

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220091684>

A Workload Based Lookup Table for Minimal Power Operation Under Supply and Body Bias Control

Article in *Journal of Low Power Electronics* August 2009

DOI: 10.1166/jolpe.2009.1018 · Source: DBLP

CITATIONS

2

READS

866

3 authors, including:



Bharadwaj Amrutur

Indian Institute of Science

187 PUBLICATIONS 2,123 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



10G ethernet [View project](#)



Mohamed Bin Zayed International Robotics Challenge (MBZIRC 2020) [View project](#)

A Workload Based Lookup Table for Minimal Power Operation Under Supply and Body Bias Control

A Thesis
Submitted for the Degree of
Master of Science (Engineering)
In The Faculty of Engineering

by
Sreejith K



Electrical Communication Engineering
Indian Institute of Science, Bangalore
Bangalore - 560 012 (INDIA)
August, 2009

I, hereby declare that I am the sole author of this thesis. I declare that the work reported in this thesis has been carried out in the Circuits and Systems Lab, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore under the supervision of Dr. Bharadwaj Amrutur and in Analog Devices India Pvt. Ltd. under the supervision of Dr. Ashok Balivada. I also declare that this work has not formed the basis for the award of any Degree, Diploma, Fellowship, Associateship or similar title of any University or Institution. I authorize Indian Institute of Science to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Sreejith K
Circuits and Systems Lab,
Department of Electrical Communication Engineering,
Indian Institute of Science,
Bangalore-560 012,
INDIA.
August 2009.

Abstract

Dynamic Voltage Scaling (DVS) and Adaptive body bias (ABB) techniques respectively try to reduce the dynamic and static power components of an integrated circuit. Ideally, the two techniques can be combined to find the optimal operating voltages (VDD and VBB) to minimize power consumption. A combination of the DVS and ABB may warrant the circuit to operate at voltages (supply and body bias) different from the values specified by the two methods working independently. Also, this VDD and VBB values for minimal power consumption varies with the workload of the circuit. The workload can be used as an index to select the optimal VDD/VBB values to minimize the total power consumption. This paper examines the optimal voltages for minimal power operation for typical data path circuits like adders and multiply-accumulate (MAC) units across various process, voltage, and temperature conditions and under different workloads. In addition, a workload based look up table to minimize the power consumption is also proposed. Simulation results for an adder and a multiply-accumulate circuit block indicate a power saving of 12-30% over standard DVS scheme.

List of Publications

Accepted for Publication

1. Sreejith K, Bharadwaj Amrutur and Ashok Balivada, "A Workload Based Lookup Table for Minimal Power Operation Under Supply and Body Bias Control" in Journal of Low Power Electronics, JOLPE - Vol. 5, N° 2, August 2009.

To My Brother

Acknowledgement

I would like to thank my advisor Dr. Bharadwaj Amrutur for encouraging me to do creative research, his invaluable guidance and above all for his patience with me. I thank my co-guide Dr. Ashok Balivada for his guidance and suggestions. I thank my colleagues Mr. Sajkapoor and Dr. Karthik for their support and encouragement to pursue my Masters.

Special thanks to my colleagues Vinoth Kumar and Aswin Srinivasa for helping to set up the simulations. I thank my colleagues Dinith, Ratish and Anoop for helping me in the layout work. I would like to single out Basavaraj Talwar for helping me in a variety of ways.

Contents

Abstract	ii
Acknowledgement	v
Introduction	1
1.1 Introduction.....	1
References.....	6
Motivation	8
2.1 Power Components and Activity.....	8
2.2 Workload Variation.....	10
2.3 Fixed Frequency operating points.....	11
2.4 Summary	17
References.....	18
Test Circuits and Key Results	19
3.1 Introduction.....	19
3.2 Test Circuits	20
3.3 Key Observations	20
3.3.1 Comparison of P and N Body Bias Schemes	20
3.3.2 MAC	23
3.3.3. Ring Oscillator	29
3.3.4 Adder	32
3.4 Summary of Results	33
References.....	36
Workload Indexed Look up Table for Blocks	37
4.1 Introduction.....	37
4.2 Workload Indexed Tables	37
References.....	41
Implementation	42
5.1 Introduction.....	42
5.2 Implementation Methods.....	43
5.3 Algorithm to generate the workload based lookup table for any block	51
References.....	53
Conclusion	54

Tables

Table 2.1: Workloads for different Data Path Units in Sharc DSP for various signal processing algorithms	10
Table 3.1: Power Comparisons when different power saving methods are employed ...	34
Table 4.1: Workload Indexed Lookup table for MAC for Typical process corner	38
Table 4.2: Workload Indexed Lookup table for MAC for Fast process corner	38
Table 4.3: Workload Indexed Lookup table for Ring Oscillator	39

Figures

Fig 2.1: Schematic representation of our VDD-VBB control System	12
Fig 2.2: VDD Vs VBB for a ring oscillator working at a fixed frequency 500 MHz.....	13
Fig 2.3: Leakage and Dynamic Power components for a ring oscillator operating at a fixed frequency 500MHz.....	14
Fig 2.4: Leakage, Dynamic and Total Power components for a ring oscillator at fixed frequency	15
Fig 3.1: Total Power (Ring Oscillator) Vs VDD with adjusted VBB for at-speed operation. Comparison of P and N Body Bias schemes. Equal rise and fall delay	21
Fig 3.2: Total Power (Ring Oscillator) Vs VDD with adjusted VBB for at-speed operation. Comparison of P and N Body Bias. Unequal rise and fall delay.....	22
Fig 3.3: Total Power (MAC) Vs VDD with adjusted VBB for at-speed operation	24
Fig 3.4: Total Power Vs VDD for various workloads at constant speed for MAC	24
Fig 3.5: VDD for Minimal Power operation Vs Workload for MAC.....	27
Fig 3.6: Dynamic, Static and Total Power (MAC) for different power saving methods...	28
Fig 3.7: Total Power Vs VDD for different workloads-Ring oscillator	31
Fig 3.8: Dynamic, Static and Total Power for Ring Oscillator.....	31
Fig 3.9: Total Power (Adder) Vs VDD with adjusted VBB for at-speed operation	32
Fig 5.1: Block diagram of VDD/VBB Control using DC/DC Conversion	45
Fig 5.2: Block diagram of VDD/VBB Control using local quantized supply voltages.....	46
Fig 5.3: Stdcell cell with substrate contacts shorted to the local power supplies.....	48
Fig 5.4: Stdcell cell with substrate contacts brought out	49
Fig 5.5: Layout of a 16 bit adder in 65 nm with substrate contacts shorted to the local power supplies	50
Fig 5.6: Layout of a 16 bit adder in 65 nm with substrate contacts brought out.....	50

Chapter 1

Introduction

1.1 Introduction

Minimizing operating power is a necessity for all the battery operated electronic devices as well as high performance computing systems. One of the traditionally followed methods for power reduction is dynamic voltage scaling (DVS) [4]. In this technique, the supply voltage is reduced to be just enough to meet the target performance specifications, thus minimizing the dynamic power consumption. DVS is also employed together with Dynamic frequency scaling (DFS) to reduce the total switching power for a range of operating speeds [2].

DVS has mostly been implemented by using programmable DC-DC converters for generating the supply voltage. The output voltage of these converters could be either fixed by programming or by continuous adaptive closed loop control. Local voltage dithering has been proposed as an alternative to the varying DC-DC regulator [3]. Here the system can operate from one of two different supply values. The circuits are connected between the pairs as required to guarantee a certain throughput while minimizing dynamic power.

Even though DVS as a technique is very powerful in reducing the power consumption, it is still sub-optimal since it reduces the dynamic power quadratically, but the leakage power only linearly. This is because the leakage current is only a weak function of the supply voltage beyond a few 100mVs of supply. DVS achieves close to optimal power in technology nodes greater than 0.18 μ m since the dynamic power is the main source of power loss in those technologies. However in the generic and high-speed flavors of 0.13 μ m and smaller technologies, leakage power constitutes a major component of the total power loss [5]. Hence an optimal power reduction technique needs to also address the problem of leakage power.

Leakage reduction during stand by mode of operation is well understood. In this mode, no switching activity takes place and hence the logic block needs to be put into the lowest leakage mode. Techniques like power supply gating [6], off-off stacking [7] etc., have been developed to achieve low stand by power consumption. However leakage power during active mode is also a significant concern. It has been theoretically shown that a good heuristic to achieve close to minimum power operation is to have the leakage power to be about half the dynamic power [8] [3]. While dynamic power can be modulated via adjusting the supply voltage, the leakage power can be modulated via body bias. In fact adaptive body bias has been demonstrated by [2], but mainly to achieve design centering, that is, to slow down and reduce leakage of chips in the fast, leaky process corners.

Motivated by the results of [8][3], Nomura, et. al. in [1] have proposed a supply and body bias adjustment technique to balance leakage and dynamic power in a closed loop system. They have replica circuits to emulate dynamic and leakage power consumption of the actual circuit. Measurements from the replica circuits are then used to adjust the supply voltage and body bias such that the leakage power of the replica is kept to be about half the dynamic power of the replica.

However there are two drawbacks in this approach. One is that the control loops rely on measurements from the replica circuits, and hence are limited by the mismatch of the replica to the actual circuit, especially with varying workloads and activity factors. The second limitation is that the control loop tries to maintain a fixed ratio of the leakage to dynamic power regardless of the activity factor. In fact the optimum ratio to minimize total power does depend on the activity factor. In [3], the authors propose to make use of the local workload variations to achieve total minimum dynamic power. However, their approach considers only the dynamic component. They don't employ DVS and ABB simultaneously to minimize total operating power

To the best of our knowledge, there have not been any studies reported on the effect of the workload on the minimal power point under simultaneous application of DVS and ABB. This thesis quantifies the power saving by using the DVS and ABB techniques simultaneously on typical data path elements like adders and multiply-accumulate (MAC) units under different

workloads in different operating conditions. Adders and MACs are critical components in Digital Signal Processors. We evaluate the relative power savings between DVS only and DVS + ABB scheme at various process and temperature corners of operation for different workloads. Based on this study, we propose a workload indexed lookup table to provide the optimum supply and body bias values to minimize the total power consumption of the circuit under different operating conditions. A lookup table based DVS scheme is already used in many chips [9]. We also discuss an algorithm to generate the look up table efficiently.

The rest of thesis is organized as follows. Chapter 2 discusses the problem formulation. This includes a detailed analysis of the various power components and their dependencies on the activity. It also discusses the constant frequency VDD-VBB relationship and the variation of the power components for constant operating frequency.

Chapter 3 discusses the test circuits used for this study while. The detailed description of the key results and observations from the application of these techniques to the test circuits are also part of the chapter. The leakage, dynamic and the total power components are quantified for the test circuits and the results are summarized at the end of the chapter.

Chapter 4 explains our proposal for a look up table for supply and body bias along with. Chapter 5 describes the implementation details and a simulation based approach to generate the lookup table for any block.

Chapter 6 presents the conclusions drawn from the work covered in this thesis.

References

- [1] Masahiro Nomura, Yoshifumi Ikenaga, Koichi Takeda, Yoetsu Nakazawa, Yoshiharu Aimoto, and Yasuhiko Hagihara, "Delay and Power Monitoring Schemes for Minimizing Power Consumption by Means of Supply and Threshold Voltage Control in Active and Standby Modes" in IEEE Journal of Solid State Circuits, April 2006, pp. 805-814.
- [2] Masakatsu Nakai, Satoshi Akui, Katsunori Seno, Tetsumasa Meguro, Takahiro Seki, Tetsuo Kondo, Akihiko Hashiguchi, Hirokazu Kawahara, Kazuo Kumano, and Masayuki Shimura, "Dynamic Voltage and Frequency Management for a Low-Power Embedded Microprocessor" in IEE Journal of Solid-State Circuits, January 2005, pp. 28-35.
- [3] Benton H. Calhoun and Anantha P. Chandrakasan, "Ultra-Dynamic Voltage Scaling (UDVS) Using Sub-Threshold Operation and Local Voltage Dithering" in IEEE JSSC, vol. 41, no. 1, January 2006.
- [4] Burd T.D., Pering, T.A., Stratakos, A.J., Brodersen, R.W., "A dynamic voltage scaled microprocessor system", IEEE Journal of Solid State Circuits, vol. 35, no. 11, pp.1571-1580, November 2000.
- [5] ITRS Roadmap at <http://www.itrs.net/>
- [6] Mutoh S., Douseki T., Matsuya Y., Aoki T., Shigematsu S., Yamada J., "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS", IEEE Journal of Solid State Circuits, vol. 30, no. 8, pp. 847-854, August 1995.

- [7] Harter J. P., Najm F., "A gate-level leakage power reduction method for ultra-low-power CMOS circuits", CICC 1997, pp. 475-478, October 1997.
- [8] Nose K., Sakurai T., "Optimization of VDD and VTH for low-power and high-speed applications", Proceedings of the Asia and South Pacific Design Automation Conference, 2000, pp. 469-474.
- [9] Intel White Paper on "Enhanced Intel Speed Step Technology for the Intel Pentium M Processor",
<ftp://download.intel.com/design/network/papers/30117401.pdf>

Chapter 2

Motivation

This chapter briefly describes the current state of the art in the related fields. It also explains the motivation for research. Analysis is done on the dynamic and leakage power components on an Integrated circuit. Analysis is also done on how workload influences the power components. Sample algorithms are run on a Digital Signal Processor (DSP) and the workload variations for various algorithms are studied on different units of the processor. Fixed frequency operating points with different supply and body bias are studied and the variation of the dynamic and leakage power components at these different operating points are also examined.

2.1 Power Components and Activity

Total Power consumed by any integrated circuit has a static as well as a dynamic component. The static or the leakage component is independent of the circuit activity. It just depends on the supply voltage and the threshold voltage of the underlying transistors. The dynamic component depends only on the supply voltage, the operating frequency and is modulated by the circuit activity. The total power consumed by the circuit can be expressed as

$$P_T = P_L + A_F P_D \quad (1)$$

where P_T is the total power, P_L is the leakage power and A_F is the activity factor. P_D is the maximum dynamic power which corresponds to the highest activity. Note that P_D is also proportional to the total switching capacitance, square of the supply voltage and the operating frequency. In this paper, we

will focus on fixed frequency circuits – which are common to many digital signal processors.

The activity factor captures the effective switching capacitance and is related to the ratio of the toggling node capacitances to the total switching capacitance in the circuit. Since the dynamic component is modulated by the activity factor, the contribution of dynamic component in the total power varies with activity. The dynamic component is zero when the circuit is idle (when there is no activity) and it increases with the circuit activity.

Since the number of nodes in the circuit and the number of toggling nodes are usually quite huge, an accurate, or even an approximate estimation of the activity is usually quite involved. Switching capacitances are a function of circuit sizes and wire loads, and hence an accurate estimate can only be obtained in conjunction with placed and routed net lists of the circuit blocks. However activity can be approximated to a certain extent if the workloads of the different blocks of the circuit are known. By workload, we mean the ratio of the number of cycles for which the block (unit) is active to the total number of cycles. If the chip (integrated circuit) runs for N cycles and one of the blocks, say B is active for M out of the total N cycles, the workload for the block B is M/N . Once we map an algorithm to a given hardware, workloads of various blocks of the hardware can be easily estimated by profiling the code, and this doesn't need the detailed placed and routed netlist of the circuit block. Thus we will use the workload as a proxy for the activity of a block.

2.2 Workload Variation

Many general purpose processors run different signal processing algorithms in the same hardware. Workloads for various hardware blocks of the processor differ from algorithm to algorithm. Some algorithms use certain hardware units almost every cycle while using some other hardware units of the same processor sparingly. Or in short, workload varies from unit to unit over time and from time to time for the same unit. Table 2.1 shows the workloads for different data path units in the Analog Devices Sharc DSP [1] for various signal processing algorithms. We can observe a large variation in the work load of active blocks based on the algorithm being run. The workloads vary from less than 1% to about 70% depending upon the algorithm and the block.

Table 2.1: *Workloads for different Data Path Units in Sharc DSP for various signal processing algorithms*

Block	Algorithm				
	FFT	FIR	FM Modulation	Phaser Effect	Mono-Single Coefficients
MAC	0.3042	0.6840	0.0070	0.1098	0.1617
Floating Point ALU	0.3588	0.6733	0.0000	0.0000	0.1617
Fixed Point ALU	0.0089	0.0003	0.0390	0.0281	0.0420

Shutting down the unused blocks might not be a viable option in many cases especially when the blocks need to be brought up once every few

cycles. Also the latency to bring up and shut down the units might not be acceptable in many applications.

DVS using Local voltage dithering (or using local quantized power supplies) to achieve maximum energy savings for varying workload is discussed in [2]. This is particularly useful for application specific processors whose frequency can be varied depending upon workload. But this cannot be applied to a fixed frequency system.

2.3 Fixed Frequency operating points

We propose a scheme which is particularly useful for fixed frequency operation. For a class of applications that cannot trade off operating frequency for power, the minimum supply voltage has a lower bound. This lower bound varies with the applied body bias. Similarly, for a given frequency, there is a maximum value of the body bias beyond which the critical path would fail. This maximum body bias varies with the applied supply voltage. The constant operating frequency criterion can be met with various combinations of supply voltage (V_{DD}) and body bias (V_{PB}/V_{NB}) values.

Fig 2.1 shows a schematic representation of our system with two controllable voltage sources for supply and body bias. In this research, we consider only reverse body bias to either of the NMOS or PMOS transistors. This allows one to increase the threshold voltage as the body bias voltage is increased. It is also possible to forward bias the body to a small extent

without turning on the body diodes [3]. We don't consider this case here, though our method can be easily extended to cover this case too. Higher forward body bias (FBB) can lead to enhanced leakage due to increased diode conduction. In the next chapter we discuss how only reverse biasing one of PMOS or NMOS body suffices.

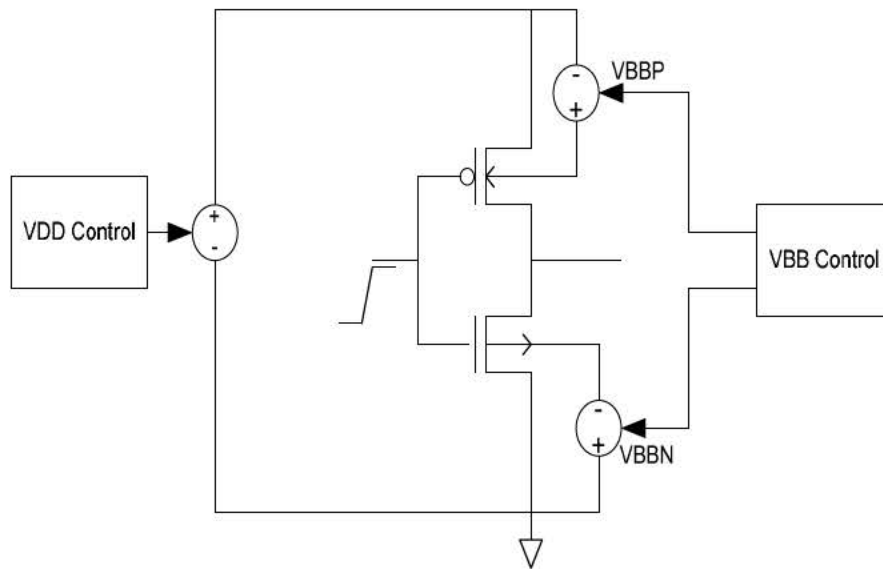


Fig 2.1: *Schematic representation of our VDD-VBB control System*

Fig 2.2 shows the VDD Vs VBB curves for a 31 stage ring oscillator operating at a constant frequency in a 90nm process node[4]. Any point in the plot will result in a constant fixed operating frequency.

As the body bias increases, the supply voltage also needs to increase to achieve the constant fixed frequency. The slowing down of the transistors by the increase in VBB (due to a higher threshold voltage) is compensated by the increase in speed due to the increase in VDD. The intent is to operate the block at a fixed frequency. For a higher supply voltage, a higher body

bias would guarantee that we provide the voltages just barely enough to meet the frequency criterion. The Y-axis range or the maximum permissible VBB value limits the operating range. This maximum possible VBB value is usually process specific. As is evident from the plot, VDD has a much bigger influence on the speed compared to VBB.

Fig 2.3 shows the power plots of the ring oscillator for fixed frequency. The X-axis is the supply voltage and the Y-axis the two power components. For each VDD, there is a corresponding VBB voltage which can be obtained from Fig 2.2.

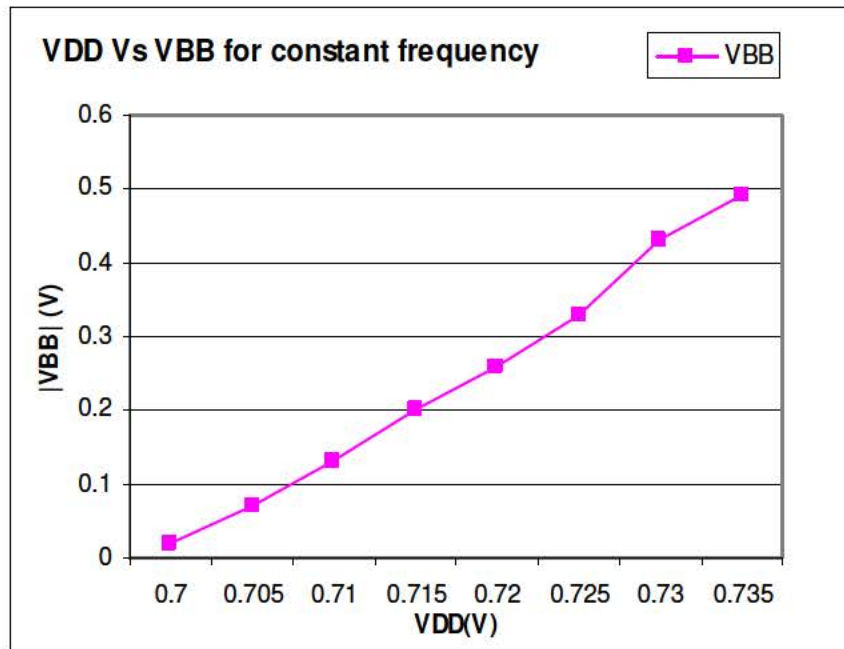


Fig 2.2: VDD Vs VBB for a ring oscillator working at a fixed frequency 500 MHz

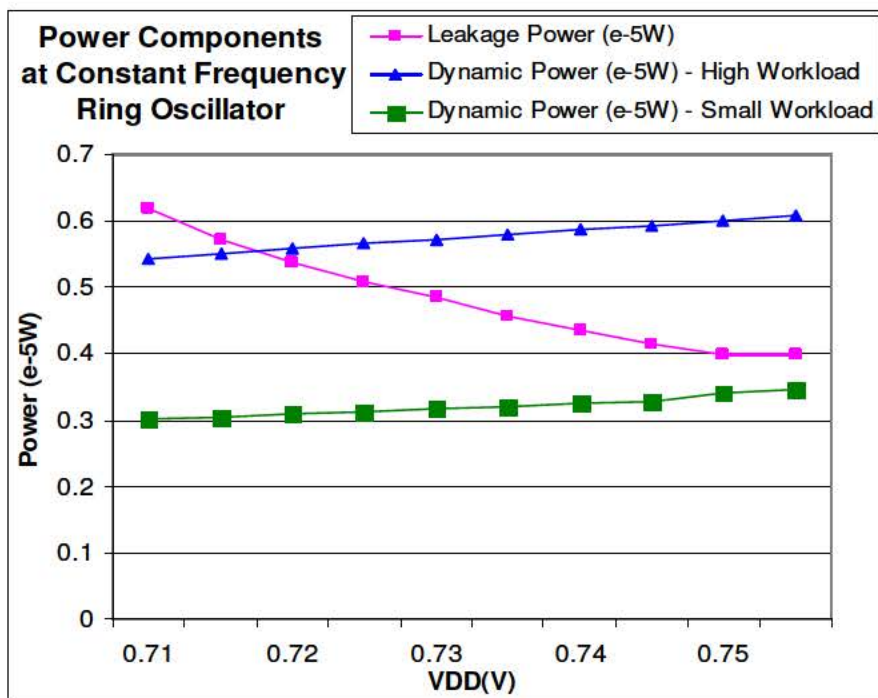


Fig 2.3: Leakage and Dynamic Power components for a ring oscillator operating at a fixed frequency 500MHz

The power plots are for two different workloads of 10% and 60%. The workload here signifies the proportion of the total time for which the ring oscillator is on. If the supply is on for “x” seconds and the oscillator is gated **on** for “y” seconds, the workload would be y/x. We control the workload by gating the oscillator **on** using an “AND” gate with an enable.

The leakage power remains same irrespective of the workload. The dynamic component increases with workload. Hence total power also increases with workload. The dynamic component is minimal at the lowest VDD point which also corresponds to the DVS point. But the leakage is high here as the VBB is zero. As VDD increases, the dynamic component also increases but now the same frequency operation is possible at a higher body bias value thereby giving a lower leakage power. So from left to right,

dynamic power increases and leakage power decreases. The total power may increase or decrease depending on the relative contribution of the dynamic and static components to the total power.

For a low workload, the leakage would be the major contributor to the total power. Hence it would be more power efficient to choose an operating point towards the right side of Fig 2.2 and Fig 2.3 where leakage is less and dynamic component is more. For a high workload, the dynamic component becomes dominant and hence reducing the dynamic component at the expense of leakage could be the better option. Choosing an operating point towards the left of the Fig 2.2 and Fig 2.3 would reduce the dynamic power but results in higher leakage.

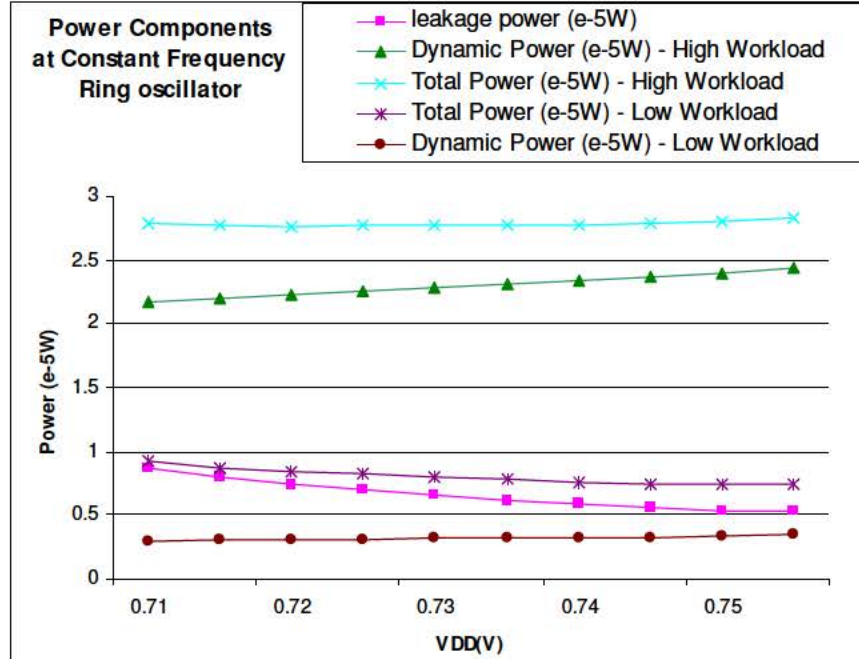


Fig 2.4: Leakage, Dynamic and Total Power components for a ring oscillator at fixed frequency

Fig 2.4 shows the total power as well in addition to the individual power components. Here the workloads are chosen different enough to highlight the effect of workload on the operating point. For the higher workload case, the best operating point is towards the left (at a lower VDD and lower VBB) but for a lower workload, the optimal operating point is towards the right of the graph (at a higher VDD and higher VBB).

For a processor which runs a fixed function algorithm, the workloads of each block would be different. Also the same block might have a different workload for a different algorithm. There is a certain amount of power savings by re-adjusting the VDD and VBB values with changing workload. This is extremely useful especially in processors which run different fixed function algorithms.

In our scheme, both the supply and body bias can vary from block to block in the same processor. All the blocks would work at the same frequency thereby providing the same data latency. But each block works with different combinations of VDD and VBB depending on the workload of the respective block.

The choice of VDD and VBB for each local block can be done in two ways. If the workload of the block for the algorithm remains almost same throughout the application, the VDD and VBB values corresponding to the workload are chosen from the pre-characterized table. This would fix the VDD and VBB for a particular block for the application. Or if the workload

for the same block varies from time to time in the application, the VDD and VBB of the block can be adjusted continuously from the table information. Due to the localized switching, the switching capacitance is small. Hence fast switching times can be achieved.

We haven't explicitly considered the impact of reverse body biasing (RBB) in deterioration of short channel behavior, but this effect is implicitly taken into account in our device models.

2.4 Summary

In short, for a fixed frequency operation, there exists a multitude of operating points (combinations of VDD and VBB). Each results in a different dynamic and leakage power. The total power would also vary depending on the relative contributions of the dynamic and the leakage components. This relative contribution is dependent on the workload of the block. Hence for a fixed operating frequency, the workload determines the optimal VDD-VBB point. For each workload, a certain optimal VDD-VBB point gives the minimal total power. This VDD-VBB optimal point for each block can be estimated and pre-characterized.

References

- [1] Analog Devices Sharc DSP Reference Manual,
<http://www.analog.com/en/embedded-processing-dsp/sharc/processors/manuals/resources/index.html>
- [2] Benton H. Calhoun and Anantha P. Chandrakasan, "Ultra-Dynamic Voltage Scaling (UDVS) Using Sub-Threshold Operation and Local Voltage Dithering" in IEEE JSSC, vol. 41, no. 1, January 2006.
- [3] H. Ananthan, C. H. Kim, K. Roy, Larger-than-Vdd forward body bias in sub-0.5V nanoscale CMOS", IEEE International Symposium on Low-Power Electronics and Design, pp. 8-13, August 2004
- [4] Bhushan M., Gattiker A., Ketchen M. B., Das K.K., "Ring oscillators for CMOS process tuning and variability control", IEEE Trans. On Semiconductor Manufacturing, vol. 19, no. 1, pp. 10-18, February 2006.

Chapter 3

Test Circuits and Key Results

This chapter presents the test circuits used in our research. The detailed description of the key results and observations from the application of these techniques to the test circuits are also part of the chapter. The leakage, dynamic and the total power components are quantified for the test circuits and the results are summarized at the end of the chapter. We compare the P and N body bias schemes individually and the results are quantified. The variation of VDD and VBB with workload for minimal power operation is also presented.

3.1 Introduction

The test circuits used in our research are data path elements from DSP Cores. We have chosen these data path elements for sample simulations since these elements constitute about 60% of the total logic in DSP cores. Some units are from an actual DSP core [1]

The experiments employ standard IC implementation sign-off methods for design and analysis. The circuits are laid out and the extracted parasitics are used in simulation along with the netlist so that the dynamic and leakage power components are as close to the actual silicon values as possible. Synopsys HSIM and synopsys hspice circuit simulators are used for simulations.

3.2 Test Circuits

The test circuits considered for this work are

- a) A 16-bit Multiply Accumulate unit (MAC) implemented in 90nm CMOS process.
- b) A 31 stage Ring Oscillator implemented in 90nm CMOS process.
- c) A 16-bit adder implemented in 90nm CMOS process.
- d) A 16-bit adder implemented in 65nm CMOS process.

The adder and MAC units are from a real DSP Core. The MAC is implemented as a Carry Save Wallace Tree array. The 16-bit CMOS adder has a ripple carry architecture which employs carry look ahead at a 4-bit modularity. Vectors are applied at the input of the circuits and the dynamic and leakage power monitored. Since leakage varies with input vector [2], leakage power is monitored for various sets of input vectors and the average is used in the leakage tables. The measurements are made for all the supported process and temperature conditions.

3.3 Key Observations

3.3.1 Comparison of P and N Body Bias Schemes

We could choose either all the P transistors or all the N transistors or all the P and N transistors in the circuit for body biasing. Fig 3.1 shows the total power Vs VDD curves at constant frequency for a ring oscillator working at 500 MHz. The P and N transistor sizes are adjusted in such a way that the rise and fall delays of the inverters in the oscillator are equalized. The P

transistor body bias curves and the N transistor body bias curves are shown separately. ABB on P or N transistors give almost the same power reduction. Fig 3.2 shows the total power Vs VDD curves for a ring oscillator in which the P and the N transistors are sized in such a way that rise and fall delays are not same. The P transistor has more delay than the N transistor or in other words, the N transistor is relatively stronger here. Hence the N transistor leaks more. Therefore we get more power savings by back biasing the N transistor.

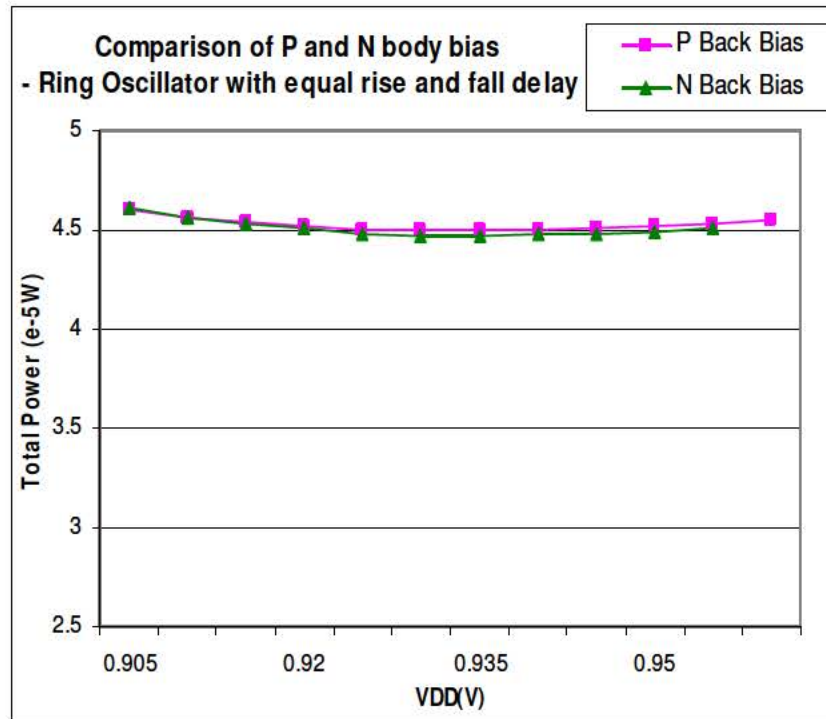


Fig 3.1: Total Power (Ring Oscillator) Vs VDD with adjusted VBB for at-speed operation. Comparison of P and N Body Bias schemes. Equal rise and fall delay

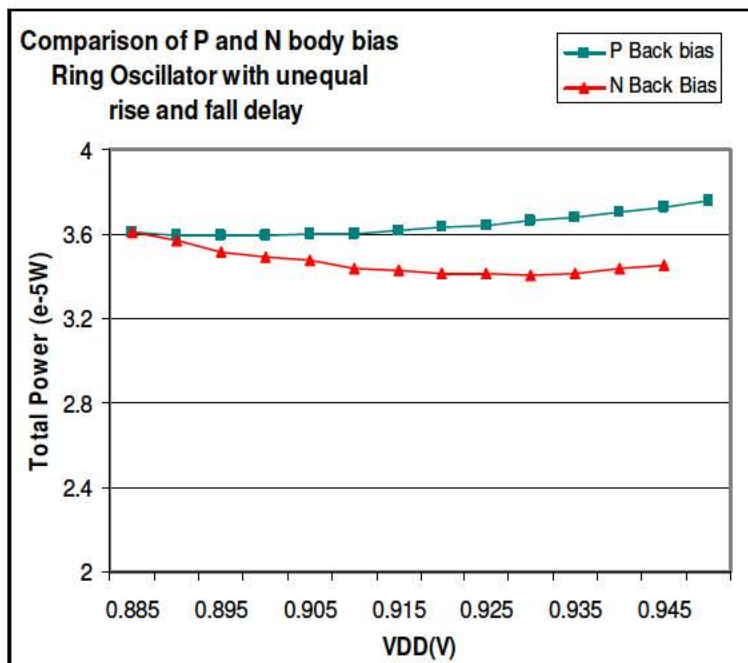


Fig 3.2: Total Power (Ring Oscillator) Vs VDD with adjusted VBB for at-speed operation. Comparison of P and N Body Bias. Unequal rise and fall delay.

We experimented with body biasing both P and N transistors. This gives a slight improvement in power savings compared to the either P or N back bias scheme. But this has the additional area overhead of an extra power route to back bias the substrates of the N as well as the P transistors. But body biasing both P and N transistors has one specific advantage that there is lesser chance of the maximum allowed limit of body voltage being reached here since we need to apply only half the body bias for the same speed reduction. This is assuming both P and N transistors contribute equally to speed of the circuit. Hence this is especially useful for low workload circuits where the total power reduction is always limited by the inability to apply a high body bias due to the circuit reliability constraints. Since we obtain more power reduction by using both P and N body bias, all

the analysis done for a single body bias is valid for the double body bias also.

In short, if the rise and fall delays of the circuit are same, there is not much difference in power savings if we control the body bias of the P or the N transistors. Hence in this work, we only control the body bias of the N transistors. However this does require the process to be a triple well process. In the case of a double well process, only the PMOS well biases can be adjusted as all the NMOS transistors will share the same body.

3.3.2 MAC

The MAC is implemented as a Carry Save Wallace Tree array in 90nm CMOS process. Fig 3.3 shows the plots of the average power (leakage plus dynamic) consumed by the MAC unit for a given fixed frequency, f of operation. The x-axis is the supply voltage (VDD). VDD is increased gradually from the DVS Voltage (V_{DVS}) which is the minimum voltage below which the critical path fails for a zero body bias. For each VDD value, the body bias is increased until when the circuit slows down to such an extent that the MAC unit can no longer run at the frequency f or the maximum allowed body voltage limit is reached.

In the plot, x-axis is the supply voltage and y-axis is the average total power drawn from the supply at the maximum back bias point. The maximum body bias is shown on the right side y-axis. The experiments are done by

independently varying P transistor body bias and N transistor body bias. Fig 3.3 shows the results when the N transistor body is back biased. This is repeated for various values of the block workload in Fig 4(b).

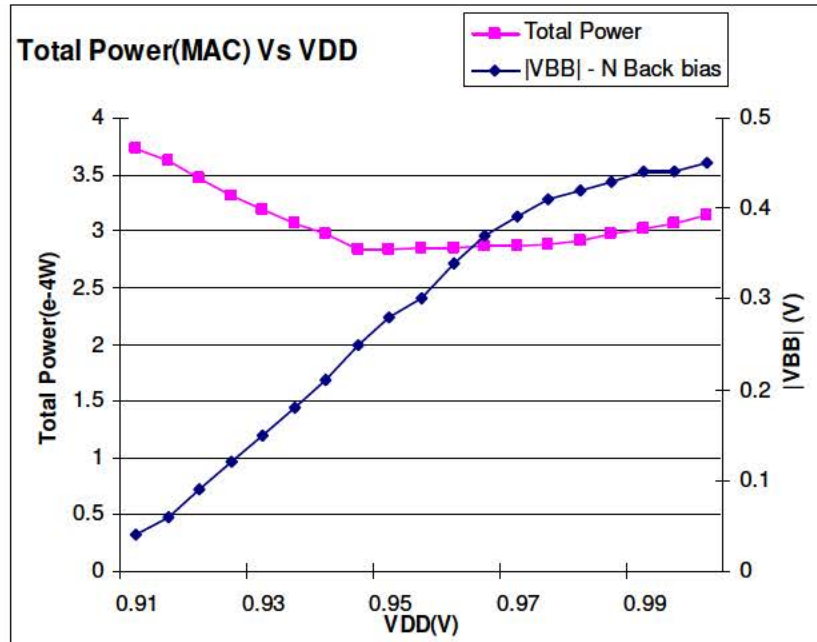


Fig 3.3: Total Power (MAC) Vs VDD with adjusted VBB for at-speed operation

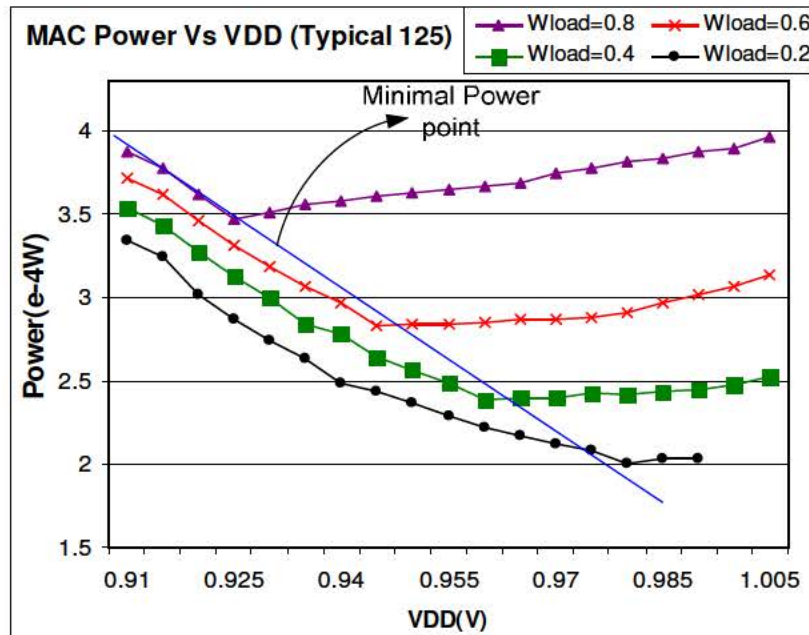


Fig 3.4: Total Power Vs VDD for various workloads at constant speed for MAC

From the figure, with the constant frequency constraint, the minimum power point doesn't coincide with the minimum voltage point (DVS point). In other words, for fixed frequency operation, DVS does not necessarily guarantee minimal power operation of the circuit. For voltages higher than the minimum possible voltage, the body back bias voltage at the current measurement point increases with VDD.

The dynamic component is minimal at the V_{DVS} point [3]. But the leakage is high there as the body bias is zero. Leakage is minimal at the highest VBB point. This highest VBB is limited by reliability considerations of the circuit. The total power is minimal at an intermediate point which corresponds to the trough of the power curve in Fig 3.3.

Fig 3.4 shows the power consumption for various workloads of the MAC circuit. The top most curve is for the highest workload. And the lower curves are for progressively lower workloads. As is obvious, when the workload increases, the curves shift up. i.e. The total power increases since the circuit is active for more time.

Another inference is the shift in the trough (the minimum total power point) to the left with increasing workload. For a small workload, the leakage power dominates the total power. Hence more power saving is obtained by applying higher VBB than by reducing the VDD. So the trough of the curve is more towards the higher VBB values (or towards the right). When workload increases, the dynamic power becomes more dominant

and hence more power saving is obtained by reducing VDD than by applying a higher VBB. So the trough shifts to the left. For still higher workloads, the trough of the curve continues to shift to the left until it reaches a point where the trough completely vanishes. At this high workload point, the DVS point itself becomes the minimal total power point.

Fig 3.5 shows the shift in the VDD for minimal power operating point with increasing workload. For the small workload = 0.1, the least total power is achieved by operating the circuit at 0.98V VDD and -0.48V VBB (N transistor back bias). This VBB is close to the highest possible absolute value of the substrate back bias that can be applied. This upper limit of the substrate bias arises from the circuit reliability constraints. For higher workloads, the least total power would be achieved at lower VDD (and hence lower absolute VBB) values since the total power is dominated by the dynamic component. For a workload = 0.7, the least total power operating point is VDD=0.92V and VBB=0V. This corresponds to the DVS Voltage operating point. This implies that the dynamic power is definitely the major component for workloads greater than 0.7. Any gain in leakage by increasing VBB cannot compensate for the increase of dynamic power for workloads greater than 0.7. For any workload higher than 0.7, it would be most power optimal to operate at this operating point itself. The VDD cannot be reduced further down than the DVS voltage.

So the power savings are limited at one end by the DVS voltage (below which the critical path fails to function). The body bias voltage at this end is zero. At the other end, the limiting factor is the maximum substrate bias that the process technology permits. The supply voltage at this end is the voltage which just makes the critical path work at this maximum body bias. Usually, the minimal power operating points for large ranges of operating workloads fall between these limits.

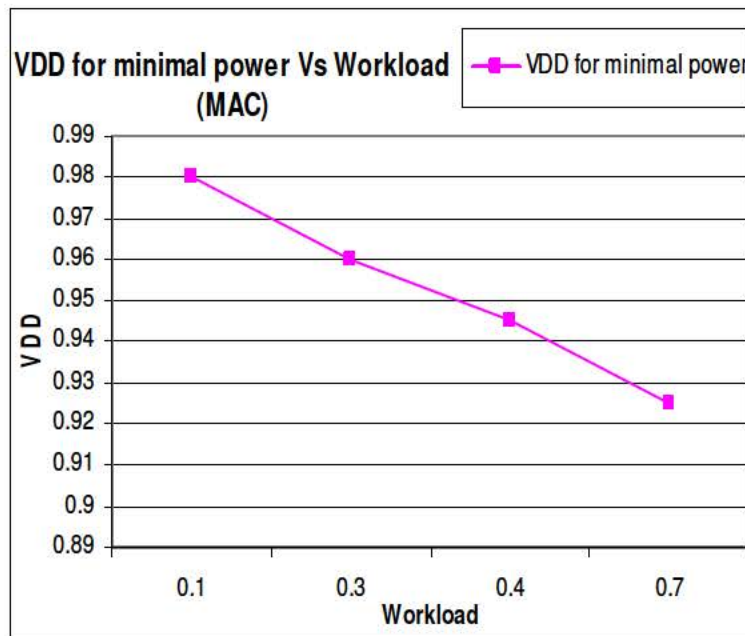


Fig 3.5: *VDD for Minimal Power operation Vs Workload for MAC*

Fig 3.6 shows the relative power savings by employing different methods. All the savings are compared to the “only DVS” scheme which has become almost the default standard for comparison. When DVS and ABB are employed without any workload indexing, the power consumption is about 73% of the “only DVS” point power consumption.

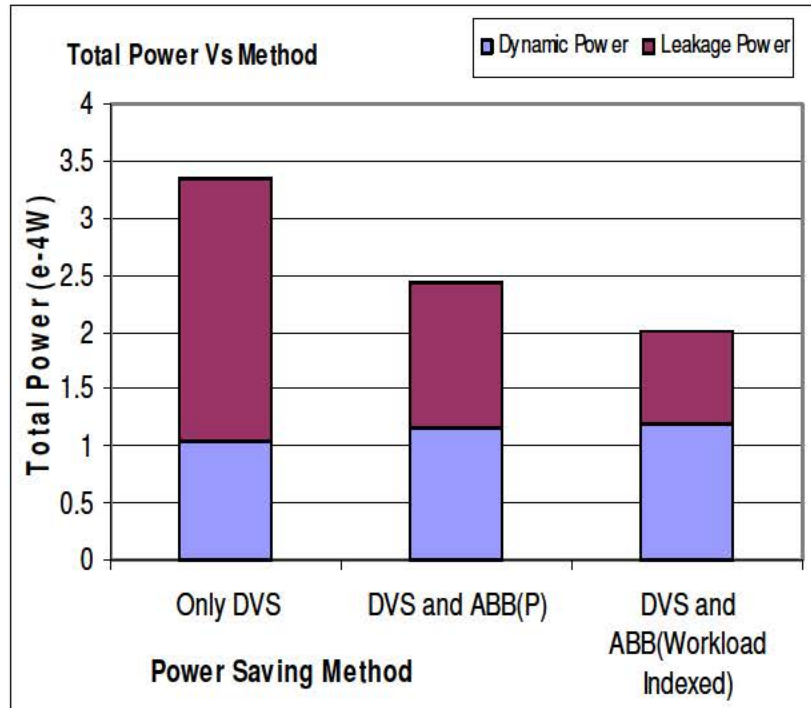


Fig 3.6: *Dynamic, Static and Total Power (MAC) for different power saving methods*

This 27% reduction compared to the “only DVS” case is attributed to the reduction in leakage with the use of ABB. As can be seen from the figure, the dynamic power is higher in the DVS-ABB scheme compared to the “only DVS” scheme. But this small increase in dynamic power is more than compensated by the reduction in leakage achieved by the body biasing.

When workload indexing is used, the power consumption reduces to about 59%. Or in other words about 14% savings is obtained in the total power when the workload is used as an index to select the VDD and VBB values. In the figure, the leakage component is less and the dynamic component is more in the workload indexed scheme as well compared to the simple “DVS-ABB” scheme. But this could as well end up the other way round in

some cases. This just depends on the initial workload used to arrive at the VDD-VBB operating point in the simple “DVS-ABB” scheme.

In Fig 3.6, the estimated workload was higher and the actual workload was lower. Hence when the VDD and VBB was adjusted to correspond to the actual lower workload, the VDD value increased and the absolute VBB value also increased. Hence the leakage power reduced and the dynamic power increased. But if the estimated workload was lower than the actual workload, a lower VDD and a lower VBB would have given a smaller total power than the estimation. In such a situation, the dynamic power would be less and the leakage power would be more in the workload indexed VDD-VBB scheme. The 14% power savings are obtained at a workload of 0.2. A workload of 0.5 is used to estimate the power for the non workload indexed scenario. The savings would also vary depending upon the difference between the workload used to arrive at the optimal VDD-VBB point and the actual workload of the circuit.

3.3.3. Ring Oscillator

The Ring Oscillator has 31 stages and is implemented in 90nm CMOS process. Fig 3.7 shows plots of the average power drawn by the Ring Oscillator circuit. Similar to the MAC, for each of the supply voltage values, the body bias is varied (increased for P and decreased for N) until the point where the oscillating frequency falls below the DVS frequency. This is repeated for a range of VDD values for which the body bias falls in

acceptable range from a circuit reliability point of view. The x axis is the supply voltage and the y-axis shows the average power drawn from the supply at the maximum body back-bias point. The different curves are for different workloads. The workload and activity for a ring oscillator are exactly same. The workload here signifies the proportion of time for which the ring oscillator is on. If the supply is on for “x” ns and the oscillator is gated on for “y” ns, the workload would be y/x . The lowest is curve for the least workload and the top curve for the highest workload.

Similar to the MAC, there is a shift of the trough of the curve to the right for lower workloads. This is due to leakage dominating the total power for lower workloads. Hence a higher VBB and higher VDD would dissipate less total power than a lower VDD and a less VBB. Fig 3.8 quantifies the power savings obtained by employing both DVS and ABB together.

In the “worst-leakage/highest speed” process corner, the ratio of the minimum power (power at the trough point) to the power at the DVS point (minimum VDD point) is lesser than the corresponding ratio at the “best-leakage/slowest speed” and “nominal leakage/nominal speed” process corners. Also this ratio is the least at the highest temperature. This indicates that the power savings obtained by employing DVS and ABB together is more for faster devices and at higher operating temperature due to the dominance of the leakage power in these operating conditions.

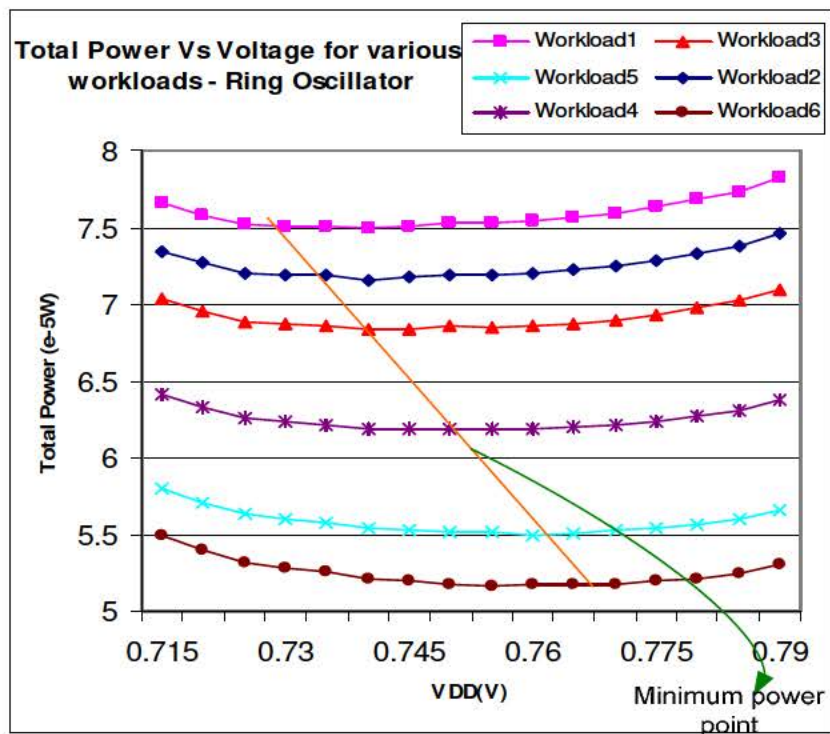


Fig 3.7: Total Power Vs VDD for different workloads-Ring oscillator

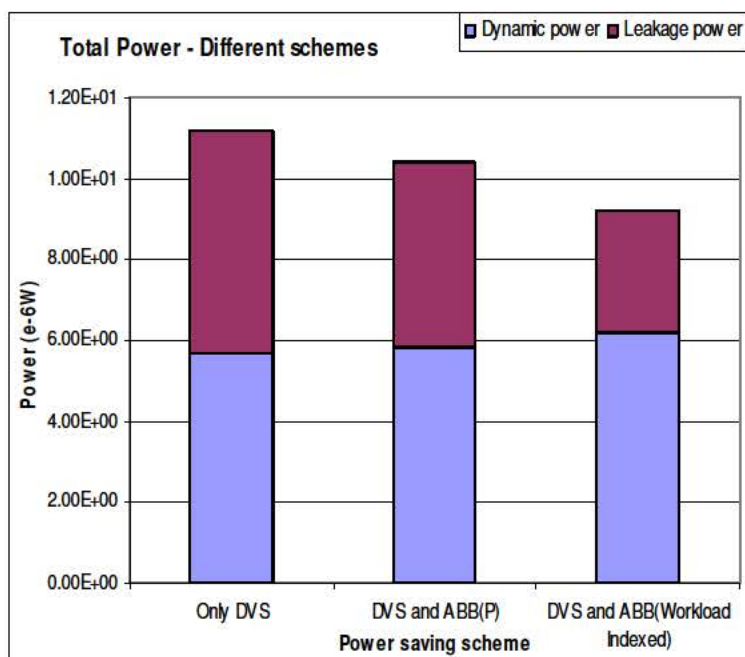


Fig 3.8: Dynamic, Static and Total Power for Ring Oscillator

3.3.4 Adder

A 16 bit adder is implemented in both 90nm and 65 nm. The adder has a ripple carry architecture with carry look ahead implemented at 4 bit modularity. Vectors are applied at the input and both static and dynamic powers are measured for various input vectors.

Fig 3.9 is the constant frequency power curve for the adder. The two curves are the total power and the VBB values for the constant operating frequency. The trough corresponds to the minimal power operating point for the given workload. The highest VDD is limited by the largest possible substrate bias value which is 0.5V for here.

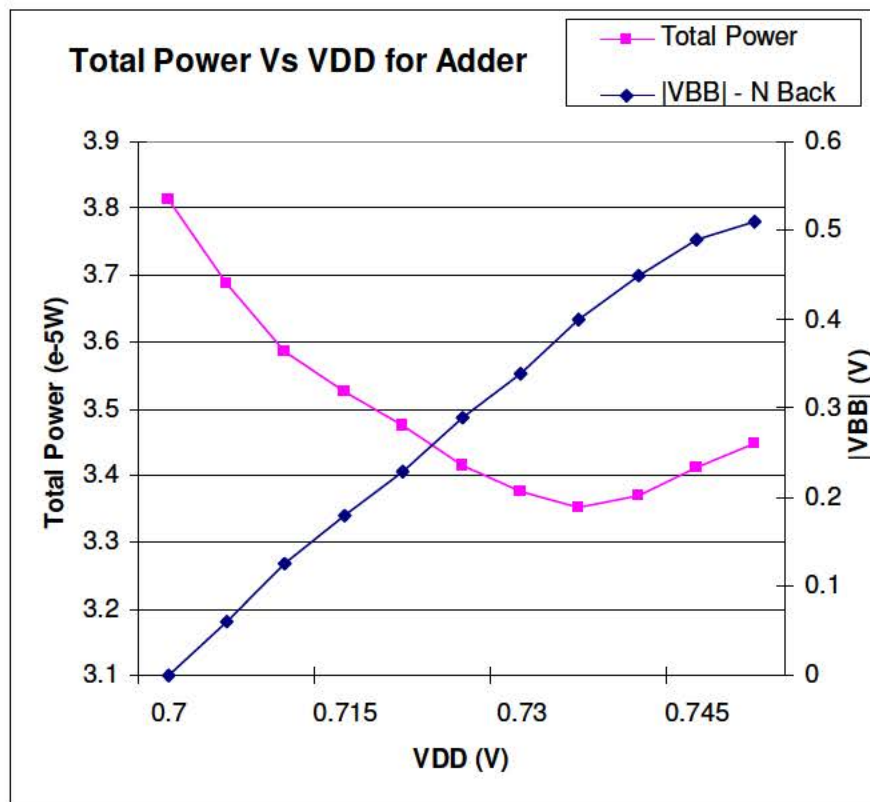


Fig 3.9: Total Power (Adder) Vs VDD with adjusted VBB for at-speed operation

This experiment is repeated for different workloads. The results show a shift in the minimum power point to lower VDDs for higher workloads similar to the MAC unit and the ring oscillator.

3.4 Summary of Results

The results for all the three blocks are summarized in Table 3.1. The table quantifies the power savings for the blocks in the leaky process corners where the power savings by threshold adjustments are maximal. The MAC and the Ring Oscillator are tested in 90nm process while the adder is done both in 90nm and 65 nm process nodes. For the savings without workload indexing, we have chosen a nominal workload of 50% and have used the corresponding VDD and VBB values. For the savings with workload, we have chosen the most optimal VDD and VBB corresponding to the actual workload. The results show more power savings at the “worst leakage/highest speed process corner.

Table 3.1: *Power Comparisons when different power saving methods are employed*

MAC 90nm 125C	Total Power FF (e-3W)	FF (%)	Total Power TT (e-4W)	TT (%)
Only DVS	1.19	100	3.34	100
DVS and ABB, no WL*	0.87	73.1	2.36	70.65
DVS and ABB, WL* indexed	0.692	58.15	2.01	60.17

ADDER 90nm 125C	Total Power FF (e-4W)	FF (%)	Total Power TT (e-4W)	TT (%)
Only DVS	8.62	100	4.02	100
DVS and ABB, no WL*	6.92	80.2	3.60	89.55
DVS and ABB, WL* indexed	5.83	67.63	3.40	84.57

ADDER 65nm 125C	Total Power FF (e-5W)	FF (%)	Total Power TT (e-5W)	TT (%)	Total Power SS (e-6W)	SS (%)
Only DVS	3.903	100	1.46	100	9.32	100
DVS and ABB, no WL*	3.44	88.1	1.35	92.4	8.89	95.3
DVS and ABB, WL* indexed	2.97	76.1	1.22	82.6	8.20	87.9

ADDER 65nm 25C	Total Power - Fast (e-6W)	Fast Process (%)	Total Power - Typical (e-6W)	Typi cal (%)	Total Power Slow (e-6W)	Slow Process (%)
Only DVS	4.117	100	1.39	100	1.08	100
DVS and ABB, no WL*	3.741	90.8	1.34	96.4	1.05	97.2
DVS and ABB, WL* indexed	3.595	87.3	1.29	92.8	1.01	93.5

Ring Oscillator 90nm 125C	Total Power FF (e-5W)	FF (%)	Total Power TT (e-5W)	TT (%)
Only DVS	5.37	100	2.21	100
DVS and ABB, no WL*	4.72	87.9	2.11	95.4
DVS and ABB, WL *indexed	3.97	73.9	2.049	92.3

*WL stands for workload.

References

- [1] Analog Devices Sharc DSP Reference Manual,
<http://www.analog.com/en/embedded-processing-dsp/sharc/processors/manuals/resources/index.html>
- [2] Siva G Narendra, Anantha Chandrakasan, "Leakage in Nanometer CMOS Technologies", Springer 2006.
- [3] Benton H. Calhoun and Anantha P. Chandrakasan, "Ultra-Dynamic Voltage Scaling (UDVS) Using Sub-Threshold Operation and Local Voltage Dithering" in IEEE JSSC, vol. 41, no. 1, January 2006.

Chapter 4

Workload Indexed Look up Table for Blocks

A workload based lookup table for VDD/VBB control is one of the many methods which can be used for maximum power saving in a general purpose processor. The workload indexed table is an extension to the normal process/temperature based look up tables used by chips for dynamic voltage scaling. This chapter describes the lookup table and provides examples for the same. The lookup table generation and details regarding implementation are reserved for the following chapter.

4.1 Introduction

Based on the results from the previous sections, we can see that a supply and body bias control scheme which takes into account the actual work load, can result in an additional 12 to 30% in power saving, beyond what DVS can provide. One practical implementation of DVS is based on table lookup [1]. For different process/temperature corners, a table provides the correct voltage value to use for a desired target frequency. We propose to extend this approach to also include indexing by work load.

4.2 Workload Indexed Tables

As an example, the workload indexed look up table for a MAC and ring oscillator to arrive at the minimal power consumption point for various workloads is shown in Table 4.1 & Table 4.2 respectively. Here VDD is the supply voltage and VPB is the body bias of the P transistors in the design.

Table 4.1: *Workload Indexed Lookup table for MAC for Typical process corner*

Workload	VDD (V)	VNB(V)
0.8	0.925	-0.12
0.6	0.95	-0.25
0.4	0.965	-0.37
0.2	0.985	-0.43

Table 4.2: *Workload Indexed Lookup table for MAC for Fast process corner*

Workload	VDD (V)	VNB(V)
0.8	0.715	-0.02
0.6	0.745	-0.18
0.4	0.76	-0.27
0.2	0.79	-0.46

The above tables illustrate the workload indexing for a MAC unit. If “only DVS” is employed, there would be just one value of VDD for every temperature and process corner.

When ABB is also employed along with DVS, there would be one VDD and one VBB value for each process and temperature corner. We go one step ahead and provide a VDD and a VBB value for each workload (obviously for each process and temperature corner).

Table 4.3: *Workload Indexed Lookup table for Ring Oscillator*

Workload	VDD (V)	VBB on P transistor, VPB(V)
0.7	0.715	0.745
0.5	0.72	0.78
0.3	0.725	0.855
0.2	0.725	0.855
0.1	0.74	1
0.06	0.74	1
0.04	0.755	1.145
0.03	0.76	1.21
0.025	0.76	1.21
0.015	0.765	1.255
0.01	0.765	1.255
0.005	0.775	1.365

Table 4.3 is the workload indexed table for a ring oscillator. Here VDD is the supply voltage and VPB is the body bias voltage for PMOS transistors in the oscillator.

For each workload, there is a pair of values, one VDD and one VBB of supply and body bias voltage which will give the minimum total operating power. When the workload tends to 1 (100%), VDD would tends to the DVS voltage and VPB would be VDD itself. Similarly, when the workload tends to zero, the VPB would tend to the maximum body bias that can be applied and the

VDD would be the voltage that can make the critical path work at this body bias voltage. Note that the speed of the circuit is the same at all the operating points given in the above table. One lookup table could be provided for each process-temperature corners. The table entries are obtained from detailed circuit simulation of extracted layouts at that process-temperature operating point. Standard ring oscillators can be used to sense the process corner for the chip [3]. Similarly, temperature sensors can be used to identify the operating temperature [2]. Thus the outputs of these two sensors can be used to select the correct table corresponding to the particular process-temperature operating point.

The size of the LUT for each process, temperature corner would vary from 16 to 32 bytes depending upon the granularity of the workloads stored. These can be implemented in the internal memory of the processor. The searching power is minimal since the search needs to be done only once at the beginning. If the algorithm that runs on the processor (IC) is known, the workloads for the major units are also usually known. In such cases, these LUTs need not be stored on the chip at all. A one time settings of the VDD, VBB values at the beginning would do.

References

- [1] Intel White Paper on “Enhanced Intel SpeedStep Technology for the Intel Pentium M Processor”,
<ftp://download.intel.com/design/network/papers/30117401.pdf>
- [2] Bakker A., Huijsing J.H., “Micropower CMOS Temperature sensor with digital output”, IEEE Journal of Solid State Circuits, vol. 31, no. 7, pp. 933-937, July 1996.
- [3] Bhushan M., Gattiker A., Ketchen M. B., Das K.K., “Ring oscillators for CMOS process tuning and variability control”, IEEE Trans. On Semiconductor Manufacturing, vol. 19, no. 1, pp. 10-18, February 2006.

Chapter 5

Implementation

This chapter discusses how to implement the simultaneous supply and body bias control based on workload. It also describes how to fit our novelty in the implementation schemes currently being followed. Some of the recent research [1] describes the simultaneous VDD and VBB adjustments based on a closed loop control. But the drawback of those being the difficulty to reproduce replica circuits to measure the dynamic and leakage components. In our scheme, there are no replica circuits. The simulations can be run on the layout extracted netlists or tests can be run on the real chip to characterize for the relative contributions of the dynamic and the leakage components in the total power consumed. These power numbers can be used to generate a lookup table which is used to select the VDD and VBB values for any algorithm that is run on the chip.

5.1 Introduction

The simultaneous supply and body bias control based on workload can be implemented without any additional overhead in all the systems which already have a provision to control VDD and VBB. In processors which run predefined fixed function algorithms, this VDD and VBB can be easily chosen from the precharacterized lookup table. If the algorithm varies from time to time, the VDD and VBB also needs to be continuously readjusted based on the workload.

5.2 Implementation Methods

In many general purpose signal processing integrated circuits, the algorithm that runs on the processor is predefined. The algorithm could vary from application to application, but for a particular end application, the algorithm is always fixed. Examples of such algorithms are the FFT, FIR, Phaser Effect etc. mentioned in Table 2.1. Once the algorithm is fixed, the workload of all the major data path blocks in the system can pre-computed from simulations. Once the workload is available, the corresponding VDD and VBB values can be obtained from the pre-characterized table mentioned above. On application of these VDD and VBB values, there is minimal total power consumption due to an optimal balance between the leakage and dynamic power components.

The above scheme can be used if there is only one algorithm running on the processor and the workload of the unit (block) for that algorithm is fixed which is true in many signal processing applications. If multiple algorithms are being run on the processor or if the same algorithm would use the units at different workloads at different times, the VDD and VBB of the block could vary from time to time. There are two different ways to achieve this. The VDD control is different in the two methods and is described below. The body bias control remains same in both the schemes and can be achieved with charge pumps since the substrate currents are small [4].

In the first method (M1), the VDD of a block can be changed by using a DC-DC converter when the processor moves from one algorithm to another. A

representative diagram of this method is given in Fig. 5.1. Here, the granularity of the supply voltage change can be as high as allowed by the DC-DC Converter. The standard available converters allow switching of the order of 20mV and have efficiencies of about 97% [2]. While the instantaneous VDD can bounce around 10% (or even more) of the nominal value during transients, the nominal (or average) value can be controlled to within 15mV accuracy using DC-DC converters available commercially. Hence we can obtain almost the full power savings as given in Table 3.1.

The second method (M2), shown in Fig. 5.2 is similar to the local voltage dithering described in [3]. In this, we have implemented VDD and VBB switching based on PMOS switches. In M2, we choose two supply voltages and the corresponding body bias voltages. Only two workloads, namely a high workload and a low workload can be accommodated in this method. Depending on the workload, either of the PMOS switched is turned on and the other is turned off. The Body bias is chosen from the pre-characterized table. If the workload is high, the VDD low switch is turned on and the VDD high switch is turned off since the dynamic power component is dominant here. If the workload is low, the VDD high switch is turned on and the VDD low switch is turned off. The magnitude of the VDD changes here are much smaller compared to the changes in [3] since the operating frequency for the block is unchanged here. Hence very fast VDD switching times of the order of a few nanoseconds can be achieved.

For the Adder and the MAC units, using PMOS switches which occupy less than 5% of the area of the computing unit, we could switch the supplies in about 4-5ns. VBB is also changed accordingly using charge pumps. Even though the magnitudes of VBB changes are large, the capacitances associated with VBB lines are very less. This is because the substrate currents are small and hence the VBB lines can be narrower compared to the VDD lines. Similar to [3], our Method 2 also provides a local power gating essentially free. There is a finite energy overhead to switch the supplies. For the 65nm Adder, this is about 190pJ. This energy can be recovered in about 47 adder cycles from the savings obtained by adjusting the VDD and VBB to the correct workload values. In other words, not adjusting the VDD and VBB based on workload would have resulted in a loss of 190pJ in 47 cycles of operation. This in turn limits the granularity of switching between different voltage settings.

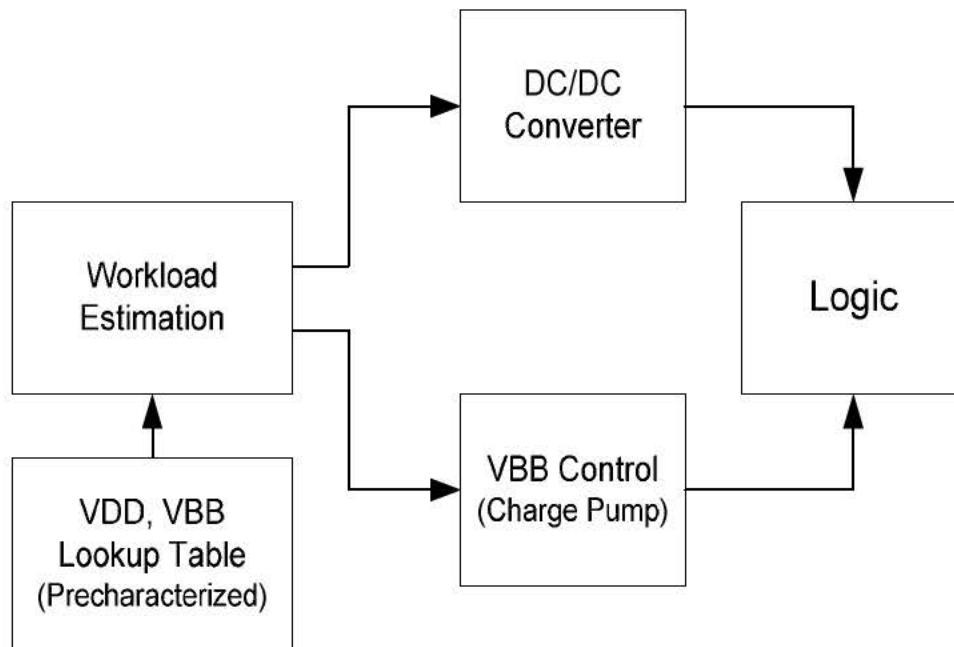


Fig 5.1: Block diagram of VDD/VBB Control using DC/DC Conversion

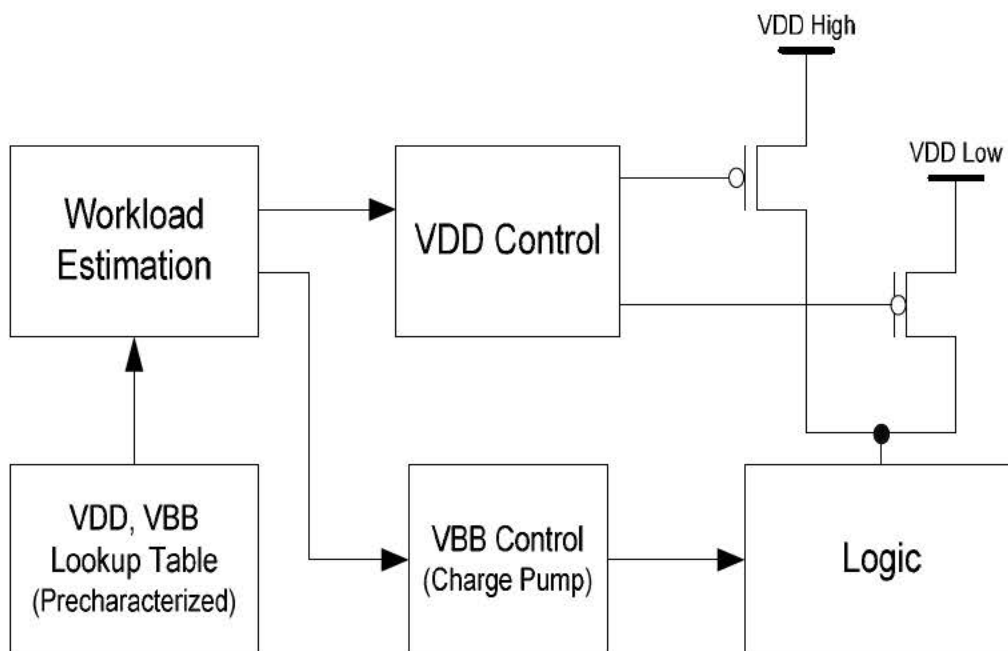


Fig 5.2: Block diagram of VDD/VBB Control using local quantized supply voltages

If M2 is employed, the power savings would certainly be lesser than the maximum possible due to the fewer granularities in VDD and VBB control. Only two VDD and VBB values are possible here. But still the two operating points can be chosen in such a way that one corresponds to a high workload and the other corresponds to a low workload. Assume the two supplies are chosen in such a way that VDD High corresponds to a workload of 0.3 and VDD Low corresponds to a workload of 0.7 and similarly the VBBs. If there was only one VDD and VBB, the voltages chosen should ideally correspond to workload of 0.5. Now there would be savings for all workloads in the ranges (0 0.4) and (0.6 1) compared to the non work load indexed scheme. The magnitude of the actual savings would depend on the actual workload.

In both the methods M1 and M2, while changing the supplies and body bias voltages, a particular order has to be followed to ensure that the operating frequency doesn't fall below the fixed frequency at any time. When changing from a high VDD to a low VDD, the VBB has to be reduced first before the VDD change. When changing from a low VDD to a high VDD, the VDD change should be done first before the VBB change. These steps ensure that the unit can work correctly at the operating frequency.

Separate ABB for P and N can lead to degraded setup-hold times as well as larger susceptibility to supply noise and can lead to more jitter in the clock nets. However, we don't change the body bias of flops and clock driver nets. All the savings currently projected are with the flops and clock drivers getting zero body bias.

The layout of the gates is modified to bring out the substrate contacts for giving external body bias, for both P transistor substrate and N transistor substrate. This increases the area of the adder unit by about 11%. But this overhead is there even if we just employ the combined VDD-VBB power optimization without any voltage adjustments based on workload. i.e Our workload based voltage adjustments don't add any extra overhead compared to the existing VDD-VBB scheme. With our workload based lookup table changes, additional change is either the need for a different set of switches for each block if use the "M2" or a different DC-DC converter for each block if we use "M1". "M2" definitely has less overhead but the penalty would be the lesser granularity in VDD-VBB adjustments.

The area overhead for the switches and the power routes in M2 would be around 4% of the area for a 16 bit adder in 65nm technology.

Fig 5.3 shows the layout of a standard cell with no substrate contacts. Fig5.4 is the layout of the same cell with the substrate contacts brought out. This cell has two additional power routes, VPB and VNB in addition to the VDD and VSS. The area of the cell increases by about 10%-11% with the change.

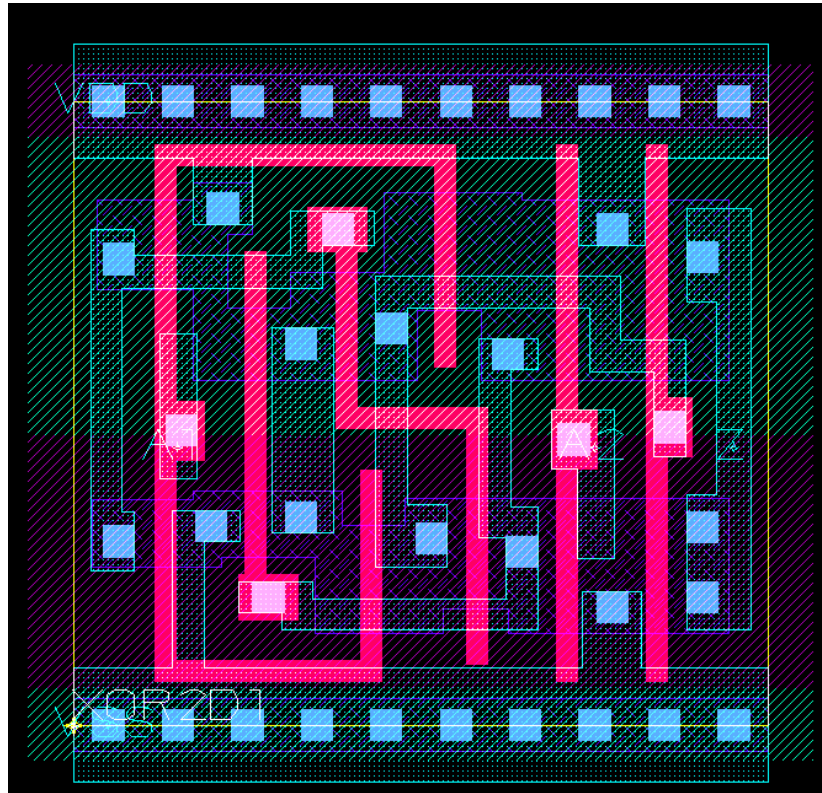


Fig 5.3: Stdcell cell with substrate contacts shorted to the local power supplies

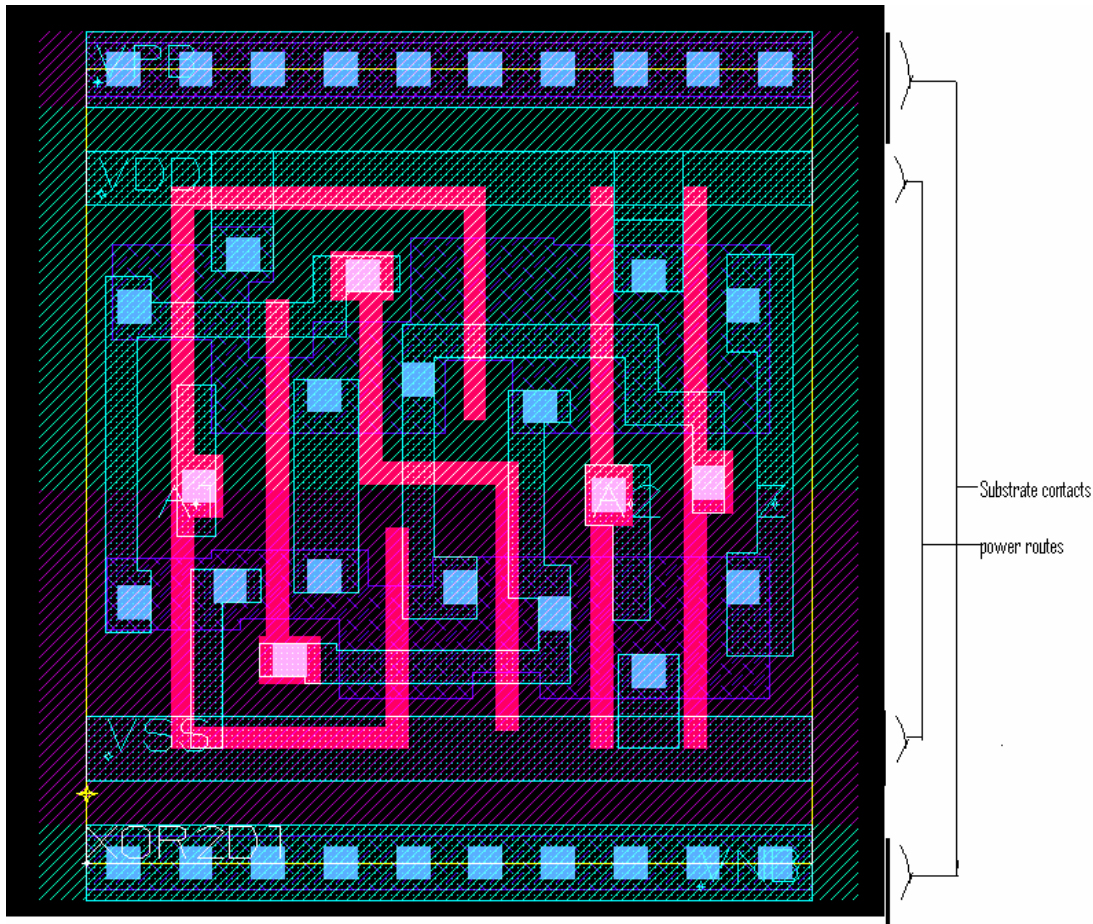


Fig 5.4: Stdcell cell with substrate contacts brought out

A normal ASIC synthesis flow wouldn't be affected much with this since just two additional power routes are needed and this can be taken care in floor planning. The cell libraries would need to change to accommodate VBB as a new parameter if we were to use the ASIC design flow. Hence this would need a recharacterization of timing.

Fig 5.5 shows the layout of an adder with all the substrate contacts shorted to the 3 local power supplies and Fig 5.6 shows the same adder with the substrates brought out. There are two extra power routes in Fig 5.6 compared to Fig 5.5. The logic area (both height and width) is same in both

the figures. The modified adder in Fig 5.6 consumes approximately 11.5% more area than the original adder in Fig 5.5. The horizontal dimension is exactly same but the vertical dimensions are different due to the increase in area due to the extra power routes.

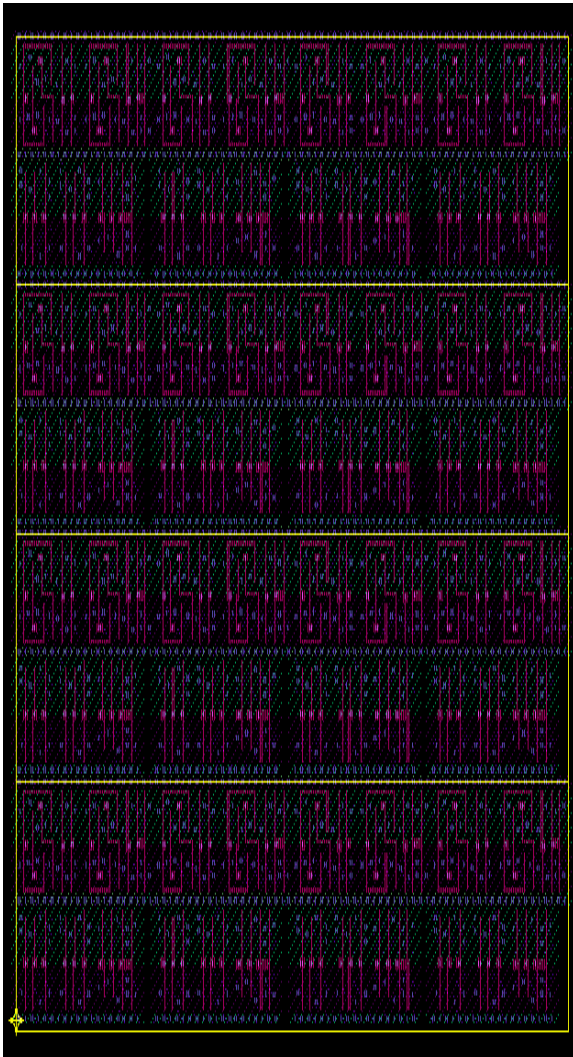


Fig 5.5: Layout of a 16 bit adder in 65 nm with substrate contacts shorted to the local power supplies

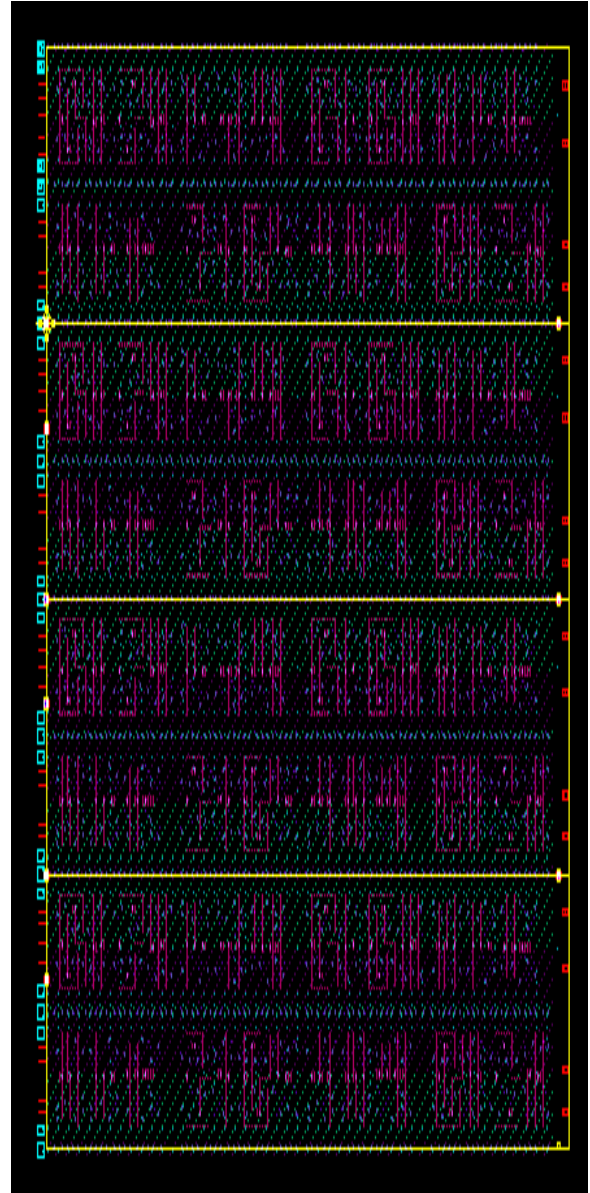


Fig 5.6: Layout of a 16 bit adder in 65 nm with substrate contacts brought out.

Systematic variability will be compensated by the inherent feedback nature of this technique. We haven't explicitly discussed how to sense the process condition, but this is routinely done using ring oscillators and we expect this to be done once during manufacturing test. We expect random variability to reduce the gains reported but we haven't quantified this in our study.

5.3 Algorithm to generate the workload based lookup table for any block

The following series of steps can be employed to generate the workload indexed lookup table for a random block.

- (i) Do Supply voltage scaling and find the minimum voltage (V_{DVS}) below which the operating frequency criterion, f' cannot be met. This corresponds to the supply for zero body bias. Measure the total power drawn from the supply at this operating point.
- (ii) Increase the supply voltage in small steps. For each supply voltage, the body bias is varied until the point where the frequency of operation of the block falls below f' . Measure this body bias voltage. Also measure the total drawn from the supply at this point.
- (iii) The experiments are done by independently varying P transistor body bias and N transistor body bias.
- (iv) Repeat steps (i) to (iii) for a range of VDD values for which the body bias is less than the maximum allowed by the process technology. Measure the total power at each operating point.
- (v) From the power measurements of step (iv), choose the supply and body bias corresponding to the least power operating point.

- (vi) Repeat steps (i) to (v) for a range of values of the block workload. Measure the Supply and Body Bias voltages corresponding to the lowest power operating point in each case.
- (vii) Construct the workload indexed table similar to Table 3 and Table 4.

References

- [1] Masahiro Nomura, Yoshifumi Ikenaga, Koichi Takeda, Yoetsu Nakazawa, Yoshiharu Aimoto, and Yasuhiko Hagihara, "Delay and Power Monitoring Schemes for Minimizing Power Consumption by Means of Supply and Threshold Voltage Control in Active and Standby Modes" in IEEE Journal of Solid State Circuits, April 2006, pp. 805-814.
- [2] Analog Devices DC-DC Converters. <http://www.analog.com/en/power-management/switching-regulators-integrated-fet-switches/products/index.html>
- [3] Benton H. Calhoun and Anantha P. Chandrakasan, "Ultra-Dynamic Voltage Scaling (UDVS) Using Sub-Threshold Operation and Local Voltage Dithering" in IEEE JSSC, vol. 41, no. 1, January 2006.
- [4] Dong-Jae Lee, Yong-Sik Seok, Do-Chan Chi, Jae-Ifyeong Lee, Young-Rae Kim, Hyeun-Su Kim, Dong-Soo Jun, and Oh-Hyun Kwon, "A 35 ns 64 Mb DRAM using on-chip boosted power supply," in Symp. on VLSI Circ. Dig. Tech. Papers, June 1992, pp. 64-65

Chapter 6

Conclusion

A combination of DVS and ABB can optimally provide the least power operating point for a general circuit. This could be different from the operating point when only DVS or only ABB are employed. This least power operating point varies with workload of the circuit. For each workload, there is an operating point sweet spot which is a particular combination of supply and body bias voltage which gives the minimum total power. The power savings vary for different workloads of the circuit.

As a summary, if neither DVS nor ABB is employed, there is maximum power consumption. If only DVS is employed, the dynamic power component is reduced since the voltage applied is just enough for the critical path to operate correctly. But this reduces the leakage component to a small extent. When ABB is also used along with DVS, the leakage component is reduced at the expense of dynamic power component. The dynamic power would be more now but total power which is the sum of the leakage and dynamic components would be less. The leakage reduction over weighs the dynamic power reduction for some combinations of VDD and VPB values.

The VDD VBB optimal power point varies with workload. Hence workload can be used as an index to select the optimal VDD VBB pair of values from a set of values. The operating frequency criterion is met at any of these VDD

VBB pairs, but the minimal power for a given workload is obtained only at a specific VDD-VBB value pair. The actual implementation and analysis of this scheme is straightforward and fits into any standard IC design methodology.

There are issues while extending the simulation based approach to a large hardware block. We propose two different ways to overcome this.

Many larger blocks of hardware are usually hierarchical. i.e Most are made up for smaller similar units. So our simulation based approach can be used on a smaller building block of the larger hardware and the results for the larger hardware block can be predicted.

There is still another way around if the hardware is a fixed function component like a hardware accelerator. The entire hardware can be provided with a single VDD and VBB control. The fabricated hardware can be characterized at different workloads. For each workload, the supply and the body bias voltages can be varied to obtain minimal total power and the coo-ordinates (VDD, VBB) for each workload can be recorded. This can be done for a range of workloads which the chip would be subject to. These minimal power coordinates can be indexed by the workload and can be distributed along with the hardware data sheet. The characterization work needs to be done only on one single part. For a general purpose programmable hardware with many units, it would be difficult to define a

workload for the entire hardware and hence the results would be sub optimal.

With scaling device geometries, the control of body bias on threshold voltage is weakened. However with multigate FETs, it might be possible to use one of the gates to achieve threshold control. A full exploration of these issues is beyond the scope of this paper. But this will be an interesting future work.