee cup size (1.6x1.6x3.5 inches).

These were essentially the conclusions reached in a recent Phillips Laboratory research program that studied methods of re-packaging the entire electronics functionality of a next-generation interceptor platform in the most efficient way possible. The program, referred to as the Monolithic Interceptor Processor (MiP), evaluated the integration of imaging sensors and inertial guidance systems within the same coffee cup form factor that housed 500 megaflops of digital processor, analog interfacing, and high-speed communications linkages. As shown in Figure 5, an integrated packaging system can eliminate entire levels of the packaging hierarchy. The advantages of the boardless, even box-less system packaging approach are clear for miniature systems, but improved performance and physical robustness are other potential benefits of interest to a great number of other systems. PL research concluded that the 3-D approaches required the flexibility to accommodate a great variety if not all possible electronics components of interest to a system designer, even mechanical components, such as MEMS devices.

Given the possibility of creating a three-dimensional compact packaging system, if sufficiently flexible, it is possible to integrate modules from a variety of industry and laboratory processes. Functions may be implemented by different vendors and integrated at one location. This packaging framework is extensible to the most aggressive technologies, as well as the most primitive.

Towards an Ultra-High Density Interconnect (UHDI)

Another area of exciting research is the drive toward the outer edges of packaging density. The working definition of UHDI processes may be provided as follows:

advanced packaging approaches and related infrastructure to achieve reliable and affordable assemblies with densities substantially beyond and properties atypical to the practiced state-of-the-art in practical multi-chip module (MCM) technologies. The delineation of layer thickness, volume efficiency, and price per cubic inch, as well as better metrics, are yet to be defined. UHDI processes are expected to achieve

greatest volume efficiencies through their ability to be stacked directly, and they are thought to be most versatile by virtue of an expected conformable abilities.

One representation of a UHDI process is shown in Figure 6. The conformable properties and associated metrics for UHDI assemblies are among its most intriguing and speculative features. Other potential benefits to UHDI process include:

- Tremendous volume densities for solid-state storage systems (e.g., > 10^9 bits/in^3)
- Enhanced potential for exploiting monolithic wafer scale integration
- Complex, heterogeneous subsystem construction for embedding into sensors, structures
- Greater reliability and operating potential due to improved thermal transport
- Increased radiation tolerance, similar to that provide by dielectrically isolated technologies.

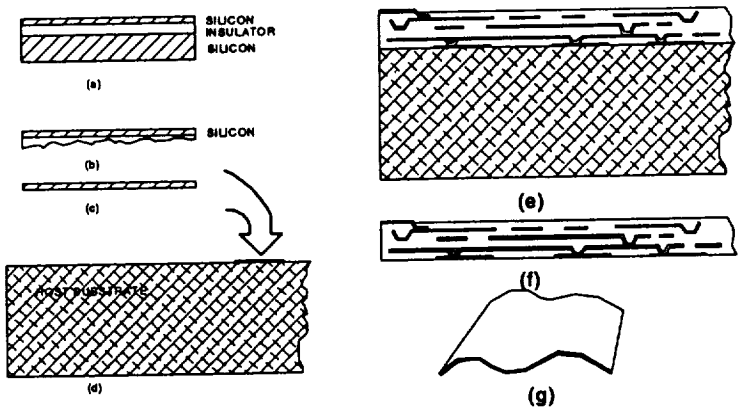


Figure 6 A potential UHDI process sequence. (a) Starting material, in this case an SOI wafer/component. (b)-(c) Epitaxial removal. (d) Attachment to temporary host substrate. (e) Patterned overlay formation. (f) Removal of host substrate. (g) Resultant UHDI membrane[4].

UHDI may well represent the limit argument in advanced packaging. It is predicated on the combination of patterned overlay multichip modules (MCMs) and integrated circuit (IC) thinning approaches to form electronic decal-like membranes. This intriguing next-step in packaging research has far-reaching consequences for almost any form of electronics system, particularly those with intense miniaturization goals.

The weight reduction potential of UHDI and an integrated 3-D packaging are tremendous: in combination, 1000x reductions of system functions are possible for certain classes of system functions. It is conceivable with this approach to re-engineer satellites that are collapsible to an extremely small form for launch and are expanded during deployment to provide the surface area needed for solar panels and communications equipment. Launch payloads could field ten times the number of satellites in a given mission, enabling a dramatic increase in the number of scientific objectives serviced by a single launch opportunity.

Integrated 3-D packaging and UHDI represent limit arguments in advanced packaging. The former establishes a 3-D packaging framework, allowing the commodization of component layers, predictability of physical characteristics, and a true leap ahead of most other 3-D packaging concepts, which typically focus on one level of packaging instead of the entire hierarchy. UHDI can operate inside or outside of this framework. Inside the framework, UHDI represents a preferred layer construction method. A typical UHDI layer would represent a slice less than 0.005" in thickness, compared to 0.040" - 0.200" for normal hybrid/MCM layers. This would represent an eightfold worst case density improvement. A 3-D implementation of UHDI would result in highly dense "pucks", which could be inserted as thicker layers, each of which might consist of several or many component layers, but less in aggregate thickness than a single 2-D MCM layer. Hence, a UHDI re-implementation of a non-UHDI system could be reduced to a credit card sized system in some cases. Outside the 3-D framework, UHDI could conceivably be conformed to structures of opportunity within the satellite, and directly integrated in multi-functional structures under research at Phillips Laboratory (PL/VTS) to eliminate cables and harnesses throughout a satellite.

Breaking Barriers for Rapid Development of Application Specific Integrated Microinstruments: Constant-Floor Plan Multichip Module Design Concept

For reasons that should be evident by now, MCM implementations of integrated MEMS systems represent an intelligent approach for extracting and preserving this performance and density benefits. Of course, the aforementioned barriers to routine MCM implementations, most notably those associated with cost, hinder progress towards establishing prototypes. A common complaint of engineers who would like to experiment with MCMs is the difficulty of justifying or raising the capital necessary to build even a simple design. Even design tools alone can exceed the capability of small companies, not to mention the problems associated with component selection, procurement, pre-test, MCM interconnect design, layout, fabrication, test, and assembly. Little standardization exists in the industry, but even if it did, only part of these issues are ameliorated. Let's examine these issues in a little more detail:

Computer Equipment for MCM Design. Buying a CAD system capable of doing MCM layouts is commonly perceived as a significant hardware/software expense. While it is of some comfort that an EDA infrastructure is emerging, such that it is possible for one to develop MCM designs in a structured and supported framework, the price for this capability is often well over $50,000. Beyond the ability to perform "mere" electrical layout, would-be MCM designers must confront decisions about investing even more money to perform linked thermal, mechanical, electrical, and reliability analyses through additional electronic design automation (EDA) packages sold by OEM and third-party suppliers. MCM designers who must also perform IC designs for a project is in double jeopardy, as he must also acquire IC EDA equipment as well. As such, the price tag to achieve entry design capability can be exorbitant, which is especially trying for "fence straddlers" who are as apt to find "better" uses for the money. It must be remembered that not every one is inexorably convinced of the necessity or even the viability of MCMs.

Component selection. Assuming that one is even in the position to begin an MCM design from an EDA standpoint, floorplanning requires that the components be nominally identified and enough information be secured to permit preliminary assessments of physical footprints. Little guidance exists for this process. Even companies such as Motorola who have supported the KGD Task Force specification activity are not monolithic, giving novice MCM designers mixed signals about die availability. He quickly learns that only a subset of the ICs in his databooks may be procured in die form. Moreover, he finds that even when there is a hint of availability, he will most likely have to construct die bond maps from third generation faxes or will have constructed a physical representation only to learn that the vendor has performed what is commonly referred to as a "die shrink". Passive components are sometimes as bad as integrated circuits on these bases. An added dimension of complexity is the practice of "thinking MCM" when making passive component selections, as not every type of capacitor or transformer, etc.

The situation has been improved considerably by the advent of MCM foundry services (resulting from ARPA sponsorship of the MOSIS "brokerage" of nChip and IBM substrate fabrication services, combined with third party assembly). Even with this service, however, die procurement, rapid design changes, and MEMS component integration are not easily accommodated. Another possibility exists that involves the use of patterned overlays to create rapidly customizable designs that amortize the cost of ubiquitous commodity components across multiple designs and lends itself to ready integration of MEMS devices. This concept is referred to as a *constant floor plan* MCM.

Constant floor plan (CFP) MCMs provide an MCM analog to gate array ICs with fixed underlayers. Simply put, a CFP MCM is based on a custom interconnection of standard, commonly required components, embedded within the substrate of a patterned overlay MCM. It is based on the premise that a judiciously chosen set of ICs and discrete components can satisfy a large percentage of designs (say 65-80%) of a particular class. For example, most microinstruments require analog conditioning circuitry, such as operation amplifiers, switches, multiplexers, and analog-to-digital converters (ADCs). A more complete, stand-alone instrument may require the addition of a microcontroller, miscellaneous logic functions (which could be accommodated using a field programmable gate array), and memory. Chances are that no tow designs, however, will connect these components together in the same manner. By fixing an ensemble of commodity components into a substrate in a pre-determined (i.e., a

constant floor plan), patterned overlay designs can be designer-specified to complete an almost arbitrary interconnection pattern of these components long after the substrate is formed. A production lot could contain many substrates with identical floor plans, but each with a completely different, user-specified interconnection scheme. It is also likely that by allowing a few components t be added that are unique to a particular design, that an even larger number of designs could be accommodated.

Since patterned overlays form interconnections after substrate assembly, it is possible to inventory a large number of identical substrates in unpatterned form for various customers, reducing the expense of many separate component purchases (each die within a design -- up to several dozen -- may be subject to $5,000-$100,000 minimum purchase) by amortizing them across a group of individual customers. The effective turn-around time for a CFP MCM would clearly be less than a full-custom MCM, since many of the die preparation steps as well as component assembly step would have already been performed. For commodity CFPs, it would in principle be possible to pre-test components in known-good-die style test fixtures. Furthermore, it is conceivable that a custom probe apparatus could be fashioned to test the entire substrate prior to locking in its personalization.

Rather than creating a single, "mega"-floorplan in a large substrate containing many components (a clear impracticality), it would be logical to establish a family of constant floor plan designs. A speculative set of constant floor plan MCMs would include: a memory CFP (for creating various parity and block organizations); an all-digital CFP; an mixed-signal, medium quality instrument CFP, a precision instrument CFP; and a digitally controlled microwave CFP. With this range of CFP options, a large number of general purpose instrumentation, memory module, digital logic, and radio-frequency designs could be readily accommodated. While even this range of options will not address every possibility, a great many possibilities for prototype MCMs can realized at the fraction of the cost of a normal MCM prototype.

The constant floor plan MCM concept has several novel features. A symbolic example of the process is shown in Figure 7.

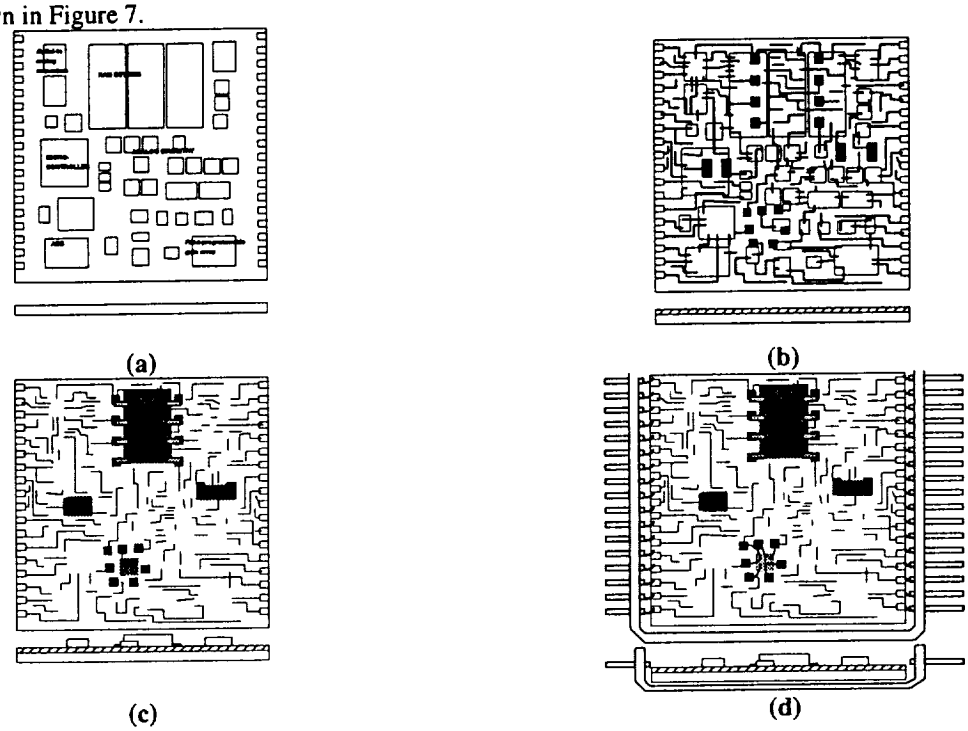


Figure 7. Constant floor plan (CFP) MCM concept. (a) Substrate, containing constant floor plan arrangement of components. (b) Formation of user-programmed interconnections, yielding "smart substrate". (c) Addition of surface-mount components, including MEMS devices. (d) Final bonding and assembly, showing optional kovar package, delidded.

First, design is simplified. Since the initial component placement is optimized at the factory, the user is not forced to assume this task. Pre-defined CAD representations of the CFP MCM would greatly simplify the routing of interconnection patterns, a task which may be automated with a variety of CAD systems, ranging from a PWB design tool on a PC to a full-featured design system on a UNIX-based workstation. Only the final routing patterns need be transmitted electronically to a foundry for processing.

A second feature of CPF is that it establishes a richer infrastructure for a class of a designs than previously possible. The instrument designer is freed from the burden of locating components, securing purchases, capturing die-specific parameters, verifying process compatibility, and of course, floor-planning and die attachment. It is possible to bundle CFP CAD files with more complete and standard information on its specific ensemble of components and applications than would be possible for many quick-turn designs. Driver models and data sheets for each IC could be included in a designer's kit. As more applications are constructed, the experience base would result in a richer accumulation of proven examples in both paper and electronic form. Even as macro cells are defined for gate array ICs, it is also possible to establish libraries of interconnect patterns corresponding to frequently used sub-circuits, such as an instrument amplifier. These macro patterns could be "cut and pasted", greatly reducing design latency. Alternately, a complete pre-worked design could be loaded into a CAD program and edited as necessary to adjust functionality for a related application. The designer's kit could also include pre-worked signal integrity and canned thermal analyses based on the floor plan itself. New thermal analyses could be re-accomplished in a small fraction of the time normally required, simply by adjusting component activity duty factors and re-executing a canned but validated thermal model.

The greatest single novelty of the constant floor plan approach for MEMS-based designers in particular is the fact that the entire top surface of the module is uncommited and could be used for mounting additional components. These components range from additional ICs to augment functionality; passive components to tailor analog functions; programmable memory devices to configure field programmable devices and computers, controllers, and micro-sequencers; and MEMS devices. In this manner, the "real estate" of the MCM can be "double-booked". Alternately, the CFP MCM can be viewed as a "smart substrate", similar to a bare, unassembled hybrid which has a completely unobstructed surface area to work with. As such, the CFP could be configured as an instrument with amplifiers, signal conditioners, digitizers, a computer, and serial interface built inside of the substrate, and completed by surface mounting a few MEMS-based sensors on the top surface. It is possible that the interconnect system, if based on certain polyimides, could serve as a process shield used to release the MEMS devices in final processing.

Next Steps Towards a Viable Heterogeneous 3-D Packaging

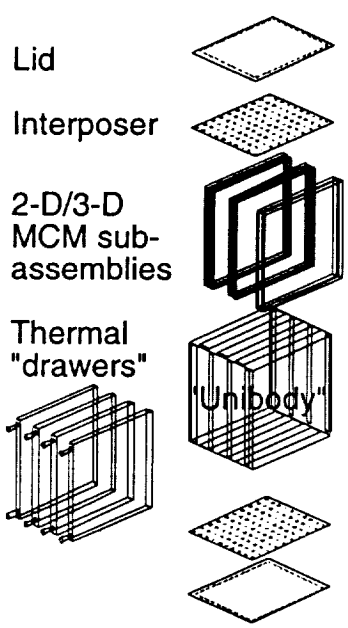


Figure 8. A prospective integrated 3-D packaging system.

Based on the precedents explored in the Monolithic Interceptor Processor (MiP) program, a new effort, termed the Highly Integrated Packaging and Processing (HIPP) program, has been established to promote the development of a heterogeneous packaging framework for 3-D systems, as previously suggested. Two visions for this framework, presented in **Figure 8**, illustrate the principle of an integrated 3-D architecture.. In one case, packaging systems are not only made more efficient, but they are demystified. A systems designer could specify the overall structure and compete the construction of the constituent layers. In this case, a unibody package provides structural reinforcement for the overall assembly (other approaches use a sliding frame to expand or contract accordion-style as layers are added or removed). MCM substrates of many forms may be inserted into the assembly. Thermal management is facilitated by access plenums in one plane, while electrical interface is achieved on the other. Thermal design is a cookbook exercise in this vision of packaging: if a designer exceeds the standard heat-handling capacity, he chooses an improved thermal material or method and slides it into a "thermal drawer".

Conclusions

The increasing sophistication of packaging technologies makes it possible to consider not only the inclusion of MEMS within what has been referred to as "physical packaging frameworks", but also a great variety of other functions, including analog instrumentation, power control, and microwave, in addition to the normal application base on complex digital functions. The trend of megafunctionality is suggested as a future paradigm, which favors in part the trend of micro-spacecraft, even though it is not currently considered an important trend in USAF space systems.

The virtues of reduction in size, weight, power, and cost, regardless of the size of the spacecraft, however, are important. USAF will increasingly rely on these technologies to prevent "growing" satellites to next, more expensive classes of launch vehicles. When DoD space is ready for micro-spacecraft, the technologies described here will enable their advent. To achieve the greatest benefit, optimization at all levels of the packaging hierarchy will be required. As long as MCMs are treated as fancy single-chip packages and merely mounted onto circuit boards, the realization of the paradigm is incomplete. On the other hand, heterogeneous 3-D MCM technologies, MEMS devices, and multifunctional structures will provide the technology infrastructure necessary to enable the construction of true micro-spacecraft without compromise.

---

[1]Neugebauer, C. *Electronics Packaging Forum*, Morris, J.E. (editor). Van Nostrand Reinhold, New York, 1990
[2]Little, M.J. and C. Thomas Moberly. *Signal Processing Systems Packaging -1*, Rome Laboratory Technical Report *RL-TR-92-95*, April 1992.
[3]Lyke, J.C. and R. Wojnarowski et al., "Three-dimensional Patterned Overlay High Density Interconnect (HDI) Technology", Journal of Microelectronic Systems Integration, 1(2), 1993.
[4]Lyke, J.C. "Ultra-High Density Interconnect" briefings and notes, 1992-1995.

THE IMPACT OF SPACE RADIATION REQUIREMENTS AND EFFECTS ON ASIMS

C. Barnes, A. Johnston and G. Swift
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109*

71165
230717
12P

ABSTRACT

The evolution of highly miniaturized electronic and mechanical systems will be accompanied by new problems and issues regarding the radiation response of these systems in the space environment. In this paper we will discuss some of the more prominent radiation problems brought about by miniaturization. For example, autonomous microspacecraft will require large amounts of high density memory, most likely in the form of stacked, multichip modules of DRAMs, that must tolerate the radiation environment. However, advanced DRAMs (16 to 256 Mbit) are quite susceptible to radiation, particularly single event effects, and even exhibit new radiation effects phenomena that were not a problem for older, less dense memory chips. Another important trend in microspacecraft electronics is toward the use of low voltage microelectronic systems that consume less power. However, the reduction in operating voltage also carries with it an increased susceptibility to radiation. In the case of application specific integrated circuits (ASICs) that are an integral part of application specific integrated microinstruments (ASIMs), advanced devices of this type, such as high density field programmable gate arrays (FPGAs) exhibit new single event effects (SEE), such as single particle reprogramming of anti-fuse links. New advanced linear bipolar circuits have been shown recently to degrade more rapidly in the low dose rate space environment than in the typical laboratory total dose radiation test used to qualify such devices. Thus, total dose testing of these parts is no longer an appropriately conservative measure to be used for hardness assurance. We also note that the functionality of micromechanical Si-based devices may be altered due to the radiation-induced deposition of charge in the oxide passivation layers.

INTRODUCTION

Risk mitigation through the application of quality assurance techniques to microelectromechanical systems (MEMS) will pose special problems for future spacecraft and satellites. The rapid evolution of new technologies that are attractive for space applications, the increased emphasis on use of commercial microelectronic parts to save cost and schedule, the very small market represented by both military and civilian space systems, and the emergence of new device failure phenomena all pose particular difficulties for insuring flight mission success. Nowhere is this more apparent than in the case of radiation effects, and the effort to establish radiation hardness assurance (RHA) for MEMS. In the case of other reliability phenomena, one can take advantage of other high volume, high reliability applications, such as automotive and medical, to leverage statistical measures of reliability, minimum cost and quick schedule, but radiation represents a unique requirement not encountered by other high volume application areas. In addition, the downturn in DoD hardened electronic part development, testing and acquisition has resulted in reduced availability of radiation hardened or even radiation tolerant electronic parts. With regard to the fabrication of microelectromechanical parts, most reliability phenomena can be detected or prevented by prudent use of statistical process control and various screening techniques. In the case of radiation effects, however, very often "improvements" in device performance characteristics achieved by altered processing steps can lead to increased radiation vulnerability which only becomes known after the fact during radiation testing. Lastly, many new, emerging technologies exhibit new and unexpected radiation effects that must be taken into account prior to insertion in space systems. In this paper, we briefly explore some of the radiation effects issues that will confront designers and users of advanced, microelectronic and microelectromechanical parts.

There are several ways this discussion can be organized; we have chosen to discuss radiation effects according to part type, beginning with a review of radiation effects in advanced bipolar devices.

---

*The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). Work was funded by the NASA Code QW sponsored Microelectronics Space Radiation Effects Program (MSREP).

ENHANCED LOW DOSE RATE EFFECTS IN BIPOLAR CIRCUITS

Relatively simple bipolar circuits such as operational amplifiers and comparators, form an essential part of many electronic circuits, hybrids and systems. Such devices will continue to play an integral role in the functionality of advanced devices such as MEMS. Thus, their radiation response will influence the ability of MEMS to survive the hostile space radiation environment.

Recently, it has been established [1-10] that many bipolar integrated circuits are much more susceptible to ionizing radiation at low dose rates (0.001 to 0.005 rad(Si)/sec) than they are at the high dose rates typically used for radiation testing of parts in the laboratory. Since the low dose rate regime is equivalent to that encountered in space, for these devices the standard laboratory radiation test at moderate to high dose rates is no longer conservative. The seriousness of this problem has led the Air Force to issue an Alert Concern for this effect. Because of the greater radiation sensitivity at very low dose rates, the only way to provide radiation hardness assurance to designers is to perform a radiation test at low dose rates which by its nature is very time consuming. Consequently, it is imperative that an RHA test be developed which can be performed at moderate to high dose rates. Because the physical mechanism for the enhanced low dose rate effect is not yet completely understood, it is not possible to propose a reliable RHA test. Herein, we will merely provide some examples of this effect, which serve to emphasize the seriousness of the enhanced low dose rate susceptibility problem.

Initial work [1] on the enhanced low dose rate (ELDR) effect suggested that as the dose rate was decreased, the enhancement effect saturated at around 10 rad(Si)/s and did not become any stronger at lower dose rates. However, expanded studies [2-10] of the ELDR effect clearly indicate that for many devices, saturation, if and when it occurs, must come at very low dose rates. For example, Figure 1 shows how input offset voltage, V_{os}, of the LM324 operational amplifier depends on total dose at different dose rates. Note that while there is little change in V_{os} for dose rates as low as 0.005 rad(Si)/s, there is a dramatic decrease in offset voltage at low doses when the LM324 is irradiated at 0.002 rad(Si)/s. The change in V_{os} may come at even lower doses if one were to expose the part at 0.001 rad(Si)/s. To understand how impractical an RHA test becomes under these conditions, suppose that a mission has a total dose requirement of 15 krad(Si) and testing must be done at 0.001 rad(Si)/s to obtain the lowest failure dose because of the type of behavior shown in Figure 1. In order to test to the mission requirement, the radiation exposure would last for nearly 6 months, a test time that would be prohibitive for many fast, aggressive flight projects.

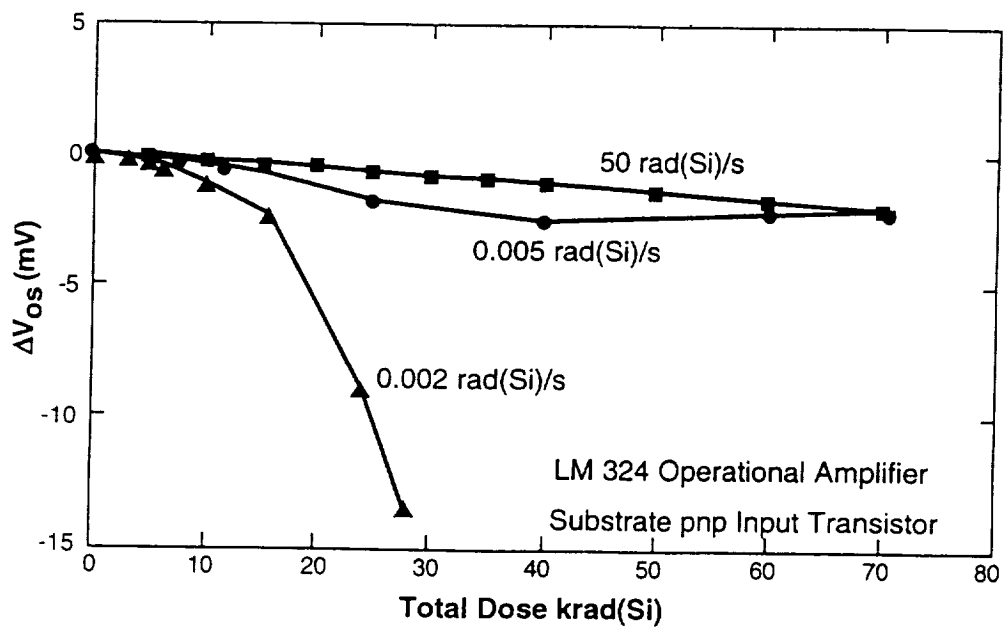


Figure 1. Degradation of input offset voltage of the LM324 operational amplifier at various dose rates.

Studies of the response of individual bipolar transistor types to different dose rates have shown that the problem is most severe for *pnp* transistors used in many linear circuits with conventional junction isolation. Damage in *pnp* devices can be 6 to 7 times greater at low dose rates than at high dose rates [4]. In contrast, *npn* transistors from the same fabrication processes are generally not sensitive to dose rate effects below approximately 1 rad(Si)/s. Thus, circuits which use both types of components may exhibit different

failure modes at low and high dose rates because of the different amount of relative damage that occurs in the two types of component transistors at low dose rates. In addition, different bipolar circuits that use varying mixes of the two transistor types can exhibit different failure characteristics, even for parts from the same process line. For example, Figure 2 compares input bias current degradation of two bipolar circuit types with similar input stage designs from the same manufacturer. Dose rate effects are relatively easy to evaluate in linear circuits with *pnp* input transistors because the input bias current provides a straightforward way to measure input transistor gain degradation. For the LM111 and LM324 shown in Figure 2, both devices use substrate *pnp* transistors, but the typical value of input bias current is three times greater for the LM111 than for the LM324. Although both circuits are more damaged when they are irradiated at low dose rate, degradation in the LM324 is much greater. Damage in the LM111 saturates at relatively low total dose levels, reducing the significance of enhanced damage. Input bias current of the LM324 continues to degrade at low dose rate as the radiation level increases, and consequently it is well above the specification limit even at 10 krad(Si). Even higher damage occurred in this device at 0.002 rad(Si)/s, although this is not shown in Figure 2. These results show that large differences can occur between different circuit types produced by the same manufacturer, and that it is risky to make blanket assessments about dose rate effects on the basis of tests on a small number of device types.

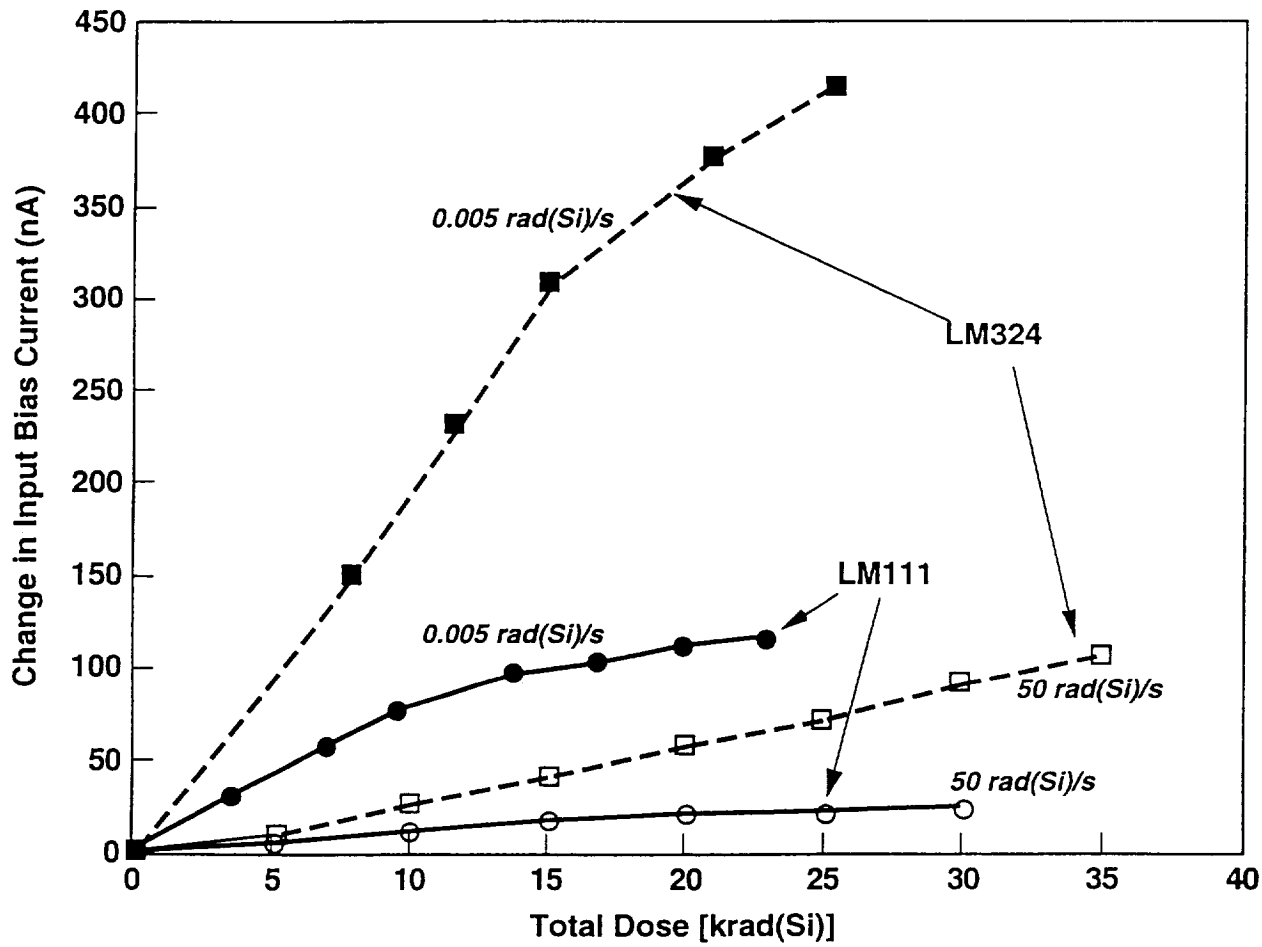


Figure 2. Degradation of input bias current of two different bipolar circuits with similar input stages from the same manufacturer.

It has also been shown [4] that the same device type from different manufacturers can exhibit significantly different ELDR effects. For example, Figure 3 shows input bias current degradation for LM111 voltage comparators, which use substrate *pnp* input transistors, procured from three different vendors, and tested at two widely different dose rates. For two of the manufacturers, damage of the input transistors is about six times greater at low dose rates so that they exhibit rapid increases in input bias current at the lower dose rate of 0.005 rad(Si)/s. Devices from the third manufacturer (vendor A) show only a small increase in damage at the lowest dose rate, even though the geometry of the input transistors of this vendor are identical to that of the vendor with the highest damage at low dose rates. Thus, determination of the extent of the ELDR effect for a particular device type is not sufficient to provide RHA if a different vendor is selected than the one used for radiation test samples.

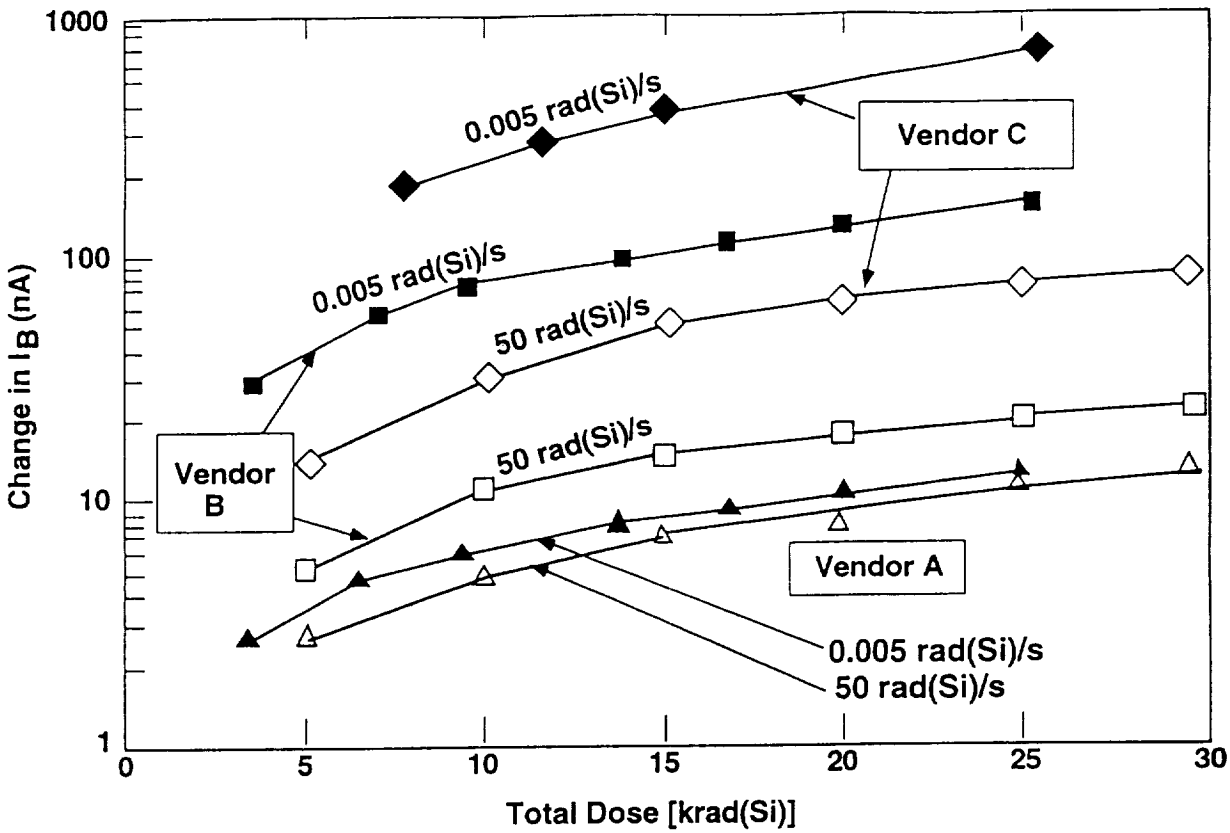


Figure 3. Total dose degradation of LM111 comparators from three different manufacturers at high and very low dose rates.

The remaining challenge for the radiation effects community is to devise an RHA test for the ELDR effect other than direct testing at very low dose rates, an expensive and time consuming alternative which is unacceptable, but necessary in some cases at this time. While the physical mechanism of the ELDR effect has not been completely defined, promising models are being developed [3,6] that are associated with the dynamics of radiation-induced electrons and holes in thick oxides with weak electric fields. One of the approaches suggested by these models is irradiation at moderate to high dose rates at elevated temperature. Recent work [3,7] has shown that the increased damage at low dose rates can be at least partially reproduced by irradiating at high dose rates and at 60°C to 100°C for certain device types from particular vendors. However, it is not clear what the temperature should be for a given process technology, or whether or not saturation of the increase in damage will occur at a reasonable temperature. Until the high temperature RHA test is better defined, another way to deal with this problem is to require tests at two different dose rates, selecting the lower dose rate so that it is sufficiently high to allow tests to be completed in days or weeks instead of the extremely long time periods imposed by the very low dose rates discussed above. Although this is not a substitute for doing tests at very low dose rates for devices that have severe enhanced damage at low dose rate, it is a more pragmatic way to identify devices that have minimal sensitivity to dose-rate effects. JPL has implemented this approach for several devices, using 0.02 rad(Si)/s for the lowest dose rate.

CATASTROPHIC SINGLE EVENT EFFECTS IN FPGAs

Field programmable gate arrays (FPGAs) are enjoying rapidly expanding usage in advanced electronic systems because of their performance and versatility. In many cases these devices can replace an entire board populated with discrete devices, thus leading to weight and power savings in spacecraft systems. In addition, field programmable devices offer a versatility that is important for low volume users such as civilian and military spacecraft and satellites. The increased insertion of FPGAs in space systems will require that their performance in radiation environments be well-established. In this section we describe a new single particle-induced dielectric rupture phenomenon that may become an increasingly important single event effect as scaling reduces the characteristic dimensions of FPGAs and similar devices. Because many types of MEMS include dielectric layers (in capacitors and transistors, for example) with voltages across them, this rupture mechanism is relevant to radiation hardness of MEMS technologies.

Recent SEE testing and research [11-13] on FPGAs has revealed a new catastrophic failure mechanism in these devices, termed single event dielectric rupture (SEDR), an effect similar to single event gate rupture (SEGR) in power MOSFETs [14,15]. In the Actel A1280 FPGAs that were studied, roughly half the silicon real estate on the chip is devoted to logic modules, while the other half consists of the interconnection matrix. As shown in Figure 4, the matrix consists of horizontal and vertical conductors with an anti-fuse at each crossing point. The anti-fuse structure, shown in Figure 5, consists of a thin (approximately 120 Å) sandwich of oxide-nitride-oxide (ONO). The conductors are connected at crossing points (anti-fuses) by electrically inducing dielectric breakdown of the ONO to establish a low resistance connection. A typical programmed FPGA design will have a few percent of the possible total number of anti-fuses (about 650,000 for the A1280) connected to establish the intended functionality of the chip. A random unconnected anti-fuse will be biased when the logic levels on the two crossing conductors are different, the occurrence of which depends on the duty cycle and phase of the two signals on these lines. When the bias is present, say at a level of 5.5 V, the electric field across the thin ONO anti-fuse is approximately 6 MV/cm. As in the case of a power MOSFET gate oxide, this field is large enough to result in rupture of the oxide due to the passage of a heavy, energetic charged particle. The effect of the resulting partial connection of the conductors depends on the surrounding circuitry and can be either benign or can compromise the functionality of the FPGA circuit.

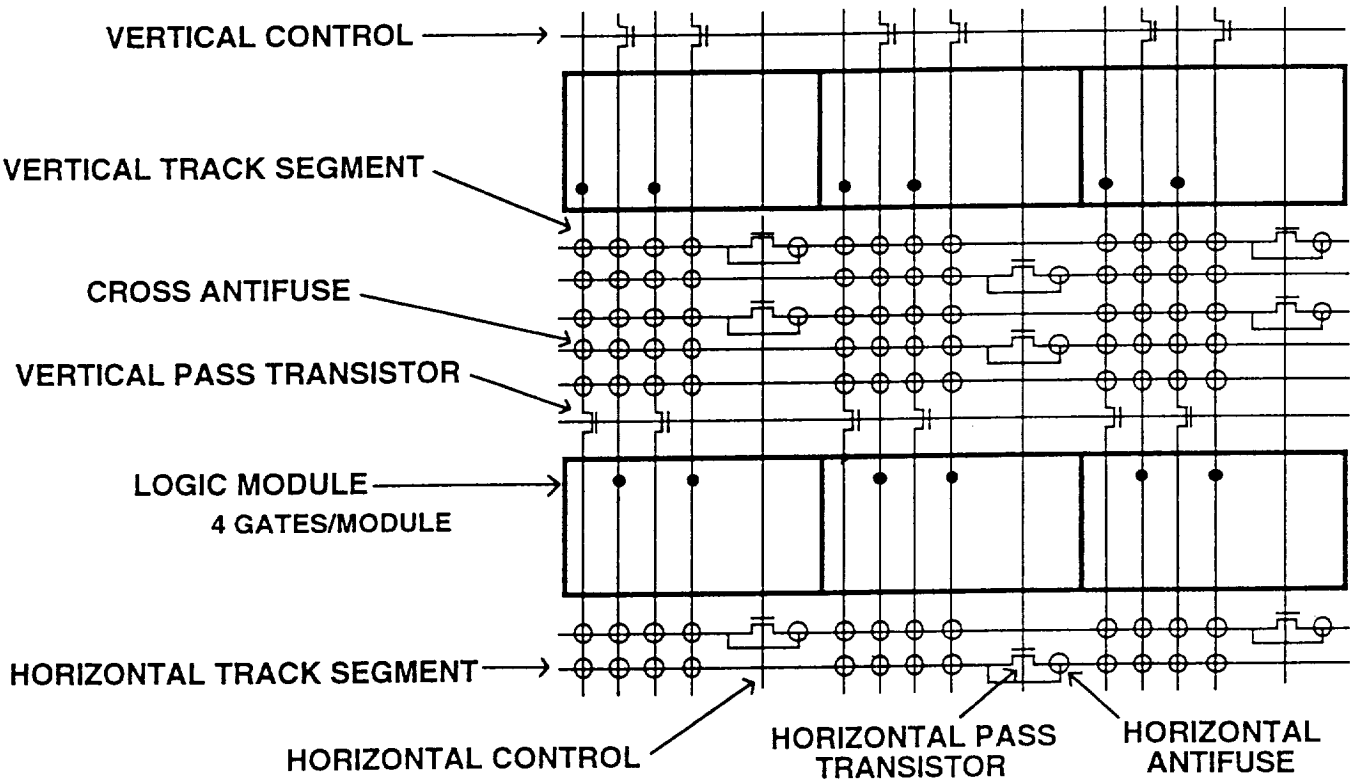


Figure 4. Anti-fuse connection architecture for Actel A1280 field programmable gate array.

Extensive accelerator testing with heavy ions has revealed the important features of the SEDR effect in the Actel A1280 FPGAs [11-13]. As one might expect, an ion-induced anti-fuse rupture is accompanied by an increase in total current as shown in Figure 6, taken during bombardment of an A1280 with iodine ions (linear energy transfer = LET = 60 MeV-cm^2/mg = 0.6 pC/μm). A rough count of SEDR anti-fuse events can be taken by merely counting the steps in the current curve shown in Figure 6. The presence of single ion-induced anti-fuse ruptures was also confirmed by locating high current flow nodes with emission microscopy for multilayer inspection (EMMI) on bombarded FPGAs. In addition, a VLSI circuit tester was used to perform I_{DDQ} tests which isolate current increases in each of the shift registers in the device. The experimental matrix demonstrated that anti-fuse rupture depends strongly on ion type (LET), bias and angle. Interestingly, and fortunately for space applications of FPGAs, the angular dependence is the opposite of that of traditional SEU: anti-fuse rupture falls off sharply with angle. The rupture effect does not appear to depend strongly on temperature, operating frequency, burn-in, lot-to-lot, or wafer-to-wafer variation.

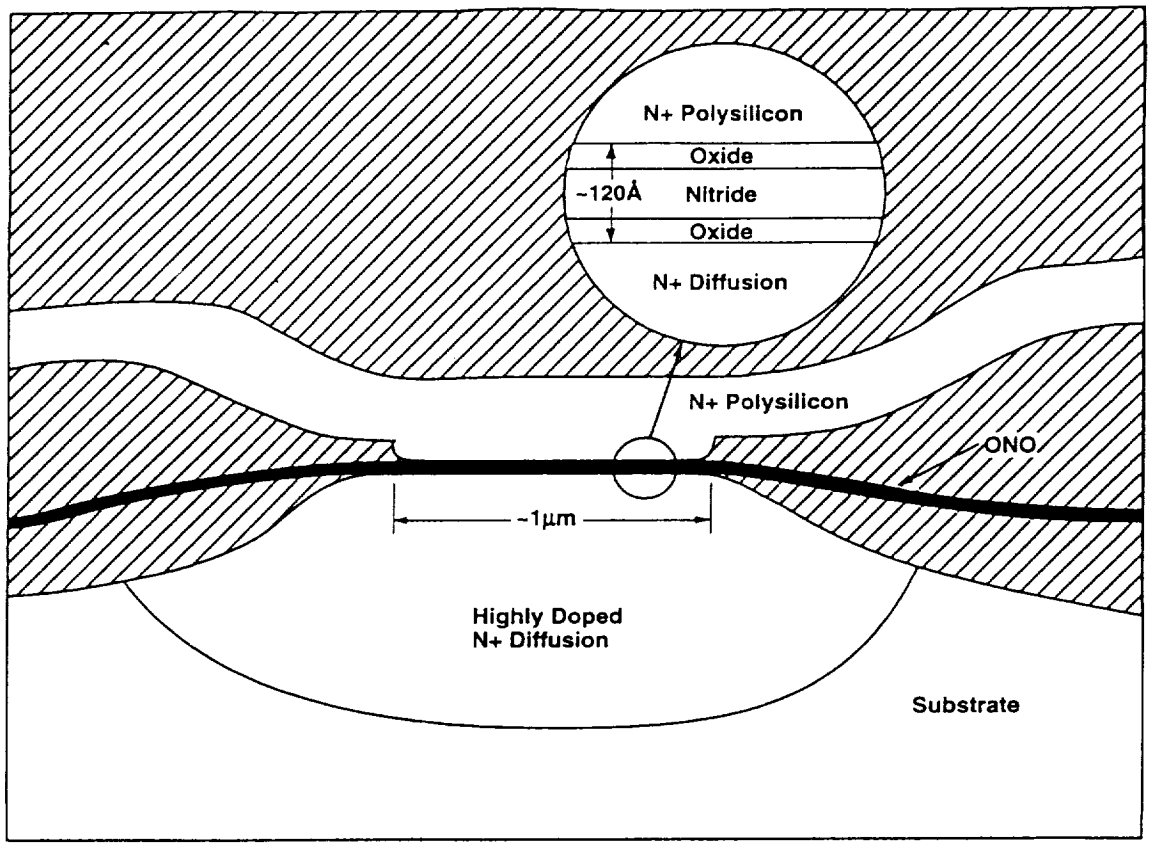


Figure 5. Structure of oxide-nitride-oxide (ONO) anti-fuse link in the Actel A1280 field programmable gate array.

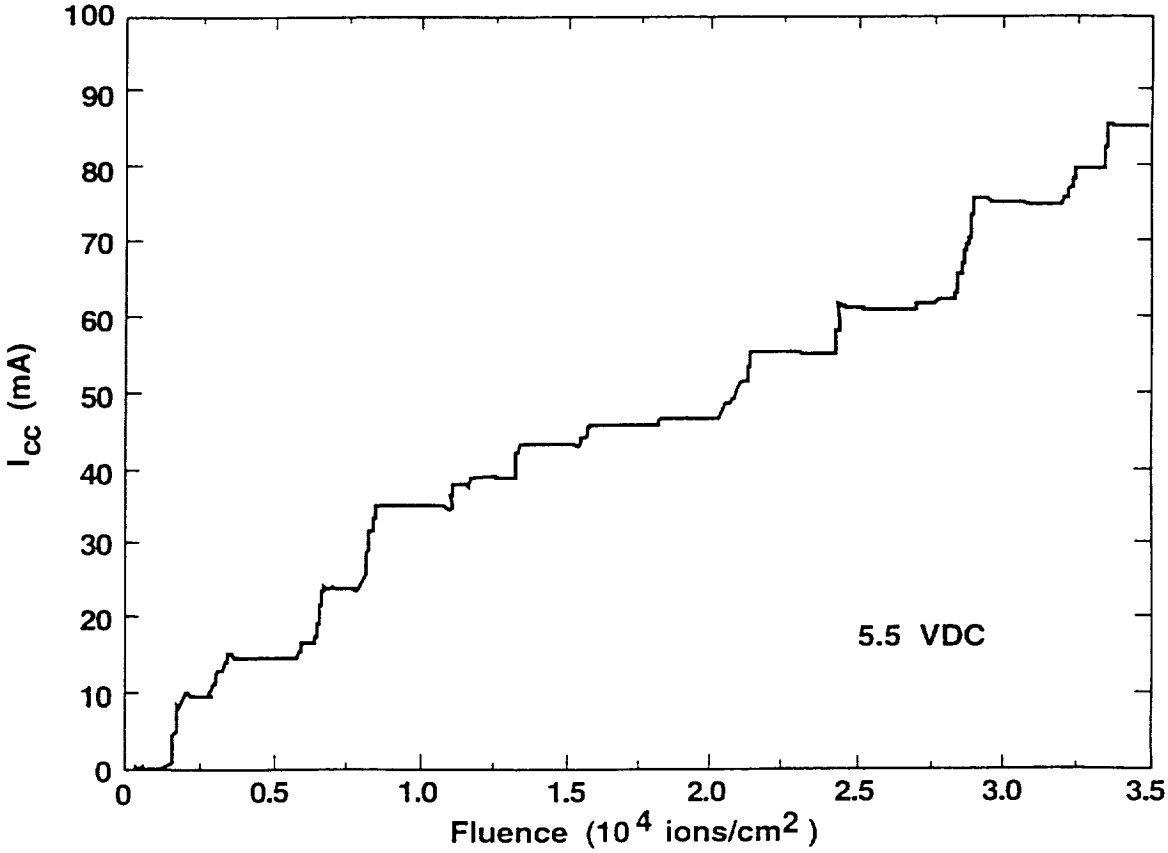


Figure 6. Current increases due to anti-fuse rupture in the Actel A1280 field programmable gate array during ion bombardment with iodine ions. Each step corresponds to rupture of an anti-fuse.

NEW RADIATION PHENOMENA IN ADVANCED DRAMS

Perhaps more than any other class of advanced microelectronic device, DRAMs represent a challenge with regard to their insertion in space systems. Advanced DRAMs are particularly attractive for mass storage systems on board spacecraft and satellites, so that it is tempting to incorporate them in large memory designs. However, because the radiation tolerant space applications market is vanishingly small compared to the commercial market for DRAMs, the designer is restricted to using commercial devices with little chance of device or process modification to accommodate radiation requirements. Thus, it is imperative to perform extensive radiation testing to determine if the commercial device under consideration will meet specific mission requirements. Unfortunately, DRAM advancements are proceeding so rapidly that by the time one has expended considerable effort and money performing Single Event Effects (SEE) and Total Ionizing Dose (TID) effects radiation tests, the DRAM under test is obsolete and essentially unavailable. In addition, scaling effects in newer devices can render radiation test results inapplicable to the latest DRAM technology. Add to this the new phenomena we discuss below and one is faced with a very challenging assurance problem for space applications of advanced DRAMs.

In the past few years, several studies [16-26] of DRAMs and SRAMs have revealed new radiation effects phenomena that can affect solid state memory performance in radiation environments. As we will show, technology advancements in the form of scaling to achieve reduced feature sizes and greater memory cell densities will probably exacerbate these effects.

State-of-the-art DRAMs are complex devices that often incorporate much more circuitry than simple DRAM cells and the necessary peripheral circuitry for accessing and writing/reading these cells. In particular, several operating modes may be built into the device in order to facilitate various test modes of operation. Normally a command sequence is used to place the DRAM in a test mode, but recent SEE testing [16] has revealed that an energetic heavy ion can throw the DRAM out of normal operating mode and into one of its test modes if the particle hits the circuit device that controls operating mode. These so called Single Event Functionality Interrupt (SEFI) events render the device inoperable until it is placed back in normal mode by specific commands. Unlike a traditional Single Event Upset (SEU), which results in a single error in one logic element, SEFIs manifest themselves as bursts of large numbers of errors due to a single particle hit. Thus, although the interaction cross section for a SEFI is less than that for a typical SEU, the effect has much greater impact on DRAM functionality. Fortunately, the effect is not permanent and can be removed by commanding the device to reenter standard operating mode. Thus, it is not permanently catastrophic in contrast with an important new class of single event hard errors we will discuss next.

During the last few years, several laboratories have observed single particle induced errors which are permanent in nature in contrast with the more traditional SEU event which is a temporary change in logic state of a circuit logic element. These "stuck bits" or "hard errors" remain in the device despite power cycling or reloading of a memory. More recently, work at various heavy ion accelerators [18-21,23-25] has led to a tentative identification of the nature of these hard errors. Earlier work [18,19] suggested that the only mechanism for hard error formation was a so-called "microdose" effect. This effect is the single particle/single transistor equivalent of a Co^{60} or total dose irradiation, and is due to microscopic ionization damage from the passage of a single ion (or a small number of ions) within the transistor gate region. As the result of continued scaling, transistors in advanced memories are now small enough so that a single ion can deposit enough ionizing energy to create the equivalent of a "total dose" effect within a single transistor. This effect is primarily important for DRAMs or 4-transistor memory cell SRAMs that use dynamic storage, but is not expected to be significant for random logic or 6-transistor memory cell SRAMs. In addition, it has been established that at least some fraction of these microdose-induced hard errors will anneal, as one might expect based on the characteristics of typical TID damage. Thus, these hard errors will recover slowly, unlike the second type described below.

The second mechanism causes catastrophic shorting of the gate oxide in the specific transistor that is struck by the energetic heavy ion [21]. It is attributed to gate rupture, and occurs because the high charge density produced by the ion track reduces the electric field needed to cause dielectric breakdown of the gate inslator. The result is destructive, permanent damage caused by the combination of applied field and heavy ion strike. Similar effects have been studied for several years in power MOSFETs [14,15], but they have only recently been observed for lower voltage transistors in VLSI devices. Theoretically, gate rupture can occur for random logic as well as memories, and may prevent the use of extremely scaled (miniaturized) devices in space. Note also that this effect is similar to the anti-fuse SEDR effect described above for FPGAs.

Heavy ion data, shown in Figure 7, taken on commercial 4-Mb DRAMs illustrates the striking difference between the two hard error mechanisms [21]. The distribution of lost bits as a function of DRAM retention time for iodine bombardment agrees with observations after total dose irradiation in which the distribution "walks" to the left in Figure 7 so that fewer and fewer bits have long retention times above the part specification of 360 ms at 45°C. In contrast, for gold bombardment two different types of data are observed. In addition to results similar to the iodine induced microdose distribution, the Au also causes permanent damage that is characterized by extremely short retention times, essentially zero. As in the case of power MOSFET gate rupture and the FPGA rupture effect described above, the angular dependence of the permanent damage is unlike traditional SEU. The natural separation into "lost ones" and "lost zeros" is also suggestive of a permanent effect that requires a bias on the transistor to cause damage. This second mechanism is particularly important for several reasons: 1) it is permanent and does not anneal so that over the course of a long mission, these errors will accumulate; 2) this effect causes catastrophic failure of the transistor in which it occurs; 3) this type of hard error could occur in any MOS transistor with bias on the gate so that it will also affect microprocessors and random logic circuits which are not amenable to software-based error detection and correction; and 4) this effect will become worse with continued part scaling that is sure to take place as circuits with greater and greater performance are brought into the market place.

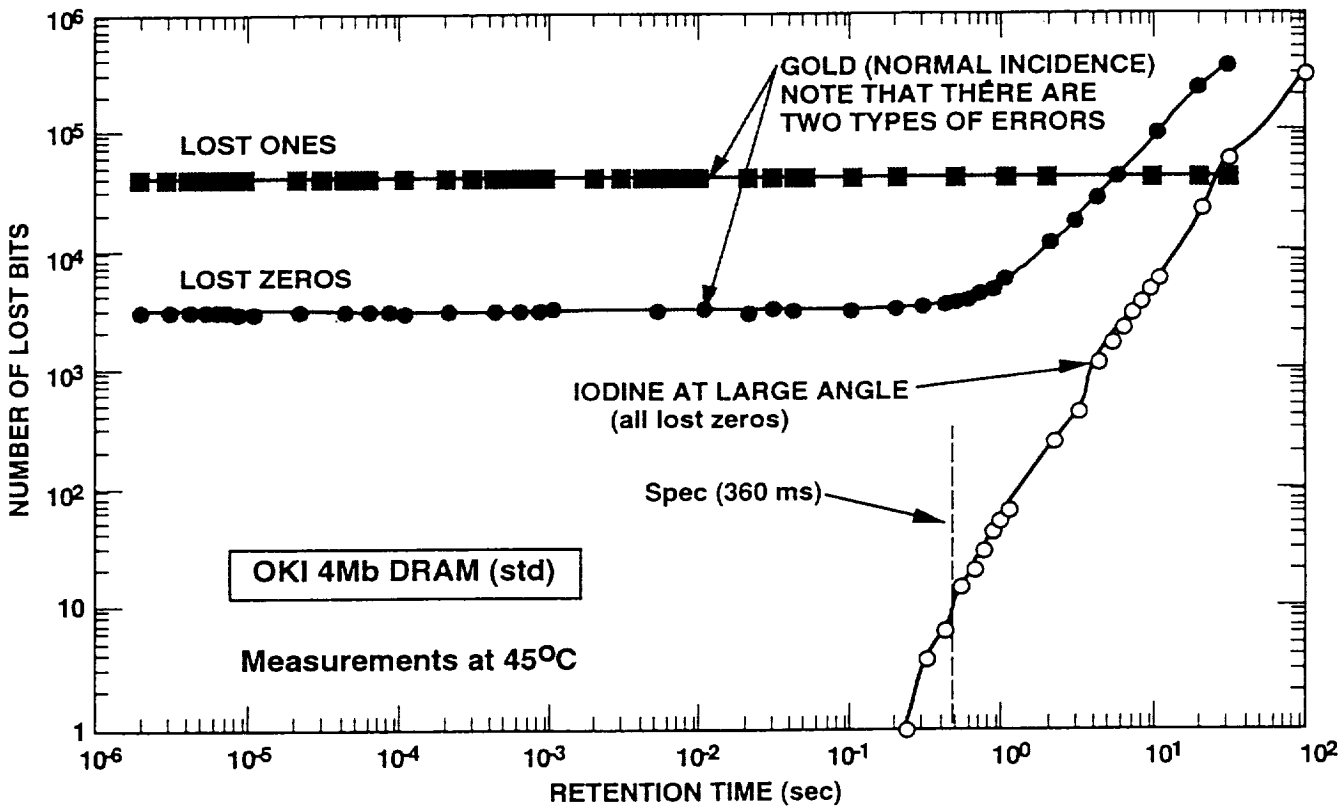


Figure 7. Retention time distributions for iodine (LET = 0.6 pC/μm) and gold (LET = 0.8 pC/μm) bombarded OKI 4 Mb DRAMs.

To illustrate the potentially negative effects of the permanent hard error mechanism on scaled devices, JPL examined a second set of 4 Mb DRAMs from the same manufacturer that had 20% smaller dimensions and thinner gate oxides. The scaled devices were more sensitive to the permanent damage effect in that the effect was observed for iodine at a lower LET than for gold. The damage thresholds, expressed in terms of the linear charge density deposited by the ions, are shown in Figure 8. The dashed line shows a prediction of the effect of future scaling changes, assuming that the threshold for gate rupture varies as the reciprocal of the square of the electric field across the oxide, using oxide fields typical of high speed scaling techniques.

In order to understand the implications of these results, one must take into account the distribution of the ions making up galactic cosmic rays (GCRs), which decrease rapidly for high linear charge densities. Above 0.3 pC/μm, the distribution falls abruptly by more than three orders of magnitude, the so-called iron ion threshold that is characteristic of the GCR spectrum. These results have been used to calculate the catastrophic hard-error rate for devices with different feature sizes, as shown in Figure 9. The error rate in

Figure 9 represents the number of failed devices for a VLSI circuit with approximately one million transistors. There is a pronounced difference in the error rate for the high speed scaling approach versus the low power scaling method because of the differences in the way the gate oxide field scales in each case. The very rapid increase in the error rate of the high speed scaling curve occurs when the threshold charge density falls below 0.3 pC/µm, where there is a large jump in the number of GCR particles. The curve suggests that it will be very difficult to use devices with 0.25 µm feature size because of the hard error limit, and that devices optimized for low power operation will have a much lower error rate in space applications.

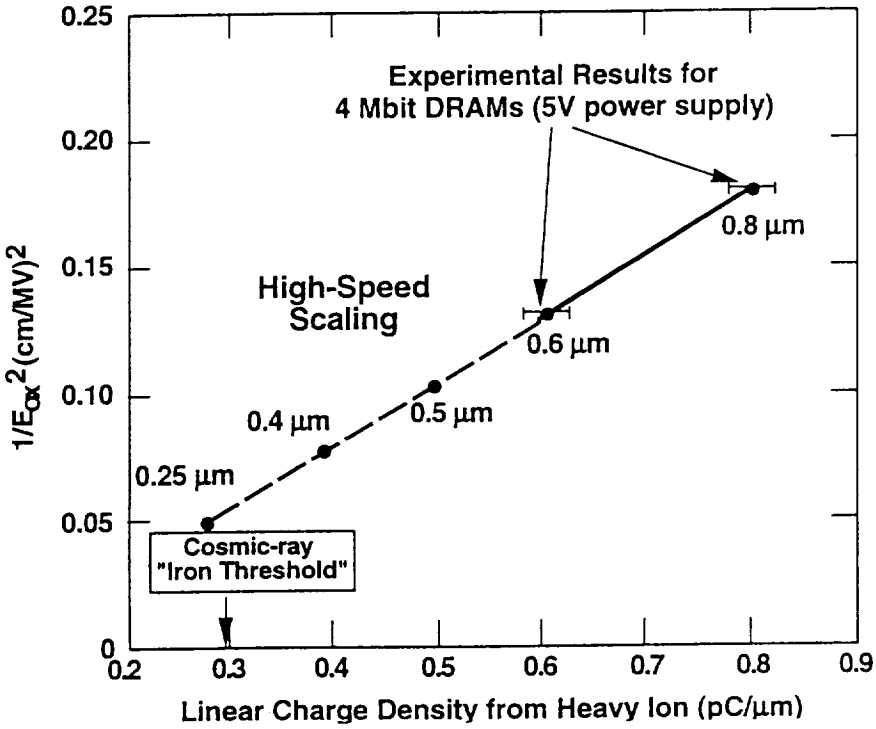


Figure 8. Effect of device scaling (miniaturization) on hard error threshold for various feature sizes.

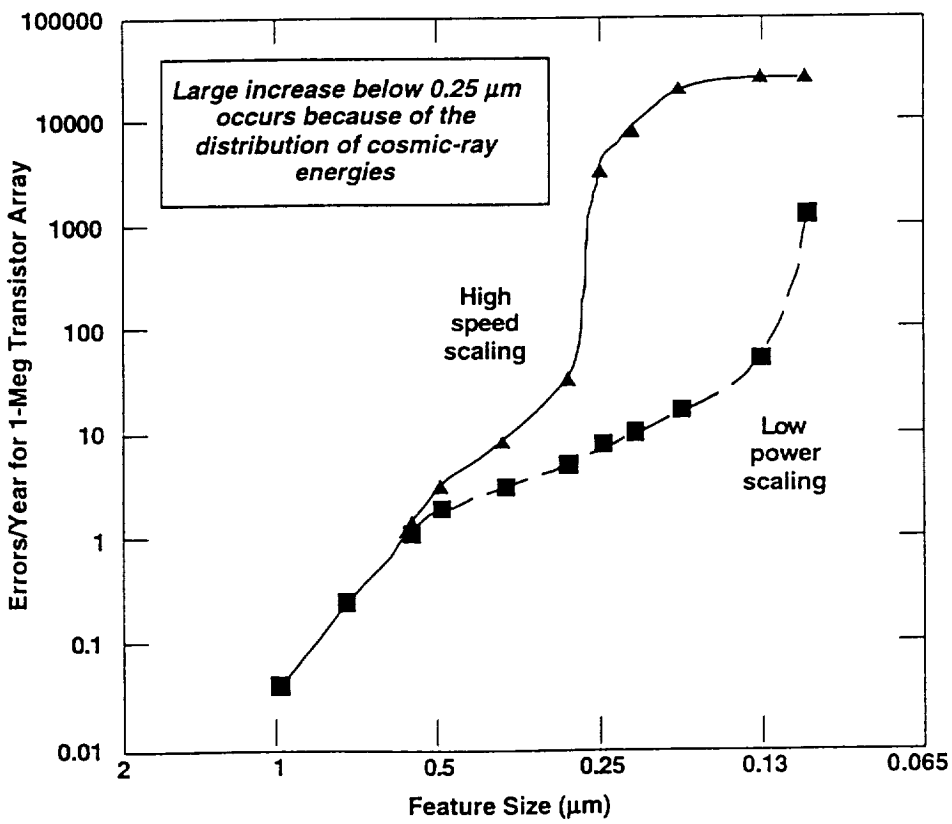


Figure 9. Hard errors induced by cosmic rays for a 1-M transistor array as feature sizes are reduced.

IMPLICATIONS FOR EMERGING TECHNOLOGIES

Thus far, we have discussed important new radiation effects phenomena that will affect the radiation susceptibility of advanced microelectronic and electromechanical systems in spacecraft and satellites. In addition, we have seen that advances in performance in traditional electronic technologies, such as DRAMs and microprocessors, which are achieved through scaling (miniaturization) of feature sizes, will result in devices and circuits that are more sensitive to radiation effects. We wish to close our discussion by emphasizing that care must also be taken in the insertion of new, emerging technologies in space systems in order to avoid radiation effects problems that may jeopardize mission success. In many cases, the use of concurrent engineering to examine potential radiation problems at early design stages will catch problems before they reach a stage that will result in considerable cost and schedule penalties to repair. One example concerns the insertion of fiber optic data links in high data rate space applications. The use of fibers and Si detectors at the so called "first window" at a wavelength of approximately 850 nm is inappropriate for a radiation environment because these fibers and detectors are more sensitive to radiation than the fibers and III-V-based detectors used at the second (1.3 μm) and third (1.55 μm) windows.

As we have noted above, many MEMS technologies involve the use of insulators in some fashion, and these insulators may have electric fields across them. In such cases, ionizing radiation and heavy, energetic ions can alter the operating characteristics of devices that depend on these insulator structures for functional operation. In addition, in most MEMS the core element, such as a microaccelerometer or microgyro, is on a chip that also includes a variety of more traditional electronic devices. Thus, even though the heart of the MEMS may not be sensitive to radiation, it is possible that the MEMS functionality will be disrupted in a radiation environment. In the rush to fly these powerful new miniaturized technologies we must not forget that their success can be jeopardized by radiation effects that can sometimes occur in the MEMS themselves, or in the more mundane technology devices that facilitate the operation of the MEMS.

REFERENCES

1. R. Nowlin, D. Fleetwood and R. Schrimpf, "Saturation of the dose rate response of bipolar transistors below 10 rad(SiO_2)/s: implications for hardness assurance", *IEEE Trans. Nuc. Sci.* NS-41, 2637 (1994).

2. S. McClure, R. Pease, W. Will and G. Perry, "Dependence of total dose response of bipolar linear microcircuits on applied dose rate", *IEEE Trans. Nuc. Sci.* NS-41, 2544 (1994).

3. D. Fleetwood, S. Kosier, R. Nowlin, R. Schrimpf, R. Reber, Jr., M. DeLaus, P. Winokur, A. Wei, W. Combs and R. Pease, "Physical mechanisms contributing to enhanced bipolar gain degradation at low dose rates", *IEEE Trans. Nuc. Sci.* NS-41, 1871 (1994).

4. A. Johnston, G. Swift and B. Rax, "Total dose effects in conventional bipolar transistors and integrated circuits", *IEEE Trans. Nuc. Sci.* NS-41, 2452 (1994).

5. D. Schmidt, R. Graves, R. Schrimpf, D. Fleetwood, R. Nowlin and W. Combs, "Comparison of ionizing radiation induced gain degradation in lateral, substrate, and vertical PNP BJTs", presented at the *32nd International Nuclear and Space Radiation Effects Conf.*, Madison, WI, July, 1995; to be published in *IEEE Trans. Nuc. Sci.* NS-42 (1995).

6. A. Johnston, B. Rax and C. Lee, "Enhanced damage in linear bipolar integrated circuits at low dose rate", presented at the *32nd International Nuclear and Space Radiation Effects Conf.*, Madison, WI, July, 1995; to be published in *IEEE Trans. Nuc. Sci.* NS-42 (1995).

7. R. Graves, D. Schmidt, R. Schrimpf, D. Fleetwood, R. Nowlin, W. Combs, R. Pease and M. DeLaus, "Hardness assurance issues for lateral PNP bipolar junction transistors", presented at the *32nd International Nuclear and Space Radiation Effects Conf.*, Madison, WI, July, 1995; to be published in *IEEE Trans. Nuc. Sci.* NS-42 (1995).

8. A. Johnston, C. Lee, B. Rax and D. Shaw, "Using commercial semiconductor technologies in space", presented at the *Third European Conf. On Radiation and Its Effects on Components and Systems (RADECS 95)*, Arcachon, France, September, 1995.

9. R. Schrimpf, "Recent advances in understanding total dose effects in bipolar transistors", presented at the *Third European Conf. On Radiation and Its Effects on Components and Systems (RADECS 95)*, Arcachon, France, September, 1995.

10. J. Bosc, G. Sarrabayrouse and F. Guerre, "Total dose effects on elementary transistors of a comparator in bipolar technology", presented at the *Third European Conf. On Radiation and Its Effects on Components and Systems (RADECS 95)*, Arcachon, France, September, 1995.

11. R. Katz, R. Barto, P. McKerracher, B. Carkhuff and R. Koga, "SEU hardening of field programmable gate arrays (FPGAs) for space applications and device characterization", *IEEE Trans. Nuc. Sci.* NS-41, 2179 (1994).

12. R. Katz, G. Swift and R. Barto, "Anti-fuse reliability in the heavy ion environment", presented at the *32nd International Nuclear and Space Radiation Effects Conf.*, Madison, WI, July, 1995.

13. G. Swift and R. Katz, "An experimental survey of heavy ion induced dielectric rupture", presented at the *Third European Conf. On Radiation and Its Effects on Components and Systems (RADECS 95)*, Arcachon, France, September, 1995.

14. H. Joseph, "Radiation-induced effects on special superlarge *n*-channel power MOSFETs for space applications", p. 128 in *Proc. of First European Conf. On Radiation and Its Effects on Components and Systems (RADECS 91)* (1992).

15. D. Nichols, J. Coss and K. McCarty, "Single event gate rupture in commercial power MOSFETs", p. 462 in *Proc. of Second European Conf. On Radiation and Its Effects on Components and Systems (RADECS 93)* (1994).

16. G. Swift, L. Smith and H. Schafzahl, "Single event functionality interrupt: a new failure mode for semiconductor memories", presented at the *9th Single Event Effects Symposium*, Manhattan Beach, CA, April, 1994.

17. R. Koga, W. Crain, K. Crawford, D. Lau, S. Pinkerton, B. Yi and R. Chitty, "On the suitability of non-hardened high density SRAMs for space", *IEEE Trans. Nuc. Sci.* NS-38, 1507 (1991).

18. C. Dufour, P. Garnier, T. Carriere, J. Beaucour, R. Ecoffet and M. Labrunee, "Heavy ion induced single hard errors on submicronic memories", *IEEE Trans. Nuc. Sci.* NS-39, 1693 (1992).

19. T. Oldham, K. Bennett, J. Beaucour, T. Carriere, C. Polvey and P. Garnier, "Total dose failures in advanced electronics from single ions", *IEEE Trans. Nuc. Sci.* NS-40, 1820 (1993).

20. S. Duzellier, D. Falguere and R. Ecoffet, "Protons and heavy ions induced stuck bits on large capacity RAMs", p. 468 in *Proc. of Second European Conf. On Radiation and Its Effects on Components and Systems (RADECS 93)* (1994).

21. G. Swift, D. Padgett and A. Johnston, "A new class of single event hard errors", *IEEE Trans. Nuc. Sci.* NS-41, 2043 (1994).

22. D. Shaw, G. Swift, D. Padgett and A. Johnston, "Radiation effects in five volt and advanced lower voltage DRAMs", *IEEE Trans. Nuc. Sci.* NS-41, 2452 (1994).

23. P. Calvel, P. Lamothe, C. Barillot, R. Ecoffet, S. Duzellier and E. Stassinopoulos, "Space radiation evaluation of 16 Mbit DRAMs for mass memory applications", *IEEE Trans. Nuc. Sci.* NS-41, 2267 (1994).

24. D. Shaw, G. Swift, D. Padgett and A. Johnston, "Radiation evaluation of an advanced 64 Mb, 3.3 V DRAM and insights into the effects of scaling on radiation hardness", presented at the *32nd International Nuclear and Space Radiation Effects Conf.*, Madison, WI, July, 1995; to be published in *IEEE Trans. Nuc. Sci.* NS-42 (1995).

25. R. Harboe-Sorensen, S. Fraenkel, R. Muller and T. Braunschweig, "Heavy ion, proton and Co-60 radiation evaluation of 16 Mbit DRAM memories for space application", presented at the *32nd Interna-*

tional Nuclear and Space Radiation Effects Conf., Madison, WI, July, 1995; to be published in *IEEE Trans. Nuc. Sci.* NS-42 (1995).

26. A. Johnston, G. Swift and D. Shaw, "Impact of CMOS scaling on single event hard errors in space systems", presented at the *1995 Symposium on Low Power Electronics*, San Jose, CA, October, 1995, p. 88 in Digest of Technical Papers from 1995 SLPE Conf., October, 1995.

3D thin film microstructures for space microrobots

Isao Shimoyama

Department of Mechano-Informatics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
Phone: 81-3-3812-2111 X6317 or 6318 Fax: 81-3-3818-0835
e-mail: isao@leopard.t.u-tokyo.ac.jp
URL: http://www.leopard.t.u-tokyo.ac.jp

51167
1p

Micromechanisms of a locomotion and a manipulator with external skeletons like the structure of an insect are proposed. These micromechanisms can be implemented using rigid plates and elastic joints. A large scale model consisting of plastic plates, springs and solenoids is shown to demonstrate this motion. Moreover, several microsized models were built on silicon wafers by using polysilicon as rigid plates and polyimide as elastic joints.

Due to scale effects, friction in micromechanical components is dominant compared to the inertial forces because friction is proportional to L^2 while mass is proportional to L^3. Therefore, a rotational joint that exhibits rubbing should be avoided in micromechanical components to ensure efficient motion.

Insects have external skeletons, elastic joints, distortions of the thorax, and contracting-relaxing muscles. The combination of an external skeleton and elastic joints produces friction-free movement of the skeletons at the elastic joints. In addition, there are several insect characteristics that provide clues for designing microrobots.

In this paper, paper models of a robot leg and a micromanipulator are initially presented to show the structures with external skeletons and elastic joints. Then, the large scale implementations using plastic plates, springs, and solenoids are shown. Since the assembly technique shown here is based on paper folding, it is compatible with thin film microfabrication and IC planar processes. Finally, several micromechanisms have been fabricated on silicon wafers to demonstrate the feasibility of building a 3D microstructure out of a single planar structure. These 3D structures can be used for space microrobots.

Text of full paper not available at time of publication.

Session 6: Space Applications I (Wednesday, AM)
Session Chair: Charles Sawin (NASA/JSC)

Development of a Space Bioreactor Using Microtechnology
by Philippe Arquint, Marc A. Boillat, Nico F. de Rooij, Sylvain Jeanneret and Bart H. van der Schoot, University of Neuchâtel,
Birgitt Bechler, Augusto Cogoli, and Isabelle Walther, Space Biology Group ETHZ,
Volker Gass and Marie-Thérèse Ivorra, Mecanex SA

Integrated Microreactor for Chemical and Biochemical Applications
by Norbert Schwesinger, L. Dressler, Th. Frank, and H. Wurmus,
Technical University of Ilmenau, Germany

Fluid Flow Volume Measurements Using a Capacitance/Conductance Probe System
by T. Nguyen and G. Arndt, NASA/Johnson Spacecraft Center,
and J. Carl, Lockheed

Biomorphic Architectures for Autonomous "Nanosat" Designs
by Brosl Hasslacher and M. Tilden, Los Alamos National Laboratory

DEVELOPMENT OF A SPACE BIOREACTOR USING MICROTECHNOLOGY

71169

Philippe Arquint[1], Birgitt Bechler[2], Marc A. Boillat[1], Augusto Cogoli[2], Nico F. de Rooij[1], Volker Gass[3], Marie-Thérèse Ivorra[3], Sylvain Jeanneret[1], Bart H. van der Schoot[1] and Isabelle Walther[2]

[1]Institute of Microtechnology, University of Neuchâtel,
Rue Jaquet Droz 1, CH-2007 Neuchâtel, Switzerland

[2]Space Biology Group ETHZ, CH-8005 Zürich, Switzerland

[3]Mecanex SA, CH-1260 Nyon, Switzerland

Abstract
A miniature bioreactor for the cultivation of cells aboard Spacelab is presented. Yeast cells are grown in a 3 milliliter reactor chamber. Supply of fresh nutrient medium is provided by a piezo-electric silicon micropump. In the reactor chamber, pH, temperature and redox potential are monitored and the pH is regulated at a constant value. The complete instrument is fitted in a standard experiment container of 63×63×85mm. The bioreactor has been used on the IML-2 mission in July 1994 and is currently being refurbished for a reflight in the spring of 1996.

Introduction
Experiments with cells in space have shown that important cellular functions are changed in microgravity [1]. These findings are of great interest for fundamental research as well as for possible biotechnological applications. For the cultivation of cells aboard spacelab, a miniature bioreactor is developed. The objective is to evaluate in a controlled bioreactor experiment (chemostat cultivation) the effects of mixing and stirring on the growth characteristics of yeast cells. The experiment has been selected by ESA for the IML-2 mission which took place in July 1994. The experiment was located in Biorack, a multi-user facility developed by ESA to host biological experiments in Spacelab. The individual experiments are mounted in standard containers of various dimensions. The Type II container allocated for this experiment has internal dimensions of 63×63×85mm.

The design of the Space Bioreactor includes the following features;

- Supply of fresh medium to cells in the reactor chamber (working volume 3 ml) over a period of 9 days at an adjustable rate (0.2 - 1.5 ml/h)
- Measurement of pH, temperature and redox potential of the culture
- Control of pH of the culture
- On-line data transfer to ground control

In view of the complexity of the instrument and the limited dimensions of the Type II container, the application of silicon technology to provide both the sensors and a micromachined pump and flow sensor are of a distinct advantage.

During its first flight on IML-2 in July 1994, the technical feasibility of the bioreactor concept has clearly been demonstrated. Due to excessive temperatures in the laboratory where the flight experiment had to be prepared, the experiment has not yet given all the expected biological results. Therefore, a reflight opportunity has been granted and the bioreactor is currently being prepared for

the STS-76 mission in the spring of 1996. In order to make the liquid handling more robust, a controlled flow system will be implemented.

Experimental Setup

Figure 1 shows the individual elements of the bioreactor and their interconnections. In the following section, these elements will be briefly described. It is interesting to note that the reactor chamber, the heart of the instrument, occupies less than 1% of the volume available in the experiment container. The largest volume is reserved for the nutrient medium reservoir. Both the fresh and used medium reservoir are flexible and while the first is emptied during the experiment, the second takes its place while filling up.

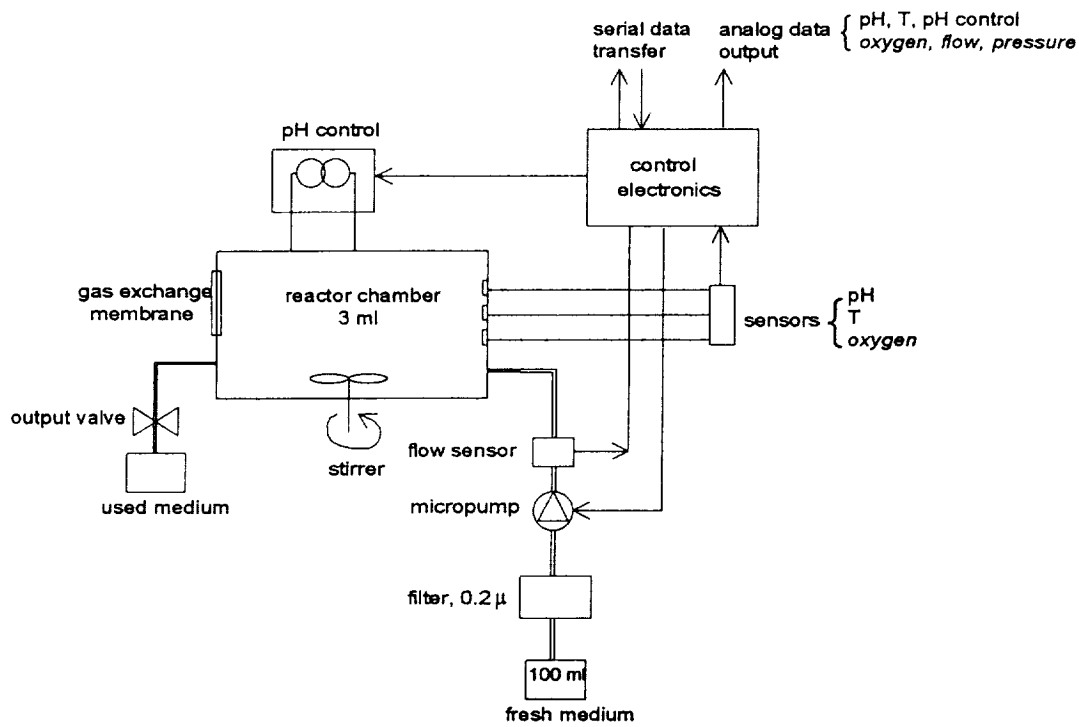


Figure 1. Principal elements of the miniature bioreactor

Pumping system

The supply of fresh medium to the reactor chamber is accomplished with a microfabricated pump [2, 3]. It is a membrane pump consisting of a structured silicon part sandwiched between two glass plates. The silicon forms two passive one-way valves that close on the thicker of the two glass layers. The thinner glass forms the membrane that is driven by a ceramic piezo-electric disk. When used as such, the flow rate of the pump is depending on the output pressure that has to be provided. Unforeseen variations in the bioreactor chamber pressure were the cause that during the first flight fluid delivery was not sufficiently controlled. For the reflight, a flow sensor is added so that the pump can operate in a closed loop control system and supply the required flow independent of output pressure [4].

Sensors

The sensors for the bioreactor, a pH-ISFET, a temperature sensitive diode and a thin film platinum redox electrode have all been integrated on a single chip that measures 3.5×3.5 mm. The sensor chip is mounted on a carrier that is inserted in the reactor chamber so that the sensor surface becomes part of the chamber wall without creation of dead angles. The reference electrode used in connection with the chemical sensors is a gel-filled Ag/AgCl electrode.

pH-control

For optimal growth conditions for the cells, the pH in the reactor has to be maintained at a constant value. During normal growth, yeast cells will acidify the medium for which a compensation is required. In view of the risk connected to the use of concentrated NaOH for this purpose, an electrochemical pH control is developed. A titanium electrode in the reactor chamber acts as a cathode for the electrolysis of water to produce hydroxyl ions. As a counter electrode a silver wire is used, embedded in a KCl loaded gel. The counter electrode compartment is separated from the reactor chamber by a Nafion membrane. Electrochemical compensation of the pH can be controlled very precise and avoids the need for an additional pump for NaOH dosage.

Electronic Circuits

The operation of the bioreactor is controlled by an Intel 87C51 microcontroller. The circuitry consists further of amplifiers for the sensors, a high voltage driver circuit for the micropump and a current source for pH-control. The pump operation is governed by an analog control system, the set-point of which is controlled by the microprocessor. The current source for pH-control is galvanically isolated from the rest of the circuit to avoid interference with the sensor signals. Relevant data are available on one multiplexed analog output channel. For preparation of the experiment before the flight a serial data connection is provided and the instrument can be directly interfaced to a standard personal computer. Figure 2 shows the basic electronic connections in the instrument. The circuits are built up with surface mounted components on two printed circuit boards

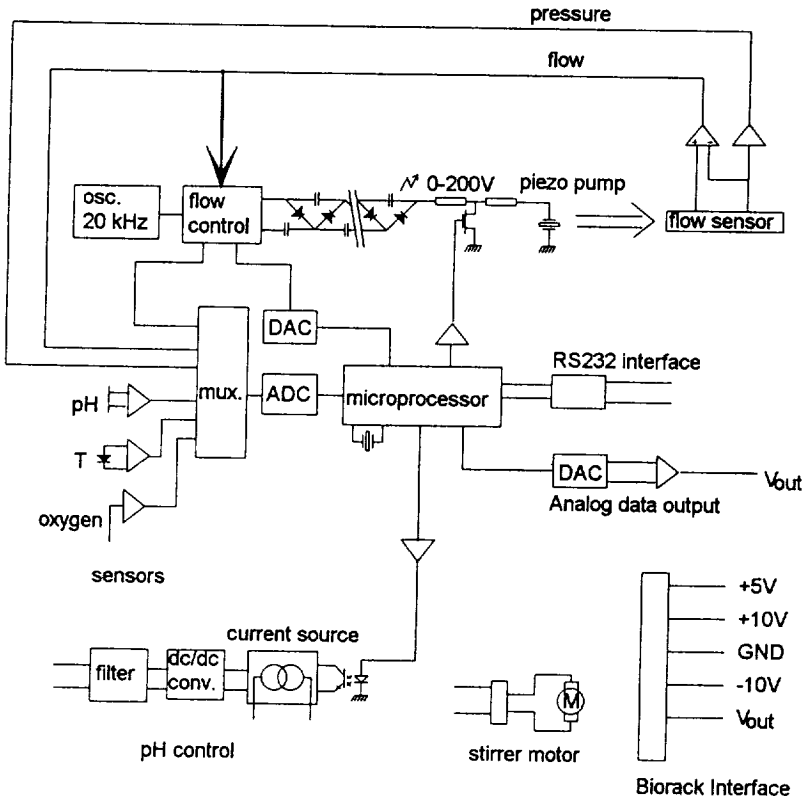


Figure 2. Electronic circuit of the bioreactor

Results

This project has started in 1991. After 15 months a breadboard model was presented to show the feasibility. After that the bioreactor has been qualified in vibration and EMC tests and, above all, in biological performance experiments as described in [5]. Figure 3 shows a photograph of an engineering model of the bioreactor that has been used for a number of the qualification tests. The first flight of the bioreactor has been relatively successful and currently the second flight is in preparation.

Conclusion

The development of a miniature bioreactor with microfabricated components for fluid handling and silicon sensors shows a promising outlook for their application in micro gravity experiments. In view of the special requirements for this kind of application, the bioreactor is an attractive showcase for microsystem technology.

Acknowledgment

The authors wish to thank the teams of the PRODEX program of ESA and of the Swiss Federal Office of Science and Education for their support in this project.

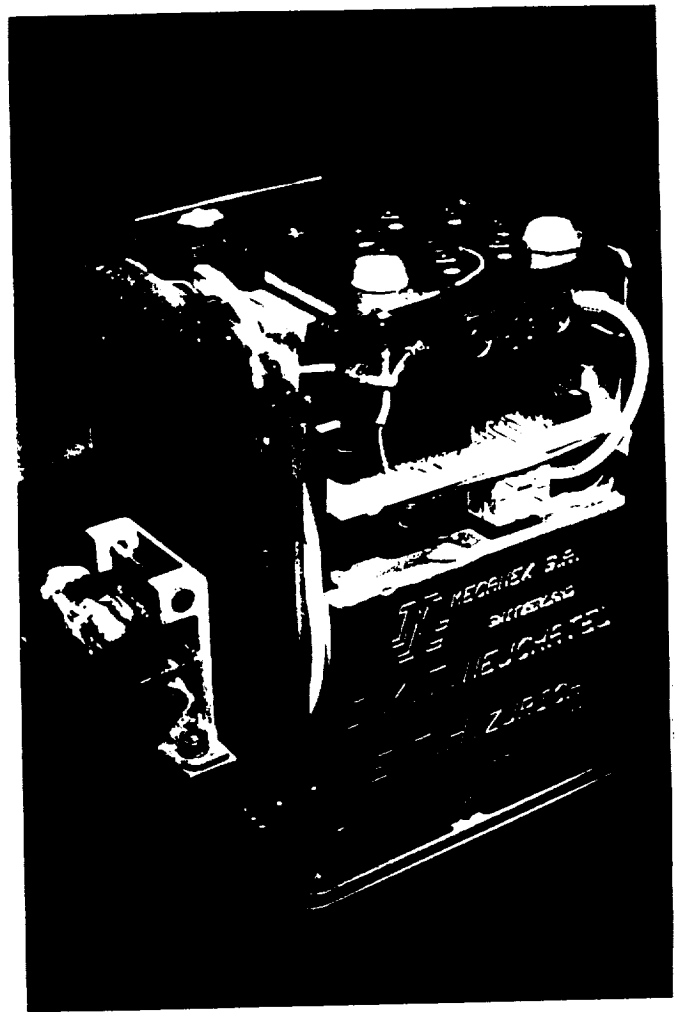


Figure 3. Engineering model of the space bioreactor mounted on the base of a Type II experiment container.

References

1 F. K. Gmünder, C.-G. Nordau, A. Tschopp, B. Huber and A. Cogoli, *J. Biotechnol.* 7 (1988) 217-228

2 H. T. G. van Lintel, F. C. M. van der Pol and S. Bouwstra, *Sensors and Actuators*, 15 (1988) 153-167

3 B. H. van der Schoot, S. Jeanneret, A. van den Berg and N. F. de Rooij, *Sensors and Actuators* B, 6 (1992) 57-60

4 M. A. Boillat, A. J. van der Wiel, A. C. Hoogerwerf, N. F. de Rooij, *Proceedings IEEE MEMS 1995*, pp 350-352, Amsterdam, the Netherlands, January 29-February 2, 1995

5 I. Walther, B. H. van der Schoot, S. Jeanneret, Ph. Arquint, N. F. de Rooij, V. Gass, B. Bechler, G. Lorenzi, A. Cogoli, *J. Biotechnology* 38 (1994) 21-32

INTEGRATED MICROREACTOR FOR CHEMICAL AND BIOCHEMICAL APPLICATIONS

N. Schwesinger, L. Dressler, Th. Frank, H. Wurmus

Technical University of Ilmenau, Faculty of Mechanical Engineering,
Department of Microsystemtechnology

7171
030030
12 P

Abstract

A completely integrated microreactor has been developed, that allows the processing of very small amounts of chemical solutions. The whole system comprises several pumps and valves that are arranged in different branches as well as a mixing unit and a reaction chamber. The streaming path of each branch contains two valves and one pump between them. The pumps are driven piezoelectrically using pizoceramic elements mounted on thin glass membranes.
Each pump has a dimension of about 3.5mm x 3.5mm x 0.7mm. A pumping rate up to 25 µl/h can be achieved. The operation voltage is in a range between 40 V and 200V. A volume stroke up to 1.5µm is achievable for the membrane structures
The valves are designed as passive valves. The sealing is made by the use of thin metal films. The dimension of a valve unit is 0.8mm x 0.8mm x 0.7mm.
The ends of the separate streaming branches are arranged to meet in one point. This point acts as the begin of a mixer unit. The unit contains several fork-shaped channels. The arrangement of these channels allows the division of the whole liquid stream into partial streams and their reuniting. A homogeneous mixing of solutions and/or gases can be observed after having passed about 10 fork elements.
A reaction chamber is arranged behind the mixing unit to support a chemical reaction of special fluids. This unit contains heating elements placed at the outside of the chamber.
The complete system is arranged in a modular structure and built up of silicon. It comprises three silicon wafers bonded together applying the silicon direct bonding technology. The structures in silicon are made only by the use of wet chemical etching processes. The fluid connections to the outside are realised using standard injection needles that are glued into v-shaped structures of the silicon wafers.
An integration of further components, like sensors or electronic circuits, is possible due to the use of silicon as basic material.

1. Introduction

The development and synthesis of special agents and fine chemicals were connected in the past very often with a low efficiency. Big amounts of different chemicals and complicated processes were required to get a very small volume of a specific material. One of the basic reasons for that is the present state of the development of chemical devices. They are very often big, unable to work with small amounts of agents and not stable enough during process steps. Although they possess a relatively high accuracy with big amounts of inserted materials the accuracy is very low at insert volumes below 1µl. The handling of the devices is complicated, too. A direct control of the reaction

parameters is mostly not possible. A detection of the reaction products requires a great effort. These disadvantages lead to a strong increase in costs for new and specific products and methods in the field of chemistry and biochemistry.

The micromachining technology can open new fields in the area of chemistry and biochemistry. Different new modules, like pumps and valves have been developed during the last years [1-4]. These small devices are able in principle to handle very small amounts of fluids with a very high accuracy. Although these components have been characterised very often as individual components some of them were designed and manufactured as integrated systems [5-8]. The aims of these works were focused on the development of total analysis systems in microtechnology, so called μ-TAS. An integration of different components leads to an advantage of these systems in comparison to commonly used devices. In-situ measurements are possible due to the integration of specific sensor systems. Direct interconnections of these groups to electronic building groups are further advantages of these systems.

While the work of different groups world wide was focused on the development of specific sensor interfaces in μ-TAS devices, we tried to pay more attention to the design of micropumps, valves, mixing units and reaction chambers. We were mainly interested in integrated microreactor systems that can be produced by using related process technologies. We have designed and produced modular micropumps, valves, mixing units and reaction chambers to investigate their parameters. A technology for the production of completely integrated microreactor systems was developed.

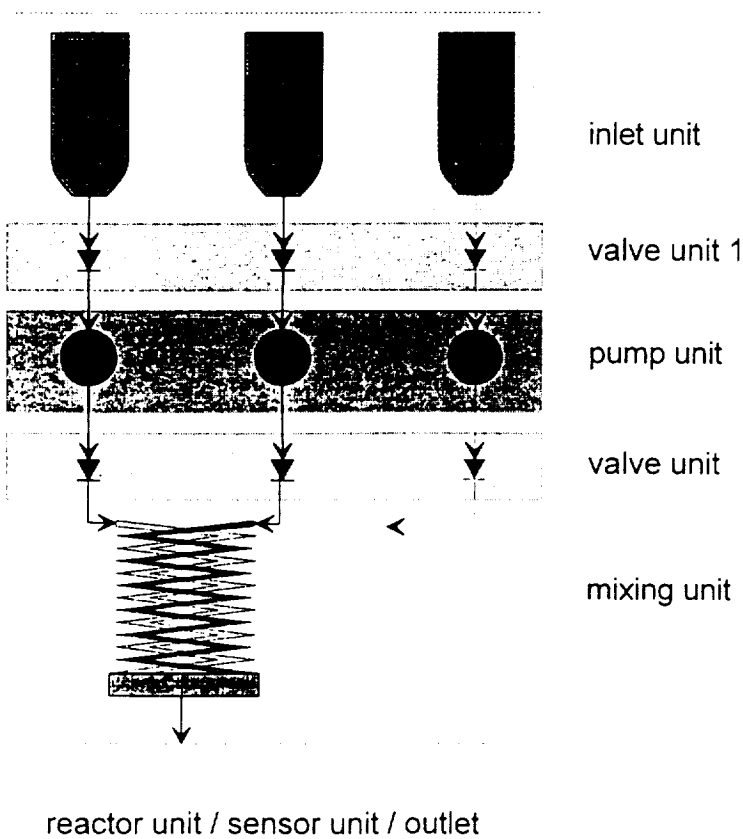


Figure 1: Integrated micrreactor

2. Concept of the microreactor system

The integrated microreactor system was designed to get the possibility of adding more microcomponents. The basic structure of a system consisting of five different groups is shown in figure 1.

The inlet unit is arranged on the top of the system. This unit connects the microsystem with the macroscopic environment. The best way to realise this is a direct connection of the microsystem with media to be observed.

The second group consists of valves. These valves prevent the reflow of the liquid. A reflow could happen when the pumps start their operation.

The pump unit is the next group of the system. The pumps drive the liquid flow from the inlet through the first and the second valve units and the mixing unit as well.

The valves of the fourth unit have to prevent a reflow of liquids into wrong branches of the system.

The fifth group of the system contains the mixing unit. The different liquids should be mixed or, if required, react during their flow through this unit.
The last group contains a reaction chamber and optionally a sensor unit or an outlet of the system. All groups of the system should be integrated. The basic material should be silicon. The number of process steps should be reduced to a minimum.

3. Micropump

3.1 Design of the micropump

The micropumps for application in microreactor systems were designed as piezoelectrically driven membrane pumps. The pump unit as shown in figure 2 consists of two silicon wafers bonded together by the use of a low temperature silicon direct , bonding technology. The wafer on the top is structured applying wet chemical anisotropic etching procedures. The thickness of the membrane is about 50µm. The membrane has a size of 4mm x 4mm. A piezo plate with a size of 3.5mm x 3.5mm x 0.2mm is mounted on top of the membrane. A voltage can be applied onto the top and the bottom of this plate. The silicon base plate covers the inlet and the outlet channels as well as the pressure chamber.

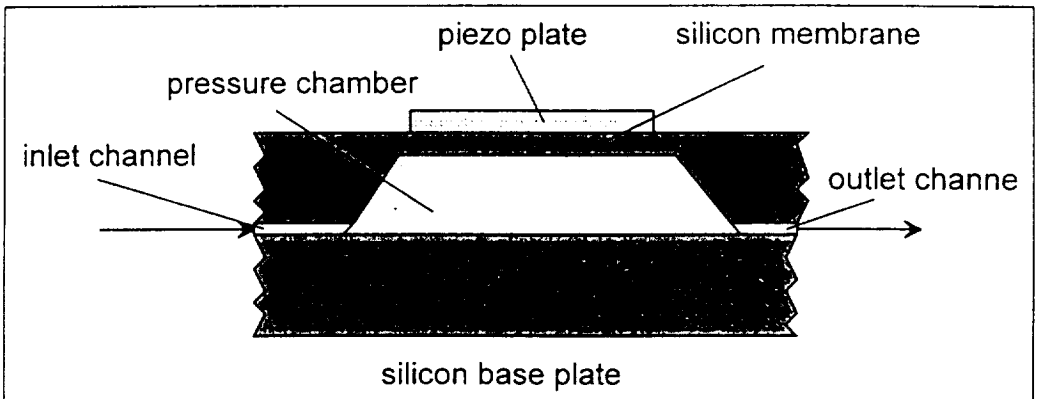


Figure 2: Basic structure of the micropump

3.2 Measurements on the pump

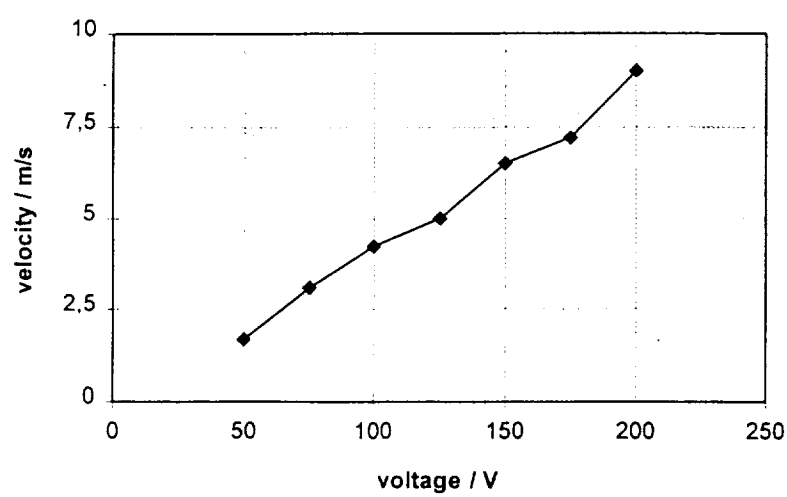


Figure 3: Velocity of the droplets in dependence from the height of voltage

To investigate the performance of the micropump we have prepared several examples according to the above principle. The inlet channel was directly connected with a fluid reservoir. The outlet channel was connected to a gaseous volume. Due to this arrangement it was possible to work without valves in the streaming path. Applying a voltage to the piezo plate one can observe a bending of the bimorph structure that consists of the piezo plate and the silicon membrane. This deformation causes an underpressure inside the

pressure chamber. The liquid tends to stream into the pressure chamber. A turn back of the applied voltage leads to an overpressure in the chamber. The liquid will be pushed out of the chamber. If the outlet is connected directly with a gaseous volume, the streaming resistivity along this path is much lower than that one of the inlet path. By applying cyclic voltage one can observe a cyclic output of the liquid at the outlet. In dependence of the height of the applied voltage we have investigated the size and the velocity of the droplets formed at the outlet. One can see as shown in figure 3that an increase of the applied voltage leads to an increase of the droplet size and an increase of the velocity of the droplets respectively. Besides this we could also found that the volume of the droplets and their velocity depend on the frequency of the applied voltage. The droplet size can also be influenced also by the geometry of the outlet channel.
The best results were achieved in a voltage range between 50V and 200V. It was possible to work in a frequency range up to 6000Hz. The minimum amount that can be pumped was about 60 pl. A maximum flow rate of about 25µl/h is achievable. We have considered a maximum stroke of the bimorph system of 1.5µm. The power consumption of the micropump is lower than 10mW.

4. Microvalve - Design and fabrication

Pump operations in closed liquid systems require the existence of valves. Otherwise the flow and the reflow to the pressure chamber are the same for the inlet and the outlet channels. Microvalves have been well known for several years [3,4]. The integration of such individual devices into closed micro reactors can lead to different problems that are connected with sealing behavior, the dead volumes and the assembling technologies. Therefore we have designed a microvalve that is comparable to the known technologies of the other components and that possesses low dead volumes. The fabrication steps of the valve are shown in figure 4. By means of this fabrication technology it was possible to solve different problems. The valve body is made of a flexible metallisation tong that allows a nearly perfect sealing of the valve seat. Sticking effects of the valve body can be prevented due to the chosen material combinations. Direct bonding technologies can be carried out at low temperatures. .
We have prepared samples of the valves to investigate their behavior under different working conditions. The cross section of the orifice was varied from 300µm x 300µm to 1000µm x 1000µm. The metallisation layer was a sandwich structure that consists of 100nm chromium and 200nm gold.

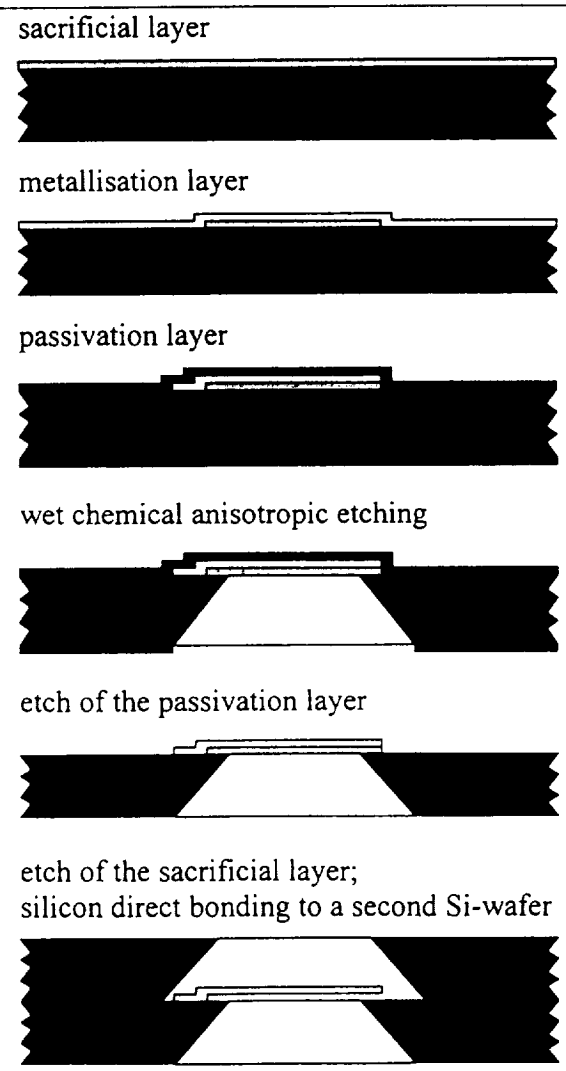


Figure 4: Process steps of the passive microvalve

Into the flow direction all valves were working perfectly up to a pressure of 1 bar. The opposite direction of the flow leads to a streaming breakdown at pressure differences greater than 5000 Pa. This low value indicates a mechanical instability. A better selection of the metals and noticeable thicker layers could lead to improved properties of the valve. An influence of the size of the valve body was not found. Due to the low streaming rate it was impossible to estimate the leakage rate of the valves below this limitation pressure.

5. Mixing unit

5.1 Design principle

The mixing of fluids in small channels is a technical problem. The reason for that is the behavior of the liquids. In very small channels one can observe only a laminar flow of the liquid. This flow is not disturbed by any turbulences. During our experiments we found out that obstacles built in the channel do not lead to any turbulences in the flow. A good mixing of different liquids requires turbulences in the flow. Therefore the answer to the question "How is it possible to achieve a homogeneous mixing of different liquids in microsystems ?" is very important.
The integration of driven obstacles into tubes is known from the macroscopic area. Some rotating devices are known too in the microscopic area [9-11] . They are driven electrostatically and built up out of silicon. Unfortunately these systems were only observed in a gaseous atmosphere. The application with liquids is limited due to the high voltages required. A further disadvantage is the construction

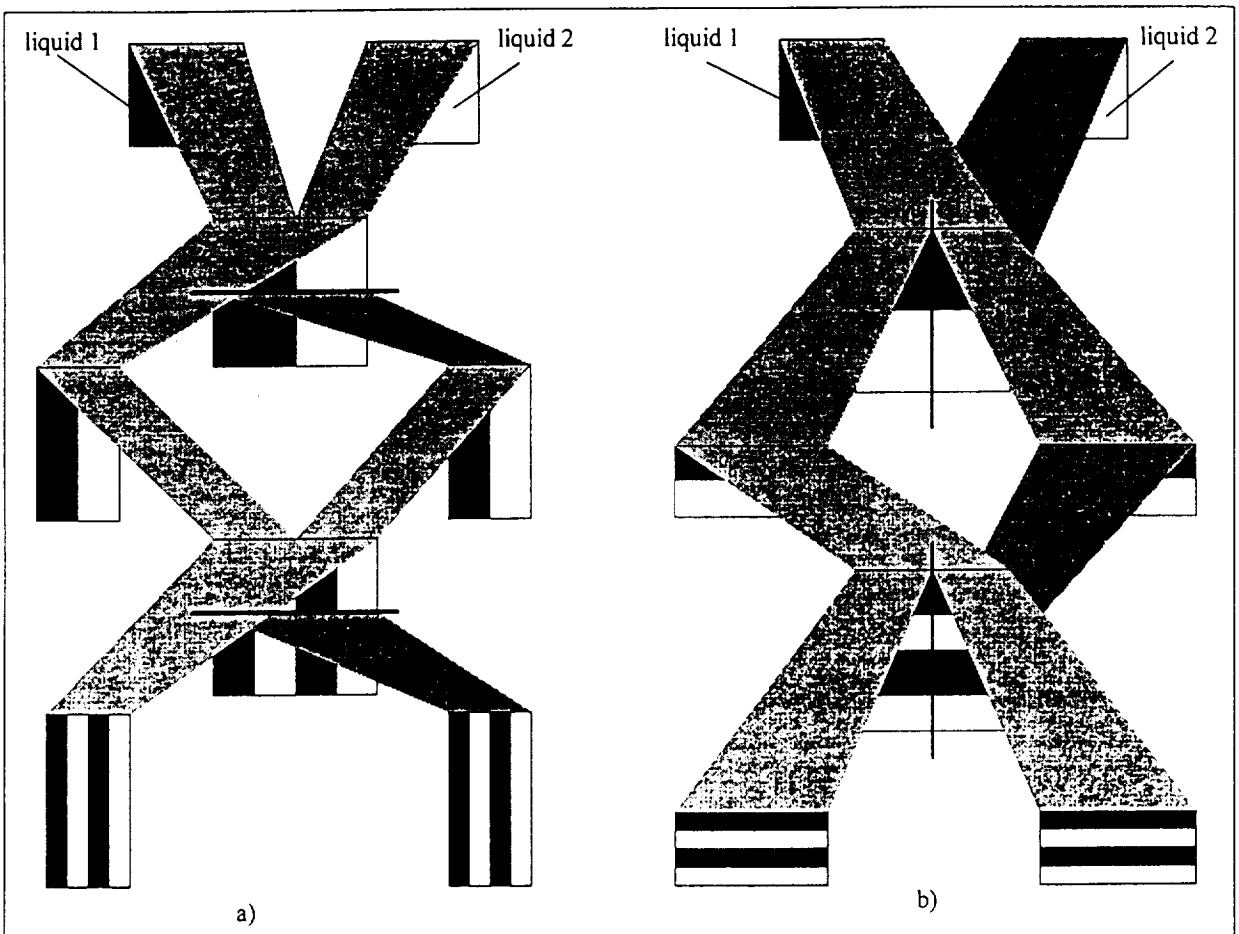


Figure 5: Mixing principles with an enlargement of the active surfacees
a) horicontal adding - vertical dividing - horicontal reuniting of two different liquids
b) vertical adding - horicontal dividing - vertical reuniting of two different liquids

principle of these micromachines. A lot of more process steps would be necessary to build up completely integrated systems using these micromachines.
Static mixing units meet all requirements of the process technology and open the possibility to work with a small amount of basic materials in contrast to moveable obstacles .
The mixing of fluids in small channels can occur by a mechanical exchange of liquid elements due to streaming obstacles and / or by diffusion processes. To get a mixed solution after a short distance it is necessary to enlarge the boundaries between the fluids. A twisting of the fluid layers can not occur due to the laminar flow. An enlargement of the surface boundaries can be achieved in different ways. A simple solution is dividing of each flow into several partial flows. The partial flows of the different solutions will be stacked. Then the partial flows will be reunited. The disadvantage of this method is the size of the system. A lot of space is required due to the division of two or more flows into several partial flows. Therefore we developed a mixing principle that is suited for the use in microsystems and that requires only a small amount of space. Two basic principles of the mixing are shown in figure 5. In the first case (Figure 5 a)) the two liquids will be added together in a horizontal plane. The whole flow will then be divided along a vertical line. As result one gets two partial flows that are located in two different planes. They will be reunited again in a horizontal plane. Then the procedure starts again. In the second case all procedures are turned around 90°. In both cases only two planes are necessary. This mixing procedure allows the mixing of two different fluids. A mixing of more than two liquids is possible by a simple increase in the number of the inlets only.
We calculated the Reynolds-number for the streaming in the channels using water as a liquid. At a cross-section area of $1600\mu m^2$ we found a Reynolds number of about Re=18. To achieve a staple of fluid layers we took into account that the streaming resistivities in different parallel streaming channels should be in the same order. Otherwise one gets a streaming path through the system along the way of the lowest streaming resistivity. In such case no mechanical mixing can occur and the mixing effect is only caused by diffusion processes. We developed a mixing unit that consists of fork-shaped mixing elements. These forks are arranged in two layers. A liquid stream through such a unit will be divided perpendicular to the boundary layer, united in a stapled formation and separated again perpendicular to the boundary layers. Different cross sections of the channels in the fork lead to adapted streaming resistivities. A scheme of the mixing unit and a demonstration of its function is shown in figure 6. We have designed a mask that possesses several mixing units. The mixing units contain

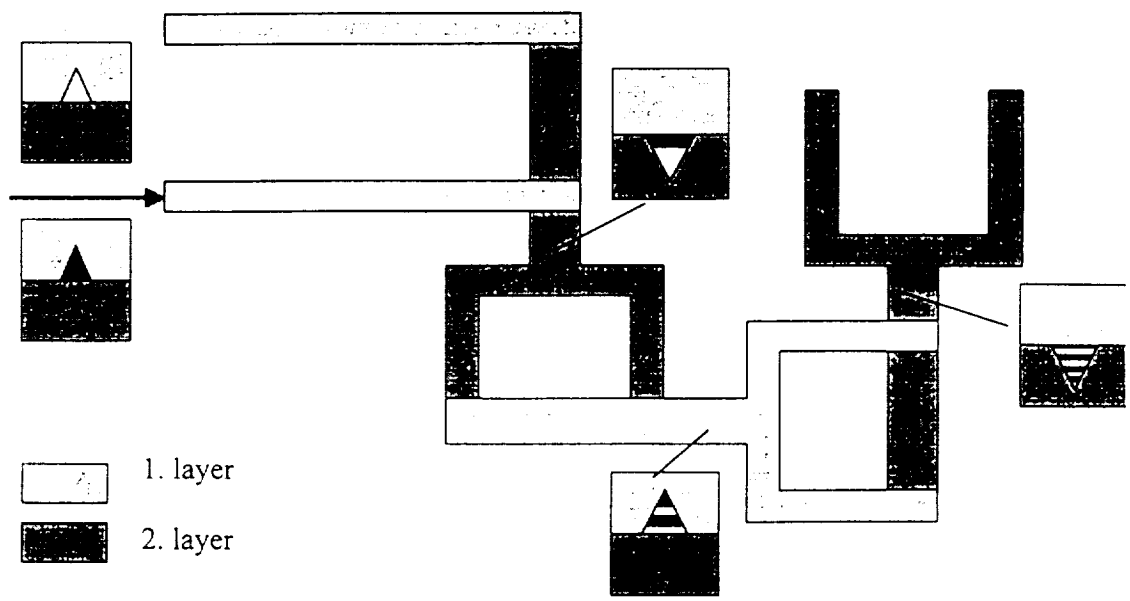


Figure 6: Scheme of the mixing unit with several states of mixing

either 5, 10, 15 or 20 mixing elements. The openings of the fluid channels of the mixing elements had a size of 150μm and 200μm respectively at the surfaces of the substrates.

5.2 Experimental investigations

For the experiments we used 3" - (100) oriented - p-doped silicon wafers. The wafers were cleaned in a standard cleaning procedure. A thermal oxide was grown with a thickness of about 1.2μm. After the lithography the oxide was etched. Then a anisotropic chemical etching was carried out in a 30% KOH solution at 80°C. All structures were etched until the natural etch stop at the (111) planes of the silicon were reached. The oxide was then etched away and the wafers were cleaned again. Then the wafers were pre-treated in an oxygen plasma for 30 sec. Immediately after this process they were rinsed in DI-water, aligned and prebonded. The bonding procedure was carried out at a temperature of 400°C for 4 hours. The wafer was severed into dies. Stainless tubes were slipped into the openings of the channels that were opened by the sawing process. Then the whole system was coated with an epoxy. The steel tubes were connected with transparent tubes made of polymers. To proof the efficiency of the mixers we used a micropump that delivers a cyclic flow rate of the fluids. The average flow rate of about 10 μl/h was constant during the experiments for both liquids. We observed the optical properties of the mixed fluids at the outlet of the mixer. This work was done by means of an optical microscope. We judged the homogeneity of the solution.

5.3 Experimental results and discussion

During our experiments we worked with mixing units that consist of different numbers of mixing elements. Four different types of fluids were used to examine the function of the micromixer.

1. 1 molar chloric acid + methylorange solved in water
2. water + air
3. oil + air
4. water + oil

1. The first liquids are soluble in each other. A mixing of these two fluids can be observed by a change of the colour of the solutions. Methylorange solved in water changes the colour from yellow to pink after a reaction with acids. We found a pink solution at the output of the mixing unit. The change of the colour was independent of the number of the mixing elements. Mixer units with 5 mixing elements show the same behaviour as units with 20 elements. In general the solution was homogeneous mixed. Due to the good solubility of the liquids it was not possible to estimate the real mixing process - diffusion or mechanical mixing.

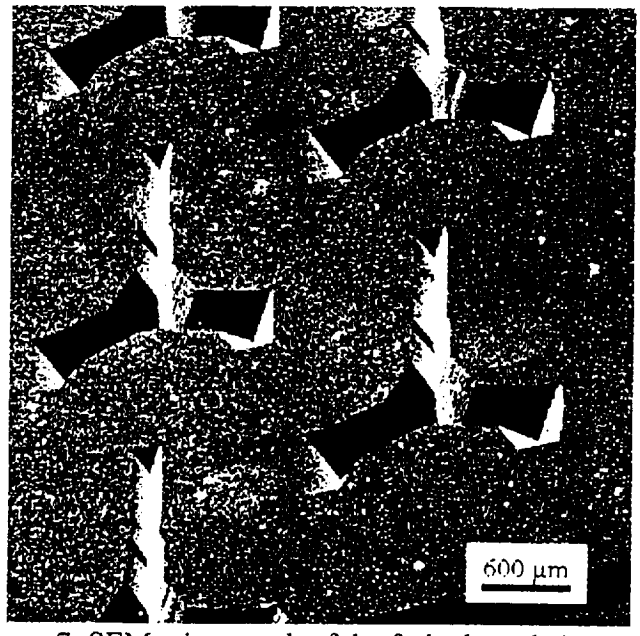


Figure 7: SEM micrograph of the fork-shaped channel structure of some mixing elements on top of one wafer

2. Air has a restricted solubility in water. One can observe bubbles of air in the water if the concentration of the air is too high. Mixing units with 5 elements show bubbles of air and between them columns of water. The volumes of the bubbles and the columns were nearly the same. The dimensions were in a range of 0.2-0.5mm. We observed a growing of an air bubble that was fixed for a short time immediately at the output of the steel tube. This bubble was removed from the outlet after it had reached a determined size. An increase in the number of the mixing elements leads to a change of the behavior. More than 10 mixing elements lead to a development of very small bubbles of air. These bubbles were not fixed at the edge of the tube. So the water became more frothy. The diameters of the bubbles were in a range of less then 0.1mm. The growing of bigger air bubbles was observed minutes after the mixing procedure. This new quality indicates a mechanical mixing of the two fluids. Due to the mixing principle the number of the fluid layers will be doubled and the thicknesses of the layers will be divided into half after each mixing element. Therefore the bubbles get smaller as more mixing elements are flowed through.

3. The mixing of air and oil leads to a similar behavior as described before. The growth of bubbles at the end of the steel tube was restricted due to the high viscosity of oil . The bubbles were removed from the edge of the tube if their diameter was in a range of about 0.1mm. The mixed liquid has changed the colour from yellow (colour of the oil) to a yellow - white. After the mixing unit the liquid was not transparent. A growth of bigger bubbles outside the mixer was observed, too. The oil mixed with the air by five mixing elements was nearly free of air bubbles after about 60 min. We got an oil free of air when the mixing was carried out in a 20 element mixer after 2-3 days. The velocity of the bubble growth was much lower than in water.

4. The mixing of oil and water is very difficult. There is no solubility of the fluids in each other. In a mixing unit with 5 mixing elements we found a growth of small water droplets at the edge of the steel tube. These droplets were surrounded by oil. The droplets were removed from the tube when they had reached a diameter of about 0.5mm. So we could observe an alternating raw of columns of oil and columns of water in the transparent tube. The liquids were separated very quickly inside a bottle made of glass. Fifteen or more mixing elements lead to a decrease of the diameter of the water droplets. These droplets are spherically shaped and surrounded by the oil. The solution is not transparent and seems to be in an emulgated state. A demixing of the two liquids was observed about 3-4 hours after the mixing procedure. This behavior indicates clearly a mechanical mixing of the two components.

6. Reactor chamber

The mixing of reagents does not lead in all cases to a chemical reaction. A supply of energy is very often required to initiate chemical processes. This can be done by light, electricity, radiation or temperature. We chose an energy supply in form of temperature. An integration of a resistor network is simple and can also be done by the use of microtechnologies. Another advantage is the relatively low power consumption. To initiate a reaction of the mixed fluids a reactor chamber was designed. This chamber is directly connected with the output of the mixing unit. To prevent a chemical reaction or a catalytic influence of the resistor materials with the reagents the resistors were arranged outside the reactor chamber. For the heating elements we used platinum as material. The structures were prepared applying a sputtering technology for deposition and a lift-off-process to create the elements. Because of the very high thermal conductivity of silicon it is possible to increase the temperature inside the reactor chamber applying a current to the heating elements.

We carried out some experiments and found some disadvantages of this solution. The high temperature gradient caused by a heat up of the resistors leads to a spreading of temperature over the whole wafer. An increase in the temperature is observable not only inside the reactor chamber but also in the mixing unit, the micropumps and the microvalves as well. It was not possible to achieve the

preestimated temperature inside the reactor chamber due to the temperature losses in the whole system. Further investigations are necessary to reduce the temperature losses and the heating up of the whole system.

7. Design and process steps of the microreactor

The microreactor contains several individual components as shown in figure 1. All these components except the liquid reservoirs should be built up in an integrated form. The basic material for all components is silicon. All process steps should be compatible to commonly known microtechnologies. On the basis of this concept we designed and prepared an integrated microreactor module that meets all requirements in liquid handling and process technologies, respectively.
The microreactor consists of three structured wafers that are bonded together. The first wafer contains square deepenings. One kind of deepenings solves as coverage for the heating elements arranged on top of the second wafer. The other deepenings are required to get a membrane structure. Piezoceramic plates are mounted on top of the membranes. These plates act together with the silicon membrane below it as a bimorph structure. Applying a voltage to the piezoelectric element the whole bimorph structure will be bended. Due to this deflection a pressure will be built up in the chamber below the membrane. Caused by the pressure difference a directed flow of the liquids inside the chamber will be initiated. The directed streaming can be achieved by means of two passive valves. The second wafer contains the two valve units, a part of the mixing unit, the process chamber and the heating elements as well. The valves act as passive valves. They have thin metallic membranes as valve bodies. The flow of the liquid in one direction is possible due to holes etched through the wafer. The direction of the streaming is forced by the inverted arrangement of the holes and the valve bodies, respectively, in each streaming path. A part of the mixing unit is formed as an arrangement of "fork-shaped" v-grooves on the bottom of the wafer. A reactor chamber is arranged at the output of the mixing element. This chamber contains v-shaped grooves arranged in parallel. Deeply etched v-grooves act as partial shape for stainless steel tubes of the inlet and the outlet connections in this wafer. The third wafer contains an arrangement of v-grooves. These grooves act as channels for the streaming path, as second part of the mixing unit (fork-shaped) and as shape to take in the outside connections. The advantage of this design is the low consumption of basic materials, short ways of the flow inside the system and therefore very small dead volumes and only a few required process steps.

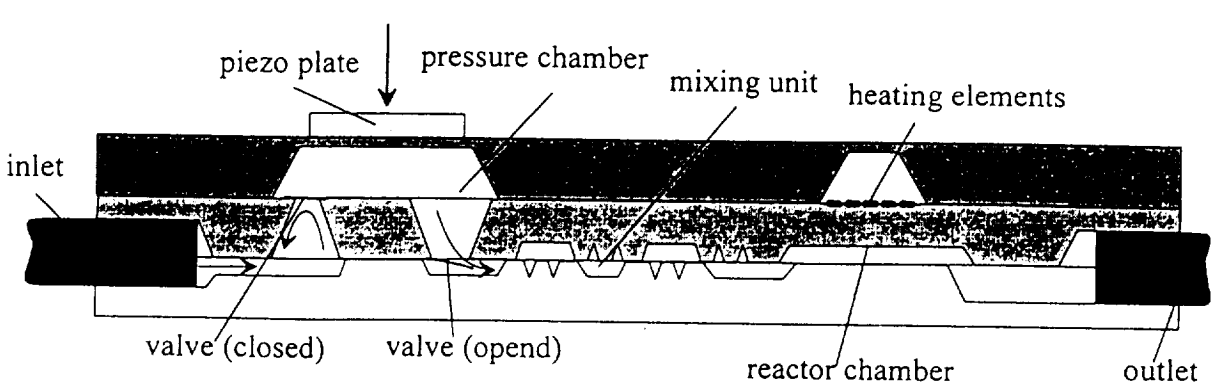


Figure 8: Cross-section of the microreactor

The microreactor can be realised by the use of silicon micromachining technologies. All three dimensional structures in the silicon can be etched anisotropically. The depth of the structures is defined by the natural etch stop at {111} planes of the silicon. Due to the etch stop it was possible to etch through the whole wafers in some required regions. The valve bodies of the passive valves and the

heating elements were produced by a metallisation process before the etching procedure. The wafers can be bonded together after structuring using the silicon direct bonding technology. This process step was carried out after a short treatment of the wafers in a pure oxygen plasma for about 30s. The bonding temperature was 450°C. The advantage of this low temperature is the possibility of an integration of microelectronic circuits. The piezoceramic plates were mounted after the bonding process with a glue on the top. The tubes of the inlet and the outlet were slipped into the deeply etched channels and fixed by the use of an elastic glue. All open surfaces of the system are made of silicon. For special chemical reagents it is possible to cover the surfaces with inert materials, like SiO_2 or Si_3N_4.
Having applied a voltage to the piezo plates a flow rate up to 20µl/h was observed.

8. Conclusion

An integrated microreactor was developed that consists of several individual components. Each individual component was investigated separately. The pumps can force a flow rate up to 25µl/h applying a voltage of 200V. Passive valves offer resistance in closed state up to 5000 Pa. The behavior of the mixing unit was proved by the mixing of different fluids. The possibility of the mixing of liquids with liquids as well as the mixing of gases with liquids was shown. A homogeneous mixture can be achieved for liquids that are soluble in each other after 5 mixing elements. Fluids that are not soluble in each other will be emulgated after more then 10 mixing elements. The mixer possesses a very high efficiency in the mechanical mixing of two fluids. The process chamber was arranged with heating elements on their outside. Due to the temperature losses in the whole system it was not possible to achieve the preestimated temperature inside the chamber. Furthermore investigations are required to find out a better energy supply inside this component. The integrated microreactor was designed and built up using commonly known process steps of the microtechnology. A total flow rate at the output of 20µl/h was observed. The microreactor is suited in principle for applications in the chemical and the biochemical area. The advantages of the system are low reagent consumption, very high accuracy, low power consumption, small and light and the possibility to integrate sensors and microelectronic circuits.

9. Acknowledgement

The authors would like to acknowledge the guidance and help of the researchers in the Laboratory of Microstructuring Technologies, particularly J. Burghold for discussions about the streaming behavior of liquids in capillaries; M. Fischer for discussions about the design of micropumps;
O. Marufke for the direct bonding procedures; B. Hartmann and G. Harnisch for preparing various types of micropumps, microvalves, mixing units and integrated microreactors.

10. References

[1] Zengerle, R.; Kluge, S.; Richter, M.; Richter, A.
A bidirectional silicon micropump
Procceding MEMS'95, Amsterdam, p. 19-24

[2] v.d. Pol, F.C.M.; van Lintel, H.T.G.; Elwenspoek, M.; Fluitman, J.H.J.
A thermo-pneumatic micropump based on micro-engineering techniques
Sensors & Actuators, A21-23, 1990, p. 198-202

[3] Esashi, M.; Shoji, S.; Nakano, A.
Normally closed microvalve and micropump fabricated on a silicon wafer
Sensors & Actuators, 20 (1989), p. 163-169

[4] Miles, R.; Weber, W.
Thermopneumatically actuated microvalves and integrated electro-fluidic circuits
4. Int. Conf. on new Actuators "Actuator 94" Bremen 1994, p. 56-60

[5] Branebjerg, J.; Fabius, B.; Gravesen, P
Application of miniature analyzers: from microfluidic components to μTAS
Micro Total Analysis Systems, Enschede 1994, Proceedings, p. 141-151

[6] Elwenspoek, M.; Lammerink, T.S.J.; Miyake, R.; Fluitman, J.H.J.
Towards integrated microliquid handling systems
Journ. of Micromech. and Microengineering vol. 4(1994), no. 4, p. 227-245

[7] Mensinger, H.; Richter, T.; Hessel, V.; Döpper, J.; Ehrfeld, W.
Microreactor with integrated static mixer and analysis system
Micro Total Analysis Systems, Enschede 1994, Proceedings, p. 237-243

[8] van der Schoot, B.; Verpoorte, E.; Jeanneret, S.; Manz, A.; de Rooij, N.
Microsystems for analysis in flowing solutions
Micro Total Analysis Systems, Enschede 1994, Proceedings, p. 181-190

[9] Fan, S.-L.; Tai, Y.-C.; Muller, R.S.
Integrated Moveable micromechanical Structures for Sensors and Actuators
IEEE Tr. on ED, vol 35(1988), no. 6 p.724-730

[10] Tai, Y.-C.; Muller, R.S.
IC-processed Electrostatic Synchronous Micromotors
Sensors & Actuators 20(1989), p. 49-55

[11] Fan, L.-S.; Tai, Y.-C.; Muller, R.S.
IC-processed Electrostatic Micromotors
Sensors & Actuators 20(1989), p. 41-47

NASA

Fluid flow volume measurements using a capacitance/conductance probe system

T. X. Nguyen

NASA Johnson Space Center
2101 NASA Road 1, MS EV3
Houston, TX 77058
Phone: (713) 483-1464 Fax: (713) 483-5830

G. D. Arndt

Electromagnetic Systems and Navigation Sensors Branch
NASA Johnson Space Center

J. R. Carl

Lockheed Engineering and Sciences Company

71173
030734
1P

A probe system has been developed to measure the flow volume of a single fluid passing through an orifice or flow line. The system employs both capacitance and a conductance probe at the orifice, together with phase detection and data acquisition circuitry, to measure flow volume and salinity under low or zero gravity conditions.

A wide variety of frequencies can be used for the RF signal source and is chosen primarily by the capacitance of the orifice probe and the fluid passing through the orifice. Rapid measurements are made using the reflected signal from the orifice probe to determine the "instantaneous" permittivity of the fluid/gas mixture passing through. The "instantaneous" measurements are integrated over time to determine flow volume.

Analysis reveals that a narrow orifice helps to reduce non-linearities caused by differing flow rates. The geometry of "deflectors" and "directors" for the flowing fluid are important in obtaining linearity. Measured data shows that a volume measurement accuracy of approximately four percent can be constantly achieved. The prototype hardware system and associated software have been optimized and are available for further applications. The system has immediate applications in low or zero gravity environments for measurements of urine or other liquid volumes.

Text of full paper not available at time of publication.

Biomorphic architectures for autonomous "Nanosat" designs

Brosl Hasslacher

Theory Division, LANL

and

Mark W. Tilden

Biophysics Division, LANL

71175

Modern space tool design is the science of making a machine both massively complex while at the same time extremely robust and dependable. We propose a novel nonlinear control technique that produces capable, self-organizing, micron-scale space machines at low cost and in large numbers by parallel silicon assembly.

Experiments using biomorphic architectures (with ideal space attributes) have produced a wide spectrum of survival-oriented machines that are reliably domesticated for wok applications in specific environments. In particular, several one-chip satellite prototypes show interesting control properties that can be turned into numerous application-specific machines for autonomous, disposable space tasks.

We believe that the real power of these architectures lies in their potential to self-assemble into larger, robust, loosely coupled structures. Assembly takes place at hierarchical space scales, with different attendant properties, allowing for inexpensive solutions to many daunting work tasks.

The nature of biomorphic control, design, engineering options, and applications will be discussed.

Text of full paper not available at time of publication.

Session 7: Space Applications II (Wednesday, AM/PM)
Session Chair: Ernest Y. Robinson (Aerospace)

MEMS Technology for Space Applications
by A. van den Berg, V. L. Spiering, T. S. J. Lammerink, M. Elwenspoek, and P. Bergveld, MESA Research Institute, Univ. of Twente, The Netherlands

Thermal Switch for Satellite Temperature Control
by H. Ziad, P. Van Gerwen, and K. Baert, IMEC, Belgium,
T. Slater, Wildcat Micromachining,
E. Masure and F. Preud'homme, Verhaert Design & Development, Belgium

Optically-Switched Resonant Tunneling Diodes for Space-Based Optical Communication Applications
by T. S. Moise, Y.-C. Kao, D. Jobanovic, and P. Sotirelis, Texas Instruments Inc.

A Micromechanical INS/GPS System for Small Satellites
by N. Barbour, T. Brand, R. Haley, M. Socha, J. Stoll, P. Ward, and M. Weinberg, Charles Stark Draper Laboratory

Micro Scanning Laser Range Sensor for Planetary Exploration
by Ichiro Nakatani, Hirobumi Saito, Takashi Kubota, Takahide Mizuno, Hiroshi Katoh, Satoru Nakamura, Kenji Kasamura, Institute of Space and Astronautical Science (ISAS), and Hiroshi Goto, OMRON Corp, Japan

NanoSat Electric Power Systems
by Michael Robyn and Larry Thaller, The Aerospace Corporation

Low-Mass Inflation Systems for Inflatable Structures
by Daniel Thunnissen, Mark S. Webster, and Carl S. Engelbrecht, Jet Propulsion Laboratory (NASA)

Miniature Telerobots in Space Applications
by S. C. Venema and B Hannaford, University of Washington

/
71176
230739
10P.

MEMS TECHNOLOGY FOR SPACE APPLICATIONS

A. van den Berg, V.L. Spiering, T.S.J. Lammerink, M. Elwenspoek and P. Bergveld

MESA Research Institute, University of Twente

P.O. Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT

Microtechnology enables the manufacturing of all kind of components for miniature systems or microsystems, such as sensors, pumps, valves, and channels. The integration of these components into a complete Micro Electro Mechanical System (MEMS) drastically decreases the total system volume and mass. These properties, combined with the increasing need for monitoring and control of small flows in (bio)chemical experiments, make MEMS attractive for space applications.

The level of integration and applied technology depends on the product demands and the market. The ultimate integration is process integration, which results to a one-chip system. An example of process integration is a dosing system of pump, flow sensor, micromixer, and hybrid feed back electronics to regulate the flow [1, 2]. However, for many applications a hybrid integration of components is sufficient and offers the advantages of design flexibility and even exchange of components in case of a modular set-up.

Nowadays we are working on hybrid integration of all kind of sensors (physical and chemical) and flow system modules towards a modular system: the micro Total Analysis System (μ TAS) [3, 4]. The substrate contains electrical connections as in a Printed Circuit Board (PCB) as well as fluid channels for a Circuit Channel Board (CCB) and integrated they form a Mixed Circuit Board (MCB).

1. INTRODUCTION

During the past few decades a large variety of chemical microsensors has been developed. A large amount of sensor principles such as optical, electrochemical, mass-sensitive and calorimetric have been developed [5]. However, few of these sensors have made the way to the market. One of the main reasons for this was the inherent limitation of chemical sensor performance caused by sensor drift and loss of selectivity and sensitivity. For this reason, actuators were added which enabled calibration of the sensor. Thanks to revolutionary developments in silicon microtechnology, in the last decade many so-called fluid-handling elements have been developed. Using these elements, complete so-called Micro Total Analysis Systems (μTAS) could be built. One of the problems encountered with such systems, however, is that due to their high complexity it is virtually impossible to develop and fabricate all necessary components alone. For this reason we have developed a generic hybrid concept, a "Micro Fluidic System" (MFS) enabling the composition of a complex analysis system by integrating different components on one "motherboard" [6].

In this paper several components for such Micro Total Analysis Systems (μTAS) will be described, such as an electrochemical micotitrator and components for fluid handling. Furthermore, the MFS concept for integration of these components into a system will be described. For an electrochemical sensor-actuator system (microtitrator) it will be shown how incorporation of this in a Micro Total Analysis System (μTAS) improves its performance.

2. SENSOR-ACTUATOR SYSTEM

The coulometric acid/base titrator consists of an electrochemical actuator combined with an ISFET pH-sensor (see fig. 1). With this sensor-actuator device fast titrations can be carried out, and the titration time t_{end} is directly related to the acid/base concentration to be determined via the formula [7,8]:

$$\frac{\partial\sqrt{t_{end}}}{\partial C_{acid}} = \frac{F\sqrt{\pi D_{acid}}}{2j_c}, \qquad (1)$$

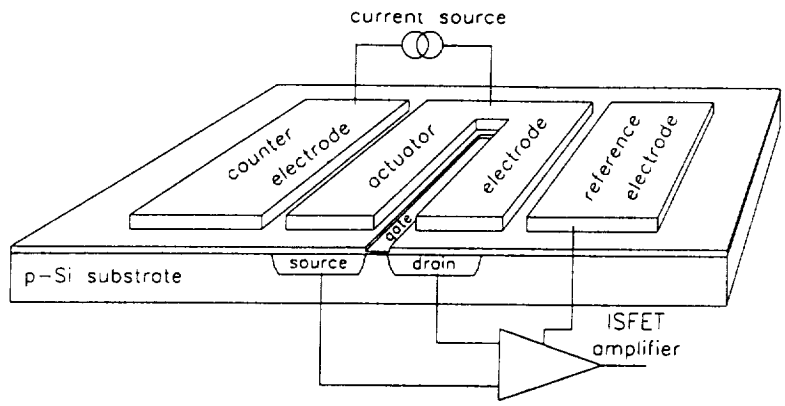


Fig. 1. Basic elements of the coulometric sensor-actuator device.

with j_c is the cathodic current density through the actuator, C_{acid} and D_{acid} respectively the concentration and diffusion coefficient of the acid and F is Faraday constant. A typical plot of the square root of t_{end} *vs.* the concentration of lactic acid is shown in figure 2.

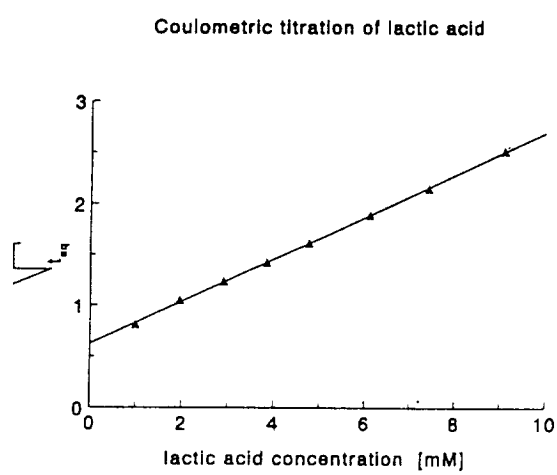


Fig. 2 Square root of t_{end} as a function of lactic acid concentration.

In practice, it appears that the coulometric actuator device, if operated as such, shows some disadvantages. First of all, there is no linear relation between titration end point and the required concentration. Secondly, the proper functioning of the device relies on the presence of only one type of mass transport: diffusion. Effects of migration may strongly influence the measurement, and to

overcome this an excess concentration of supporting electrolyte is needed. Finally the operational range is limited to a maximum of approximately 10 mM, caused by limitations in the current density. This upper limit turns out to be problematic with respect to the much higher total acid concentrations in samples such as fruit juice or wine (50 to 150 mM). The problems mentioned above can be overcome by incorporation of the microtitrator in a Micro Total Analysis System (μTAS), where the sample solution can be diluted if necessary (see figure 3).

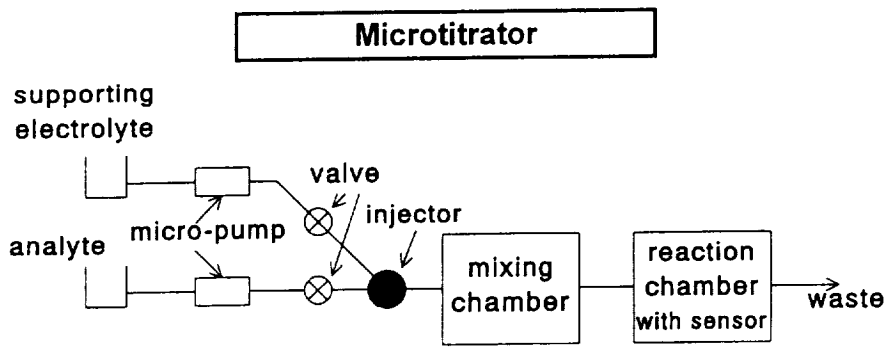
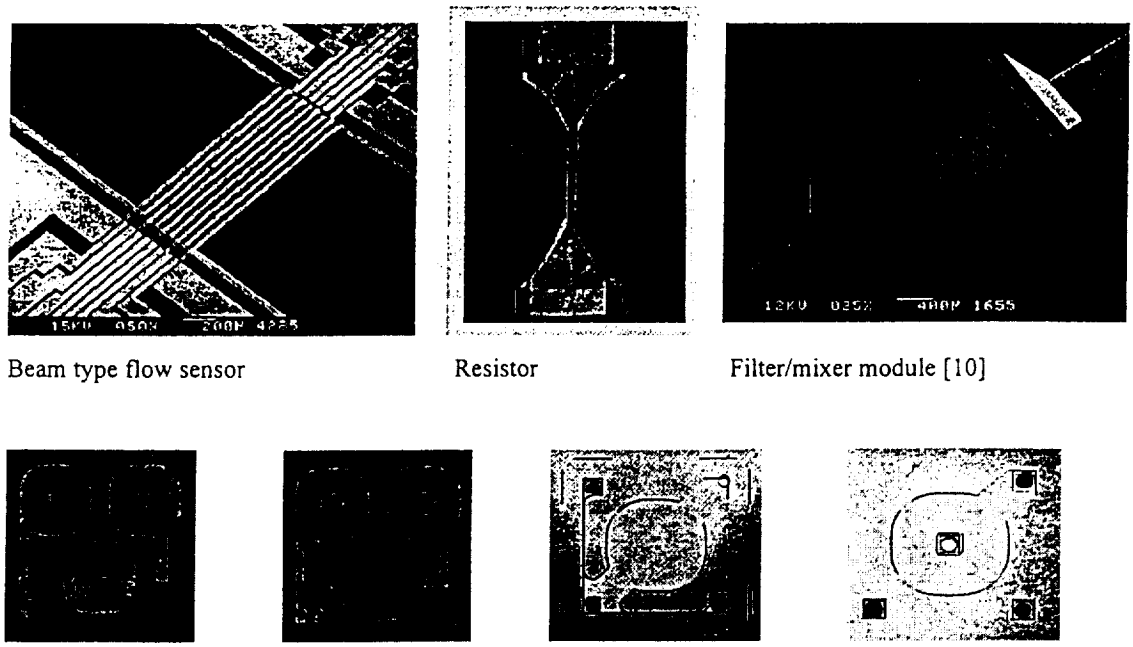


Fig. 3. The proposed μTAS for improved coulometric sensor-actuator performance.

The μTAS thus proposed, including subsystems, is schematically shown in figure 3. In reality the mixing and reaction chamber might be one chamber only. Also, the micro-pump itself might function as a valve.

3. COMPONENTS FOR FLUID HANDLING

For the realisation of Micro Fluidic Systems (MFS) a wide variety of components is needed. Many of them, like pumps, flow-sensors and filters, were already developed at MESA [9]. For integration in the MFS new designs have been made of a number of components that made them compatible with mentioned concept. In Fig. 4 examples are shown of a flow sensor, a resistor, a filter/mixer module, a micropump, and a microvalve specially designed for integration in MFS.



Beam type flow sensor | Resistor | Filter/mixer module [10]

Pump (top) | Pump (bottom) | Pressure Controlled Valve (top-bottom)[11]

Fig. 4. Photographs of a number of different components for the Micro Fluidic System (MFS).

The integration of the abovementioned components into microsystems can be realised in two ways: hybrid and monolithic. The hybrid form will be discussed in section 4. An example of a monolithically integrated dosing system, comprising a micropump and a flow sensor, is shown in figure 5. The advantage of this type of integration is the possibility to realise very small dead volumes, which is of interest for chromatography applications.

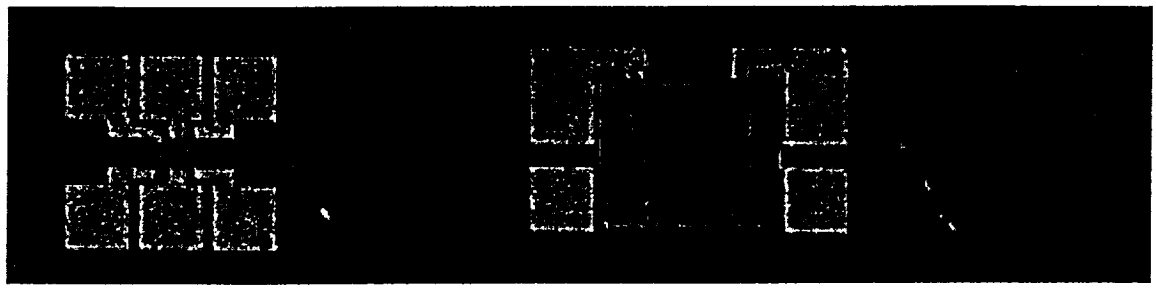


Fig. 5. Photograph of a monolithic integrated dosing system containing a pump and flow sensor.

4. MICRO FLUIDIC SYSTEM (MFS) CONCEPT

For the integration of components for fluid-handling into a system, several approaches have been proposed. Van der Schoot *et al.* [12] proposed a vertical, stackwise arrangement of components which has the advantage of efficient use of surface area but the disadvantage of being rather inflexible. An alternative was presented by Fiehn et al. [13], who presented a Fluidic ISFET-Microsystem (FIM) based on a planar integrated system. The main disadvantage of this system is that a whole processed glass-bonded silicon wafer is used as system substrate. Recently, an alternative was presented in which the components are at least partly, reversibly mounted in a way perpendicular to the substrate [14]. In the latter system, which uses a silicon sealing for hermeticity, the advantage is that each component can be replaced. The disadvantage is that the fabrication of the system is not easily automated.

The Micro Fluidic System (MFS) we propose is composed of a so-called Mixed Circuit Board (MCB) containing the fluid channels as well as the electronic circuitry in combination with the silicon-based fluidic components (modules). These modules have a standardised connection to the planar MCB both for fluids and electrical signals. The MCB consists of a glass-bonded silicon backplate in a first stage, whereas in the second stage (laminated) plastics are used (see fig.6). In fig. 7 an example is given of how the electrical and mechanical layout of such a microanalysis system looks like.

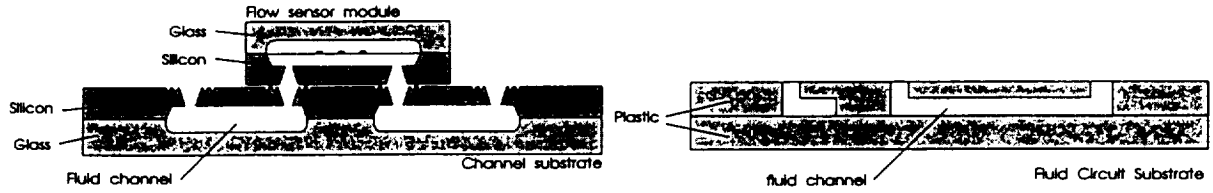


Fig. 6. Flow sensor module on Si-glass bonded substrate (left) and plastic Mixed Circuit Board (right).

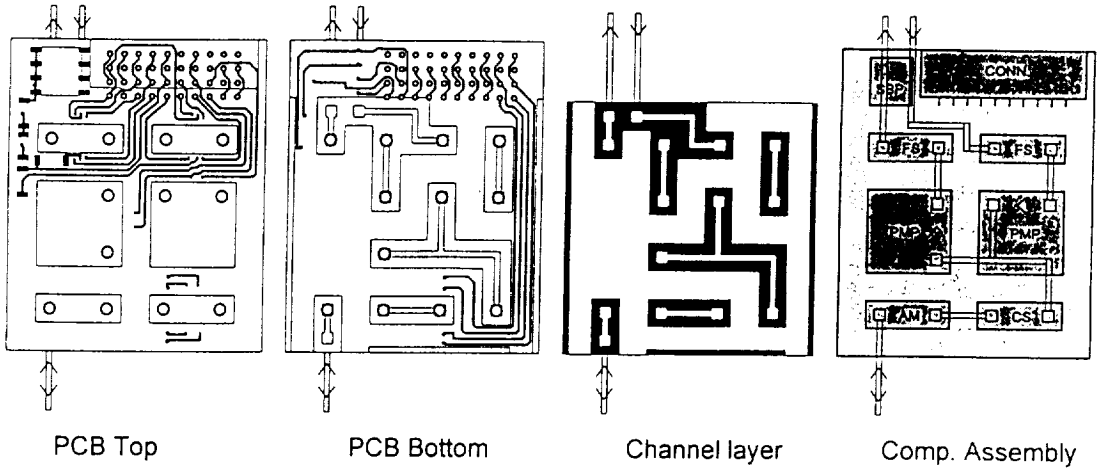


Fig. 7. Schematics of floorplan with mixed electrical and fluid connections

An example of a microanalysis system based on the MFS concept, containing two micropumps , two flow sensors and an optical detection cell is shown in figure 8. With this system, color changes of Congo red (a color indicator) upon pH changes between pH 3 and pH 9 is readily detected.

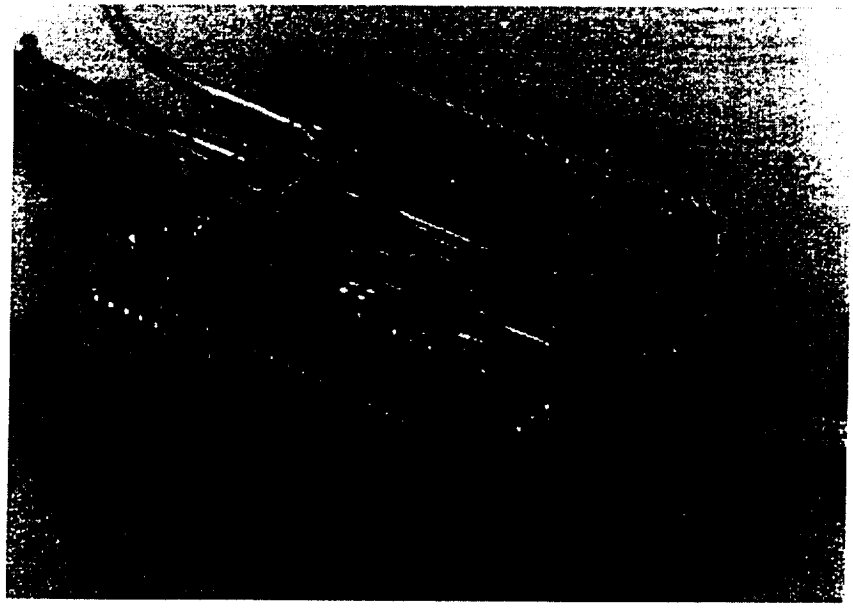


Fig.8. Micro Analysis System demonstrator comprising two pumps, flow sensors and optical detector cell.

5. CONCLUSIONS

A modular system concept for fluid handling (Micro Fluidic System, MFS) has been proposed for the realization of Micro Total Analysis Systems and micro chemical systems. MFS enables the use of standard components or modules to be integrated on a planar base plate that contains fluid channels as well as electronic circuitry. Through the standardization the exchange of different components from different suppliers is stimulated. An example of a microreactor in the form of a microtitrator is presented and it is illustrated how incorporation of this microreactor in a micro total analysis system improves it performance and avoids some of it disadvantages.

ACKNOWLEDGEMENTS

The authors want to thank Aquamarijn BV for supplying photographs of the microfilter/mixer module. The assistance of LC-Packings in the characterisation of the pressure resistance is gratefully acknowledged.

REFERENCES

1 H.T.G. van Lintel et al., A piezoelectric micropump based on micro-machining of silicon, *Sensors and Actuators 15*, (1988), 153-167.

2 T.S.J. Lammerink et al., Integrated micro-liquid dosing system, *Proc. MEMS '93, Fort Lauderdale, Florida, U.S.A., Feb. 7-10 1993*, pp. 254-259.

3 T.S.J. Lammerink et al., Modular concept of miniature chemical systems, *Symp. Microsystem techn. for chem. and biol. microreactors, Mainz, Germany, Feb. 20-21 1995.*

4 A. van den Berg, P. Bergveld, D.N. Reinhoudt, M.Elwenspoek and J.H.J Fluitman, "Miniaturized Chemical Analysis Systems", *Proc.5th Int. Symp. on Micro Machine and Human Science"*, Nagoya, Japan, (1994), 181-185.

5 W. Göpel, J. Hesse and J.N. Zemel (ed.): *"Sensors. A comprehensive review."*, **Vol. 2**: Chemical and Biochemical Sensors., VCH, Meinheim, Germany, (1991).

6 T.S.J. Lammerink, "Micro Fluid System Concept", internal MESA report, (1994).

7 W. Olthuis, B.H. van der Schoot, F. Chavez and P. Bergveld, *Sensors and Actuators*, **17**, (1989), 279-283.

8 W. Olthuis, J. Luo, B.H. van der Schoot, P. Bergveld, M. Bos and W.E. van der Linden, *Anal. Chim. Acta*, **237**, (1990), 71-81.

9 J.H.J. Fluitman, A. van den Berg and T.S.J. Lammerink, in: A. van den Berg and P. Bergveld (eds.), "Micro Total Analysis Systems", Kluwer Academic Press, The Netherlands, (1994), 289-293.

10 C.J.M. van Rijn and M.C. Elwenspoek, Proc. MEMS '95, Amsterdam, The Netherlands, (1995), 83-87.

11 T.S.J. Lammerink, N.R. Tas, J.W. Berenschot, M.C. Elwenspoek and J.H.J. Fluitman, Proc. MEMS '95, Amsterdam, The Netherlands, (1995), 13-18.

12 B.H. van der Schoot, S. Jeanneret, A. van den Berg and N.F. de Rooij, "A Silicon Integrated Miniature Chemical Analysis System", *Sensors and Actuators*, **B6** (1992), 57-60.

13 H. Fiehn, S. Howitz, M.T. Pham, T. Vopel, M. Bürger and T. Wegener, in: A. van den Berg and P. Bergveld (eds.), "Micro Total Analysis Systems", Kluwer Academic Press, The Netherlands, (1994), 289-293.

14 R. Hintsche, M. Paeschke, A. Uhlig, C. Kruse, F. Dittrich, M.T. Pham and S. Howitz, Proc. Transducers '95, Stockholm, Sweden, (1995), 772-775.

THERMAL SWITCH FOR SATELLITE TEMPERATURE CONTROL

H. Ziad*, T. Slater**, P. Van Gerwen*, E. Masure***, F. Preud'homme***, K. Baert*

*IMEC, MAP MS-division, Kapeldreef 75, 3001 Leuven, Belgium.
**Wildcat Micromachining, 1032 Irving Box 227, San Francisco CA 94122 USA.
***Verhaert Design & Development, Hogenakkerhoekstraat 9, 9150 Kruibeke, Belgium.

ABSTRACT

An Active Radiator Tile (ART) thermal valve has been fabricated using silicon micromachining. Intended for orbital satellite heat control applications, the operational principle of the ART is to control heat flow between two thermally isolated surfaces by bringing the surfaces into intimate mechanical contact using electrostatic actuation. Prototype devices have been tested in vacuum and demonstrate thermal actuation voltages as low as 40 volts, very good thermal insulation in the OFF state, and a large increase in radiative heat flow in the ON state. Thin anodised aluminum was developed as a coating for high infrared emissivity and high solar reflectance.

INTRODUCTION

There are applications for devices which can control the radiative transfer of heat into or out of an object; the application of interest for this study is orbital satellites. A novel approach for controlling the radiative heat flow between the satellite and the environment has been recently proposed [1]. Active Radiator Tiles (ARTs) rely on a variable mechanical contact to control the heat flow between two surfaces, creating a thermal switch or valve. Desired properties of the thermal switch device are a high thermal resistance at rest, a high thermal conductance when activated, and low power operation. In this paper we report the construction of ART prototypes and the experimental verification of the principle of operation. A generalised side view of the proposed device is shown in Figure 1. The operating principle is to control the flow of heat between a hot side (Base) and a cold side (Radiator) of the device by direct mechanical contact between the two sides, with electrostatics supplying the driving force.

In the OFF state (no actuation), the two plates are separated by thermal insulators (polyimide) from 10 to 20 μm tall, which minimises the conductive heat transfer Q_{ins}. The interior cavity surfaces of the complete ART device (upper surface of the Base plate and lower surface of the Radiator plate)

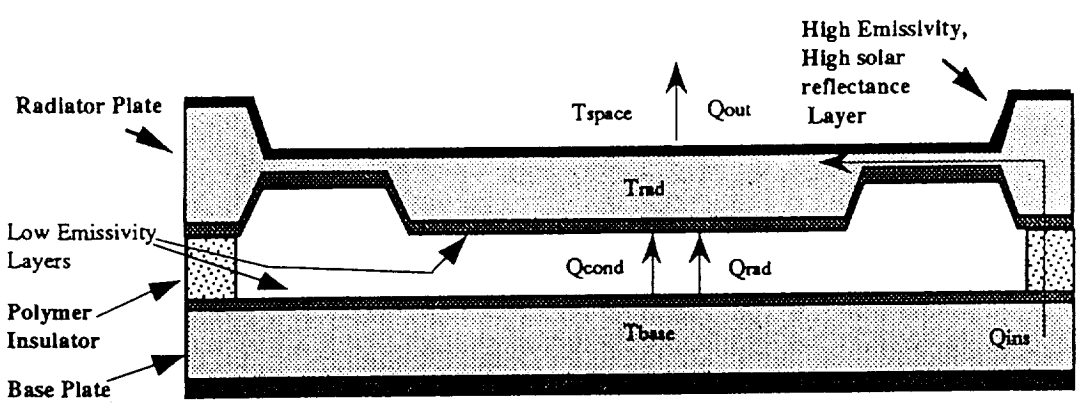


Figure 1. Schematic side view of Active Radiator Tile.

are coated with low-emissivity layers (aluminum and PECVD nitride) to minimise radiative heat transfer Q_{rad} between the plates. In this state the heat flow from the Base to the Radiator is very low and the device presents a large thermal resistance from the Base to the environment.

The silicon Radiator plate is etched anisotropically to form a large membrane with a large central contacting boss. Application of a suitable voltage across the two plates will cause a deflection of the Radiator plate into intimate contact with the Base plate, allowing heat transfer by conduction Q_{cond} from the Base to the Radiator, where the heat can then be radiated to the environment as Q_{out} (the ON state). The exterior surface of the Radiator plate, which will directly face the operating environment, is coated with a high-emissivity layer of paint for maximum radiative heat flow to the environment in the ON state. The combination of dual facing low-emissivity layers and a high-emissivity exterior surface allows for high thermal resistance when the device is at rest, and a high thermal conductance to the environment when the device is actuated. Thus the ART device acts as a thermal valve or switch, actively controlling the flow of heat to the environment in response to an input voltage.

THERMAL/OPTICAL CONSIDERATIONS

The basic desired properties of the ART device are a good radiative heat flow in the ON state and a minimum heat flow in the OFF state. To achieve a good radiative heat flow in the ON state, we require a high-emissivity exterior surface. To minimise the heat flow in the OFF state, the interior cavity surfaces should have as low an emissivity as possible. Since the ART will operate in high-vacuum there is no convective heat transfer, and if the support posts have a negligibly high thermal resistance all of the heat flow between the plates is radiative. A simple theoretical model of the OFF state thermal characteristics can be developed if the conductive heat flow Q_{ins} through the support posts can be ignored. Also assumed is the total reflection of solar energy by the radiator outer coating. Referring to Figure 2, the radiative heat transfer from the Base to the Radiator plate can be written in the form

$$Q_{r,cav} = A \cdot \sigma \cdot E_{br}\left(T_{base}^4 - T_{rad}^4\right) \tag{1}$$

where A is the area of the plates, σ is the Stephan-Boltzman constant, and

$$E_{br} = \left(\frac{1}{\varepsilon_{bc}} + \frac{1}{\varepsilon_{rc}}\right)^{-1} \tag{2}$$

with ε_{bc} and ε_{rc} the emissivities of the Base and Radiator surfaces of the cavity [1]. The interchange factor E_{br} accounts for reflection and absorption within the optical cavity. A similar equation describes the heat flow from the Radiator to the environment, thus

$$Q_{r,out} = A \cdot \sigma \cdot \varepsilon_{re}\left(T_{rad}^4 - T_{space}^4\right) \tag{3}$$

where ε_{re} is the emissivity of the Radiator exterior. In equilibrium, $Q_{r,cav} = Q_{r,out}$. Combining eq. 1 and eq. 3, we can derive the following two equations:

$$Q_{r,out} = A \cdot \sigma \cdot \frac{E_{br}\varepsilon_{re}}{E_{br} + \varepsilon_{re}}\left(T_{base}^4 - T_{space}^4\right), \tag{4}$$

$$T_{rad} = \sqrt[4]{\frac{E_{br}T_{base}^4 + \varepsilon_{re}T_{space}^4}{E_{br} + \varepsilon_{re}}}. \tag{5}$$

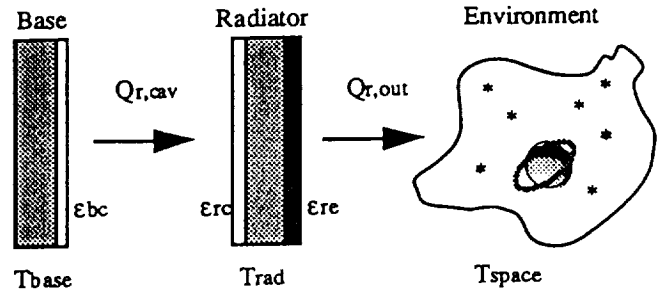


Figure 2. Optical cavity of ART. The net heat flow is from the Base to the Environment.

Since we require a high external emissivity, $\varepsilon_{re} \approx 1$, for small E_{br} the quantity $[(E_{br}\varepsilon_{re})/(E_{br}+\varepsilon_{re})]$ in eq. 4 reduces to simply E_{br}. Considering eq. 2 in conjunction with eq. 4 illustrates the potential for very high thermal isolation in the OFF state. For instance, if we set $\varepsilon_{bc} = \varepsilon_{rc} = \varepsilon$, with $\varepsilon \ll 1$, $E_{br} \approx \varepsilon/2$. The effective emissivity to the environment in the OFF state can be about half the emissivity of the cavity layers, which translates to a factor of two increase in effective thermal resistance between the heat source and the environment; while still retaining a high emissivity in the ON state.

CONSTRUCTION AND TESTING

The ART prototypes are constructed using typical silicon micromachining techniques with 5 inch 625 μm thick silicon wafers. For Base plates, the bare silicon wafers are coated with aluminum and a PECVD nitride insulating layer, then photoimageable polyimide is used to create support posts from 10 microns up to 50 microns in height. The heat radiation when actuated depends on good thermal contact between the two plates of the device when in contact, and the level of thermal contact depends on applied pressure [2], so it is necessary to have a very high quality insulating layer on the interior surfaces to allow high applied voltages (for high applied electrostatic pressures) without breakdown. This is essentially related to the mechanical stiffness of the membrane supporting the radiator plates: the Radiator plates are constructed using timed double-sided anisotropic etching to create a membrane 5 to 10 μm thick with a large central contacting area approximately 300 μm thick. The size of the prototypes in this study is approximately 1 inch square. The voltage required for good thermal contacting would have been much less with a corrugated [3] membrane supporting the radiator.

The prototypes were tested in a small turbopumped vacuum chamber at pressures between 10^{-6} mbar and 10^{-5} mbar. Higher background pressures (> 10^{-4} mbar) would noticeably degrade the thermal isolation between the two plates. The Base of each device was attached with silver epoxy to an aluminum heater block assembly; the heater assembly was then mounted via thermal insulators in the vacuum chamber. A scanning infrared camera thermometer was used to monitor the temperatures of both the Radiator and the Base (where the temperature of the Base is equal to the temperature of the heater block) through an infrared window. The temperature of the heater block was also monitored with a thermocouple. Calibration of the infrared camera to correct for the infrared window absorption was accomplished by viewing the infrared signal from both the Radiator and the heater block as the block was ramped to various temperatures in both air and vacuum. The Base temperature range used for testing was approximately 40°C to 165°C.

Two different styles of prototypes were tested; assemblies which were only suitable for determining the OFF state thermal isolation, and assemblies which could be actuated and demonstrate both ON and OFF states.

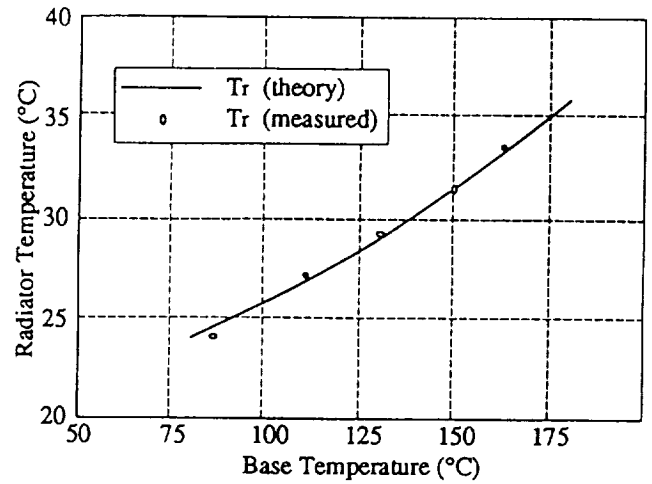


Figure 3. Temperature of the Radiator exterior versus the Base temperature. Values for the theoretical curve are: ε_{bc}=0.14, ε_{rc}=0.07, ε_{re}=0.94, and T_{space} = 20°C.

For testing the OFF state, ART devices were built which resembled as closely as possible the idealized device, including a high-emissivity layer of black paint on the exterior. Stress in the black paint applied to the exterior of the Radiator plates would curve the Radiators to the point that they could not be used for actuation.

The testing of the ON state required ART devices with a non-ideal exterior surface of bare aluminum, which had a serious impact on the heat flow which could be obtained to the environment. Since the amount of heat radiated to the environment is proportional to the emissivity of the Radiator exterior, in the actuatable devices the heat differential between the Radiator and the Base in the OFF state was much less than for the idealized Radiators.

The ART prototypes were actuated using capacitance-voltage (CV) measurement equipment which could apply up to 40 volts DC. This made it possible to monitor the motion of the ART device once a voltage was applied, via the change in capacitance. The maximum voltage which could be applied was limited by defects in the insulating PECVD layer; however the voltage range was sufficient to clearly demonstrate a variable thermal conductivity as a function of applied voltage.

In parallel, a process was investigated that allows for integration of a high emissivity coating with thin micromachined devices. Anodised aluminum was selected since it combines high emissivity ε_{IR} in the infrared with high solar reflectance R_S, e.g. low solar absorptance α_s=1-R_S [4, 5, 6].

RESULTS

OFF state. The ART prototypes achieved an OFF state thermal isolation very close to that predicted from the simple thermal model, demonstrating a thermal behaviour dominated by radiative heat transfer from the Base to the Radiator plate. Figure 3 shows a graph of the Radiator temperature as a function of the Base temperature calculated using eq. 5, and the measured values. The emissivity values used to derive the theoretical curve were measured on the actual device layers. The agreement between the simple model and the measured values is excellent. Note that the Radiator plate, which is spaced only 10 μm from the Base, can remain at a temperature only 15°C above room temperature while the Base is heated to over 160°C.

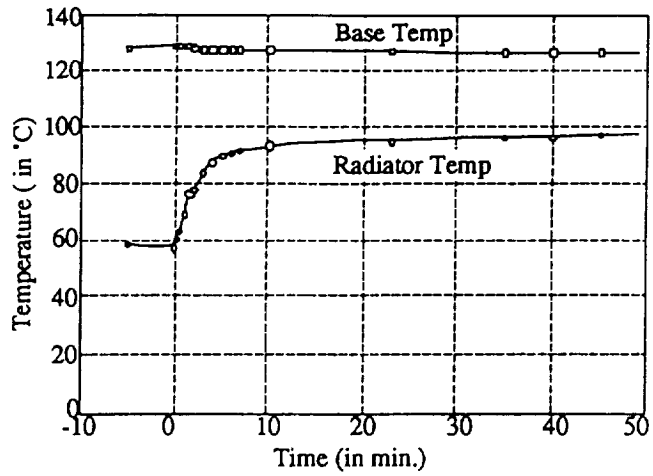


Figure 4. Temperature of Radiator after actuation.

ON state. The ability to control the heat flow to the environment was demonstrated. The ART Radiator temperature as a function of time after device actuation can be seen in Figure 4. As expected, the temperature of the Radiator increases once the device is actuated. In this case, there is a rather large temperature difference between the Radiator and the Base due to the low actuation voltage. Note that the temperature of the Base, which is heated with a constant power, drops slightly due to the heat flow into the Radiator and from there to the environment.

Actuating the device at a higher applied voltage would result in a smaller temperature difference between the Radiator and the Base and a faster response time during the actuation. This can be seen in Figure 5, which shows the temperature response of the device as a function of the applied voltage. Note that appreciable turn-on voltages are much higher than the voltages required for simple physical contact; this is expected as the thermal contact resistance decreases with increasing applied pressure[2]. Figure 5 also indicates that the current study gathered data at the very minimum voltage required for appreciable thermal switching.

Mechanical. Representative results of CV testing are also shown in Figure 5. The relative capacitance (arbitrary units) at low voltages shows a parabolic increase in capacitance attributed to

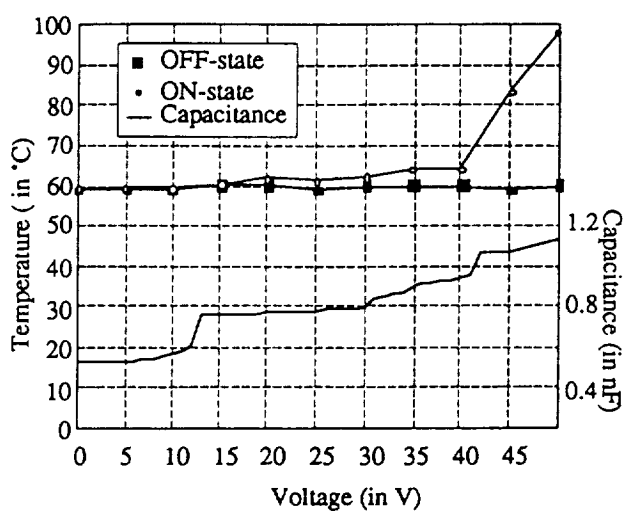


Figure 5. Temperature of Radiator after 5 min. actuation at indicated voltage, and relative capacitance.

the movement of the Radiator plate in response to the applied potential; then a sudden increase in capacitance as the Radiator plate comes into contact with the Base plate. The initial plateau at 12 volts was verified as the point of initial contact at the center of the contacting boss by observing the motion of the Radiator under an optical microscope. Several smaller disparities can be seen in the capacitance as the voltage is increased; this could be due to the different corners of the square membrane boss coming into contact. Note that thermal switching does not occur until the applied voltage is much higher than the voltage needed for simple contact.

The DC current into a device actuated at 50 volts was measured at 10 nA. The steady-state power required is therefore 500nW for the prototype, which compares with a transmitted heat flow of about 20 milliwatts; so the temperature rise during actuation is not due to joule heating. This power consumption is less than 1 milliwatt per square meter.

High emissivity coating. The high IR emissivity (ε_{IR}) coating was developed based on a standard process for the anodisation of aluminum in a sulphuric acid solution [7]. This process results in the formation of alumina with a reported value of $\alpha_s/\varepsilon_{IR}$ that can be as low as 0.19 [4, 5, 6]. Parameters for our study were the thickness of the as evaporated aluminum and the anodization time, which determines the thickness of grown alumina. Spectral reflectance in the 0.3 to 2 µm range, where most of the solar energy is concentrated, was measured using an integrating sphere set-up for diffuse reflectance measurements. Infrared spectral emissivity ε_{IR} was measured separately between 2.5 µm and 25 µm using a specular reflectance set-up (the spectral emissivity is calculated as $\varepsilon_{IR}=1-R_{IR}$). It can be seen from figure 6, where data for wavelenghts that are typical of the spectra of concern in satellite cooling are plotted, that the IR emissivity saturates about 0.9 when the alumina thickness reaches a few microns.

However, the solar reflectance drops very quickly for increasing alumina thickness, so that we could not reach the 0.19 value published for $\alpha_s/\varepsilon_{IR}$. This undesirable trend appears to be related to the thickness of as evaporated aluminum. Our efforts to increase Rs in further developments should should be concentrated on the aluminum itself from which alumina is grown.

CONCLUSION

The ART prototypes have proven the principles of operation of a thermal switch for satellite temperature control. The prototypes demonstrate a high degree of thermal isolation in the OFF state, a low actuation voltage and a corresponding low operating power, and when actuated drastically decrease their thermal resistance and radiated heat to the environment.

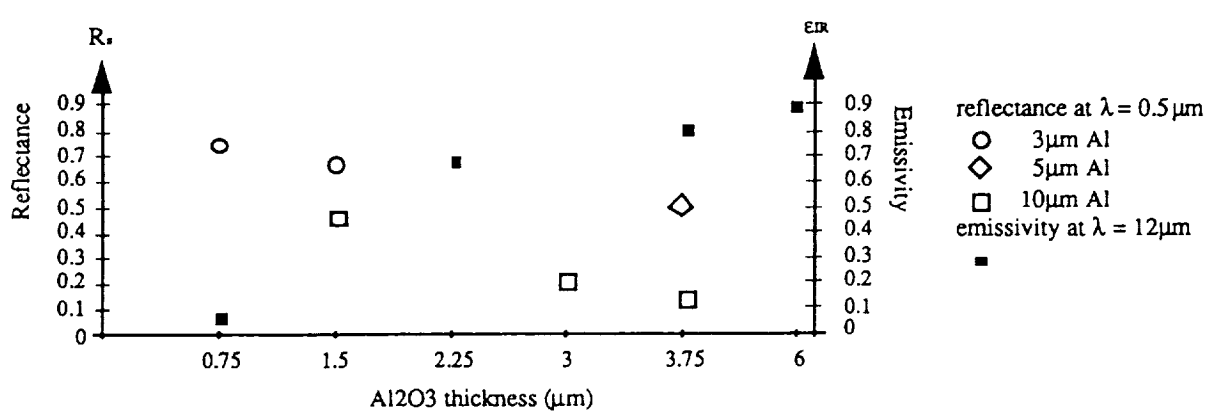


Figure 6. Measured solar reflectance and infrared emissivity as a function of the thickness of grown alumina for various thicknesses of the as deposited aluminum

A new design of thermal switches with lower actuation voltage, fully processed in a micromachining technology and that could resist to launch conditions is presently under development.

ACKNOWLEDGEMENTS

The authors wish to thank K. Said, Luc and Alain for experimental set-up and measurements, Agnes and Andre for invaluable technical assistance and advice. This work was supported by IWT/ Verhaert D&D.

REFERENCES

[1] Slater et al; "ART: A Novel Thermal Valve for Temperature Control Applications". *J. Micromechanics and Microengineering*, vol.5, no.2 (1995) pp.186-188

[2] Yüncü et al, "Thermal Contact Conductance - Theory and Applications". *Proc. NATO Adv. Study Inst. on Cooling of Electronic Systems*, 1993, pp. 29-54

[3] J. H. Jerman. "The fabrication and use of micro machined silicon diaphragms". *Sensors and actuators*, vol. **A21-A23** (1990)

[4] D. Weber, "Spectral emissivity of solids in the infrared at low temperature". *J. Opt. Soc. Am.*, **49-815** (1959)

[5] G. Haas, "temperature stabilization of highly reflecting spherical satellites". *J. Opt. Soc. Am.*, **49-918**(1959)

[6] L. Johnson, "Spacecrafts radiators". Space/Aeronautics, 37,76 (January 1962)

[7] F. A. Lowenheim "Deposition of inorganic films from solution", Thin film processes, Academic Press (1978), p251

OPTICALLY-SWITCHED RESONANT TUNNELING DIODES FOR SPACE-BASED OPTICAL COMMUNICATION APPLICATIONS

71180

T.S. Moise, Y.-C. Kao, D. Jovanovic, and P. Sotirelis
Corporate Research and Development
Texas Instruments Inc., Dallas, TX 75265

ABSTRACT

We are developing a new type of digital photoreceiver that has the potential to perform high speed optical-to-electronic conversion with a factor of 10 reduction in component count and power dissipation. In this paper, we describe the room-temperature photo-induced switching of this InP-based device which consists of an InGaAs/AlAs resonant-tunneling diode integrated with an InGaAs absorber layer. When illuminated at an irradiance of greater than 5 Wcm^{-2} using 1.3 μm radiation, the resonant-tunneling diode switches from a high-conductance to a low-conductance electrical state and exhibits a voltage swing of up to 800 mV.

1. Introduction

Relative to conventional wire or cable technologies, optical fiber communication systems offer reduced weight, higher bandwidth and freedom from electromagnetic interference.[1] These attributes are particularly important for space-based applications in which reduced weight and power dissipation as well as electromagnetic interference immunity are critical. In recent years, satellite-based optical links have been demonstrated[2] and high-speed fiber optic networks for space-based data bus applications have been developed.[3]

The bandwidth, sensitivity, cost, and power consumption of a fiber optic communication system are controlled to a large extent by the characteristics of the optical receiver unit. Conventional receivers for these applications employ hundreds of transistors and consist of three primary subsystems: a photodetector, a low-noise amplifier, and a level-restoring comparator. For high-speed, long distance telecommunication systems, it is necessary to use a detector made from InGaAs which efficiently absorbs photons at the minimum dispersion and loss wavelengths of the optical fiber. A high speed, transimpedance amplifier, typically made from GaAs-based circuits, converts the optically induced photocurrent to a voltage signal which is then digitized. For high speed systems (> 1 Gb/s), more than 1 W of electrical power is required to convert the incident optical signal to its electronic equivalent.

In this paper, we describe the room-temperature photoinduced switching of an InP-based device which consists of an InGaAs/AlAs resonant-tunneling diode integrated with an InGaAs absorber layer. When illuminated at an irradiance of greater than 5 Wcm^{-2} using 1.3 μm radiation, the resonant-tunneling diode switches from a high-conductance to a low-conductance electrical state and exhibits a voltage swing of up to 800 mV. The switching characteristics are reversible and, in the absence of light, the detector returns to its original high conductance operating state.[4]

2. Description of Device Operation

A schematic device cross-section and computed energy band diagram[5] for the ORTD are shown in Figs. 1 and 2, respectively. The epitaxial layers employed in these experiments have been grown by molecular beam epitaxy using elemental group III and group V sources.[6] As shown in Fig. 1, a thick 0.5 μm n^+ (5×10^{18} cm^{-3}) $In_{0.53}Ga_{0.47}As$ layer is epitaxially grown on the InP semi-insulating substrate. Following this heavily doped region, a lighter doped, 30 nm thick n-type (1×10^{18} cm^{-3}) $In_{0.53}Ga_{0.47}As$ layer is deposited followed by 100 nm of undoped $In_{0.53}Ga_{0.47}As$ which serves both to absorb the incident photons and to increase the peak-to-valley voltage swing of the diode.

An AlAs/ $In_{0.53}Ga_{0.47}As$/ AlAs (28 Å/ 50 Å/ 28 Å) RTD is then grown on the undoped $In_{0.53}Ga_{0.47}As$. Following the RTD, a thin (2 nm) $In_{0.53}Ga_{0.47}As$ spacer layer, a 30 nm n-type (1×10^{18} cm^{-3}) $In_{0.53}Ga_{0.47}As$ layer, and 250 nm of n-type (5×10^{18} cm^{-3}) $In_{0.52}Al_{0.48}As$ are subsequently deposited. A thin (400 Å) n+ (5×10^{18} cm^{-3}) InGaAs cap layer is then grown on top of the device. This emitter structure serves to minimize optical absorption within the top contact layers while maintaining a relatively low contact resistance of approximately 1×10^{-5} Ωcm^2, as derived by mesa-etched transmission line data. To minimize parasitic effects, we have employed airbridge contacts to the mesa-isolated ORTD devices.

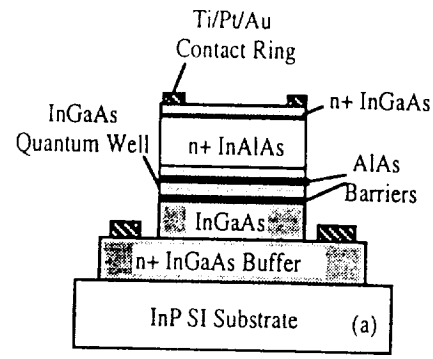


Fig. 1 Schematic cross-sectional diagram for the optically switched resonant-tunneling diode. The epitaxial layers are grown by molecular beam epitaxy within the material science laboratory of Texas Instruments.

The physical basis for ORTD operation can be explained with reference to the energy band diagram shown in Fig. 2 with 1.5 V applied bias. The external voltage is dropped primarily across the 100 nm undoped InGaAs layer because the width of this undoped layer is much larger than that of the double barrier structure (approximately 10 nm). For photon energies greater than 0.75 eV, electron hole pairs are created within this undoped layer. Under the influence of the external electric field, the photo-generated electrons are accelerated toward the n+ collector while the holes accumulate at the collector barrier of the ORTD. This process is similar to that described by England et. al.[7] The accumulated charge partially neutralizes the electric field within the undoped layer which, in turn, leads to a local enhancement of the electric field within the double barrier structure. As a result, the current-voltage characteristics for the illuminated ORTD are shifted to lower voltage relative to the dark response.

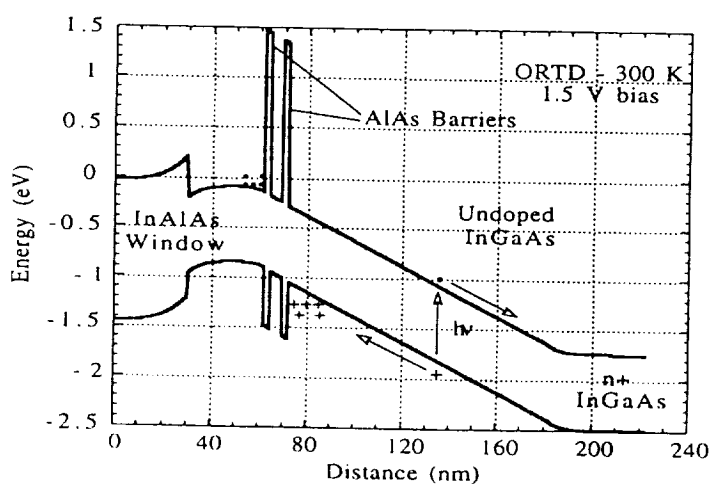


Fig. 2. Computed energy band diagram[5] for the optically switched resonant-tunneling diode. Under bias, electron-hole pairs created within the undoped InGaAs layer screen the electric field present within this layer.

3. Experimental Results

Under dark conditions, the 60 μm diameter ORTD exhibits a peak voltage and current of 1.7 V and 24 mA, respectively, as depicted by the solid curve in Fig. 3. This bias polarity corresponds to electron injection from the surface towards the substrate. With light incident upon the ORTD, the magnitude of both the peak and the valley voltage are reduced as the result of electric field screening within the undoped intrinsic region. Thus, with an incident power of 1.0 mW at 1.3 μm, corresponding to an irradiance of 30 Wcm^{-2}, the peak voltage and the valley voltage are reduced by approximately 140 mV and 120 mV, respectively.

From this reduction in peak voltage, we can obtain an order-of-magnitude estimate for the steady state density of photogenerated holes in the following manner. First, we estimate the photo-induced reduction in electric field by dividing the observed voltage shift by the undoped spacer layer thickness. Second, we calculate the density of accumulated holes by using an infinite sheet charge approximation. Thus, from the measured 100 mV shift in peak voltage, we determine that the electric field is reduced by 1×10^4 V/cm within the 100 nm spacer layer leading to an estimated accumulated hole density of 6×10^{10} cm^{-2}. To a first approximation, we anticipate that the accumulated hole density and the corresponding voltage shift will vary linearly with irradiance. However, at high irradiance (> 100 Wcm^{-2}), the experimental results suggest that hole tunneling through the double barrier structure, as evidenced by an increased valley current during illumination, leads to a saturation of the observed voltage shift. For negative polarity, we note that the peak voltage and current are nearly independent of irradiance as the result of electron accumulation at the InGaAs/AlAs interface.[8]

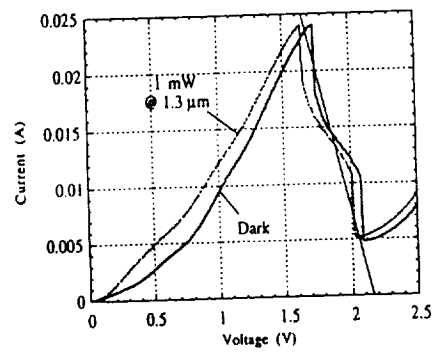


FIG. 3. Room-temperature current-voltage characteristics for the 60 μm ORTD under dark and illuminated conditions at a wavelength of 1.3 μm.

When biased with a resistor load, this photo-induced voltage shift is used to switch the device from a high-conductance to a low-conductance operating state, as illustrated by the load line drawn in Fig. 3. The bias values, $V_{applied}$ = 2.2 V and R = 20 Ω, are selected such that the load line intersects the dark current voltage characteristic near the peak voltage and intersects the illuminated characteristic near the valley voltage. By biasing the ORTD in this manner, we have created a bistable optical-to-electronic converter which operates at 1.7 V (23 mA) in the dark and 2.1 V (6 mA) when illuminated with an irradiance above the threshold value (determined by the load line and photo-induced voltage shift) of 20 Wcm^{-2}. For this measurement, the ORTD was illuminated at approximately 5 Wcm^{-2} using 1.3 μm radiation. We note that the plateau region observed in the current-voltage characteristics between 1.7 V and 2.0 V results from measurement circuit instability (i.e. oscillation) and does not constitute a stable operating point of the detector.

A timing diagram showing the measured output voltage of an ORTD together with the input laser drive signal is shown in Fig. 4. Under dark conditions, the ORTD is in a high-conductance state leading to a low output voltage (1.7 V).

With illumination, the resonant-tunneling structure switches into a low-conductance state leading to a high output voltage (2.1 V). Thus, this single device exhibits an output voltage characteristic similar to the more complex multi-transistor receiver circuits which require an order of magnitude more active components.

We have measured the large signal switching characterisics of the ORTD at higher frequency with the results shown in Fig. 5. For this measurment, the ORTD is biased through an RF bias tee which leads to a pulse width of approximately 7 ns. On this scale, the transition times are found to be less than 1 ns. Additional experiments are currently underway to understand the pulse width dependence upon RF bias conditions. The high speed measurements yield an output voltage swing of nearly 800 mV, which is double the voltage swing achieved at low frequency. This occurs because, at high frequency, the output is capacitively coupled and the voltage swing is the product of the optically induced current change through the ORTD (16 mA - see Fig. 3) and the 50 Ω input impedance of the oscilloscope.

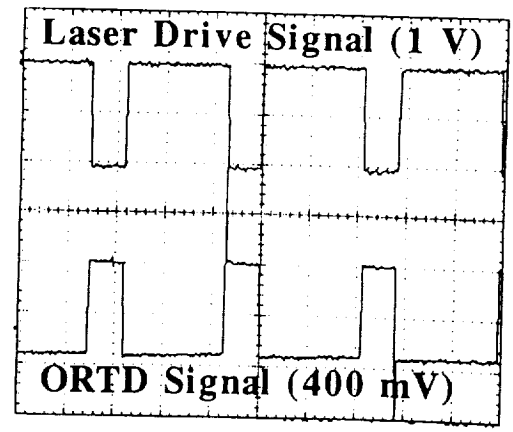


FIG. 4. A low-frequency (1 KHz) timing diagram of the measured output voltage of the ORTD together with the input laser drive signal . The bias conditions for this measurement are 2.1 V through a 100 W resistor.

In an attempt to estimate the upper limit for the ORTD large-signal switching speed, we have measured the on-wafer small signal response using a 20 μm diameter ORTD biased at 1.0 V through a 30 W resistor. The results are shown in Fig. 5. For these impulse measurements, the ORTD was illuminated with a 250 fs pulse at 1.3 μm generated using optical parametric amplifier system. The 3 dB bandwidth was found to be approximately 4 GHz. Additional tests with higher bandwidth emitters are currently underway to further characterize the frequency response limits of the ORTD.

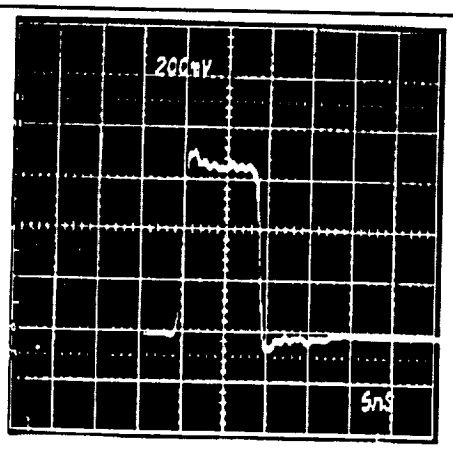


FIG. 5. A high-frequency timing diagram of the measured output voltage of the ORTD when irradiated with a 250 fs, 1.3 μm optical pulse. The frequency response is limited bias elements associated with the RF bias tee. These measurements were performed at the University of Texas microelectronics center.

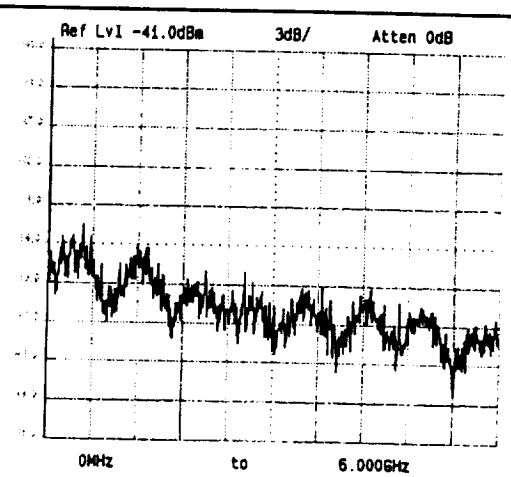


Fig. 6. The measured small-signal frequency response of the 20 μm ORTD when illuminated with a 250 fs, 1.3 μm impulse signal. These measurements were performed at the University of Texas microelectronic center.

4. Conclusions

In summary, we have developed an optically-switched, resonant-tunneling diode that exhibits up to a 800 mV output voltage swing when illuminated at an irradiance on the order of 5 Wcm^{-2} using 1.3 μm radiation. The switching characteristics are reversible and, in the absence of light, the detector returns to its original high conductance operating state. Small-signal optical measurements of the ORTD biased prior to resonance demonstrate a 3 dB bandwidth of approximately 4 GHz.

5. Acknowledgments

We gratefully acknowledge the assitance of Professor Joe Campbellof the University of Texas in performing the ultrafast impulse measurements. We also acknowledge many useful discussions with T.P.E. Broekaert, G.A. Frazier, and A.C. Seabaugh along with the outstanding technical assistance of E. Pijan and R. Thomason all of Texas Instruments.

REFERENCES

1) E.J. Friebele, M.E. Gingerich, and D.L. Griscom, SPIE **1791**, p. 177 (1992).
2) E.W. Taylor, J.H. Berry, A.D. Sanchez, R.J. Padden, S.P. Chapman, 1992 DOD Fiber Optics Conference Proceedings, p. 25 (1992).
3) J. Bristow and J. Lehman, SPIE **1953**, p. 159 (1993).
4) T. S. Moise, Y.C. Kao, L.D. Garrett, and J.C. Campbell, Appl. Phys. Lett., **66** (9) p. 1104 (1995).
5) BANDPROF heterojunction device simulator, W.R. Frensley, 1993.
6) Y.-C. Kao, A.C. Seabaugh, H.Y. Liu, T.S. Kim, M.A. Reed, P. Saunier, B. Bayraktaroglu, and W. M. Duncan, Proc. S.P.I.E. **1144** p. 30 (1989).
7) E.T. Koenig, C.I. Huang, and B. Jogai, J. Appl. Phys. **68** (11) p. 5905 (1990).
8) P. England, J.E. Golub, L.T. Florez, and J.P. Harbison, Appl. Phys. Lett., **58** (9) p. 887 1991).

A MICROMECHANICAL INS/GPS SYSTEM FOR SMALL SATELLITES

N. Barbour, T. Brand, R. Haley, M. Socha, J. Stoll, P. Ward, M. Weinberg

Charles Stark Draper Laboratory, Cambridge, MA 02139

Abstract

The cost and complexity of large satellite space missions continue to escalate. To reduce costs, more attention is being directed toward small lightweight satellites where future demand is expected to grow dramatically. Specifically, micromechanical inertial systems and microstrip GPS antennas incorporating flip-chip bonding, ASIC, and MCM technologies will be required.

Traditional microsatellite pointing systems do not employ active control. Many systems allow the satellite to point coarsely using gravity gradient, then attempt to maintain the image on the focal plane with fast-steering mirrors. Draper's approach is to actively control the line-of-sight pointing by utilizing on-board attitude determination with micromechanical inertial sensors and reaction wheel control actuators.

Draper has developed commercial and tactical-grade micromechanical inertial sensors. The small size, low weight, and low cost of these gyroscopes and accelerometers enable systems previously impractical because of size and cost. Evolving micromechanical inertial sensors can be applied to closed-loop, active control of small satellites for microradian precision-pointing missions.

An inertial reference feedback control loop can be used to determine attitude and line-of-sight jitter to provide error information to the controller for correction. At low frequencies, the error signal is provided by GPS. At higher frequencies, feedback is provided by the micromechanical gyros. This blending of sensors provides wide-band sensing from dc to operational frequencies.

First-order simulation has shown that the performance of existing micromechanical gyros, with integrated GPS, is feasible for a pointing mission of 10 microradians of jitter stability and ~1 milliradian absolute error, for a satellite with 1 meter antenna separation. Improved performance micromechanical sensors currently under development will be suitable for a range of micro-nano-satellite applications.

1. Introduction

The cost and complexity of conducting space missions with individual, highly integrated, large satellites continue to escalate. To reduce costs, more effort and attention is being directed toward small, light-weight satellites where future demand is expected to grow dramatically. During the past few years significant effort has been made to develop the "smaller, better, faster" approach being embraced by the aerospace industry as a means to prepare for space missions of the future. Recent advances in silicon microfabrication technology have led to the development of low cost micromechanical inertial sensors. The small size, low weight, and low cost attributes of these sensors permit on-board insertion of gyroscopes and accelerometers for space applications previously considered impractical because of size and cost considerations.

A micro-nanoclass spacecraft with a precision pointing requirement is a unique design challenge for the hardware, instrument, and payload designers. This class of satellite is an ideal application for micromechanical gyros to provide the inertial reference information needed to perform image motion compensation for stabilization and the reduction of image smear in an extremely small and light weight package.

Existing microsatellite pointing systems do not employ active control to improve pointing performance. Some systems allow the bus to point coarsely, then attempt to maintain the image on the focal plane with fast-steering mirrors. GPS alone will not provide sufficient accuracy. However, line-of-sight pointing can be controlled by on-board attitude determination from micromechanical inertial sensing, coupled with GPS, and closed-loop error correction and attitude control actuators. Others have not undertaken this approach since traditional inertial reference systems are significantly heavier and larger than the micromechanical package, and could consume a significant fraction of the mass budget for an entire micro-class satellite.

The concept described herein consists of a three axis stabilized system using on-board GPS, next-generation micromechanical inertial instruments, small reaction wheels for attitude control, and miniaturized phased array antennas for line-of-sight communications. The basic structure will be advanced graphite-epoxy composite materials. Material fabrication is kept cost effective by using simple shapes and optimizing the overall configuration. Body-fixed solar cells, as well as batteries, are used. This flexibility permits selection and adjustment of the center of gravity coincident with the center-of-pressure to minimize rotational torques due to atmospheric drag.

The feasibility of a micromechanical gyro to provide sufficient performance to accomplish a pointing mission has been analyzed for a bandwidth of 0.1 Hz, based on estimated external and induced disturbance frequencies. From a first-order simulation, it was determined that the performance of existing micromechanical gyros with integrated GPS is capable of providing pointing jitter stability better than 10 microradians and absolute pointing error of approximately 1 milliradian, for a satellite with 1 meter antenna separation (i.e., 1m baseline). This result is encouraging and gives credence to the overall concept and approach. For a satellite with ~100 mm baseline, pointing accuracy will be degraded by more than a factor of 10 unless more accurate micromechanical instruments are used.

2. Pointing Scenario

A typical pointing scenario assumes that the satellite will process GPS-derived attitude measurements (and perhaps attitude-rate) and gyro-derived attitude rate measurements while in orbit to establish estimates of gyro drift bias and scale factor (SF). Satellite maneuvers can be used to separate the error components due to bias and SF. Accelerometer measurements can be used to provide delta velocity in cases where the satellite may require orbit changes.

The satellite is pitched forward from a nominal nadir-pointing attitude to some new attitude, and then pitched in reverse during imaging to reduce the transverse velocity of the line-of-sight at the surface of the earth. Without this counter motion, the satellite velocity over an integration period will result in too much blur of the picture. The amount by which the satellite must be initially pitched forward in preparation for the imaging slew depends upon the amount of time required to settle out initial pitch slew transients once the imaging slew begins. The time requirement, in turn, depends upon the controller bandwidth. A higher controller bandwidth will result in transients subsiding quicker but will result in a greater response to sensor noise. A 0.03 Hz controller bandwidth requires an initial attitude of a least 60° of the nadir and a 0.1 Hz controller bandwidth requires less than 30° in order to satisfy the requirement that the additional drift (i.e., in addition to the nominal drift due to satellite motion) over an integration period should not exceed the equivalent of 1 pixel motion.

During this large angular rotation in a short period of time there will likely be several changes of GPS satellites and, possibly, a large variation of attitude precision. Therefore, attitude measurements from GPS will be effectively suppressed during the imaging slew and attitude inputs to the controller will rely essentially upon calibrated gyro measurements. Calibration of the gyros depends upon the effectiveness of the integrated GPS/gyro filter.

If gyro calibration is not sufficiently accurate, it may be necessary to pause at the end of the preparatory pitch-forward maneuver to regain the required absolute attitude accuracy by again processing GPS attitude measurements at the new pitch attitude. If this angle is large, a second set of GPS antennas might have to be used to obtain satisfactory satellite visibility. On the other hand, if the angle is small, as required for the 0.1 Hz controller, a single set of antennas may suffice for both the nadir and off-nadir satellite attitudes.

Table 1 shows the component requirements for a conceptual imaging mission with an absolute pointing requirement of 1 mrad and a jitter requirement of 0.5 microradian, for a satellite baseline of 1 meter.

Table 1. Component Requirements

Function	Requirements	
Attitude Control System	Jitter Slew Rate Absolute	0.5 microradian over 14 ms integration 0.75°/s 1 milliradian (1σ)
Inertial Sensor*	Angle Random Walk Bias Drift Scale Factor	0.1 deg√hr 4 deg/hr (1σ) over 10 minutes 1000 ppm (1σ)
Control Actuators	On-board Disturbance	Negligible between 0.01 - 50 Hz 0.0001 N-m at any single freq. between 50 to 500 Hz

***Calibrated integrated system performance.**

3. INS/GPS Role

Miniature INS/GPS systems are currently under development at Draper. The system typically contains a Micromechanical Inertial Measurement Unit (MMIMU) comprising three orthogonal micromechanical accelerometer instruments and three orthogonal micromechanical gyro instruments with individual conditioning electronics and high-resolution digital output quantizers. The system also contains a GPS receiver as well as a guidance and navigation processor. This entire system is miniaturized further using low power mixed signal CMOS ASICs, high-performance bump bonding techniques, shared/multiplexed electronics functions and advanced high-density packaging. A single clock, voltage reference, A/D converter, and DSP are used to service the micromechanical INS/GPS system, resulting in a compact, low-power, multisensor, multi-axis system. The complete INS/GPS system occupies less volume than a hockey puck. Power dissipation is minimized both through the use of low power CMOS and also by using power conserving sleep modes.

As new fabrication technologies reduce the size of the electronic assemblies in GPS receivers and other portable communications equipment, the antenna and battery become limiting factors in determining the minimum size of the overall receiver package. Miniaturized GPS microstrip antennas, fabricated on high-dielectric ($k \geq 80$) constant ceramic substrates, are being developed at Draper. These permit significant reduction in size compared to antennas fabricated on common low-dielectric substrates. New analytical models that are unique in their application to both antenna design and test are required.

GPS Role

GPS provides position with accuracy on the order of a few meters, and provides a highly reliable technique for providing attitude using interferometry techniques. Attitude determination using GPS is based on a simple geometric principle: two distinct points uniquely define a line and three noncolinear points define a plane. Treating GPS antennas with known relative positions as the noncolinear points, the orientation of the plane in which they reside can be determined, as well as the orientation of the antennas within the plane. This is achieved by using the differences in phase of the incoming GPS carrier signal between the antennas, hence the term "GPS interferometry".

The ability to determine the translational and rotation state of a satellite is dependent on the number of GPS satellites in view and the resulting measurement geometry. Since the GPS system was designed for "Earth-bound" users, the GPS satellite radiation signal is directed toward Earth. Satellites

at altitudes up to approximately 1000 km will generally find 10 or more satellites "in view". At higher altitude, the user may be outside the main radiation beam of some GPS satellites. At altitudes above roughly 3000 km, the number "in view" may frequently drop below 4. Although 4 satellites in view is generally considered the minimum set to solve for both the three components of position and the use clock bias, studies have shown that acquisition of a single GPS satellite can be adequate to maintain orbit accuracy. Thus, satellites in either high altitude or highly elliptic orbits can maintain accurate position information.

The determination of attitude using GPS is based upon the carrier phase difference between two antenna tracking the same satellite. Figure 1 illustrates the basic measurement geometry. The angle θ is determined by knowing the baseline (**b**) in body axis and measuring the GPS signal delay (Δ**r**). Thus, θ can be determined as follows:

$$\theta = \cos^{-1}(\Delta r/b)$$

Extending this to three axes using more than one baseline allows complete determination of vehicle attitude.

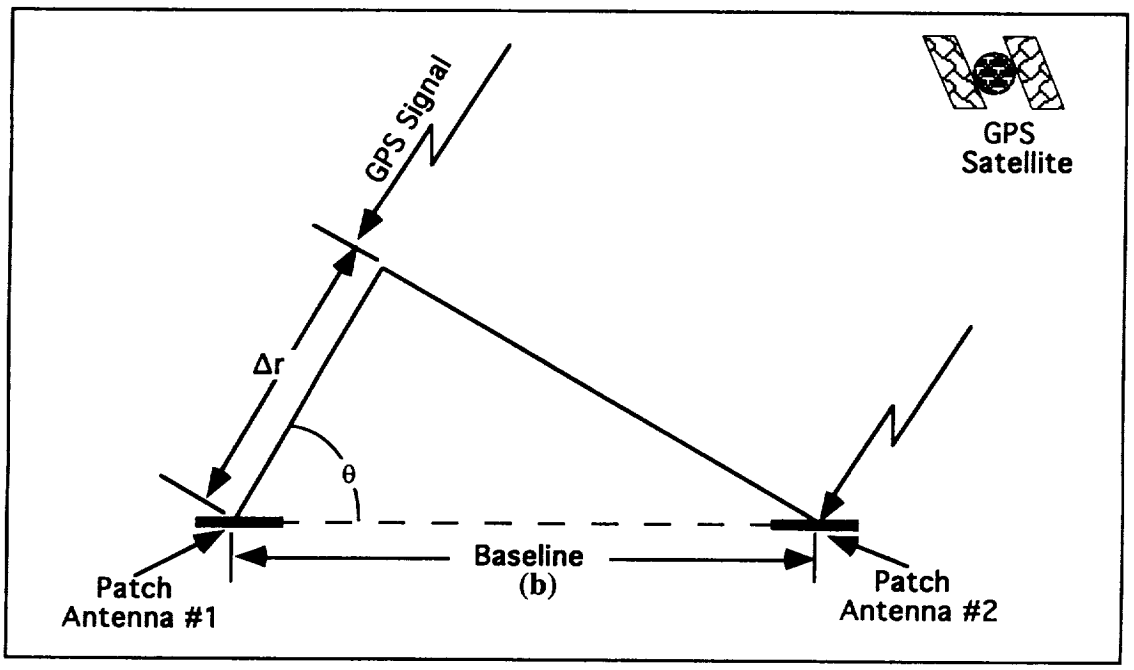


Figure 1. GPS Interferometry

One example of this technique is the GPS Attitude and Navigation Experiment (GANE), which will test a GPS interferometer intended for International Space Station Alpha (ISSA). It is currently scheduled to fly onboard the Shuttle in April, 1996. Mounted on a 1.5m by 3m platform in the cargo bay will be a four-antenna interferometer, along with an Inertial Reference Unit (IRU) for attitude verification. The requirements for this stand-alone GPS interferometer are to estimate the station's attitude to within 0.3 deg (3σ) per axis and attitude rates to within 0.01 deg/s (3σ) per axis at 0.5 Hz.

Micromechanical IMU Role

In a small satellite, the micro-mechanical IMU data can be blended optimally with the GPS data, thereby taking advantage of the IMU's ability to measure high-frequency dynamics and the GPS's ability to bound the error growth due to gyro drift.

A small satellite should have attitude determination capability at any orientation in space. The ability to determine attitude from a "cold start" without any prior knowledge is also important. Furthermore, the ability to measure rotation rates will be crucial to achieving stable conditions at any desired attitude. The micromechanical IMU can play a major role in alleviating these problems.

The ability to instantaneously measure via GPS all components of vehicle attitude can place severe requirements on spacecraft geometry and antenna mounting to ensure observability. Antennas must be mounted such that both antennas making up a baseline can observe the same satellite without masking problems. Furthermore, this condition must be met for more than one baseline and satellite. Using the micromechanical IMU (or gyro package) as a reference permits us to bridge small gaps in interferometer coverage in both a spatial and temporal sense. This can reduce the number of antennas required and mitigate problems arising from field-of-view masking due to factors such as solar arrays or other instrumentation.

Two other key roles for the IMU are control feedback for both rotational and translational maneuvers. First, attitude pointing and stabilization requires high-speed, low-noise measurements for effective control. GPS measurements alone (without an IMU) will be too noisy for effective attitude control if antenna baselines are short. Second, translational maneuvers can be provided with instant 3-axis measurement of Δv maneuvers via the IMU to perform orbit placement or adjustment. GPS has the ability to measure velocity changes very precisely via carrier tracking, and therefore could be used in place of the IMU accelerometers. Use of the IMU is more straightforward and eliminates antenna masking and satellite observability concerns.

4. INS/GPS Subsystem Integration

Figure 2 describes the satellite INS/GPS subsystem in which the flight processor navigation blends the INS/GPS data to determine translational and rotational state. Alternately, INS data could be supplied to the receiver for optimal estimation, or a cascaded design could be used. In this option, the GPS and INS data are blended at the translational/rotational state level.

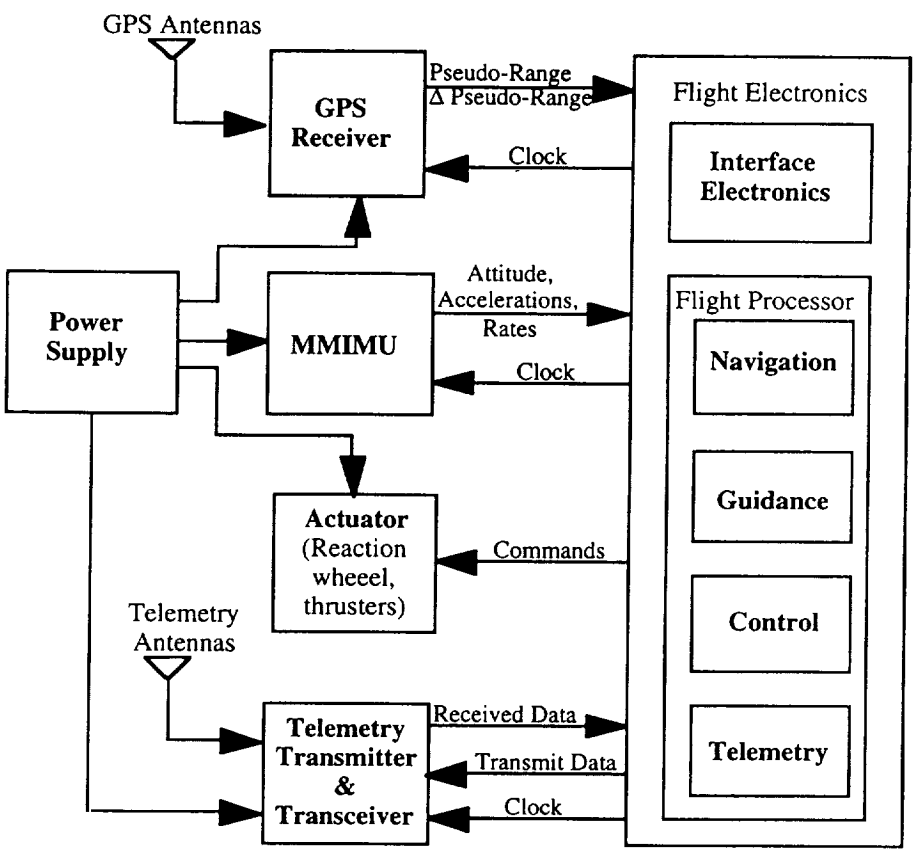


Figure 2. INS/GPS Subsystem

5. Micromechanical IMU

Draper Laboratory has been developing micromachined silicon gyroscopes and accelerometers since 1984. Micromachining uses process technology developed by the integrated circuit industry to fabricate tiny sensors and actuators. Very small-size and low-cost instruments have been microfabricated at Draper from single-crystal silicon anodically bonded to a glass substrate (dissolved wafer process). These instruments are described in more detail in the references.

These sensors are extremely rugged, and easy to fabricate. The gyro senses the Coriolis acceleration of the proof masses operating as a double-ended tuning fork. Sense and drive are electrostatic. The present units demonstrate 0.5°/s bias stability over a temperature range of -40 to +85°C without thermal compensation or control; stability at room temperature is <30°/h; angle random walk is $0.3°/\sqrt{h}$. The accelerometer is a pendulous seesaw. At room temperature, bias stability is 1 milli g and velocity random walk is $0.05\ m/s/\sqrt{h}$.

The next evolution of these sensors is in progress towards goals of 1°/hr gyro bias stability, 0.1% scale factor error, and 100 μg accelerometer bias stability. These performance goals are sufficient for several micro-nano-satellite applicatons, including the pointing scenario described in Section 2.

6. Microminiature Electronics

The micromechanical sensors are of extremely small size, with each sensor occupying an area of less than 4 square millimeters. It is therefore consistent with the applications for these small sensors to develop mating electronics which are of minimum size, power, and cost.

Draper has developed multi-chip modules (MCMs) and mixed-signal application-specific integrated circuits (ASICs). Both technologies are necessary for the implementation of micromechanical (MEMs) systems. Flip-chip (or bump) bonding is required for alignment accuracy between sensors and to fiducials (reference lines and edges or surfaces of circuit board upon which components are mounted). Presently, the optimal in small size, low power and low production cost can be achieved using mixed-signal CMOS ASICs. A first-iteration of the rate-gyroscope electronics has been designed and implemented on a single mixed-signal CMOS ASIC. The gyroscope sensor, ASIC, and supporting components are placed in a one inch square flat-package as shown in Figure 3. Total power dissipation for the gyroscope ASIC is a small fraction of a watt.

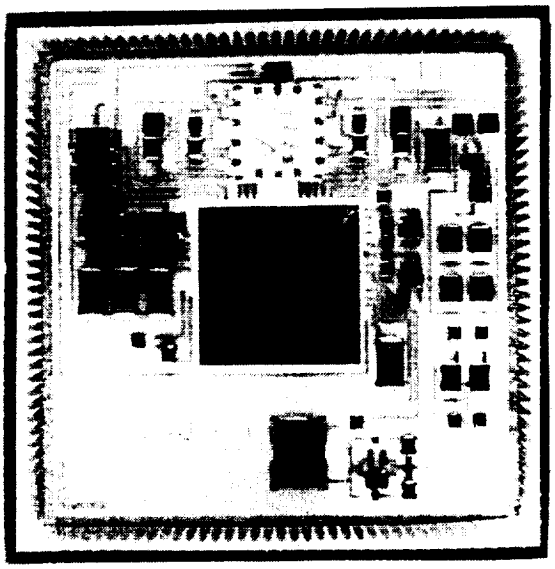


Figure 3. 1 inch by 1 inch Gyro and ASIC Package

Draper has developed a mixed-signal ASIC which will support a number of micromechanical accelerometer designs ranging from micro-g accelerations to guided munitions applications requiring measurement of 100,000 g accelerations. The accelerometer ASIC includes a high resolution digital output as well as continuous automatic self-test.

Electronics to support an entire inertial measurement unit can be placed on a single ASIC. In conjunction with the efforts toward increased miniaturization, are efforts toward improved performance and lower power. Near term goals are to develop a wide-bandwidth DC-coupled miniature gyroscope instrument with <10 degree/hour bias error and 0.1% scale-factor error which uses a fraction of the power of the present ASIC.

7. INS/GPS Accuracy

Figure 4 shows how the integrated INS/GPS system improves absolute pointing accuracy for a 1 meter baseline satellite. The GPS attitude errors are highly dependent on multipath environments. As sampling time increases, the integrated system reaches the calibrated performance of the inertial sensors.

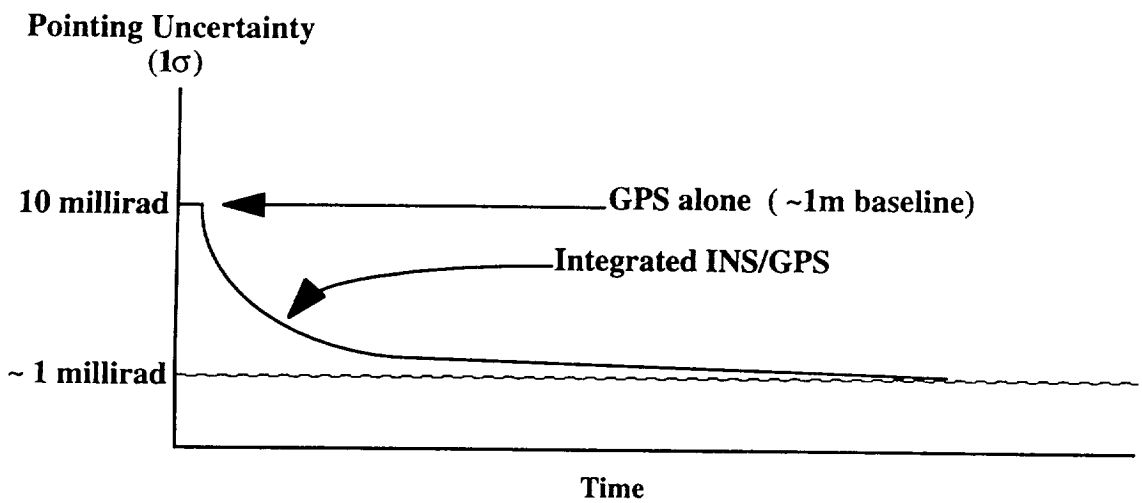


Figure 4. Pointing Accuracy with Integrated INS/GPS

The attitude accuracy potential using GPS is dependent on many factors. A very small satellite on the order of 0.5m or less can present a special challenge due to short measurement baselines, but small size can also mitigate problems from vehicle structural flexibility and carrier wave ambiguity. Major error sources in the GPS observed attitude solution are discussed below; areas where the micromechanical IMU can alleviate these problems are identified.

Receiver Noise

The calculation of the incoming signal's phase angle is subject to error induced by the hardware itself. The phase error depends on the receiver design quality, as well as the vehicle dynamics. If the satellite dynamics are minimal, the risk of losing carrier lock is minimal and narrow-band tracking loops can be used to extract very accurate phase difference measurements.

The biases from the frequency standard and receiver hardware cancel in differencing these measurements. Thus phase difference measurements should be corrupted by less than a millimeter after passing through the RF switch at the receiver's front end. One millimeter is ~2° of tracking error. If we assume an antenna baseline of 1 meter, then the angular uncertainty introduced by a 1 mm tracking error at each end of the baseline is roughly 0.08° (about 3 mrad). If we halve the baseline, this contribution to the overall error will double. Fortunately we can use the micromechanical IMU as a stable reference

over which many GPS phase-difference measurements can be averaged. This filtering process allows the receiver noise to be significantly reduced.

Vehicle Flexibility

Knowledge of the spacecraft attitude can only be as good as that of the baseline vectors, for it is these vectors which define the orientation. A trade-off exists in choosing the baseline length. One might expect that greater angular resolution would be obtained with longer baselines, but more integer ambiguity and unaccounted vehicle flexing between the antennas might negate that benefit. An acknowledgment that the baselines are subject to flexures is therefore required for realistic attitude determination. One source of flexure would be thermal stresses experienced by the craft when passing in and out of eclipse. The effect of the flexure can be to change the baseline length, as well as its direction. The latter result may not be easily differentiated from a change in vehicle attitude. On a small satellite, this should not be a problem unless the antennas are mounted on flexible appendages. If the configuration of the satellite makes this a particularly severe problem, the inclusion of a micromechanical attitude reference at each of the antennas could measure the relative rotational dynamics, and calibrate thermal bending as the satellite moves in and out of eclipse.

Line Bias

The line bias over a baseline results from the difference in electrical path lengths from each antenna to the receiver. This "bias" may not be constant, however, since the effective path lengths may change with thermal variations, similar to the baseline variations previously mentioned. Minimizing the error can be accomplished by keeping the path lengths as short and as symmetrical as possible, and by configuring the vehicle such that the cabling is subject to small temperature gradients. Fortunately, the phenomenon affecting the path delay to one antenna should be similar to that affecting the path delay to another; the changes in path delay are therefore mitigated somewhat. The line bias can be eliminated by utilizing the between-satellite double difference observable.

Multipath

Multipath is the undesired reflection of the incoming GPS signal from the antenna's surroundings. The antenna may receive both the direct and reflected signal, or could conceivably receive only the reflected one. Since the reflected signal travels a different path, its phase is shifted from the direct signal. The signal reflected then appears as an additive bias to the primary transmission. For small baselines on a spacecraft, the multipath error experienced by each antenna could have a common component.

Multipath error can be diminished by various techniques. A low-multipath environment is most desirable; antennas, therefore, should obviously not be placed near multipath sources. If the satellite has a simple shape without appendages (such as a sphere or cylinder), then small patch antennas on the surface will not experience multipath problems. A suitable choice of coating on the mounting surface can also reduce reflectivity. The antenna gain pattern can be appropriately shaped to mask out signals entering from directions of suspected multipath sources. Calibration of the repeatable part of multipath is another alternative. The only source of multipath in a spacecraft application is the spacecraft itself. A wave front from a given direction will reflect off vehicle surfaces in a repeatable way, provided the reflectivity does not change and there are no moving parts in the viewing antenna. The multipath can then, in theory, be calibrated out as function of incidence direction, and whatever error remains is characterized as receiver noise. It may be possible to perform this calibration in orbit using the micromechanical IMU as an alternate attitude reference. Using the IMU as an independent attitude reference, severe multipath conditions can easily be detected and edited out of the blended GPS/INS solution.

Antenna Phase Center Variations

Another error source is the asymmetry of the antenna gain pattern. This asymmetry can cause significant differences in phase based on the signal's incidence direction, resulting in an apparent migration of the phase center. Two antennas of the same make can have similar gain patterns, and the error can be small if the two are similarly oriented with respect to the incoming signal. But if the signal arrives from two different antenna-relative directions caused by, say, the antennas being canted away from each other, the error in differencing the two phases can be significant. This area should receive special attention on a small satellite since short baselines increase the sensitivity. Again, the micromechanical IMU can aid an inflight calibration process. By changing satellite attitude over a short timeframe such that gyro drift is not a concern, the GPS line-of-sight with respect to the satellite will change. Using the IMU as a reference, this phenomena can be measured and calibrated, within the limitations of receiver noise, etc. Both phase center movement and multipath effects would be contained in the observable.

Dilution of Precision

The attitude solution is also compromised by the GPS satellite geometry, itself. For translational state solutions, the familiar Geometric Dilution of Precision (GDOP) figure-of-merit is used to assess the impact of satellite geometry on position and time determination. A smaller GDOP corresponds to a greater volume of the poly-hedron formed by connecting the vertices of the unit vectors from the user to each GPS satellite. For attitude determination, the best resolution occurs when the line of sight is perpendicular to the baseline vector. An alternate figure of merit, Attitude Dilution of Precision (ADOP), is therefore required to rate the effect of satellite geometry on the attitude solution. Proper antenna placement can help minimize this concern.

Integer Ambiguity and Cycle Slip

Obviously, the GPS receiver cannot distinguish one cycle of the carrier signal from any other. Thus, for baselines longer than half of a wavelength (L1 carrier wavelength = 19 cm), an ambiguity in the integer number of cycles between the two antennas exists. Translating the phase difference, Δf, into the range difference, Δr, and ultimately solving for the relative positions of the antennas requires resolution of this integer ambiguity. For small satellites with short baselines this should not be a problem. Furthermore, the micromechanical IMU can easily aid in the detection of any cycle slip occurrences.

Antenna

At a minimum, 3 antennas are necessary to solve for the vehicle attitude. A single baseline between two antennas observing phase difference measurements from a single GPS satellite can determine one component of vehicle attitude. A second GPS satellite observed over the same baseline, can completely resolve the direction of the baseline (i.e., two components of attitude). Ideally, the two GPS satellites should be nearly orthogonal to obtain maximum precision. However, rotation about the baseline is not observable, thus requiring a second baseline (3rd antenna) to completely resolve three-axis attitude.

8. Summary

This paper presents a compilation of technologies, hardware, and control methodology that may contribute to future micro-nanoclass space applications. Draper's development of commercial and tactical-grade micromechanical inertial sensors spearheads this effort. The work outlined here addresses a fairly precise pointing challenge and proposes a solution using micromechanical gyros, integrated with GPS, and active control.

Among the emerging opportunities for micromechanical inertial sensor insertion is the application to closed loop, active control. These instruments combined with a GPS receiver, provide complete inertial navigation for attitude control and precision pointing applications.

Recent advances in silicon microfabrication technology have led to the development of low cost, tactical performance grade, micromechanical inertial sensors. The inherent small size, low weight, and low cost of these sensors permit on-board insertion of gyroscopes and accelerometers for inertial instrument application previously impractical because of size and cost considerations.

The micromechanical sensors under development at Draper Laboratory are of extremely small size, with each sensor occupying an area of less than 4 square millimeter. Mating electronics consistent with the applications for these small sensors are being developed for minimum size, power, and cost. Miniaturized GPS microstrip antennas, fabricated on high-dielectric ($k \geq 80$) constant ceramic substrates, are being developed at Draper. These permit significant reduction in size compared to antennas fabricated on common low-dielectric substrates.

REFERENCES

1. *Performance Analysis of a GPS Interferometric Attitude Determination System for a Gravity Gradient Stabilized Spacecraft*, J. Stoll, Draper Laboratory, MIT Masters Thesis, CSDL-T-1253.

2. *A Micromachined Comb Drive Tuning Fork Rate Gyroscope*, M. Weinberg, J. Bernstein, S. Cho, A. T. King, A. Kourepenis, and P. Maciel, Draper Laboratory, Proceedings of the ION 49th Annual Meeting, "Future Global Navigation and Guidance", June 21-23 1993.

3 *High Dynamic-Range Microdynamic Accelerometer Technology*, J. Sitomer and J. Connelly , Draper Laboratory, Proceedings of the ION 49th Annual Meeting, "Future Global Navigation and Guidance", June 21-23 1993.

4. *Micromechanical Sensors for Kinetic Energy Projectile Test and Evaluation*, A. Hooper, U. S. Army Yuma Proving Grounds, and R. Hopkins and P. Greiff, Draper Laboratory, U. S. Army TECOM Test Technology Symposium VII, March 29-31, 1994.

5. *Precision Strike Concepts Exploiting Relative GPS Techniques,* G. Schmidt and R. Setterlund, Draper Laboratory, NAVIGATION, Journal of the ION, Vol. 41, No. 1, Spring 1994, pp. 99-112.

6. *A Micromechanical Comb Drive Tuning Fork Gyroscope for Commercial Applications,* M. Weinberg, J. Bernstein, S. Cho, A. T. King, A. Kourepenis, P. Ward, and J. Sohn, Draper Laboratory, Sensors Expo, September 20-22, 1994.

7. *A Micromechanical INS/GPS System for Guided Projectiles*, D. Gustafson, R. Hopkins, N. Barbour, J. Dowdle, 51st ION Annual Meeting, June 1995.

8. *Micro-electromechanical Instrument and Systems Development at Draper Laboratory*, J. Connelly, J. Gilmore, M. Weinberg, International Conference on Integrated Micro-Nanotechnology for Space Applications, Houston, TX, November 1995.

9. *An Innovative Bonding Technique for Optical Chips Using Solder Bumps That Eliminate Chip Positioning Adjustments*, T. Hayashi, IEEE Transactions on Components, Hybrids, and Manufacturing Technology, Vol. 15, No. 2, April 1992.

NB/lah

MICRO SCANNING LASER RANGE SENSOR FOR PLANETARY EXPLORATION

71184

*Ichiro Nakatani, *Hirobumi Saito, *Takashi Kubota,
*Takahide Mizuno, *Hiroshi Katoh, *Satoru Nakamura,
*Kenji Kasamura, **Hiroshi Goto

*The Institute of Space and Astronautical Science
3-1-1, Yoshinodai, Sagamihara-shi, Kanagawa-ken 229, JAPAN

**OMRON Corporation, Central R&D Laboratory
45, Wadai, Tsukuba-city, Ibaraki 300-42, JAPAN

Abstract

This paper proposes a new type of scanning laser range sensor for planetary exploration. The proposed sensor has advantages of small size, light weight, and low power consumption with the help of micro electrical mechanical system technology. We are in the process of developing a miniature two dimensional optical sensor which is driven by a piezoelectric actuator. In this paper, we present the mechanism and system concept of a micro scanning laser range sensor.

1. Introduction

As increasingly many missions are being proposed for the moon and planet explorations, autonomous navigation technology for spacecraft in deep space is getting more important than ever. In recent years, many researchers have been extensively studying and developing planetary landers or rovers for unmanned surface exploration of planets[1][2][3][4][5].

In a planetary exploration mission, it is important for a lander to land in a plain and safe place. A planetary rover is also required to travel safely over a long distance in unknown terrain without collision with big stones, rocks etc. Hence, it is necessary for a lander or a rover to recognize the terrain on the planetary surface.

There have been developed various kinds of sensors for this purpose[6][7][8]. However conventional sensors have many problems with weight, size, power consumption, reliability, etc. In this paper, we propose a scanning laser range sensor using micro electrical mechanical system (MEMS) technology. A two dimensional optical scanner is one of the most promising devices to recognize the three dimensional terrain. So we are in the process of developing a miniature two dimensional optical sensor which is driven by a piezoelectric actuator. The scanner system consists of two elements, a resonator and a piezoelectric actuator. The resonator has two typical vibration modes, twisting and bending mode of the torsional spring. Actuating the resonator with piezoelectric element at the resonance frequency leads to the resonant rotating vibration of the resonator with large amplitude. When the piezoelectric actuator is driven simultaneously with both the resonance frequencies, optical beam reflected on the mirror of the vibrating resonator is scanned in two perpendicular directions. The proposed technology will remarkably reduce the weight, size, and power consumption compared with the conventional sensor.

This paper is structured as follows. In Section 2, the concept of recognition sensor is discussed. Then a scanning mechanism is described in Section 3. In Section 4, a method to obtain range image by a micro scanning laser range sensor is explained. Final Section is for discussion, conclusion, and future work of the research.

2. Concept of Recognition Sensor

Several sensors and methods have been proposed to recognize the terrain on the planetary surface; for example, a laser range finder, stereo vision etc. However, conventional laser range finders have problems with weight, size, power consumption, reliability, etc. for the purpose of scanning the terrain area. It is difficult and takes a long time to find correspondent points in stereo vision methods. Hence, we have proposed a new type, of two dimensional optical scanner sensor.

Figure 1 shows the concept model of the proposed recognition sensor, which consists of an optical scanner device, a micro lens, a laser diode, a photo detector, a piezoelectric micro actuator to drive the scanner, driver circuits, and signal processing circuits. Target size, weight, and power consumption are shown in Table 1.

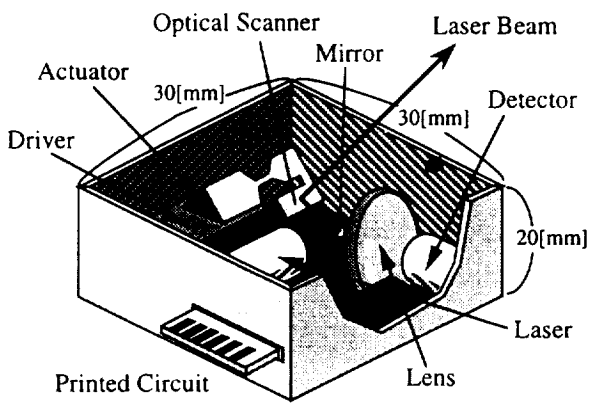


Figure 1. Concept of Recognition Sensor

Table 1. Specification of LRF

Laser Range Finder	size	weight	power
Conventional LRF (Rotating Mirrors)	20[cm] (L) 20[cm] (W) 20[cm] (H)	15[kg]	100[W]
Advanced LRF (Polygon Mirrors)	15[cm] (L) 15[cm] (W) 15[cm] (H)	3[kg]	60[W]
Micro LRF (Piezo-Mechanism)	target 3.0[cm] (L) 3.0[cm] (W) 2.0[cm] (H)	target 0.5[kg]	target 3[W]

3. Scanning Mechanism

3.1 Two Dimensional Optical Scanner

A miniature two dimensional optical scanner is shown in Figure 2. The developed optical scanner[9] consists of two elements, a resonator and a piezoelectric actuator. Here the center of gravity is shifted from each rotational axis P and Q. So the resonator has two typical vibration modes, twisting and bending of the torsional spring as shown in Figure 3. Actuating the resonator with a piezoelectric actuator at each resonance frequency leads to the resonant rotating vibration of the resonator with a large amplitude. The resonance frequency f is described by the following equation.

$$f = \frac{1}{2\pi}\sqrt{\frac{K}{\ell}} \qquad (1),$$

where K and ℓ denote stiffness of torsional spring and rotational inertia moment of resonator respectively. As a result, optical beam reflected on the mirror of the vibrating resonator can be scanned in the two directions of θ_T and θ_B.

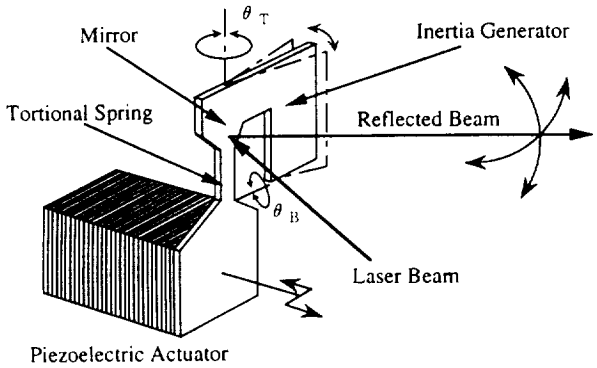


Figure 2. Structure of Scanner Sensor

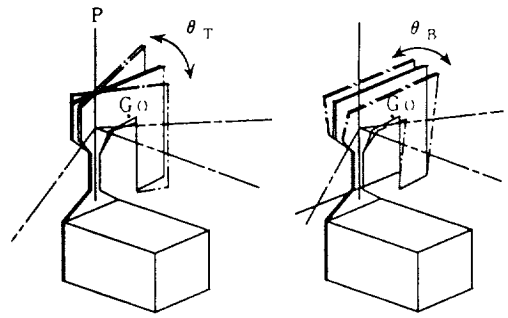


Figure 3. Resonance Vibration Modes

3.2 Piezoelectric Micro Actuator

A multilayered piezoelectric actuator is often used because it is easy to realize miniaturization and high speed response of the actuator. However the multilayered type requires high voltage over 100 volts to obtain several microns amplitude of the resonator. It is needed to avoid generating heat at the miniature devices. The new actuator[10], which is called moonie, is developed as shown in Figure 4. The actuator consists of a multilayered piezoelectric actuator and metal plates fixed on the top of the actuator. The plates has shallow cavity to transform and magnify the radial strain of the actuator into the axial strain.

When the piezoelectric actuator generates the strain in X direction as the axial strain by supplying electrical field, the contraction strain in Y direction is produced. The plate is deformed by the contraction strain as shown in Figure 4(b). So a magnified strain in X direction can be obtained.

Applying the moonie actuator in the two dimensional optical scanner, low voltage driving, less than 10 volts can be realized. It can reduce generate heat at the actuator and driving circuits.

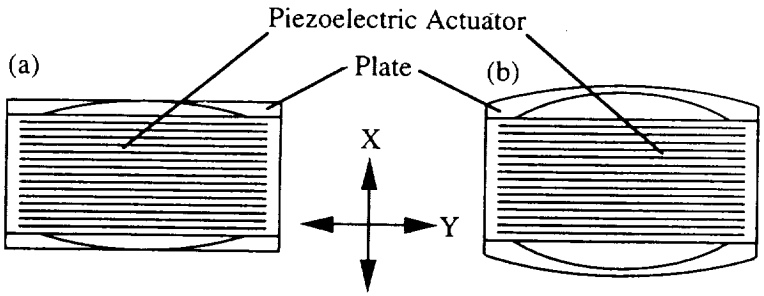


Figure 4. Structure of Moonie Actuator

3.3 Scanning Angle Detection

It is necessary to detect the scanning beam angle for making three dimensional terrain map of planetary surface. In the developed resonance type sensor[11], the phase difference between the scanning orientation and driving voltage wave form is 90[deg] which is constant regardless of the oscillation frequency and amplitude. So it is possible to detect the scanning angle by monitoring the driving voltage to the actuator. The scanning angle is calculated by the following equations:

$$\theta_T(t) = \pm\theta_0\sqrt{1-\left(\frac{V_T(t)}{V_{0T}}-1\right)^2} \qquad (2),$$

$$\theta_B(t) = \pm\theta_0\sqrt{1-\left(\frac{V_B(t)}{V_{0B}}-1\right)^2} \qquad (3),$$

where V0 is maximum amplitude of the driving voltage wave form to the actuator and θ_0 is maximum scanning angle of the scanner.

3.4 Performance of the scanner

The performance of the scanner angle is shown in Figure 5. The developed scanner can scan an optical beam in two directions with approximately 40[deg] at 4[μm] piezoelectric actuator amplitude. In this case, the resonance frequency of the resonator is 121[Hz] for twisting mode and 288[Hz] for bending mode. So the scanning speed of 242[scan/sec] and 576[scan/sec] is obtained if a to-and-fro scanning is applied. The scanning speed is adjustable by changing the resonator shape. The maximum speed is over 2000[line/s] by reducing air viscous resistance against the resonator. Two dimensional scanning pattern is determined by the ration of the resonance frequencies. Figure 7 shows an example of the scanning pattern.

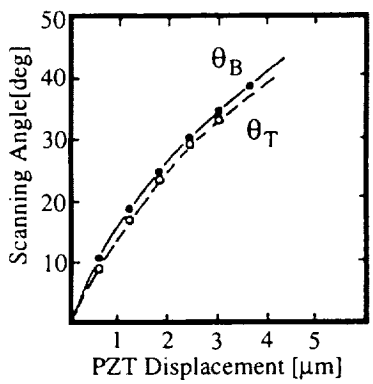


Figure 5. Scanning Angle Data

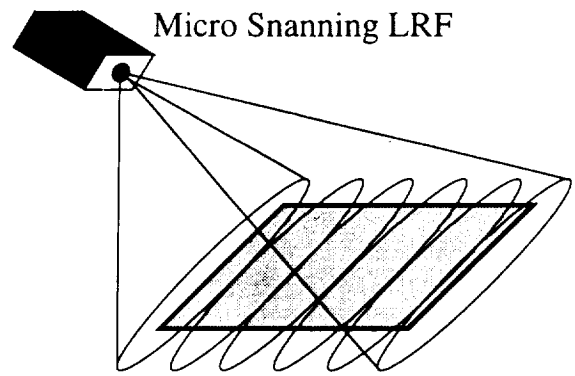


Figure 6. Scanning Pattern

4. Range Image

Laser range sensors for planetary exploration require high accuracy and high frame rate performance. This leads to using phase-comparison system by intensity modulated CW laser. The laser beams are radiated to target objects reflected and returned to the sensor with some phase shift proportional to the target distance. The intensity of the semiconductor laser is modulated at 10.7 [MHz]. The reflected beam from the object is received by an APD photo detector. The signal received by the detector passes into the detection circuit of the range and reflectance. The range is determined by the phase difference of returned signal and the reference signal as shown in Figure 7. In order to obtain sufficient sensitivity and accuracy, nearly 100 waves are integrated for each range data. The data of range and reflectance are converted into digital data by an A/D converter. Optical scanning in two dimensional directions help to obtain three dimensional terrain recognition.

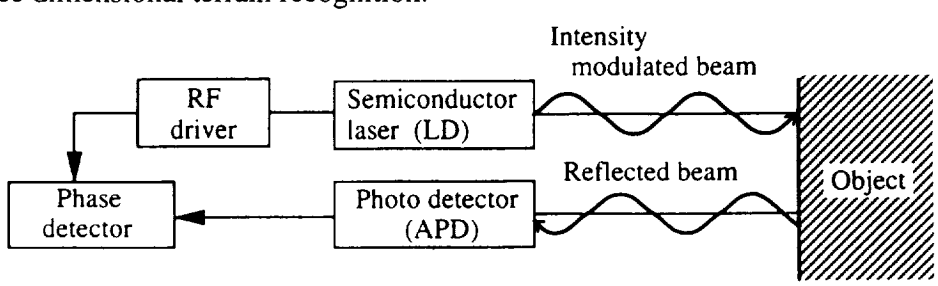


Figure 7 Range Data Measurement

5. Conclusion

This paper presents a scanning laser range sensor for terrain recognition of planetary surface. The concept of micro scanning laser range sensor is proposed with the help of micro electrical mechanical system technology. The proposed micro sensor is expected to solve some problems with weight, size, power consumption. Experimental study is under going. When the proposed sensor is used in space environment, there are some items for further study as follows: range drift due to temperature change, vibration and shock problems at launch. Those problems are under study.

Reference

[1] C.Champetier, J.Serrano, J.de Lafontaine, "Autonomous and Advanced GNC Techniques for the Interplanetary Rosetta Mission," Proc. of 1st ESA Int. Conf. on Spacecraft Guidance, Navigation and Control Systems, pp.559-570, 1991.

[2] J.de Lafontaine, "Autonomous Spacecraft Navigation and Control for Comet Landing," Journal of Guidance, Control and Dynamics, Vol.15, No.3, pp.567-576, 1992.

[3] W.T.Fowler, E.Pinon III, et.al. "Mini spacecraft for a Multiple-Asteroid Rendezvous Mission," AIAA Aerospace Design Conf., AIAA 93-0998, 1993.

[4] J.Kawaguchi, Y.Morita, T.Hashimoto, T.Kubota, H.Yamakawa, and H.Saito, "Nereus Sample Return Mission," Proc. of the 45th Congress of International Astronautical Federation, IAF-94-Q.5.354, 1994.

[5] I.Nakatani, T.Kubota, T.Yoshimitsu, "Path Planning for Planetary Rover using Extended Elevation Map," Proc. of Int. Symposium on Artificial Intelligence, Robotics and Automation in Space, pp.87-90, 1994.

[6] J.R.Kerr et al., "Real Time Imaging Range finder for Autonomous Land Vehicle," SPIE Vo.1007 Mobile Robots III, pp.349-356, 1988.

[7] I.S.Kweon et al., "Experimental Characterization of the Perception Laser Range finder," Technical Report CMU-RI-TR-91-1, 1991.

[8] Y.Wakabayashi, M.Honda, T.Adachi, T.Iijima, "Small Image Laser Range Finder for Planetary Rover," Proc. of Int. Symposium on Artificial Intelligence, Robotics and Automation in Space, pp.201-204,1994.

[9] H.Goto, K.Imanaka, "Super Compact Dual Axis Optical Scanning Unit Applying a Torsional Spring Resonator Driven by a Piezoelectric Actuator," Proc. of SPIE, Vol.1544, pp.272-281, 1991.

[10] H.Goto, H.Kamoda, M.Yoneda, K.Imanaka, "Miniature 2-Dimensional Optical Scanner Utilizing a multilayered Piezoelectric Actuator with New Displacement Expansion Mechanism," Proc. of Actuator'92, pp.68-70, 1992.

[11] H.Kamoda, H.Goto, K.Imanaka, "Two Dimensional Optical Sensor Using a Dual Axis Miniature Optical Scanner," Proc. of SPIE, Vol.1751, pp.280-288, 1992.

[12] H.Goto, H.Ohta, K.Imanaka, T.Yada, "Scanning Optical Sensor for Micro Robot," Proc. of IARP Workshop on Micro Machine Technology and Systems, pp.1-6, 1993.

Nanosatellite Power System Considerations

M. Robyn, L. Thaller, D. Scott

INTRODUCTION

The capability to build complex electronic functions into compact packages is opening the path to building miniature satellites in the order of 1 kg mass, 10 cm across, packed with the computing machines, motion controllers, measurement sensors, and communications hardware necessary for operation.[1] Power generation will be from short strings of silicon or gallium arsenide-based solar photovoltaic cells with the array power maximized by a peak power tracker (PPT).[2] Energy storage will utilize a low voltage battery, with nickel cadmium or lithium ion cells as the most likely selections for rechargeables and lithium (MnO2-Li) primary batteries for one shot short missions.

Based on a spacecraft requirement of 2-W orbit average power for a low Earth orbit (LEO) application with a 60/30 minute sun/dark ratio, the battery would need to deliver 2 W for 30 minutes, and to achieve satellite energy balance would need to be teamed with a solar array able to produce 3 to 4 W.

The spacecraft primary power bus (PBUS) which is the combined output of SA and battery, could be sized to directly power a transmitter for a communications intensive satellite and still be high enough in voltage for efficient conversion to the levels needed by the logic subsystems. These satellites would also use the standard techniques to save power: (1) dynamic adjustment of processing speeds and (2) powering down sections not in use.

GENERAL POWER SYSTEM CONSIDERATIONS

A typical power system is composed of the solar array for producing power from the sun, the battery for storing energy for use during the dark portions of the orbit, the circuitry for controlling and limiting the charge and discharge of the batteries, and power filters, converters, and switches that distribute the raw or conditioned power to the spacecraft subsystem loads.

Typically, one-third of the total weight of the spacecraft is taken by the power subsystem, with the batteries weighing one-third of the power system. The batteries not only supply energy during the dark portions of the orbit but provide the reserve energy to maintain the bus voltage during peak loads.

Selection of the cell type is dictated by the number of cycles and depth of discharge required over the mission life, operating temperature range, and average and peak discharge levels.

Sophisticated charge control can be done efficiently to implement protocols such as constant potential current-limited charging or by trickle charging after reaching a predetermined battery voltage.

ENERGY STORAGE

Nanosatellites would consume only 1 or 2 W of power, compared to medium to large contemporary satellites which draw 100 to several kilowatts of power. For smaller spacecraft requiring less than 200W, nickel cadmium remains a popular choice for reasons which include volumetric energy density, cost, and availability.

Now undergoing extensive evaluation for space applications, Lithium Ion or Li+ cells look to be promising alternates to NiCds with twice the energy density and 3 times the nominal terminal voltage per cell.

Li+ cell chemistry is based on intercalated lithium ions. Although the use of intercalated lithium reduces the energy density relative to the use of pure lithium foils, the cycle life is much greater and the voltage levels are almost as high. Many possible cathode materials are still being investigated in the search for a flatter discharge curve. Li+ cells must be protected from over voltage during charging and this requires each cell to have a voltage clamp which acts as a constant voltage current shunt around the cell. Given the low voltage bus requiring a few cells, the over voltage circuitry is modest overhead for the gains realized. There still remains important work to determine cycle life and optimum charge techniques before using Li+ cells for space use.

The battery type selected for an application can be based on a number of factors. Although cycle life is the most usual factor, other factors include volumetric energy density, and gravimetric energy density.

The Nanosat designer also has a choice between rechargeable and primary batteries. Lithium primary batteries have been flown on Shuttle Payloads and have an energy density tenfold greater than NiCds. They also can deliver significant number of 1C current pulses, generate no gas during discharge, are usable from -20C to +60C and have less than 0.5% self discharge per yr. for missions where long storage time is needed such as a NanoSatellite waiting for release from the mother ship to perform an Observer mission.

Another emerging cell type is based on thin-film technology.[3] Thin, low capacity devices based on successively plating layers of conductive base, cathode, electrolyte, and anode have already demonstrated encouraging cycle life performance.. Unfortunately, the capacities available from these cells are too low for consideration in NanoSatellite applications

For NanoSat applications, it is felt that a more traditional power bus architecture with centralized power source and charge control functions for the entire spacecraft will be more weight effective.

Energy Density

The energy density of a single cell is usually determined by fully discharging it to an appropriate cutoff voltage over a 5-hour period. The energy density is a function of the cell capacity. Large nickel cadmium cells are usually assigned the number of 40 Whr/kg in cell sizes above 10 Ah. As cell sizes become smaller, the energy density is reduced as well as an increasing percentage of the cell consists of case, feed thrus, and other non-energy-producing components.

The energy density of a battery is reduced by 20 to 40% compared to the energy density of the cells because of the structures, wiring, and controls associated with the completed battery. These numbers are further reduced as the batteries are typically cycled to only 20 to 40% of their full capacity. The depth of discharge to which a battery is cycled is set by a combination of factors including temperature limitations, current strain rate limitations, life cycle requirements, and minimum end of discharge limitations. As an example, when 40 Wh/kg nickel cadmium cells are used in a 22-cell battery that is cycled to 20% DOD, the battery will have a usable energy density of 6.4 Wh/kg.

Although the cycle life of a complete battery system is viewed to be a strong function of the particular chemistry under consideration, it is often affected by other factors. Cell design codes, recharge

protocols, DOD, battery temperature, etc., can result in factors-of-10 changes in the cycle life of an aerospace battery. These factors are usually under the control of the power system designer and/or satellite operator. Their impact on cycle life is investigated during the course of ground-based life cycle testing programs. As a result of extensive data bases, cycle life vs. depth-of-discharge relationships have been developed for each cell chemistry. This information can help guide the satellite design engineers.

Given the power requirements of a suggested nanosatellite and the cycle life requirements in low Earth orbit, only power systems based on commercially available nickel cadmium cells have a proven predictable flight history

Table 1. Summary of Battery Characteristics in Small Cell Sizes
(More information on these cell types can be found in Ref. 8.)

Cell Chemistry	Cycle Life Rating	Size Availability	Energy Density
Lead Acid	Insufficient	Should Be	Medium*
Lithium Ceramic	Encouraging	Not Likely	Low
Lithium Ion	In Development	Available	High
Lithium Organic	Insufficient	Available	High
Lithium Polymer	Insufficient	Should Be	High
Nickel Cadmium	Encouraging	Available	Medium
Nickel Hydrogen	Encouraging	Not Likely	Medium
Nickel Me Hydride	In Development	Should Be	Medium
Silver Cadmium	Insufficient	Not Likely	Medium
Silver Zinc	Insufficient	Should Be	High
Li MnO2	Primary Cell	Should Be	High+

*Low, < 15 Wh/kg; Med., 15-40 Wh/kg; High, > 40 Wh/kg; High+ >200 Wh/kg

If we project 5 to 10 years into the future, systems based on nickel metal hydride as well as intercalated lithium ion chemistries may be considered. The cell sizes and projected energy densities will be discussed in a later section. The relative energy density compared to nickel cadmium will be 1.5 for nickel metal hydride and 2.5 for lithium ion devices. For these small cell sizes, the energy densities at 100% DOD are approximately 20 Wh/kg, 30 Wh/kg, and 50 Wh/kg, respectively.

POWER SYSTEM OPTIMIZATION FOR NANOSATELLITES

Based on the current state of development of power system related technologies, the following subsystems form the basic blocks for a nanosatellite power system:

- Solar Arrays
- Solar Array Bus (SAB)
- Satellite Primary Bus (PBUS)
- Batteries
- Power Converters
- Battery Charge Circuits
- Satellite Logic Voltages
- Power Switching

Solar Arrays

Present plans are to use body-mounted solar cells. Small deployed arrays are a possibility if more power is required. The solar array, if confined to an area 10 cm in diameter on the top

surface, will be inadequate to produce the required energy to operate the 2-W satellite when the recharge of the battery and the solar cell degradation factors are taken into account. Even with the very highest efficiency solar cells, the array needs to be kept pointing to the sun in order to obtain 2.6 W, which is about 1 W short of required. Dual junction cells operating at 25% are suggested for this application as being available in reasonable quantities. In addition, the packing factor of the cells needs to be as close to unity as possible. The circular outline also presents problems, as most solar cells are rectangular. The solar array arrangement is still under consideration.

Solar Array Bus

The solar array (SA) output supplies the battery charge regulators and the Primary Bus (PBUS) voltage regulator, the outputs of which are combined to form the PBUS voltage, which in turn is the primary power for the spacecraft loads. The array voltage is set so the battery charger and PBUS regulator operate near their highest efficiency when the SA's approach their end of life; that value is typically in the range of 6-8 V above the battery voltage. For an 8-V battery bus, the SA would be configured for an EOL output of 14-16 V, which would require a series string with about 10 high-efficiency, multi junction solar cells.

Maximizing Solar Array Power

Conventional solar array interfaces do not continuously set the array operating point at the "knee" of the voltage-current (V-I) curve where maximum power is produced. As a result, recharging deeply discharged batteries tends to depress the voltage reflected or "seen" by the array, cutting SA efficiency just when it is needed most. By using a peak power tracker (PPT) in series with the array, the maximum SA power can be generated regardless of the array's illumination, environmental conditions, or effects of aging. 2

The PPT technique uses a switchmode converter in series with the array to dynamically adjust the array output impedance to match the load. The PPT works by slowly increasing the series converter pulse width, so as to "load" the array, causing the array voltage to decrease. At the same time, the PPT measures the rate of SA voltage change and when the knee of the curve is passed, the rate of change starts to drop significantly. The PPT control loop detects this change and reverses course slightly to keep the system stable. The cycle is then repeated. Small PPTs have been demonstrated that realize a 10 to 15% power improvement.

Primary Bus (PBUS)

The PBUS is the spacecraft primary voltage bus formed by combining two sources: (1) the battery output and (2) the solar array output converted by the PBUS regulator to slightly above the maximum battery voltage. The PBUS voltage would be set high enough to directly power the primary loads on a communications or an imaging nanosat, e.g., a sensor array or a transmitter in the range of 0.3 W output. PBUS would also offer a voltage which could be efficiently converted up for higher power transmitters and converted down for the low voltage subsystems such as telemetry sensors and TT&C and ACS computers.

One possible PBS configuration would use either lithium ion or nickel cadmium cells in a nominal 7.2-7.5 V battery, resulting in a PBUS output of 9 V during sunlight, dropping to nominal 7 V as SA power drops out during umbra. A back up feature of PBUS is if the battery or charge regulator should fail, the satellite could still operate from the solar array but only

of course during daylight.

Power Converters

Switch mode power converters are pulse-width-modulated down-converters, which are used in the PPT, in the battery charge regulator and in the PBUS regulator. The converters can be built using monolithic ICs and a few discrete parts with a 1-2 W output at 80-85% efficiency. More complex converters running at 600 kHz and using synchronous rectification to reduce switching device losses have 85% efficiency. However, for these low power applications, the small size and low overhead of the IC converters make them the preferred devices.

Batteries

The battery voltage is selected to minimize the number of cells, since the cell energy efficiency drops as cells get smaller, but still have high enough voltage to efficiently power the satellite subassemblies. The two battery types suggested as a result of the review presented in the Energy Storage section of this chapter are as follows: (1) six series-connected nickel cadmium cells and (2) two cells based on the emerging lithium ion chemistries. These batteries have nominal voltages of 7.5 V and 7.2V, respectively.

To supply the required power would require nickel cadmium cell sizes referred to as AA size (500 mAh). Test data for C and D sizes of nickel cadmium cells have shown reliable operation over the 5 to 15,000 charge and discharge cycles needed for 1 to 3 years of operation. Additional analysis and testing would be needed for the AA applications since performance comparable to the bigger cells has yet to be demonstrated for the AA size. Although there is insufficient cycle life data on the lithium manganese dioxide type of lithium ion cells, testing is in progress which holds promise for extended operation at 40-50% DOD. If lithium ion cells prove reliable, a battery consisting of 2 lithium ion cells at 40% DOD would have a weight savings of 30-40 gm, compared to a nanosat built with 6 nickel cadmium cells cycling to 50% DOD.

Battery Charge Regulator

Long-term lithium ion battery testing is needed to determine how to charge lithium ion cells for long cycle life. Since it is well established that present lithium-based cells have no inherent overcharge tolerance, the charger will need to protect each cell from over-voltage by performing precision end of charge (EOC) control. The charge regulator will also be able to adjust charge rates and EOC voltages based on battery temperatures or other environmental conditions. By comparison, the battery charger for nickel cadmium cells, which have inherent overcharge tolerance, can be a single unit for all the cells employing simple battery temperature compensation.

Regardless of battery selection, the charge regulator will be a switch-mode hybrid package, 80+% efficient, and capable of temperature-compensated constant current/constant voltage charging with lithium ion cells or with nickel cadmium cells.

Charger telemetry would include standard voltages, current, and temperature data and might include onboard fuel gauge computation as part of the charge regulator to aid in planning on-orbit operations. A typical value for the round trip energy efficiency for discharging and charging nickel cadmium batteries at the DODs under consideration here is 65-70%.

Logic Bus

The logic system will need regulated power stepped down from the PBUS and will use linear and digital integrated circuits (ICS) that run from 2.7 V to 3.6 V. The development of low

voltage ICs substantially reduces the power requirements over 5 V systems and has the added benefit of lower effective switching noise. Since the complementary metal oxide semiconductor (CMOS) logic used for all the computing elements draws power directly in proportion to the clock rates, power savings can be realized from control firmware which adjusts the clock rates of the computing and control sections to just meet the processing task.

Lower voltage logic is expected to be more susceptible to single event upsets, but less susceptible to single event latchups compared to 5 V logic. Critical RAM data or program code memory upsets could be corrected by error detection and correction circuits (EDACs). EDACs assign check and correction codes to each data word allowing the EDAC to do a background check to correct single bit errors and to detect double bit errors.

Radiation tests of random access memories (RAMs) from select vendors[4] have shown that a single SEU event rarely hits two bits in a single byte. Thus EDAC is an effective approach to repairing memory SEUs in critical requirements and for protecting the satellite should a radiation blitz occur.

Other promising techniques to improve radiation tolerance include making relatively small changes to the basic gate layouts during the chip masking phase. Post wafer fabrication radiation hardening is also a possibility to improve SEU tolerance and total dose performance. However, for LEO orbits, the radiation levels after shielding by the silicon and aluminum satellite structures are well within present silicon-based IC tolerances.

Power Switching

Power switching is done by the power controller (PC), which routes and switches PBUS power or logic power to the various subsystems. Discrete switches within the PC need not be centralized but could be distributed onto the subassemblies, receiving on/off commands over a control bus. The PC would also generate the needed voltage, current, switch status and temperature telemetry. Power switching would employ smart IC switches, devices which integrate the FET switch and driver along with current limit and thermal protection to form a nearly indestructible switch.

SUMMARY

The power system technology requirements and architecture for a NanoSatellite design have been reviewed and a conceptual power system defined. For short mission life, Li primary batteries have high energy density and eliminate the solar array and charger overhead. For long missions, batteries based on the well established nickel cadmium chemistry or the much newer lithium ion chemistry would be used to store energy generated by high efficiency multi junction solar cells. To minimize the number of cells used in the battery as well as maximize the energy density of the cells used in the battery, a very low bus voltage should be used. The increasing availability of analog and digital devices and sensors able to efficiently operate below 5 V has greatly facilitated the practicality of low voltage bus architecture. There appears to be no insurmountable problems precluding building NanoSatellites on a production line which handles miniature parts, such as Laptop PCs, resulting in high quality low cost NanoSatellites.

REFERENCES

1. E. Y. Robinson, *ASIM and Nanosatellite Concepts for Space Systems, Subsystems, and Architectures*, The Aerospace Corp., Report No. ATR-95(8168)-1 (1 Feb 1995).
2. D. J. Caldwell, L. T. Bavaro, and P. J. Carian, "Advanced Space Power System with Optimized Peak Power Tracking," *Proc 26th* 6. *Lithium Batteries, New Materials, Developments and Perspectives*, edited by G. Pistoia, (Elsevier, 1994)
3. *Handbook of Batteries and Fuel Cells*, edited by David Linden, (McGraw-Hill, 1984).
4. R. Koga, W. R. Crain, K. B. Crawford, D. D. Lau, and S. D. Pinkerton, The Aerospace Corp.; B. K. Yi, Fairchild Space Co.; R. Chitty, CTASS; "On the Suitability of Non-Hardened High Density SRAMs for Space Applications," *IEEE Trans. Nuclear Science* (6 Dec 1991).

LOW-MASS INFLATION SYSTEMS FOR INFLATABLE STRUCTURES

Daniel P. Thunnissen, Mark S. Webster, Carl S. Engelbrecht
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

Abstract

The use of inflatable space structures has often been proposed for aerospace and planetary applications. Communication, power generation, and very-long-baseline interferometry are just three potential applications of inflatable technology. The success of inflatable structures depends on the development of a low mass inflation system. This paper describes two design studies performed to develop a low mass inflation system. The first study takes advantage of existing onboard propulsion gases to reduce the overall system mass. The second study assumes that there is no onboard propulsion system. Both studies employ advanced components developed for the Pluto Fast Flyby spacecraft to further reduce mass. The study examined four different types of systems: hydrazine; nitrogen and water; nitrogen; and xenon. This study shows that all of these systems can be built for a small space structure with masses lower than 0.5 kilograms.

Introduction

The purpose of an inflation system is to successfully inflate an inflatable structure in a space environment. This design study performs two analyses each with its own major assumption. The first assumes an existing propulsion system onboard the spacecraft. The second assumes there is none. The former further assumes that the inflation system can run off an existing onboard propulsion system. It was for this reason that the four following inflation systems were studied: hydrazine; nitrogen and water; nitrogen; and xenon. The focus of this design study was on mass considerations while volume and complexity considerations were considered to a lesser extent. Each analysis is discussed in detail.

The inflation system consists of pressurant or pressurants (such as nitrogen and water), system components (such as valves and filters), tanks, and tubing. For this design study, the total of these four represent the total mass of the inflation system. Of the four inflation systems considered, the total mass is heavily dependent on the size of the space structure and the length in time of the mission. Nitrogen and xenon systems are the least massive for short missions and small structures. Hydrazine and nitrogen and water systems are the least massive for long missions and large structures. Hence, the final choice of an inflation system is mission dependent. However, unless mass is a critical design parameter, the nitrogen and xenon systems are recommended for their comparatively simple system design.

Major assumptions in the design study will first be discussed. A brief discussion of the inflation sequence follows, with the equations governing the pressure and volume of the inflatable structure. The four inflation systems will then be explained in detail for the design study that assumes an onboard inflation system followed by a summary of the propulsion system itself. A brief discussion of the differences between the two design studies is then discussed as are other important assumptions made. The mass, volume, tank impact, and tubing analyses follow and remaining open issues are listed. Lastly, recommendations and conclusions end the paper.

Major Assumptions

The major assumption made in the first design study is the existence of a propulsion system onboard the spacecraft. Indeed, it was this assumption that led to the design study of four inflation systems that could be run off the spacecraft's existing propulsion system: hydrazine; nitrogen and water; nitrogen; and xenon. This assumption was made primarily to save tankage mass, but also to significantly reduce the complexity

of the inflation system. A combined system would already have the hardware associated with loading, testing, monitoring, fluid pressure and temperature, etc. as part of the propulsion system. Such a combined system would have this hardware available to both the propulsion and inflation system without a mass penalty to the latter. The second analysis assumes there is no onboard propulsion system. This analysis results in slightly higher overall masses due to the addition of tanks and hardware.

The type of inflatable structure that was used as a baseline for this study is a concentrator such as that shown in Figure 1.

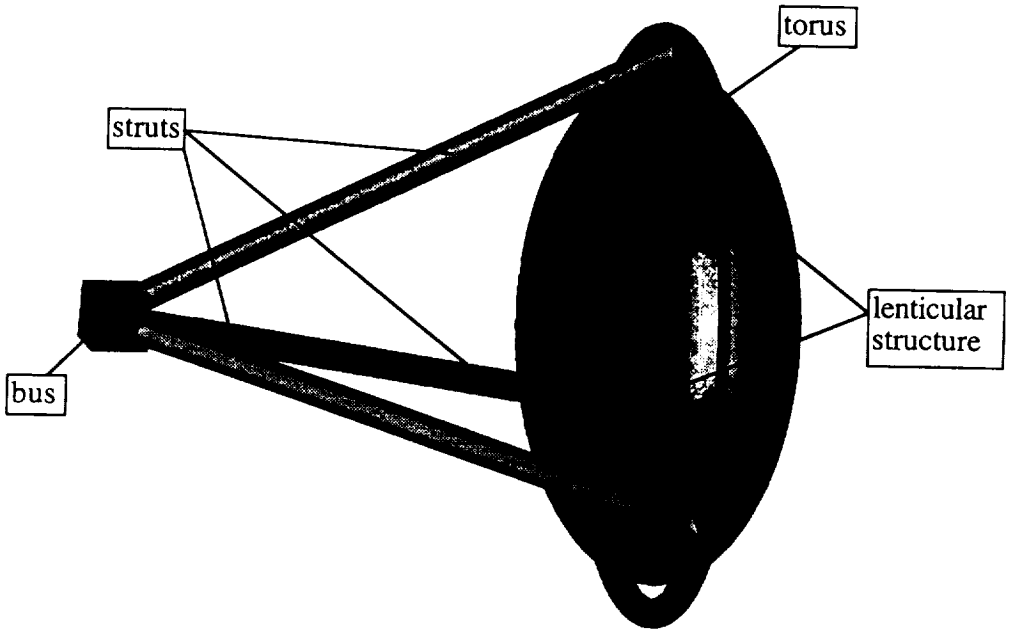


Figure 1. Inflatable Concentrator Used in this Design Study.

This structure could serve as an antenna or a solar concentrator. An inflation system that has been designed for this type of structure should be adaptable to any type of inflatable structure. The basic structure includes the lenticular structure with a transparent front surface and a reflecting back surface; a torus around the outer edge of the lenticular structure; and struts that connect the torus to a focal structure. The focal area, or bus, is where the spacecraft would be located.

The three parts of the concentrator that inflate are the struts, torus, and lenticular structure. The three primary values that were varied in both design studies were the diameter of the lenticular structure (D), the focal length over lenticular structure diameter ratio (f/D), and the length in time of the mission (t). Mission requirements size the structure. Although a concentrator might have a diameter of 25 meters, an f/D ratio of 1, and a 2 year mission time, a very-long-baseline interferometry experiment might have a diameter of 10 meters, an f/D ratio of 4, and a ten year mission time. Figure 1 illustrates a concentrator with a 15 meter diameter and an f/D of 1.

The assumption that the total mass of the inflation system does not include the mass of cables, electronics, and structure is significant. Furthermore, the total mass does not include the interface hardware.

Inflation Sequence

The actual inflation sequence for this type of structure is as follows:

1. Release a small amount of inflatant into struts, torus, and lenticular structure to start the deployment sequence.
2. Just prior to full erection (the amount of time it takes to do this is on the order of minutes), release gas into the torus and struts until the material reaches yield stress to rigidize (in the case of aluminum film) or near yield stress to remove wrinkles (for other types of rigidization methods).

3. Release gas into the lenticular structure to achieve near yield stress in the material to remove wrinkles. It should be noted that the inflation rate in this step has to be fast enough to overcame any initial leaks in the system or structure.
4. Vent the lenticular structure to a predetermined maintenance pressure.
5. Vent the torus and struts to space.
6. Maintain pressure in lenticular structure to compensate for losses in inflatant due to micrometeor holes.

The venting in step 5 removes inflatant from the structure that could leak out at a later time due to micrometeor impacts. Such leaks would create forces and moments that would affect the structure's attitude and control.

Governing Equations

Equations that determine the volume and pressure needed for a concentrator with a paraboloid reflector and a transparent cone that holds the focus apparatus at the tip were used.[1] These equations are modified for a lenticular structure with two mated paraboloids and struts that connect the focal structure to the torus.

Lenticular Structure
The equation for the pressure to achieve near yield pressure (p_l) in the lenticular structure is:

$$p_l = \frac{2S_l t_l}{R} \tag{1}$$

where:

$$R = D \cdot (0.48K + 0.11)$$

S_l = yield stress of lenticular material

t_l = thickness of lenticular material

D = lenticular structure diameter

$$K = 4\left(\frac{f}{D}\right)$$

f = focal length

The pressure needed to retain the shape of the lenticular structure is a function of the structure size:

$$p_m = \frac{2S_m t_l}{R} \tag{2}$$

where:

S_m = maintence stress of lenticular structure

The volume of the lenticular structure is:

$$V_l = \frac{\pi D^3}{16K} \tag{3}$$

The mass (in kg) of inflatant needed to replace gas loss due to micro-meteor holes is:[2]

$$m = \sqrt{MW}\, p_m A \sqrt{\frac{\gamma}{RT}} \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \frac{a}{2} t^2 \tag{4}$$

where:

$$MW = \text{molecular weight of inflatant}$$

$$p_m = \text{maintenance pressure (in Pa)}$$

$$A = \text{projected area (in m}^2) = \frac{\left(\pi + \frac{K}{2} + \frac{5}{6K}\right)D^2}{8}$$

$$\gamma = \text{ratio of specific heats for inflatant}$$

$$R = \text{universal gas constant} = 8314.3 \ \frac{J}{kmol \cdot K}$$

$$T = \text{temperature of inflatant (in K)}$$

$$a = 1.08692462718E-15 \ \frac{1}{s} \left(0.000343 \ \frac{cm^2}{m^2 \cdot yr}\right)$$

$$t = \text{mission time (in s)}$$

Torus
The diameter of the torus is given by the equation:

$$d^3 = \frac{FS \cdot \left(\frac{2S_s t_1}{R}\right) D^4 K}{6E_t \left(\pi + \frac{2}{3}\right) t_t} \tag{5}$$

where:

$$FS = \text{factor of safety (4 for this study)}$$

$$S_s = \text{stress to size torus}$$

$$E_t = \text{torus material modulus}$$

$$t_t = \text{thickness of the torus material}$$

The volume of the torus is:

$$V_t = \frac{\pi^2 d^2 D}{4} \tag{6}$$

The pressure required to yield the torus is given by:

$$p_t = \frac{2S_t t_t}{d} \tag{7}$$

where:

$$S_t = \text{yield stress of torus material}$$

Struts
The strut material and diameter are assumed to be the same as the torus. Hence the pressure required to yield the struts is identical to the pressure required to yield the torus ($p_s = p_t$). The total volume of the struts is given by:

$$V_s = 3 \cdot \frac{\pi D d^2}{8} \sqrt{\frac{(K - \frac{1}{K})^2}{4} + 1} \tag{8}$$

The '3' in Equation (8) signifies the number of struts in the concentrator.

In this design study the torus and struts were assumed to be made out of aluminum polyester film composite (S_t = 2.000E7 Pa; t_t = 6.350E-6 m). The elastic modulus of the torus is 5.860E9 Pa. The lenticular structure was assumed to be an aluminum polyester film composite with a slightly different make-up and properties (S_l = 7.750E7 Pa; t_l = 9.000E-6 m). The maintenance stress of the lenticular structure (S_m) and the stress to size the torus (S_s) are 3.447E5 Pa and 6.895E5 Pa, respectively.

Existing Propulsion System

As stated earlier, the first design study in this paper assumes an existing propulsion system onboard the spacecraft. Four different types of inflation systems were studied: hydrazine; nitrogen and water; nitrogen; and xenon.

Hydrazine System

The hydrazine pressurization system (shown in Figure 2) uses gases created when the liquid monopropellant hydrazine decomposes as an inflatant. Please note from Figure 2 that the latch valve LV1 and filter F1 (to the left of the dashed line) are normal parts of the propulsion system, but are shown here because they also perform a function in the inflation system.

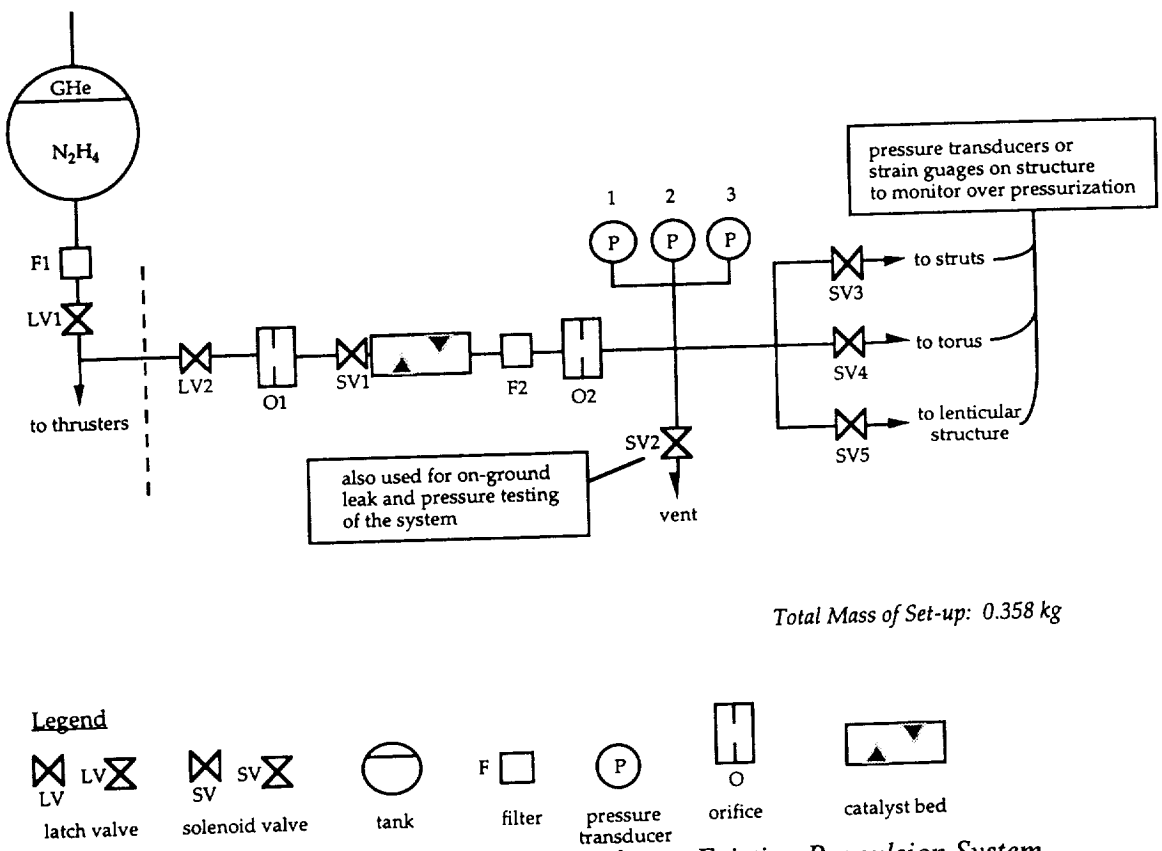


Figure 2. Hydrazine System Set-Up for an Existing Propulsion System.

After launch, when the propulsion system is activated, LV1 is opened to allow the flow of hydrazine out of the propellant tank. When inflatant is required, latch valve LV2 is opened, resulting in pressurized hydrazine filling the inlet of the gas generator (which is solenoid valve SV1 in conjunction with a catalyst bed). Flow control orifice O1 has the function of limiting the flow rate of propellant into the gas generator so that very small amounts of gas can be generated if desired. This is important during the initial inflation and later in the mission when maintenance pressure is required.

Inflatant gas is created when solenoid valve SV1 is opened and liquid hydrazine passes into the catalyst bed. Hydrazine decomposes in the following two-step reaction:[3]

$$3 \cdot N_2H_4 \rightarrow 4 \cdot NH_3 + N_2 + Q$$
$$4 \cdot NH_3 \rightarrow 2 \cdot N_2 + 6 \cdot H_2 - Q \tag{9}$$

While the first reaction is exothermic, the second is endothermic. The amount of ammonia (NH_3) that dissociates depends on, among other things, the length of the catalyst bed. Equation (9) can also be expressed as a function of the amount of ammonia dissociation, X:

$$3 \cdot N_2H_4 \xrightarrow[\text{cat}]{} 4 \cdot (1 - X) \cdot NH_3 + 6 \cdot X \cdot H_2 + (2 \cdot X + 1) \cdot N_2 \tag{10}$$

For an inflation system, it would be preferable to maximize the amount of dissociation possible, both because more moles of gas are created (thus requiring less hydrazine for the same inflation amount) and because the final temperature of the created gas is lower, reducing the task of cooling the gas once it is generated.

Once nitrogen, hydrogen, and possibly ammonia gas are created, they pass through filter F2 and flow control orifice O2 into a manifold, where the pressure of the gas is sensed by redundant pressure transducers (P1, P2, and P3). Note that three transducers are required to allow voting, in case one transducer fails. Different parts of the inflatable structure can be pressurized by opening SV3, SV4, and/or SV5.

Pressure transducers can also be mounted downstream of solenoid valves SV3 through SV5, or strain gauges could be mounted on the inflatable structure to indicate if the pressures are indeed at the correct level. In case of over-inflation, solenoid valve SV2 can be opened. SV2 is also used for relieving the pressure within the inflatable structure once it has already been rigidized.

It should be noted that no service valves are required for the inflation system because SV2 can be used for leak and functional testing of all up-stream components. It is assumed that leak testing of SV3 through SV5 can be performed prior to the final installation of the inflatable structure.

The mass of the various components used in the hydrazine system are summarized in Table 1.

Table 1. Mass Breakdown of Hydrazine System Set-Up.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference/Comments
catalyst bed	1	150	150	discussions with Olin Aerospace
filter	1	50	50	typical mass for capacity and flow rate assumed
heater	0	0	0	
latch valve	1	73	73	Pluto Fast Flyby latch valve[4]
manual valve	0	30	0	
orifice	2	10	20	Viscojet (Lee Company)
pressure transducer	3	10	30	Entran (Fairfield, NJ); no electronics
solenoid valve	5	7	35	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	358 g	

Nitrogen and Water System

The nitrogen and water system shown in Figure 3 assumes a cold-gas nitrogen system already exists which is regulated to produce relatively constant thrusts throughout the mission.

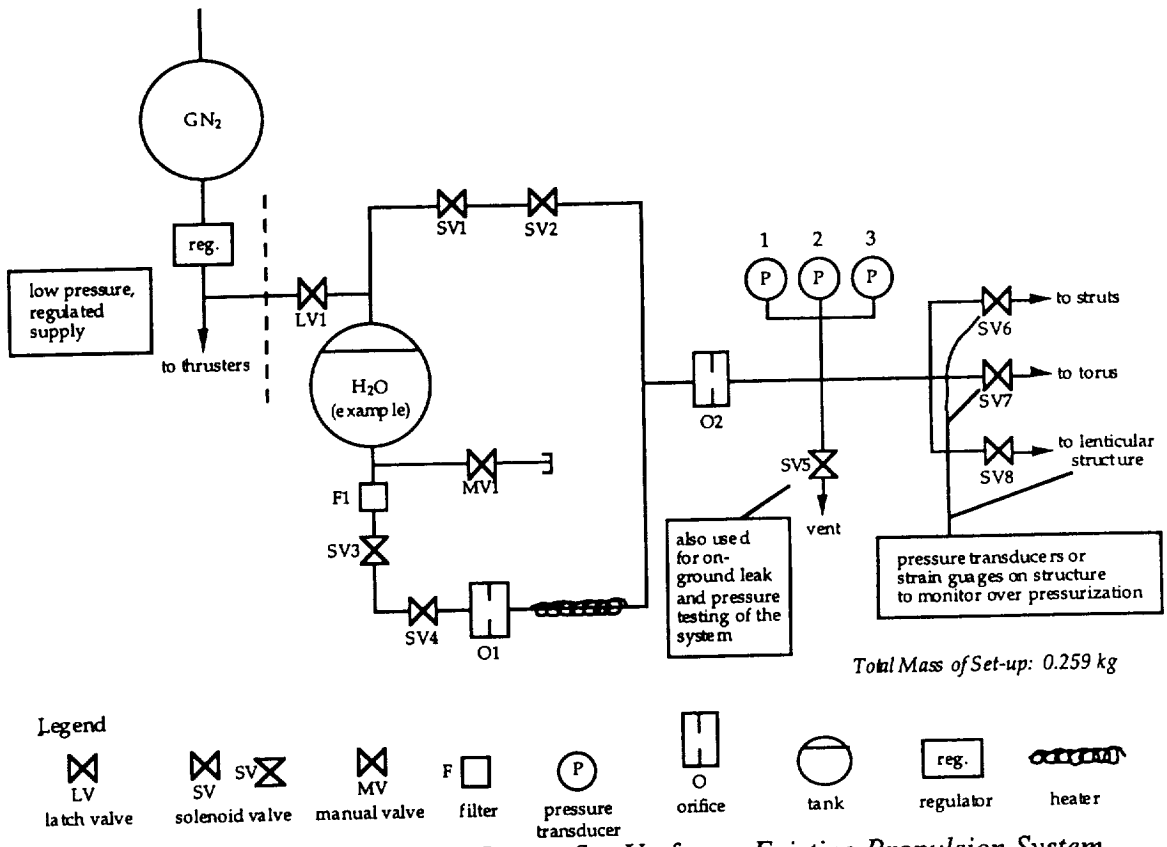


Figure 3. Nitrogen and Water System Set-Up for an Existing Propulsion System.

Figure 3 further assumes a regulated pressure high enough to yield the structure. A separate water tank is incorporated to take advantage of storing the inflatant gas as a liquid and thus minimizing tankage size and mass. The nitrogen and water system maximizes water use in the inflation process. That is to say, water is used as the pressurant up to its vapor pressure. The vapor pressure of water is a function of its temperature. The vapor pressure of water is 2333.14 Pa at 293 K.[5]

Note that although water was used in this study, any liquid with a high vapor pressure and density could be used. Freon would be an ideal choice but was not selected for this design study for environmental reasons.

After launch, latch valve LV1 is opened to pressurize the ullage of the water tank, which is assumed to be a normal bladder-type tank. When pressurizing gas is required, solenoid valves SV3 and SV4 are opened, allowing water to flow through flow orifice O1, after which it vaporizes due to the low pressure downstream of O1. A heater is likely required to keep the water from freezing, due to cooling caused by rapid pressure decrease.

Once the water is vaporized, it passes through flow orifice O2 into a similar manifold described for the hydrazine system. If pressures higher than the vapor pressure of water are required (such as to fully inflate the lenticular structure), high pressure gaseous nitrogen can be introduced directly into the inflatable structure by opening solenoid valves SV1 and SV2.

Manual valve MV1 is incorporated into the system to allow loading and testing of water system plumbing and tank. The mass of the various components used in the nitrogen and water system are summarized in Table 2.

Table 2. Mass Breakdown of Nitrogen and Water System Set-Up.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference
catalyst bed	0	150	0	
filter	1	50	50	typical mass for capacity and flow rate assumed
heater	1	0	0	negligible mass
latch valve	1	73	73	Pluto Fast Flyby latch valve[4]
manual valve	1	30	30	VACCO (developed for the Ballistic Missile Defense Office)
orifice	2	10	20	Viscojet (Lee Company)
pressure transducer	3	10	30	Entran (Fairfield, NJ); no electronics
solenoid valve	8	7	56	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	259 g	

Nitrogen Only System

The nitrogen only system (sketched in Figure 4) works the same way as the nitrogen pressurization part of the nitrogen and water system, and has no water tank or associated hardware.

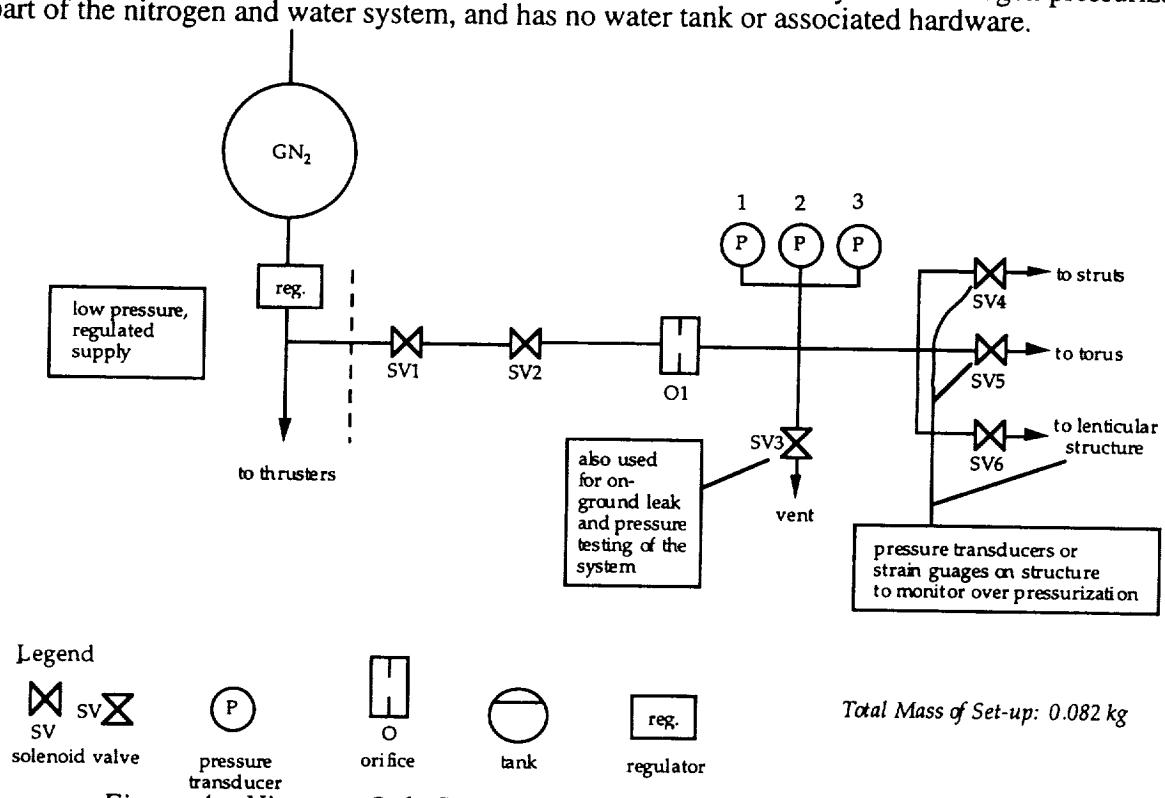


Figure 4. Nitrogen Only System Set-Up for an Existing Propulsion System.

The mass of the various components used in the nitrogen system are summarized in Table 3.

Table 3. Mass Breakdown of the Nitrogen and Xenon System Set-Ups.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference
catalyst bed	0	150	0	
filter	0	50	0	
heater	0	0	0	
latch valve	0	73	0	
manual valve	0	30	0	
orifice	1	10	10	Viscojet (Lee Company)
pressure transducer	3	10	30	Entran Sensors and Electronics (Fairfield, NJ); no electronics
solenoid valve	6	7	42	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	82 g	

Xenon System
The xenon system works the same way as the nitrogen only system with the only difference being the use of xenon propellant instead of nitrogen. A xenon system is appealing for use in conjunction with an electric propulsion system. The mass of the various components used in the xenon system are summarized in Table 3.

Propulsion Tanks

As stated earlier, since the first design study assumes that an inflation system piggy-backs on the propulsion system, an existing propulsion system had to be used. Such an assumption would allow mass impacts of the inflation system on the spacecraft to be calculated provided reference tanks were available.

For this design study the propulsion systems proposed for the Pluto Fast Flyby and NASA SEP (Solar Electric Propulsion) Technology Application Readiness program (NSTAR) were used.[6] The hydrazine system assumed 24.34 kg of liquid hydrazine at initial pressure of 3.447E6 Pa (500 psi). The nitrogen and nitrogen and water system assumed 1.25 kg of gaseous nitrogen at 4.137E7 Pa (6,000 psi). The xenon system uses two tanks each holding 20 kg of supercritical xenon at 1.379E7 Pa (2,000 psi). Note that although the Pluto Fast Flyby and NSTAR programs may change their tanks, it does not affect this study since only a rough baseline was needed for tank impact analysis. This design study assumed T-1000 graphite/epoxy tanks with an aluminum liner for calculations involving nitrogen and xenon. Titanium tanks were used for calculations involving hydrazine or water.

A spreadsheet was used to calculate the resulting size and mass of these tanks. Table 4 summarizes this information.

Table 4. Propellant Tank Information.

System	State	Mass of Propellant (kg)	Pressure of Propellant (Pa)	Initial-to-Final Pressure Ratio	Resulting Mass of Tank (kg)	External Tank Diameter (m)
Hydrazine	Liquid	24.34	3.447E6	3-to-1*	1.149	0.277
Nitrogen	Gas	1.25	4.137E7	120-to-1	0.679	0.178
Xenon (2)	Supercritical	20.00	1.379E7	40-to-1	1.135	0.271

* blowdown ratio for hydrazine

The nitrogen tank would be used in both the nitrogen and water system and the nitrogen only system. The former requires the addition of a water tank whose size depends on the amount of water needed.

No Propulsion System

The second design studied performed assumed no onboard propulsion system. This assumption affects the total mass in two ways. The first is that since there is no onboard propulsion system, there are no onboard tanks. Although tank impact is no longer an issue, entirely new tanks must be determined for each case. The second is the addition of components used to control the inflation process that were part of the propulsion system. Valves and transducers used to monitor the inflatant are part of the propulsion system in the first design study and must be added for the second design study. Both of these changes increase the overall mass of the system.

Hydrazine System
The hydrazine inflation system for a spacecraft with no existing onboard propulsion system is shown in Figure 5.

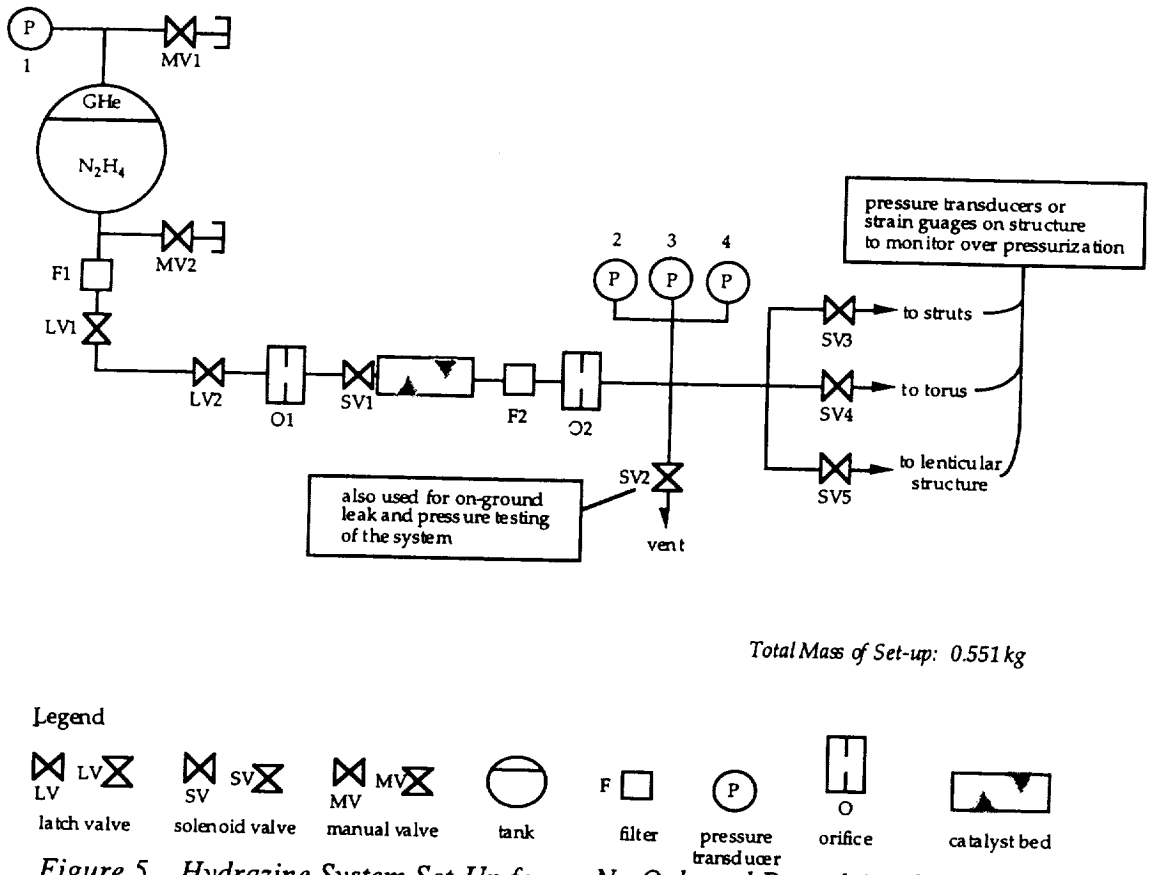


Figure 5. Hydrazine System Set-Up for an No Onboard Propulsion System.

Figure 5 differs somewhat from Figure 2 through the addition of several components that were assumed to be part of the propulsion system in Figure 2. The manual valves MV1 and MV2 are used to load the ullage gas and propellant, respectively. These valves are also used in combination with other manual valves for leakage testing. The pressure transducer P1 is needed for monitoring the amount of propellant remaining in the tank. Filter F1 is needed at the outlet of the tank for contamination control. Such a filter protects down-stream valves from contamination that could lead to leakage. Table 5 summarizes the mass of the various components used in this system.

Table 5. Mass Breakdown of Hydrazine System Set-Up.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference/Comments
catalyst bed	1	150	150	discussions with Olin Aerospace
filter	2	50	100	typical mass for capacity and flow rate assumed
heater	0	0	0	
latch valve	2	73	146	Pluto Fast Flyby latch valve[4]
manual valve	2	30	60	VACCO (developed for the Ballistic Missile Defense Office)
orifice	2	10	20	Viscojet (Lee Company)
pressure transducer	4	10	40	Entran (Fairfield, NJ); no electronics
solenoid valve	5	7	35	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	551 g	

Nitrogen and Water System

The nitrogen and water inflation system for a spacecraft with no existing onboard propulsion system is shown in Figure 6.

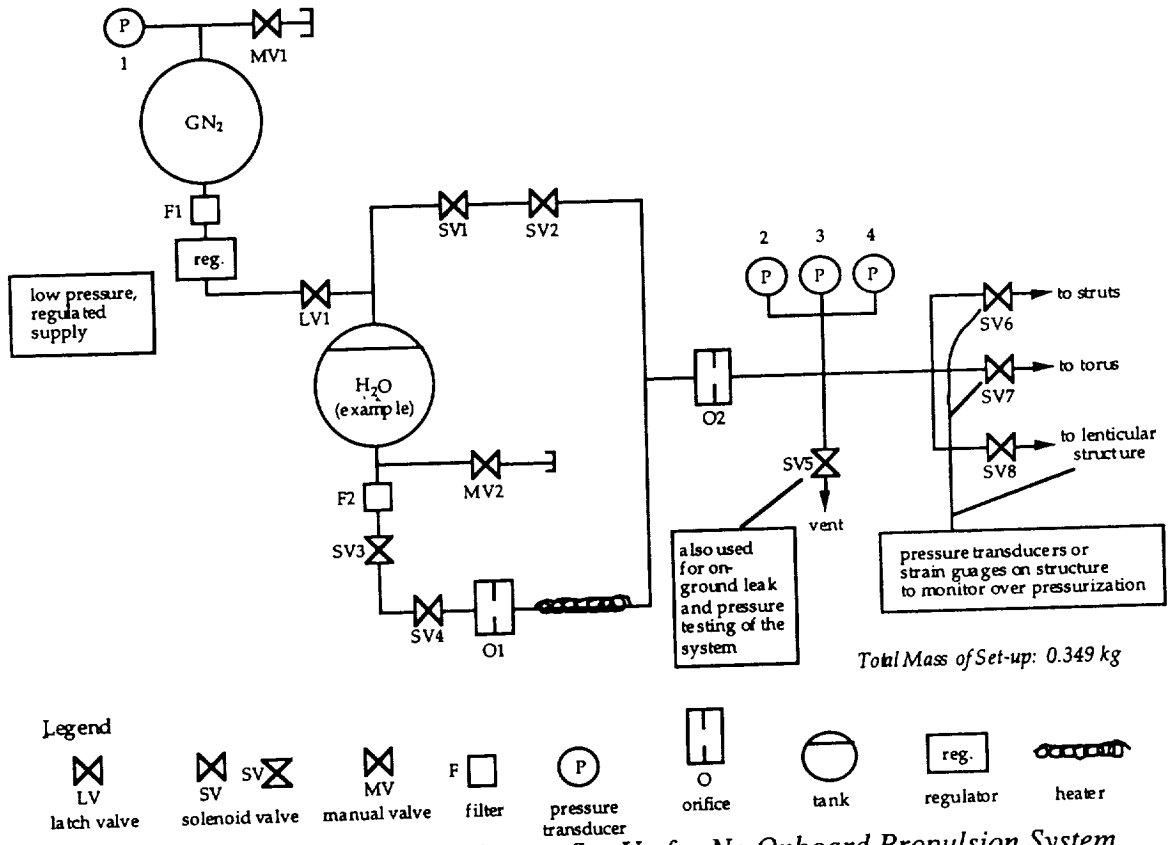


Figure 6. Nitrogen and Water System Set-Up for No Onboard Propulsion System.

The nitrogen and water system in Figure 6 adds several components to the schematic shown earlier in Figure 3. As with the hydrazine system, the pressure transducer P1 is added to monitor the amount of propellant in the tank. Only one additional manual valve, MV1, is needed for loading the propellant. Filter F1 is added for contamination control. Table 6 summarizes the mass of the various components used in this system.

Table 6. Mass Breakdown of Nitrogen and Water System Set-Up.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference
catalyst bed	0	150	0	
filter	2	50	100	typical mass for capacity and flow rate assumed
heater	1	0	0	negligible mass
latch valve	1	73	73	Pluto Fast Flyby latch valve[4]
manual valve	2	30	60	VACCO (developed for the Ballistic Missile Defense Office)
orifice	2	10	20	Viscojet (Lee Company)
pressure transducer	4	10	40	Entran (Fairfield, NJ); no electronics
solenoid valve	8	7	56	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	349 g	

Nitrogen System

The nitrogen inflation system for a spacecraft with no existing onboard propulsion system is shown in Figure 7.

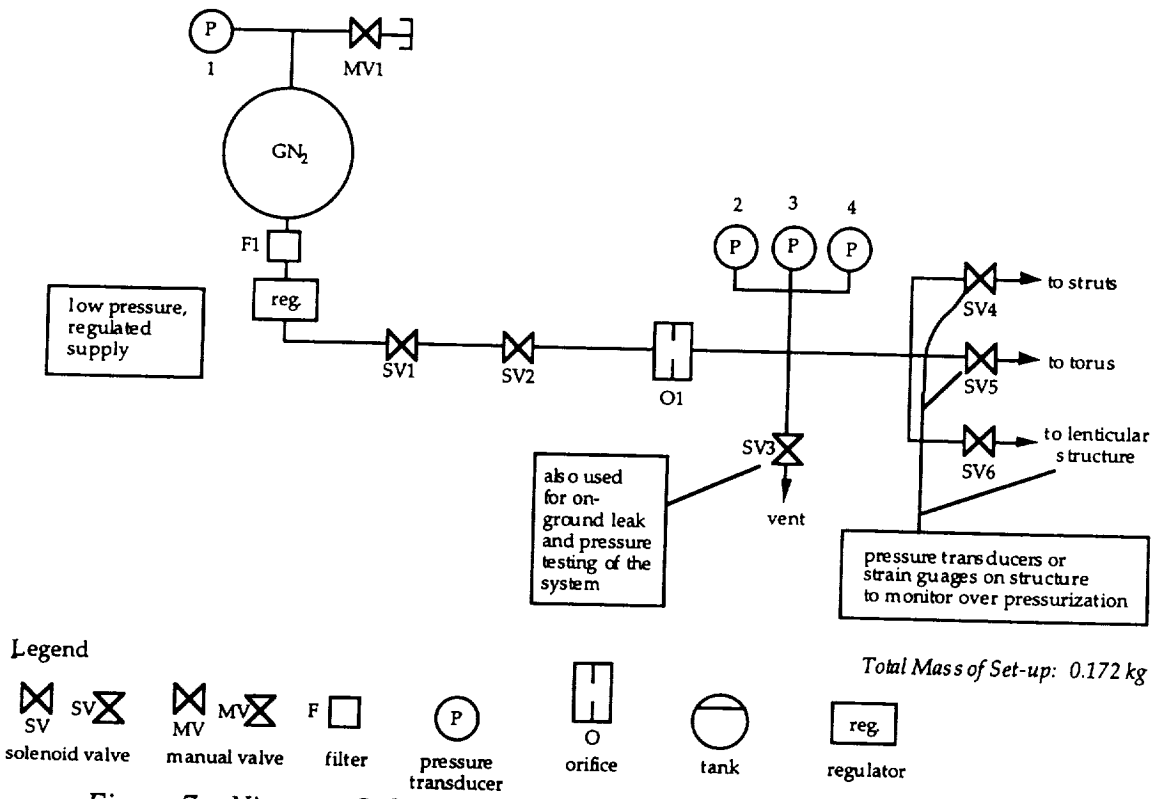


Figure 7. Nitrogen Only System Set-Up for No Onboard Propulsion System.

The nitrogen system in Figure 7 adds three components to the schematic shown earlier in Figure 4. The pressure transducer P1, manual valve MV1, and filter F1 provide the same purpose as described in the previous nitrogen and water system. Table 7 summarizes the mass of the various components used in this system.

Table 7. Mass Breakdown of the Nitrogen and Xenon System Set-Ups.

Item	Quantity (#)	Mass (g)	Total Mass (g)	Reference
catalyst bed	0	150	0	
filter	1	50	50	typical mass for capacity and flow rate assumed
heater	0	0	0	
latch valve	0	73	0	
manual valve	1	30	30	VACCO (developed for the Ballistic Missile Defense Office)
orifice	1	10	10	Viscojet (Lee Company)
pressure transducer	4	10	40	Entran (Fairfield, NJ); no electronics
solenoid valve	6	7	42	Pluto Fast Flyby thruster valve[4]
		TOTAL MASS	172 g	

Xenon System

The xenon system works the same way as the nitrogen only system with the only difference being the use of xenon propellant instead of nitrogen. The mass of the various components used in the xenon system are summarized in Table 7, above.

General Assumptions

In addition to the major assumptions stated earlier, the following general assumptions have been made in this design study. These assumptions apply to both design studies unless otherwise stated.

- An operational temperature of 293 K (20 °C) throughout. The temperature of the spacecraft is kept high enough such that the hydrazine and water do not freeze.
- The dissociation of hydrazine to be 90% (X = 0.9).
- A factor of safety of 4 in the calculation of the torus diameter.
- An existing tank was used with the exception of the water tank (first design study).
- A burst factor of 2 for tank calculations involving water and hydrazine; a burst factor of 1.5 for tank calculations involving nitrogen and xenon.
- A 15% fitting factor in tank calculations.
- A 10% margin in the hydrazine and water tank masses for the bladder.
- Pressure monitoring of the propellant in the tanks is performed by the propulsion system (first design study). Temperature transducers have negligible mass.
- The gaseous nitrogen put out by the cold-gas system is regulated to 3.447E5 Pa (50 psi). This pressure was randomly selected to allow tank sizing. Hence, the initial-to-final pressure ratio is 120-to-1 for nitrogen while 40-to-1 for xenon.
- A 33.3% initial ullage volume for hydrazine and a 6% initial ullage volume for water.
- 20% mass margins for liquids (hydrazine and water) and 50% mass margins for gases (nitrogen and xenon). Such margins account primarily for leakage. They also take into account the scenario in which more pressurant is needed during the initial inflation than originally calculated. A larger margin is required for a gas since it leaks more easily than a liquid.

Mass Analysis

The ideal gas law relates the pressure and volume of a gas to its mass and temperature:

$$PV = mRT \tag{11}$$

where:

P = pressure

V = volume

m = mass

R = gas constant

T = temperature

Equation (11) can be rewritten in terms of the mass:

$$m = \frac{PV}{RT} \tag{12}$$

Equation (12) and deviations on Equation (12) were used throughout this design study in determining the amount of pressurant needed for various stages of the inflation process. For example, the mass of nitrogen required to fully inflate the lenticular structure of a 10 meter power antenna (f/D = 1) would be:

$$m = \frac{(68.719\ \text{Pa})(49.087\ \text{m}^3)}{\left(296.749\ \frac{\text{J}}{\text{kg}\cdot\text{K}}\right)(293\ \text{K})}$$

$$m = 0.0388\ \text{kg}$$

The equation governing the amount of pressurant in kilograms needed for maintenance was stated earlier in Equation (4). Continuing with the example earlier, the amount of nitrogen (to three digits) needed to maintain the pressure (0.306 Pa determined from Equation (2)) for 5 years would be:

$$m = \sqrt{28.016 \tfrac{kg}{kmol}}(0.306\ Pa)(66.874\ m^2)\sqrt{\frac{(1.4)}{(8314.3\ \frac{J}{kmol \cdot K})(293\ K)}}\left(\frac{2}{(1.4)+1}\right)^{\frac{1.4+1}{2(1.4-1)}} \cdot \frac{a}{2} \cdot t^2$$

$$m = 0.642\ kg$$

Where a is 1.087E-15 s^{-1} (defined below Equation (4)) and t is 1.578E8 s (5 years). It is apparent from Equation (4) and the calculation above that a significant portion of the total mass for long missions is replacement gas.

Volume Analysis

After summing the mass for the inflation steps and maintenance, the resulting volume (V) is determined for liquids (such as hydrazine and water) by rewriting the definition of density (ρ):

$$V = \frac{m}{\rho} \tag{13}$$

Note that xenon is supercritical at 1.379E7 Pa and 293 K. That is, xenon at this pressure is neither a liquid nor a gas, but a state in between the two. The density of xenon at this pressure and temperature is assumed to be 2012.7 kg/m^3.[7] The volume of xenon was calculated using this value in Equation (13).

The equation of state is used for gases (such as nitrogen):

$$p = \rho RT \tag{14}$$

With Equation (13) substituted in, Equation (14) reduces to:

$$V = \frac{mRT}{p} \tag{15}$$

Further continuing with the nitrogen example, for a total mass (including margins) of 1.233 kg, the resulting volume at 4.137E7 Pa (6000 psi) would be:

$$V = \frac{(1.233\ kg)(296.749\ \frac{J}{kg \cdot K})(293\ K)}{(4.137E7\ Pa)}$$

$$V = 0.00259\ m^3$$

A final resulting internal tank volume can be obtained by applying the initial ullage volumes defined earlier (120-to-1 in the case of nitrogen).

Tank Impact Analysis

A spreadsheet was used to determine the impact that such an internal volume increase would have on an existing tank. As stated earlier, propellant amounts from the Pluto Fast Flyby and NSTAR design were assumed. The mass and diameter of these tanks were calculated by using the spreadsheet and were summarized in Table 4.

The mass and diameter of tanks were also calculated with the internal volume increase of the inflation pressurant. The mass and diameter impact is quite simply the difference between these two values. For the nitrogen example, the mass and diameter of the tanks with the pressurant included are 1.268 kg and 0.244 m, respectively. Hence, the mass impact would be 0.589 kg (1.268 kg - 0.679 kg = 0.589 kg). The diameter impact would be 0.046 m (0.224 m - 0.178 m = 0.046 m).

It should be noted that if the tanks for the propulsion system are larger (that is, if more propellant is required than documented here), then the mass impact of the inflation system is smaller (a smaller percent change in the tank size required). Also, it is typical for flight projects to select tanks which are "off-the-shelf" to save money. This means that often tanks that are too large are selected. If the mass total of the pressurant required for the inflation system is small enough that the gas can be loaded into the selected tank without affect, then no tankage mass impact would result.

Tubing Impact Analysis

A rough estimate of the mass of tubing that would be required for such an inflation system was also carried out. The design parameters of the 0.01 inch tubing are summarized in Table 8:

Table 8. Tubing Design Parameters.

Material	Stainless Steel
Length	2 m
Outer Diameter	0.003175 m (0.125 in)
Inner Diameter	0.002667 m (0.105 in)

The volume of the tubing is calculated to be:

$$V = \pi\left(\frac{d_o}{2}\right)^2 l - \pi\left(\frac{d_i}{2}\right)^2 l$$

$$= \pi\left(\frac{0.003175\ \text{m}}{2}\right)^2 (2\ \text{m}) - \pi\left(\frac{0.002667\ \text{m}}{2}\right)^2 (2\ \text{m})$$

$$V = 4.66170\text{E-}6\ \text{m}^3$$

Recalling the density of stainless steel to be 8000 kg/m^3, the mass of this tubing can be calculated to be:

$$m = \rho V$$

$$= \left(8000\ \tfrac{\text{kg}}{\text{m}^3}\right)\left(4.66170\text{E} - 6\ \text{m}^3\right)$$

$$m = 0.03729\ \text{kg}$$

Mass Totals

Existing Propulsion System

The sum of the masses of the set-up components, the total pressurant required, the tank impact, and the tubing equals the total mass of the inflation system. This value represents the amount of mass that would need to be added to an existing propulsion system onboard a spacecraft. This value can be obtained for any system or configuration using the spreadsheet developed for this design study. Recall that this mass total does not include electronics, cables, and structure. Figure 8 plots the total mass as a function of lenticular structure diameter for a 5 year mission with a structure having an f/D ratio of 1.

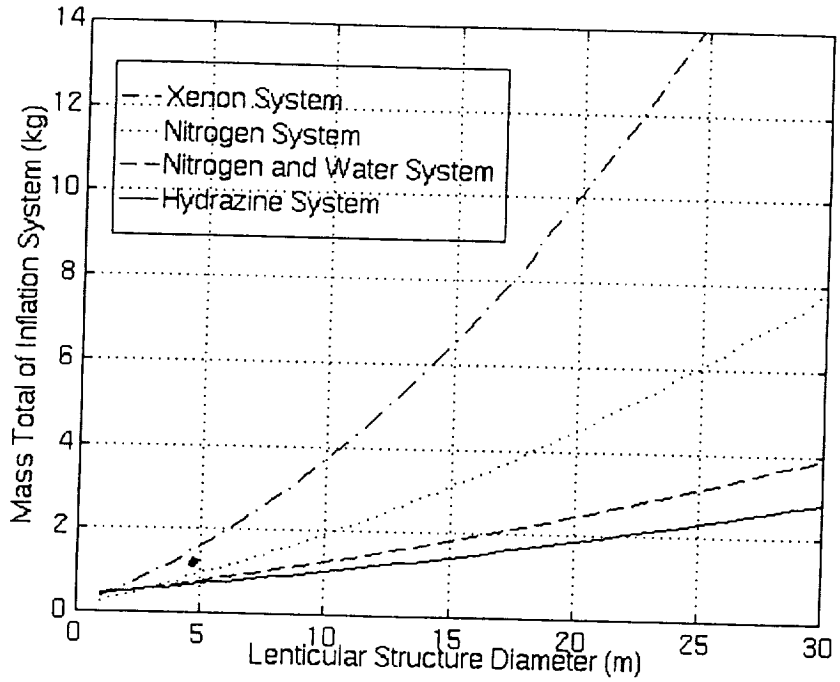


Figure 8. Mass Total for an Existing Propulsion System as a Function of Lenticular Structure Diameter (f/D = 1 and t = 5 years).

From Figure 8 it is apparent that the total mass of the inflation system increases dramatically with lenticular structure diameter. This is especially pronounced for the xenon and nitrogen systems. While these systems are the least massive for small diameters, they are the most massive for large diameters. The difference in mass totals between systems for small structures can be attributed the higher overall component mass for the hydrazine and nitrogen and water system. This difference is not noticeable for large structure since the mass totals of the inflatant and tanks, on the order of kilograms, overshadow the mass total of the components, on the order of grams. The hydrazine and nitrogen and water systems display a competitive mass for all sizes. Figure 9 plots the total mass as a function of time of mission for a structure having a 25 meter lenticular structure diameter and an f/D ratio of 1.

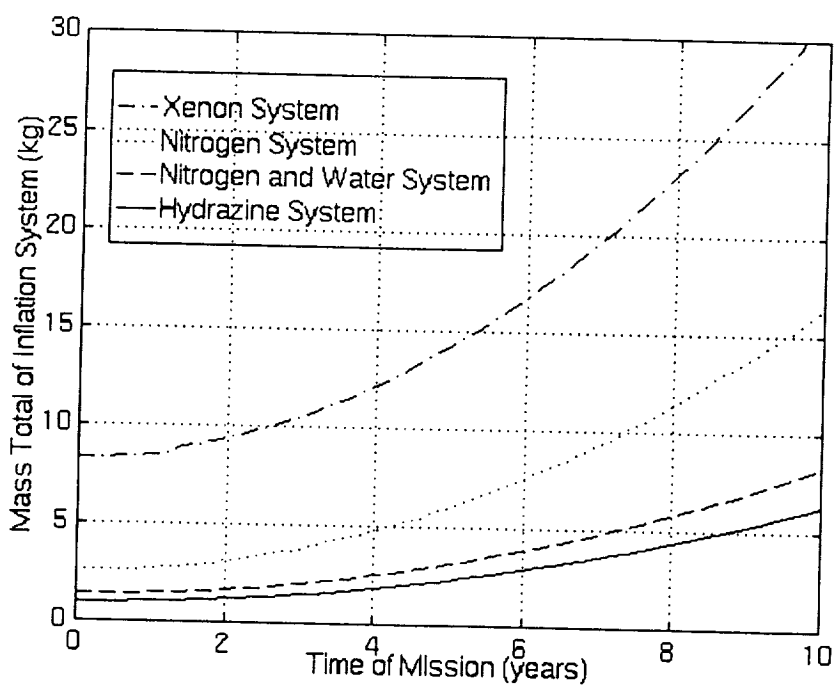


Figure 9. Mass Total for an Existing Propulsion System as a Function of Time of Mission (f/D = 1 and D = 25 m).

Figure 9 illustrates the affect the time of mission has on the total mass of the inflation system. As the time increases, more pressurant is needed to replace gas lost through leakage. As in Figure 8, the xenon and nitrogen systems are more massive than the hydrazine and nitrogen and water systems. This can be attributed mostly to the molecular weight of the pressurants and the resulting tank impact. Figure 10 plots the total mass as a function of f/D ratio for a 5 year mission of a structure with a 25 meter lenticular structure diameter.

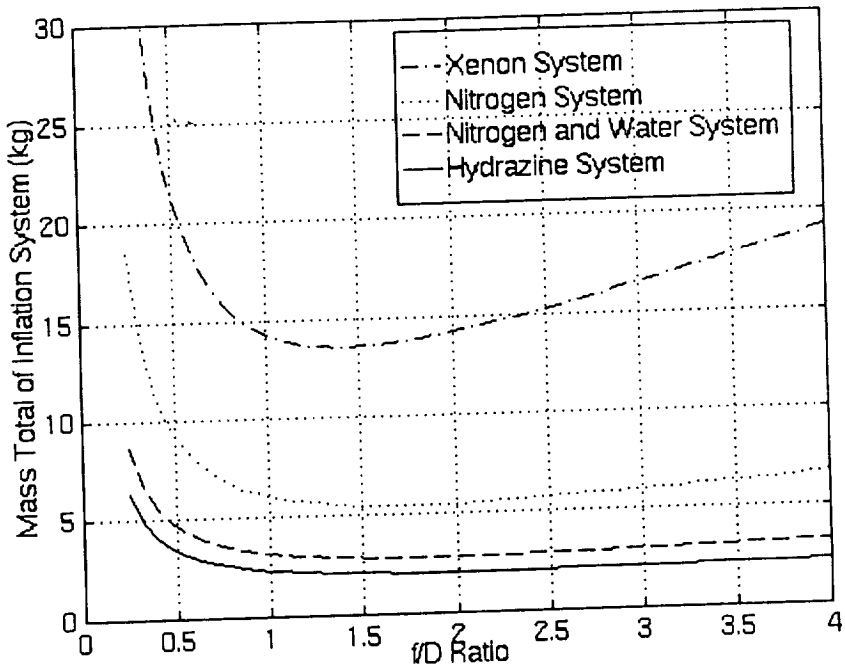


Figure 10. Mass Total for an Existing Propulsion System as a Function of f/D Ratio (t = 5 years and D = 25 m).

Figure 10 illustrates the effect the f/D ratio has on the total mass of the inflation system. It appears that for each system there is an "optimal" f/D for a given diameter. For example, the "optimal" f/D is in the region of 2 for the hydrazine system. Although this is interesting, it is not particularly useful. Mission objectives will decide the f/D ratio as opposed to mass considerations. Once again the xenon and nitrogen systems are more massive than the hydrazine and nitrogen and water systems.

No Propulsion System

The sum of the masses of the set-up components, the total pressurant required, the tank(s), and the tubing equals the total mass of the inflation system. This value represents the amount of mass that would need to be added to a spacecraft which has no existing onboard propulsion system. This value can be obtained for any system or configuration using a modified spreadsheet developed for the first design study. Recall that this mass total does not include electronics, cables, and structure. Figure 11 plots the total mass as a function of lenticular structure diameter for a 5 year mission of a structure having an f/D ratio of 1.

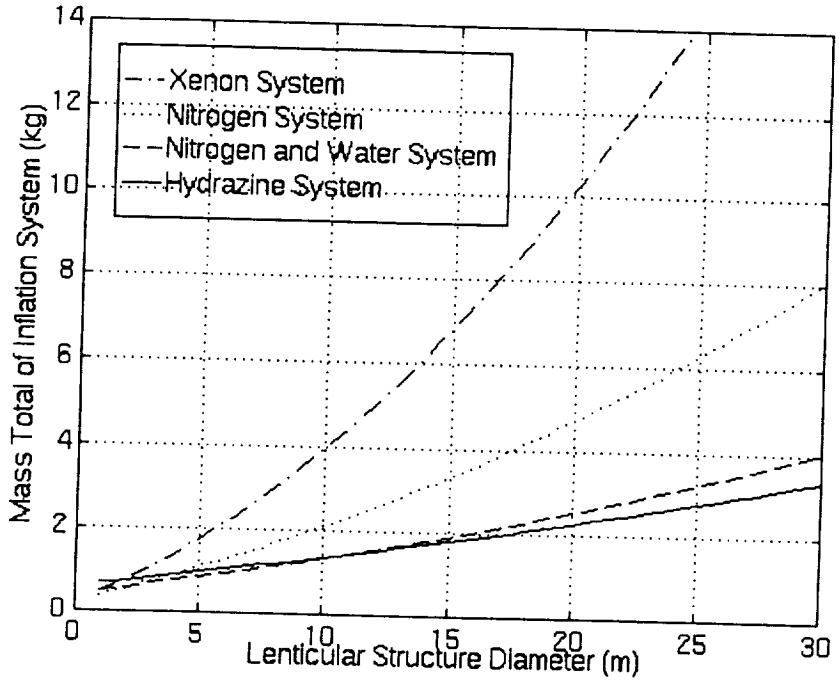


Figure 11. Mass Total for No Onboard Propulsion System as a Function of Lenticular Structure Diameter (f/D = 1 and t = 5 years).

Figure 11 displays a similar trend to Figure 8. The mass total is slightly higher for all cases. Once again, the most promising mass totals are those of the hydrazine and nitrogen and water systems. Figure 12 plots the total mass as a function of time of mission for a structure having a 25 meter lenticular structure diameter and an f/D ratio of 1.

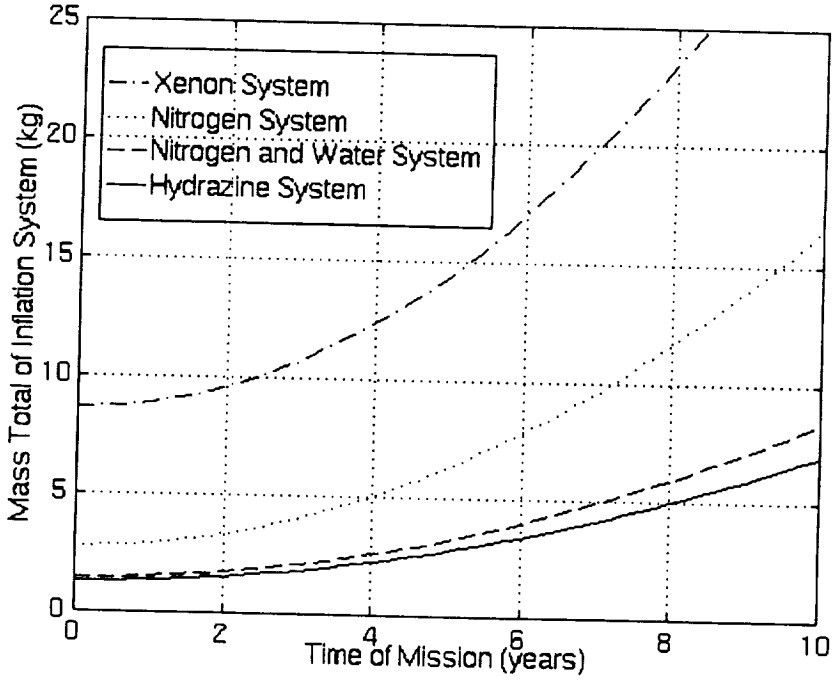


Figure 12. Mass Total for No Onboard Propulsion System as a Function of Time of Mission (f/D = 1 and D = 25 m).

Figure 12 illustrates a similar trend to Figure 9 with slightly higher masses. Figure 13 plots the total mass as a function of f/D ratio for a 5 year mission of a structure with a 25 meter lenticular structure diameter.

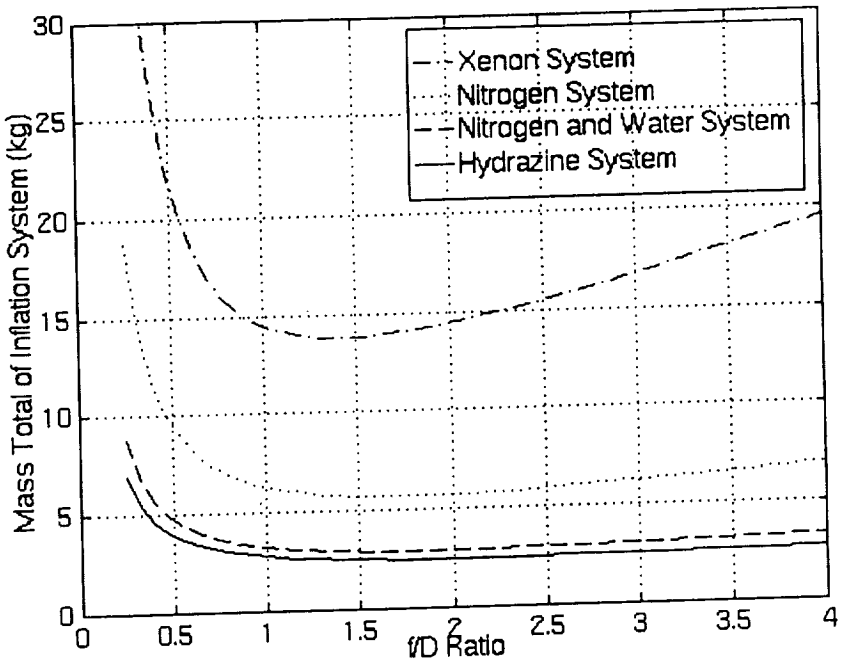


Figure 13. Mass Total for No Onboard Propulsion System as a Function of f/D Ratio (t = 5 years and D = 25 m).

As with the previous plots, the mass total for no onboard propulsion system displays a similar trend to the mass total for an existing onboard propulsion system.

Open Issues

Although each inflation system is different, all have open issues that must be addressed prior to their selection and answered during the development of the inflation system. These issues are summarized for each of the four systems below:

All Systems

- Should the gas flow-control orifice and vent (O2 and SV2 in Figure 2 for example) be downstream of the isolation solenoid valves (SV3, SV4, and SV5 in Figure 2 again)? If so, then three times more solenoid valves are needed. One reason to do this is over-pressure control in case of thermal ratchet (current design allows venting of only one system at a time unless all systems are at the same pressure).
- Will it be possible to generate small enough bursts of pressurant during the maintenance stage of the mission?
- The absorptivity in the wavelength of interest of the gas used to maintain the shape of the lenticular structure.

Hydrazine System

- Can the gas generator itself supply controlled and repeatable pulses of gas when the demand is very small (during the maintenance phase of the mission)?
- Is a catalyst bed heater required? Can the catalyst bed survive the thermal shocks associated with pressurization without being heated to some high temperature first? Will sufficient catalysis occur for very small pulses unless the catalyst bed is heated?
- Will the catalyst bed be poisoned due to long residence time of exhaust gasses caused by the downstream orifice?
- Does the gas coming out of the gas generator have to be actively cooled, or is thermally shorting the tubing to the spacecraft structure enough?
- What is the compatibility of the inflatable structure material with ammonia and possibly hydrazine vapor? This includes the structure material as well as any epoxies or other adhesives used.
- Will ammonia or hydrazine vapor condense on the inflatable structure? If so, what then?

- How well can we get the hydrazine to decompose, especially during maintenance pressure phase when the amount of hydrazine to decompose is very small? At this stage, what amount of ammonia dissociation should be assumed?
- Do we need an orifice up-stream of the gas generator (O1 in Figure 2)?

Nitrogen and Water System

- Compatibility of inflatable structure material with water vapor.
- Will there be a problem of water condensing on the inside of the structure if it gets cold? What is the consequence of this happening? Can the inflated structure be kept warm enough?
- Is single point failure possibility of water tank outlet valve acceptable? If it leaks, the system fills with water to some extent. Will this freeze when the downstream solenoid valves open?
- How much mass can be saved by thermally shorting the water tank outlet tubing to the tank itself?

Xenon System

- Xenon thrusts require very pure xenon to prevent erosion of the thruster points. Can the xenon point of use purity requirement be maintained with hydrocarbons present in the inflation system? Will out-gassed hydrocarbons from the inflatables permeate into the propulsion system?

The actual inflatable structures have many open issues that are beyond the scope of this memorandum.

Recommendations and Conclusions

The two design studies performed attempted to be conservative in mass calculations. That is, mass margins of 20% for liquids and 50% for gases are probably excessive. The design studies revealed three areas of the analysis process that could use improvement. An accurate method of determining leakage would be needed for more accurate overall mass totals. A more accurate tank sizing spreadsheet is needed for improved tank calculations. The current spreadsheet is sufficient for this study yet becomes increasingly inaccurate for larger tanks. Lastly, a more realistic estimate of the tubing mass is needed.

Although this paper dealt primarily with mass considerations, reliability considerations are in many ways as important. There is a much higher chance of failure through leakage in any of these inflation systems than a single-point failure. However, such a single-point failure could occur in one or more of the many valves in a particular inflation system. By minimizing mass (redundant devices) there is an increased chance of failure. Hence, the trade-off with having a low mass inflation system is an increased chance of overall failure. Although the nitrogen and water system displays some of the most promising mass totals, it is a complex system. The hydrazine system also displays promising mass totals but the complexity of this system also raises the question of reliability. There is a potential for something to wrong in such complex systems. The nitrogen and xenon systems, although more massive for many conditions, are simple and reliable in comparison.

It is also important to recall the assumptions made in this analysis. If the operating temperature of the spacecraft is lower than 293 K, all four systems will be more massive. The nitrogen and water system will be most affected by temperature since the vapor pressure of water decreases significantly with decreased temperature. The assumption of near complete ammonia dissociation in the hydrazine system is also worth mentioning. In all likelihood, 90% ammonia dissociation would be possible but an assumption of 40% to 90% might be more realistic. A lower ammonia dissociation would not only raise the overall mass of the hydrazine but also further magnify the open issues surrounding the hydrazine system.

In general, the choice of an inflation system will depend on mission objectives. While a short and small mission would favor the nitrogen system, a large and long mission would favor the hydrazine system. If mass is not a crucial constraint in design, it is recommended the nitrogen or xenon system be used for their simplicity and reliability.

References

1. Friese, G., Bilyeau, G., and Thomas, M., "Initial '80's Development of Inflated Antennas," NASA CR 166060, January 1983.

2. M. Webster and P. Washabaugh, "An Estimate of the Upper Bound of Mass Loss from Inflatable Structures due to Micrometeorite/Debris Collisions," JPL Interoffice Memorandum, September 1995.
3. "Monopropellant Hydrazine Design Data," Rocket Research Company Handbook.
4. Morash, D. and Strand, L., "Miniature Propulsion Components for the Pluto Fast Flyby Spacecraft," AIAA 94-3374.
5. Nebergall, W., Holtzclaw, H., and Robinson, W.. *General Chemistry, 6th Edition*. Lexington, Massachusetts: D.C. Heath and Company, 1980.
6. "Pluto Mission Development; FY94 Final Deliverables Package, Book One," JPL Document, September 29, 1994.
7. G. Ganapathi, "Xenon Thermodynamics: PVT data source comparisons (MIT and NIST)," JPL Interoffice Memorandum, April 1995.

MINIATURE TELEROBOTS IN SPACE APPLICATIONS

S.C. Venema and B. Hannaford
Department of Electrical Engineering
Box 352500
University of Washington
Seattle, WA 98195-2500

71190
10

ABSTRACT

Ground controlled telerobots can be used to reduce astronaut workload while retaining much of the human capabilities of planning, execution, and error recovery for specific tasks. Miniature robots can be used for delicate and time-consuming tasks such as biological experiment servicing without incurring the significant mass and power penalties associated with larger robot systems. However, questions remain regarding the technical and economic effectiveness such mini-telerobotic systems. This paper addresses some of these open issues and the details of two projects which will be able to provide some of the needed answers. The Microtrex project is a joint University of Washington/NASA project which plans on flying a miniature robot as a Space-shuttle experiment to evaluate the effects of microgravity on ground-controlled manipulation while subject to variable time-delay communications. A related project involving the University of Washington and Boeing Defense and Space will evaluate the effectiveness of using a minirobot to service biological experiments in a space station experiment "glove-box" rack mock-up, again while subject to realistic communications constraints.

I. INTRODUCTION

Humans have a central imperative to explore and to expand their physical domain. Increasingly this is directed to the exploration and eventual colonization of space. However, the hazards of vacuum, high temperature fluctuations, and radiation require humans to live and work in highly protected artificial space environments. The high costs of providing reliable protection for humans in space motivates the use of robotic tools to augment and enhance human activities in space.

In the future, fully automated robots may be utilized to perform coordinated work in space. Near-term activities such as the International Space Station Alpha may also be able to benefit from robots by allowing more activities to be accomplished without placing additional time demands on the limited human crew availability. However, the feasibility of fully automated robots has been limited due to a lack of confidence in Artificial Intelligence technology. Teleoperated robots offer many of the same benefits as automated robots by shifting much of the decision-making intelligence to human controllers. In addition, if the telerobots are controlled from Earth-based control stations, many space activities may be accomplished with little or no orbital crew time requirements. However, the communications requirements for ground-controlled telerobots present difficulties to current and near-term planned space activities.

While it is desirable to be able to do "human-scale" work with human-sized robots in space, the high launch weight, cost and complexity of such systems has limited their near-term viability. In contrast, small robots offer the chance to perform many types of IVA tasks such as environmental maintenance, housekeeping, and the operation of scientific and biological payloads without paying the penalties of higher mass and complexity associated with larger robots. In the future, these minirobots may also be employed for EVA activities such as exterior surface inspection and repair.

Direct-drive actuation may be used to minimize mass, friction, and backlash commonly present in geared drive robot systems. These properties allow open-loop force control to be used, so that delicate manipulation tasks may be done without requiring force sensors on the robot end-effector. However, some questions remain as to the effect of microgravity on the friction properties of precision bearings. These questions must be answered so that satisfactory friction models can be used for open-loop force control in microgravity.

II. Background

An important factor of the space environment is the complex nature of the communication links between ground station and this environment. Although in principle, this phenomenon can be simulated, in practise, realistic simulation of the communication link is very difficult to obtain because the existing link technologies were designed primarily for data-logging and involve many interfaces with ground and flight subsystems. Each of these interfaces imposes time delays and bandwidth limitations which vary with time and as a function of the overall mission activities. This is especially true for the Shuttle data links through the TDRSS system of relay satellites. Most operational experience with this data relay system has been during use for archiving (logging) of research data.

Time delays between a human operator and a slave robot have been studied since the sixties. In the original work, Sheridan and Ferrell [1] found that when confronted by time delays greater than about 300 ms., operators adopted a "move-and-wait" control strategy in which he/she performed a task by commanding single simple movements, and waited for visual or other feedback of the results before making another small command. This strategy is very sensitive to the magnitude of the delay and causes dramatically reduced manipulation performance. These and other studies assumed a constant time delay between operator and remote manipulator. Realistic space communication systems will have more complex properties—notably, variable delay, intermittent communication, de-synchronized delays on multiple feedback channels, and bandwidth limitations. These links were designed for archival accumulation of data and have rarely been used for reactive, real-time, decision making and control by a human in the loop.

Friction properties of robot joints in microgravity environments remain uncharacterized. This information is important because it is needed to allow better understanding and control of robot dynamics and

is also needed for precise open-loop force control. In 1-g, bearings in each joint of a serial mechanism are loaded with radial and axial forces which are dominated by a relatively constant gravitational component. In microgravity, this component will be gone and what remains will be a fluctuating force due to manipulator dynamics and contact. Thus micro-surface contact in the bearings will be rapidly changing. This phenomenon cannot be studied on the ground because air bearing tables do not simulate weightlessness in enough dimensions, and floatation tanks supply too much viscous damping.

III. Previous Work

STS-RMS System

The Space Shuttle Remote Manipulator System (RMS) — nominally the first robot manipulator in routine space use — has been an essential and extremely useful component of the Space Transportation System. It has repeatedly been used in powerful and unforeseen ways. However, its utility is also limited because it requires a manned mission, and cannot be controlled from the ground. Subsequently, NASA studied the Flight Telerobotic Servicer as a remotely controlled telerobot for servicing satellites in geosynchronous orbit (unreachable by manned missions). However, this project was canceled due to funding limitations.

DLR-ROTEX Experiment

The ROTEX experiment was conducted by Dr. Gerd Hirzinger in May of 1993 [2]. In this experiment, a human scaled, gear driven arm, flown in the Space-Lab D2 was successfully controlled from the Space-Shuttle flight deck, as well as from a ground control station in Germany. The robot was controlled with a "space-ball" hand controller, a device which measures cartesian forces and moments applied by an operator's hand while grasping a ball mounted on the end of a force/torque sensor. The measured forces and moments were used to control the robot's 6-axis cartesian velocities in proportion to the forces applied by the operator's hand. A sophisticated end effector provided an extensive array of end point sensing.

Visual feedback from the ROTEX robot worksite to a ground-based operator was provided via video cameras. However, the presence of substantial round-trip communication delays (5-7 seconds) between the Space Shuttle and the German-based ground station necessitated the use of a graphics simulation system which provided the operator with a simulated real-time view of the remote worksite to facilitate robot operations. Command sequences generated by user interaction with the graphics simulation system were sent up to the ROTEX robot for a later execution.

The ROTEX experiment has generated significant excitement for the potential of space teleoperation, but does not address the high cost and low mission frequency issues associated with large bulky space

robots. A significant new initiative is the ETS-VIII, a Japanese satellite exclusively dedicated to telerobotics experiments [3]. This project is scheduled for launch in 1997.

UW Mini Direct Drive Robot

The University of Washington has recently developed a prototype, 5-axis, mini direct drive robot [4]. The robot (Figure 1) has an overall length of 10 cm and a workvolume of 10 cm^3. It has 5 motion axes

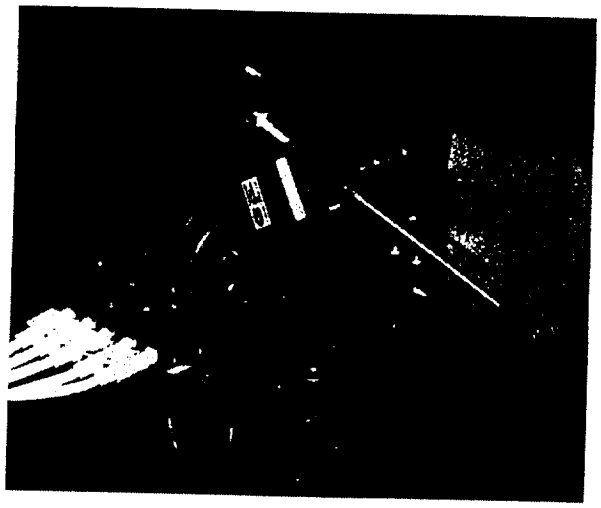


Figure 1: University of Washington Mini Direct-Drive Robot

giving complete freedom to position and orient an axially symmetric tool within the manipulator's joint limits. Analog optical position encoders give positioning repeatability to 1 micron at the end-point. The design makes use of flat coil actuators taken from miniature hard disk drives for very low friction levels, absence of torque ripple, and design flexibility [5]. The current prototype uses actuators from 5 1/4", 3 1/2", and 1.8" disk drives to provide a range of masses and torque capabilities for the various joints in the mechanism. This prototype has already been used in teleoperation tasks that manipulate objects ranging from 5mm to 80μm in size.

IV. MICROTREX

Objectives

The Microtrex (Micro Telerobotic Experiment) project proposes to use a small flight telerobot in low earth orbit to perform ground-controlled telerobotics experiments. Microtrex will have two primary experimental objectives: First, mathematical models of human/machine system performance [6,7,8,9,10] which have been extensively developed and tested in ground experiments must be calibrated for actual flight operations. Many aspects of the flight environment, such as the detailed characteristics of the contemplated communications channels, have not yet been simulated on the ground.

Secondly, Microtrex will obtain data from which a model of robot joint friction in micro-gravity can be constructed. These factors, unique to the space environment, must be studied with a flight experiment

before planners can guarantee successful performance of manipulation tasks involving contact with rigid objects.

Finally, we intend to make Microtrex a motivational and educational experience for children to encourage and prepare them for study in engineering and science. MicroTrex will include a live interactive video link with elementary and secondary schools to create a motivational and educational experience for children.

Experiment Description

The Microtrex experiment will consist of a flight module and ground station. The Microtrex flight module will consist of the mini-direct-drive manipulator, a control computer, three CCD cameras, and a task board for experimental manipulation. All manipulation experimentation will take place within the fully enclosed flight module.The task board will be mounted within the work volume of the flight module manipulator and will contain several experiments designed to resolve the technical uncertainties described above. First, in order to measure micro-gravity friction characteristics, a surface of controlled hardness and smoothness properties will be provided for constant force and controlled sliding contact tests. Second, to test system capabilities in more application related tasks, and measure operator performance, a peg-in-hole task will be provided. The peg will be held in a restraining fixture, grasped by the manipulator, and inserted into two holes with different tolerances. To be consistent with previous ground studies (e.g. [6]), tolerances of 5% and 0.05% (clearance over diameter) will be used. Finally, additional modules will be included in the task board to perform biotechnology related demonstration tasks. These experimental tasks will allow in depth model calibration for a wider variety of end-user application elements.

The launch carrier for this experiment has yet to be chosen, but will most likely be the STS with the experiment mounted in a cannister in the Hitchhiker G or M configuration. A mock-up of this flight configuration is shown in Figure 2. In this case, two downlink channels (1200bps and 1.4Mbps) and a single uplink channel (1200bps) will be available for communication. This bandwidth will be sufficient for the proposed experiment. However, latencies will be severe (2-4 seconds, variable) due to the TDRSS system. Since the majority of the data bandwidth is required for the video downlink, one option being considered is a direct radio line-of-site downlink from the top of the GAS cannister. This places some additional constraints on Shuttle orientation. If available, the direct radio link will have very low latency but will require tracking antennae on the ground and intermittent operation. To make maximum use of whatever communication link is chosen, we will incorporate data compression for the returning sensor information from wrist force torque sensor and video cameras.

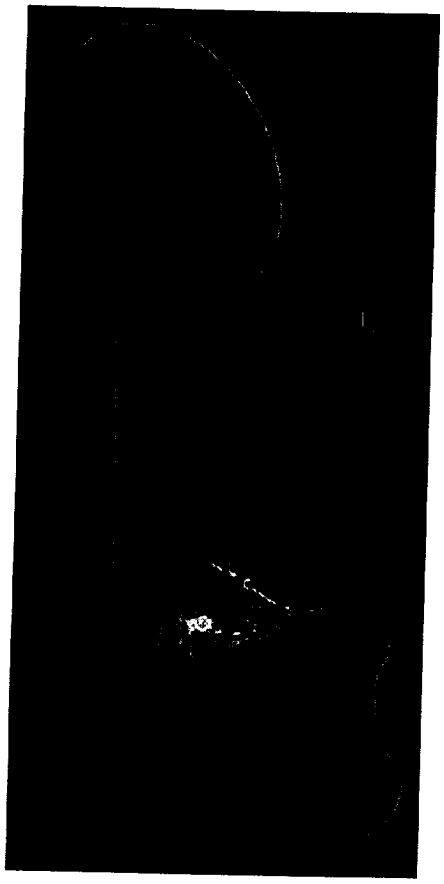


Figure 2: Mock-up of Microtrex flight module in Space Shuttle Hitchhiker configuration.

During experiment operations, an operator situated at the ground station will send commands to the flight module using a multi-axis hand controller. These commands will consist of a data stream of incremental motion commands in the six axes (x, y, z, roll, pitch, yaw) relative to a coordinate system on the flight module task board. The operator will also command the manipulator into a pre-planned sequence of tasks designed to test remote manipulation capabilities and measure joint friction in micro-gravity.

Three CCD cameras will image the end effector and task board of the flight module. Their signals will be compressed and sent back to the operator along with a stream of force and position information. The operator will view the operation using the downlinked video signals, and also be able to view a realistic real-time computer animation of the manipulator from any desired angle.

In order to resolve the friction uncertainty, the Microtrex robot manipulator will be equipped with a micro 6-axis force-torque sensor either at the wrist or in the task board. During contact tasks, contact forces and torques will be both measured by the wrist sensor, and calculated from the joint actuator torques. The difference between the two will be due to the friction. By measuring the difference between actuator and wrist sensor force/torque signals during contact, and transforming that back to the joints via the transpose of the Jacobian matrix, a measurement will be obtained of the joint friction.

Data Analysis

The Microtrex experiment will be the culmination of at least a decade of ground studies in the performance of human-in-the-loop teleoperation systems [11,12]. These and other studies have yielded extensive performance data on ground teleoperation both with and without simulated time delays. Performance of a variety of tasks has been extensively analyzed through a variety of mathematical measures such as completion time, sum-of-squared contact force, RMS force, and observed error rate. From this extensive database, analytical models have been devised through which performance can be estimated on new tasks before experimental teleoperation. One such model is the Hidden Markov Model of human-machine task performance [9] which can derive completion times, and contact force signatures from engineering descriptions of tasks.

However, none of these models has been calibrated against actual space flight. A central aim of the Microtrex experiment is to obtain calibration of model parameters for the unique conditions of microgravity and the complex conditions of orbital communication links. All of the Microtrex proposed experiments would be performed on the ground using the control station, ground prototype manipulator, and a simulated time-delay communication link. The results of the Microtrex experiment will be calibrated analytical models of human-machine performance based on previously developed Hidden Markov Model methodologies and software. This software package will enable a mission planner to design a model for his/her contemplated application, and predict human-machine performance of a telemanipulation system.

Existing work on friction in robot manipulators has focused on a stationary model consisting of three components: stiction, Coulomb friction, and viscous friction. Although these models are difficult to identify with great precision, in the 1-g environment, adequate results have been obtained to justify a stationary model - that is one which does not vary with time [13]. Results from the friction testing component of the experiment will be analyzed to develop a new model which can quantify possible discontinuous, non-stationary friction behavior.

V. Space Station Biomedical Research

Many space station research activities will require significant amounts of object manipulation activities. Many of these steps require delicate, high-accuracy motions with small glass "pipette" tubes in order to pick and place small fluid samples or objects floating within small fluid samples. Human crew can perform these activities but current estimates of crew time availability show that only about 2 hours of crew time will be available per day to service all space station experiments after all other required activities such as housekeeping, maintenance, etc. have been accomplished.

A small, ground-controlled telerobot could be mounted within a space-station "glove-box" rack to service multiple experiments without requiring significant human crew interaction. We are working with

Boeing Defense and Space Company to develop a "glove-box" robot for experiment servicing using ground-controlled teleoperation. The robot will be mounted on a mobile platform which allows it to be positioned in front of various experiments in the glove box that require servicing.

During the first stage of this project we are integrating the existing 5-axis manipulator into a Space Station glove-box. Figure 3 shows a mock-up of this configuration using a ~1m wide glove-box. Control

Figure 3: Mock-up of Space Station "glove-box" with mini direct-drive robot positioned for servicing biomedical experiments.

of the robot from across the country will be demonstrated between our laboratory (Seattle, WA) and Boeing (Hunstville, AL) via an internet connection. The eventual goal is to allow individual experiment PI's to service their own experiments by "driving" the robot from their own offices anywhere in the world via internet connections.

A secondary challenge is the development of an efficient teleoperation communication protocol that maintains safe operating conditions at all times in the face of significant communication errors and disruptions. We have developed preliminary versions of this protocol [14] which have already been used for teleoperation tasks with force feedback over internet links between sites as distant as Seattle, USA and Padua, Italy.

VI. Conclusions

Miniature robot manipulators have the potential to make a significant contribution to both future and near-term space activities. Ground-controlled manipulators can significantly reduce crew time requirements for experiment servicing tasks on the Space Station. While some important technical challenges remain, significant inroads will be made in the near future. We have described experiment activities

which will address some of these concerns and help promote the use of miniature robots in future space activities.

VII. References

[1] T.B. Sheridan, W.R. Ferrell, "Remote Manipulative Control with Transmission Delay," IEEE Transactions on Human Factors in Electronics, vol. HFE-4, pp. 25-29, 1963.

[2] G. Hirzinger, B. Brunner, J. Dietrich, J. Heindl, "Sensor-Based Robotics -- ROTEX and its Telerobotic Features," IEEE Transactions on Robotics and Automation, vol. 9, pp. 649-663, October 1993.

[3] M. Oda, "On the Dynamics and Control of ETS-7 Satellite and its Robot Arm," Proceedings, IROS-94, Munich, Sept. 1994.

[4] B. Hannaford, P.H. Marbot, M. Moreyra, S. Venema, "A 5-Axis Mini Direct Drive Robot for Time Delayed Teleoperation," In Press, Proc. Intelligent Robots and Systems (IROS 94), Munich, Sept. 1994.

[5] P. Buttolo, D.Y. Hwang, B. Hannaford, "Experimental Characterization of Hard Disk Actuators for Mini Robotics," Proc. SPIE Telemanipulator and Telepresence Technologies Symposium, Boston, October 31, 1994.

[6] B. Hannaford, L. Wood, B. Guggisberg, D. McAffee, H. Zak, "Performance Evaluation of a Six-Axis Generalized Force-Reflecting Teleoperator," JPL Publication 89-18, Pasadena, CA 91109, June 15, 1989.

[7] B. Hannaford, L. Wood, D. McAffee, H. Zak, "Performance Evaluation of a Six Axis Generalized Force Reflecting Teleoperator," IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, pp. 620-633, 1991.

[8] B. Hannaford, P. Lee, "Hidden Markov Model Analysis of Force/Torque Information in Telemanipulation," Proceedings 1st International Symposium on Experimental Robotics, Montreal, June 1989.

[9] B. Hannaford, P. Lee, "Multi-Dimensional Hidden Markov Model of Telemanipulation Tasks with Varying Outcomes," Proceedings IEEE Intl. Conf. Systems Man and Cybernetics, Los Angeles, CA, Nov. 1990.

[10] B. Hannaford, P. Lee, "Hidden Markov Model of Force Torque Information in Telemanipulation," International Journal of Robotics Research, vol. 10, pp. 528-539, 1991.

[11] B. Hannaford, W.S. Kim, "Force Reflection, Shared Control, and Time Delay in Telemanipulation," Proceedings, IEEE Intl. Conference on Systems, Man, & Cybernetics, Cambridge, MA, Nov. 1989.

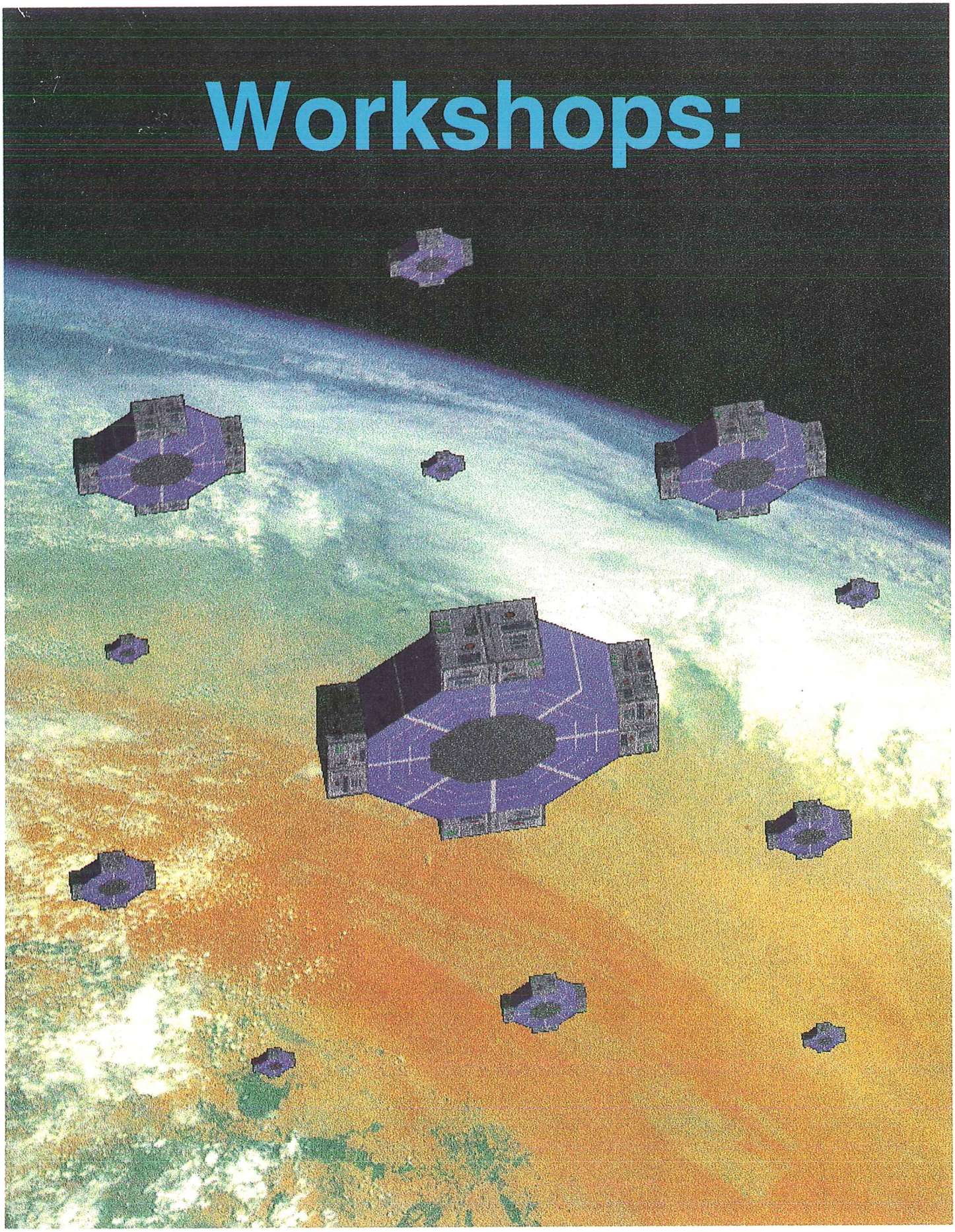
[12] W.S. Kim, B. Hannaford, A.K. Bejczy, "Force-Reflection and Shared Compliant Control in Operating Telemanipulators with Time Delay," IEEE Transactions on Robotics & Automation, vol. 8, pp. 176-185, 1992.

[13] B. Armstrong, "Control of Machines with Non-linear Low-velocity Friction: A Dimensional Analysis," In "Experimental Robotics I: The First International Symposium", V. Hayward & O. Khatib, Ed., Springer Verlag, Montreal, June, 1989.

[14] P. Buttolo, J. Hewitt, R. Oboe, B. Hannaford, "Force Feedback in Virtual and Shared Environments," Poster-Exhibit, IEEE SYstem, Man and Cybernetics, Vancouver, October, 1995.

OMIT TO END

Workshops:



Forward to Workshop Reports

The program for The first International Conference on Integrated Micro-Nanotechnology for Space Applications included two days of focused workshop deliberations on the opportunities for insertion of key aspects of microtechnology into space systems. Each Workshop with the exception of Workshop 9 had two tasks:

1. Identify and develop technology roadmaps for applications of microtechnology in space

2. Develop conceptual designs and requirements specific to the development of one or more subsystems of an Untethered Flying Observer (UFO).

The UFO is conceived to be a free flying vehicle hosted on a larger satellite. Its mission, activated on demand or on predetermined conditions, is to detach itself, capture detailed images of the host and transmit these images to the ground. This mission is to last no more than 48 hours. Workshops 1 through 8 were assigned one or more of the UFO's subsystems to define in a manner consistent with the mission and the specified size, weight and power budgets. Workshop 9 was tasked to integrate these subsystems and act as system manager during the definition process. Those subsystems requiring extensive development to meet the stated mission objectives were flagged and technology road maps were designed to foster their growth. In addition, Workshop 9 identified and defined alternate missions for untethered, parasitic, microengineered spacecraft. The Panels and the individuals serving as Cochairs are listed below.

Panel 1: Sensors & Transducers for Space Applications
Chair: S. Amimoto (The Aerospace Corp.)

Panel 2: Software
Chair: M. Gorlick (The Aerospace Corp.)

Panel 3: Low Power Electronics for Communications
Cochairs: J. Hurrell and W. Bloss (The Aerospace Corp.)

Panel 4: Biomedical Applications for Manned Space Flight
Chair: C. Sawin (Johnson Space Center)

Panel 5: Nanosatellite
Chair: S. Janson (The Aerospace Corp.)

Panel 6: ASIM Applications for Current Space Systems
Chair: E. Y. Robinson (The Aerospace Corp.)

Panel 7: Low Power Electronics for Data Processing
Cochairs: N. Sramek (The Aerospace Corp.) and G. Frazier (Texas Instruments)

Panel 8: Materials, Manufacturing and Design
Chair: H. Helvajian (The Aerospace Corp.)

Panel 9: Specific ASIM Development (Untethered Flying Observer)
CoChairs: D. Sutton and R. Stroud (Aerospace Corp.)

All of the Workshops followed the agenda outlined in the following table.

Agenda for Workshops

Day	Time	Workshop
Wednesday	2:50 pm	Introduction to Workshop Panels
Wednesday	3:30 pm	Session 1: Outline specific workshop goals. Panel Chairs meet for dinner.
Thursday	8:00 am	Session 2: Workshop Member discussions (use view charts to educate participants:. Participants listen.
Thursday	11:00 am	Session 3: Open discussion session; participants can make presentations (5 min each)
Thursday	12:00 pm	Panel Chairs meet working lunch (exchange information; get input from Sutton panel).
Thursday	1:00 pm	Session 4: Workshop Member discussions (use viewcharts to educate participants). Participants listen.
Thursday	4:00 pm	Session 5: Open discussion session; participants can make presentations (5 min each).
Thursday	5:30 pm	Panel Chairs meet
Friday	8:00 am	General Assembly
Friday	8:10 am	Panel Presentations 1 through 9
Friday	11:40 am	Open Discussion
Friday	12:00 pm	Workshop/Wrap up

However, many of the Workshop Chairs met in general session during the evenings of Wednesday and Thursday in order to participate in the integration of the UFO design. The following Section of these Proceedings contains the completed Workshop reports.

Workshops
Wednesday PM to Friday

- **Panel 1: Sensors & Tranducers for Space Applications**
  - Cochair: S. Amimoto (The Aerospace Corp.)
- **Panel 2: Software**
  - Cochair: M. Gorlick (The Aerospace Corp.)
- **Panel 3: Low Power Electronics for Communications**
  - Cochairs: J. Hurrell and W. Bloss (The Aerospace Corp.)
- **Panel 4: Biomedical Applications for Manned Space Flight**
  - Cochair: C. Sawin (JSC)
- **Panel 5: Nanosatellite**
  - Cochair: S. Janson (The Aerospace Corp.)
- **Panel 6: ASIM Applications for Current Space Systems**
  - Cochairs: E. Y. Robinson (The Aerospace Corp.)
- **Panel 7: Low Power Electronics for Data Processing**
  - CoChairs: N. Sramek (The Aerospace Corp.) and G. Frazier (TI)
- **Panel 8: Materials, Manufacturing and Design**
  - CoChair: H. Helvajian (The Aerospace Corp.)
- **Panel 9: Specific ASIM Development (Untethered Flying Observer)**
  - CoChairs: D. Sutton and R. Stroud (Aerospace Corp.) and D. Welles (JSC)

Workshop 1: Sensors and Transducers

Chairman: Sherwin Amimoto, The Aerospace Corporation

I. Introduction

The scope of this panel covers sensors and transducers that could be used in future space applications and missions. The primary emphasis was placed on the payload or imaging system for the Untethered Flying Observer(UFO) constructed using MEMS technologies. The UFO mission is to observe a mothership from which the UFO is launched. The optical train, the focal planes, shutter, focus control, A/D converters, data compression are discussed within the context of the UFO mission. Technology risks were deemed low. Advanced technologies such as micro-optics and high temperature superconductors are briefly reviewed.

Many sophisticated transducers and sensors are commercially available. But many do not appear to be space qualified for radiation hardness. Several novel sensors were discussed for advanced applications.

The major participants for this workshop were:

Robert Duncan	Sandia Laboratory
Adolfo Guitierrez	InterScience, Inc.(ISI)
Will Trimmer	Belle Mead Research
George Rossano	The Aerospace Corporation
Ali Habibi	The Aerospace Corporation
Hirobumi Saito	Institute of Space & Astro. Sciences (ISAS)
Lou Hermans	IMEC
M. Robyn	The Aerospace Corporation

II. Technology Assessment for Space Systems

A. Near Term

1. Current State-of-the-Art

The current state of the art of many sensors and transducers is at a high level of maturity. For simple sensors and transducers, a large variety of components are offered commercially by a large number of sources. These companies represent many of the major US and foreign semiconductor producers. In Table 1, we summarize a quick look at the CD ROM-furnished database called DATA/PAL[1] which is used as a selection guide for electronic components. Although the designer may think he is blessed with a plethora of choices, many available components do not meet military specification standards nor are they adequately radiation hardened. The smorgasbord of electronic components is shaped by their commercial earth-bound applications. Many companies are not interested in the development of military standard or hardened components since these sales represents far less than 10% of commercial sales. Hence the selection of semiconductor sensors and transducers for use in a severe space environments may be limited by availability, funding to develop hardened electronics and sensors, or by

modifications necessary to successfully use available sensors and transducers. However for non-stressing applications, the breath of sensors is very good indeed.

Sensor/Electronic Type	No. of Entries (devices)	No. of Manufacturers
Accelerometer	76	3
Fluid/liquid sensor	37	5
Magnetic/Hall effect sensors	638	17
Pressure	2588	*
Temperature	313	12
Microcontroller, 8bit A/D	1552	*
Focal Planes, CCD	149	9

* Very large number of manufacturers

Table 1. Availability of Commercial Sensors and Transducers.

Many other devices may be useful for future space observations. These include millimeter wave pulse formers, digital demodulator, mixers, etc. which have been constructed from high temperature superconducting materials with revolutionary improvements over conventional devices[2]. They have been space qualified and are considered available. For example Superconducting Quantum Interference Devices(SQUIDs) have been qualified for a Space Shuttle flight and were used on STS-52 to measure temperature. The sensitivities of SQUIDs, approaching 10^{-21} Weber or a field sensitivity of one femtoTesla in a one Hertz bandwidth, exceed the capability of conventional electronics by 3 orders of magnitude.

2. Attractive Opportunities for Space

The small compact size and inherent mass producibility favors the use of MEMS technologies for certain space applications. For remote sensing applications, imaging at low spatial resolution or large field of view are readily accommodated. Other missions include mapping of magnetic fields, gravity fields, charged particles in space, etc. For high resolution imaging applications due to aperture size limitations, the imaging range will be restricted. A small imager will be well-suited for close range view of space objects. Short term missions are conceivable such as imaging around a mother ship, NASA's Shuttle, or the Space Station. NASA is developing a SPRINT robot and planning experiments for observations and transport of tools and other objects while assisting astronauts on missions serving as preparatory steps to assembly of a Space Station in space. For long term observation, more station keeping by the nano-sat is necessary for formation flying and is subject to constraints on the size of the propulsion system of the nano-sat and the differences in the drag coefficients of the mothership and the remote observer. At high orbits, the drag coefficients are favorable for longer missions but ground communication links may require more power for the remote sensor or communication via a nearby relay.

3. Technology Development Issues

Improvements in sensors and the development of new sensors will speed the use of MEMS and nano-technologies in space. Programs with these goals would advance performance characteristics well beyond present radiation hardness, thermal cycling, lifetime, and other reliability levels. But improvements in sensors and transducers by themselves will be insufficient. The outputs of sensors alone are seldom of use by

themselves. The sensor output in the form of data needs to be correlated with other observational data for example position or angles of observation, distances, time of observation, duration of observation, etc. To achieve this fusion, emphasis should be placed on the development of an infrastructure to allow the rapid prototype development of sensors fused to data recording, fusion with other sensors, data storage, processing, and subsequent communication all practiced with power management to keep system size, weight, and power appetite small. This infrastructure will encourage end users to propose systems that are customized for their applications. One strategy is to foster use of high density packaging concepts[4], a powerful and ingenious technique that embodies these properties. These concepts should be specifically encouraged due to its numerous advantages of compactness, short interconnects leading to lower power consumption, and lower power storage, and lighter mechanical enclosures that are required. For theses high density packing techniques, access to known good dies(KGD) are also needed. These KGD represent the semiconductor integrated circuits that are the building blocks for the sensor platform that fuses sensors to all other functions required for their use in space.

B. Far Term (5 to 20+ years)

1. Emerging Technologies/Development Roadmap

Optical fabrication using micro-optics is a technology that should be encouraged[5]. Fabrication methods for micro-optics including lenses, steering lens arrays, diffractive and replicated optics are common to the fabrication of many semiconductor electronics for example, photo-lithography, etching, ion milling, mask fabrication, LIGA processing, X-ray lithography, electroforming, diamond turning, direct electron beam write, etc. Hence it may be possible to fabricate micro-optical components using the identical foundry as is employed for electronic components. This could lead to production of monolithic optical/electronic components. The status of design methods and theory are well established.

Development of prototypes over the next few years is projected by G. Gal to culminate in insertion of useful components near the turn of the century. For future advancements in micro-optics, vector diffraction theory and software development will be needed to address feature sizes and spectral and optical design. Developmental programs in the technologies of diffractive/refractive micro-lens and arrays, diffractive and reflective elements, and liquid filled groups are envisioned. They could be used to fabricate components consisting of microlenses, multifocal lenses, dispersive lenses, corrector optics, beam combiners and diverters, aspheric generators, wavefront samplers and micro-reflectors. This could enable hybrid integrated systems such as optical beam steerers, smart focal planes, compound eye sensors, wide angle imagers, and display devices for the operator and allow their use in a space system.

Robert Duncan reported on the proposed use of differential gravity measurement methods coupled with precision navigational sensing, to map out mineral and fossil fuel reserves based on their associated gravitational anomaly. This would require a large fleet of low orbit sensors using differential measurements of gravity gradient information. Measurements at 50 miles (not 50 km) should be adequate to obtain a 50 km^2 'pixel' size (7 km x 7 km) on the earth geodesy map. This swarm of nanosats concept should prove to be better than the current EOS proposal since it would involve up to 50 nanosats, and hence up to 50 times more data than the single orbiter concepts. Also, real-time differencing between the nanosats could provide much more accurate gradient information, provided that this differencing could be done with high precision. The development of this superconducting sensor is presently under development at Sandia Laboratory in Albuquerque.

High temperature SQUIDs have been recently fabricated at Los Alamos National Laboratory and are in the early stages of testing and characterization. But other high temperature SQUIDs designed to operate at microwave and millimeter wave frequencies are mature and exhibit performance improvements of over a factor of 100 times conventional devices. Vector magnetometers using a MEMS version of a flux gate or a high temperature SQUID would have sufficient sensitivity to detect the vector direction of the earth's field in even high orbits. This could be used for navigation and control or to detect deviations in the earth's magnetosphere. In turn induction charging effects of large man-made structures such as power grids and conducting metal pipelines on earth could be predicted to prevent power grid outages and corrosion or safety issues of a pipeline.

For long term missions, cryogens may not be feasible to provide cooling even for high temperature superconductors. The use of thermal radiators on low-earth orbit using heat pipes and phase change materials to maintain near isothermal conditions is thought to be limited to 105 K which is presently higher than is desired for YBCO superconductors[2]. Thallium-based superconductors have Tc = 125 K and may prove useful on such a thermal radiator. However the thallium superconductor material science is not anywhere near the level of development of YBCO superconductors. Sandia Laboratory is using this approach. Another opportunity is to use a small cryocooler such as the Inframetrics unit to achieve cooling to about 60 K. This could prove useful for the inertial units proposed for use in the oil/mineral exploration.

III. UFO Subsystem Summary

A. Recommendations

1. Prioritized Technology Requirements

A systems study in some detail would be recommended to identify areas for further development. A demonstration program may also identify additional issues.

Power management of all electrical components is necessary to achieve minimally low battery weights. Strategies to implement power management include turning on subsystems such as communication links and sensors only when needed and placing controllers and processors in a safe sleep mode with suitable wake-up modes.

2. Ground and Space Qualification Issues

No show stoppers were identified for the UFO mission. In part the selection of the orbit and the active mission time which were assumed for the UFO mission resulted in very benign requirements regarding radiation hardening for total dose and single event upset. Power for the focal plane, laser rangefinder/focus control, shutter, and low resolution imager is low.

B. Imaging Payload Requirements for a UFO

Specific design requirements for the imaging function or payload of the UFO were discussed in the workshop. These includes the subsystems of the optical train, focal plane, shutter, alternate imaging strategy, pointing, and focus control mechanism. Commercial video cameras use many of the identical subsystems. For the purpose of designing a camera system, a limitation of a 4 inch size envelope of the imaging system

(imposed by the size of available silicon wafers) and a UFO weight limitation of 1 kg were imposed. The range of sizes and weights of existing commercial video camera systems are comparable to the design performance despite differences in performance criteria and design guidelines discussed below.

The UFO will be deployed from a mothership. The life of the mothership is assumed to be 7 years as is typical of most satellites. The mission duration of the UFO following deployment is 48 hours. The UFO is stored in a canister to increase the survivability of the UFO to radiation, contaminants, and space environment prior to its release for its observation mission. Self test functions, a communication relay, and navigation data will be supported or provided by the canister. Following its release from the canister, the UFO will be maneuvered using its own propulsion system into a decentered orbit that roughly parallels the mothership to enter a phase of flying together in formation. This orbit can be tilted to allow vertical viewing of the mothership above and below the plane of the mothership orbit. The UFO will also be maneuvered into a slow spin. This will allow the UFO camera system to pan over the entire surface of the mothership. The concept of decentered orbits will allow the UFO to circle the mothership once for every orbital revolution of the flying formation.

The orbital altitude of the mothership was fixed at 800 nm. The corresponding orbital period is 110 minutes. For a 48 hour mission duration 26 mutual orbits of the UFO about the mothership will take place which more than adequate to cover the motherships surface.

The design of the camera system was predicated on acquiring images of the surface of a large satellite. The UFO is launched from a mothership and through maneuver circles the mothership to photograph the exterior of the mothership. Typically the mothership satellite may have a cylindrical shape with large flat solar panels located to its side that are oriented towards the sun to generate power. A preliminary and some what arbitrary resolution of 0.3 cm on the mothership surface was selected. Its surface area was estimated to be 100 m^2 . A factor of 10 was assumed in terms of imaging area to account for overlapping of the images and to ensure complete coverage of the satellite surface area. If the resolution is matched to a single pixel, then 10^8 pixels are needed to provide the large satellite's surface area. A CCD or similar camera was selected for use at visible wavelengths. Ideally the design and fabrication of the camera would be compatible with MEMS technologies.

What type of information can be extracted from a complete stereoscopic imagery of the mothership? We can speculate that the following categories of information may be extracted as shown in Table 2. Often confirmation of the listed phenomena will provide spacecraft designers a much higher degree of certainty in proposing future improvements and assisting in design specifications.

Optics

A three element, reflective optical system with a folding flat was designed to fit within a 4 inch circle as shown in Fig. 1. The radii and conics describing the mirror surfaces are shown in Fig. 1. A transparent window located at the entrance aperture which is not shown will be required to prevent contamination(scattering and absorption) of the surfaces of the optics and focal plane. The effective focal length of the system is 12.8 cm. Reflective optics were chosen to minimize chromatic aberrations that would result when using the wideband visible wavelengths that matches the spectral response of CMOS or CCD focal planes. The square aperture size 1.8 cm is somewhat larger than necessary to meet the resolution requirement at 50 m range. At 50 m the resolution is

0.18 cm at an MTF of 0.4. At the focal plane the resolution is 4.5 micron. However with a focal plane pixel size of 7.5 micron, the object plane resolution will be 0.3 cm. At object-to-camera distances above 83 m the resolution will be limited by the optics. The spot diagram indicates that over the central zone of 512 x 512 pixels, the performance is very good. The analysis was not performed for the full 2k x 2k pixels.

General Category	Components
Structural, dynamical, thermal	• Actuation--antennae, solar panels, shrouds, ordinance, covers, doors, tethers, latches • Structures--cracks, impact holes, broken cabling • Thermal flexing--solar panels, antennae • Vibration--thermal heating, gyro vibration • Cryogen leaks
Propulsion related	• Leaks of propellants • Plumbing malfunction--valves, joints, seals • Delamination, frozen--skirts, nozzles

Table 2. Observable phenomena.

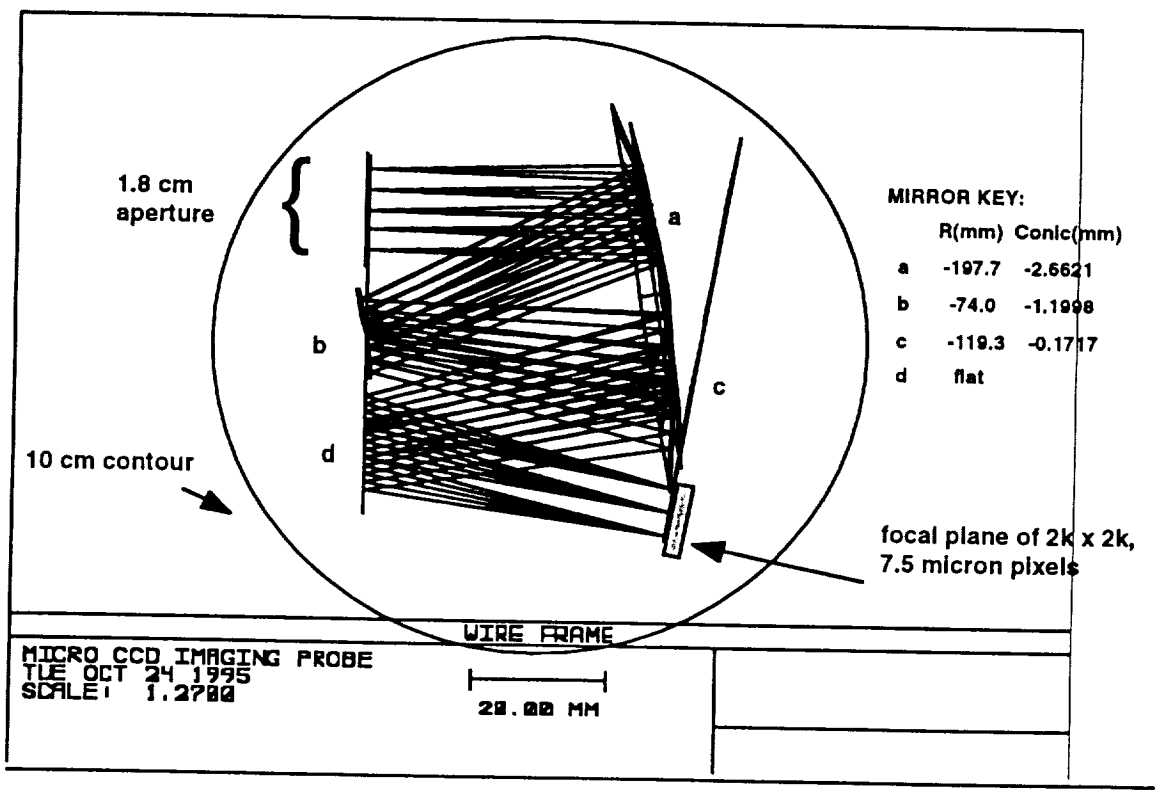


Fig. 1 Ray trace for a 1.8 cm aperture imaging optical system

Commercial video cameras are produced annually in very large numbers. It is expected that if a large number of UFOs are to be produced, standard manufacturing methods will suffice. Optics can be produced by conventional polishing or by diamond turning or some combination thereof. Assembly and alignment of the complete optical train using the reflective optics can also be accomplished using standard manufacturing techniques. Surface figures of approximately 1/8 wave at 0.5 microns, readily achievable for each surface will be necessary to maintain high imaging performance.

Angular slew and jitter pointing requirements can be inferred from the exposure time for a single frame. For either CCDs or CMOS focal planes and the optical system described previously, exposure durations of 50 ms will be necessary. During that exposure, assuming a criteria of motion no greater than 1/4 pixel, a stable pointing requirement of 3×10^{-4} radians can be derived. Shorter exposures times are possible if a faster imaging system consisting of larger and heavier optics can be tolerated to relax the pointing stability.

Focusing Mechanism

At the nominal range of 50 m the depth of field of the imaging system is 25 microns. For the object distance this translates to $\pm$ 12 m which is quite adequate. But the outer contour of the mothership is not shaped as the orbit of the UFO about the mothership. This causes the average focus to move as much as 0.3 mm as shown in Fig. 2. A focusing mechanism is needed. This need is meet in commercial video cameras with a focusing mechanism. Many commercial cameras employ a mechanism patented by Honeywell. The focusing travel distance can be readily spanned using a MEMS actuator. The translational adjustment will be optimized for small travel and have little effect on aberrations if the folding flat is translated. Positional accuracies better than 5 m in the object plane or 7.5 microns for the folding mirror are needed to avoid image degradation.

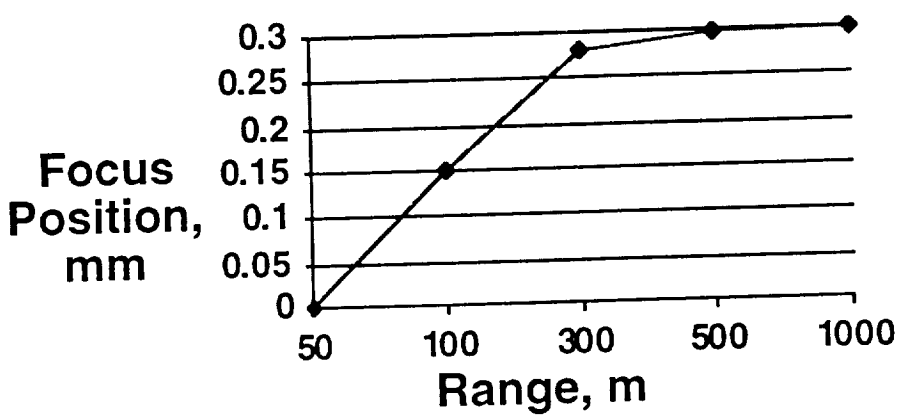


Fig. 2. Depth of field of imaging system as a function of camera imaging range.

Several mechanisms used for translational motion have been used in MEMS devices. These include a piezo drive, a comb drive, a diaphragm deflection, voice coil magnetic deflection, and a fluidic drive of a piston. Of these the comb drive and voice

coil approaches appear to meet the ~150 micron throw distance and would be easiest to implement.

Although an autofocus method is available(using an electric motor and gears) knowledge of range from the UFO to mothership could serve a useful purpose for maneuver or specialized imaging needs. Commercial laser range finders are available but may require modifications to reduce their present weight. For example Hamamatsu is apparently selling a rangefinder for $400 US and weighs 500 g. H. Saito reported on their efforts to develop a laser range sensor for planetary exploration as reported in this proceeding[6]. Using a beamsplitter, a laser diode as a source, an avalanche photodiode, and a electronics package to modulate and measure the phase shift of the return signal beam it is estimated that a scaled accuracy of 3 m will be possible for a range of 100 m. The laser power will be 0.3 w for a measurement time of 70 μs.

Shutter and coarse imaging camera

A conventional shutter mechanism will be needed to prevent the focal plane from accidental exposure to the sun. This control could be accomplished using a small camera with a large field of view, small aperture and low resolution. A fish eye lens will be sufficient. Additional attenuation or bandwidth filtering can be used to increase protection for the focal plane. This camera can be aligned with the higher resolution system described above to determine if the sun will be within its field of view . If desired this camera system will provide additional assistance during maneuver to reacquire the sun or the earth or the mothership to determine the orientation of the UFO. It will also assist the UFO to center the camera at the precise orientation towards the mothership to allow exposure of the higher resolution camera. With a 120 ° field of view, positioning accuracy to less than 0.2 ° may be easily achieved. Exposure control of the higher resolution camera will also be possible provided the camera responses have been calibrated against one another.

An alternate camera system was proposed using a compound eye(fly's eye) designed using a biological model of an insect eye. This system provides a very large field of view coverage that could readily cover 2 π steradians. Novel versions using diffractive lenslets or fiber optics coupled to an array sensor have been developed

Focal Planes

Using a 2k x 2k pixel focal plane, approximately 25 images will be needed to photograph the surface of the mothership. Use of a large focal plane minimizes the number of exposures and maneuvers that are necessary to efficiently cover the mothership's surface. A smaller number of maneuvers will also reduce the required propellant mass. There is no impact on the memory storage since the entire set of images will be stored to facilitate transfer of the data to a limited number of passes over available ground stations.

The IMEC camera is particularly well-suited for the UFO mission. Although the camera will not be available until 1996, it has a wide-dynamic range of 10^4, a large format array size, comparable sensitivity to a CCD, and has little blooming due to its direct current measurement feature. In addition it has a logarithmic response which reduces the data accuracy but compresses the data into a digital data size of only 7 bits. Normally 14 bits are needed to describe a value up to 10^4. For space application where large differences in reflected sunlight are anticipated this scheme is a clever solution to reduce data storage with no compression/decompression necessary.

Very recently a low noise CCD focal plane(KAF-168000) is now available from Eastman Kodak. It has a 4096 x4096 pixel array consisting of 9 μ x 9 μ pixels, a 100% fill factor, and a 76 dB high dynamic range. Although it is limited by the single slow read-out gate the addition of more read out gates could decrease the time to access the image. For this technology, improvements are possible to increase overall total dose hardening to 10 to 30 krads.

Camera Parameter	UFO Camera Specifi-cations	VVL Camera	ESPRIT3 Nr 7101 MINOSS	Prop'd JPL	JPL Digital Wire-less	IMEC FUGA 15	Develop mental IMEC focal plane
Spatial Res. (Pixels)	2k x 2k	312x287	128x128	128x128	245x256	512 x 512	2k x 2k
Frame Rate, f/s	30	25	25	30	30	10	10
Pixel Resolution at 50 m distance	0.3 cm	0.5 cm	1.2 cm	1.2 cm	1.0 cm	0.3 cm	0.3 cm
Max. Power, mw	100	200	50	50		<125	
Volume, Cu. inch		1.0		1.0		1.0	
Weight(g)		30(focal plane)	10(focal plane)	300		75	90(focal plane)
Techno-logy		CMOS	CMOS/ APS	MOS Photoga.	CMOS/ APS	CMOS	CMOS, direct read, 1og. res-ponse
Single Image Frame Memory, Mbits	64	0.9	0.17	0.17	0.63	2.	28
No. of Images	25	1.1k	6.1 k	6.1 k	1.6 k	382	25

Photogate - CCD like Capacitor Charge Storage.
APS- Active Pixel Sensor.
Gray Level Resolution = 10 bits per sample
VVL - VLSI Vision Limited Intelligent Micro Camera.

ESPRIT3 - Nr 7101 MINOSS (IRST, Italy), Also Model Nr 6490 TRIMOD is being developed at University of Sheffield, UK.
Digital Wireless Camera is based on JPL CMOS/APS technology under development and funded by ARPA.
FUGA 15 is available from IMEC, Leuven, Belgium
IMEC developmental focal plane will be available in 1996

Table. 3. Focal plane and camera survey.

A survey of camera/focal planes was reported by L. Hermans as shown in Table. 3. These include available cameras as well as cameras under development. In some instances the data for these cameras are not available.

The analogue signal output from the focal plane need to be processed beginning with A/D converter/data compression and merged with the angular and/or positional coordinates the UFO and mothership to enable stereoscopic reconstruction of the mothership. The data is stored to allow the data to be transmitted whenever favorable fly over communication link conditions exist to download the image from the UFO to a ground station.

Many cameras are available for the UFO mission. Phillips offers a 7000x9000 pixel array CCD focal plane. The development of these larger arrays is driven by commercial interests for the high definition TV applications. An issue for size scaling is the cost and yield of these devices. CMOS technologies are increasingly being used for low cost applications and are better suited for small pixel applications than CCDs. CMOS devices can operate at lower voltages to save power and are more easily integrated to A/D converters. For this UFO mission with power management, the energy budget will be small compared to other UFO subsystems.

Radiation Hardness

Radiation hardness requirements are set primarily by the orbital placement of the mothership, the length of time in orbit or use, and the weight allocation for radiation shielding to achieve total dose hardening. The anticipated radiation level of 10^{+5} rads was the estimated cumulative radiation dosage. Single event upset was also considered a significant problem if it occurs but for the orbit, it is thought to be very infrequent. Single event upset during storage was not considered a problem since the camera and other subsystems need not be turned on to avoid latch-up. For the long storage time before deployment, the simple and attractive strategy of shielding the storage housing for the UFO before its deployment was recommended.

Push Broom Scanning

An alternative was proposed to the focus mechanism. In this strategy a series of linear arrays would be used. Each array would be offset from one another but located at separate focal distances. From either inspection of the images or by using range finding, the appropriate array could be multiplexed into a A/D converter and the data stored in memory. This requires a relatively smooth and constant slewing across the object to obtain its image. It is conceivable that during a scan, the focusing conditions(distances) could change. Inputs to the A/D converter would then be switched from one array to another and the seam between them would be corrected by post processing of the images. This concept does offer a very efficient method to minimize the number of overlapping images to scan the mothership. It is conceivable that a set of arrays could easily be used to obtain the entire image of the mother ship in a single scan. Very large arrays of 7000 to 10000 pixels long are available from Japanese companies. Long integration times are possible for low intensity applications. However the scans should be accomplished rapidly compared to the relative translational motion of the UFO to the mothership to avoid on ground processing necessary to compensate the resulting image distortion from this motion during the scan. It will seem prudent to keep the laser range finder for this push broom scanning method to select the best image produced by the set of arrays.

Image Steering Option

G. Gal proposed a set of lens arrays to provide beam steering capability[5,7]. He indicated that image steering could be accomplished over a large angles by merely moving one lens array over a small distance. This steering array would be located at the entrance pupil of the telescope where the lights rays are essentially parallel. However the image at the focal plane is subject to scattering effects arising from each of the microlens structures in the lens array. There will be some degradation of the image quality and the MTF(modulation transfer function) of the lens array which will reduce image quality by approximately a factor of two. Despite this shortcoming it remains an attractive solution if image steering or stabilization is necessary.

Size, Power and Weight Estimate for the UFO

Estimates for the size, volume, weight, power, and energy for the UFO are reported in the Table 4. The volume is dominated primarily by the imaging optics. Weight is limited by the rangefinder(primarily electronics), the focal plane, and the telescope imager. Although the power is small, with proper power management, the overall payload energy requirement is trivially small compared to processing or communication requirements. The weight and power for the UFO is considered low risk.

Subsystem	Size (cm)	Volume (cc)	Weight (g)	Power (w)	Tot. time Use(s)	Energy (w-hr.)
High Res. Optics	6x5x2	60	50	NA		NA
Range Finder	3x2x2	18	100	1.5e-06	60	2.5e-08
Focus Control	2x2x1	4	15	.05	20	2.8e-04
Focal Plane and Electronic s	2x3x3	12	75	.125	200	.0069
Mech. Solar Shutter	2x2x0.2	16	15	.05	5	.00083
Wide angle imager	2x2x4	16	15	0.05	5	6.9e-05
Sub-Total		126	265	0.325		0.0081
Image Steering option	2x2x0.3	1.2	10	.1	30	0.00083

Table 4. Physical Characteristics of Imaging and Related Subsystems

References

1. DATA/PAL is available from D.A.T.A. Business Publishing, Englewood, CO 80155-6510.

2. Private communication, Robert Duncan, 1995.

3. C. Price and K. Grimm, "Sprint: The first flight demonstration of the external work system robots", Proceedings of the International Conference on Integrated Micro-Nanotechnology for Space Applications, Oct. 30-Nov. 3, 1995, Houston, TX

4. Jim Lyke, "Packaging Technologies for Space-Based Microsystems and Their Elements", in Microengineering Technology for Space Systems, ed. by H. Helvajian, 1995, Aerospace Corporation Report no. ATR-95(8168)-2.

5. G. Gal, Private communication, 1995.

6. I. Nakatani, H. Saito, T. Kubota, T. Mizuno, H. Kato, S. Nakamura, "Micro-scanning Laser Range Finder Sensor for Planetary Exploration", Proceedings of the International Conference on Integrated Micro-Nanotechnology for Space Applications, Oct. 30-Nov. 3, 1995, Houston, TX

7. G. Gal, "Diffractive and Miniaturized Optics", Critical Reviews of Optical Science and Technology, CR49, 329-359(1994).

Workshop 2: Software for Nanosatellites

Chairman: Michael M. Gorlick, The Aerospace Corporation

1 Introduction

From the perspective of a computer scientist a nanosatellite is a (physically) small distributed computing platform with exotic peripherals. Are these devices just "more of the same" in the sense that the main line disciplines of real-time systems, software architecture, and software engineering will suffice, or do they represent a qualitative change that challenges computer system engineering? In the past significant alterations in the structure or performance of computing systems precipitated the development of new software architectures and algorithms. Does the wafer scale integration of computation with electromechanical devices represent the next wave of qualitative change?

We endeavored to address these issues in the context of the UFO (Untethered Flying Observer), a solid silicon nanostatellite assembled from a wide variety of MEMS devices. We envision that a UFO is mass produced using production techniques comparable to those for the manufacture of large scale integrated circuits and consumer electronics. It contains a mix of standard subsystems common to all nanosatellites (power management, propulsion, guidance, and the like) and mission-specific elements such as cameras or specialized sensors.

The goal of the workshop was to identify, from a software perspective, the critical issues in the design, deployment and use of these devices and produce an outline of a research and engineering agenda that addresses these issues. Topics of discussion included:

- Rough order of magnitude estimates for the amount of software required for a UFO and the computational resources that it will require.
- The degree to which a UFO can function autonomously.
- Software architectures suitable for a broad mix of standard subsystems plus mission-specific elements.
- Cooperation within swarms of nanosatellites.
- Dynamic reconfiguration to cope with changes in mission.
- Real-time requirements for critical subsystems such as guidance or attitude control.
- Algorithms for the control of nanosatellites.

The workshop participants were:

Ronald Arkin	Georgia Tech
Don Batory	U. Texas, Austin
Tim Cleghorn	NASA
Robert Davis	NASA
Tom Diegelman	NASA
Michael Gorlick	Aerospace
Hank Johansen	NASA (retired)
Robert Mandelbaum	U. Pennsylvania
Robert Savely	NASA
Will Stackhouse	Consultant
Kevin Sullivan	U. Virginia

The workshop concluded that the software for the UFO is feasible; however, constructing the software in a flexible and economical manner is a significant software engineering challenge that will require substantial forward progress on a number of research and engineering issues. More generally, nanosatellites are merely one example of the revolution in software engineering that will be precipitated by micro- and nanotechnology.

We attempted to address both the general issues and the UFO-specific ones and that effort is reflected in the organization of this report. Section 2 presents a broad overview of the software issues for nanosatellites in the context of the impact of anticipated advances in micromachining and digital electronics while Section 3 focuses on the much narrower issues of producing the software for the first prototypical UFO. Section 4 forecasts software developments over the next twenty years as we learn to control and harness swarms of nanosatellites for a variety of missions. Finally, Section 5 summarizes the basic issues that we face for the near- and medium-term development of nanosatellite software.

2 Major Issues

The technology to manufacture a solid silicon nanosatellite is almost within grasp and we can see a future where nanosatellites are manufactured in high volumes and low cost just as disk drives, VCRs and laptops are today. It is not hard to imagine a manufacturing technology that would allow a nanosatellite to be fabricated from a mixed collection of stock and custom component parts, for example, a custom sensor married to stock "nanosatellite bus." It is unthinkable that the onboard software for such a device be handcrafted anew for each member of the nanosatellite family. Ideally it would be wholly generated from high level descriptions and models that accompany the stock parts (just as a specification sheet is now packaged with an electronic component). At worst it is a mix of (predominantly) stock code with minor custom additions or modifications.

Software reliability for these devices is a critical issue. No mission looks forward to discovering on orbit a crippling bug in the software of the nanosatellite bus. Furthermore, satellite manufacturers, integrators, and end users must be assured that bugs appearing in custom software can not give rise to a propagating wave of faults that incapacitates the craft. To the extent that nanosatellites are assembled from a "menu" of subsystems and components, the software will be increasingly diverse with a strong emphasis on plug and play. Those implementing custom subsystems should be confident that if they conform to the interface and behavioral constraints then at worst the spacecraft will fail safely.

Achieving this degree of reliability and flexibility will require an integrated set of software tools of a high order, and software reuse will be a fundamental strategic goal. Research has demonstrated that large scale comprehensive software reuse can not be done manually. It will require automated support at all steps in the software lifecycle. The ARPA Domain-Specific Software Architectures Program [4, 8] illustrates the sweeping range of tools and models required to adequately support a domain. The challenge for nanosatellite software is in the same vein since the only rational path is massive component-based reuse.

However, nanosatellite software is only one element of a larger pattern. The history of computer science is demarcated by a series of dramatic shifts in the technology of computation. In each case that shift precipitated a major realignment as computer scientists were forced to rethink the structure of systems as their prior assumptions were invalidated. For example, the extraordinary improvement in the price/performance ratio of commodity microprocessors and the dominance of network bandwidth over memory speeds are two fundamental qualitative shifts among many over the past fifteen years.

We are on the verge of yet another realignment that arises from the confluence of three powerful forces:

- the extraordinary economic pressure for the development of low power digital electronics (cellular telephones, laptops, portable games, and the like);
- the inexorable forward march of Moore's Law will continue to assure a doubling of the price/performance ratio of digital electronics every eighteen months for well into the next century; and
- the rapid rise of micromachines and their physical integration with digital computation.

To appreciate the impact of the coming change count the number of devices that are controlled by the average desktop personal computer. They number well under a dozen: keyboard, mouse, screen, modem, one or more hard drives, a CD-ROM player, printer, and perhaps a scanner. Even in domains rich with sensors and controls (such as process control or automated manufacturing) the number of sensors and actuators is on the order of hundreds. The intimate integration of micromachines with digital controllers means that computational elements may have thousands of microdevices under their control in a space possibly no larger than a sugar cube. In addition these constructions will have significant amounts of processing power and memory (on the order of MIPS and megabytes).

The three simultaneous trends of more densely integrated low power commodity architectures, hybrid architectures of processors and micromachines, and the merging of hardware and software in the form of ROM, programmable gate arrays, and custom digital circuitry (generated directly from software descriptions) will present novel software challenges. The evolutionary trend has been to move computation closer and closer to "where the action is" (as we migrate from mainframe to desktop to portable). Micromachining is the next step in that evolution as computation is intimately commingled with sensors and actuators in a wide variety of fixed and portable devices.

Since micromachines and their controllers will be ubiquitous, software will assume a primary role. Consequently issues of reliability, integration, ease of construction, and ease of evolution will come to dominate.

These micromachines and their digital controllers will be the "golden screws" of engineering — appearing as piece parts or subsystems in larger, more complex devices.

It is only recently that software has become a commodity item. While many software systems are hand-crafted, one-of-a-kind, the commercial marketplace is increasingly populated by shrinkwrap software. The introduction of a variety of integrative technologies (OLE, CORBA, Java) suggest the possibility of commodity applets that are specifically designed to seamlessly integrate as piece parts into a much larger whole.

At this point it is quite unclear what tools, languages, or integrative technologies are required and in what combination to achieve this goal. The commodity worlds of machine parts or electronic components are predicated on a "degree of simplicity" and methods of uniform description that seem difficult to achieve for software.

3 Software for the Untethered Flying Observer

We now turn to the much narrower problem of developing and implementing the software for a first generation UFO (Untethered Flying Observer). It was the consensus of the attendees that the onboard software for a UFO did not present any overwhelming obstacles however, there were a number of technical and programmatic concerns.

3.1 Algorithmic Concerns

In order for the mission to be successful close integration is required between the spacecraft sensors and two vital functions: localization and pointing. Localization is the ability of the UFO to determine its location relative to the mothership, while pointing is the ability of the UFO to correctly orient its camera in the right direction at the proper time. Both of these capabilities have real-time constraints and are critical to the mission. Matching localization algorithms to the capabilities of the UFO sensors can be troublesome and will require close cooperation between algorithmists and sensor designers. Similarly the pointing maneuvers required for the photographic survey of the mothercraft dictates that close attention be paid to the integration of sensors and processing.

The constraints on weight and size require extremely aggressive power management during all phases of the mission. Given its pervasiveness the working group felt that power management should be consolidated as a service layer in the system architecture rather than scattered among the various subsystems. The specification and design of this layer presents some interesting problems in descriptive techniques and automated software generation.

Finally, it is tempting to employ image processing in the aid of navigation as the UFO conducts its survey. Simple linear CCD algorithms would have low computational demand and could provide useful navigational data. Coarse digitization, obtained by subsampling the image, could be useful. Nonetheless, for the first generation UFO, the working group feared that the computational demands of sophisticated image processing, scene recognition, and machine vision were far more than could be reasonably supported by the computational devices that we expect to find on a UFO deployed over the next five years.

However, in the single case where the UFO loses all "lock" on the mothership and must conduct a general search to reacquire attitude it is feasible to employ some simple, well known image processing techniques that would permit the UFO to detect the mothership against a star field or backlit by earthglow. No doubt there are other situations in which simple image processing may be of benefit. We expect that future generations of UFOs will have sufficient computational capability to use machine vision as a navigational aid.

3.2 Software System Architecture

The software system architecture for a UFO is sketched in Figure 1. In general it is a classic layered architecture that bears a strong resemblance to the form of modern avionics systems particularly helicopter avionics. Boxes represent basic subsystems; physical adjacency denotes direct interactions (for example, the Power Management subsystem communicates directly with the Navigation, and Safety Kernel subsystems while the Safety Kernel is the only layer that communicates directly with sensors $S_1 \ldots S_n$ and actuators $A_1, \ldots, A_m$).

Sensors provide raw data to higher level system components. Navigational sensors, for example, report raw data on the position, attitude, and velocity of the spacecraft. The navigation layer refines these data values into best-guess estimates of the true position, attitude, and velocity of the craft. The guidance layer knows the destination of the UFO and uses the estimates from navigation to create a flight path that will take the spacecraft from its current position to its destination position. The autopilot subsystem uses this flight path to plan, initiate and control thruster

firings so that the spacecraft traverses the projected flight-path at the proper rate and with the correct orientation. The mission subsystem monitors external events and the status of the spacecraft to alter mission objectives and final flight positions.

Two of the layers, power management and the safety kernel, deserve special mention. The power management layer provides power management services to the craft as a whole and is responsible for power apportionment and timing to all electronic and electromechanical subsystems. For example, during the survey portion of the mission the power requirements of the camera are so high that all other subsystems except for critical maneuvering and monitoring functions are completely shutdown. In addition since power management is pervasive it is collected together in a single layer that supports all of the functional behavior of the nanosatellite.

There are three basic safety issues that arise in the UFO mission. First and foremost is collision avoidance with the mothercraft since the UFO has sufficient mass to damage the antennas, solar panels, or instrument booms of the mothership. Secondly, the only suitable propellant, given the mission profile and limitations on spacecraft volume and weight, is liquid ammonia which may corrode or foul the mothercraft. Consequently, the nanosatellite must always stand off at a safe distance (approximately 50 meters) from the mothership. Finally, when its mission is complete a UFO's fuel load is nearly spent and it becomes a navigational hazard in the orbit of the mothercraft. Therefore, once its survey is complete and the images have been downlinked, the UFO must deorbit where it will incinerate in the Earth's atmosphere upon reentry.

One promising approach to software safety for complex systems is the use of a safety kernel [17, 16] that is responsible for tracking the system state and ensuring that the system never transitions to an unsafe state. Because it is infeasible to verify large bodies of code, it is infeasible to verify system safety if it depends on large amounts of code. By localizing safety-related issues in a safety kernel whose size is small relative to that of the overall system, one restricts the needed verification to a relatively small amount of code. Verifying the safety of all of the code is thus reduced to verifying the safety kernel. Evidence to date suggest that safety kernel software architectures can contribute significantly to system safety. However, ultra-high assurance software remains an elusive goal.

All control of the nanosatellite actuators (microthrusters and a single solid thruster for the final deorbit burn) is mediated by the safety kernel. Thus

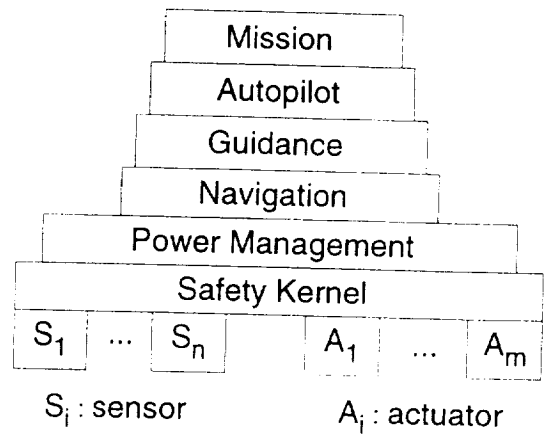


Figure 1: A generic software architecture for a UFO

the safety kernel will block any maneuver that would bring the nanosatellite too close to the mothership or exceed other safety limits (for example, maximum speed during maneuvers or rate of propellant expenditure). In addition the safety kernel will guarantee that the nanosatellite is properly oriented for deorbit prior to ignition of the solid booster.

Software for specific nano-satellite missions will be an instantiation of Figure 1. Subsystems $S_1 \ldots S_n$ will be replaced by specific sensor subsystems; $A_1 \ldots A_m$ will be replaced by specific actuators. Similarly, the safety kernel, power management, guidance, and other subsystems will be replaced with customized versions that are designed specifically to the target nanosatellite mission.

We believe that the process of customizing nanosatellite subsystems will be accomplished by composing prefabricated components (building blocks). Compositions of these primitives will define customized power management, navigation, guidance, and other subsystems, much like the way avionics software is specified, customized, and built in ADAGE [4]. Figure 2 shows an expansion of Figure 1 to reveal the structure of nano-satellite subsystems as component compositions. To illustrate, sensor subsystem S_1 is a composition of three components while subsystem S_n is defined by a single component.

Our general approach is to follow the work of ADAGE by defining component libraries (building blocks) of nanosatellite software and show how they can be composed to form customized, efficient, mission-specific systems. Weaves [7] are one form of software integration that can be used to combine ADAGE-like components into a highly flexible, dy-

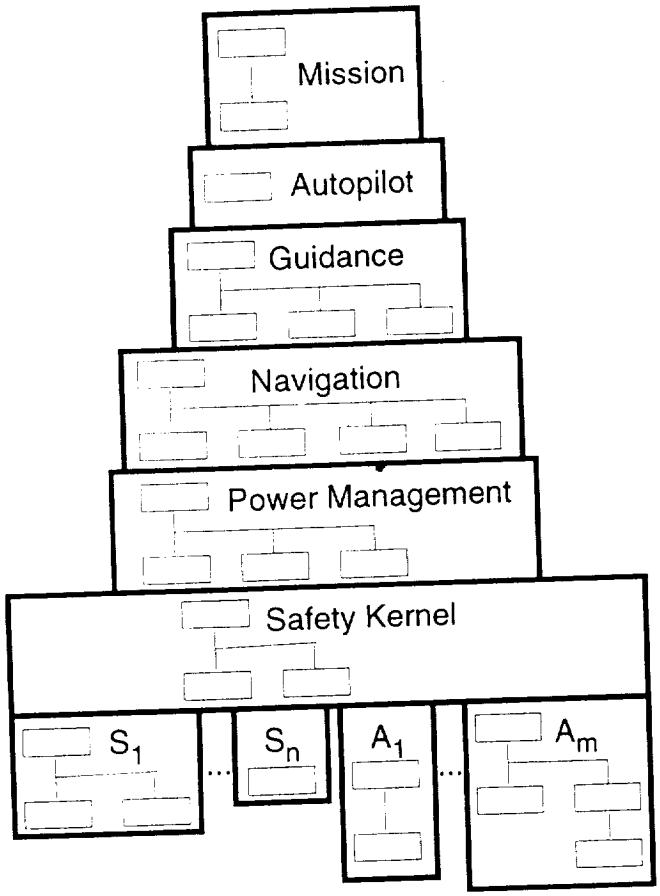


Figure 2: A specific software architecture for a UFO

namically reconfigurable whole. The end result will be an ability to produce customized software at a fraction of the time, man-power, and cost it currently takes to build such software.

3.3 Research Areas

The workshop identified several research areas that are relevant to software systems for nanosatellites. The general problem of easily constructing software systems for large families of nanosatellites poses profound problems of software engineering and software integration that are poorly understood at this time. The development of the software for the UFO is a golden opportunity to address these problems head-on in a way that will have lasting benefits.

Particularly promising are "Batory-style" generators [3] that use a combination of domain modeling and component-based software construction techniques to dramatically reduce the time and effort required to produce custom software systems. One of the critical elements required for the application of this form of software technology is a deep understanding of the domain and the likely variation among members of the family of nanosatellite software systems. In other words this approach is successful only if domain engineers can establish one or more base configurations from which families of extended configurations are derived. One positive sign is that there is a large body of knowledge about the various subsystems (power management, guidance, navigation and control, attitude) that can be used to construct subgenerators for each major subsystem. The real challenge lies in integrating those subgenerators into a cohesive whole. Knowing "what is likely to change" will be the key to a clean and extensible software design.

Integrative architectures were another area identified by the workshop that may have significant positive impact. This work comes in a variety of flavors including the abstract behavioral types and mediators of Sullivan [14, 13], the weaves architecture of Gorlick [6], and the hybrid robotic architectures now under investigation [9, 1, 5]. The workshop repeatedly identified the major issues as architecture and integration rather than algorithms and opined that the successful exploitation of nanosatellites rested more on success in software architectures than on any other single issue.

Given the power constraints these small devices face computational services must be mindful of power budgets. Recent work in low-power computing has begun to address the problem of code generation for high-level programming languages that optimizes for power consumption [15]. In other words, the code sequence generated computes the specified function but in such a manner as to minimize power consumption. Optimizers of this sort may be a powerful low-level tool in the arsenal of the nanosatellite software engineer.

Finally, the workshop identified software safety as a critical essential technology. If the software is to be assembled automatically from high-level specifications then users must have assurance that the nanosatellite will not endanger the mothercraft. If, as we predict for the future, swarms of nanosatellites are used for space missions, then the safety concerns will multiply accordingly. The interplay among software architectures, automated software generation, and software safety is complex and its resolution may lead to significant advances in the development of safe and reliable software for a enormous range of devices that rely on embedded computers and embedded micromachines.

3.4 Programmatic Concerns

In addition to the technical issues the workshop members also commented on the process of developing the nanosatellite software. Two concerns were raised. First, it is tempting to regard the nanosatellite software as just another development item rather than a research issue in its own right like microthrusters or space-qualified packaging for micromachined components. It is well within the reach of current software technology to design, develop, and deploy a "one-of" software system that satisfies the mission requirements of the UFO. However, we feel that it would be a grave mistake to ignore the software research issues outlined above in favor of a development plan that narrowly focuses solely on the software itself and not on the technologies or processes that were used to develop it. The software engineering technology that allows us to economically and conveniently produce reliable and flexible software for large families of nanosatellites will be of immense value and can be broadly applied in an enormous number of domains. Substantial progress in this area would have profound implications for our economic well being and military capabilities.

Secondly, we strongly recommend that the nanosatellite development team employ the process of concurrent engineering. Historically, spacecraft subsystem designers almost always succumb to the temptation of leaving some difficult or messy element to the "software jockies." Given the intimate connection between micromachines and computation and the extraordinary dependence of the nanosatellite on a wide variety of computational services it would be a grave error to allow any subsystem design and implementation to proceed without the close cooperation of digital designers, computer scientists, and software engineers.

In particular we recommend that each subsystem team include at least one software specialist throughout the development of the UFO. The hardware designs should not impose unnecessary burdens on achieving software reuse and modularity. Ill-considered hardware designs often make modularity, interchangability, and reuse difficult if not impossible to achieve. This situation often arises in commercial and military avionic systems where the failure to employ concurrent engineering leads to software and systems that are unique and can not be reused in related craft.

4 Future Technology

Design families of nanosatellites sharing common stock subsystems is now within the grasp of current software technology. Recent developments in robotics point the way towards a more ambitious goal of societies of loosely coupled nanocraft [2, 10, 11]. The individual members of such societies would engage in a variety of cooperative behaviors. For example, the UFO mission might be carried out by several nanocraft simultaneously working together in concert. Cooperative behavior has numerous advantages. In this context multiple UFOs could conduct a complete photographic survey more quickly than one craft if each UFO concentrates on only a portion of the mothercraft. The UFOs could deliver multiple perspectives to analysts, for example, a three-dimensional view could be reconstructed on the ground from the views contributed by two or more nanosatellites. It would be possible to conduct multiple missions simultaneously such as both a general global survey and a specialized local survey. A small society can be more robust than a single individual for if one member of the society fails for any reason then the burden of its mission can be shared among the remaining nanocraft.

The exploitation of groups of nanosatellites is closely coupled to volume manufacturing of these devices. Volume manufacturing, assuming that the software issues outlined earlier can be resolved, has numerous advantages. First, over a large manufacturing run it is possible to amortize development and design costs and as the manufacturer scales the learning curve the per unit cost can drop dramatically. To appreciate the economies of scale just consider the decreases in price over the years of complex consumer electronics such as VCRs, CD players, cellular telephones, and the like. In addition high volume manufacturing permits continuous evolution and improvement of the product. Given constant output it is comparatively inexpensive to incorporate modest changes or improvements in the device. Notice also that an ongoing volume production line will lead to dramatically reduced delivery times with all of the accompanying program savings. Finally nanosatellites can be launched in batches of tens to hundreds using significantly smaller launch vehicles. Consequently devices of this sort will also enjoy substantially lower launch and replenishment costs.

4.1 Research Themes

If societies of cooperating nanocraft are to become a reality then research must be directed toward societal and integrative architectures that will bind together

multiple craft into a cohesive, purposeful whole. Important research issues include the following:

- the number of individual agents (nanosatellites) required to cooperate to accomplish a task in a given amount of time. Three UFOs might complete the survey in a third the time it would take one UFO but would twenty UFOs accomplish it in a twentieth of the time?

- the degree, form, and frequency of interagent communication required for the accomplishment of a mission. If two or more UFOs cooperate in a survey what must they communicate to one another and how often?

- the degree to which agents are homogeneous or heterogeneous. For the UFO mission it might be advantageous to have a nanosatellite devoted to communications, that is, instead of cameras it carries larger antennas, a more powerful transmitter and a more sensitive receiver. Its job is to act as a tiny relay station for the UFO conducting the photographic survey. A specialized "communications nanosatellite" could achieve higher downlink rates and may be able to stay in constant communication with the ground. This form of specialization (or heterogeneity) might allow ground controllers to significantly extend or modify the UFO mission as need arises.

- the form of the task specification. Detailed reprogramming of a nanosatellite for each new mission is tedious and error-prone and the job becomes much more complicated when trying to coordinate a group of mobile robots. Therefore high-level task specification languages will be required for detailing the mission and its contingencies.

- the longevity (or lifetime) of agents. For long duration missions it is unclear whether the mission is better served by multiple generations of short-lived nanosatellites or by comparatively few long-lived nanocraft. Agent lifetime will be an important parameter in the optimization of mission performance.

- replenishment strategies. Related to the notion of longevity is the rate of replenishment required for various missions. As nanosatellite technology improves it may be of benefit to replace still unused UFOs with newer, more capable models. For other longterm missions it may be necessary to regularly restock the society.

- the adaptivity of agents. Related to the issues of homo- and hetero-geneity is the degree of adaptivity of individual members of the group. Not only is this an issue of societal control but it is also an issue of complexity and economics. Highly adaptable and general devices may be significantly more complicated and expensive to build than comparatively simpler, more specialized nanosatellites. If appropriate societal controls can be imposed then it may be more cost-effective to conduct a mission with large numbers of specialized members than with a few general, adaptable individuals.

All of these research themes speak to the larger issue of *societal design*, the deliberate engineering of individuals and their rules of interactions to produce a cooperative society that collectively accomplishes a given task. Thus societal design would strive to exploit the dynamics of local interactions between agents and the world in order to create complex global behaviors.

4.2 Five Year Projection

Over the next five years we expect to see a number of significant developments in the software engineering of robotic systems. Indicative of the trend is the research conducted under ARPA sponsorship of autonomous land vehicles that led to the development of systems capable of driving and navigating a motor vehicle on- and off-road without human intervention. This, and related work, will lead the way toward sophisticated reactive systems that, for example, can undertake safely and autonomously a "nap of the earth" tour of a mothercraft suitable for a highly detailed closeup view (see for example [12]). In addition we expect to see the marriage of conventional and reactive systems in a hybrid architecture that combines the best properties of both. This architectural concept is illustrated in Figure 3 where the reactive layer is responsible for autonomous navigation and mobile behaviors while the deliberative layer is responsible for mission planning, directed execution, and other higher-order functions. Finally, we expect to evolve control laws and specification methods adequate for small societies of less than ten agents.

4.3 Ten Year Projection

At the ten year mark we anticipate the development of hierarchical "multicaste" societies comprised of three to four castes where each caste is itself a small society of ten agents or less. This concept is

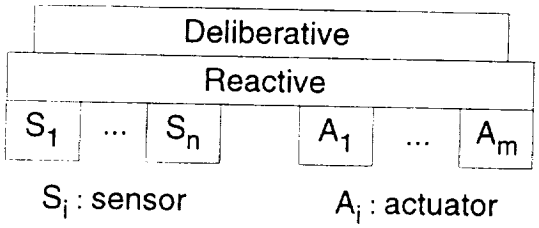


Figure 3: A hybrid architecture containing both reactive and deliberative layers.

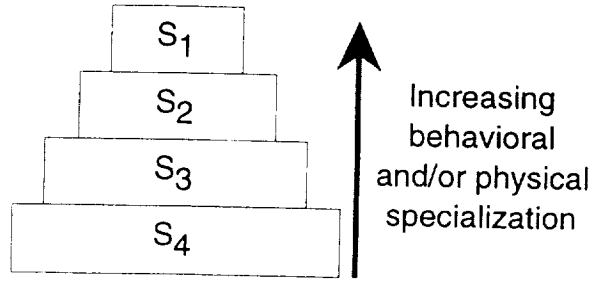


Figure 4: A mission architecture formed from a small multicaste society.

diagrammed in Figure 4 which illustrates a multicaste society of four levels with a total population of twenty to thirty agents. Each caste S_i would have significantly fewer members than caste S_{i+1}. These societies would be moderately heterogeneous with the members of any one caste S_i either identical or largely similar with only modest variations. However, there may be significant differences between the members of two adjacent castes S_i and S_{i+1}. In all likelihood each caste may have significant behavioral and/or physical specialization.

Agents in the higher levels of the society (S_1, S_2) are "viewers" or "brains" while the population of the lower levels (S_3, S_4) are "doers" or "brawn." Each layer is capable of independent behavior. However, the range of behaviors of the lower layers may be sharply restricted, either by design, or by directives issued by the higher layers of the multicaste society. Small hierarchical societies of this form could perform a wide variety of useful tasks in space such as: the automated assembly of larger structures (including the on-orbit assembly of conventional satellites), the maintenance, repair, and inspection of existing space assets, and the disassembly, disposal or garbage collection of obsolete or failed assets. The elimination of space debris in popular orbits may be an important mission for collective societies of this form.

An alternative likely development is the evolution of specialized agents that act as negotiators and mediators between two small independent societies. Nanosatellite development will, over the short term, evolve into the creation of small homogeneous societies and mission planners will be strongly motivated to take advantage of the capabilities of pre-existing societies. This will lead, in a natural way, to the introduction of negotiators that are capable of directing and coordinating the activities of two more more small societies to accomplish a given mission.

4.4 Twenty Year Projection

Over a twenty year period we predict the development of large scale societies (on the order of 10^3 members) and sub-societies. "Societal engineering" will be a recognized engineering discipline combining elements of computer science, software engineering, biology, anthropology, sociology, and communications. While it is impossible to predict the detailed form of these societies they will be heterogeneous, highly distributed, and capable of inter-society cooperation.

The creation and deployment of these large-scale societies will revolutionize the exploration and exploitation of space. At the very least these societies will form space-based "tapestries" of sensing and communication that will dwarf our present capabilities in terms of coverage, resolution, and bandwidth. Likely applications are: space salvage and large-scale manufacturing (including the manufacture of additional nanosatellites in space), space and planetary exploration such as robotic insect-like rovers on Mars and the moon, large cellular or mesh-like structures of nanosatellites for detailed observations and communications, webs of sentinels to provide early warning of meteorites or solar flares, and a massive deep space communications, sensing and scientific network.

5 Conclusions

There are no technical "showstoppers" that will prevent the development of the software for the first generation UFO. However, there are a significant number of software issues that must be addressed if large, reusable families of nanosatellites are to become a reality:

- It is not economically feasible to handcraft the software for each new family of nanosatellites. In-

stead, it must be composed of pre-existing software components using a variety of generative and integrative technologies. To the extent that specialization is required it must be minimized and contained through the use of standard frameworks and interfaces;

- It will be desirable to upgrade software capabilities dynamically (that is, post deployment). For example, it makes little sense to inflexibly fix on a compression algorithm today for a mission lifetime of five to ten years when the algorithm will be obsolete in a year or two. The ability to incrementally update the nanosatellite software on orbit will be invaluable in extending the useful lifetime of these assets. The capability to recast the software dynamically must be designed into these systems from the beginning;

- Nanosatellite hardware will evolve into families with distinct capabilities, each supporting a variety of missions. The software technologies that we adopt should demonstrate the potential, over the long run, of keeping pace with the anticipated degree of evolutionary change and variability; and

- The software community has little experience building reusable architectures for families whose initial systems have not yet been built, thereby exacerbating the problem of constructing suitable domain models. The modeling efforts should begin immediately however, there may be exemplars of like domains (such as avionics) from which we can learn.

The resolution of these issues, in a general setting, will have profound implications for the design, development, and deployment of large scale software systems in a number of domains including education, manufacturing, entertainment, and defense.

6 Acknowledgements

I am indebted to Ron Arkin, Don Batory, Robert Mandelbaum, and Kevin Sullivan for their contributions of prose, diagrams, and references to this report. All errors, omissions or misrepresentations are solely my responsibility and occurred in spite of the best efforts of my colleagues.

References

[1] Ronald C. Arkin. Integrating behavioral, perceptual, and world knowledge in reactive navigation. *Robotics and Autonomous Systems*, 6:105–122, 1990b.

[2] T. Balch and R. C. Arkin. Motor schema-based formation control for multiagent robot teams. In *Proceedings 1995 International Conference on Multiagent Systems*, pages 10–16, 1995.

[3] D. Batory and B.J. Geraci. Validating component compositions in software system generators. In *Proceedings of the International Conference on Software Reuse*, Orlando, Florida, 1996. To appear.

[4] L. Coglianese and R. Szymanski. DSSA-ADAGE: An environment for architecture-based avionics development. In *Proceedings Advisory Group for Aerospace Research and Development*, 1993.

[5] E. Gat. Integrating planning and reaction in a heterogeneous asynchronous architecture for controlling real-world mobile robots. In *Proceedings of the AAAI*, 1992.

[6] Michael Gorlick and Alex Quilici. Visual programming in the large versus visual programming in the small. In *Proceedings of the 1994 IEEE Symposium on Visual Languages*, St. Louis, Missouri, October 1994. IEEE Computer Society Press.

[7] Michael M. Gorlick and Rami R. Razouk. Using weaves for software construction and analysis. In *Proceedings of the 13th International Conference on Software Engineering*, pages 23–34, Austin, Texas, May 1991. IEEE Computer Society Press.

[8] M. Graham and E. Mettala. The domain-specific software architecture program. In *Proceedings of DARPA Software Technology Conference*, 1992. also appears in Crosstalk: The Journal of Defense Software Engineering, October 1992.

[9] R. Mandelbaum. *Sensor Processing for Mobile Robot Localization, Exploration and Navigation.* PhD thesis, University of Pennsylvania, 1995.

[10] Maja J. Mataric. Designing and understanding adaptive group behavior. *Adaptive Behavior*, 4(1), December 1995.

[11] F. Noreils. Coordinated protocols: An approach to formalize coordination between mobile robots. In *Proceedings 1992 IEEE International Conference on Intelligent Robots and Systems*, pages 717–725, 1993.

[12] T. M. Rao and R. C. Arkin. 3d navigational path planning. *Robotica*, 8:195–205, 1990.

[13] Kevin J. Sullivan. *Mediators: Easing the Design and Evolution of Integrated Systems*. PhD thesis, University of Washington Department of Computer Science and Engineering, 1994. available as Technical Report UW-CSE-94-08-01, August, 1994.

[14] Kevin J. Sullivan and David Notkin. Reconciling environment integration and software evolution. *ACM Transactions on Software Engineering and Methodology*, 1(3):229–268, July 1992.

[15] V. Tiwari, S. Malik, and A. Wolfe. Compilation techniques for low energy: an overview. In *Proceedings of the 1994 IEEE Symposium on Low Power Electronics*, October 1994.

[16] Kevin G. Wika and John C. Knight. A safety kernel architecture. UVA-CS-TR 94-04, University of Virgina, Department of Computer Science, 1994.

[17] Kevin G. Wika and John C. Knight. On the enforcement of software safety policies. In *Proceedings of the 10th Annual IEEE Conference on Computer Assurance*, June 1995.

Workshop 3: Low Power Communications

Co-Chairmen: John Hurrell and Walter Bloss, The Aerospace Corporation

I. Introduction: (subject, participants, and focus)

This workshop considered the communication requirements for crosslink and downlink nanosatellite communications and to what extent near term technology will be able to support these missions. The least challenging communication scenario is merely a crosslink either to a nearby mother ship which acts as a relay to transmit data to the ground or to another local satellite in a constellation. This scenario is attractive for two reasons. First, the free space loss $(\lambda/4\pi R)^2$ is minimized by a short range R, and second, the intended receiver is essentially always available. Both benefits are absent if the nanosatellite communicates directly to the ground. Link budgets and appropriate technology requirements for these scenarios were established. Current technology capabilities and future developments to support such missions were reviewed.

Members of the workshop were:

Charles Barnes	JPL
Wayne Fenner	Aerospace
George Haddad	University of Michigan
David Ksienski	Aerospace
Clark Nguyen	University of Michigan
Gary Rummelsburg	Hughes
John West	JPL
Peter Zdebel	Motorola

II. Technology Assessment for Space Systems

A. Near Term (1 to 5 years):

1. Current State-of-the-Art

Several chip sets have been developed for the Global System for Mobile (GSM) Communications that illustrate the current performance of commercial, low power 1 GHz transceiver technology. Transceivers built in silicon BJT technology operate at 2.7 V and consume about 60 mA in the transmitter channel and 50 mA in the receiver channel. The transceiver chip is preceded by a low noise amplifier (LNA) and followed by a small power amplifier. Some off-chip components, such as a master oscillator and rf filters, are also used to complete the rf system. Similar power budgets are achieved with 3 V biCMOS chips.

Personal Communication Services (PCS) are being developed that will use transceivers operating near 2 GHz. Single chip transceivers have yet to be demonstrated. Gallium arsenide technology will probably provide the lowest power systems.

2. Attractive Opportunities for Space

Low power, short hop crosslinks could be developed with a radiation tolerant version of the GSM type transceiver. Gallium arsenide monolithic microwave integrated circuits (MMICs) for power and low noise amplifiers have already been developed for space-based radar applications. Some development would be required to optimize efficiency for operation at 1 GHz.

Low power downlinks require a simple miniaturized SGLS (space ground link subsystem) transceiver. The receiver would only need to decode a few uplink commands and no range information would be required, greatly simplifying the digital signal processor. There exists a large commonality between SGLS and PCS requirements that suggest that a SGLS transceiver module could be produced with low power PCS technology. The data rate that could be supported on the link would be limited by available transmitter power. Fifty percent power added efficiency would be a good goal for the power MMIC chip.

3. Technology Development Issues

In order to minimize costs and amortize NRE over a large number of units, it is essential to develop chip sets that have broad application and are based on proven commercial technology. The wireless communication industry will provide ground-based systems. The space community needs to make this technology survivable and reliable for space.

B. Far Term (5 to 20+ years)

During the next 10 years, the semiconductor industry will encompass lower power approaches for digital electronics. The roadmap to achieve this improvement is already in place. The space community needs to track this development and learn whether it can be made survivable in space. This development in technology will impact both DSP and transceiver power budgets. For example, biCMOS will evolve to 1.5 V battery operation with 4x less power consumption and 2x speed increase. Complementary heterojunction fet and heterojunction bipolar transistor technologies in gallium arsenide will provide competing capabilities that may be directly applicable to space application. Also, resonant tunneling diodes (RTD) are easily implemented in gallium arsenide type technology and promise to decrease power consumption in signal processing functions even further.

The MMIC technology for low noise and power microwave amplifiers is unlikely to change much in the long term. However, increasing use of lighter weight packages and more automated production of rf modules will occur.

III. UFO Subsystem Summary:

The strawman UFO mission we considered involved the collection of 1 Gbit of image data, gathered during a 48-hour free flight around a mother ship orbiting the earth at 800 Km, and the subsequent dissemination of this data to the ground. We considered both a direct downlink and a crosslink relayed to the ground from the mother ship.

A. Recommendations:

A 2 Km crosslink can be closed with a transmitter power of a few milliwatts for data rates below a megahertz. The frequency should lie near the GSM communication band at 950 MHz to

take advantage of the commercial investment in wireless telephone technology. At this frequency, it should be possible to develop a complete communication subsystem that may dissipate less than 150 milliwatt and support two simultaneous channels to ensure uninterrupted communications. The receiver would operate continuously and the transmitter would be used intermittently to transmit data.

A low power downlink using a miniaturized SGLS-type transceiver could support a data rate of 256 Kbps with a transmitter power of 0.5 Watt. Technology is being developed for the PCS market that will be directly applicable to SGLS. The UFO mission could be supported with a simplified SGLS transceiver constructed from a reduced chip set by excluding any DSP requirements.

B. System Impacts on the UFO: (mass, volume, power, etc.)

The most efficacious way to transmit UFO data directly to the ground is to use the S-band Space Ground Link Subsystem (SGLS) in the Air Force Satellite Control Network (AFSCN). This system provides full duplex communication for commanding, tracking and telemetry between satellites and (autonomous) remote tracking stations (ARTS). The stations are equipped with large diameter telescopes (up to 60 feet) and provide (G/T) values between 21 and 26 dB/K. The receivers are designed to support a BER of 10^{-5} with E_b/N_o = 12.3 dB. Using SGLS would allow the nanosatellite transmissions to be tracked and collected at several ground stations as part of normal satellite maintenance operations.

The following link budget calculation was performed for a transmitter power of 0.5 Watt and a slant angle range of 2700 Km. It was assumed that the transmitter had an omnidirectional antenna characterized by a loss of 1 dB, the space loss was 168 dB and the atmospheric attenuation was 1 dB. Then the received isotropic power is -143 dBm and (C/N_o) lies between 76.6 and 81.6 dB Hz at the different sites. This carrier-to-noise range results in an (E_b/N_o) value between 22.5 and 27.5 dB for a data rate of 256 Kb/s. The resulting link margin of 10 - 15 dB is further reduced by about 3 dB due to the complexities of the SGLS waveform. Tracking is performed on the carrier and the data is carried as a binary PCM waveform on a subcarrier offset by 1.7 MHz. Nevertheless, this link budget does suggest that 27 dBm of transmitter power will support the maximum data rate permitted on the mandatory subcarrier. It would take 65 minutes to download the data on this link, corresponding to about 5 overhead passes of the nanosatellite. This projection will support the UFO mission if a BER of 10^{-5} can be tolerated.

For a crosslink scenario, we assumed that the propagation distance had been reduced to 2 Km, and the frequency reduced to 1 GHz for better compatibility with existing wireless communication technology. Then the space loss is reduced to - 98.5 dB. However, the receiver sensitivity is now poor. Assuming an omnidirectional antenna and a system noise temperature of 365 K, the (G/T) value is reduced to - 26 dB/K. Consequently, the link improves by 70 dB from the space loss reduction and degrades by 47 dB from reduced receiver sensitivity, leaving a net gain of 23 dB. If the data rate of 256 Kbps is kept, the transmitter power can be reduced to a few milliwatt and still close the link.

The crosslink has further attractive features that allow error-free transmission. The small latency of a few microseconds permits the data to be sent as packets on demand without incurring significant penalty in reduced data rate. Then detected errors can be corrected on-the-fly by requesting that the packet be sent again. The overhead in bits to form the packet structure is also small. Clearly the packet size should be approximately 0.1 times the inverse BER, e.g., 1Kbyte

for a BER of 10^{-5}.

Bulky components in the communication system will be the master oscillator, rf filters (if required) and the antenna; 14 cm^2 is required for a patch antenna at 2 GHz. The antenna may be designed to provide integrated front-end filtering and transceiver chip designs handle most IF filtering without the need for off-chip filters. Consequently separate filters may not be essential. Nevertheless, a crystal controlled master oscillator will be required.

Workshop 4: Biomedical Applications for Manned Space Flight

Chairman: Charles Sawin, JSC

I. Introduction

This workshop was truly considered exploratory. As the lineage to the UFO project for this workshop could not be considered obvious & overt, the members felt enjoined to investigate multiple applications avenues, using a "clean-sheet-of-paper". The workshop participants quickly focused on building scenarios that would utilize the potential of the technology in providing greater insight to the status and condition of the astronauts in a truly non-invasive fashion. Beyond duplication of standardized telemetry gathering and processing, the workshop also explored how nano/MEMS technology could be applied to providing unique insights to the inner-space of the human body.

Members of the workshop:

Charles Sawin	NASA/Johnson Space Center (JSC)
Neal Pellis	JSC
Dennis Morrison	JSC
Dick Sauer	JSC
Mark Holderman	JSC
Brosl Hasslacher	Los Alamos National Laboratories
Mark Tilden	Los Alamos National Laboratories
Henry Halvejian	Aerospace Corporation

II. Technology for Space Systems

1. Nano/MEMS applications for measuring blood flow & pressure, in-situ organ temperature, and muscular activity were chosen as areas that would have immediate application. However, development of these areas would be considered an iteration upon existing efforts and not representative of the true potential that this technology could deliver.

2. Fluid Delivery systems are viewed as extremely promising. The associated valves, conduits/tubing , actuators, point dispensing tips, metering sensors, and micro reservoirs would have dual applications in both human/bio-medical and Reaction Control Systems (RCS) for nano-satellites. Antibiotics, anti-inflammatories, pain medications and other selected fluids could be autonomously administered from either embedded systems within the astronaut (emergency mode) or from micro-surface packs that are always attached to an external injection port and/or mechanism.

II. Technology for Space Systems (cont'd)

3. Packaging of these and related systems introduces some challenging operational environment requirements. Each component/System would have to be resistant to the harsh and extreme enviornent of space. Additionally, if deployed within a human subject, the environment would be both corrosive and require that any devices or packaging designs consist of materials that are known to present a non-hostile profile to the immune system of the body. Qualification for the space environment could benefit from the minimal (micro) surface areas involved , which could also mitigate any perceived deleterious impacts or affects to the "human system'as insignificant.

III. Development Areas

1. An identified priority for sensor development surfaced in the area of RAD dose detection. This capability could prove to be of critical importance for astronauts performing Extra Vehicular Activities (EVA) or for extended duration space flight, such as that intended for International Space Station Missions. The potential for utilization of an extremely compact device, that is fully integrated with power and telemetry capability, is viewed as very high. The micro size and operational flexibility would allow for location points to be placed at critical points on the body surface, as well as possibly locating some of the devices directly adjacent to certain organs, so that a very accurate full-body (3-dimensional, in effect) RAD dose map could be made. Scheduled updates, in near real-time fashion, to the dose map would also be a natural outgrowth of this type of capability.

2. The same detection device could also become a benign hitchhiker payload on a UFO type vehicle that could "map" the radiation "hot-spots" of the ISS. This would provide some semblance of an optimized EVA route relative to minimizing RAD exposure.

3. Near-Term application outside of space applications could prove quite remarkable. A nano/MEMS RAD dose sensor could be placed directly on the surface of a cancerous tumor (at a number of locations), as well as directly within the "live" center portion of the tumor. This would provide a real-time feedback loop that could maximize the efficiency and accuracy of X-ray treatment protocols. Precise shaping of the focused X-ray beam, along with incident angle orientation and duration of the loitering beam, could prevent localized "brown-tissue" from forming along the X-ray path (primarily on the exit side of the tumor). The subjects recovery time would be markedly decreased and the assessment of the X-ray protocol (i.e., determining the condition of the tumor, post treatment) also occurring in a more timely fashion.

IV. Recommendations

1. Micro/MEMS fluid systems and components should continue to be developed as well as functionally characterized in the space environment.

2. The RAD Nano/MEMS micro device (functionally an ASIM) shows merit. The established and large terrestrial applications of the Medical Community, along with the potential for utilization in the on-orbit environment, indicate that this could be a rewarding pathfinder development.

Workshop 5: Nanosatellites

Chairman: Siegfried Janson, The Aerospace Corporation

I. Introduction:

For this workshop, nanosatellites are defined as satellites with masses between 1 gram and 1 kilogram. Modern digital microelectronics, monolithic microwave integrated circuits (MMICs), and microelectromechanical systems (MEMS) can be integrated together to produce nanosatellites which rival larger microsatellites and minisatellites in capability. While batch-fabricated all-silicon construction is the ultimate goal for nanosatellites, near-term versions could integrate MMICs and MEMS into more conventional designs. Nanosatellites offer low per-unit launch costs and the ability to launch hundreds of satellites using a single small launch vehicle. Nanosatellites should be extremely useful for monitoring a man-made or environmental parameter at a large number of locations in space, for "high delta-V" missions such as interplanetary probes, and for "disposable" satellite missions.

This workshop also focused on propulsion and deployment issues for the UFO. A number of general issues relating to the power system and imaging payload were also discussed.

The major participants for this workshop were:

A. Dorian Challoner	Hughes
Brosl Hasslacher	Los Alamos National Laboratories
Jacky Jouan	Matra Marconi Space
Howard MacEwen	General Research Corporation
Hirobumi Saito	Institute of Space & Astro. Sciences (ISAS)
Mike Socha	Charles Stark Draper Laboratories
Michel Thoby	CNES
Mark Tilden	Los Alamos National Laboratories
Isaiah White	Boeing

II. Technology Assessment for Space Systems:

A. Near Term (1 to 5 years):

1. Current State-of-the-Art

Microsatellites in the 10 kg mass range are already on-orbit, and 5 kg spacecraft, such as the SpaceQuest "nanosat" will be launched within a few years. These spacecraft typically have orbit-average power levels of less than 10 Watts, use simple magnetic stabilization, and carry communications payloads that operate through omnidirectional antennae. Limited Earth observation capability (100 meter or larger ground resolution) with Nadir-pointing within a few degrees is possible using gravity-gradient stabilization. Webersat, a 12 kg microsatellite, carries a color CCD camera but its usage is limited by spacecraft rotation oriented about the local magnetic field vector; it spends most of its time looking away from the Earth.

The 1 kg "Bitsy" satellite bus, under development by AeroAstro and the USAF Phillips Laboratory, is the only current example of a nanosatellite. This bus uses a single circuit board for

communications, command and data handling, power control, etc. The Bitsy offers cold gas propulsion and attitude control with 50 milliradian (3°) accuracy.

2. Attractive Opportunities for Space

Most nanosatellites in the near term will be power-limited and have crude pointing capability. This directly impacts communications link throughput; continuous communications are limited to low data rates and high data rates can only be supported intermittently. Appropriate missions include:

- single-channel voice and data (~10,000 bits/second) relay,
- store-and-forward communications,
- space environment monitoring (radiation, rf, air density, etc.), and
- snapshot imaging of the Earth at low resolution;
  - news services (weather, oil-field fires, oil spills, forest fires, etc.) and
  - Earth resource users (agriculture, mining, etc).

The low cost of launch for nanosatellites enables a number of "disposable" missions with lifetimes of a few days to a few weeks. These short on-orbit lifetimes may allow use of primary batteries (no solar cells, rechargeable batteries, or battery charging system) and cold-gas thrusters for 3-axis attitude control. Communications links will be significantly improved by using directional antennae on the spacecraft. Appropriate missions include:

- air-density monitoring at low altitudes via orbital decay measurements,
- the untethered flying observer;
  - U.S. space shuttle, MIR, international space station and
  - geosynchronous satellites (array deployment, antenna deployment, etc.),
- tactical military satellites (communications and medium-resolution imaging),
- flyby interplanetary probes to near-Earth objects (Venus, Mars, and asteroids),
- lunar orbiters and impacters,
- materials processing, and
- entertainment and advertising ("eye in space").

3. Technology Development Issues

Simple nanosatellites can be constructed using existing flight-tested and commercial technologies. More capable nanosatellites will require gravity-gradient or 3-axis stabilization for high-gain communications antennae, Earth sensor pointing, and solar arrays. Constellations of nanosatellites will require propulsion for deployment, maintenance, and disposal. The major technology issues for near-term nanosatellite and microsatellite missions are:

- low-cost micropropulsion systems for maneuvering and attitude control;
  - micropyro valves or non-leaking latch valves,
  - on/off valves with 1,000,000 : 1 or greater on/off flow ratios, and
  - microthrusters,
- space-qualified, low-power (less than 1 Watt) GPS units,
- high-efficiency (greater than 20% without concentrators) solar cells,

- space-qualified high-energy density batteries in "AA" or "C" size;
  - primary batteries for disposable nanosatellite missions and
  - secondary batteries for long-term missions,
- micro GN&C;
  - programmable plug-and-play GN&C electronics module for modular sensors,
  - low-cost, batch-fabricated sun, Earth, and star sensors, and
  - propulsive, magnetic, or momentum-based actuators,
- active micro-thermal control devices (microlouvers, micro heat pipes, etc.),
- integrated active phased-array antennae for S-band and higher frequencies, and
- low-cost ground stations.

B. Far Term (5 to 20+ years):

1. Emerging Technologies

Nanosatellites in the far term should be highly-integrated yet modular. All-silicon construction based on monolithic batch-fabricated wafers is one design option. Another option is to use other substrates such as silicon carbide or a ceramic for mechanical rigidity, thermal control, and radiation shielding combined with silicon or gallium-arsenide dice that provide electronic, electromechanical, and electrooptic functionality. The monolithic wafers or MCMs are application-specific integrated microinstruments (ASIMs) that perform one or more spacecraft functions. A third option is to use a large light-weight structure, i.e. a rigidized sheet or a balloon, that physically supports a number of interconnected (hard-wired or wireless) ASIMs. The latter approach enables large solar arrays and large phased-array apertures for communications and ranging while maintaining low mass. The emerging technologies that will support one or more of these hypothetical nanosatellites are:

- MEMS;
  - attitude sensors,
  - attitude actuators (micropropulsion and momentum wheels), and
  - active thermal control systems,
- low-power digital electronics;
  - radiation-hardened silicon-on-sapphire CMOS and
  - quantum electronics
- low-power integrated communications circuits;
  - integrated phased-array antennae,
  - single-chip CMOS transceivers for 1 GHz and lower frequencies, and
  - gallium arsenide, SiGe and low-resistivity Si MMICs for higher frequencies,
- multi-chip modules (space-qualified),
- inflatable space structures,
- active control of spacecraft structures, and
- distributed processor architechtures.

2. Technology Development Roadmap

Most of the technologies described above are commercially-driven. Some modifications to COTS MEMS devices will be required for the vacuum, thermal, and radiation environment on-orbit.

Modifications to COTS electronics, either through process, design, or technology changes, will be required for various mission orbits. Some technologies, such as microelectric propulsion, may not be commercially-driven and require targeted development. In all cases, some standardization early on can help accelerate nanosatellite integration and flight qualification. The technology development roadmap for nanosatellites should include the following elements:

- standardize as many interfaces as possible and get consensus from the major contractors and professional (AIAA, IEEE, etc.) societies;
  - ASIM interfaces (mechanical, signal, power, rf, etc.),
  - microsatellite and nanosatellite bus design, and
  - nanosatellite communications protocols,
- continue development of radiation-hardened, low-power electronics;
  - modify COTS (digital, analog, and rf),
  - design new CPUs, signal processors, memory, etc., and
  - develop quantum electronics with radiation hardness in mind,
- space qualify selected terrestrial MEMS;
  - accelerometers and gyros,
  - fluid and gas valves, and
  - magnetometers,
- develop space-specific MEMS;
  - attitude (Earth, sun, and star) sensors,
  - thrusters and micro momentum wheels,
  - thermal control components, and
  - structure dynamic control components,
- develop autonomous control techniques for large constellations, and
- develop low-cost launchers for 100 kg and smaller low Earth orbit payloads;
  - reusable or air-launched expendable rockets and
  - gun launchers.

3. Long Term Issues

Long-term issues include accurate estimation of fabrication costs as a function of time, choice of spacecraft architecture (all-silicon monolithic construction vs. multi-material MCM), choice of mission architecture (constellations or local clusters), process standardization, design standardization, and the degree of modularity required. Nanosatellites require inexpensive mass production of spacecraft systems using integrated MEMS, MMICs, and semiconductor electronics, which in turn requires a large initial R&D effort. Low cost per function is achieved when large numbers of nanosatellites or their individual systems are produced. Some missions can benefit from a large number of dispersed nanosatellites and some can benefit from a large number of nanosatellites in close proximity; free-flying local clusters or physically-connected satellites functioning as a dispersed larger satellite. The major long term issues are:

- What missions are most cost-effective using nanosatellite technology?, and
- When are these missions technologically feasible?

III. UFO Subsystem Summary:

A. Recommendations:

1. Prioritized Technology Requirements

The UFO requires propulsion for maneuvering and attitude control. Attitude control could be performed using magnetic solenoids (torque rods) but this is usually limits angular accelerations to very low levels. As spacecraft shrink in size, propulsive attitude control becomes more attractive due to rapidly diminishing moments of inertia; the moments scale as s^5 for constant geometry and density where s is a characteristic scale length. Cold gas thrusters are appropriate for maneuvering and attitude control while a solid rocket is appropriate for deorbit. The solid rocket is essentially off-the-shelf and in principle, a complete cold gas thruster system could be constructed using existing hardware. The dry mass of a conventional cold gas thruster system, however, would be many times greater than the propellant mass. It makes sense to microfabricate as many of the propulsion system components as possible to minimize the system dry mass. A prioritized list of technology requirements for the propulsion system follows:

- a GN&C module to control the thruster system,
- micro-pyrovalve or the equivalent (latch valve, diaphragm valve, etc.);
  - capable of supporting 50 bar pressures over 10 years without leaking and
  - compatibility with ammonia,
- micro-pressure regulator or reducer;
  - capable of supporting 50 bar pressures and
  - compatibility with ammonia,
- low-cost, low power cold gas thruster modules;
  - 2 or 4 thruster nozzles and microvalves per module with a common gas input,
  - low leak rate, low power, normally-closed microvalves, with
  - integrated pressure monitor, and
- electric micro thrusters for more advanced UFOs with higher ΔV requirements;
  - micro-resistojets for specific impulses between 100s and 200 s, and
  - micro ion engines for specific impulses above 1000s.

2. Ground and Space Qualification Issues

The main issues for the UFO are:

- life jacket design;
  - size and location(s) on host and
  - launch loads,
- radiation hardness and/or tolerance (depends on mission orbit and life jacket design) of command and control, payload, and communications electronics,
- contamination potential of deployment mechanism and thrusters,
- thermal environment inside and outside the life jacket, and
- hermetic sealing of accelerometers and gyros that require 1 bar atmospheric pressure for proper operation

B. System Impacts on the UFO:

Mission requirements strongly influence the UFO design. We assumed that the host vehicle is in LEO, the UFO is initially mounted on the host in a solar-powered "life jacket", the "life jacket" provides differential GPS for UFO position determination, the UFO is operative for 48 hours, the UFO mass is about 1 kg, and the UFO should be de-orbited after 48 hours of use. The UFO really has two mission possibilities:

- the host is fully or mostly functional (it is stabilized and cooperative), or
- the host is significantly disabled (tumbling, spinning, or dead) due to faulty payload separation or catastrophic system failure.

In the first case, slow ejection (a few cm/s) of the UFO followed by injection into an observation "orbit" about the host vehicle is appropriate. In the second case, a fast ejection (~0.5 m/s) may be required to clear the host without getting hit by an appendage, followed by deceleration and either a series of "fast" (about 5 cm/s velocity changes and rotation rates of ~10° per second) maneuvers or injection into the observation "orbit".

The following basic propulsion system parameters were estimated based on the previous assumptions:

- Less than 1 m/s of ΔV is required for the imaging phase if a series of slow "orbits" about the host is acceptable. This requires accurate knowledge of the UFO position relative to the host and accurate knowledge of the host's orbital elements for initial orbit insertion.
- About 10 m/s of ΔV is required for a series of "fast" maneuvers that provide complete host vehicle scans on a time scale much shorter than the host's orbit period.
- About 170 m/s is required for deorbit (700 km altitude)
- A cold gas propulsion system should be used for the imaging phase.
- A solid propulsion system should be used for deorbit.

Cold gas propulsion systems are low in specific impulse (50 to 120 s) and typically require more tank mass than propellant mass yet they are relatively simple and can provide small impulse bits for attitude control requirements. A small solid is very simple, has a specific impulse between 200 and 280s, but can only be fired once. The rocket equation, shown graphically in Fig. 1, shows that the cold gas system will require a propellant mass fraction less than 2% (50 s I_{sp} and 10 m/s ΔV as the worst case) while the solid will require a propellant mass fraction of about 8% (200 s I_{sp} and 170 m/s ΔV).

A schematic diagram of a cold gas propulsion system for complete 3-axis attitude control is shown in Fig. 2. The pyrovalve (or equivalent) isolates propellant from the rest of the system until UFO deployment, which prevents propellant loss through leakage during a possible 15 year dormancy. The fill/drain valve must also be leak-free, but since it is operated only on the ground, it can be a simple mechanical valve with a soft seal. Propellant storage pressure, which is a function of volume limitations and propellant choice, determines the complexity of the pressure reducer. The ideal propellant should have a low molecular weight for high specific impulse, a high density at room temperature under moderate (up to 1 MPa) pressure, and not react with the host spacecraft. No single propellant satisfies all three requirements, so propellant choice must result from a complete constrained systems analysis. Possible choices are ammonia (0.6 gm/cc at 1.1 MPa, molecular weight of 17, but chemically reactive with many materials), nitrogen (0.06 gm/cc at 5 MPa,

molecular weight of 28, chemically inert) and nitrous oxide (0.64 gm/cc at 7 MPa, molecular weight of 44, mostly inert). A ΔV of 10 m/s will require a propellant volume of 160 cm^3 for nitrogen at 100 bar pressure or only 33 cm^3 for ammonia at 10 bar pressure. Propellant volumes for the slow observation orbit mission are an order-of-magnitude lower.

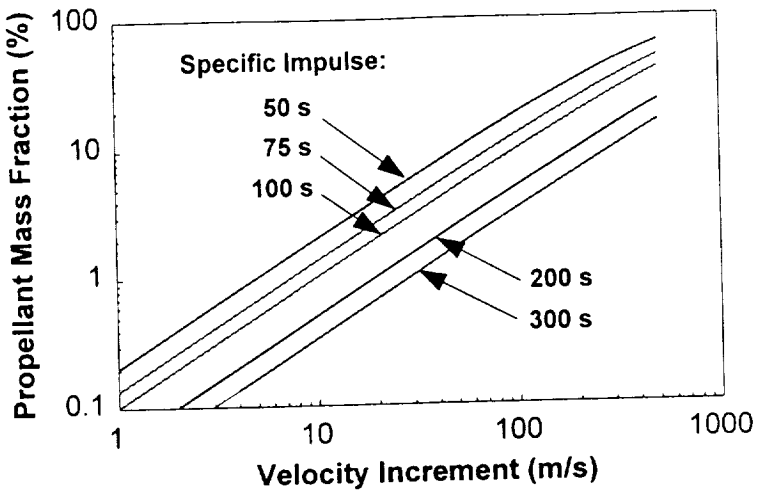


Figure 1. Propellant mass fraction required to generate a given velocity increment.

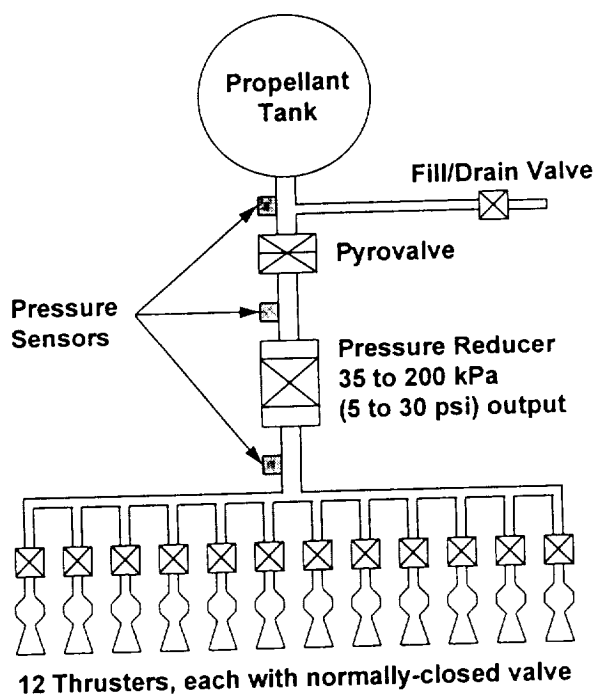


Figure 2. Schematic of a simple cold-gas propulsion system to provide 3-axis attitude control.

Note that a valve is required for each thruster. Current micromachined valves can tolerate inlet pressures up to 200 kPa but have leak rates of at least 0.02 sccm at an inlet pressure of 140 kPa. A UFO with 20 grams of nitrogen, 12 thrusters, and a leak rate of 0.02 sccm per thruster would loose only 4% of its propellant due to leakage during a 48 hour mission. Each cold gas thruster should

produce about 10 mN, which requires a throat diameter of a few hundred microns and an exit plane diameter of between 1 and 2 mm.

Working Group 6: ASIM Applications in Current Space Systems

Chairman: E. Y. Robinson, The Aerospace Corporation

I. Introduction:

A. Background

We defined the concept of an ASIM (application specific integrated microinstrument) as an integrated device providing custom application specific, functions as a stand-alone wireless sensor, or as modular subsystem, for space application. It has all necessary elements for a mass-producible stand-alone microinstrument, with one or more sensors (for example, a geolocation module such as augmented GPS on a microchip), data processing and memory, power, transmit-receive communications, and signal processing. Other implicit elements of such a device concept include a clock, power generation module, and power management. The ASIM combines MEMS with several related fields of microelectronic processing.

A standardized basic platform is envisioned to which may be added modular overlays that provide application-specific capabilities. Therefore, this device could range from a fairly straightforward single or multi parameter microinstrument, to an assembly of modular subsystems that comprises every element found on a satellite, including propulsion and navigation. This would then become a mass-producible subminiature satellite, the nanosatellite, discussed at length elsewhere in this symposium and in the workshops. The key is achieving a producible integrated system, available at moderate cost in low volume production runs (hundreds to a few thousands).

ASIMs offer a technology path to future low cost space systems. Their development could be aided by finding interim beneficial application on existing space systems.

This workshop focused on hypothetical beneficial application of ASIMs in existing space system, and where these potential applications might become stepping stones toward a future, highly integrated, mass producible nanosatellite.

B. Workshop Participants and Prior Tasks

In preparing for this symposium this workshop was organized prior to the meeting and several tasks identified for presentation and discussion here. The general perspective of possible applications to existing systems was based on a recent publication by the Aerospace Corporation, "Microengineering Technology for Space Systems" Aerospace report No. ATR-95(8168)-2, edited by H. Helvajian, in the chapter entitled: "ASIM Applications in Current and Future Space Systems", by E. Y. Robinson. Certain applications were selected for study prior to this symposium. Each task was carried out by the charter members and application leads as noted below.

Charter members of the workshop, preparing specific proposals ahead of the meeting

Ernie Robinson, Chairman	Aerospace Corp.
Jim Collins, Multiparameter Logistic Sensor Task Leader	USAF/Rome Lab
Mark McCallum, Multiparameter Logistic Sensor	USAF/Rome Lab
Dick Schulz, Multiparameter Logistic Sensor	Honeywell Inc.
Charlie Klimcak, Fiber Optic Composite Vessel Sensor	Aerospace Corp.
Geof Smit, GN&C Task Leader	Aerospace Corp.
Jerry Gilmore, GN&C Task	Draper Lab
Mike Robyn, Power	Aerospace Corp.
*Bernardo Jaduszliwer, Fiber optic sensor	Aerospace Corp
*Lonny Smith, Fiber optic sensor	SCI
*Larry Thaller, Power	Aerospace Corp
*Elric Saaski, Fiber optic sensor	RI

* did not attend the symposium/workshop

The proposed development plans were structured to address four key questions:

1. Need for the ASIM application;
2. Benefit of the proposed ASIM application;
3. The general approach to development and demonstration of the proposed application; and
4. Estimate of schedule and cost to achieve the application.

The workshop team was joined by conference participants:

Antonio Martinez de Aragon	ESA/ESTEC, Nordwijk NL
Dean Collins	NIST
Henry Woo	Rockwell
Nick Davinic	NRL/NCST
Vit Babuska	MRJ
Joachim Schulz	FZK/IMT, Germany
John Hayes	Lockheed-Martin
Dorian Challons	Hughes Space & Comm
Rodney Rocha	NASA/JSC
Chuck Sawin	NASA/JSC
Volker Gass	MECANEX/Suisse
Ronald Quinn	Honeywell Inc.

C. General Observations

The workshop objectives were defined as follows:

- Identify technology applications that reduce cost to design, build and operate current and future space systems
- Enhance end-to-end information flow
- Provide flexibility for space missions

The applications of new technology, such as ASIMs, to existing space systems poses special problems. The real advantage of the ASIM lies in cost effctive utilization of large numbers of producible replicate elements, that are optimally used in a total systems approach which fully exploits the new technologies. The application of some aspects of this technology on existing space systems, where the basic systems design is unchanged, limits the cost benefit. Furthermore, achieving such "graft-on" applications requires acceptance by agencies and program managers that presently operate such systems. These folks are traditionally very cautious about adding elements to their space systems, and possibly having to accommodate these graft-ons by changes to existing operations and doctrines. The add-on must therefor have compelling technical benefits and be essentially transparent to the host space system. The ASIM concept could meet such requirements.

The purpose of this workshop is, therefore, to seek value-added interim ASIM demonstrations with minimal disturbance to the host system. The ultimate purpose is to use these demonstrations as steps toward harnessing and integrating these emerging capabilities for totally new space systems, for example, highly distributed parallel networks, and the nanosatellite constellation.

The motivations for ASIMs and related new technology applications are:

- New technology does a job significantly better than present methods
- A valuable new function is enabled by new technology
- Significant cost reduction can be achieved via COTS utilization
- Implementation of highly distributed systems for greater "awareness" and the accommodation of corresponding data proliferation and data quality issues

- Systems level identification of key cost elements by microengineering

D. The Microengineering Connection

The design and development of ASIMs places demands on cooperation among several engineering disciplines: 2D and 3D micromachining, microprocessing foundry services, microelectronic integration (such as Multi-chip modules), MEMS, microcircuit design, low power electronics, quantum electronics, complex separable interconnect schemes and packaging design. The confluence of these disciplines and technologies into an integrated design and development function is the field of microengineering.

E. General Guideline Comments

- These technologies have a variety of possible applications for improved space systems, but specifics must be driven by specific mission requirements for the device or microsystem.
- Formal requirements must be articulated and response solicited from suppliers
- Cost trades must be done to rate the cost-benefits for the specific space applications
- Industry commercial producers will not be motivated by space needs (small number)
- Alternatives for affordable devices to achieve specific special space needs with limited numbers must be pursued through user adaptation and integration of available elements, through captive production facilities, through national micro-machining centers, and possibly through university centers

This leads to a very interesting possibility. Since small quantities of such ASIMs for application to existing space systems are of passing interest to commercial producers, the development and manufacture of such devices for space systems could be achieved within the growing network of university micro-machining centers and foundry services. This may be a highly viable alternative, or adjunct, to piecing together COTS devices.

II. Technology Assessment for ASIM Applications in Current Space Systems

A. Near Term (1 to 5 years)

The near term potential applications may be categorized for each phase of the space system as shown in the figure below, a breakdown of the phases and the general types of ASIMs that might find beneficial current application.

Production and Logistics	Multi-parameter logistics surveillance, ID tags
Launch Base Logistics	Multi-parameter logistics surveillance, T & E
Pre Launch Ops	Multi-parameter logistics surveillance, T & E
LVs and Ascent	Environments for lift-off and ascent
Orbital Operations	**Sub-systems (all stages)** - Ordnance/separation - GN&C - Power - Propulsion - Avionics - Communications - Fluid systems - Structures/mechanical

1. Current State-of-the-Art

All of the discussions in this workshop addressed ASIM application to existing systems and fall into the category of near term application, provided that technology investments are made to realize the potential suggested by the ASIM concept. The ASIM concept and its evolution is tied to each of the workshop areas addressed in this symposium, especially that of data processing, communications, low power electronics, and improved signal detection and processing.

The state of the art at the present includes several MEMS products that utilize accelerometers, microfluid systems (e.g. the HP Inkjet package, and Redwood Systems fluid control microregulators), pressure sensors, chemical sensors etc. A variety of novel concepts are being pursued in R&D activities at government laboratories and university centers that may become commercial items in the near term. Some of these products are intended for use in the specific proposals presented at this workshop (e.g. the Guidance, Navigation and Control Package). In some cases integrated devices currently exist at a large scale, board level, that are good candidates for miniaturization and integration (e.g. The Multi-Parameter Stress Sensor).

Technology Development Proposals, for near term application to existing space systems, were prepared before the conference, and presented at the workshop:

- GN&C	G. Smit
- Multi-parameter logistics sensor	J. Collins
- Integrated fiber optic sensor for composite vessels	C. Klimcak
- Power options for miniature platforms	M. Robyn

Additional candidate application areas were defined during the workshop:

- Leak detection ASIM to address ubiquitous leaks	J. Gilmore (acoustics, SiC)
- SuperSensor (commercial leverage)	D. Collins (multi-sensor)
- T&E System level applications (launch availability)	H. Woo (find cost drivers)
- New standard integrated modules within MOSIS	D. Collins (low cost ASIMs)

2. Technology Development Plans

The following are detailed plans, according to the prearranged format, that were presented to the workshop and recommended for development and demonstration of the ASIM concept.

a. Guidance Navigation and Control Package — G. Smit, Aerospace

NEED

- Where small inertial measurement items are required
- Small spacecraft and low cost missions that have excluded gyros ($$)
- Single gyro missions have failed, without a backup
- New technology may enable non-gyro GN&C, but still need a low cost gyro to recover from tumble
- Low cost redundancy

BENEFITS

- Short life missions may not need rad hard, can use COTS element
- Low cost backups to high value primary devices
- Small size adapted to inertial tracking of large space structures and as part of dynamic feedback control of large space structures
- Useful 5 gm, 0.25 W devices are achievable by available technology integration

ISSUES

- Longer missions will need rad hard
- Variety of GN&C reqm'ts: LRV, ELV, long range, rendezvous, docking...
- LRV reqm'ts differ from expendable, and from SV
- Capabilities not sufficient for highly accurate pointing
- Current commercial translation is not space oriented

b. Multi-parameter Logistics Sensor — J. Collins, USAF/Rome Lab

b. Multi-parameter Logistics Sensor (ad hoc Variant: The SuperSensor — J. Collins, USAF/Rome Lab; D.Collins, NIST)

NEED
- Environment stress states cause equipment failures
- Discriminating fault signature data needed to avoid unverified failures
- Reduce cost of failure (time delay, retest, replace, redesign)
- Accurate information about life cycle and operational environments
- Existing models, and extrapolations have led to inaccurate design criteria
- Reduced costs of integration and test, reduced operational time lines
- Prevent acceptance and processing of flawed hardware
- Improve launch availability; informed maintenance
- Multi-sensor for high and low temperatures (engines, cryo propellants)

BENEFIT
- Builds on experience base in Rome Lab micro TSMD project (for airborne)
- Integrated system can be achieved at low cost
- Generic platform can contain ALL stress parameters:
3-axis vibration, acceleration, acoustics, temperature, pressure, humidity
power quality, strain, event timer and counter, etc.

CHARACTERISTICS
- CPU, non-volatile memory, real time clock, ADC, and optional wireless comm, alternative bus or battery power

c. Fiber Optic Sensor for FW Pressure Vessels — C. Klimcak, Aerospace

NEED
- Flexible and convenient technology for distributed sensing and data processing
- Practical monitor for handling of graphite composite FW pressure vessels

BENEFIT
- Fiber optic fits naturally with composite manufacturing methods
- Bragg reflectors can be conveniently spaced
- Number of sensitive areas is compatible with accurate de-multiplexing
- Extremely simple sensor system
- Fiber optic sensor can detect steady state strain (dc) as well as vibration
- A time based record of environmental history will be resident on each composite vessel or structure
- 3D micromachined spectrometer can be used effectively in the de-multiplexer
- Fiberoptic sensors can be used to sense a variety of distributed parameters

APPROACH
- Development team includes hardware producer, sensors physicists, and electronics supplier
- Phased plan proceeds if all gates met at each phases

d. Leak Detection (ad hoc item) — Jerry Gilmore, Draper Lab

NEED
- Leak detection in cryogenic propellant tankage for RLV, LV, Upper stages, SV
- All pressurized fluid systems for ground and flight are susceptible to leak
- Vibration monitoring of turbopumps

BENEFIT
- Improved launch readiness
- Fault detection and isolation
- Reduce repetitive maintenance to as needed (reduce prelaunch time line)
- Expanded data awareness about system hardware condition

APPROACH

- Integrate current advanced technology in MCM package:
  - micromechanical acoustic sensor
  - micromechanical accels
  - New SiC substrates and devices (for high/low service temperature)
- Develop reference set of detailed requirements
- Specify environmental ranges
- Define a proof of principle demo project
- Define a road map to incorporate with vehicle health monitoring doctrine

3. Recommended for Evaluation and Developments

- Stand alone Integrated GN&C package
  - CMOS
  - Rad Hard
- Multiparameter Logistics Sensor
  - Stand alone (battery power, wireless)
  - "Standardized" interconnects
  - SuperSensor
  - Explore translation to SiC substrate
- Fiberoptic Smart Composite Sensor
  - For strain history and impact events
  - commercial potential
- Power for ASIMs
  - Approach new devices and ASIMs with view to full service life under battery power
- Leak detection ASIM on SiC substrates
  - Surveillance of cryo systems
  - Surveillance of hot engine sections
- System level T&E focus on autonomous operability and integrated health management systems, with models to key on main cost driver assessments (program specific)
- Explore MEMS and high level integration within MOSIS
  - Possible low cost development and fabrication of complex integrated devices in short runs

Workshop 7: Low Power Electronics for Data processing

Co-Chairman: Nick Sramek, The Aerospace Corporation
Gary Frazier, Texas Instruments

I. Introduction

Low power electronics are the key for a near term or even a farther term nanosat. Near term nanosats require a minimal amount of data processing, including station keeping, command and control of all subsystems and interfaces, and data storage, compression and communication. Near term technologies for data processing offer minimal processing for low power, therefore power conservation or power management is critical. For the farther term nanosat, much more processing will be required. The capability to process and store tremendous amounts of data, as well as increased satellite command, control, and autonomy, will be available. Development of new and advanced low power integrated circuit technology will make future nanosats viable.

The major participants for this workshop were:

Nick Sramek	The Aerospace Corporation
Gary Frazier	Texas Instruments
Hector De Los Santos	Hughes
Rob Duncan	Sandia
Doug Matzke	Texas Instruments
Paul van der Wagt	Texas Instruments
Ken Smith	Rice University
Jerry Johnson	Lockheed - Martin
Jim Luscombe	Naval Postgraduate School
P. Mazumder	University of Michigan
Deb Newberry	Computing Devices Int.
Cal Laurvick	Naval Research Labs.

II. Technology Assessment for Low Power Data Processing:

A. Near Term (1 to 5 years):

1. Current State-of-the-Art

Low power microelectronics are currently available which will operate at 3.3 volts. Technologies are emerging that will operate at 2.5 volts. These technologies offer reduced power over 5 volt technologies. Also technologies are being worked on that will operate at voltages lower than 2.5 volts. The 2.5 volt and lower volatge technologies either are available or will be available in the near term.

2. Attractive Opportunities for Space

Low power data processing can only be accomplished today for near term nanosats by utilizing concepts of power management. Electronics can be turned on and off as necessary as well as reducing clock rate for lower power operation. Cleaver power management schemes can significantly reduce the overall operating power of the data processing system.

3. Technology Development Issues

Technology development for the near term nanosats lies in reducing the operating voltage, while increasing the density. There are no emerging technologies that will be available near term that can significantly reduce power over current technologies.

B. Far Term (5 to 20+ years):

1. Emerging Technologies

Technologies are being developed that will reduce the required data processing power significantly. Resonant Tunneling Diodes (RTD) is an emerging technology that can increase the circuit density per unit power . The following chart shows options for the next 20 years.

UFO
NANO TECHNOLOGY OPTIONS

- **MEMORY (2000+)**
  - **SRAM**
    Silicon RTD, MOS peripherial support, DRAM Process
  - **ROM**
    Silicon RTD, MOS-EEROM support, EEROM Process
- **Logic (2005+)**
  - **Control Processor**
    GaAs RTD, Silicon RTD, HBT/MOS Process
  - **Payload Processor**
    GaAs RTD, HBT Process
- **ADC (2000+)**
  - **GaAs RTD, HBT/HFET Process (1 GHz version)**
  - **Silicon RTD, MOS Process (A & D) (1MHz version)**

2. Technology Development Roadmap

The following roadmap shows what the technology can be capable of with adequate funding for technology development.

LOW POWER ELECTRONICS FOR SPACE
NANOTECHNOLOGY ROADMAP

TECHNOLOGY	DEVELOPMENT	PROTOTYPE	TESTBED	FLIGHT
High Speed ADC	1994-1996 2 GSPS @ 4 Bits 100 mW	1997	1998	2003
High Speed ADC	1996-1997 2.5 GSPS @ 6 Bits 500 mW	1998	1999	2004
Low Power ADC	1996-1998 2 MSPS @ 12 Bits 100 mW	1999	2000	2005
Low Power ADC	1997-1999 2 MSPS @ 12 Bits 10 mW	2000	2002	2007
High Speed SRAM	1994-1996 2 Ghz Access 2 mW/bit	1997	1998	2003
Low Power SRAM	1994-1996 20 pW Bit Cell	1997	1998	2003
Low Power SRAM	1996-1998 1 Mbyte IC 20 mW Standby	1999	2000	2005
Low Power SRAM	1998-2000 1 Gbyte IC 20 mW Standby	2001	2002	2007
Low Power SRAM	2000-2010 1 Terabyte IC 1 Watt Standby	2012	2013	2018
Low Power CPU	1997-2000 100 MIPS/Watt	2002	2003	2008
Low Power CPU	2000-2005 1000 MIPS/Watt	2006	2008	2014
Low Power CPU	2005-2010 10,000 MIPS/Watt	2011	2013	2018

3. Long Term Issues

Long term issues involve both technology development and manufacturing. Continued technology development needs to be funded to achieve the goals above. Funding also needs to be allocated to take the emerging technology and make it producible on one or more manufacturing lines for insertion into future nanosats and other applications. Radiation hardening may be an issue. For the near term UFO, electronics can be shielded in a canister with no power applied. Operation is expected to be short term. For the future UFO, longer term operation will necessitate radiation hardening of nanotechnology.

III. UFO Subsystem Summary:

A. Recommendations:

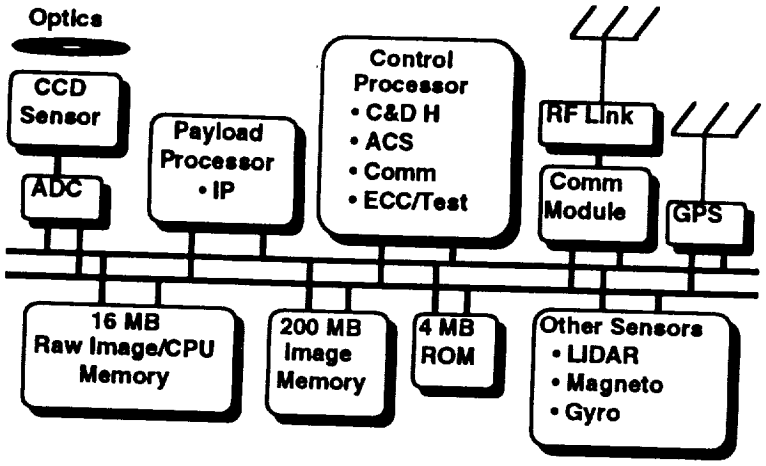
1. Prioritized Technology Requirements

Near term development of RTD technology should focus on the development of memories, in both the volatile and non-volatile memory areas. These memories will both drive the development of the technologies, and provide critical components for future nanosats because of the requirement for increasing amounts of memory for data processing. Follow-on development should be in the processor and analog to digital converter areas.

B. System Impacts on the UFO:

This group discussed design and development of the data processing for the untethered flying observer. A strawman architecture was developed to support near term data processing. This architecture is shown below. Near term power and weight for the data processing will be driven by current technology and current high density packaging technology. The architecture included the data processing for the spacecraft as well as the image processing and data storage.

UFO
COMPUTATION SYSTEM ARCHITECTURE



Concept of Operation

Today: Power management with software is critical to maintain lower power for the UFO. Subsystems need to be turned on and off as needed. An example of this is the payload processor. The payload processor is used for data compression of each picture. This processor needs to be turned on only when required, the data then going into mass storage, and the processor then is turned off for power reduction.

5 year operation: Lower power electronics will be available allowing all of the data processing electronics to run simultaneously . Integrated power management should be available to manage the power within the data processing system.

> 5 year operation: More functions will be available for less power. Electronics will be capable of integration of all processing functions within a small footprint with reduced power.

Workshop 8: Materials, Manufacturing & Fabrication

Chairman: Henry Helvajian, The Aerospace Corporation

I. Introduction:

The topics of this workshop relate to the space qualification of materials, the processing of these materials for microinstrument applications, the packaging of these devices for space applications, the means for prototyping and testing devices under development and the best cost effective manufacturing approaches, given that the procurement lot sizes will never be large. The workshop also explores optimum packaging for a microinstrument suite that will permit the construction of small satellites like the Untethered Flying Observer by assembling modular blocks.

The major participants for this workshop were:

David Altemir	NASA/JSC
James Doscher	Analog Devices Corporation
Craig Friedrich	Louisiana Technical University
Henry Helvajian	The Aerospace Corporation
David Koester	MCNC
James Lyke	Air Force Phillips Laboratory
Adrian Michalicek	Air Force Phillips Laboratory
Bart van der Schoot	University of Neuchatel, Swiss
Norbert Schwesinger	Technical University Ilmenau, Germany
Wayne Stuckey	The Aerospace Corporation

II. Technology Assessment for Space Systems:

A. Near Term (1 to 5 years):

1. Current State-of-the-Art

Current state-of-the-art is driven by commercial applications. Companies that have heavily invested in development costs want to recoup some of these losses. As such they would prefer to focus, in the near term, on their core program (i.e. production for the mass market). Consequently, the manufacturing of one-of-a-kind space-qualified components is thought to be more likely to be possible at "service" centers (e.g. MCNC, Draper Laboratories, BSAC, U. Michigan, Sandia Laboratories). Regardless of where the parts are manufactured, the fact that the number of required units is small will most certainly drive up the cost per unit. Some of this cost may be ameliorated as service centers begin to collaborate in what is called a "virtual factory" design. In the virtual factory each node has an area of expertise and provides a value-added segment to the piece. The part is circulated among the various node facilities for processing. The NSF is testing the "virtual factory" concept in the National Nanofabrication User Facility (NNUF) Program. The concept has merit. No one center could afford to own all the advanced materials fabrication tools currently in use (e.g. LIGA, silicon fabrication + CMOS, Laser processing, micromilling, advanced packaging). It is expected that each site would own, at best, two or three advanced fabrication tools. The advantage of the virtual factory concept is the ability for rapid prototyping of new designs and the customization of components. For space applications, the virtual factory concept could offer a low cost solution for manufacturing of

specialized space-qualified components. The federal government could accelerate the insertion of microengineering technology into space systems by establishing a *collaborative infrastructure* among fabrication and manufacturing centers with specific protocols for developing space qualified parts. For this to be viable, the panel felt that the collaboration aspect must be made a "deliverable" and written into the contract. The panel also felt that the "virtual factory" concept could likely serve as a model for future commercial manufacturing. One element in manufacturing reliability whether via the virtual factory or otherwise is the maintaining of the process reliability. This entails a minimum process-run schedule to insure maintenance of high standards. The panel felt that for MEMS based systems a recommendation of at least 200 wafers/yr per fabrication facility per wafer design be maintained.

In a similar vein, the choice of materials currently available for microinstrument development is solely driven by the fabrication tools which are available. Microelectronics processing techniques work for the semiconductors materials and much has been done with silicon. The development of SiC is a material important for space applications. However, there is a need to utilize material properties which are inherently found for example in, ceramics, polymers, and diamond. For space applications these materials offer a robust alternative to the use of silicon in microinstruments. In the near term non silicon materials can be machined via contact micromilling techniques(e.g. Louisiana Technical University) or for fabricating high precision structures via focused ion beam milling. Non silicon materials can also be processed with lasers. Laser based processing permits rapid prototyping, which has been especially useful in micro-optics fabrication[1]. To date the cost of processing by laser has been uneconomical, However, establishing of the DOE laser material processing facility in Newport News, VA is expected to reduce the cost of processing with light.

The area of MEMS or microinstrument packaging was discussed at great length. Current state-of-the-art in MEMS packaging resembles that used for electronics packaging in the 1960's. It is difficult to predict how microinstrument packaging might evolve in the near term. Microinstrument packages have requirements which are not necessary for the packaging of electronics. In microinstruments some microtransducers need to be hermetically sealed (e.g. high Q resonant structures), while others must interact with the environment (e.g. chemical sensors). The interface between die in a microinstrument may require connections for mass transport, force transduction and electrical continuity. Also, for some MEMS based transducers the physical bonding of the die to the substrate affects performance which entails a post assembly calibration process. Finally, unlike electronic circuits which can be thinned (<50 microns) for ultra high density interconnect packaging, the thinning of MEMS die increases stress which affects repeatability and stability. However, it is anticipated that in 5 years microinstrument packaging technology would be a 3D heterogeneous multichip module with demountable MEMS. Plastic substrates with local shielding as required.

2. Attractive Opportunities for Space

In the near term silicon and silicon carbide have attractive applications for space systems. These materials would primarily be used as transducers in microinstruments. However, if large area (wafer scale) MEMS devices could be fabricated then it is conceivable that these wafers can be used as tiles on space-craft surfaces for applications as simple as "microVelcro™", to more complex applications as in dynamically altering surface properties for thermal management, signal propagation, and electrical permittivity. As non silicon

micromachining technology matures, other materials will also be available for use in microinstruments or in the packaging of these instruments.

3. Technology Development Issues

The panel spent some time deliberating the issue of materials, the response of microstructures to the space environment and the value of validating-materials property databases. The use of bulk material property parameters are not necessarily valid when fabricating microstructures. The panel felt strongly that the government should set standards (i.e. "yardsticks") for materials property database qualification including those in the literature and those which are sold as proprietary. The panel also recommended the materials property characterization of non isotropic materials (i.e. Young's Modulus and Poission's ratio). Sorely lacking is radiation hardness testing of microinstruments. Many current microinstrument designs are capacitively coupled. The effect of radiation and the potential for subsequent charge trapping at the insulating gaps must be addressed prior to any space application.

An area of material processing technology which would enhance the packaging of microinstrument and microstructures is the ability for creating microhermetic seal volumes. Since most MEMS structures are made of silicon, a key element is finding processes which permit the low bulk temperature silicon fusion bonding. Such a technology would enable encapsulation of MEMS structures (e.g. accelerometers, gyros, mechanical "clocks" and filters) and the development of single crystal "thick" MEMS structures.

With regards to specifically advancing microvalving technology. The panel felt that the current leak rates (0.1 μliter/min for liquids @ 2 psi) for MEMS valves need to be reduced by factors of 100 or better to make it viable for many space applications. Of special concern is that most current microvalves are designed for operation at few psi and the leak rates dramatically increase when they are operated at higher pressures (100s psi). For space applications, the required standoff pressures vary but can be as high as thousands of psi (e.g. cold gas thrusters). Although, valves can be designed to withstand such pressures, there is some concern on the physical actuation force necessary to overcome the high pressure. Electrostaticaly driven actuation may not have sufficient force, thermoactuated poppett valves may work but will require lots of power. One design recommended, though no analysis was done, was an iris-style "poppett" valve which is driven by a wobble-screw motor. The panel recommended technology development of low-leak rate valve designs, with patternable elastomers or compliant alloys as valve "seats"

B. Far Term (5 to 20+ years):

1. Emerging Technologies in Materials Manufacturing & Fabrication

The panel used as guide current work being done at various universities and other laboratories world wide. In assessing the emergence of a particular technology let alone predicting its maturity date, the panel ran into some difficulty. The dilemma was not in identifying which from the select group of technologies would make it, but whether some would fail as a result of the inequality in the funding support. Some general trends that were identified were that materials are now "engineered" for the application, in manufacturing the cost of properly disposing of "waste" or byproducts is a significant percentage of the product cost, and in fabrication the process reliability and repeatability is the driving factor. Consequently, any

emerging technology must provide a cost effective alternative while addressing the above issues. However, new technologies and application areas, at least initially, seem to not be confined by these dictums. This nominal advantage does not hold in the far term and as such the panel to that into consideration. Below are listed some of the far term emerging technologies which may support microengineering space applications. The order is not of consequence.

- For materials, novel plastics and other "engineered" ceramics.
- In manufacturing facilities: integrated super fabrication at your workstation, parts are made at distributed facilities, the "broker" is the workstation via the internet.
- Materials processing software: VHDL for microengineered systems (VHSIC hardware description language currently in use for IC design and fabrication).
- For optics: large scale phase-mostly flexure beam micromirrors.
- For surfaces, nozzle, fluid channels: Dynamically reshaping surfaces.
- Packaging: true 3D packaging including manifold for MEMS structures and lines.
- Materials processing: processing with light (i.e. tunable, high repetition rate lasers)
- Materials processing: processing with all additive techniques no subtractive (i.e. etching).
- Interconnects, seals: MEMS assisted

2. Technology Development Roadmap

The development of a roadmap without clear understanding of the specific application is difficult to accomplish. It is even more so for the geneal topic of materials, manufacturing and fabrication. The logical roadmap for this panel is to outline concepts which lead to acceleration in the development of technologies. In this regard and most important is the establishing of an infrastructure which enables rapid prototyping. This entails funding of service centers which have "value added" capabilities, it entails the developing of alliances among service centers to permit manufacturing in "virtual factory" like environments and it entials educating the users in the use of these facilities. The latter concept is facilitated by the development of user friendly CAD/CAE/CAM software. In the case for space applications where reliability is of the essence and the quantities produced small, the roadmap must also include the maintaining of a minimum run-schedule for process reliability.

III. UFO Subsystem Summary:

A. Recommendations:

The following general recommendations are based on the following list of assumptions regarding the UFO and its mission. The UFO must weigh < 1Kg and total power available is 2 watts. The UFO mission is 48 hours, with a 7 year total life. The UFO maneuvers by using ammonia cold gas thrusters (12 in all). The fuel volume is 30ml and is stored at 200 psi. On orbit, the UFO is stored in a cocoon on the outside of the mother ship. The cocoon has solar cells with power available via RF to the UFO. At the end of the mission the UFO deorbits via ignition of a solid fuel.

The panel approached the manufacturing aspects of the UFO by first identifying it as a system and then as an assembly of microcomponents. In some cases it was felt that microtechnology may not be the best approach. The recommended approach for outlining the

manufacturing roadmap is to partition the UFO vehicle into manufacturable segments. The sequence of general thought processes to be followed are outlined below.

- Group UFO into functions.
- Collect elements in group which have particular design embodimenta cluster.
- Check for processing incompatibilities in the cluster
- Fabricate the cluster as an integrated unit.

The assembly of the UFO, then tantamount to assembling the clusters. In this regard, the panel felt that alignment features or alignment structures would be critical for automated assembly. In this regard, LIGA or MEMS components can be used as alignment "pins" or interconnects. This insures proper assembly by the "pick-and-place" machines and if the interconnect is a MEMS device, actuation by externally supplied power could permit the locking of the assembled units. The design of the interconnects could also include sensing elements which enables the tracking and identification of failed sub-assemblies.

The panel felt that the overall system was not the UFO but the UFO in the cocoon. The cocoon will be used to protect the UFO from radiation effects. The radiation environment for a 360 nm, 68 degree inclination orbit is such that an cocoon made of aluminum and 100 mils thick (2.5 mm) transmits 1000 rads/year. This radiation level is thought to be reasonable for the UFO dormant phase. During the active 48 hour mission, damage from either radiation or space debris/meteoride impacts is thought to be negligible (space-debris data for an 852 km altitude, 67 degree inclination orbit show that there are 10^5 impacts/m^2/year for a particle of mass 10 pico-gms. This value reduces to 1 impact /m^2/year for particle of 1 μgm mass).

The panel felt that since the UFO may be developed in a "virtual factory", the cocoon could also serve as the intrafacility clean-box package and carrier system. If this concept is incorporated into the UFO manufacturing program, then there is no reason why the UFO cannot be processed, undergo bakeout, and testing in a clean room environment. The sealed cocoon could then be directly shipped for attachment to the mother ship.

As with regards to the materials which can be used on the UFO, the panel found no reason why the UFO could not be made of plastics or aluminum. The 48 hour limited mission permits many transgressions on established canons. Low outgasing plastics should be used and the UFO optical elements should be protected during the dormant life of the vehicle. The use of plastics significantly reduces the manufacturing cost since injection mold techniques are well established. However, silicon is the better choice because of the long term goal of monolothic packaging with electronics. Silicon has the appropriate stiffness to serve as a UFO structure member. The contact of the fuel (ammonia) with silicon is the only concern. Ammonia is compatible with silicon if water is not present. The presence of water induces the heterogeneous chemistry which leads to silicon etching.

One technological barrier which the panel could identify as affecting the UFO development is the lack of appropriate valves for the propulsion subsystem. Two types of valves are necessary. A main-seal valve which holds the fuel in storage for 7 year life and the "metering" valves for thruster control. The main-seal valve can just be a rupture-seal type (e.g. metal membrane and piston) which is initiate just prior to cocoon exit. Following this action the metering valve must hold off the fuel pressure (200 psi) and on demand release an appropriate gas volume. Given that the mission is 48 hours, a small amount of fuel loss to leak is tolerable.

Assuming that a 10% loss (3 ml) to leaks is manageable, the maximum volume which can be leaked per valve (6 total) is 500 μliters. For the 48 hour mission the resulting leak rate per valve is 0.17 μliters/min. Current MEMS microvalves can deliver on this leak rate at 2 psi backing pressure. However, for the fuel backing pressures of the UFO 200 psi the leak rates are significantly higher. A second aspect of current microvalves is that they are thermoactuated. The necessary power for moving the sealing membrane is around 1 watt for a liquid at 300K. As the temperature of the microvalve is reduced more energy is required to actuate the membrane.

The panel did identify a list of issues specific to the UFO which might be feasible in the 20 year time frame. These concepts are in addition to what was listed above in section II-B-1. For the case of the UFO fuel tank, sensors (pressure, flow, chemistry) would line the inside wall with integrated electronics on the outer wall. The "smart-skin" fuel tank could be interrogated periodically for reliability. For propulsion, integrated ion propulsion platforms should be feasible. That capability arising from the technology development of flat-panel active matrix display technology. A more ambitious concept is the development of a propellant "fabric" (e.g. solid fuel) which is shed (i.e. discharged) after use. The UFO carries a tray of sandwiched propellant fuel "cards". For thrust vector perterbation control, the use of nozzles with dynamically reshaping orifices. This capability would arise out of the MEMS membrane optical switching technology.

1. Prioritized Technology Requirements
   - The development of low-leak rate primary and "metering" valves
     Patternable elastomers which are compatible with MEMS processing.
   - The development of high pressure stand-off valves
     Actuation force compatible for shuttering valve
   - The development of low-power (<<1 Watt) valves which can operate in cryogenics
     Current thermoactuated MEMS valves require more power the lower the temperature
   - The development of low-bulk temperature silicon-silicon fusion bonding technology.
     Fabrication of large structures (e.g tanks) in silicon by sandwiching wafers.
   - The establishing of "good" materials property "yardsticks".
   - Improving the integrated CAD/CAE/CAM framework software.

2. Ground and Space Qualification Issues
   - The pyro-devices which open the cocoon.
   - The RF power transfer to the sealed UFO

B. System Impacts on the UFO:

- Power will be needed to periodically exercise microthruster valves while in dormancy.
- In the near term MEMS microvalves will leak, expect some loss of fuel to leaking.
- Current MEMS valves may require up to 1 watt of power per valve.
- Leaks from microvalves my contaminate UFO optics and mother ship during mission.

---

[1]G. Gal, "Micro-Optics Technology" Tutorial Course Jerusalem College of Technology October 26-27 1994, Jerusalem, Israel.

Workshop 9: Untethered Flying Observer

Chairmen: David G. Sutton and Robert Stroud
The Aerospace Corporation

I. Introduction:

This workshop is central to the others in that it's main task is to coordinate the definition and design of the Untethered Flying Observer. The UFO is conceived to be a free flying vehicle hosted on a larger satellite. It's mission, activated on demand or on predetermined conditions, is to detach itself, capture detailed images of the host and transmit these images to the ground. This mission is to last no more than 48 hours. All of the other workshops have been assigned one or more of the UFO's subsystems to define in a manner consistent with the mission and the specified size, mass and power budgets. This Workshop integrated these subsystems and acted as system manager during the definition process. Those subsystems requiring extensive development to meet the stated mission objectives were flagged and technology road maps designed in cooperation with the appropriate panels. In addition, Workshop 9 identified and defined alternate missions for untethered, parasitic, microengineered spacecraft.

In addition to the Chairmen the major, full-time participants for this workshop were:

Munson Kwok	The Aerospace Corporation
Dennis Wells	Johnson Space Center
Antonio Martinez de Aragon	European Space Agency

The workshop also benefited from contributions by members of the other workshops. Specific subsystems were conceived and outlined for the workshop by the following individuals.

Sherwin Amimoto	Imaging Sensor
Michael Gorlick	Processing Software
John Hurrell	Low Power Communication
Siegfried Janson	Cold Gas and Solid Propulsion
Geoffrey Smit	Guidance and Control

II. Technology Assessment for Space Systems:

A. Near Term (1 to 5 years):

1. Current State-of-the-Art

The UFO's mission is real. A satellite with the identical mission to the UFO is currently under development at Johnson Space Flight Center for use on Shuttle flights. NASA's version is named the Aercam. It is a 15 Kg spacecraft being developed for a demonstration that includes man-controlled circumnavigation of an orbiting shuttle to provide real time images of shuttle tiles to the crew. In addition to the Aercam, several versions of a parasitic spacecraft, with the identical features of cold gas propulsion and an imaging sensor, have been proposed by Daimler-Benz (Inspector), JPL (Roboball), and Space Industries (Nanosatellite). The Roboball employs a low power imaging sensor especially developed by JPL; otherwise none of these spacecraft is currently conceived to employ microtechnology. They range in mass from the 15 Kg of Aercam to 176 Kg for the Inspector.

2. Attractive Near Term Opportunities for Space

The UFO workshop in collaboration with personnel from the Workshop on Biomedical Applications identified a number of near term missions suitable for spacecraft that would naturally evolve from the UFO. The following table lists single use missions with their associated issues and chief enabling technologies that were identified in this process.

Table 1. Near term single use missions for the Untethered Flying Observer

Mission	Payload	Issues	Enabling Technologies
Environmental mapping			
Radiation	Radiation sensors	Space qualification	Low power radiation microsensors
Chemical (H_2, H_2N_2, HCl, etc.)	Chemical microsensors		Chemical microsensors
Physical (mechanical, temperature)	Vibration sensors, thermometers	Attach UFO to various points on host vehicle	Soft docking, attachment points
Probes			
Atmospheric	Atomic oxygen sensor		
Magnetic	Magnetometers		
Calibration			
Optical	Radiation source, calibrated reflectors	Precise positioning	Ranging, station keeping
RF	Receivers	Precise positioning	Ranging, station keeping

A number of additional mission were identified that required a capability to retrieve and reuse the UFO. These missions are described in Table 2. The technologies that enable safe retrieval of the UFO are common to all of these missions, but are not listed in the table. These were identified by the panel as replentishable propellants (methane), soft docking, modular design (for payload, propellant and subsystem replacement), low power state for extended life, and robotics.

B. Far Term (5 to 20+ years):

The UFO Workshop also identified several missions that could be achieved in the long term. Table 3 outlines these missions and the capabilities that would need to be acquired to make them feasible.

Table 3. Far term missions for an Untethered Flying Observer

Mission	Payload	Issues	Enabling Technologies
Cooperative missions (2+ UFOs)			
Stereo imagery	Imaging sensors	Same scene viewing	Synchronized imaging, precision pointing
Geolocation of emitting sources	Receivers		Precise navigation and pointing
Interferometry	Receivers	Orbital phasing	Precise navigation and pointing
Long Range Missions			
Assessment of ailing spacecraft, other that the host	Imaging sensors	Radiation hardening, orbital transfer	Ion propulsion

III. The UFO Spacecraft

The UFO Workshop spent considerable effort developing a conceptual design and establishing the baseline configuration that is consistent with the mission and the specified size, mass and power budgets. For this exercise the mission was to provide a complete image survey of a host spacecraft within 48 hours of activation. For a host in low earth orbit the UFO would be capable of independently telemetering these images to earth. The UFO spacecraft itself was limited to a mass of 1 kg, a volume of 350 cm^3 and an average power consumption of 2 watts. Figure 1 is a conceptual drawing of the UFO and host after deployment.

Table 2. Near term missions for a reusable Untethered Flying Observer

Mission	Payload	Issues	Enabling Technologies
Communications around large space structures	Transponder, repeater	Power	Smart trajectories
Damage survey	Video	Real time flight control	
Contamination assessment	Chemical sensors, mass spectrometer	Decontamination of UFO	Sensor package
Robotic assistant	Tools, parts		High level controls, Robotics

3. Technology Development Issues

Simple nanosatellites can be constructed using existing flight-tested technologies, including cold gas propulsion, standard solar cells and patch antennas. However, for the near term missions outlined above several technologies will need to be pushed to a state of flight readiness not now attained. The three areas listed below are currently under development and if they continue to show progress will be available for UFO-like missions.

- Low power imaging systems
- Combined GPS-microgyro navigation and control systems
- Low power data processing

In addition to these developing technologies several areas need specialized attention if we are to achieve this mission in the next five years. These areas are not currently being pursued, nor are they likely to be developed in any sphere not specifically space oriented. These technologies are primarily related to the UFO platform. They include:

- Development of microvalves suitable for operation in the space environment
- Validation of a GN&C scheme for close-in navigation
- Assembly procedures for microsatellites using facilities that are geographically scattered
- Microthruster development
- Power management

Finally there are several general capabilities that can be classified as enabling both for the UFO mission and for more advanced microsatellite missions. Most of them are currently being developed under ARPA's infrastructure program or one of the commercial communications satellite programs. Nevertheless, the essential nature of these capabilities requires that they be monitored closely and augmented as necessary in order to ensure a full capability is available for microsatellite development. This latter class of capabilities includes the following:

- Fabrication capabilities in non-silicon materials
- Custom production of application specific microinstruments for ranging, attitude sensing and control, and sensor payloads
- Mass production of satellites

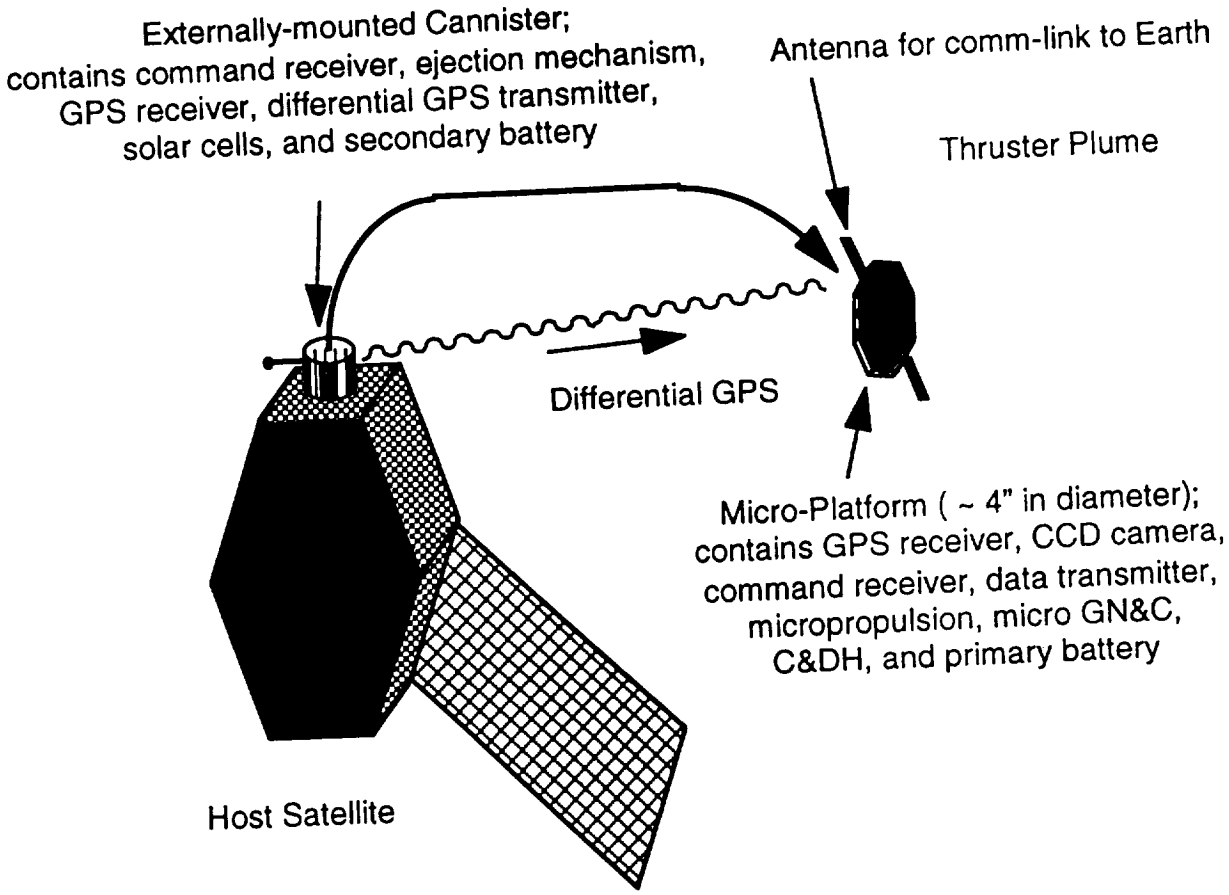


Figure 1. The LEO Untethered Flying Observer

Working in conjunction with the other panels a baseline design was derived that would be capable of performing this mission within the defined constraints of total mass, power and volume. However, the derived design requires a storage canister on the host to provide long term shielding for the UFO. In addition, the canister has an independent power source, can be triggered to release the UFO from a ground signal (or predefined spacecraft conditions) is equipped with a GPS receiver and can provide its position to the UFO for navigation by differential GPS. The canister's systems are not included in the budgets allotted for the UFO.

The following table lists the subsystems that comprise the UFO and summarizes their requirements for mass, power and volume.

Table 4. UFO subsystems and budgets

Subsystem	Vol. (cm^3)	Mass (gms)	Peak Power (mw)	Average Power (mw)	Energy (w-hrs)	Issues
PROPULSION						
DE-ORBIT	50	110				SOLID ROCKET
MANEUVER	30	20	100		5	COLD GAS - ORBIT MANAGEMENT
		30				10 THRUSTERS
SENSORS						
CAMERA	140	50	300		1.4	REDUCE POWER: SLEEP MODE, COUPLED A/D
BATTERIES	70	220				48 HRS @ 220 W-HR/KG
GN&C						
GPS	15	50	250		12	
MICROGYRO	2	5	250		12	
MAGNETOMETER	1	1	10		0.5	USE IMAGE PROCESSING TO ACQUIRE HOST S/C
COMMUNICATIONS	100	100	200	2000	5	DONE IN SERIAL WITH DP
DATA PROCESSING	10	30	500	1500	25	"
STRUCTURE	80	250				INCLUDES PACKAGING
TOTALS	498	866	1610		60.9	

These subsystems comprise the main elements of the UFO. Each will be described briefly in turn.

The propulsion system has two components. A solid rocket for deorbit and a cold gas system for maneuvering. The cold gas system uses ammonia for propellant in order to minimize tankage mass. The required ΔV to provide a scan of the host spacecraft is minimized by the use of asymmetric orbits. Properly executed a well designed flight mode provides a single path scan of the host every half orbit. Dennis Wells has conducted a study of flight modes for the Aercam surveillance of the space station. One such mode suitable for the UFO mission requires 0.34% of the vehicle mass (3 grams for the UFO) to execute. Given a tankage and system fraction of 1/2, we have allowed for 10 grams of propellant or enough to

accommodate three of these maneuvers. Most of the mass of the propulsion system is allotted for deorbit, consistent with the much higher ΔV requirements. A small power allocation is provided for valves and sensors related to the propulsion system.

The payload is the largest subsystem and requires the second largest power allotment. A detailed account of the optical system is included in the report of the Workshop on Sensors and Transducers. It consists of a 125 mm focal length telescope designed to fit in a 10 cm diameter circle. The sensor package is a visible CCD sensor with 512X512 pixels. The sensor will require some development in order to achieve the low power consumption and data throughput rates. Several such devices are under development at various locations including JPL. Nevertheless, careful integration and aggressive power management will be required to meet the allotted 300 mw average power.

Instead of solar cells as originally planned. The short duration of the mission enables the use of batteries as the sole power source. Lithium batteries provide the necessary shelf life and specific energy to meet this mission. We have allotted mass and volume based on current performance characteristics.

The guidance navigation and control system is unconventional. Primary location is accomplished by differential GPS. Both the canister and the UFO have GPS units. These units will provide the vector and distance to the host. A combination of microgyros and magnetometers are sufficient to stabilize the UFO and point the telescope. However, in order to provide for acquisition of the host, it was necessary to invoke an image processing mode based on the visible camera system. This puts some additional, but not prohibitive, requirements on the data processing system. The combined mass and power consumption of these units is estimated at 56 grams and 510 milliwatts respectively.

Because of the high peak power requirements the communications (data transmission to the ground) and data processing are done serially. In addition, data transmission must be done when overflying ground stations with high gain dishes. Patch antennas on the UFO are sufficient for this task thereby eliminating any pointing requirements or high gain antennas on the UFO. These systems are described in their respective workshops. It suffices here to point out that low power, high speed data processing and memory are required to enable the large data flows generated by imaging and that aggressive power management will be a major enabling factor.

III. Summary and Recommendations

The UFO's mission is both real and achievable. Employment of microtechnology will enable UFO like spacecraft to be hosted on unmanned, relatively small hosts. Many of the technologies required for this mission are already under development. These include microgyros and GPS receivers, low power sensors and data processing, thin film-high specific power batteries. Many alternate payload sensors are also under development including, chemical microsensors and magnetometers. Once developed, such devices could greatly expand the range of feasible missions.

However, propulsion and platform control systems need to be developed individually, since they have no sponsors from other fields. These technologies could enable long term and long range UFO missions and ultimately establish the platform for independent micro/nanosatellites.

Appendix I

Attendance List for the Conference

First Name	Last Name	Affiliation	Address	City	State or Country	Zip Code	Phone Number
J. Douglas	Adam	Westinghouse	555 Technical Square, Mail Stop 03	Cambridge	MA	02139	(617) 258-2832
Byong	Ahn	C.S. Draper Lab	Johnson Space Center, Mail Code EV3	Houston	TX	77058	
David	Altemir	NASA					
Vincent	Amatucci	Sandia	P. O. Box 92957, Mail Stop M5/753	Los Angeles	CA	90009	(310) 336-7660
Sherwin	Amimoto	Aerospace Corp.	College of Computing, Mail Stop 0280	Atlanta	GA	30332	(404) 894-8209
Ronald	Arkin	Georgia Tech	Johnson Space Center, Mail Code EV3	Houston	TX	77058	
G. Dickey	Arndt	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Scott	Askew	NASA	Dept of Physics	Stony Brook	NY	11794	(516) 632-8177
Dmitri	Averin	SUNY	10560 Arrowhead Dr	Fairfax	VA	22030	(703) 277-1320
Vit	Babuska	MRJ	P. O. Box 534, S-751 21	Uppsala	Sweden		(46) 18 18 3023
Ylva	Backlund	Uppsala Univ.	4800 Oak Grove Drive	Pasadena	CA	91109	(818) 354-4467
Charles	Barnes	JPL					
Don	Batory	U. of Texas	One Space Park, Mail Stop D1-1050	Redondo Beach	CA	90278	(310) 814-6452
Andrew	Benjamin	NASA	Linder Hoehe D-51147	Cologne	Germany		(713) 483-5478
John	Berenz	TRW	P.O. Box 92957, Mail Stop M2/244	Los Angeles	CA	90009	(310) 336-9269
Hans	Blome	DLR					
Walter	Bloss	Aerospace Corp.	2451 Crystal Drive	Arlington	VA	22245	(202) 767-9603
T.	Brand	C.S. Draper Lab	P.O. Box 655936, MS 134	Dallas	TX	75265	(214) 995-4312
William	Breedlove	Space & Naval Warfare Sys	Johnson Space Center, Mail Code EG	Houston	TX	77058	
Tom	Broekaert	Texas Instrum.	Johnson Space Center, Mail Code DM4	Houston	TX	77058	
Don	Brown	NASA	Johnson Space Center, Mail Code NS4	Houston	TX	77058	
Philip	Burley	NASA					
Kent	Castle	NASA	P.O. Box 902, Mail Stop E0/E1/D110	El Segundo	CA	90245	(310) 416-5219
Mary	Cerimele	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
A. Dorian	Challoner	Hughes Space & Com	Johnson Space Center, Mail Code EV3	Houston	TX	77058	
John	Chladek	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Andrew	Chu	NASA	Room H425, Bld 101	Gaithersburg	MD		(301) 975-4712
Tim	Cleghorn	NASA					
Dean	Collins	NIST	Johnson Space Center, Mail Code EA	Houston	TX	77058	
James	Collins	USAF/Rome Lab	Johnson Space Center, Mail Code EV13	Houston	TX	77058	
Phil	Congdon	Texas Instrum.	1 Cyclotron Road, Mail Stop M550A5131	Berkeley	CA	94720	(510) 486-5438
Ken	Cox	NASA	4555 Overlook Ave S.W., Code 8222	Washington	D.C.	20375	(202) 404-7410
William	Culpepper	NASA	Johnson Space Center, Mail Code EV13	Houston	TX	77058	
Glen	Dahlbacka	Lawrence Berkeley Natl Lab					
Nicholas	Davinic	Naval Research Lab					
Robert	Davis	NASA					
Hector	De Los Santos	Hughes					

Phillipe	Dejour	NASA	Johnson Space Center, Mail Code EV2	Houston	TX	77058	
David	Denniston	Texas Instruments	2451 Crystal Drive	Arlington	VA	22245	(703) 502-6029
Dwight	Denson	SPAWAR 40	Johnson Space Center, Mail Code DF15	Houston	TX	77058	
Shen	Dexin	Chinese Academy of Sciences	Johnson Space Center, Mail Code EA63	Houston	TX	77058	
Tom	Diegelman	NASA	831 Woburn St.	Wilmington	MA		(617) 937-1534
Mark	Diogu	NASA	Johnson Space Center, Mail Code EV3	Houston	TX	77058	
James	Doscher	Analog Devices	21743 Pinewood Cr	Sterling	VA	20164	(703) 430-6071
Rob	Duncan	Sandia National Laboratory	7525 Colshire Dr., Mailstop Z761	McLean	VA	22102	(703) 883-5930
Tim	Early	NASA	Johnson Space Center, Mail Code HA	Houston	TX	77058	
Stephen	Edwards	USAF	P.O. Box 92957, Mail Stop M1-136	Los Angeles	CA	90009	(310) 336-7604
W.	Ehrfeld	Inst of Microtech	P. O. Box 92957, M2/247	Los Angeles	CA	90009	(310) 336-6000
James	Ellenbogen	Mitre Corp.	Johnson Space Center, Mail Code EA63	Houston	TX	77058	
Kyle	Fairchild	NASA	Johnson Space Center, Mail Code SD5	Houston	TX	77058	
Wayne	Fenner	Aerospace Corp.	Gower Street	London	United Kingsdom	WC1E6 BT	(44) 171-391-1583
Seymour	Feuerstein	Aerospace Corp.	2010 Hoeback Rd, P.O. Box 1130319	Ann Arbor	MI	48113	(313) 971-4422
David	Fletcher	NASA	P.O. Box 655936, Mail Stop 134	Dallas	TX	75265	(214) 995-2844
Suzanne	Fortney	NASA	P.O. Box 10348	Ruston	LA	71272	(318) 257-2777
Terence	Fountain	University College London	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
James	Fox	Canopus Systems Inc	Z. I. Nord	Nyon	Switzerland	1260	(41) 22-361-77-33
Gary	Frazier	Texas Instruments	Johnson Space Center, Mail Code EA4	Houston	TX	77058	
Craig	Friedrich	Louisiana Tech Univ	555 Technology Square	Cambridge	MA	2139	(617) 258-1271
Michael	Furlong	Boeing	Johnson Space Center, Mail Code EA4	Houston	TX	77058	
George	Gal	Lockheed Martin	Johnson Space Center/EG	Houston	TX	77058	(713) 483-8244
Raphael	Garcia	NASA	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
Volker	Gass	Mecanex SA	P.O. Box 92957, M1-102	Los Angeles	CA	90009	(310) 336-8661
Ron	Gerlach	NASA	JSC, Mail Code ER	Houston	TX	77058	
Jerold	Gilmore	C. S. Draper Laboratory	1301 Beal Ave, Rm 3303	Ann Arbor	MI	48109	(313) 764-7139
Chuck	Goldsmith	Texas Instruments	Johnson Space Center, Mail Code NS3	Houston	TX	77058	
Winston	Goodrich	NASA	Johnson Space Center, Mail Code NB	Houston	TX	77058	
Kenneth	Goodwin	C. S. Draper Laboratory					
Tom	Goodwin	NASA					
Michael	Gorlick	Aerospace Corp.					
Keith	Grimm	NASA					
Erich	Grossman	NIST					
Adolfo	Gutierrez	InterScience					
George	Haddad	Univ of Michigan					
Max	Haddock	NASA					
Bill	Harris	NASA					
Brosl	Hasslacher	LANL					

John	Hays	Lockheed Martin	P.O. Box 179, Mail Stop 8048	Denver	CO	90009	(310) 336-7680
Henry	Helvajian	Aerospace Corp.	P. O. Box 92957, Mail Stop M5/753	Los Angeles	CA		(32) 16-281-463
Lou	Herhans	IMEC	Kapeldreep, Mail Stop 75	Leuven, 3001	Belgium		
Mac	Himel	NASA	Johnson Space Center, Mail Code NS3	Houston	TX	77058	
Takayuki	Hirano	JETRO Chicago	401 N. Michigan Ave., Suite 660	Chicago	IL	60611	(312) 527-9000
Ben	Hocker	Honeywell	12001 State Highway 55, MN 14-4B55	Plymouth	MN	55441	(612) 954-2745
Mark	Holderman	NASA	Johnson Space Center, Mail Code MA5	Houston	TX	77058	
Alan	Holt	NASA	Johnson Space Center, Mail Code OD	Houston	TX	77058	
John	Hurrell	Aerospace Corp.	P.O. Box 92957, Mail Stop M2-238	Los Angeles	CA	90009	(310) 336-1636
John	James	NASA	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
Siegfried	Janson	Aerospace Corp.	P. O. Box 92957, Mail Stop M5/754	Los Angeles	CA	90009	(310) 336-7420
Ken	Jenks	NASA	Johnson Space Center, Mail Code PC	Houston	TX	77058	
Mike	Jensen	NASA	Johnson Space Center, Mail Code EG3	Houston	TX	77058	
David	Jih	NASA	Johnson Space Center, Mail Code EV13	Houston	TX	77058	
Jerry	Johnson	Lockheed Martin					
Jacky	Jouan	Matra Marconi Space	31 Rue des Cosmonautes	Toulouse	France	31077	(61) 39-66-50
Richard	Juday	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Dayon	Kane	NASA	Johnson Space Center, Mail Code FA	Houston	TX	77058	
Melvin	Kapell	NASA	Johnson Space Center, Mail Code EV!3	Houston	TX	77058	
Charles	Klimcak	Aerospace Corp.	P.O. Box 92957, Mail Stop M2-253	Los Angeles	CA	90009	(310) 336-6069
David	Koester	MCNC	3021 Cornwallis Rd.	RTP	NC	27709	(919) 248-1455
Edward	Kolesar	Texas Christian University					
Carl	Kotila	NASA	Johnson Space Center, Mail Code MA5	Houston	TX	77058	
Kumar	Krishen	NASA	Johnson Space Center, Mail Code HA	Houston	TX	77058	
Dave	Ksienski	Aerospace Corp.	P.O. Box 92957, M1-135	Los Angeles	CA	90009	(310) 336-0181
Munson	Kwok	Aerospace Corp.	P.O. Box 92957, M5-753	Los Angeles	CA	90009	(310) 336-5441
C.	Kyriacou	JPL (NASA)	4800 Oak Grove Drive, Mail Stop 161-213	Pasadena	CA	91109	
Mark	Lacombe	Boeing					
J. F.	Lam	Hughes Research Labs	Johnson Space Center, Mail Code EV3	Houston	TX	77058	
Jim	Lamoreaux	NASA	Johnson Space Center, Mail Code EV	Houston	TX	77058	
Ken	Land	NASA	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
Helen	Lane	NASA					
Gene	Lang	Lockheed Martin	P.O. Box 2242	Woodbridge	VA	22193	(713) 502-6076
Conrad	Laurvick						
Byeung Leul	Lee	Samsung AIT	Johnson Space Center, Mail Code EM	Houston	TX	77058	
Gary	Lee	Boeing	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Lubert	Leger	NASA	13588 N. Central Expressway	Dallas	TX	75252	(214) 995-7742
Larry	Li	NASA	3554 Chain Bridge Road	Fairfax	VA	22030	(703) 273-7010
Tsen-Hwang	Lin	Texas Instruments					
Dino	Lorenzini	Spacequest					

Chris	Lovchik	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Forrest	Lumpkin	NASA	Johnson Space Center, Mail Code EG3	Houston	TX	77058	
James	Luscombe	Naval Postgrad Sch					
James	Lyke	Phillips Lab					
Howard	MacEwen	Gen Research Corp					
Marc	Madou	UC Berkeley					
Charlie	Mallini	NASA					
Kin	Man	JPL (NASA)	Johnson Space Center, Mail Code EA6	Houston	TX	77058	
Robert	Mandelbaum	Univ of Penn	4800 Oak Grove Drive, Mail Stop 301-456	Pasadena	CA	91109	(818) 393-0255
Alvin	Marks	Advanced Research Dev					
Karen	Markus	MCNC MEMS Tech App Ctr	359R Main Street	Athol	MA	1331	(508) 249-4696
Antonio	Martinez de Aragon	ESA/ESTEC	P.O. Box 299, Mail Stop FSA	2200AG Nordwijk	Netherlands		(31) 71-565-3698
Jim	Masciarelli	NASA					
Andrew	Mason	Univ of Michigan	Johnson Space Center, Mail Code EA63	Houston	TX	77058	
Sumner	Matsunaga	Aerospace Corp.	1301 Beal, 1246 EECS, Mail Stop 2122	Ann Arbor	MI	48109	(313) 764-3352
Douglas	Matzke	Texas Instruments	13873 Park Center Rd, Mail Stop Hern850	Herndon	VA	22071	(703) 709-9749
Pinaki	Mazumder	University of Michigan					
Mark	McCallum	USAF/Rome Lab	Dept of EECS, Mail Stop 2215	Ann Arbor	MI	48105	(313) 930-1953
Timothy	McHale	TASC					
Elric	McHenry	NASA	14100 Prake Meadows Dr	Chantilly	VA	22021	(703) 830-4444
M. Adrian	Michalicek	Phillips Lab	Johnson Space Center, Mail Code EA6	Houston	TX	77058	
Ted	Moise	Texas Instruments	3550 Aberdeen Ave SE, Bld 887	Kirtland AFB	NM	87117	(505) 846-4955
Leo	Monford	NASA	13588 N. Central Expy, Mail Stop 134	Dallas	TX	75243	(214) 995-2781
Robert	Moreland	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Naoki	Motoshima	JETRO Chicago	Johnson Space Center, Mail Code EP4	Houston	TX	77058	
David	Nagel	Naval Research Lab	401 N. Michigan Ave., Suite 660	Chicago	IL	60611	(312) 527-9000
Don	Nelson	Aerospace Corp.	4555 Overlook Ave., S.W.	Washington	D.C.	20375	(202) 767 2931
Don	Nelson	NASA	P.O. Box 92957, Mail Stop M1-928	Los Angeles	CA	90009	(310) 336-4244
Deb	Newberry	Computing Devices Int	Johnson Space Center, Mail Code DF15	Houston	TX	77058	
Clark	Nguyen	Univ of Michigan	8800 Queen Ave, Mail Stop BLCSIP	Bloomington	MN	55431	(612) 921-6692
Hai	Nguyen	NASA					
Thanh	Nguyen	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Sid	Novosad	NASA					
James	Park	NASA	Johnson Space Center, Mail Code EV	Houston	TX	77058	
Neal	Pellis	NASA	Johnson Space Center, Mail Code EG3	Houston	TX	77058	
Dave	Petri	NASA	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
Sam	Pool	NASA	Johnson Space Center, Mail Code EG5	Houston	TX	77058	
Joseph	Prather	NASA	Johnson Space Center, Mail Code SD	Houston	TX	77058	
James	Preslock	Univ Tech Sci Ctr	Johnson Space Center, Mail Code EM	Houston	TX	77058	

Charles	Price	NASA	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Andris	Pumpurs	DOD	13610 Deerwater Dr.	Germantown	MD	20874	(703) 502-6128
S.	Rajic	Oak Ridge National Lab					
Dick	Ramsell	NASA	Johnson Space Center, Mail Code EG5	Houston	TX	77058	
John	Randall	Texas Instruments	P.O. Box 655936, MS 134	Dallas	TX	75265	(214) 995-2723
James	Randolph	JPL (NASA)					
Bob	Reinicke	Marotta Sci. Controls, Inc.	17671 Irvine Blvd, # 208		CA	92680	
Millard	Reschke	NASA	Johnson Space Center, Mail Code SD5	Houston	TX	77058	
Chuck	Reynerson	SPAWAR					
Will	Reynolds	Space Community Found.	P. O. Box 580387	Houston	TX	77258	(713) 532-2000
Ed	Robertson	NASA	Johnson Space Center, Mail Code EA63	Houston	TX	77058	
Ernest	Robinson	Aerospace Corp.	P. O. Box 92957, Mail Stop M2/248	Los Angeles	CA	90009	(310) 336-7824
Michael	Robyn	Aerospace Corp.	P. O. Box 92957, Mail Stop M5/122	Los Angeles	CA	90009	(310) 336-1009
Rodney	Rocha	NASA	Johnson Space Center, Mail Code ES2	Houston	TX	77058	
Matt	Rosenthal	NASA	Johnson Space Center, Mail Code DT66	Houston	TX	77058	
Richard	Rosenthal	TRW Space & Electronics	One Space Park, Mail Stop R4/2050E	Redondo Beach	CA	90278	(310) 814-8939
Gary	Rummelsburg	Hughes					
Hirobumi	Saito	Inst of Space & Astro Sci	3-1-1, Yoshinodai, Sagamihara	Kanagawa	Japan	229	(81) 3-427-51-3911
Clarence	Sams	NASA	Johnson Space Center, Mail Code SD4	Houston	TX	77058	
Darryl	Sargent	C. S. Draper Laboratory	555 Technology Square, Mail Stop 65	Cambridge	MA	02139	(617) 258-2296
Bob	Savely	NASA	Johnson Space Center, Mail Code PA	Houston	TX	77058	
Charles	Sawin	NASA	Johnson Space Center, Mail Code SD5	Houston	TX	77058	
Joachim	Schulz	TZK	Inst. f. Mikrostrukturtechnik, Postfach 3640	Karlsruhe	Germany	76021	(49) 7247-82-4438
Richard	Schulze	Honeywell	3660 Technology Dr, Mail Stop 65-2500	Minneapolis	MN	55418	(612) 951-7895
Norbert	Schwesinger	Tech Univ of Ilmenau	PT 0565	Ilmenau	Germany	98684	3677-682878
Sonya	Sepahban	NASA	Johnson Space Center, Mail Code EG	Houston	TX	77058	
I.	Shimoyama	University of Tokyo					
Meenam	Shinn	Samsung AIT	Johnson Space Center, Mail Code EV4	Houston	TX	77058	
Robert	Shuler	NASA	7 Avenue Colonel Roche	Toulouse	France	31077	(33) 61-33-6362
Jean	Simonne	LAAS-CNRS	3550 Aberdeen S.E.	Kirtland AFB	NM	87117	(505) 846-0484
Babu	Singaraju	Phillips Laboratory					
T.	Slater	Wildcat Micromachining	P. O. Box 92957, Mail Stop M4/972	Los Angeles	CA	90009	(310) 336-1602
Geoffrey	Smit	Aerospace Corp.	3550 Aberdeen Ave, SE	Kirtland AFB	NM	87117	(505) 846-8407
Francis	Smith	USAF					
Ken	Smith	Rice University	Johnson Space Center/Mail Code NS3	Houston	TX	77058	(713) 483-3263
Robert	Smith	NASA	555 Technology Square, Mail Stop MS-23	Cambridge	MA	02139	(617) 258-2126
Michael	Socha	Draper Lab	P. O. Box 92957, Mail Stop M4/983	Los Angeles	CA	90009	(310) 336-9235
Nick	Sramek	Aerospace Corp.	P. O. Box 534, S-751 21	Uppsala	Sweden		(46) 18 18 3015
Carola	Strandman	Uppsala University	42 Acorn Ct.	Sterling	VA	20164	(703) 709-9749
Robert	Stroud	Aerospace Corp.					

James	Stuart	Teledesic	P. O. Box 92957, Mail Stop M2/250	Los Angeles	CA	90009	(310) 336-7389
Wayne	Stuckey	Aerospace Corp.	Thornton Hall	Charlottesville	VA	22903	(804) 982-2206
Kevin	Sullivan	Univ of Virgina	P.O. Box 92957, Mail Stop M2/248	Los Angeles	CA	90009	(310) 336-5049
David	Sutton	Aerospace Corp.	Centre Bld, Suite 1450, 8455 Colesville Rd	Silver Spring	MD	20910	(301) 427-2121
Hilmer	Swenson	Aerospace Corp.	Johnson Space Center, Mail Code EV	Houston	TX	77058	
Bill	Teasdale	NASA	Johnson Space Center, Mail Code EA63	Houston	TX	77058	
Charles	Teixeira	NASA	18 Av Edouard Belin, Mail Stop 2532	Toulouse	France	31055	(33) 61-27-3628
Michel	Thoby	CNES					
Daniel	Thunnissen	JPL (NASA)	Jamison Rd	East Aurora	NY	14052	(716) 687-4281
Mark	Tilden	LANL	8801 Edward Gibbs Place	Alexandria	VA	22309	(703) 360-5969
Hans	Toews	MOOG Inc	Johnson Space Center, Mail Code NS5	Houston	TX	77058	
William	Tolles	Miniaturization S&T					
John	Trainor	NASA	2900 Semiconductor Dr, Mail Stop A1500	Santa Clara	CA	95052	
William	Trimmer	Belle Mead Research	P/O. Box 217	Euicmed		7500AE	
Chris	Tubis	National Semiconductor	P.O. Box 655936, Mail Stop 134	Dallas	TX	75265	(214) 995-6968
A.	van den Berg	MESA Research Institute	Rue Jaquet Drpz 1, P.O. Box 3	Neuchatel	Switzerland		(41) 38-205-387
Paul	van der Wagt	Texas Instruments					
Bart	van der Schoot	University of Neuchatel	865 Chang Ning Road	Shanghai	P.R. China	200050	(086) 21-625-11070
Steven	Venema	Univ of Washington					
Roland	Walker		11550 Water Oak Ct	Woodbridge	Va	22192	(703) 590-7000
Wei Yuan	Wang	Chinese Acad of Sci	Johnson Space Center/CB	Houston	TX	77058	(713) 244-8915
Robert	Warrington	Louisiana Tech University	4555 Overlook Ave, S.W., Mail Stop 8220	Washington	D.C.	20375	(202) 767-1714
Steve	Watson	Scientific Res & Dev Office	Johnson Space Center, Mail Code SD5	Houston	TX	77058	
Mary Ellen	Weber	NASA					
Ronald	Weber	Allied Signal	Johnson Space Center, Mail Code ER	Houston	TX	77058	
Laurie	Webster	NASA	4800 Oak Grove Dr	Pasadena	CA	91009	(818) 354-2632
Mark	Webster	Jet Propulsion Lab					
Dennis	Wells	NASA	12214 Lakewood Blvd, Mail Stop AA87	Downey	CA	90241	(310) 922-0638
John	West	Jet Propulsion Lab					
Isaiah	White	Boeing					
David	Whittle	NASA					
Henry	Woo	Rockwell					
Lars	Yoder	Texas Instruments					
Peter	Zdebel	Motorola					
Mark	Zdeblick	Redwood Microsystems					