L861892-9

NASW-4066

IN-91

69614-CR

P.80

MECA WORKSHOP ON

DUST ON MARS II

MECA

(NASA-CR-180545) MECA WORKSOP ON DUSI CN MARS 2 (Lunar and Planetary Inst.) 80 p
Avail: NTIS HC A05/MF A01 CSCL 03B

N87-30236

Unclas
G3/91 0069614

LPI Technical Report Number 86-09

LUNAR AND PLANETARY INSTITUTE 3303 NASA ROAD 1 HOUSTON, TEXAS 77058-4399

MECA WORKSHOP ON

DUST ON MARS II

Edited by
Steven Lee

Sponsored by
The Lunar and Planetary Institute
National Aeronautics and Space Administration

Hosted by
Arizona State University
Tempe, Arizona
February 24-26, 1986

Lunar and Planetary Institute   3303 NASA Road 1   Houston, Texas 77058-4399

LPI Technical Report 86-09

Compiled in 1986 by the
LUNAR AND PLANETARY INSTITUTE

The Institute is operated by Universities Space Research Association under Contract NASW-4066 with the National Aeronautics and Space Administration.

This report may be cited as:
Lee S. ed. (1986) *MECA Workshop on Dust on Mars II.* LPI Tech. Rpt. 86-09. Lunar and Planetary Institute, Houston. 79 pp.

Papers in this report may be cited as:
Author A. A. (1986) Title of paper. In *MECA Workshop on Dust on Mars II* (S. Lee, ed.), pp. xx–yy. LPI Tech Rpt. 86-09. Lunar and Planetary Institute, Houston.

This report is distributed by:

LIBRARY/INFORMATION CENTER
Lunar and Planetary Institute
3303 NASA Road 1
Houston, TX 77058-4399

Mail order requestors will be invoiced for the cost of postage and handling.

Contents

Program

Monday, February 24
9:00 a.m. - 12:15 p.m.

Session I: Physical and Chemical Properties of Dust
Chairman: Alan Peterfreund

J. Tillman
Martian atmospheric dust over three years as inferred from Lander meteorology measurements

T. Martin
Infrared opacity of dust in the Mars atmosphere

R. Kahn
The varying properties of martian aerosols

F. Jaquin
The vertical distribution of aerosols during the growth phase of a global dust storm

J. Gooding
Condensation of frosts on martian dust particles

D. Colburn, J. B. Pollack, and R. M. Haberle
Influence of dust on water condensation at Mars

A. Banin
Constraining Mars-dust mineralogy on the basis of Viking biology simulations and Mars spectral reflectance

R. Morris, H. V. Howard, Jr., D. G. Agresti, and J. A. Newcomb
Spectral properties of "dust" produced in the dry valleys of Antarctica; A martian analog?

Monday, February 24
1:30 - 4:30 p.m.

Session II: Atmospheric Dust Transport and Redistribution
Chairman: Robert Haberle

L. Martin and P. B. James
Dust clouds of a different nature

P. James and L. J. Martin
Interannual variability in Mars' South Polar recessions and possible correlations with major duststorms

D. Paige
Dust in the spring season polar atmospheres: A north-south comparison

J. Barnes and J. L. Hollingsworth
Implications of a polar warming for dust transport into the North Polar region of Mars

R. Haberle
The development of global dust storms on Mars: The role of the mean meridional circulation

R. Zurek and R. M. Haberle
Cross-equatorial transport during a martian great dust storm

J. Pollack, R. M. Haberle, J. Barnes, M. Schlesinger, R. Zurek, C. B. Leovy, J. White, J. Schaeffer, and K. Bilski
Influence of dust on the general circulation of the martian atmosphere

Tuesday, February 25
9:00 a.m. - 12:15 p.m.

Session III: Dust Erosion, Deposition, Sources, and Sinks

Chairman: Peter Thomas

R. Greeley
Toward an understanding of the martian dust cycle

R. Arvidson
On the rate of formation of sedimentary debris on Mars

H. Moore
Martian miniature landslides, Mutch Memorial Station

J. Zimbelman
Physical properties of aeolian features in the Elysium-Amazonis region of Mars

B. Jakosky and P. Christensen
Are the Viking Lander sites representative of the surface of Mars?

P. Christensen
Surface albedo variations on Mars: Implications for yearly dust deposition and removal

S. Lee
Viking observations of regional sources and sinks of dust on Mars

L. Roth, R. S. Saunders, and T. W. Thompson
Radar reflectivity of a variable dust cover

T. Thompson, R. F. Jurgens, L. Roth, and L. Robinett
Possible Goldstone radar observations of Mars: 1986 and 1988 oppositions

Tuesday, February 25
1:30 - 4:30 p.m.

Session : General Discussion/Overview/Questions

(1) How many dust components are there on Mars, and what are their properties?
(Discussion Leader: Alan Peterfreund)

(2) How does the global atmospheric circulation affect dust redistribution?
(Discussion Leader: Robert Haberle)

(3) How do dust sources and sinks vary with time?
(Discussion leader: Peter Thomas)

Summary of Technical Sessions and Discussions

A MECA workshop, "Dust on Mars II," was sponsored by NASA through the Lunar and Planetary Institute and hosted by Arizona State University on February 24–25, 1986. Following the recommendations of the participants in the previous MECA "dust" workshop (held at ASU on February 4–5, 1985), the goal of this workshop was to discuss the following questions:

(1) How many components of dust are there on Mars, and what are their properties?
(2) How is dust ejected from the surface into the atmosphere?
(3) How does the global atmospheric circulation affect the redistribution of dust?
(4) Are there sources and sinks of dust? If so, how do they vary with time?

Forty-three people attended the meeting and engaged in the discussion sections; of these, twenty-four presented summaries of their dust-related research.

The first workshop session considered the physical and chemical properties and the distribution (both temporal and spatial) of dust and condensates. Several "observational" presentations focused on detailed analyses of Viking data. Pressure data recorded over four martian years at Viking Lander 1 and two years at Viking Lander 2 have been studied by J. Tillman with particular attention to the major transient events related to two global dust storms in 1977 (Year 1) and one in 1982 (Year 4). The 1982 event was similar in season to the first 1977 storm (L_s ~200°) but was more intense than this storm, appearing instead to be similar to the second 1977 storm. The 1982 storm was noted as having a unique pressure signature, suggesting a second surge of intense activity a week after the initial change in pressures was noted. T. Martin presented a dust opacity history based on a global data analysis using the 7, 9, and 15 μm data collected by the Orbiters' Infrared Thermal Mappers (IRTM). The resultant maps of opacity depict the spatial and temporal variations of atmospheric dust loading throughout the Viking missions. Detailed maps during the 1977 dust storm season suggest large local variability of opacity in many areas.

Evidence for H_2O ice in the Mars atmosphere has been inferred from a variety of optical observations using the Landers and Orbiters. Limb observations show that a detached haze fluctuates in height and opacity with time and location. R. Kahn modelled the properties of these hazes during the non-dust storm periods with consideration of mean optical depth, condensation level, mean particle size, and eddy diffusion. Correlation of the cloud height with MAWD column water vapor abundance implies that water controls the location of the cloud base. The mean particle sizes increase as condensation heights decrease in such a way that the average fall time is always about 1/4 day. F. Jaquin reported on analysis of the limb observations made by the Orbiter cameras over three Mars years. Profiles of reflectance vs. elevation above the surface show the buildup of dust with height as a function of time and latitude. During the 1977 dust storm season, continuous haze up to 50 km altitude is seen in the northern hemisphere with a weak detached haze above that. The detached haze (probably condensates) is better defined in the southern hemisphere.

The second set of speakers reported on "laboratory and theoretical" results. J. Gooding has been performing experiments and modelling the heterogeneous nucleation of condensates on mineral grains. Favorable conditions for condensation appear to be related to crystallographic compatibility (with respect to structure and number of "active" sites) and the nature of chemical bonds. Details for mineral classes and for given minerals, especially clays, show the complexities of the nucleation process as a function of condensate type. Furthermore, as nucleation occurs the initial condensates will change conditions associated with subsequent nucleation. D. Colburn discussed the influence of dust on water condensation using Lander optical depth measurements. Both seasonal and diurnal effects appear to be factors in the amount of water condensation. Also, most condensation appears to occur aloft as opposed to near-surface, and is strongly related to dust abundance. Simulating results from the labeled-release biology experiment aboard the Landers, A. Banin inferred properties of martian dust. Smectite minerals (montmorillonite and nontronite) that were ion-modified were found to best simulate the kinetic results of the biology experiment.

The ion-modification served to exchange calcium and sodium ions with iron. This modification is consistent with Lander chemical analyses. These experiments suggest that: (1) the pH of martian soil is less than 4.5; (2) $CaCO_3$ probably cannot be in soil at concentrations larger than ~1.0%; (3) montmorillonite is apparently essential for causing rapid decomposition on clay surfaces. According to results presented by R. Morris, the spectra of martian bright regions suggest that "dust" produced in the dry valleys of Antarctica may be a reasonable martian analogue. Weathering on Mars, assuming analogous processes to those in Antarctica, results in more highly oxidized material, but in general appears to be similar to that for ferric products that were studied. Further, the presence of hematite as a component of martian dust is strongly suggested.

The second workshop session addressed topics of atmospheric dust transport and redistribution. L. Martin led off the session discussing equatorial dust clouds imaged by Viking that differed from other dust clouds because of their persistent association with a small canyon. Downslope katabatic-type winds may be responsible for the clouds' origin; in some cases they appeared to flow into the canyon floor. The potential of the telescopic record for revealing interannual variation in the CO_2 cycle was the subject of the presentation given by P. James. The signature of this variability appears as differences in the retreat rate of the south polar cap (readily observed during favorable oppositions). Comparison of data for the 1956, 1971, 1973, and 1977 dust storm years shows some variability with the cap retreating slower than the median in 1971 and 1977 and faster than the median in 1956 and 1973. The extent to which this variability would affect pressure variations is evidently small, however, given the high degree of repeatability in the seasonal pressure data from the Lander sites. D. Paige discussed his analysis of IRTM data at the north and south poles during the spring season. By constraining a radiative equilibrium model he developed to reproduce the brightness temperatures in the various IRTM channels, he concluded that dust in both polar regions is similar in composition to that observed in the tropics (the so called "Toon-Pollack" dust). He also found that unless the dust was confined to the lowest 5 km, it was not possible to reproduce the observed 15 micron brightness temperatures. This problem was even more difficult in the north where additional cooling, perhaps due to water ice, is needed.

J. Barnes focused attention on the transport implications of the warming of the north polar atmosphere during the second global dust storm of 1977. Based on models that follow the actual motion of air parcels, the observed warming must be accompanied by sinking motion; to conserve mass, poleward transport from lower latitudes is required. Barnes also discussed other modelling that implies that vertically-propagating forced stationary waves may play an important role in the heat budget of the polar atmosphere during global dust storms. In the next paper, R. Haberle presented a mechanism for interannual variability of global dust storms. His nearly inviscid zonally-symmetric circulation model suggests the following scenario: First, global dust storms transport dust into the northern hemisphere. This dust is then available for raising by baroclinic waves in the following years. During the post-dust-storm years the cross-equatorial Hadley circulation is therefore weakened due to heating by dust suspended in its descending branch. When the supply of dust is exhausted, the full strength of the Hadley circulation, and the likelihood of global dust storms, is restored. R. Zurek then spoke about how tidal forcing of the mean meridional circulation may break up the otherwise smooth structure of the cross-equatorial Hadley cell. During relatively dusty periods, tides produce significant accelerations of the zonal and meridional wind fields, and these accelerations have considerable vertical structure. When included in Haberle's 2D model, several vertically-stacked Hadley-type cells were produced. Zurek suggested that by changing the structure of the Hadley circulation, the tides could effectively limit the flux of dust into the northern hemisphere during a global dust storm, explaining the decay patterns seen at the Lander sites. Completing the "theoretical" presentations of this session, J. Pollack gave a progress report on the Mars general circulation model (GCM) and presented some preliminary results. Work on the GCM during the past several years has focused on including dust as a radiatively active constituent, generalizing the model to an arbitrary number of layers, and improving the numerical methods for calculating infrared fluxes. All of these changes have been implemented and tested, and the current effort has shifted to determining the model's performance in a variety of configurations and examining how dust affects the behavior of large-scale atmospheric eddies.

During the final session of the workshop, the participants turned their attention to the study of dust sedimentation and erosion on Mars. The research that was discussed is being conducted in three general areas: photointerpretation of sedimentary and erosional features, surface properties determined from thermal inertia, albedo, and radar data, and laboratory and theoretical work on the processes involved in wind transport of dust. R. Greeley tied together the various laboratory and theoretical studies applicable to the martian dust cycle. In particular, the continuing question of the role of volatiles in any aggregation of dust particles and in ejection of dust from the surface was considered. R. Arvidson presented results of a study of the morphology and statigraphy of south polar pitted and layered terrains that indicates deposition in two very different phases extending over about half of martian history. The radically different layered and unlayered pitted deposits suggest substantial differences in sedimentary regimes. At another size scale, H. Moore reported that the morphology of small landslides at the VL1 site is diagnostic of surface layers of very low cohesiveness that are probably recent aeolian deposits. Slope failure of such layers may be an important mechanism for producing a surface that is not in equilibrium with martian wind regimes, thereby enhancing the possibility of aeolian erosion.

Several specific regions of the planet were discussed in detail. The physical properties of aeolian features in the Elysium-Amazonis area were examined by J. Zimbelman. Analysis of the thermal inertia of wind streaks in the region suggests a long-lived movement of fine to medium sand-sized materials in the observed depositional dark streaks and possible dust deposits of significant thickness accumulating in bright streaks. B. Jakosky reported on a study relating the characteristics of the Lander sites to the global data on thermal inertia, albedo, and color; the sites apparently do not match the general planet-wide trends in these data sets. In particular, the VL1 site is intermediate between bright areas assumed to accumulate dust and those dark areas that are thought to be swept relatively clear. P. Christensen discussed the albedo variation of several areas on the planet using the IRTM albedo channel. He found that northern hemisphere dark regions showed a long period of progressive dust removal after global dust storms, while southern hemisphere dark regions show a rapid return to pre-storm albedos. S. Lee investigated regional sources and sinks using IRTM albedo data and Orbiter images. The patterns of temporal changes indicate that Syrtis Major, through much of the year, acts as a source region for dust redeposited in neighboring Arabia. In the Solis Planum region, removal of dust is restricted to periods of major dust-storm activity.

The ability of radar to detect seasonal variability in surface dust cover was examined by L. Roth. A theoretical study of radar reflectivity as a function of dust layer thickness concludes that time-variable reflectivity data would be consistent with significant changes in dust layer thickness only under very restricted conditions of layer geometry and continuity. The session was concluded with T. Thompson's review of future opportunities for radar observations of Mars. The 1986 and 1988 oppositions will greatly extend the southern hemisphere coverage available from previous years, while upgrades in the Goldstone radar should significantly improve resolution and allow both X- and S-band observations.

The final afternoon of the workshop consisted of an open discussion of the major questions posed during the preceding two days. Much of the discussion centered around the possibilities for interdisciplinary cooperation. For example, observations of regional dust sources and sinks, and tracking the location, timing, and growth of local and great dust storms, should provide useful input and constraints for atmospheric circulation models. The mineralogical properties of dust may greatly influence the formation of condensates; hence, knowledge of such properties may play major roles in modelling the production of atmospheric hazes and polar deposits. It was agreed that defining the frequency and characteristics of local and "global" dust storms, the distribution (temporal and spatial) of dust on the surface, and the physical and chemical properties of dust along with inclusion of dust related effects in atmospheric circulation models, are all areas in which continued research and communication among the various interested groups would be fruitful for the entire community. The research discussed at the workshop is clearly moving in the direction of posing questions for the Mars Observer mission, while showing that Viking Orbiter and Earth-based data provide an excellent basis for formulation of Mars Observer experiment strategies.

ABSTRACTS

ON THE RATE OF FORMATION OF SEDIMENTARY DEBRIS ON MARS; Raymond E. Arvidson, McDonnell for the Space Sciences, Department of Earth and Planetary Sciences, Washington University, St. Louis, MO 63130

We now have enough information to place rather meaningful constraints on the current rates of aeolian redistribution of fine grained debris on Mars. For example, from the three years of Viking Lander 1 observations, typical values of sediment redistribution are micrometers per year, although centimeters of loose, disturbed material were removed during a brief interval in the third year (1). Rates of erosion of rocks were too small to be observed, either by tracking changes in morphologies or by tracking changes in rock photometric properties. Consideration of the large number of pristine-looking, small bowl-shaped craters at the Lander 1 site suggests a rate of rock breakdown and removal of only meters over the lifetime of the surface. Thus, averaged over the lifetime of the Chryse Plains, rock breakdown and removal has been meters per billions of years, orders of magnitude lower than the micrometers per year for soils (2). Most of the equatorial regions of Mars likewise preserves ancient surfaces, with craters even at small sizes seemingly in production. Thus, rock breakdown and removal over the equatorial terrains has been quite small for much of geological time. Even the fretted terrains have probably been inactive for a long while, considering that the crater abundances in much of fretted terrain are second in abundance only to the cratered terrains (3). The younger fretted areas seem to be embayed by younger deposits (4) or to be in areas of relatively recent tectonic activity, such as the chaotic terrains. By areal extent, most of the equatorial terrains of Mars have been subjected to very low erosion rates, significantly less than the debris redistribution rates witnessed by Viking Lander 1. Thus, as noted by (2), differential aeolian erosion on Mars is a major geological process, with debris deposits accumulating and eroding to depths of hundreds of meters over geological time, while rock breakdown has been occurring in most regions at vanishingly small rates. Given the low rates of production of new debris, one is forced to conclude that Mars has had a decidedly non-linear history of debris production. In particular, most sedimentary debris must have been produced relatively early, perhaps in the first billion years of geological time. Impact cratering, production of volcanic ash, and chemical corrosion may all have been important debris-forming processes. Given that the mineralogy of martian debris has apparently not come into equilibrium with the present ambient conditions (5), we may have some chance of deciphering the relative importance of various debris forming processes. For example, if analyses of Mars Observer Imaging Spectrometer data show a dominance of palagonitic materials, then early volcanism would be indicated.

REFERENCES

1. Arvidson, R. and others, 1983, Science, v.222, p.463-468.
2. Arvidson, R. and others, 1979, Nature, v.278, p.533-535.
3. Coradini, M. and R. Arvidson, 1976, Proceed., Int. Colloq. Planetary Geology, Geol. Romana, v.15, p.377-381.
4. Frey, H. and A. Seneniuk, 1985, Geol. Soc. Am., Abstracts, p.586.
5. Gooding, J., 1978, Icarus, v.33, p.483-513.

CONSTRAINING MARS-DUST MINERALOGY ON THE BASIS OF VIKING BIOLOGY SIMULATIONS AND MARS SPECTRAL REFLECTANCE.

A. Banin, Hebrew University, P.O. Box 12, Rehovot, Israel

Among the prime, unresolved and often debated issues related to Mars is the question of the composition and properties of its fine, dusty surface materials. Unfortunately, no direct and definitive mineralogical analysis has been done on the soil and dust of Mars as yet. Various indirect approaches were used, however, to arrive at Mars dust mineralogy. These included direct chemical analyses, chemical modeling, simulations of the chemical interaction with water and solutes, and a multitude of spectroscopic observations from Earth, orbiters and landers. In the following we will briefly review these, specifically emphasizing the chemical simulations and present our current thinking on the mineral composition of the Martian dust.

Chemical composition of the Martian fine surface material: Quite probably, the most directly related available evidence regarding soil and dust mineralogy, is the inorganic chemical analysis (ICA) obtained by X-ray fluorescence analyzers aboard the Viking landers. The analyses have shown elemental abundances, (given as oxides), of 40-45% SiO_2, 7-7.5% Al_2O_3, 17-19% Fe_2O_3, 5-7% MgO, 5-6% CaO, 0.4-0.7% TiO_2, traces of K, 6-9% SO_4 and 0.3-0.9% Cl (1). The chemical composition of the soil in the two Viking landing sites, and of different samples taken from each site, was found to be strikingly similar. On the basis of the chemical data, several combinations of minerals have been suggested initially (2,3,4). These characteristically contained about 60-80% clay minerals of the smectite group, mixed with various soluble salts such as kieserite ($MgSO_4$) and halite (NaCl). Suggested accessory minerals were: iron oxides(Fe_2O_3), calcite ($CaCO_3$), and quartz (SiO_2).

Simulation of the Viking Biology experiments: The Viking Biology experiments turned out to be (and still are) an important and unique source of chemical information on Martian dust and soil because of the direct "wet chemistry" that was conducted on the soil and the information obtained on the interaction of the Mars minerals with water, salts and organics. Although no clear evidence for life was found, the Viking Biology experiments showed that the minerals in the Mars soil were chemically reactive and capable of decomposing various organic acids, catalyzing photochemical fixation of CO_2/CO, and releasing oxygen upon wetting (5,6).

The simulation of the Viking Biology experiments, conducted between 1977 and 1979 gave somewhat diverse results. Levin and his colleagues reported that the Mars soil analogs provided by the ICA Team did not simulate the Labeled Release (LR) Viking experiment (7). On the other hand, Banin and his colleagues, using ion-modified clays, succeeded in the kinetic simulations of this experiment (8,9,10). The clay modification by Banin et al. (11,12) involved the exchange of calcium and sodium ions, the primary adsorbed ions of clays on Earth, with iron, a ubiquitous and abundant element in the Martian soil. In addition to affecting the chemical and catalytic properties of the clay, this exchange resulted in a pronounced color change from gray to reddish-orange, making the clay very similar in color and spectral characteristics to the Martian soil (13,14).

Our detailed experimental studies showed (10,13,14,15) that among the many mixtures of clays and various salts whose combinations give elemental compositions similar to the Mars soil, only iron smectites, and particularly montmorillonite, reasonably simulated the chemical reactivity in the Viking Biology LR experiment. These simulations showed that if any CO_2 was to be released in the LR Viking experiment, calcite ($CaCO_3$) could not be present in the dusty surface materials at concentrations higher than about 0.5%, and the pH of the soil could not be above ca. 5.

In separate simulation experiments, Hubbard (16), using the modified clays of Banin, detected photochemical fixation of atmospheric ^{14}C, and its incorporation into organics in the soil. This photochemical fixation was similar to the observations on Mars in the Viking Pyrolytic Release (PR) experiment. Other candidates for Mars soil components which were suggested using Viking Biology results as the guideline were carbon suboxide (17) and various mafic silicate minerals (18).

Spectroscopic evidence: Earth-based visible and near infrared reflectance data were accumulated through telescopic measurements of McCord's group in Hawaii (19,20). Since the termination of the Viking extended mission in 1978, this spectral information became the major if not the sole source of new experimental data on Mars soils mineralogy. The Mars reflectance spectrum was found to show opacity and band saturation (no reflectance) in the near UV-visible range (0.3-0.5 µm) with increasing transparency (increasing reflectance) in the long-wavelength visible and near infrared range (0.5-0.9 µm). However, it was lacking any clear and pronounced absorption features throughout this range. Spectra of crystalline, pure iron oxides such as hematite and goethite were found by Singer (21) and by Sherman et al. (22) to exhibit significant deviations from the Mars reflectance spectrum measured from Earth. However, several iron-containing minerals have been found to bear significant similarity to the Mars reflectance spectrum. These include amorphous iron-silica gels (23,24), palagonite (21), iron-containing aluminum oxy-hydroxide (25), and iron saturated montmorillonite clay (14). A feature common to all of the candidates is that they contain ferric iron in a matrix of oxy-hydroxide which is not well crystallized, or has only short-range ordering. Unfortunately, on the basis of the spectral comparison alone, there is yet no obvious way to choose among the various candidates, and additional properties and characteristics of the Martian soil and its Earth analogs have to be studied in order to further constrain the possible Martian mineralogical composition.

References: 1) Clark, B.C. et al. (1982) JGR 87, 10,059-10,067. 2) Toulmin, P. Jr. et al. (1977) JGR 82, 4625-4634. 3) Clark, B.C. et al. (1977) JGR 82, 4577-4594. 4) Clark, B.C. et al. (1976) Science 194, 1283-1288. 5) Klein, H.P. (1977) JGR 82, 4677-4680. 6) Klein, H.P. (1979) Rev. Geophys. Space Phys. 17, 1655-1662. 7) Levin, G.V. and Straat, P.A. (1977) Biosystems 9, 165-174. 8) Banin, A. and Rishpon, J. (1978) COSPAR, Life Sci. Space Res. XVII, 59-64. 9) Banin, A. and Rishpon, J. (1979) J. Mol. Evol. 14, 133-152. 10) Banin, A. et al. (1981) Proc. Third Intern. Coll. Mars Publ. 441 Lunar Planet. Inst., Houston, pp. 16-18. 11) Banin, A. (1973) U.S. Patent 3, 725,528, 14 pp. 12) Gerstl, Z. and Banin, A. (1980) Clays & Clay Min. 28, 335-345. 13) Banin, A. et al. (1984) Proc. Lunar Planet Sci. Conf. (15th) 31-32. 14) Banin, A. et al. (1985) JGR 90, (supplement) C771-C774. 15) Banin, A. and Margulies, L. (1983) Nature 305, 523-526. 16) Hubbard, J.S. (1979) J. Mol. Evol.14, 211-222. 17) Oyama, V.I. et al. (1978) COSPAR, Life Sci. Space Res. XVI, 3-8. 18) Huguenin, R.L. (1982) JGR 87, 10,069-10,082. 19) McCord, T.B. et al. (1982) JGR 87, 3021-3032. 20) Singer, R.B. et al. (1979) JGR 84, 8415-8426. 21) Singer, R.B. (1982) JGR 87, 10,159-10,168. 22) Sherman, D.M et al. (1979) JGR 84, 8415-8426. 23) Evans, D.L. and Adams, J.B. (1979) Proc. Lunar Planet. Sci. Conf. (10th), 1829-1834. 24) Evans, D.L. and Adams, J.B. (1980) Proc. Lunar Planet. Sci. Conf. (11th), 757-763. 25) Morris, et al. (1983) Proc. Lunar Planet. Sci. Conf. (14th), 526-527.

This work was supported in part by NASA grant NSG-7512 and NASA Cooperative Agreement NCC 2-352 (Ames Research Center).

IMPLICATIONS OF A POLAR WARMING FOR DUST TRANSPORT INTO THE NORTH POLAR REGION OF MARS;

J.R. Barnes and J.L. Hollingsworth,
Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331

The layered terrains in the polar regions of Mars have been widely interpreted as records of quasi-periodic variations in the Mars climate system driven by orbital element variations (see, e.g., Pollack and Toon, 1982). It has been suggested that the polar laminae in the north are being formed at present and that perhaps the bulk of the deposited dust and water ice is transported to the pole during global dust storms (Cutts, 1973; Pollack et al., 1979), which currently occur during the growth phase of the seasonal CO_2 frost cap. Pollack et al. hypothesized that CO_2 and water ice condensation onto dust grains would consitute a very effective mechanism for removing the dust to the surface in the polar regions. Paige (1985) carried out analyses of Viking IRTM observations and concluded that deposition of dust did occur (probably in conjunction with CO_2 condensation in the atmosphere) in the region of the north residual cap during the winter solstice (the second) dust storm of 1977. The magnitude of this deposition is uncertain, however, Jakosky (1983) and Jakosky and Martin (1984) have argued that the transport of dust (and water) onto the north residual cap during global dust storms is insignificant.

The deposition of dust onto the seasonal polar cap is also of considerable importance, because of the effects on the polar cap albedo and thus the radiation budget of the subliming cap. Additionally, the dust loading of the polar atmosphere is significant because of its effects on the radiation budget of the cap: the enhanced emissivity of the atmosphere implies enhanced downward IR fluxes.

The intense warming of the polar atmosphere which was observed by the Viking IRTM during the winter solstice dust storm of 1977 (Martin and Kieffer, 1979; Jakosky and Martin, 1984) must have been the result of substantial dynamical transports of heat into the north polar region. The warming at the pole (~50 K in magnitude in the vicinity of the 25 km level) must have been produced by compressional heating associated with downward - in a Lagrangian sense - motion of air parcels. An accompanying poleward flow is then also implied. Such a Lagrangian motion pattern implies poleward and downward transport of tracers such as dust and water.

Numerical simulations conducted with a simplified beta-plane wave-mean flow model strongly suggest that the polar warming could have been induced by a forced, vertically propagating, planetary-scale (zonal wavenumber 1) wave (Barnes and Hollingsworth, 1985). The dynamics of such a warming are most clearly viewed in the so-called transformed Eulerian-mean system, instead of the traditional Eulerian one. In the transformed system, the warming is produced by downward vertical motions - rather than eddy transport of heat. [The transformed Eulerian-mean vertical motion field is generally a good approximation to the true Lagrangian-mean vertical motion field - unlike the Eulerian-mean verical motion field.] The mean meridional circulation in the transformed Eulerian-mean system, the so-called residual mean meridional circulation, is characterized by poleward and downward flow into the polar interior during the simulated warmings. This poleward and downward flow takes place throughout a very deep region as the warming occurs, with the much weaker return flow confined to low levels. The residual mean meridional circulation should constitute a good first approximation appropriate to the transport of tracers (Holton, 1981), thus implying transport into the polar interior of a tracer initially located largely equatorward of the cap boundary (e.g., dust or water).

Full tracer (dust) transport simulations corresponding to the dynamical warming simulations are planned. These will allow a quantitative examination of the transport of dust into the polar region in association with a polar warming event.

REFERENCES

Barnes, J.R. and J.L. Hollingsworth (1985) Dynamical modeling of a Martian polar warming. Bull. Amer. Astron. Soc., 17, 3: 733.

Cutts, J.A. (1973) Nature and origin of layered deposits of the Martian polar regions. J. Geophys. Res., 78, 4231-4249.

Holton, J.R. (1981) An advective model for two-dimensional transport of stratospheric trace species. J. Geophys. Res., 86, 11, 989-11, 994.

Jakosky, B.M. (1983) Mars north polar cap: Net annual source or sink? Bull. Amer. Astron. Soc., 15, 846.

Jakosky, B.M. and T.Z. Martin (1984) Mars: Polar atmospheric temperatures and the net transport of water and dust. Unpublished manuscript.

Martin, T.Z. and H.H. Kieffer (1979) Thermal infrared properties of the Martian atmosphere. 2. The 15 μm band measurements. J. Geophys. Res., 84, 2843-2852.

Pollack, J.B., D.S. Colburn, F.M. Flasar, R. Kahn, C.E. Carlston, and D. Pidek (1979) Properties and effects of dust particles suspended in the Martian atmosphere. J. Geophys. Res., 84, 2929-2945.

Paige, D.A. (1985) The annual heat balance of the Martian polar caps from Viking observations. Ph.D. dissertation, California Institute of Technology, 207 pp.

Pollack, J.B. and O.B. Toon (1982) Quasi-periodic climate changes on Mars: A review. Icarus, 50, 2/3, 259-287.

DUST DEPOSITION AND EROSION ON MARS: CYCLIC DEVELOPMENT OF REGIONAL DEPOSITS

Philip R. Christensen, Dept. of Geology, Arizona State University, Tempe, AZ, 85287.

The cycle of dust erosion, transport, and deposition plays an important role in the evolution of the martian surface, both today and in the past. Thermal, radar, and visual remote sensing observations provide important constraints on the surface properties, and have been used to determine the location and physical properties of regional dust deposits. These deposits provide direct clues to the rate and history of dust deposition, and suggest that the martian surface is being actively reworked. The observed dust cycle may be directly related to the cycle of variations in Mars' orbit, with dust continually transported between hemispheres on time scales of 10^5 to 10^6 years.

Major dust deposits are currently located in three northern equatorial regions: Tharsis (-20°S to 59°N, 60° to 190°W), Arabia (-5°S to 30°N, 300° to 360°W), and Elysium (10° to 30°N, 210° to 225°W). They are covered by fine (~2-40 μm), bright (albedo > 0.27) particles, with fewer exposed rocks and coarse deposits than found elsewhere (1,2). These regions also have less bonded material exposed at the surface than found elsewhere (3). Dust deposits may have initially formed due to differences in wind velocity. Once initiated, the burial of sand and rocks would make removal of fines increasingly difficult, enhancing the rate of dust accumulation. Dust is currently deposited uniformly throughout the equatorial region at a rate of ~40 μm/global storm. Over geologic time the rate of accumulation may vary from 0 to 250 μm/year due to changes in atmospheric conditions produced by orbital variations (4). Dust deposited during global storms is subsequently removed only from dark regions, resulting a net accumulation in the low-inertia, bright regions. The evidence for the subsequent removal from dark regions comes from: 1) the observed higher dust content over dark regions than bright regions during clear periods; 2) the post-storm darkening of dark regions (5, this issue); 3) the removal of storm-deposited dust at the Viking lander 1 site during the year following the major 1977 dust storms (6); and 4) the historical persistence of classic dark features. Non-removal of dust from low-inertia, bright regions results in a net accumulation of dust in these areas.

The thickness of the current dust deposits can be estimated from thermal, radar, and visual observations.. The low thermal inertia of the deposits places a lower limit of ~0.1 m on their thickness, while the sparse but ubiquitous presence of exposed rocks and the degree of visible mantling indicates that the thickness is less than 5 meters (7). Dual-polarization radar observations of Tharsis reveal a very rough texture. These measurements can best be reconciled with the other observations by assuming that a relatively thin dust layer buries most of the surface rocks but that the radar samples through this layer to the rough surface below (1). The maximum thickness of this layer can be estimated from the electrical properties of rock powders. For dry powders, the radar energy should penetrate 5 to 50 wavelengths (0.6 to 6 m for the 12 cm radar used) before being attenuated by 1/e (8). The presence of adsorbed water will not affect these results (9). Thus, 12 cm radar could penetrate ~ 1 m, be reflected, and exit the surface with only a 1/e reduction in initial energy. This estimate puts a constraint of 1-2 m on the dust mantle thickness, which agrees well with the value obtained using the exposed rock abundance derived from thermal measurements.

Based on their thickness and rate of accumulation, the age of these deposits is 10^5-10^6 years, suggesting a cyclic process of deposition and removal (1). One possible cause may be cyclic variations in the magnitude and location of maximum wind velocities related to variations in Mars' orbit. At present, perihelion and maximum wind velocities occur in the south whereas regional dust deposits occur in the north, suggesting net transport from south to north. Orbital parameters oscillate with periods ranging from 5×10^4 to 10^6 years. The agreement between these periods and the dust deposit age suggests that there is a possible link. At different stages in orbit evolution, maximum wind velocities will occur in the north, with subsequent erosion and redistribution of the accumulated fines.

The model proposed here implies that material must be periodically removed from regional dust deposits in order to prevent long-term buildup of fine material in a given location. Thus, the proposed young age of these dust deposits requires mechanisms for eroding extensive deposits of fine material. The burial of sand and rocks makes it increasingly difficult to set particles in motion, with 20 μm particles requiring wind velocities a factor of 2 higher than the most readily moved particles (10). There are, however, several mechanisms for eroding these deposits, including erosion from the edge inward, and increased surface shear stress produced by increased winds (1).

A more plausable mechanism involves the formation of coarse particles as bonded aggregates of dust. Bonding of material has been observed at both Viking landing sites, and globally pervasive crusts have been detected from remote sensing observations (2). These crusts may form during transition periods between obliquity extremes as volatiles carrying adsorbed ions are cycled into and out to the surface (2). This mechanism would link the erosion of the deposits to the same process that leads to their formation. Thus, crusts could form after the deposition of a layer of fine material has been deposited after each obliquity cycle. Such crusted aggregates could provide a source of coarse particles that could be more readily moved by the wind, thereby providing a mechanism for eroding the underlying deposit of dust.

Perhaps the most important process may be surface erosion due to insolation-driven convective vortices (dust devils) of various scales. Experimental work suggests that vortices are very effective in raising particles of all sizes which can then be easily transported by much lower winds (11). There is direct evidence for the occurrence of dust devils on Mars, both through the passage of 5-950 m diameter vortices at the Viking lander sites (12), and through direct observation of dust devils from orbiter images (13). Of the 118 vortices observed at the lander sites, 4 had wind velocities greater than 30 m/sec, which may have been sufficient to raise dust (12). These vortices produced a factor of 2-3 enhancement in the ambient wind velocity. Because dust devils form due to convection in an atmosphere with a superadiabatic lapse rate, they are more frequent during periods of maximum surface heating. This erosion mechanism would be most effective over low-inertia surfaces during summer, and may provide a mechanism for eroding dust from the hemisphere that has the maximum solar insolation during summer. Thus, dust deposited in one hemisphere when the insolation maximum was in the opposite hemisphere would subsequently be eroded when obliquity variations caused the insolation maximum to reverse.

The presence and history of regional dust deposits provide evidence for cyclic processes of deposition outside the polar regions and support models of cyclic variations in martian climate over geologic time. Dust is continually eroded and redeposited, with the location of major deposits shifting on time scales of 10^5 to 10^6 years. In this model much of the uppermost martian surface is young and is being constantly reworked.

References

1) Christensen, P.R., 1986, *J. Geophys. Res.*, in press.
2) Christensen, P.R., 1983, *Icarus, 56,* 496-518.
3) Jakosky, B.M. and P.R. Christensen, 1986, *J. Geophys. Res.*, in press.
4) Pollack, J.B. and O.B.Toon, 1982, *Icarus, 50,* 259-287.
5) Christensen, P.R., 1986, this issue.
6) Arvidson, R.E., E.A. Guinness, H.J. Moore, J. Tillman, and S. D. Wall, 1983, *Science, 222,* 463-468.
7) Harmon, J.K., D. B. Campbell, and S.J. Ostro, 1982, *Icarus, 52,* 171-187.

8) Campbell, M.J. and J. Ulrichs, 1976, *J. Geophys. Res., 84*, 2795-2820.
9) Olhoeft, G.R. and D.W. Strangway, 1974, *Geophys. Res. Lett., 1,* 141-143.
10) Greeley, R., R. Leach, B. White, J. Iversen, and J. Pollack, 1980, *Geophys. Res. Lett., 7,* 121-124.
11) Greeley, R., B.R. White, J.B. Pollack, J.D. Iversen, and R.N. Leach, 1981, *Spec. Pap. Geol. Soc. Am., 186,* 101-121.
12) Ryan, J. A. and R. D. Lucich, 1983, *J. Geophys. Res., 88*, 11,005-11,010.
13) Thomas, P. and P. Gierasch, 1986,*Science, 230*, 175-177.

SURFACE ALBEDO VARIATIONS ON MARS: IMPLICATIONS FOR YEARLY DUST DEPOSITION AND REMOVAL

Philip R. Christensen, Dept. of Geology, Arizona State University, Tempe, AZ, 85287

Dust deposition and removal is an important process in the development and modification of the martian surface. Mars has been known to have variable surface markings from the earliest telescopic views of the planet. These changes have since been seen to be related to aeolian activity, primarily through the reworking of bright dust deposited following major global dust storms (1,2). Viking Infrared Thermal Mapper (IRTM) observations of albedo have also revealed significant changes in surface brightness through time, again primarily associated with major global dust storms (3,4). All of these observations indicate that there is a significant amount of dust that is deposited during the decay of global storms which is subsequently reworked and redistributed. The purpose of this study is to determine the degree, spatial distribution, and timing of the deposition and removal of dust-storm fallout, and to relate the current patterns of dust deposition and removal to the long-term evolution of the martian surface.

A model has been proposed (5, 6 this issue) for the development of regional dust deposits that form through the preferential accumulation of dust-storm fallout into specific northern hemisphere regions. In this model, dust is deposited uniformly during the decay phase of each major storm, but is subsequently removed only from regions that are seen today as classic dark areas. Thus, dark regions remain unmantled by dust, whereas bright regions have developed a 1-2 m thick mantle of fine, bright dust (5). This model can account for the high thermal inertia (coarse) material observed in dark regions, together with their relatively high rock abundance (7), and low albedo. Conversely, bright regions have fine particles (5-40 µm) and fewer exposed rocks, presumably due to mantling of the coarse material by dust.

In order to directly observe the seasonal changes in surface brightness associated with dust deposition and removal, the albedo of specific regions in both hemispheres has been determined through time. The IRTM data were collected into 1° latitude by 4° longitude bins, at 3 hour intervals for each 10° of L_s. Using these data, the albedo changes for a given area have been investigated from the beginning of the Viking mission (L_s 84°), through the first (L_s 190-240°) and second (L_s 270-340°) global dust storms that occurred in 1977. Global data are available through L_s 120° of the second year, allowing a year to year comparison of surface albedo.

The albedo variations as a function of season are shown in Figure 1 for representative bright and dark regions. All of the areas studied show a marked increase in brightness associated with the two global storms, due primarily to the presence of dust in the atmosphere. The increase in brightness, even for bright regions, indicates that the albedo and scattering phase function of suspended dust varies from dust on the surface. The maximum brightness at the peak of the second storm was nearly equal for most bright and dark regions, indicating that the atmospheric dust was optically thick. For some dark regions, however, such as Solis Planum, the albedo remained relatively low even at the height of the storm activity, suggesting that the atmospheric dust was not globally uniform nor well mixed. Many areas show a non-uniform decrease in brightness during the decay phase, again suggesting spatial variations in dust load and non-uniform mixing, possibly due to episodic injection of dust into the atmosphere locally (8).

The albedo of most regions had returned to the pre-storm value by L_s 355°, indicating that the atmosphere had cleared to pre-storm levels by that time. This conclusion is supported by Viking lander observations, which show that the opacity over the two lander sites had decreased to pre-storm levels by L_s 360° (8). Therefore, surfaces that remained brighter after L_s 360° than they were prior to the two storms are thought to be covered by a thin layer of bright dust fallout.

The distribution of surfaces that remained bright following the storms, and those where the surface quickly returned to its pre-storm albedo follow a consistent pattern. The albedo of bright regions, such as Arabia and Tharsis, rapidly returned to pre-storm values, and was close to the albedo of the previous year (Fig. 1a). Many dark regions also darkened to nearly their pre-storm levels by L_s 360° (Fig. 1b). This pattern holds particularly well for southern hemisphere dark regions. This behavior is consistent with the model of deposition described above; in dark regions the dust is rapidly removed with little net accumulation, whereas in bright regions a dust mantle already exists so that the deposition of additional bright dust does not affect the surface albedo.

There are several dark regions that differ from the general trends described above and provide insight into the level of dust activity that occurs throughout the year. Syrtis Major and Acidalia Planitia are among the few regions that remained significantly brighter at L_s 360° than they were before the global storms began. These areas did, however, continue to darken with time, returning to nearly their pre-storm albedo by L_s 120° (Fig. 1c). It is interesting to note that the albedo of these and some other regions was still slightly higher at this time than it was the previous year, suggesting that some dust still remained on the surface. This finding is consistent with observations at the Viking lander 1 site where dust was deposited following the global storms and was not removed until over a year later (9). These observations support the hypothesis that Syrtis Major and Acidalia Planitia act as local dust sources during inter-storm periods, producing enhanced dust loading in the northern hemisphere (10).

In summary, observations of seasonal changes in surface albedo reveal regional differences in the deposition and subsequent erosion of dust-storm fallout. Southern hemisphere dark areas quickly return to close to their pre-storm albedos, suggesting rapid removal of any dust that was deposited. Northern hemisphere dark regions are brighter post-storm, but gradually darken to pre-storm levels over a Mars year. In doing so they act as local sources of dust during otherwise clear periods. Dust does not appear to be removed from bright regions, resulting in the 1-2 m thick deposits observed today.

References

1) Thomas, P. and J. Veverka, 1979, *J. Geophys. Res., 84*, 8131-8146.
2) Veverka, J., P. Thomas, and R. Greeley, 1977, *J. Geophys. Res., 82*, 4167-4187.
3) Pleskot, L.K. and E.D. Miner, 1981, *Icarus, 45*, 179-201.
4) Lee, S.W., 1986, submitted to Icarus.
5) Christensen, P.R., 1986, *J. Geophys. Res*, in press.
6) Christensen, P.R., 1986, this issue.
7) Christensen, P.R., 1986, submitted to Icarus.
8) Pollack, J.B., D.S. Colburn, R. Kahn, J. Hunter, W. Van Camp, C.E. Carlston, and M.R. Wolfe, 1977, *J. Geophys. Res., 82*, 4479.
9) Arvidson, R.E., E. Guinness, H.J. Moore, J. Tillman, and S.D. Wall, 1983, *Science, 222*, 463-468.
10) Christensen, P.R., 1982, J. *Geophys. Res., 87*, 9985-9998.

1.2.1 FOURIER COLLOCATION

For problems with periodic boundary conditions, the Fourier expansion of a function $u(x)$ is given by the infinite series

$$u(x) = \sum_{k=-\infty}^{\infty} a_k e^{ikx}. \tag{8}$$

The collocation projection is defined by the discrete Fourier transform pair

$$P_N u = \sum_{k=-N/2}^{N/2-1} \hat{u}_K e^{ikx} \tag{9}$$

where the coefficients $\hat{u}_k$ are defined by

$$\hat{u}_k = \frac{1}{N} \sum_{j=0}^{N-1} u\,(\,x_j)e^{-ikx_j} \qquad k = \frac{-N}{2}\,, \frac{-N}{2} + 1, \ldots, \frac{N}{2} - 1. \tag{10}$$

The collocation points, x_j, are uniform on the interval $[0,2\pi]$

$$x_j = 2\pi j/N \qquad j = 0,1,2,\ldots N-1. \tag{11}$$

The transforms (9) and (10) are almost always computed by the use of a fast Fourier transform if N is a highly composite integer such as $N = 2^p 3^q$.

Derivatives of the function u at the collocation points are approximated by the derivatives of the interpolating polynomial. Thus, the ℓth derivative of u is approximated

$$\frac{d^\ell P_N u}{dx^\ell} = \sum_{k=-N/2}^{N/2-1} (ik)^\ell \hat{u}_k e^{ikx}. \tag{12}$$

From the form of equation (12), it is clear that the evaluation of the derivative at the collocation points can also be computed efficiently with a fast Fourier transform. See Hussaini, Streett, and Zang [4] for more information on the implementation of the Fourier collocation method.

1.2.2 CHEBYSHEV COLLOCATION

The collocation points for using a Chebyshev method to approximate a nonperiodic function are usually defined by

$$x_j = -\cos(\pi j/N) \qquad j = 0,1,\ldots,N. \tag{13}$$

These points are the extrema of the N^{th} order Chebyshev polynomial, $T_N(x)$, and are obtained from the Gauss-Lobatto integration formula (see Davis and Rabbinowitz [5]).

The collocation projection operator is defined as the interpolation

$$P_N u = \sum_{n=0}^{N} a_n T_n(x) \tag{14}$$

where the coefficients are defined by

$$a_n = \frac{2}{N} \frac{1}{\overline{C}_n} \sum_{j=0}^{N} \frac{u(x_j) T_n(x_j)}{\overline{C}_j} \quad \text{where} \quad \overline{C}_j = \begin{cases} 2 & j = 1,N \\ 1 & \text{otherwise} \end{cases} . \tag{15}$$

Again, derivatives of u at the collocation points are approximated by the derivative of the interpolating polynomial evaluated at the collocation points. The first derivative, for example, is defined by

$$\frac{du}{dx} = \sum_{n=0}^{N} a_n^{(1)}\, T_n(x)$$

$$a_{N+1}^{(1)} = 0$$

$$a_N^{(1)} = 0 \tag{16}$$

$$\bar{C}_n a_n^{(1)} = a_{n+2}^{(1)} + 2(n+1)a_{n+1} \qquad n = N-1,1,\ldots,0 \,.$$

The transform pair given by equations (14) or (16) and (15) can be efficiently computed with a fast cosine transform. Equivalently, the interpolating polynomial and its derivatives can be computed using matrix multiplication. The matrices for the Chebyshev collocation method are conveniently collected in the review by Gottlieb, Hussaini, and Orszag [6]. For $N < 32$, this approach is competitive with using a fast cosine transform, at least on serial computers.

1.3 APPROXIMATION THEORY (COLLOCATION)

1.3.1 FOURIER COLLOCATION

The problem of how well $P_N u$ approximates u for Fourier approximations has been discussed by Kreiss and Oliger [7], Pasciak [8], and by Canuto and Quarteroni [9]. See also Mercier [10]. It is most convenient to express the interpolation results in terms of a Sobolev space, $H^m(0,2\pi)$. This is a Hilbert space with the norm

$$\|u\|_q = \sum_{j=0}^{q} |u|_j^2 \tag{17}$$

defined in terms of the seminorms

$$|u|_p^2 = \sum_{k=-\infty}^{\infty} k^{2p} |a_k|^2. \tag{18}$$

The use of the discrete Fourier transform pair (9), (10) represents the projection of the Sobolev space onto the space $S_N(0,2\pi)$, the space of Fourier polynomials of degree N.

The primary interpolation result is given by Theorem 1:

Theorem 1: For any $0 \leq p \leq q$ with $q > 1/2$ there exists a constant C independent of u and N such that

$$\|u - P_N u\|_p \leq CN^{p-q} |u|_q. \tag{19}$$

Proof: See Pasciak [8].

Equation (19) states that the rate of convergence depends (through the order of Sobolev norms) only on the smoothness of the function being approximated. This type of error decay is known as spectral accuracy. In practice, one sees errors which decay exponentially and hence spectral accuracy is often called exponential accuracy. Several applications described in Section 2 exhibit exponential accuracy. The term infinite order accuracy is also used often to refer to the case as $q \to \infty$.

Exponential accuracy has been shown explicitly by Tadmor [11] for functions u which are also analytic in the complex plane.

Theorem 2: *Let* $u(x)$ *be* 2π-periodic *and analytic in a strip of width* $2s_0$. *Then for any* $0 < s < s_0$

$$\|u - P_N u\|_p \leq CM(s)/\sinh(s)N^p e^{-Ns} \tag{20}$$

where C *depends on* p *and*

$$M(s) = \operatorname*{Max}_{|\mathrm{Im} z| \leq s} |u(z)|.$$

Proof: See Tadmor [11].

If the solution is not very smooth, then the approximation may not be very good. In fact, if the function is discontinuous, the interpolant shows global oscillations (Gibbs phenomenon) and the approximation error decay is globally only first order. Smoothness is not usually a problem with the solutions of many elliptic or parabolic equations, but discontinuities are characteristic of the solutions of hyperbolic equations.

It is still possible to obtain spectrally accurate approximations to nonsmooth functions, at least away from any discontinuities, but some type of filtering is required. Two papers which address this issue are Majda, MacDonough, and Osher [12] and Gottlieb and Tadmor [13]. The first approach used to smooth discontinuous solutions was that of Majda, MacDonough and Osher

[12] whose results show that spectral accuracy can be retained if Fourier space filtering is applied. Since the main results refer directly to the solutions of hyperbolic partial differential equations, they will be discussed in the next subsection.

Gottlieb and Tadmor [13] have taken the approach of smoothing in real space to allow the accuracy to depend on the local smoothness of the function. The smoothing procedure consists of convoluting the collocation approximation with a regularization kernel which is localized in space. If we call $\tilde{P}\,u$ the smoothed approximation to the originally oscillatory interpolant $P_N u$, the convolution takes on the form

$$\tilde{P}u(x) = 2\,\frac{\pi}{N} \sum_{j=0}^{N-1} P_N u(y_j)\psi^{\theta,p}(x - y_j) \tag{21}$$

where

$$\psi^{\theta,p}(y) = \frac{1}{2\pi\theta}\,\rho_\alpha\!\left(\frac{y}{\theta}\right) \frac{\sin((p + \frac{1}{2})\frac{y}{\theta})}{\sin(y/\theta)} \tag{22}$$

is the Dirichlet kernel localized in space by the cutoff function

$$\rho_\alpha = \begin{cases} e^{\alpha\xi^2/(\xi^2-1)} & |\xi| < 1 \\ 0 & \text{otherwise} \end{cases} . \tag{23}$$

The function ρ ensures that the kernel does not interact with any regions of discontinuity. For example, for a single discontinuity at $x = \pi$, they choose $\theta = \pi|x - \pi|$. With this smoothing, they show that the error depends only on the smoothness of the cutoff function $\rho(\xi)$:

Theorem: *Let* $\rho(\xi)$ *be a* C^{2s} *cutoff function satisfying* $\rho(0) = 1$ *and having support in* $[-\pi,\pi]$. *Then for any* x *in* $[0,2\pi]$ *the smoothed function* $\tilde{P}u$ *satisfies the estimate*

$$|\tilde{P}u(x) - u(x)| \leq C_s\|\rho\|_{2s} \max_{\substack{|y-x| \leq \theta\pi \\ 0 \leq k \leq 2s}} |D^k u(y)|(1 + \theta^{-2s})N^{1-s} \qquad s > 1. \tag{24}$$

Proof: See Gottlieb and Tadmor [13].

1.3.2 CHEBYSHEV COLLOCATION

To study the approximation properties of the Chebyshev projection (14), it is practical to work in a weighted Sobolev space with weight $w(x) = (1 - x^2)^{-1/2}$. Defining the weighted L^2_w norm by

$$\|u\|^2_{0,w} = (u,u)_w = \int_{-1}^{1} u^2 w dx \tag{25}$$

and the Sobolev norm by

$$\|u\|^2_{q,w} = \sum_{i=1}^{q} \left\|\frac{d^i u}{dx^i}\right\|^2_{0,w} \tag{26}$$

the spectral approximation result is given by

Theorem 4: *Let* $q > 1/2$ *and* $0 \leq p \leq q$. *Then there exists a constant* C *such that for all* u *in* $H^q_w(-1,1)$

$$\|u - P_N u\|_{p,w} \leq CN^{2p-q} \|u\|_{q,w}. \tag{27}$$

Proof: See Canuto and Quarteroni [9].

So, like the Fourier approximation, the Chebyshev interpolation gives spectral accuracy; that is, the accuracy depends only on the smoothness of the function to be interpolated. Exponential convergence has also been proved by Tadmor [11]. This time, the function u must be analytic in an ellipse with foci at -1 and 1:

Theorem 5: Assume $u(x)$ is analytic in $[-1,1]$ and has a regularity ellipse whose sum of its semi-axes equals $r_0 = \exp(\eta_0) > 1$. Then for any $\eta, 0 < \eta < \eta_0$ we have

$$\|u(x) - P_N u\|_{H_T^1} \leq 8M(\eta)\left(\frac{\coth(N\eta)}{e^{2\eta} - 1}\right)^{1/2} Ne^{-N\eta} \tag{28}$$

where the norm is defined by

$$\|u\|_{H_T^1} = \sum_{p=0}^{\infty} (1+p)^2 |\hat{u}_p|^2. \tag{29}$$

Proof: See Tadmor [11].

If the function which is being approximated is discontinuous, it is still theoretically possible to recover a spectrally accurate solution [13] by filtering in physical space. The procedure is the same as the smoothing procedure for the Fourier case, but the Dirichlet kernel is replaced by

$$K_p(y) = \frac{(1 - y^2)}{p\pi} \frac{T_p'(y)}{y} . \tag{30}$$

1.4 THEORY OF SPECTRAL COLLOCATION METHODS FOR PDE'S

Proofs of the convergence of spectral approximations to partial differential equations are usually accomplished using energy methods which mimic proofs of the well-posedness of the original equations. Consequently, it is most convenient to discuss stability and convergence with respect to the three major types of partial differential equations separately.

1.4.1 ELLIPTIC EQUATIONS

Theoretical analysis of the convergence of Fourier collocation methods is simplified because of periodic boundary conditions. The elliptic problem is to find the function $u(x)$ which satisfies

$$\begin{aligned} &Lu = f \qquad x \in [0,2\pi] \\ &u(0)=u(2\pi) \end{aligned} \tag{31}$$

where L has the property

$$(Lu,u) \geq \alpha \|u\|_1^2 \qquad \alpha > 0. \tag{32}$$

The Fourier collocation approximation is obtained as described in section

1.2.1 and satisfies the same inequality, i.e., $\forall u \in S_N$

$$(L_c u, u) \geq \alpha \|u\|_1^2.$$

Then we have the

Theorem 6: For $\tau > 1$ there exists a constant C such that if $f \in H_p^{\tau-2}(0,2\pi)$ and $u \in H_p^{\tau}(0,2\pi)$ then the following estimate is optimal:

$$\|u - P_N u\|_1 \leq C(1 + N^2)^{(1-\tau)/2} \|u\|_\tau. \tag{33}$$

Proof: See Mercier [10].

Chebyshev methods with both Dirichlet and Neumann boundary conditions have been analyzed for the elliptic differential equation of the form

$$Lu = -(au_x)_x + (bu)_x. \tag{34}$$

The Chebyshev spectral collocation approximation is formally written as

$$L_c u_c = -\frac{d}{dx}\left(P_N\left(a \frac{d}{dx} u_c\right)\right) + \frac{dP_N(bu_c)}{dx} + \gamma u_c. \tag{35}$$

For Dirichlet problems, the equation is collocated at the interior points and boundary conditions of the form

$$u(-1) = u_l \quad \text{and} \quad u(+1) = u_r \tag{36}$$

are specified directly at the boundary points. Stability and convergence were proved by Canuto and Quarteroni [14] using a variational approach. They show

Theorem 7: Let u_c be the solution to $L_c u_c = f_c$ where L_c is defined by equation (35) with homogeneous boundary conditions, $u_r = u_l = 0$ then with suitable conditions on a, b, α the following estimate holds

$$\|u - u_c\|_{1,w} \le C_1 N^{1-r} \|u\|_{r,w} + C_2 N^{-s} \|f\|_{s,w}. \tag{37}$$

Proof: See Canuto and Quarteroni [14], Theorem 2.4.

Convergence proofs for Neumann or mixed-type boundary conditions are available for boundary conditions applied in one of two different ways. A discussion of these approaches can be found in Canuto [15], [16]. The first approach is explicit. At interior points, the equation is collocated normally as in equation (35). At the boundary points, however, the collocation approximation to the derivative is written in matrix form and the boundary conditions are used to determine the value at the boundary point. Thus, the approximation to the boundary condition

$$\begin{aligned} B_l u &= \beta u_x + \alpha u \\ B_r u &= \delta u_x + \gamma u \end{aligned} \tag{38}$$

is found by solving the system

$$\begin{aligned}(\alpha + \beta d_{00})u_0 + \beta d_{0N}u_N &= -\beta \sum_{j=1}^{N-1} d_{0j}u_j \\ (\gamma + \delta d_{NN})u_N + d_{N0}u_0 &= -\delta \sum_{j=1}^{N-1} d_{Nj}u_j\end{aligned} \tag{39}$$

where $u_j = u_c(x_j)$ and $[d_{ij}]$ is the matrix for the derivative at the collocation points (see Gottlieb, Hussaini, and Orszag [6]).

The convergence is very rapid for smooth solutions:

Theorem 8: *Let* $\sigma > 1/2$ *and let* u *and* u_c *be solutions to* $Lu = f$ *and* $L_c u_c = f_c$ *where* L *and* L_c *are defined as above. Then with explicitly applied Neumann boundary conditions the following convergence estimate holds*

$$\|u - u_c\|_{2,\eta} \leq CN^{-\sigma}\{\|u\|_{\sigma+2,w} + \|f\|_{\sigma,w}\} \tag{40}$$

where $\eta = (1 - x^2)w(x)$ *and* C *is independent of* N.

Proof: See Canuto and Quarteroni [14], Theorem 3.2.

Canuto [16] also describes how to impose Neumann boundary conditions implicitly for elliptic problems. In this way, the boundary conditions are not exactly satisfied because what is actually solved is the modification of the interior approximation. For the spectral case of a pure Neumann problem, the first derivatives are computed normally as in equation (16). At the boundary points, the derivatives are replaced by the Neumann conditions. Then the second spectral derivatives are computed by using (16) again on the modified set of derivatives. This has the advantage that all of the points are treated the same, but the boundary conditions are not exactly satisfied. The boundary error does decay spectrally, however.

Theorem 9: Let u_c be the solution to $L_c u_c = f$ with implicit Neumann boundary conditions. If $u \in H^m_w(-1,1)$ with $m > 5/2$ then

$$\left|\frac{\partial u_c}{\partial x}(\pm 1)\right| \leq C N^{4-m} \|u\|_{m,w} \tag{41}$$

where $C > 0$ is independent of N.

Proof: See Canuto [16].

The convergence in the interior is also spectral, and the estimate bounds both the solution and the collocation derivative.

Theorem 10: Under the assumptions of Theorem 9,

$$\|u - u_c\|_{0,w} + \left(\sum_{j=1}^{N-1} [u_x - (u_c)_x](x_j) w_j\right)^{1/2} \leq C N^{2-m} \|u\|_{m,w} \tag{42}$$

where the w_j are the Gauss-Lobatto weights at the points x_j.

Proof: See Canuto [16].

1.4.2 PARABOLIC EQUATIONS

The convergence and stability theory for linear parabolic equations, like the theory for elliptic equations, is fairly well developed. In particular, the theory has centered on studies of semi-discrete equations in which the spatial variation is discretized, but the time variation is left continuous. The emphasis, however, has been on application to boundary value problems--that is, on the convergence of Chebyshev collocation methods. In this section we survey theoretical results for initial-value problems of the form

$$\begin{aligned} u_t &= (Au_x)_x + Bu_x + Cu + f \\ u(x,0) &= u_0(x) \end{aligned} \tag{43}$$

where A, B, and C are n x n matries. The general collocation approximation to the first, third, and fourth terms of the right hand side of equation (43) is written in a manner identical to that of the elliptic equations in equation (35).

Stability of the Fourier approximation of the heat equation is easy to prove and is discussed in Gottlieb, Hussaini, and Orszag [6]. The more complicated case is equation (43) above. Kreiss and Oliger [17] have proved stability with two different treatments of the first order term. The first treats it in "skew symmetric form", that is, by writing

$$Bu_x + Cu = \frac{1}{2}(Bu_x + (Bu)_x) + (C - \frac{1}{2}B_x)u. \tag{44}$$

The second, discussed more fully in the next section, involves filtering the first derivative to ensure stability. A convergence estimate using the skew symmetric form for the scalar equation with $f = 0$ is

Theorem 11: *Let* $\tau \geq 1$, $T > 0$, *and assume that* $u_0 \in H_p^{\tau+1}(0,2\pi)$. *Then there exists a constant* C *such that the following estimate holds:*

$$\|u(t) - u_c\|_0 \leq C(1 + N^2)^{(1-\tau)/2}\{\|u_0\|_{\tau-1} + [\|u_0\|_\tau^2 + (\|u_0\|_\tau + \|u_0\|_{\tau+1})^2]^{1/2}\} \quad (45)$$

for all $t \in [0,T]$.

Proof: See Mercier [10] Theorem 11.2. Here H_p is defined in terms of distribution derivatives of periodic functions.

The convergence of Chebyshev approximations to parabolic equations on bounded domains has received a lot of attention recently. The spatial approximation for a Dirichlet problem will be exactly like that for the elliptic problem. Stability for the heat equation with non-constant coefficients was originally shown by Gottlieb [18]. Convergence estimates were worked out by Canuto and Quarteroni [19]. For the scalar heat equation

$$u_t = a(x)u_{xx} \qquad x \in (-1,1) \quad (46)$$

with homogeneous boundary conditions $u(-1,t) = u(1,t) = 0$ they show

Theorem 12: *Let* $\sigma > 1/2$ *and* $S > \sigma + 2$ *and* $T > 0$. *If* $u \in L^1(0,T;\, H_w^S(-1,1))$ *then there is a constant* C, *independent of* N *such* *that*

$$\|u(t) - u_c(t)\|_N \leq C\ N^{2\sigma+4-S} \tag{47}$$

where the norm $\|\cdot\|_N$ *is the discrete norm derived from the Gauss-Lobatto integration formula*

$$\|v\|_N = \sum_{j=0}^{N} \frac{w_j v(x_j)^2}{a(x_j)} . \tag{48}$$

Neumann boundary conditions for parabolic problems can also be applied either explicitly or implicitly. For the implicit treatment, convergence is similar to that of the corresponding elliptic equation:

Theorem 13: *Suppose the solution to the differential equation (46) with Neumann boundary conditions is regular to the extent that* $u \in L^2(0,T;H_w^m(-1,1))$ *and the time derivative satisfies* $u_t \in L^2(0,T;H_w^{m-1}(-1,1))$ *for* $T > 0$ *and* $m > 5/2$. *If* $u_0 \in H_w^m(-1,1)$ *then*

$$\begin{aligned} &\|u(t) - u_c(t)\|_{0,w} + [\sum_{j=1}^{N-1} (u_x - (u_i)_x)^2 w_j]^{1/2} \\ &\leq C_1 N^{2-m} \{\|u_0\|_{m,w} + e^{t/2} [\int_0^t \|u\|_m^2 d\tau \\ &+ \int_0^t \|u_t(\tau)\|_{m-1}^2 dt\} . \end{aligned} \tag{49}$$

Proof: See Canuto [16] Theorem 4.4.

1.4.3 HYPERBOLIC EQUATIONS

The study of the convergence of spectral approximations to hyperbolic equations is complicated by the fact that the straight-forward discretization of an equation of the form

$$u_t + a(x)u_x + bu = 0 \tag{50}$$

written as

$$\frac{\partial u_c}{\partial t} + P_N(a(x)\,\frac{\partial u_c}{\partial x}) + bu_c = 0 \tag{51}$$

is often unstable. In this section, we will discuss the available theory of formally stable approximations.

Fourier methods are stable if $a(x)$ is of fixed sign. If $a(x)$ in equation (50) is strictly positive and $b = 0$, the energy estimate for the approximation, (51)

$$\frac{d}{dt}\sum_{j=0}^{N-1}\frac{u_c^2(x_j)}{a(x_j)} = 0 \tag{52}$$

shows that the approximation is stable. If $a(x)$ is zero at some point, however, then this estimate is not valid and no general technique is available to show stability.

Two basic approaches have been used to devise schemes which can be shown to be stable. The first, indicated in the last section, is to write the spatial derivative in skew-symmetric form. That is, instead of computing (51), one computes

$$\frac{\partial u_c}{\partial t} + 1/2\; P_N\{\frac{a\partial u_c}{\partial x} + \frac{\partial(P_N a u_c)}{\partial x}\} - 1/2\; P_N\{\frac{u\partial a}{\partial x}\} + bu_c = 0. \tag{53}$$

Kreiss and Oliger [20], [17] showed that this discretization is stable. Mercier [10] examined the stability and convergence of the Fourier approximation to the skew-symmetric equation

$$u_t + v(x)u_x + (v(x)u)_x = 0 \tag{54}$$

and showed that the error decay is spectral.

Theorem 14: For $\tau > 1$ and $T > 0$, if the initial condition satisfies $u(x,0) \in H_p^{\tau}(0,2\pi)$ then there is a constant C independent of N such that

$$\| u(t) - u_c(t) \| \leq C(1 + N^2)^{(1-\tau)/2} \| u_0 \|_{\tau}. \tag{55}$$

Proof: See Mercier [10], Theorem 9.1.

Though approximations written in skew symmetric form are stable, there are objections to their use. The first objection is that they are less efficient since they have twice as many derivatives to evaluate. More important, conservation is lost when this is applied to conservation law equations (such as the equations of gas-dynamics) for the computation of weak solutions. Tadmor [21] has examined the skew-self adjoint form of systems of non-linear conservation laws. They can be explicitly shown to be well-posed, but the conservation property is lost.

The alternative to rewriting the equation in skew-symmetric form is to use the approximation of equation (51) and filter the solutions. Finite difference solutions are often filtered by adding an explicit low order artificial viscosity. The goal of filtering Fourier spectral solutions is to do so without destroying the accuracy of the method.

Two approaches for filtering Fourier approximations to guarantee stability have been suggested. The first was proposed by Majda, McDonough and Osher [12]. In their method, the spectral derivative defined in equation (12) is modified by filtering the computed solution. For linear problems, this can be done efficiently by modifying the Fourier coefficients of the solution and using those new coefficients when the derivative is computed. Let $\rho(x) \epsilon c^{\infty}(-\pi,\pi)$ be a "filter function". Its values are zero near $\theta = \pm \pi$ and identically one in a neighborhood of $\theta = 0$. The Fourier coefficient $\hat{u}_k$ is replaced by $\rho(2\pi k/N)\hat{u}_k$ and this is used in equation (12) to compute the derivative.

For smooth initial conditions, smoothing gives a stable approximation and spectral accuracy

Theorem 15: For $u_0 \epsilon C^{\infty}$ the error satisfies the inequality

$$\|u(x,t) - u_c(x,t)\|_s \leq C\,h^{\lambda} \quad \text{for all} \quad s,\lambda \tag{56}$$

where C depends on both s and λ.

Proof: See Majda, McDonough, and Osher [12], Corollary 1.

If the solution is discontinuous, it is still possible to obtain spectral accuracy in the sense of equation (56) if the initial condition is properly smoothed. It is not enough to smooth the discrete Fourier coefficients of the initial condition with a filter whose support is enclosed within the support of ρ. Rather, it is necessary to use smoothed versions of the exact Fourier coefficients.

A different approach to filtering was proposed by Kreiss and Oliger [17]. Instead of filtering the solution with a predefined filter, they showed that linear stability could be obtained by smoothing the space derivative with a weak filter which depends on the smoothness of the function. They arbitrarily split the frequency range of the solution into a high frequency range, $|k| > N_1$, and a low frequency range, $|k| \leq N_1$. The coefficients of the low frequency range are not modified at all. The coefficients of the high frequency range are modified only if they do not decay rapidly enough. Call $v(x) = u_x$ defined by equation (12) and define v_1 to be the derivative summing only the low frequency components $|k| \leq N_1$. The modified coefficients of the derivative are defined to be $w = Hv$ where

$$w_k = \begin{cases} v_k & \text{for } |k| \leq N_1 \\ v_k & \text{if } |k| > N_1 \text{ and } |v_k| \leq D \, \|v_1\|/|k|^j \\ D \, \|v_1\| v_k/(|v_k| \, |k|^j) & \text{otherwise.} \end{cases} \tag{57}$$

Kreiss and Oliger prove the following stability theorem:

Theorem 16: Suppose the coefficient $a(x)$ in equation (50) is smooth so that its Fourier coefficients decay at a rate $|k|^{-\beta}$. The approximation

$$u_t + P_N\left(a \frac{H\partial u_c}{\partial x}\right) = 0 \tag{58}$$

where the filter H is defined by (57) is stable and converges if $j = \beta > 2$.

Proof: See Kreiss and Oliger [17], Theorem 4.2.

For linear problems, however, it is not clear that filtering is always needed. The fact that the energy method gives only a sufficient condition for stability means that equation (52) does not prove instability if a is not of one sign. For example, Gottlieb, Orszag, and Turkel [22] show stability in the usual sense of convergence as $N \to \infty$ of the scalar equation where $a(x) = A\sin(x) + B\cos(x) + C$ for arbitrary A, B, C. The numerical solutions do, however, grow in time - just as the exact solutions do.

For non-linear problems, experience shows that filtering of the Fourier approximation is needed, particularly if there are discontinuities in the solutions. Hussaini, Kopriva, Salas, and Zang [51] discuss the application of these filtering methods and the choice of filters to a periodic transonic flow with a shock.

Proofs of the stability and convergence of Chebyshev approximations have the added complication of the boundary conditions and the weight, $w(x)$, which is unbounded at the endpoints. In particular, the case where $a(x)$ changes sign makes it difficult to show stability. Gottlieb [18] has proved stability of the straightforward Chebyshev collocation for the special cases where $a(x) = \pm x$.

To show stability of Chebyshev approximations in general, the skew-symmetric form of the equations is needed. We will survey the convergence theory of Canuto and Quarteroni [23] for the special case of the hyperbolic boundary value problem with $b(-1) > 0$

$$\begin{aligned} & u_t + (bu)_x + b_0 u = f && x \in (-1,1), \quad t \in (0,T] \\ & u(-1,t) = 0 && t \in (0,T] \\ & u(x,0) = u_0 && x \in (-1,1). \end{aligned} \tag{59}$$

The further assumtion is added that

$$1/2\ b_x + b_o - 1/2\ bw_x w^{-1} \geq 0 \qquad \text{for} \quad x \in (-1,1). \tag{60}$$

For the Chebyshev weight $w(x) = (1 - x^2)^{-1/2}$, the use of integration by parts to get an energy estimate will give an unbounded boundary term evaluated at $x = +1$ (see Gottlieb and Orszag [1]). This has led to the use of a modified weight and norm with which to prove stability and convergence. Let the new weight be $w^*(x) = (1 - x)w(x)$ so that $w^*(1) = 0$. Then the following convergence estimate holds:

Theorem 17: Suppose that $u \in L^\infty(0,T;\ H^\beta_{w^*}(-1,1))$ and $b \in H^\beta_{w^*}(-1,1)$ for $\beta > 2$. Then the skew-symmetric Chebyshev approximation to (59) satisfies

$$\| u - u_c \|_{L^\infty(L^2_{w^*})} \leq CN^{2-\beta} \| u \|_{L^\infty(H^\beta_{w^*})}. \tag{61}$$

Proof: See Canuto and Quarteroni [23], Theorem 2.3. Note: Their theorem actually allows for more general boundary conditions than we have mentioned here.

For computations which are not done in skew-symmetric form, such convergence estimates are not available in the general case. As indicated above, Gottlieb [18] has shown stability in some particular cases. However, Reyna [24] has shown that if $b(x)$ is not strictly positive in the interval that a straight-forward Chebyshev approximation need not be stable. To stabilize the solutions he proposes the use of filtering. It is not sufficient, however, to simply smooth the Chebyshev coefficients. Rather, he shows that stability can be proved if Legendre coefficients are computed from the Chebyshev ones, the Legendre coefficients are smoothed, and then transformed back.

The stability and convergence of Chebyshev approximations to the hyperbolic initial-boundary-value problem for systems

$$u_t = Au_x \qquad -1 \leq x \leq 1, \quad t \geq 0 \tag{62}$$

where u is an n-vector and A is a constant matrix has recently been proved by Gottlieb, Lustman, and Tadmor [25], [26]. Because this system is hyperbolic, the matrix A can be assumed to have been diagonalized to

$$A = \begin{bmatrix} A^I & 0 \\ 0 & A^{II} \end{bmatrix}$$

where $A^I < 0, A^{II} > 0$ are diagonal matrices.

Boundary conditions for which this system is well posed are of the form

$$\begin{aligned} u^I(-1,t) &= Lu^{II}(-1,t) + g^I(t) \\ u^{II}(1,t) &= R\, u^I(1,t) + g^{II}(t) \end{aligned} \tag{63}$$

where u^I and u^{II} represent the partition of u into inflow and outflow components (see Kreiss and Oliger [5]). Under the assumption that the boundary conditions are dissipative, the standard Chebyshev collocation is stable:

Theorem 16: Under the assumption that $|R| \cdot |L| \leq 1 - \delta < 1$, the Chebyshev collocation method is stable for the system (62) with boundary conditions (63) in the sense that there exists a weighting pair $w(x)$ and constants q and $\eta_0 \geq 0$ such that for all s with $\mathrm{Re}\ s = \eta > \eta_0$

$$(\eta - \eta_0)\|\hat{u}_c(x,s)\|^2 \leq CN^{2q}|\hat{g}(s)|^2$$

where $\hat{u}_c$ and $\hat{g}$ are the Fourier transforms of u_c and g.

Proof: See Gottlieb, Lustman, and Tadmor [25].

2. SOME APPLICATIONS OF SPECTRAL COLLOCATION METHODS

In this section, some recent developments in the application of spectral methods to problems in fluid mechanics are surveyed. Much current emphasis has involved making spectral methods more efficient and more applicable to problems with complicated geometries. This has lead to the development of spectral multidomain methods which eliminate the need for global mappings and to the development of iterative techniques for the rapid inversion of the full matrices which occur when implicit time discretizations are used.

2.1. METHODS FOR ELLIPTIC AND PARABOLIC PROBLEMS

2.1.1. SPECTRAL MULTIDOMAIN METHODS

Spectral multidomain methods have been developed in order to avoid the need for global mappings required by spectral methods in problems with complicated geometries. A complicated domain can be subdivided into several subdomains and individual spectral discretizations can be applied to each subdomain. For elliptic and parabolic problems, for handling the interfaces, early work considered explicit enforcement of continuity (e.g. Orszag [27] and Morchoisne [28]). More recently, spectral element discretizations and enforcement of global flux balance have been used. The spectral element methods retain the accuracy of spectral methods in the context of a geometrically flexible finite element formulation. Global flux conservation has been used effectively when the mappings and/or domain sizes vary radically across interfaces.

Consider first the solution of the (second-order, self-adjoint, elliptic) Helmholtz equation,

$$\nabla^2 u - \lambda^2 u = f \quad \text{in} \quad D \tag{64}$$

with Dirichlet boundary conditions on the domain boundary, ∂D. Following the lead of finite element techniques, the spectral element algorithm [29, 30] proceeds by recognizing the equivalence of (64) to maximization of the following variational form,

$$\max_{u \in H^1} \int_D \{-\nabla u \cdot \nabla u/2 - \lambda^2 u^2/2 - uf\} dx, \tag{65}$$

The variational form, (65), is preferable over the differential statement, (64), in that it requires less continuity of candidate solutions.

The spectral element discretization proceeds by breaking up the computational domain, D, into general quadrilateral elements. Within a given element k, the solution, geometry, and data are then expanded as tensor product Lagrangian interpolants through a set of specified collocation points. For instance, in two space dimensions, the solution u in element k is represented as,

$$u^k(r,s) = \sum_i \sum_j u^k_{ij} h_i(r) h_j(s) \tag{66a}$$

$$h_m(z_n) = \delta_{mm}, \tag{66b}$$

where r and s are the local elemental coordinates, the h_i are the Lagrangian interpolants, the z_n are the collocation points, and δ_{mm} is the Kronecker delta symbol. All summations run from 0 to N, where N is the order of the Lagrangian interpolants in each element.

The expansions (66a) are then inserted into (65), and the functional rendered stationary with respect to arbitrary variations in the nodal values, u^k_{ij}. Direct stiffness summation [31] (which recognizes that the global approximation space must be C^0) is then used to assemble the elemental equations into the system matrix. It should be noted that, as regards the treatment of elliptic and parabolic equations, the "spectral element" recipe presented here is very similar to earlier "p-type finite element" methods [32] and the "global element" method [33].

It is clear from the above representation, (66), that the global interpolant space is only C^0, that is, that the approximation space suffers discontinuities in derivative at elemental boundaries. Although this may appear to violate the basic smoothness required of spectral methods, this is not the case due to the fact that the variational formulation, (65), is used rather than the (unintegrated) Galerkin weighted-residual form. In particular, in the absence of "variational crimes", the spectral element method can be shown to achieve exponential convergence to smooth solutions as N, the order of (fixed) elements, is increased. For nonsmooth solutions (e.g., corner-induced singularities), high-order convergence is more difficult to obtain, however refinement techniques have been developed for the p-type finite element method [32].

Variational crimes take the form of numerical quadrature errors and interpolation of boundary data. (Nonconforming elements are not considered.) In order to insure that these errors do not dominate the approximation errors, it is important to correctly choose the collocation points of the Lagrangian interpolants. Earlier work on spectral elements used the Chebyshev collocation points, as they are simple to evaluate and amenable to fast transform techniques. However, as the variational formulation (65) has essentially a unity weighting, it appears that a better choice is the Legendre-Lobatto points from the point of view of accuracy and efficiency of numerical quadratures [6, 34]. Although Legendre polynomials are less convenient than Chebyshev polynomials, are subject to round-off errors for high-order expansions, and cannot be "fast transformed", none of these objections are particularly oppressive for the relatively low-order expansions used in spectral element methods.

As an example of the accuracy of a Legendre spectral element [34] (see [29, 30] for extensive discussion of Chebyshev-based techniques), consider the problem

$$\nabla^2 u = 0 \qquad \text{in} \quad D \tag{67a}$$

where D is the domain defined in Figure 1, $x\epsilon[0, 1]$, $y\epsilon[0, 1 + \frac{1}{2}\sin\pi x]$. Dirichlet boundary conditions are imposed such that the solution to the problem is given by,

$$u(x,y) = \sin(x)e^{-y}. \tag{67b}$$

The L_∞ error for the spectral element mesh shown in Figure 1 is plotted in Figure 2. As expected from the analytic nature of the solution (67b) in the complex plane, exponential convergence is achieved as the order of the elements is increased.

As another example of elliptic problems, consider the moving-boundary Stefan problem [34], given by

$$\nabla^2\theta = 0 \qquad \text{in} \quad D_1, D_2 \tag{68a}$$

$$\theta = 1/2 \quad \text{on} \quad \partial D_I \tag{68b}$$

$$\nabla\theta\cdot\hat{n}|_- + 3\nabla\theta\cdot\hat{n}|_+ = 0 \quad \text{on} \quad \partial D_I \tag{68c}$$

$$\nabla\theta\cdot\hat{n} = 0 \quad \text{on} \quad \partial D_0 \tag{68d}$$

$$\theta = 1 + 1/2 \cos 2\pi x \quad \text{on} \quad \partial D_1 \tag{68e}$$

$$\theta = 0 \quad \text{on} \quad \partial D_2, \tag{68f}$$

where D_1, D_2, ∂D_I, ∂D_0, ∂D_1, and ∂D_2 are defined in Figure 3. Here the evaluation of $_-$ and $_+$ refer to the D_1 and D_2 sides of ∂D_I, respectively. In point of fact, the time-dependent (parabolic) version of (68) was solved, approaching the steady-state only as $t \to \infty$; since the solution of parabolic equations involves at each time step the solution of an elliptic equation of the form (64), this aspect of the problem does not warrant separate discussion.

Solution of the Stefan problem (68) illustrates several aspects of the spectral element method. First, since the interface ∂D_I is unknown and general, it demonstrates the ability to handle complex geometry. Second, though the solution suffers a discontinuity at ∂D_I the method has the ability to resolve certain non-homogeneities without losing "spectral accuracy". Figure 4 shows the interface position obtained with a Legendre spectral element method using a two-element mesh. In Figure 5, the associated temperature (θ) distribution is given. High accuracy can be achieved with very few points.

It is critical that the spectral element schemes not only be accurate, but also efficient as regards work required for a given level of accuracy. The key to the computational efficiency of the techniques is the sum factorization which follows from the tensor product representation, (66). For instance, a typical elemental term in a two-dimensional Chebyshev spectral element equation is of the form,

$$\sum_m \sum_n \int h_i h_m \, dr \int h_j h_n \, ds \; u_{mn} \tag{69}$$

where h_i, u_{mn} are defined as in (66), and all subscripts range from 0 to

N, the order of the polynomial space in each co-ordinate direction. Naive evaluation of this sum gives an operation count of $O(N^3)$, and $O(N^4)$ in three dimensions. This sum factorization is at the heart of both direct solvers using static condensation and fast eigenfunction solvers [35] and iterative solvers using conjugate gradient algorithms [36].

Another approach to handling domain interfaces was taken by Macaraeg and Streett [37], [38]. Within subdomains, the usual collocation procedure described in Section 1.2.2 is used. The interface values are computed by requiring that the solution be continuous and that the global flux be balanced. As an example of the procedure, consider the equation

$$\begin{gathered} G(u) \equiv F_x(u) - \nu u_{xx} = S(u) \\ u(-1) = a \\ u(1) = b \end{gathered} \tag{70}$$

where an interface is placed at $x = x_i$. Integration of (70) from -1 to 1 and the requirement that the jump in the flux [G] be zero at the interface yields

$$G(u)_{x=-1} + \int_{-1}^{x_i} S(u)dx = \check{G}(u)_{x=1} - \int_{x_i}^{1} S(u)dx. \tag{71}$$

Numerical experiments show that spectral accuracy is retained. In two dimensions, the method has been used to solve Laplace's equation with discontinuous boundary conditions.

2.1.2 ITERATIVE SPECTRAL METHODS

For evolution problems, explicit time-stepping can be extremely inefficient. This is because the typical time-step limitation for spectral methods is proportional to $1/N^2$ for the advection equation and $1/N^4$ for the diffusion equation (where N is the number of modes) [39]. Hence, implicit time-stepping becomes a necessity. This results in a set of algebraic equations which are, in general, amenable to iterative solution techniques only. Also, elliptic equations governing practical problems virtually require implicit iterative techniques. Since the condition number of the relevant matrices are large, preconditioned iterative schemes including multigrid procedures are the attractive choices. In this section, the fundamentals of iterative spectral methods are discussed with reference to an elementary example.

For the purpose of illustration, consider the equation,

$$u_x = f, \tag{72}$$

periodic on $[0,2\pi]$. For the Fourier method, the standard choice of collocation points is given in Equation (11).

The Fourier collocation discretization of the equation (72) may be written

$$LU = F, \tag{73}$$

where $U = (u_0, u_1, \dots, u_{N-1})$, $F = (f_0, f_1, \dots, f_{N-1})$, and $L = C^{-1}DC$. Here C is the discrete Fourier transform operator, C^{-1} the inverse transform, and D the diagonal matrix denoting the first derivative operator in the Fourier space. Specifically,

$$C_{jk} = e^{-2\pi ik\frac{(j-N/2)}{N}}, \qquad j,k = 0,\ 1,\ \ldots,\ N-1$$

and (74)

$$d_{jj} = \begin{cases} i(j - N/2) & \text{for } j = 1,2,\ \ldots,\ N-1 \\ = 0 & \text{for } j = 0 \end{cases}$$

The eigenvalues of L are $\lambda(p) = ip$, $p = -N/2 + 1,\ \ldots,\ N/2 - 1$, and the largest one grows as $N/2$. A preconditioned Richardson iterative procedure for solving Eq. (73) is

$$V \leftarrow V + \omega H^{-1}\ (F - LV) \tag{75}$$

where the preconditioning matrix, H, is a sparse, readily invertible approximation to L. An obvious choice for H is a finite difference approximation L_{FD} to the first derivative. With the various possibilities for L_{FD}, the eigenvalue spectrum of $L_{FD}^{-1}L$ is given in Table I. Apparently, the staggered grid leads to the most effective treatment of the first derivative. This kind of preconditioning was successfully used in the semi-implicit time-stepping algorithm for the Navier-Stokes equations discussed in section 2.2 on Navier-Stokes Algorithms. The eigenvalue trends of that complicated set of vector equations are surprisingly well predicted by this extremely simple scalar periodic problem.

Next, consider the second order equation

$$-u_{xx} = f \qquad \text{on } [0,2\pi] \tag{76}$$

with periodic boundary conditions. A Fourier collocation discretization of this equation is the same as Eq. (73) except for the diagonal matrix D which represents now the second derivative operator in the Fourier space.

$$D_{jj} = \begin{cases} -(j-\frac{N}{2})^2, & j = 1, 2, \ldots, N-1 \\ 0, & j = 0 \end{cases} \quad . \tag{77}$$

The eigenvalues of L are $\lambda(p) = p^2$, $p = -N/2 + 1, \ldots, N/2-1$. To make the case for the multigrid procedure (consisting of a fine-grid operator and a coarse-grid correction) as a preconditioner, assume H to be the identity matrix I in the iterative scheme (75). The iterative scheme is convergent if the eigenvalues, $(1 - \omega\lambda)$, of the iteration matrix $[I - \omega L]$ satisfy

$$|1 - \omega\lambda| < 1.$$

Each iteration damps the error component corresponding to λ by a factor $v(\lambda) = |1-\omega\lambda|$. The optimal choice of λ is that which balances damping of the lowest-frequency and the highest-frequency errors, i.e.,

$$(1 - \omega\lambda_{max}) = -(1 - \omega\lambda_{min}) \tag{78}$$

This yields

$$\omega_{SG} = \frac{2}{(\lambda_{max} + \lambda_{min})} , \tag{79}$$

and the spectral radius

$$\mu_{SG} = \frac{(\lambda_{max} - \lambda_{min})}{(\lambda_{max} + \lambda_{min})} \quad . \tag{80}$$

In the present instance, $\lambda_{max} = N^2/4$, $\lambda_{min} = 1$, and thus $\mu_{SG} = 1 - 8/N^2$, This implies order N^2 iterations are needed for convergence. This poor performance is due to balancing the damping of the lowest frequency eigenfunction with the highest-frequency one. The multigrid procedure exploits the fact that the lowest-frequency modes ($|p| < N/4$) can be damped efficiently on coarser grids, and settles for a relaxation parameter value which balances the damping of the mid-frequency mode ($|p| = N/4$) with the highest-frequency one ($|p| = N/2$). Table II provides the comparison of single-grid and multigrid damping factors for N=64. The high frequencies from 16 to 32 are damped effectively in the multigrid procedure, whereas the frequencies lower than 16 are hardly damped at all. But then some of these low frequencies (from 8 to 16) can be efficiently damped on the coarser grid with N=32. Further coarser grids can be employed until relaxation becomes so cheap that all the remaining modes can be damped. In concrete terms, the ingredients of a multigrid technique are a fine-grid operator, a relaxation scheme, a restriction operator which interpolates a function from the fine grid to the coarse grid, a coarse-grid operator, and a prolongation operator interpolating a function from the coarse grid to the fine grid. The fine grid problem for the present example may be written

$$L^f U^f = F^f \ . \tag{81}$$

Let V^f denote the fine-grid approximation. After the high-frequency content of the error $V^f - U^f$ has been sufficiently damped, attention shifts to the coarse grid. The coarse-grid problem is

$$L^c U^c = F^c \tag{82}$$

where

$$F^c = R\,[F^f - L^f\,V^f]\ ,$$

R being the restriction operator. After a satisfactory approxmation, V^c is obtained; the coarse-grid correction $(V^c - RV^f)$ is interpolated onto the fine grid by the prolongation operator P, yielding the corrected fine-grid solution

$$V^f \leftarrow V^f + P\,(V^c - RV^f) \tag{83}$$

The details of spectral multigrid techniques are furnished in [40]. Spectral multigrid techniques have been used to solve a variety of problems including the transonic full potential equation [41, 42]. Additional applications of spectral methods to compressible flows are described in [42].

2.1.3 Convection-Dominated Flows

A model for convection-dominated flows is the viscous Burger's equation,

$$u_t + (u^2)_x/2 = \nu u_{xx} \qquad u(x,t=0) = -\sin\pi x, \tag{84}$$

with boundary conditions $u(-1) = u(1) = 0$, and "small diffusivity," $\nu = .01/\pi$ [43]. The solution to this problem develops a near shock. This near shock is characterized by the time at which the derivative at the origin attains a maximum value, t_{max}, and the value of its maximum derivative, $|\partial u/\partial x|_{max}$. The convective term is clearly dominant for short times, however the diffusion term insures that the solution will be smooth. This convection-diffusion balance is a good model for the kind of phenomena that arise in solution of the incompressible Navier-Stokes equations. The critical numerical issues are numerical dispersion and diffusion. The former leads to incorrect propagation speeds of the shock affecting t_{max}. The latter leads to smearing of the shock affecting $|\partial u/\partial x|_{max}$.

This problem has been solved by a variety of methods, including the spectral element method [43] and the explicit flux balancing method [37]. The spectral element calculations have used Crank-Nicolson in time on the diffusion term and the resulting Helmholtz equation in space was solved using the variational methods presented in Section 2.1.1. The convective term was handled by explicit third-order Adams-Bashforth. Four elements were used covering the intervals [-1., -0.05], [-0.05, 0], [0., 0.05], [0.05, 1.] which cluster points around the location of high function variation. Macaraeg and Streett [37] used three subdomains with their flux conservation method. Table III presents a comparison of various methods on this model problem.

2.2. INCOMPRESSIBLE NAVIER-STOKES EQUATIONS

This section is devoted to a description of algorithms for the solution of the incompressible Navier-Stokes equations in primitive variable form. The algorithms are based on methods discussed in the previous section in the simplest context. For example, simulation of instability and transition to turbulence in a flat-plate boundary layer have used iterative methods described in section 2.1.2. The spectral element method has been used for a variety of flow computations, including the problem of flow past a cylinder.

The Navier-Stokes equations in the so-called rotation form are

$$q_t = q \times \omega + \nabla \cdot (\mu \nabla q) - \nabla P \qquad \text{in } D$$

$$\nabla \cdot q = 0 \qquad \text{in } D \tag{85}$$

$$q(x,0) = q_0(x) \qquad \text{in } D$$

and

$$q = g \qquad \text{on } \partial D$$

where $q = (u,v,w)$ is the velocity vector, $\omega = \nabla \times q$ the vorticity, $P = p + 1/2\ |q|^2$ the total pressure, μ the variable viscosity, D the interior of the domain, and ∂D its boundary. In the stability and transition problems, the domain D is cartesian and semi-infinite: periodic in the two horizontal directions (x,z), and bounded by a wall at $y=0$. Fourier collocation can be used in the periodic directions (x,z) and Chebyshev collocation is used in the vertical (y) direction. The collocation points in the periodic directions are given by a relation similar to Eq. (11). The vertical extent of the domain $0 < y < \infty$ is mapped onto $-1 < \xi < +1$. The velocities are defined and the momentum equations enforced at the points

$$\xi_j = \cos\left(\frac{\pi j}{N_y}\right), \qquad j = 0, 1, \ldots, N_y \tag{86}$$

The pressure is defined at the half points

$$\xi_{j+\frac{1}{2}} = \cos\left[\frac{\pi\ (j+1/2)}{N_y}\right], \qquad j = 0, 1, \ldots, N_y - 1 \tag{87}$$

and the continuity equation is enforced at these points. The staggered grid avoids artificial pressure boundary conditions, and precludes spurious pressure modes.

After a Fourier transform in x and z, the temporal discretization (backward Euler for pressure, Crank-Nicolson for normal diffusion, and third or fourth-order Runge-Kutta for the remaining terms) of Eq. (85) leads to

$$\begin{gathered}[I - MDM]\, Q + \Delta t\, A_0\, \nabla \Pi = Q_c \\ - A_+\, \nabla \cdot Q = 0\end{gathered} \tag{88}$$

where

$$Q = \{\tilde{q}_0^{n+1}, \tilde{q}_1^{n+1}, \ldots, \tilde{q}_N^{n+1}\}$$

$$\Pi = \{\tilde{p}_{1/2}^{n+1}, \hat{p}_{3/2}^{n+1}, \ldots, \hat{p}_{N-1/2}^{n+1}\} \tag{89}$$

$$\nabla = \{ik_x \frac{\partial}{\partial y}, ik_x\}$$

M is the Chebyshev derivative operator, D the diagonal matrix with $1/2\mu\Delta t$ as its elements, and A_0 is the interpolation operator from the half points to cell faces, A_+ vice versa. Obviously, the equations for each pair of horizontal wave number (k_x , k_x) are indepenent and they can be written as the system

$$LX = F \tag{90}$$

where $X = [Q, \Pi]$. The iterative solution of this equation is carried out by preconditioning the system with a finite difference approximation on the Chebyshev grid, and applying a standard iterative technique such as Richardson, minimum residual or multigrid [44].

The method described above solves the implicit equations together as a set. The extension of this method to the more general cases of interest such as those involving two or more inhomogeneous directions is not straightforward. An alternative is the operator-splitting technique or the fractional step scheme [45]. This method yields implicit matrices which are positive definite and are easily amenable to iterative methods. In the first step, one solves the advection-diffusion equation

$$q_t^* = q^* \times \omega^* + \nabla \cdot (\mu \nabla q^*) \tag{91}$$

subject to the initial and boundary conditions

$$\begin{aligned} q^*(x,t_n) &= q(x,t_n). \\ q^* &= g^* \quad \text{on } \partial D. \end{aligned} \tag{92}$$

Note that g^* has yet to be defined. In the second step, one solves for the pressure correction

$$q_t^{**} = -\nabla P^{**} \tag{93}$$

$$\nabla \cdot q^{**} = 0 \tag{94}$$

subject to the conditions

$$\begin{aligned} q^{**}(x,t^*) &= q^*(x,t^*) && \text{in } D \\ g^{**} \cdot \hat{n} &= g \cdot \hat{n} && \text{in } \partial\Omega \end{aligned} \tag{95}$$

where $\hat{n}$ is the unit normal to the boundary. Further, the tangential component of the Eq. (95) holds on the boundary, i.e.,

$$q_t^{**} \cdot \hat{\tau} = - \nabla P^{**} \cdot \hat{\tau} \qquad \text{in } \partial D \qquad (96)$$

where $\hat{\tau}$ is a unit tangent vector to the boundary. Now g^* is defined [45] as (using Taylor expansion in t)

$$g^* \cdot \hat{n} = (g^n + \Delta t\, g_t^n) \cdot \hat{n} \qquad (97)$$

$$g^* \cdot \hat{\tau} = [g^n + \Delta t\, (g_t^n + \nabla P^n)] \cdot \hat{\tau} .$$

Eq. (91) is discretized in the usual spectral collocation manner. After a temporal and spatial discretization of Eq. (93), the boundary conditions are built into the relevant matrix operators, and then a discrete divergence is taken. This results in a discrete Poisson equation (with as many algebraic equations as unknowns) for pressure, which can be solved by standard iterative techniques including the multigrid method.

Spectral element methods have been applied to the incompressible Navier-Stokes equations (85). In addition to (98), the uncoupled (passive) or coupled (natural convection) energy equation is also often of interest. The time discretization used for the Navier-Stokes equations is either a Green's function technique [29] or an operator splitting scheme [30]. Both of these methods reduce (85) at each time step to an initial convective step, followed by a Stokes problem consisting of a sequence of Poisson and Helmholtz equations. The spatial discretizations discussed above in Section 2.1.1 are directly applicable to these Navier-Stokes subproblems.

Spectral element methods have been applied to the simulation of numerous flows in the Reynolds number range $0 < R < 1500$ [36, 46 - 47]. An example is provided by the classical problem of flow past a cylinder. Results are presented here for $R = 100$, based on freestream velocity and cylinder diameter, for times sufficiently large that the flow has reached a steady-periodic state. Figure 6 shows the spectral element mesh used, and Figures 7 and 8 show the streamlines and isotherms, respectively, at one time in the periodic flow cycle. The thermal boundary conditions are $T = T_\infty$ far from the cylinder, and $T = T_w$ on the cylinder surface. The isotherm pattern clearly reveals the spatial structure of the von Karman vortex street. Note the minimal numerical dispersion in the scheme, as evidenced by the clear identity of the shed packets of fluid and the absence of trailing waves in the wake. More details of these cylinder calculations, as well as comparisons with previous numerical work and experiment, can be found in [36].

2.3 HYPERBOLIC EQUATIONS

Here, the application of spectral methods to the solution of inviscid compressible flow problems is surveyed. Methods for such problems are not nearly so advanced as those for incompressible flows. The survey is limited to methods for the solution of the Euler equations of gas-dynamics governing some flows of aerodynamic interest. For the solution of the full potential equation for transonic flows, see Streett, et al. [42].

The Euler equations of gas-dynamics are a coupled system of nonlinear hyperbolic equations which (in one dimension) are usually written in the conservative form

$$u_t + F_x(u) = H. \tag{98}$$

Typically, spectral discretization in space and explicit finite difference discretization in time are used. The discontinuous solutions of this set of equations have been obtained in the case of a shock tube (Gottlieb, Lustman, and Orszag [48], Cornille [49]), quasi-one-dimensional flow in a nozzle (Zang and Hussaini [50]) and for the astrophysical problem of shocked flow in a galaxy (Hussaini, et al. [51]).

The astrophysical problem is the most challenging one-dimensional compressible flow problem for which shock capturing has been attempted with a Fourier spectral method. It contains a strong shock and an adjacent strong expansion. Unlike problems with weak shocks and expansions, it was necessary to apply strong filtering to stabilize the numerical solution. The result of this drastic filtering was a reduction of the order of accuracy. Even in the smooth parts of the solution away from the shock, the accuracy was only first order. In view of the extra work involved to compute the spectral approximations, it is not clear that spectral methods with filtering are a viable alternative to finite difference methods when strong shocks are captured.

An alternative to capturing shocks is to treat them as boundaries. In this case, it is possible to compute the solutions using the nonconservative form

$$Q_t + AQ_x = E \tag{99}$$

along with an ordinary differential equation for the motion of the shock. A number of two dimensional shock-fitted solutions are described in Hussaini, et al. [52]. These solutions include a shock/turbulence interaction.

Shock/vortex interaction and supersonic flow past a cylinder. When shocks are fitted, spectral methods do indeed outperform typical second order finite difference methods, as long as the solution is adequately resolved. Kopriva, et al. [53] compared the performance between MacCormack's method and the spectral collocation method for the shock/plane wave interaction problem and for the Ringleb problem. A comparison of the accuracy of the finite difference method vs. the spectral method is shown in Table IV.

A multidomain method for the nonconservative form of the Euler equations suitable for use with shock-fitting has been described by Kopriva [54]. In each subdomain, the usual collocation method (Section 1.2.2) is applied. At interfaces, however, a weighted average of the derivatives is used. In one dimension,

$$Q^I_t + A^L Q^L_x + A^R Q^R_x = E \qquad (100)$$

where Q^I denotes the solution vector at the interface and the derivatives superscripted with the L and R denote the left and right computed spectral approximations. For consistency, $A^L + A^R = A$. The weighting corresponds to an upwind approximation

$$A^L = 1/2(A + |A|) \qquad A^R = 1/2(A - |A|)$$

where $|A| = Z|\Lambda|Z^{-1}$ and Z is the matrix of right eigenvectors. For many applications, this can be simplified by replacing $|A|$ by a diagonal approximation $|A| \approx \lambda^* I$ where $|\lambda|_{min} < \lambda^* < |\lambda|_{max}$ is an approximation to the eigenvalues of A.

Subdividing the domain retains spectral accuracy. Table V shows the performance of a two domain computation of a one dimensional 2×2 hyperbolic system

$$\begin{bmatrix} u \\ v \end{bmatrix}_t + \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}_x = 0 \qquad (102)$$

from [54] with solution u and v for an equal number of points on each side of the interface.

For a given number of grid points, it is possible to obtain solutions with a multidomain spectral method which are significantly better than the single domain method. For the two dimensional nonlinear Ringleb problem computed by Kopriva [54], Table VI shows the effect on the error of a four domain division in which the position of the streamwise interface is varied. The accuracy is best when rapid changes in the solution are best resolved.

The shock-vortex interaction problem described by Kopriva [55] provides another example of the advantage of a multidomain method over a single domain method. A two dimensional region between the shock and an arbitrary upstream boundary is mapped onto a square. The shock moves downstream where it encounters a vortex. The interaction of the shock and the vortex creates a circular sound wave centered on the vortex. Because the physical domain is continually increasing in size, the resolution of the solution decreases with time.

The single domain solution to the shock-vortex problem cannot be computed without added smoothing. Figure 9a shows the pressure contours with no smoothing. The numerical oscillations in the pressure are of the same order as the sound pressure wave created by the interaction. If the region between the shock and the upstream boundary is subdivided and some of the subdomains are allowed to move with the shock, smoothing is not required. Figure 9b shows the pressure contours of a two domain calculation with the same number of grid points in the horizontal direction. The horizontal numerical oscillations are no longer present and the sound pressure wave is clearly visible.

REFERENCES

1. Gottlieb D. and S. A. Orszag, "Numerical analysis of spectral methods: Theory and application," SIAM-CBMS, Philadelphia, 1977.

2. Quarteroni, A., "Theoretical motivations underlying spectral methods", Proc. Meeting I.N.R.I.A.-Novosibirsk, Paris, December 1983.

3. Courant, R. and D. Hilbert, Methods of Mathematical Physics, Vol. I, Interscience, New York, 1953.

4. Hussaini, M. Y., C. Streett, and T. A. Zang, "Spectral methods for partial differential equations," ARO Report 84-1, pp. 883-925.

5. Davis, P. J. and P. Rabbinowitz, Methods of Numerical Integration, 2nd Edition, Academic Press, New York, 1984.

6. Gottlieb, D. , M. Y. Hussaini, and S. Orszag, "Theory and application of spectral methods" in Spectral Methods for Partial Differential Equations, SIAM, Philadelphia, 1984.

7. Kreiss, H.-O. and J. Oliger, "Approximate solution of time dependent problems," GARP Rept. #10, World Meteorological Organization, 1973.

8. Pasciak, J. E. "Spectral and pseudospectral methods for advection equations," Math. Comp., Vol. 35, 1980, pp. 1081-1092.

9. Canuto, C. and A. Quarteroni, "Approximation results for orthogonal polynomials in Sobolev spaces," Math. Comp., Vol. 38, 1982, pp. 67-86.

10. Mercier, B., "Analyse numerique des methodes spectrales," Note CEA-N-2278, Commisariat A L'Energie Atomique, June 1981.

11. Tadmor, E. "The exponential accuracy of Fourier and Tchebyshev differencing methods," SIAM J. Numer. Anal., Vol. 23, 1986, pp. 1-10.

12. Majda, A., J. McDonough, and S. Osher, "The Fourier method for nonsmooth initial data," Math. Comp., Vol. 32, 1978, pp. 1041-1081.

13. Gottlieb, D. and E. Tadmor, "Recovering pointwise values of discontinuous data within spectral accuracy," ICASE Report No. 85-3, 1985.

14. Canuto, C. and A. Quarteroni, "Variational methods in the theoretical analysis of spectral methods," in Spectral Methods for Partial Differential Equations, SIAM, Philadelphia, 1984, pp. 55-78.

15. Canuto, C., "Recent advances in the analysis of Chebyshev and Legendre spectral methods for boundary value problems," in Glowinski and Lions (Editors) Computing Methods in Applied Science and Engineering, VI North Holland (Amsterdam, New York), 1984.

16. Canuto, C., "Boundary conditions in Chebyshev and Legendre methods," SIAM J. Numer. Anal., Vol. 23, 1986, pp. 815-831.

17. Kreiss, H.-O. and J. Oliger, "Stability of the Fourier method" SIAM J. Numer. Anal., Vol. 16, 1979, pp. 421-433.

18. Gottlieb, D., "The stability of pseudospectral-Chebyshev methods," Math. Comp., Vol. 36, 1981, pp. 107-118.

19. Canuto, C. and A. Quarteroni, "Spectral and pseudospectral methods for parabolic problems with non-periodic boundary conditions," Calcolo, Vol. 23, 1981, pp. 197-217.

20. Kreiss, H.-O. and J. Oliger, "Comparison of accurate methods for the integration of hyperbolic equations," Tellus, Vol. 24, 1972, pp. 199-215.

21. Tadmor, E., "Skew-selfadjoint form for systems of conservation laws," J. Math. Anal. Appl., Vol. 103, 1984, pp. 428-442.

22. Gottlieb, D., S. A. Orszag, and E. Turkel, "Stability of pseudospectral and finite-difference methods for variable coefficient problems," Math. Comp., Vol. 37, 1981, pp. 293-305.

23. Canuto, C. and A. Quarteroni, "Error estimates for spectral and pseudospectral approximations of hyperbolic equations," SIAM J. Numer. Anal., Vol. 19, 1982, pp. 629-642.

24. Reyna, L. G. M., "Stability of Chebyshev collocation," Ph.D. Thesis, California Institute of Technology, 1982.

25. Gottlieb, D., L. Lustman, and E. Tadmor, "Stability analysis of spectral methods for hyperbolic initial-boundary value systems," ICASE Report No. 86-2, NASA CR-178041, 1986.

26. Gottlieb, D., L. Lustman, and E. Tadmor, "Convergence of spectral methods for hyperbolic initial-boundary value systems," ICASE Report No. 86-8, NASA CR-178063, 1986.

27. Orszag, S. A., "Spectral methods for problems in complex geometries," J. Comp. Phys., Vol. 37, 1980, p. 70.

28. Morchoisne, "Inhomogeneous flow calculations by spectral methods: Monodomain and multidomain techniques," in Spectral Methods for Partial Differential Equations, D. Gottlieb, M. Y. Hussaini, and R. G. Voigt (Editors), SIAM Philadelphia, 1984, p. 181.

29. Patera, A. T., "A spectral element method for fluid dynamics; laminar flow in a channel expansion," J. Comput. Phys., Vol. 54, 1984, pp. 468-488.

30. Korczak, K. Z. and A. T. Patera, "An isoparametric spectral element method for solution of the Navier-Stokes equations in complex geometry," J. Comput. Phys., Vol. 62, 1986, pp. 361-382.

31. Strang, G. and G. J. Fix, "An analysis of the finite element method," Prentice-Hall, New Jersey, 1973.

32. Babuska, I. and M. R. Dorr, "Error estimates for the combined h and p versions of the finite element method," Numer. Math., Vol. 37, 1980, p. 267.

33. Delves, L. M. and C. A. Hall, " An implicit matching procedure for global element calculations," J. Inst. Math. Appl., Vol. 23, 1979, p. 223.

34. Bullister, E. T., G. E. Karniadakis, E. M. Ronquist, and A. T. Patera, "Solution of the unsteady Navier-Stokes equations by spectral element methods," in Proc. Sixth Int. Symp. on Finite Element Methods in Flow Problems, Antibes, 1986.

35. Patera, A. T., "Fast direct Poisson solvers for high-order finite element discretizations in rectangularly-decomposable domains," J. Comput. Phys., Vol. 65, 1986, pp. 474-480.

36. Karniadakis, G. E., E. T. Bullister, and A. T. Patera, "A spectral element method for the two- and three-dimensional time-dependent incompressible Navier-Stokes equations," in Proc. Europe-U.S. Symp. on Finite Element Methods for Nonlinear Problems, Norway, Springer, 1986.

37. Macaraeg, M. G. and C. L. Streett, "Improvements in spectral collocation through a multiple domain technique," Appl. Numer. Math., Vol. 2., 1986, pp. 95-108.

38. Macaraeg, M. G. and C. L. Streett, "A spectral multi-domain technique with application to generalized curvilinear coordinates," NASA TM 87701, 1986.

39. Gottlieb, D. and S. A. Orszag, "Numerical analysis of spectral methods: Theory and applications," CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, 1977.

40. Zang, T. A., Y-S. Wong, and M. Y. Hussaini, "Spectral multigrid methods for elliptic equations," J. Comput. Phys., 1982, Vol. 48, pp. 485-501 and Vol. 54, pp. 489-507.

41. Streett, C. L., T. A. Zang, and M. Y. Hussaini, "Spectral multigrid methods with applications to transonic potential flow," J. Comput. Phys., Vol. 57, 1985, pp. 43, 76.

42. Hussaini, M. Y., M. D. Salas, and T. A. Zang, "Spectral methods for inviscid, compressible flows," Advances in Computational Transonics, pp. 875-912, W. G. Habashi (ed.), Pineridge Press, Swansea, 1985.

43. Basdevant, C., M. Deville, P. Haldenwang, J. M. Lacroix, J. Ouazzani, R. Peyret, P. Orlandi, and A. T. Patera, "Spectral and finite difference solutions of the Burgers equation," Computers and Fluids, Vol. 14, 1986, pp. 23-41.

44. Zang, T. A. and M. Y. Hussaini, "Numerical experiments on subcritical transition mechanisms," AIAA Paper No. 85-0296, 1985.

45. Zang, T. A. and M. Y. Hussaini, "On spectral multigrid methods for time-dependent Navier-Stokes equations," Appl. Math. Comp., Vol. 19, 1986, pp. 359-372.

46. Ghaddar, N. K., K. Z. Korczak, B. B. Mikic, and A. T. Patera, "Numerical investigation of incompressible flow in grooved channels. Part 1: Stability and self-sustained oscillations," J. Fluid Mech., Vol. 163, 1986, pp. 99-127.

47. Ghaddar, N. K., M. Magen, B. B. Mikic, and A. T. Patera, "Numerical investigation of incompressible flow in grooved channels. Part 2: Resonance and oscillatory heat transfer enhancement," J. Fluid Mech., Vol. 168, 1986, pp. 541-567.

48. Gottlieb, D., L. Lustman, and S. A. Orszag, "Spectral calculations of one dimensional inviscid compressible flows," SIAM J. Sci. Stat. Comp., Vol. 2, 1981, pp. 296-310.

49. Cornille, P., "A pseudospectral scheme for the numerical calculation of shocks," J. Comp. Phys., Vol. 47, 1982, pp. 146-159.

50. Zang, T. Z. and M. Y. Hussaini, "Mixed spectral-finite difference approximations for slightly viscous flows," Proceedings of the 7th International Conference on Numerical Methods in Fluid Dynamics, W. C. Reynolds and R. W. MacCormack (Editors), Lecture Notes in Physics, No. 141, Springer-Verlag, New York, 1981, pp. 461-446.

51. Hussaini, M. Y., D. A. Kopriva, M. D. Salas, and T. A. Zang, "Spectral methods for the Euler equations. Part 1: Fourier methods and shock-capturing," AIAA J., Vol. 23, 1985, pp. 64-70.

52. Hussaini, M. Y., D. A. Kopriva, M. D. Salas, and T. A. Zang, "Spectral methods for the Euler equations: Part II - Chebyshev methods and shock fitting," AIAA J., Vol. 23, 1985, pp. 234-340.

53. Kopriva, D. A., T. A. Zang, M. D. Salas, and M. Y. Hussaini, "Pseudo-spectral solution of two dimensional gas dynamics problems," Proc. 5th GAMM Conference on Numerical Methods on Fluid Mechanics, M. Pandolfi and R. Piva (Editors), Friedr Vieweg & Sohn, Braunschweig, Wiesbaden, 1983, pp. 185-192.

54. Kopriva, D. A., "A spectral multidomain method for the solution of hyperbolic systems," Applied Numerical Mathematics, Vol. 2, 1986, pp. 221-241.

55. Kopriva, D. A., "A multidomain spectral collocation computation of the sound generated by a shock-vortex interaction," Proc. First IMACS Symposium on Computational Acoustics, Yale University, 1986.

FIGURE CAPTIONS

Figure 1. The domain and Legendre spectral element discretization used for solution of the Laplace equation described in the text. Although results are given for a single isoparametric element, similar results are obtained with multiple elements. (Legendre results due to E. M. Ronquist.)

Figure 2. A plot of the L_∞ error as a function of the number of points in one direction for solution of Laplace's equation in the domain shown in Figure 1.

Figure 3. Description of the domain and boundaries for the Stefan problem presented in the text. (Stefan problem results due to E. M. Ronquist.)

Figure 4. Spectral element prediction for the position of the interface, ∂D_I, for the Stefan problem described in the text. The spectral element mesh uses two elements, one each in each phase.

Figure 5. Temperature (θ) distribution for the Stefan problem described in the text. Note the discontinuity of slope at the interface, ∂D_I.

Figure 6. Spectral element mesh used for simulation of flow past a cylinder. Note the flexible resolution afforded by the elemental decomposition. (Flow past a cylinder results due to G. E. Karniadakis.)

Figure 7. Instantaneous streamlines of the cylinder flow at a Reynolds number of $R = 100$.

Figure 8. Instantaneous isotherms of the cylinder flow at a Reynolds number of $R = 100$ (Prandtl number of unity). The von Karman vortex street can be clearly seen in the temperature distribution behind the cylinder.

Figure 9. Pressure contours for a shock/vortex interaction. (a) Single domain calculation. (b) Three vertical domain calculation.

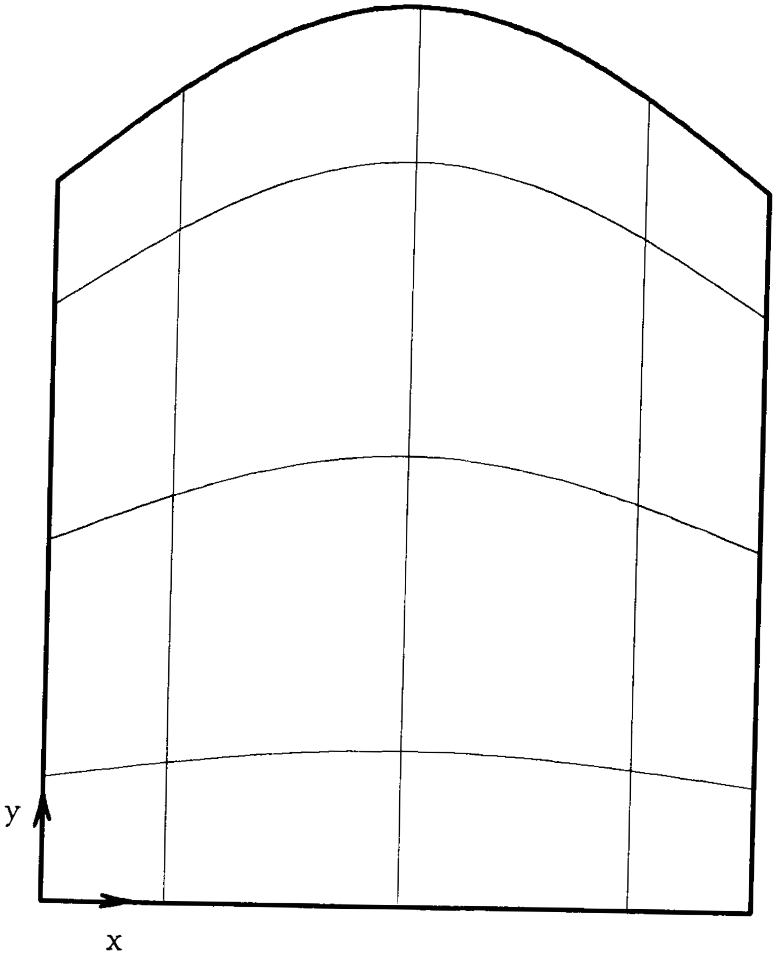

Figure 1.

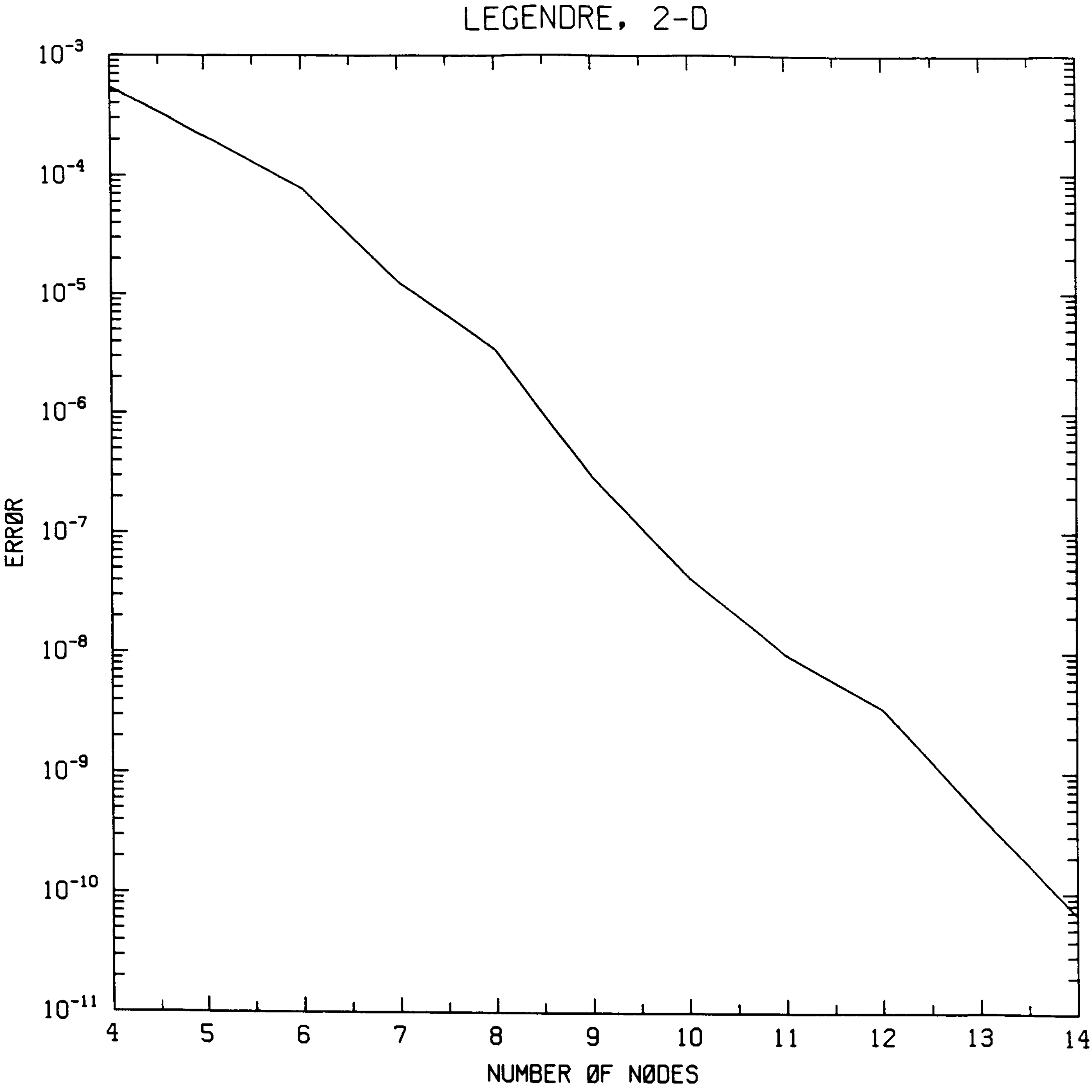
LEGENDRE, 2-D
ERRØR
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
4
5
6
7
8
9
10
11
12
13
14
NUMBER ØF NØDES

Figure 2.

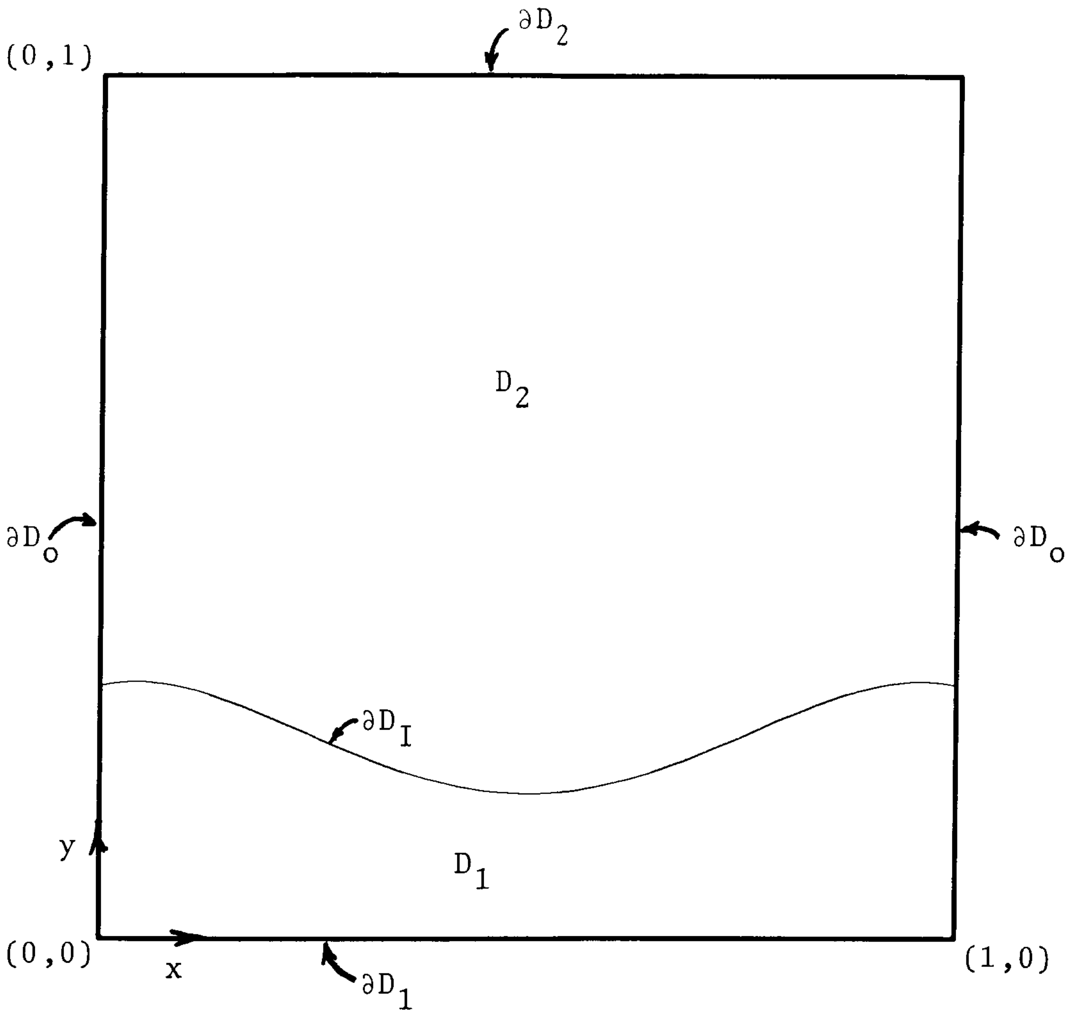
∂D_2
(0,1)
D_2
∂D_o
∂D_o
∂D_I
y
D_1
(0,0)
x
∂D_1
(1,0)

Figure 3.

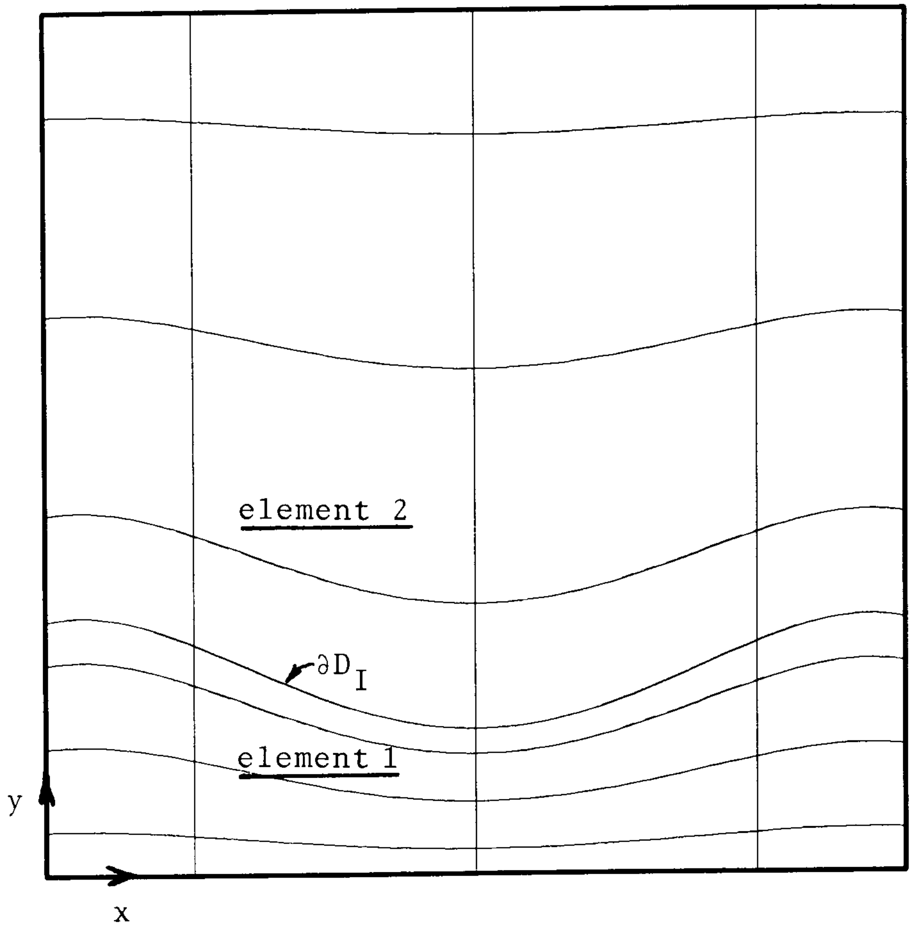

Figure 4.

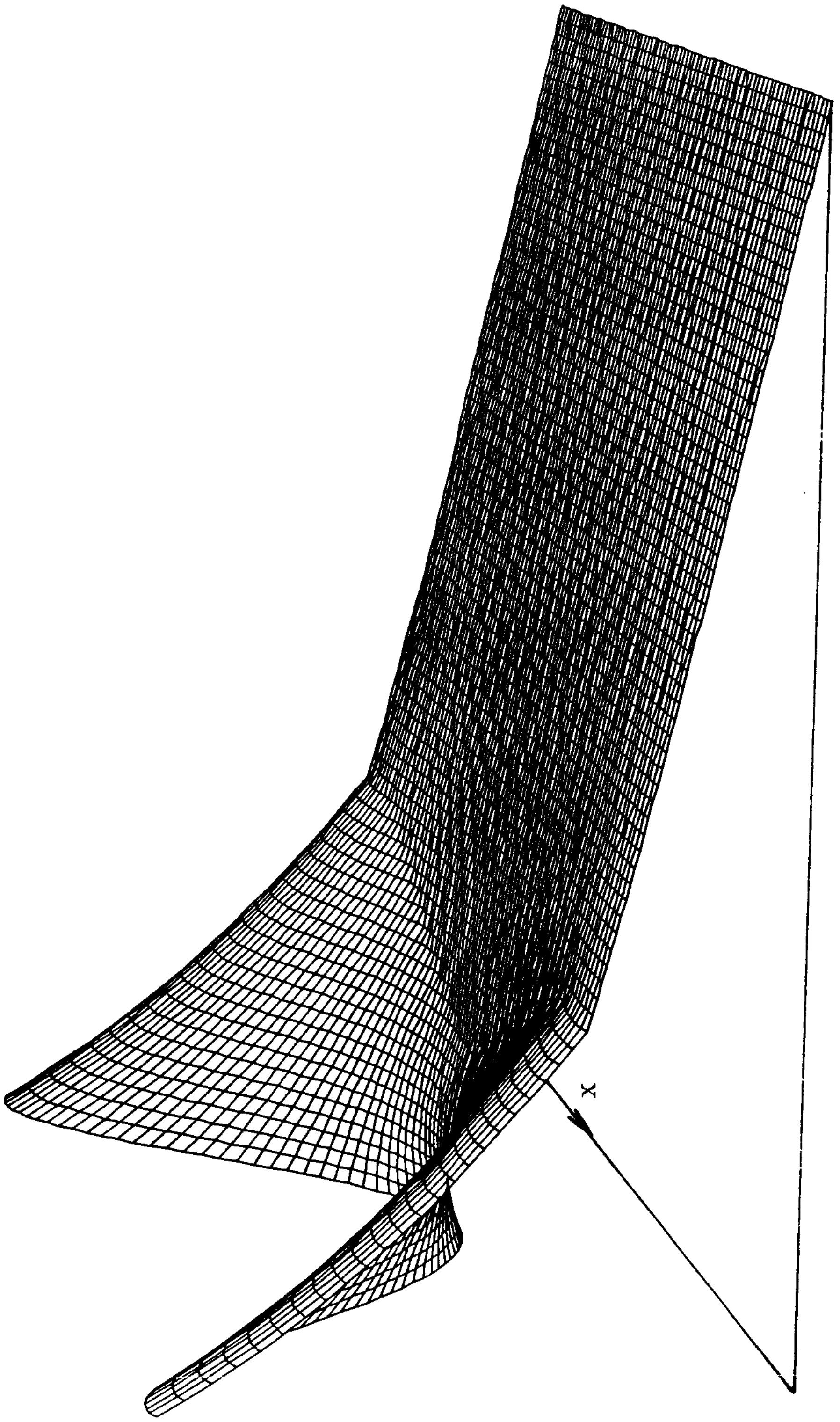

Figure 5.

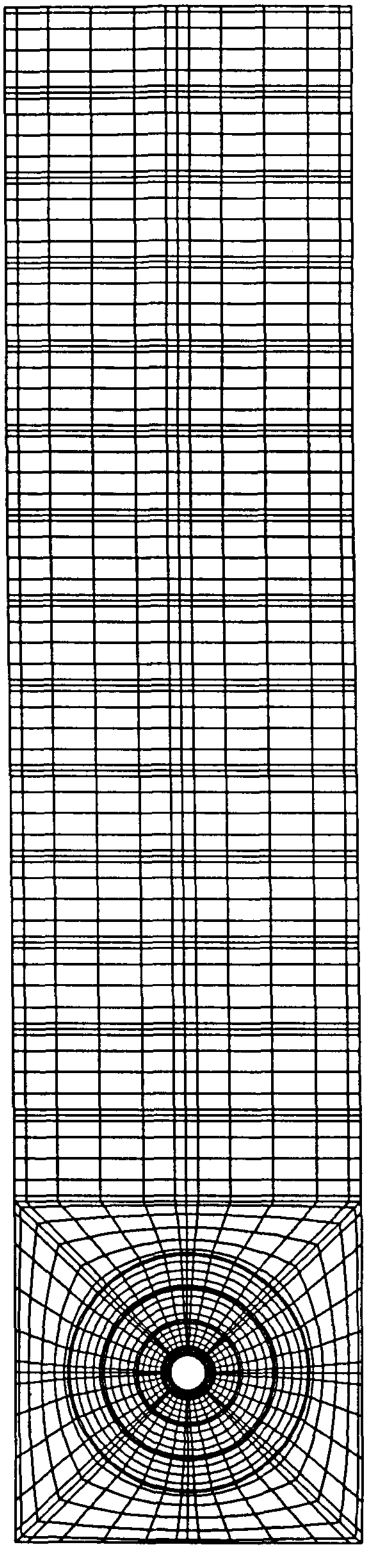

Figure 6.

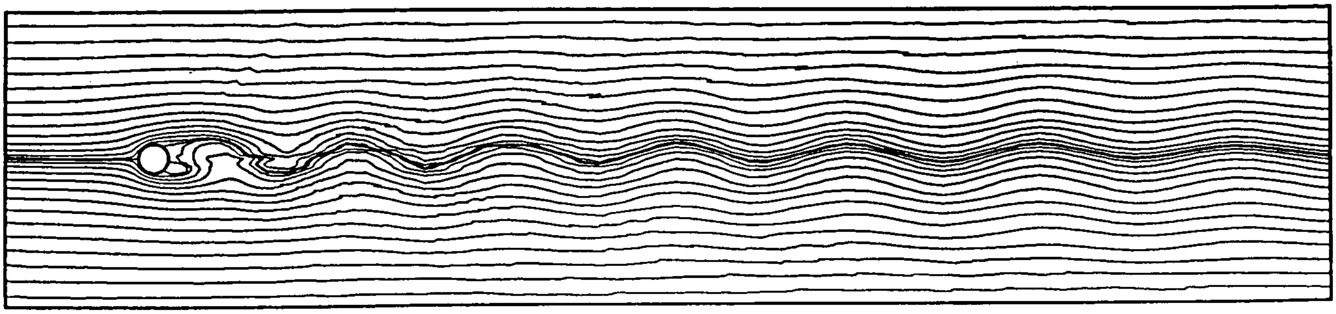

Figure 7.

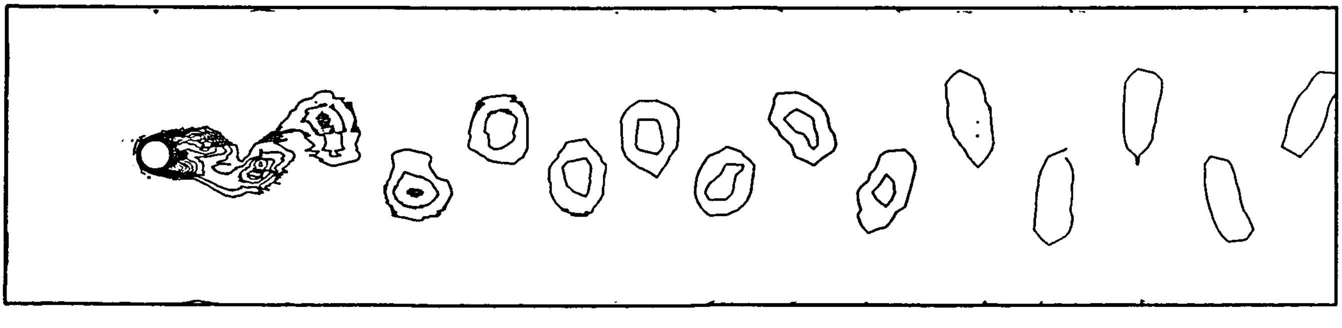

Figure 8.

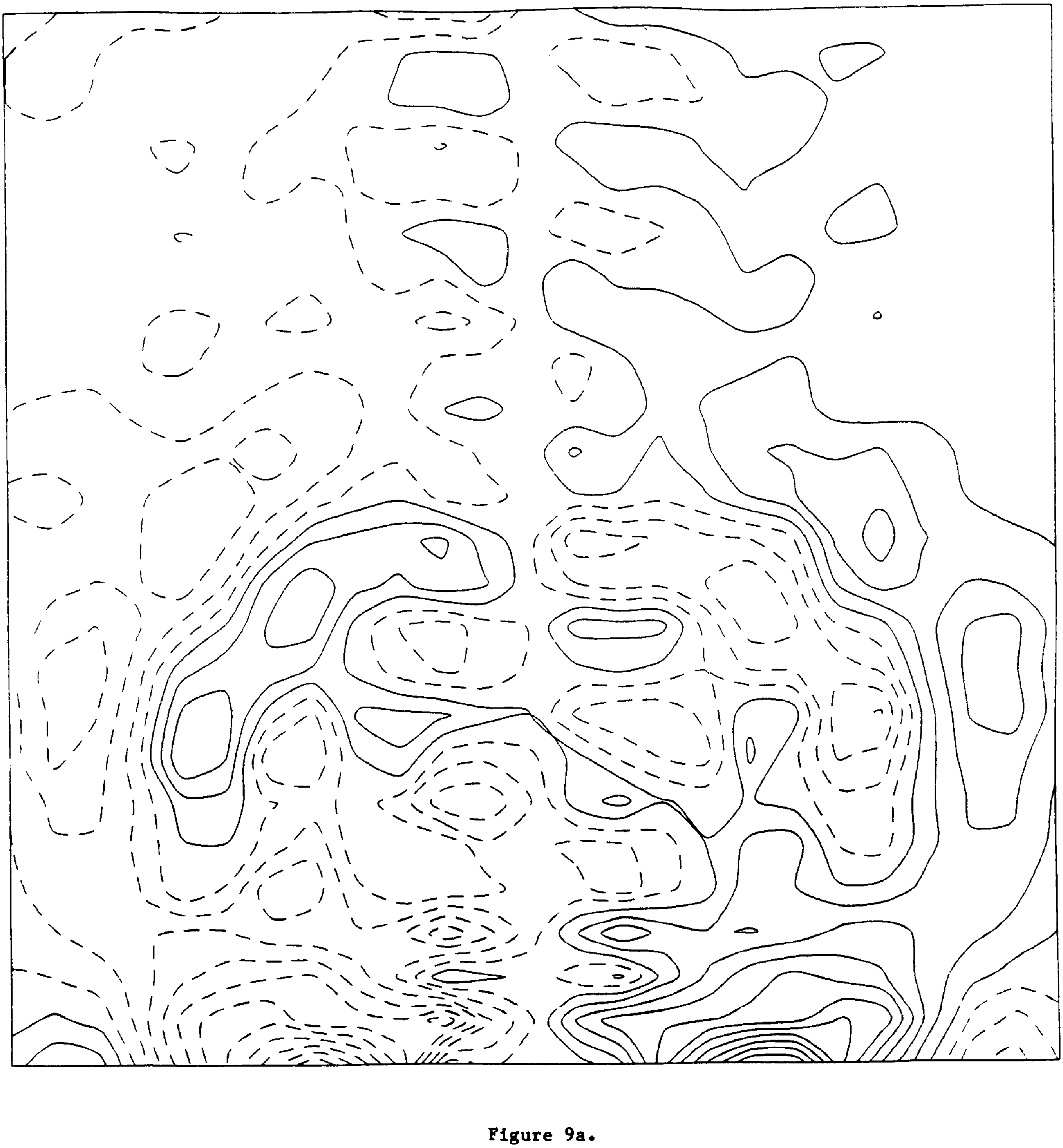

Figure 9a.

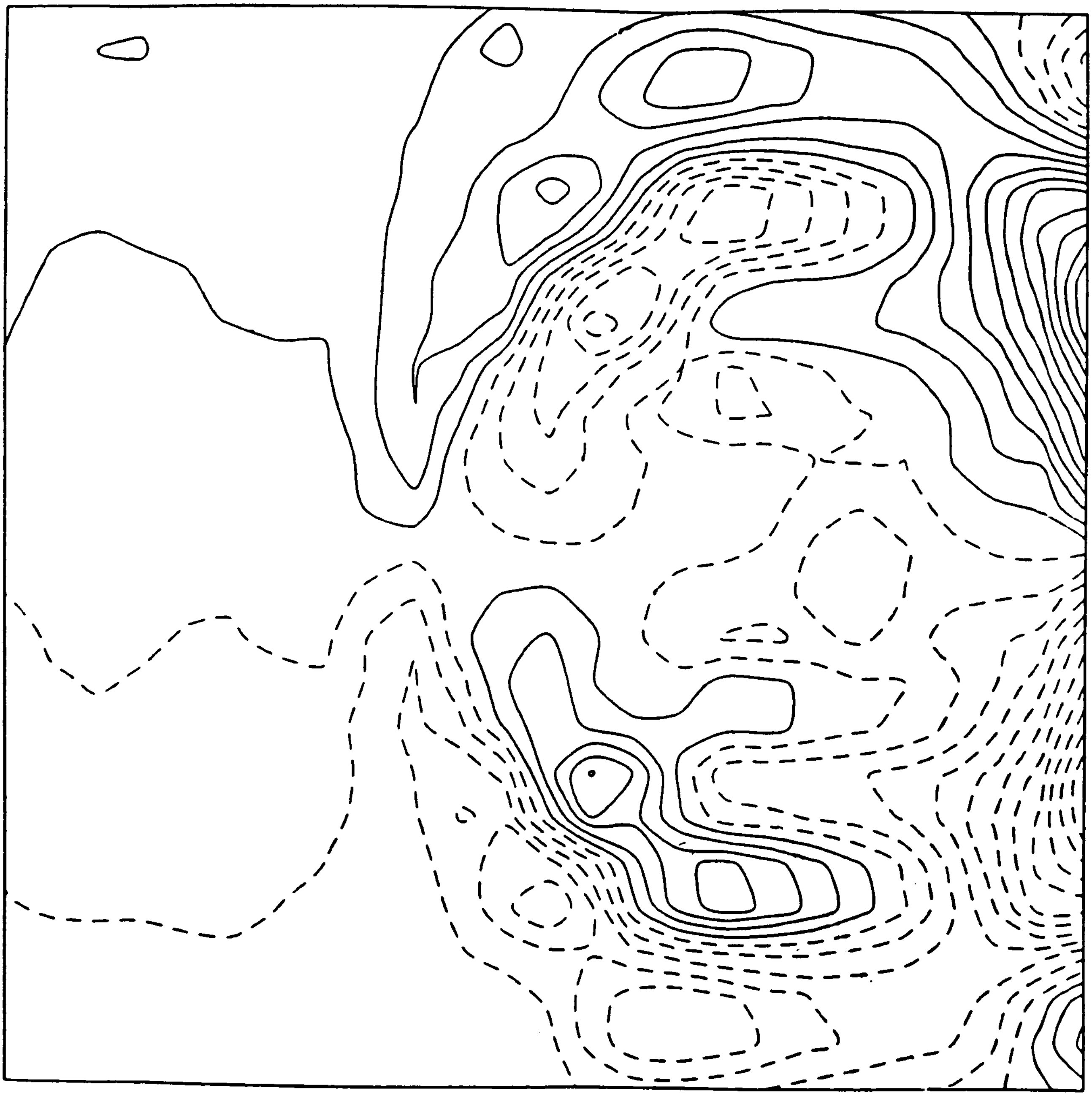

Figure 9b.

Table 1. Preconditioned Eigenvalues for One-dimensional First Derivative Model Problem

Preconditioning	Eigenvalues
Central Differences	$\frac{k\Delta x}{\sin(k\Delta x)}$
High Mode Cutoff	$\frac{k\Delta x}{\sin(k\Delta x)} \qquad \lvert k\Delta x\rvert \leq (2\pi/3)$ $0 \qquad (2\pi/3) < \lvert k\Delta x\rvert \leq \pi$
One-sided Differences	$e^{-i(k\Delta x/2)} \frac{k\Delta x/2}{\sin((k\Delta x)/2)}$
Staggered Grid	$\frac{(k\Delta x)/2}{\sin((k\Delta x)/2}$

Table 2. Damping Factors for N = 64

p	Single-Grid	Multigrid
1	.9980	.9984
2	.9922	.9938
4	.9688	.9750
8	.8751	.9000
12	.7190	.7750
16	.5005	.6000
20	.2195	.3750
24	.1239	.1000
28	.5298	.2250
32	.9980	.6000

Table III. Comparison of Methods for Solution of Burger's Equation (from Ref. 43, 37)

Method	Interval	$\left\|\frac{\partial u}{\partial x}\right\|_{max}$	$\pi \cdot t_{max}$	N/M	$\Delta t \cdot \pi$
- Fourier Galerkin	[-1,1]	151.942	1.6035	682/1024	5.10^{-4}
		142.665	1.60	682/1024	10^{-2}
		148.975	1.603	170/256	5.10^{-4}
		142.313	1.60	170/256	10^{-2}
- Fourier Pseudospectral	[-1,1]	142.606	1.60	256/256	10^{-2}
		144.237	1.60	128/128	10^{-2}
- ABCN collocation	[-1,1]	145.877	1.60	512	5.10^{-3}
+coordinate transform	[-1,1]	152.123	1.60	64	10^{-2}
- Spectral Element	[-1,1]	152.04	1.6033	16 x 4	$10^{-2}/6$
- FD	[-1,1]	150.1	1.63	81	10^{-2}
- Chebyshev					
ABCN spectral	[0,1]	152.05	1.60	64	1/300
Rosenbrook spectral	[0,1]	151.998	1.60	64	10^{-2}
	[0,1]	150.144	1.60	32	10^{-2}
ABCN collocation	[0,1]	152.126	1.60	64	10^{-2}
- Flux balance	[-1,1]	152.00011			
- Analytical		152.00516	1.6037		

Table IV. Maximum Error in p for MacCormack and Spectral Computation of Transonic Ringleb Flow

Grid	MacCormack	Spectral
9 x 5	2.6×10^{-2}	2.2×10^{-2}
17 x 9	1.1×10^{-2}	1.9×10^{-3}
33 x 17	3.2×10^{-3}	5.0×10^{-5}

Table V. Solutions to (102) with Equal Number of Points on Each Side of the Interface

N	Error in u	Error in v
8	1.57×10^{-2}	1.49×10^{-2}
16	4.15×10^{-6}	4.86×10^{-6}
32	1.91×10^{-9}	1.91×10^{-9}

Table VI. Effect of Streamwise Mesh Distribution on Ringleb Calculation

Grid	Division	Maximum Error
8 + 8	0.45 + 0.55	7.8×10^{-4}
8 + 8	0.50 + 0.50	9.3×10^{-4}
16 (SD)	-	1.9×10^{-3}
10 + 6	0.34 + 0.66	1.2×10^{-2}

National Aeronautics and Space Administration

Report Documentation Page

1. Report No.
NASA CR-178373
ICASE Report No. 87-62

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
SPECTRAL COLLOCATION METHODS

5. Report Date
September 1987

6. Performing Organization Code

7. Author(s)
M. Y. Hussaini, D. A. Kopriva, A. T. Patera

8. Performing Organization Report No.
87-62

10. Work Unit No.
505-90-21-01

9. Performing Organization Name and Address
Institute for Computer Applications in Science
and Engineering
Mail Stop 132C, NASA Langley Research Center
Hampton, VA 23665-5225

11. Contract or Grant No.
NAS1-18107

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23665-5225

13. Type of Report and Period Covered
Contractor Report

14. Sponsoring Agency Code

15. Supplementary Notes
Langley Technical Monitor:
Richard W. Barnwell

Submitted to IMACS J. Numer. Anal.

Final Report

16. Abstract

This review covers the theory and application of spectral collocation methods. Section 1 describes the fundamentals, and summarizes results pertaining to spectral approximations of functions. Some stability and convergence results are presented for simple elliptic, parabolic and hyperbolic equations. Applications of these methods to fluid dynamics problems are discussed in Section 2.

17. Key Words (Suggested by Author(s))
spectral methods, Navier-Stokes equations

18. Distribution Statement
34 - Fluid Mechanics and Heat Transfer
64 - Numerical Analysis

Unclassified - unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	70	A04

NASA FORM 1626 OCT 86

NASA-Langley, 1987