Mixed Voltage VLSI Design [1]

Ramesh Panwar[§][†], David Rennels[§][†], and Leon Alkalaj[§]

[§]Jet Propulsion Laboratory　　　　　[†]Department of Computer Science
California Institute of Technology　　　　University of California
Pasadena, CA 91109　　　　　　　　Los Angeles, CA 90024

panwar, rennels, leon@telerobotics.jpl.nasa.gov

Abstract – We present a technique for minimizing the power dissipated in a VLSI chip by lowering the operating voltage without any significant penalty in the chip throughput even though low voltage operation results in slower circuits. Since the overall throughput of a VLSI chip depends on the speed of the critical path(s) in the chip, it may be possible to sustain the throughput rates attained at higher voltages by operating the circuits in the critical path(s) with a high voltage while operating the other circuits with a lower voltage to minimize the power dissipation. The interface between the gates which operate at different voltages is crucial for low power dissipation since the interface may possibly have high static current dissipation thus negating the gains of the low voltage operation. We present the design of a voltage level translator which does the interface between the low voltage and high voltage circuits without any significant static dissipation. We then present the results of the mixed voltage design using a greedy algorithm on three chips for various operating voltages.

1   Introduction

The power consumed in CMOS VLSI primarily results from charging and discharging the node capacitances in the VLSI circuit. The CV^2f power dissipated in a VLSI chip depends on the square of the operating voltage [3]. Thus operating a chip with 3V instead of 5V results in 64% power savings. Low voltage operation is one way of achieving low power in systems. Lower voltage operation also results in improved signal integrity and reliability due to lower crosstalk and ground bounce. Space systems that require low power with high reliability should find the migration to lower voltages a very attractive proposition. Lowering the operating voltage has an adverse effect on the speed of the circuits. Thus low voltage design reduces the overall chip throughput unless some other compensating measures are taken. In this paper, we propose a mixed voltage design technique which addresses the problem of lowering the power dissipation without affecting the chip throughput. The circuits in the critical path(s) are operated with the higher voltage to maintain the chip throughput while the circuits that are not in the critical paths are operated with the lower voltage to minimize the power dissipation. The interface between circuits that operate at different voltages can have large static current dissipation because the transistors in both the p-block

and the n-block may be on resulting in a conducting path from the power to the ground rail. We present a voltage level translator that can be used in the interface. The translator is implemented using 6 transistors with appropriate transistor sizing. The translator has an excellent DC transfer characteristic and fast propagation delays with very minimal static current dissipation (of the order of pico-amperes). We then present a greedy algorithm for doing the mixed voltage design. The algorithm is not optimal but gives very satisfactory results. Finally we present the results of the mixed voltage design algorithm on three chips that were designed at JPL for various flight projects using the Honeywell RICMOS process [4].

2 Mixed Voltage Design

2.1 Low Voltage Operation

The delay ($t_d = 4C_L/\beta V$) of a VLSI circuit depends on the inverse of the operating voltage [3] while the power ($P = CV^2 f$) consumed by the circuit depends on the square of the voltage. It follows from this observation that if we decrease the operating voltage of a chip, then we must increase the clock cycle time to allow for the slower operation of the circuits on the chip. Increasing the chip clock cycle time implies a reduced overall chip throughput. A way to overcome the problem of reduced chip throughput for low voltage design is by architectural techniques like pipelining and parallelism as discussed in [2]. Two problems with the architectural solutions outlined in [2] is that they involve architectural redesign and they require more chip real-estate. Essentially, the techniques proposed in [2] trade off power for area. We propose a technique which does not require any architectural redesign. The technique relies on the fact that the chip clock cycle time (and throughput) depend on the speed of the critical path(s). The elements in the critical path(s) may be operated with a high voltage to ensure that there is no (or minimal) loss of chip throughput. The elements that are not in the critical paths may be operated with the lower voltage to ensure lower power dissipation. Let us consider two operating voltages V_{ddh} and V_{ddl} for operating a VLSI circuit where $V_{ddh} > V_{ddl}$. We begin with some definitions to illustrate the tradeoffs in mixed voltage design.

Definition 1 *The switching energy ratio r_{SE} of a circuit with respect to two voltages V_{ddh} and V_{ddl} is the ratio of the energy required for switching the circuit with the lower voltage to the energy required for switching the circuit with the higher voltage.*

Definition 2 *The switching delay ratio r_{SD} of a circuit with respect to two voltages V_{ddh} and V_{ddl} is the ratio of the switching delay of the circuit with the lower voltage to the switching delay of the circuit with the higher voltage.*

Definition 3 *The chip energy ratio r_{CE} of a chip with respect to two voltages V_{ddh} and V_{ddl} is the ratio of the switching energy of the chip with a mix pf V_{ddh} and V_{ddl} operation to the switching energy of the chip with V_{ddh} operation alone.*

Definition 4 *The chip delay ratio r_{CD} of a chip with respect to two voltages V_{ddh} and V_{ddl} is the ratio of the fastest chip clock cycle time with a mix of V_{ddh} and V_{ddl} operation to the fastest chip clock cycle time with V_{ddh} operation alone.*

Definition 5 *The MVD ratio r_{MVD} of a chip is the ratio of swtching energy ratio and the chip energy ratio.*

Given the expression for the circuit delay and power dissipation, we have the following relationships:

$$r_{SE} = \frac{V_{ddl}^2}{V_{ddh}^2}$$

$$r_{SD} = \frac{V_{ddh}}{V_{ddl}}$$

The chip delay ratio r_{CD} is a parameter that should be determined by the system architect. The parameter will depend on how much loss of throughput can be tolerated by the chip when we go from a high voltage operation to a mixed voltage operation. For example, if a compression chip is designed such that its maximum path delay is 87 ns but the chip needs to handle input data rates of 10 MHz, then the chip can withstand a chip delay ratio of approximately 1.15. Once an acceptable value for the chip delay ratio has been determined by the system architect, the technique of mixed voltage design may used to lower the chip power dissipation. The reduced power of the chip will result in a chip energy ratio which is a function of the switching delay ratio and chip delay ratio:

$$r_{CE} = f(r_{SD}, r_{CD})$$

It should be noted that the switching energy ratio is dependent on the switching delay ratio:

$$r_{SE} = 1/r_{SD}^2$$

2.2 Voltage Level Translator

In a chip with mixed voltage design, a gate that operates with the higher voltage V_{ddh} may receive inputs from gates that operate with the lower voltage V_{ddl}. In such a case, the gate input swings between V_{ss} and V_{ddl}. The input signal may produce a conducting path between V_{ddh} and V_{ss} in the gate when it is at V_{ddl} since both the p-transistors and the n-transistors in the path may be on continuously. This causes a large static current dissipation in the gate which may negate the gains of low voltage operation. Hence, we need a voltage level translator to translate the low voltage signals produced by gates that operate with V_{ddl} to the high voltage signals that can be applied as inputs to the gate operated with V_{ddh}. It should be noted that the reverse voltage level translator to translate high voltage signals to low voltage signals is not required since high voltage signals driving a gate operated at a lower voltage can never produce a conducting path between the power and ground rails (other than the transient conducting path that produces the short-circuit current found in conventional CMOS). We have designed a voltage level translator using 6 transistors which allows us to convert low voltage signals (signal swing between V_{ss} and V_{ddl}) to high voltage signals (signal swing between V_{ss} and V_{ddh}). The desireable characteristics of such a translator are the following:

- The translator should have good DC characteristics for noise tolerance.

- The translator should have reasonably fast propagation delays.

- The translator should have no (or minimal) static dissipation.

- The translator should not be too expensive in terms of VLSI real-estate.

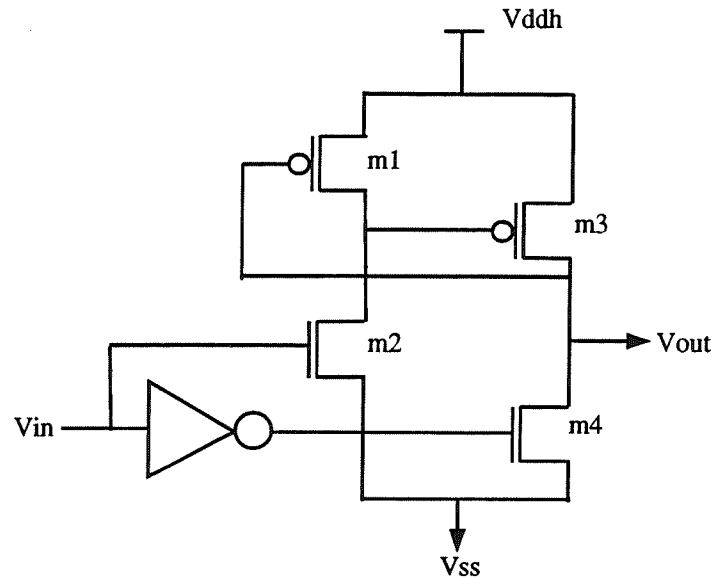- The translator should not require any special processing for VLSI fabrication.

Figure 1: Voltage level translator

The voltage level translator is shown in figure 1. The operation of the translator is as follows:

- Asssume that the input to the translator V_{in} has been steady at the logic level 0 which implies that $V_{in} = V_{ss}$. In this state, transistor $m2$ is turned off while transistor $m4$ is turned on. Since $m4$ is on, the output V_{out} is pulled low. Since V_{out} is fed back to the input of $m1$, transistor $m1$ is on which implies that the gate of transistor $m3$ is pulled to V_{ddh}. Thus, transistor $m3$ is off which allows the output V_{out} to remain at V_{ss}. It should be noted that there is no conducting path from V_{ddh} to V_{ss} in this state.

- When the input to the translator switches from V_{ss} to V_{ddl}, then transistor $m2$ is turned on while transistor $m4$ is turned off. Since transistor $m1$ is on at the same time, there is a conducting path from V_{ddh} to V_{ss} which leads to a flow of short-circuit current. However, if transistor $m2$ is larger than transistor $m1$, then it will overpower transistor $m1$ and discharge the gate input of transistor $m3$ thus forcing $m3$ to turn on. When $m3$ turns on, the output V_{out} is pulled to V_{ddh} which turns off transistor $m1$, thus allowing the translator to settle into a stable state with $V_{out} = V_{ddh}$. There is no conducting path from V_{ddh} to V_{ss} in the stable state thus eliminating the possibility of any static current dissipation.

- When the translator input changes back to V_{ss}, transistor $m2$ is turned off while transistor $m4$ is turned on. Since transistor $m3$ is on at the same time, there is a conducting path from V_{ddh} to V_{ss} which leads to a flow of short-circuit current. However, if transistor $m4$ is larger than transistor $m3$, then it will overpower transistor $m3$ and discharge the gate input of transistor $m1$ thus forcing transistor $m1$ to turn on. $Vout$ consequently goes low. When $m1$ turns on, the gate input of transistor $m3$ is pulled to V_{ddh} which turns off transistor $m3$ thus allowing the translator to settle in the stable state with $V_{out} = V_{ss}$. There is no conducting path from V_{ddh} to V_{ss} in the stable state thus eliminating the possibility of any static current dissipation.
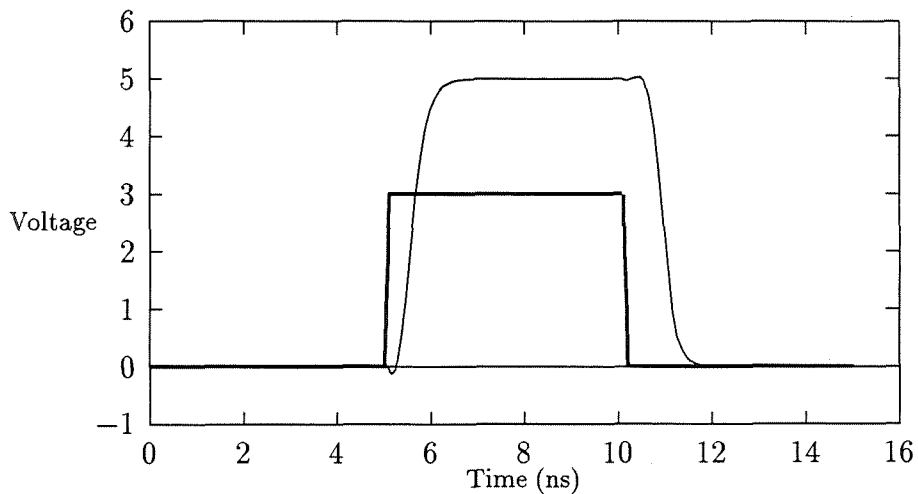

Figure 2: Pulse response

Figure 2 shows the response of the voltage level translator to a 5 ns pulse. The simulations were done using SPICE with models for 2μ devices from [1]. A critical parameter in the design of the voltage level translator is the sizing of transistors $m2$ and $m4$ relative to $m1$ and $m3$. For the simulations, we sized the transistors such that the W/L ratios of $m2$ and $m4$ were 5 times those of $m1$ and $m3$. This ratio is sufficient for attaining a reasonably fast propagation delay as shown in figure 2. The rise time delay in the simulation was 1.08 ns and the fall time delay was 1.18 ns. Figure 3 shows the DC transfer characteristic of the voltage level translator. The figure shows that the translator output has a very sharp transition when the input voltage is approximately 1.4V (assuming that the translator voltage levels are 3V and 5V). The sharpnesss of the response implies a very good noise immunity. The static current dissipation in the translator was 13.9 pA.
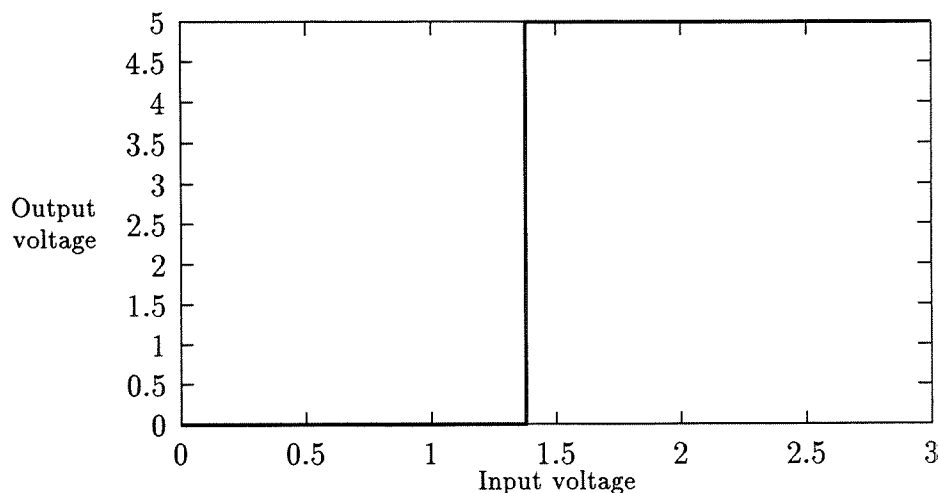
Figure 3: DC transfer characteristic

2.3 Optimization Problem

For the purpose of illustrating the optimization problem in mixed voltage design, we shall assume a standard cell based design. Each cell may have a number of inputs and outputs and may occur on a number of different paths in the chip. We denote the delay presented by cell i to path j by d_{ij}. Assume that each cell i has a weight w_i which is proportional to the internal capacitance of the cell. The weight of a cell is used to select between two cells for the optimization algorithm since a cell with a higher weight will lead to greater power savings if the cell is operated with the lower voltage. A path is essentially an ordered set of cells which occur on the path. Let us denote the set of cells representing the path j by S_j. If we use D_c to denote the chip critical path delay when the entire chip is operated at V_{ddh}, then the optimization problem can be cast as follows:

$$\text{Maximize} \qquad \sum_{i=1}^{N} \delta_i w_i$$
$$\text{subject to} \quad \sum_{i \epsilon S_j} (1 - \delta_i + \delta_i r_{SD}) d_{ij} \leq r_{CD} D_c \quad \forall j$$

where

$$\delta_i = \begin{cases} 1 & \text{if cell } i \text{ is operated with lower voltage} \\ 0 & \text{otherwise} \end{cases}$$

3 Experimental Results

3.1 Chip Benchmarks

We considered three chip designs that were done at JPL for various flight projects as benchmarks for our study of power optimization using mixed voltage design. The chips were

designed using the Honeywell RICMOS process [4] by three different VLSI designers and consequently had very different path length distributions. The chips were approximately of the same size and various statistics on the chips are shown in table 1. The table shows the

	chip0	chip1	chip2
n_{paths}	2699	2846	4123
n_{cells}	1936	2121	2038
n_{regs}	491	552	519

Table 1: Chip statistics

total number of paths, cells and registers in each of the chips. The cells used in the chip designs ranged in complexity from a simple inverter to a 4-to-1 multiplexer. The designs had a total of 39 different types of cells. The weight w_i of a cell was assumed to be the number of gate equivalents used by the cell [4]. Figures 4-6 show the distribution of path lengths in the three chip designs. The length of a path is the total number of cells in a path. The cumulative frequency (on the y-axis) is the total fraction of paths whose length exceed that of the path length shown on the x-axis. The figures and table 1 show that one of the chips (*chip2*) was significantly different from the other two chips. The largest path length in *chip2* was 34 while the largest path length for the other designs was 18. Also, *chip2* had a very large number of paths compared to the other chips inpite of the fact that all the chips had about the same number of cells and registers.
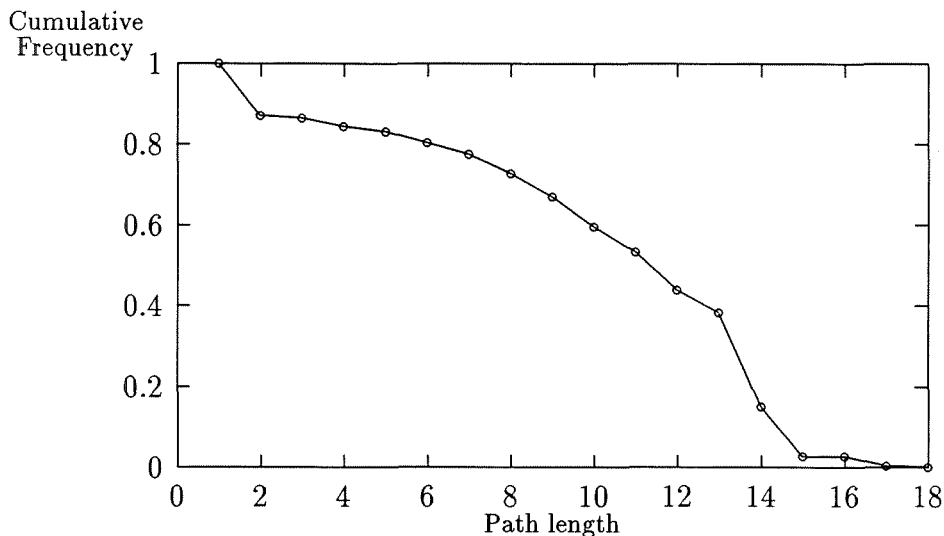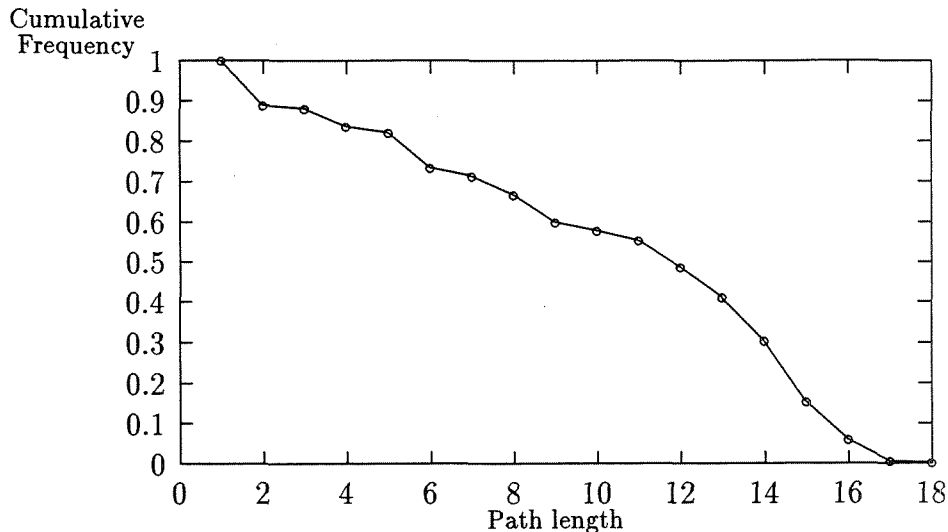

Figure 4: Path length distributions in *chip0*
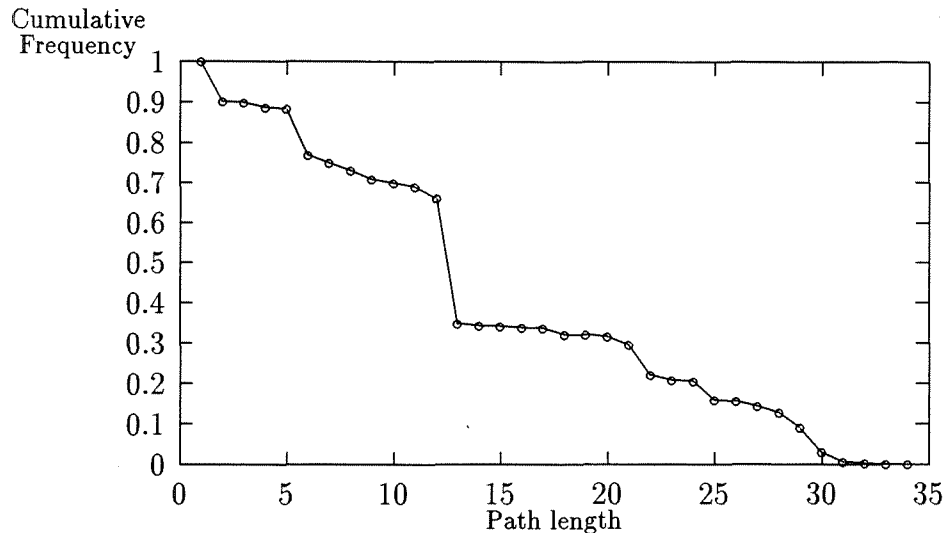
Figure 5: Path length distributions in *chip*1

3.2   Greedy Optimization Strategy

We adopt a greedy optimization algorithm for mixed voltage design. The input to the algorithm is an adjacency list representation of the chip design. Each list on the adjacency list represents a path. We sort the cells by their weights and then traverse down the sorted list of cells to see which cells can have their voltage lowered to V_{ddl}. The voltage of a cell can be lowered to V_{ddl} if the delays of all the paths that contain the cell are still bounded by $r_{CD}D_c$ when we lower the voltage of the cell to V_{ddl}. It should be apparent to the reader that a greedy strategy on the cell weights is not optimal. To prove the non-optimality of the greedy strategy, consider three cells i, j, and k on path S_l. If the weights of the cells are such that $w_i > w_j > w_k$, then during the execution of the greedy algorithm, cell i will have its voltage lowered to V_{ddl}. However, doing so may preclude cells j and k from having their voltages lowered because we may violate the timing constraint if the delay of S_j exceeds $r_{CD}D_c$. On the other hand, if we had lowered the voltages of j and k first, we may have still stayed within the bounds imposed by the timing constraint and yet have attained a lower power dissipation. It is easy to see that this will be the case if the following conditions hold:

$$d_{jl} + d_{kl} < d_{il}$$

$$w_j + w_k > w_i$$

One may consider other types of greedy strategies (involving greed on the ratio of the cell weight and the cell delay) or other approaches using dynamic programming. The greedy algorithm (which is greedy on the cell weigths) gives us very good results for power savings and we did not investigate the other greedy approaches or dynamic programming approaches for this paper. We plan to look into the other approaches for future research.

Figure 6: Path length distributions in *chip2*

3.3   Power Savings

Figures 7-9 show the results of the greedy strategy on the three chip benchmarks. The figures show the MVD ratio (r_{MVD}) as a function of the chip delay ratio (r_{CD}) for various values of the switching delay ratio r_{SD}. The MVD ratio should be a monotonically increasing function of the chip delay ratio. The non-monotonicity of the MVD ratio as shown in the figures is an artifact of the non-optimality of the greedy algorithm. The power savings is given by $(1 - 1/r_{SD}^2 r_{MVD}) \times 100\%$. It can be seen that all the chip designs have a power savings of 65-70% without any significant penalty in the chip throughput ($r_{CD} = 1$) and that all the designs have a power savings of about 83% if the chips are allowed to be slowed down by a factor of 2. The MVD ratio is almost a linear function of the chip delay ratio which implies that the power dissipated is inversely proportional to the chip delay ratio. A higher value of r_{SD} (as a consequence of lower V_{ddl}) results in lower power dissipation up to a certain point after which the power dissipated is constant. For example, for *chip0*, the power saved is approximately 65% (for chip delay ratios of $r_{SD} = 2.0$ and $r_{SD} = 2.5$) which implies that any futher increase of r_{SD} by reducing V_{ddl} will not result in any more power savings. In fact since the power savings is identical for both values of r_{SD}, the designer may prefer operating the chip with $r_{SD} = 2.0$ since the MVD ratio r_{MVD} is higher for $r_{SD} = 2.0$ as compared to the MVD ratio for $r_{SD} = 2.5$. The higher MVD ratio implies less translators in the design though we have not quantified this in our study yet. Future work will address this relationship along with analytical models of the MVD ratio and the power savings that can result.
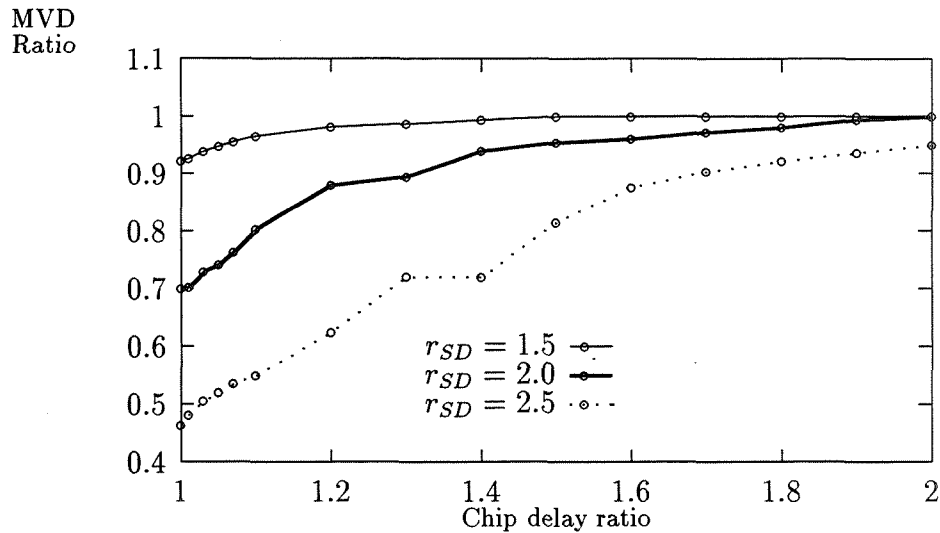
MVD
Ratio

Figure 7: Mixed voltage design results for *chip*0

4 Conclusions

We have presented a technique for minimizing the power dissipation in CMOS VLSI by operating the VLSI chip with two voltages. The higher voltage is used to ensure that the chip throughput isn't adversely affected by the low voltage operation while the lower voltage is used for the majority of the cells in the chip to ensure low power dissipation. The key to designing chips with two operating voltages is an innovative voltage level translator that we designed using 6 transistors that allows us to translate low voltage signals to high voltage signals without any significant static power dissipation. Results of using the mixed voltage design technique with a greedy optimization strategy were presented for three VLSI chips with considerably different designs showing that it is possible to save about 65-70% of the power dissipated in the chips without paying any penalty in the chip throughput.

5 Acknowledgements

We would like to thank Gary Bolotin who provided the chip designs used as benchmarks in this study.
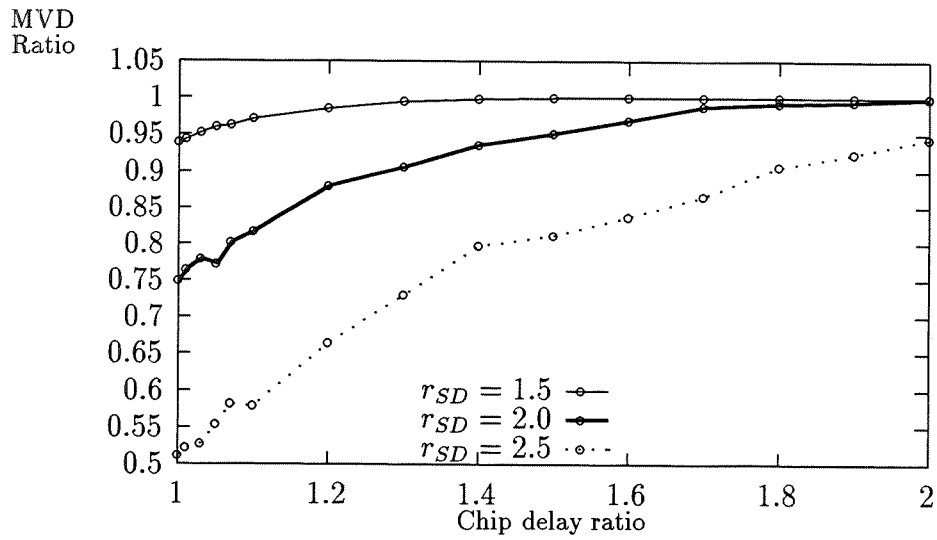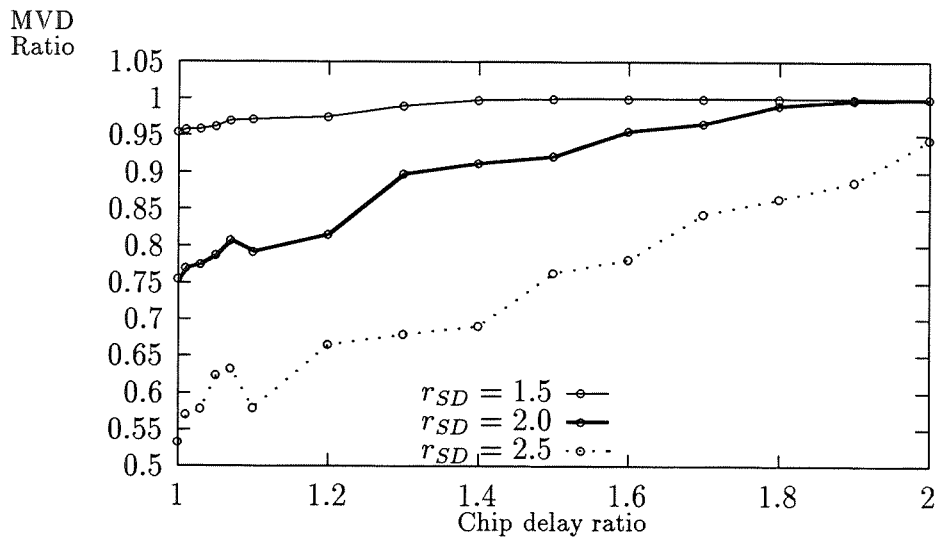
Figure 8: Mixed voltage design results for *chip*1

Figure 9: Mixed voltage design results for *chip*2

References

[1] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

[2] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-Power CMOS Digital Design," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, Apr. 1992.

[3] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*. Reading, MA: Addison-Wesley, 1988.

[4] *The Honeywell RICMOS ASIC Library Data Book*. 1991.