

January 2007

Current-sensed Interconnects: Static Power Reducation and Sensitivity to Temperature

Sheng Xu

University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Xu, Sheng, "Current-sensed Interconnects: Static Power Reducation and Sensitivity to Temperature" (2007). *Masters Theses 1911 - February 2014*. 57.

Retrieved from <https://scholarworks.umass.edu/theses/57>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

CURRENT-SENSED INTERCONNECTS:
STATIC POWER REDUCTION AND SENSITIVITY TO
TEMPERATURE

A Thesis Presented

by

SHENG XU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

September 2007

Electrical and Computer Engineering

© Copyright 2007

All Rights Reserved

CURRENT-SENSED INTERCONNECTS:
STATIC POWER REDUCTION AND SENSITIVITY TO TEMPERATURE

A Thesis Presented

by

SHENG XU

Approved as to style and content by:

Wayne P. Burlison, Chair

Maciej Ciesielski, Member

Csaba Andras Moritz, Member

C.V. Hollot, Department Head
Electrical and Computer Engineering

DEDICATION

To my parents....

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Wayne P. Burleson, for his patient guidance and continued support. He guided me through the area of VLSI and gave me ample opportunities to be specialized in current sensing interconnect circuit design.

I thank my committee members, Prof. Maciej Ciesielski and Prof. Csaba Andras Moritz for their suggestions and guidance.

I want to express my appreciation to Jinwook Jang, Vishak Venkatraman, Ibis Benito and other co-workers in the Circuit and System group for their great support and inspiration.

I wish to express my gratitude to the Semiconductor Research Corporation (SRC) and Intel Corporation for supporting this work.

I want to thank Cadence and Synopsys for CAD tools support through their university program.

ABSTRACT

CURRENT-SENSED INTERCONNECTS:

STATIC POWER REDUCTION AND SENSITIVITY TO TEMPERATURE

SEPTEMBER 2007

SHENG XU

B.S., SHANGHAI UNIVERSITY OF ENGINEERING SCIENCE

M.S.E.C.E, UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Wayne Burleson

Global on-chip interconnects in deep sub-micron CMOS present challenges in satisfying delay constraints in the presence of noise and dramatic temperature variations, while minimizing energy consumption due to leakage and static power. Although repeaters are typically used to reduce delay and maintain signal integrity in long interconnects, they introduce significant area, power (both dynamic and leakage), delay, noise and design overhead as well as exacerbating variations due to their local power supply noise and temperature. Current-Sensing is an alternative to repeaters that transfers signals with no intermediate circuits by sensing current rather than voltage at the end of a long interconnect. Among the current sensing circuits, Differential Current-Sensing (DCS), which uses conventional CMOS inverters to drive differential signal, is preferred because of its high common-mode noise rejection. The DCS circuit is fast and simple in layout compared to repeater insertion despite significant static and leakage power which remains a barrier for broad

application. Temperature variation throughout the chip also causes the timing uncertainty on interconnects to increase.

This thesis addresses current-sensing interconnect circuit design in several aspects. First, it provides an improved differential current-sensing circuit called the differential leakage-aware sense amplifier (DLASA), that uses local power gating that results in 39.6% reduced leakage and static power compared to conventional differential current sensing. Secondly, thermal impact on interconnect is studied and temperature sensitivity is analyzed for interconnect circuits. Theoretical analysis is discussed as a base design guideline, then accurate simulation based experiments in 65nm, 45nm and 32nm CMOS technologies are used for verification from 25^oC to 150^oC. Thus this project provides a view of the year of technology toward 2013.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	v
ABSTRACT.....	vi
LIST OF TABLES	x
LIST OF FIGURES.....	xi
CHAPTER	
1 BACKGROUND	1
1.1 Interconnect Circuit Challenge	1
1.2 Existing Interconnect Circuits.....	3
1.3 Differential Current Sensing	4
1.4 Thermal Impact on Interconnect Circuit.....	5
2 ANALYTICAL AND EXPERIMENT APPROACH.....	7
2.1 HSPICE Model	7
2.1.1 Wire Model	7
2.1.2 Device Model.....	12
2.2 Circuit Simulation.....	14
2.2.1 Repeater Circuit	14
2.2.2 Repeater Insertion Optimization	15
2.2.3 Differential Current Sensing	17
2.3 Experiment Setup.....	23
3 ENERGY-AWARE DIFFERENTIAL CURRENT SENSING	24
3.1 Energy-aware Differential Current Sensing	24
3.2 Experimental Setup.....	29
3.3 Repeater Optimization	29
3.4. Results and Discussion	32
3.4.1 First Order Comparison	32
3.4.2 Activity Factor Impact	33
3.4.3 Wire Length Impact	34
3.4.4 Driver Size Impact	37

3.4.5 Technology Scaling Impact.....	39
3.4.6 Signaling Complexity and Area Efficiency	41
3.5 Conclusions.....	42
4 INTERCONNECT CIRCUITS UNDER THERMAL CHALLENGE	44
4.1 Thermal Challenge in DSM Integrated Circuits	44
4.2 Temporal Temperature Variation on Interconnect.....	46
4.2.1 Impact on Wire Segment and Single Transistor	46
4.2.2 Impact on Repeated Line	48
4.2.3 Impact on DCS and DLASA Circuit.....	52
4.3 Spatial Temperature Variation on Interconnect	57
4.4 Analytical Model for Repeated Line.....	68
4.5 Summary and Conclusion	72
5 SUMMARY	74
6 BIBLIOGRAPHY	76

LIST OF TABLES

Table	Page
2.1 Top Metal Layer Dimensions from PTM.....	11
2.2 Calculated Wire Parameters for 65nm, 45nm and 32nm.....	12
3.1 Interconnect And device parameters.....	31
4.1 Temperature Variation Effects on Delay	50
4.2 Spatial Temperature Variation Impacton Repeated line	74

LIST OF FIGURES

Figure	Page
1.1 ITRS trend of Interconnect delay, wire spacing and resistivity	2
2.1 5-pi distributed RC wire.....	9
2.2 Inductance model for simulation.....	10
2.3 Repeater Insertion Line	15
2.4 (a) left: Voltage Mode Configuration	18
2.4(b) right: Current Mode Configuration	18
2.5 Current Sensing Circuit in interconnect.....	19
2.6 Maheshwari's Differential Current Sensing Amplifier.....	22
2.7 Differential Current Sensing Circuit	23
2.8 Simulated waveform of DCS	24
2.9 Experimental setup and flow	25
3.1 Delay and Energy comparison between DCS and repeater.....	27
3.2 Differential Leakage-Aware Sense Amplifier (DLASA).....	29
3.3 Simulation waveforms of DCS and DLASA	30
3.4 Delay and leakage power for HVT and NVT repeaters	33
3.5 Delay and energy of repeaters and DLASA	34
3.6 Energy Comparison under different Activity Factors	36
3.7 Leakage power of HVT, NVT repeaters and DLASA	37
3.8 Static and leakage power in DCS and Energy-aware DCS from 1mm to 10mm	38
3.9 1mm-10mm wire Energy Versus Delay.....	38

3.10 Leakage power of high Vt repeater, normal Vt repeater, and Energy-aware DCS varying driver size on 5mm wire	39
3.11 5 mm wire Energy Versus Delay on Driver Size Varying	40
3.12 5mm wire Static+leakage power in DCS and Energy-aware DCS on varying driver sizes	41
3.13 Technology scaling impact on DCS and DLASA respect to propagation delay from 1mm to 5mm.....	42
3.14 Technology scaling impact on DCS and DLASA respect to propagation energy from 1mm to 5mm	42
3.15 Technology impact on DCS and DLASA, 5 mm wire Energy Versus Delay on Driver Size Varying.....	43
3.17 Circuit area comparison among DCS, DLASA and repeater.....	44
4.1 Percentage of delay increase for temporal thermal variation in 65nm,45nm and 32nm repeated interconnects.....	50
4.2 Percentage of energy increase for temporal thermal variation in 65nm,45nm and 32nm repeated interconnects.....	51
4.3 Temporal thermal variation impact on delay for 1mm-5mm repeated interconnects.	52
4.4 Temporal thermal variation impact on energy for 1mm-5mm repeated interconnects.	53
4.5 Temporal thermal variation impact on delay for different repeater numbers in 65nm, 45nm and 32nm repeated interconnects.....	54
4.6 Percentage of delay increase for temporal thermal variation on a 3mm DCS wire.....	55
4.7 Temporal thermal variation impact on delay for 65nm, 45nm and 32nm for 1mm-5mm DCS.	56
4.8 Temporal thermal variation impact on energy for 45nm, 1mm-5mm repeated interconnects.....	56
4.9 Impact on delay and energy due to temporal thermal variation on a repeated interconnect compared to DCS for a 45nm, 3mm wire.....	57

4.10 Impact on delay due to temporal thermal variation on a DCS and DLASA for 3mm wire.	58
4.11 Impact on delay due to temporal thermal variation on a DCS and DLASA for 3mm wire.	58
4.12 Spatial distribution profiles applied on a repeated interconnect and a current-sensed interconnect.....	58
4.13 Impact of spatial thermal variation on delay for 65nm, 45nm and 32nm repeated interconnects.....	60
4.14 Impact of spatial thermal variation on energy for 65nm, 45nm and 32nm repeated interconnects.....	60
4.15 Impact of two nonuniform thermal distribution profiles on the delay of 65nm, 45nm and 32nm repeated interconnects.	61
4.16 Impact of two nonuniform thermal distribution profiles on the energy of 65nm, 45nm and 32nm repeated interconnects.	61
4.17 Impact of two nonuniform thermal distribution profiles on the delay of 65nm, 45nm and 32nm repeated interconnects.	63
4.18 Impact of two nonuniform thermal distribution profiles on the energy of 65nm, 45nm and 32nm repeated interconnects.	63
4.19 Impact of two nonuniform thermal distribution profiles on the delay of 65nm, 45nm and 32nm DCS.	64
4.20 Impact of two nonuniform thermal distribution profiles on the energy of 65nm, 45nm and 32nm repeated interconnects.	65
4.21 Impact of spatial thermal variations on delay and energy for varying number of repeaters on 65nm, 45nm and 32nm repeated interconnects.	66
4.22 Comparison between the impact on delay and energy due to spatial thermal variation on a repeated interconnect and differential current sensing.....	67
4.23 DLASA/DCS delay under different temperature profiles.....	68
4.24 DLASA/DCS energy under different temperature profiles.....	68
4.25 DLASA/DCS delay with different temperature variation.....	69

4.26 DLASA/DCS energy with different base temp..... 70

CHAPTER 1

BACKGROUND

This chapter introduces the background and motivation on this thesis. Section 1.1 explains the trend and challenge of VLSI circuit design. Section 1.2 introduces several existing interconnect circuits. Among the interconnect circuits that have been discussed in 1.2, Differential Current Sensing (DCS) has its advantages on speed and common mode noise rejection while it has several drawbacks such as high static energy dissipation. Section 1.3 explores both advantages and drawback of DCS. Temperature variation on interconnect have been discussed in section 1.4. The organization of this thesis is introduced at the end of this chapter.

1.1 Interconnect Circuit Challenges

Challenges have been presented in global on-chip interconnects in deep sub-micron CMOS. Delay constraints have to be satisfied under harsh conditions including noise and dramatic temperature variations, while energy consumption due to leakage and static power is required to be minimal.

As the geometry of wires shrinks and routing density increases, wire resistance is increasing due to reduced cross-sectional area, and the coupling capacitance is also

increasing due to reduced line spacing and insulator thickness. The resulting delay of long interconnects becomes a major component of the timing budget in VLSI circuits. Repeater insertion is a standard interconnect optimization method. By using buffers to break the wires into short segments, the quadratic relation between the wire length and delay will be decreased toward near linear order. In [1, 2] and many other works, various repeater insertion methods have been explored. Repeaters are usually very large since they need to drive the wire fast enough to meet the timing budget. The increase in die size and the shrinking of geometries result in the rapid increase of the relative length of global interconnects. The total number of repeaters remains significant even as the absolute wire length tends to decrease. [3] The significant amount of repeaters gives a challenge to the circuit design since:

1. The delay of the wire is sensitive to the placement of the repeaters. Since the available layout space is very limited, the ideal placement of certain repeaters may not be satisfied and hence lead to a sub-optimal result.
2. The dynamic power will increase as the repeater sizes increase.
3. Even if the interconnect has a low activity factor which result in less dynamic power consumption, the leakage power is still an issue.

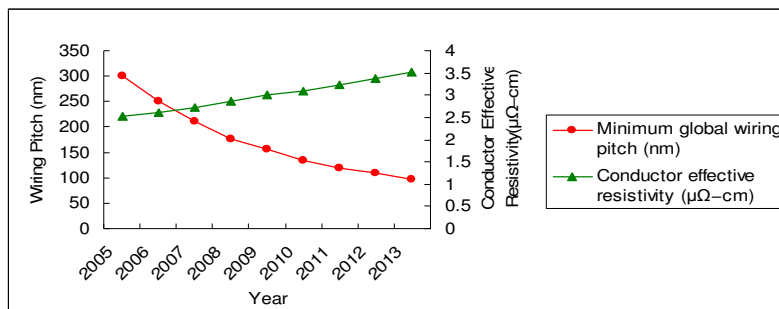


Figure 1.1 ITRS trend of Interconnect delay, wire spacing and resistivity

1.2 Existing Interconnect Circuits

There are several alternatives to repeater insertion. The examples of these are booster insertion, phase coding, differential current sensing and multi-level current signaling.

A booster detects a transition earlier than a conventional inverter and then accelerates it to a full logic swing level. A booster attaches along the wire rather than interrupting it. Booster can be used for driving bidirectional signals. Layout placement is not an issue for boosters, and these results in layout simplicity. The drawback is that it can not be combined with logic. It is also not suitable for interconnects that require buffering [4].

The width of the pulse can reflect the actual analog value of the signal transmitted. The phase coding technique extends the Pulse Width Modulation (PWM) principle to digital signal lines. It enables power savings because signal transitions in the encoded group will be translated to only two transitions according to the modulated pulse. It provides a means of transmitting multiple bits on a single wire which improve the bandwidth. Phase coding has some drawbacks the additional encoder and decoder area, and its susceptibility to noise. Additionally, the sizing of the encoder and decoder is not trivial [5].

The multi-level signaling system is a current-mode system that consists of a driver, a receiver and a decoder. The driver encodes the two bits of signals into four current levels and transmits them. The currents propagate through the interconnect and are compared at the receiver to a reference current. The receiver converts the four current levels into thermometer codes. Finally, the decoder recovers the original signal. This

method realizes multi-bit signaling in one clock cycle. The speed is comparable to repeater insertion. The observations are similar to phase coding since it prunes to noise and process variation [6].

1.3 Differential Current Sensing

Among the alternative interconnect circuits, differential current sensing is a promising option. A differential current sensing circuit consists of a pair of drivers at the beginning of the wire and a receiver/amplifier at the end. Instead of using voltages as the signal, it transfers currents to the receiver, and in turn, it amplifies the currents to full swing voltage output.

The advantages of the differential current sensing are

1. It is not sensitive to coupling capacitance
2. It is fast compared to voltage sensing circuit
3. It does not break the wire into segments. Thus it provides more layout flexibility.

Differential current sensing overcomes the non-trivial but common problem, the sensing of the current mode circuit. Another advantage of differential signaling is the immunity to noise due to its high common-mode rejection.

Differential inputs and outputs will increase the routing area and the extra clock is an overhead compared to repeater insertion. Another major drawback of differential current sensing is the static and leakage power consumed by the receiver. Since the current is used as signaling parameter, there will be a path to ground from the driver.

As a result, high static power dissipation is expected in current-mode signaling. Meanwhile, voltage mode circuits such as repeated lines are very good at static power reduction since they turn off the current when there is not signal switching. The merit of low static power consumption in repeaters makes it more attractive to designers than traditional DCS. As the technology shrinks to 65nm and beyond, leakage power becomes more dominant in integrated circuits. This trend puts a new power cross-over point between repeater insertion and current sensing. Thus, current mode circuits become more competitive in terms of power consumption. Still, the total energy dissipation of the DCS circuit is not affordable. In [7], an energy-aware differential current sensing circuit is proposed and will be further refined and discussed in Chapter 3. The proposed circuit effectively prevents leakage and static current by using power gating technique and hence reduces the total energy considerably.

1.4 Thermal Impact on Interconnect Circuit

Chip temperature is becoming more difficult to handle in deep-sub-micron regimes. Consequently, temporal and spatial hotspots across chip induce various performance and reliability problems. Efforts have been made to correct this in all fields of semiconductor technology, from an architectural standpoint down to material science. This thesis investigates the mechanism of thermal surge in digital microprocessor, and reviews techniques on thermal analysis and management. Recent advances in architecture and circuit are explored. Advantages and limitations of the existing strategies are demonstrated.

Thermal sensitivity could be as important as other aspects when choosing an appropriate interconnect circuit. It is beneficial to understand the performance and power change of the circuit in presence of different temperature environments. This temperature variation could be spatial, which means one wire goes through several different temperature regions. Or the variation could be temporal, which means that the interconnect circuits experience different temperatures in a time domain. Temporal hotspots should be relatively manageable since the temperature patterns can be clear and the prediction techniques are somewhat developed. Managing a circuit across several temperature regions could be complicated even when the temperature changes gradually through the area. Both temporal and spatial variations would result in an unpredictable output and signal degradation. Chapter 4 analyze the thermal impact on DSM interconnect both theoretically and experimentally. Repeater, DCS and DLASA are compared as different interconnect circuit implementation under different thermal profiles.

This thesis will be organized as follows: In Chapter 2, analytical approaches and experimental methodology are explained. In Chapter 3, an energy-aware differential current sensing circuit is proposed and analyzed. Chapter 4 reviews the thermal related research. Both repeated and DCS interconnect circuits under spatial and temporal thermal distribution profiles will be addressed. A summary is given in Chapter 5.

CHAPTER 2

ANALYTICAL AND EXPERIMENT APPROACH

This chapter explains the analytical approach and experimental setup of the proposed project. Interconnect/wire and transistor/device have been analyzed in section 2.1. These models are used for HSPICE simulation. Repeater insertion line and differential current sensing circuit have been built and verified in section 2.2. The optimization methods have been discussed and hence the advantage of simulation-based approach has been recognized as the optimization strategy. The experimental and data extraction process is explained in 2.4.

2.1 HSPICE Model

2.1.1 Wire Model

It is not practical to model on-chip wires without knowing the trend of semiconductor materials and fabrication, while it is also critical to keep the circuit model to a certain degree of abstraction. Interconnect wires can be categorized into three types. For a seven metal layer microprocessor, the top two or three layers are used for global

wires. Several middle layers are classified as intermediate interconnect layers. The bottom layers are local interconnect layers. Among these three, global wires are the most challenging layer for designers. Global wires are usually long (3 mm to 10 mm in 65nm) in order to transfer signals between blocks, e.g. on-chip buses. The activity factor on these wires is usually not as high as in a local wire which means the low leakage circuit will exert its advantage. Repeaters appear attractive to designers on these layers since they have lower leakage and static power. Global wires are usually slow due to the capacitive coupling between the lines and the large load capacitance due to the long wire. There are several techniques including shield wires, use of low dielectric materials and fat wires. Shield wires, i.e., either Vdd or Gnd, are intently put between every other metal wire or between alternative wires. Thus, it prevents the noise resulting from the coupling capacitance. Another approach is to use fat wires for global interconnect. Since resistance is inversely proportional to the wire width and height, fat wires will decrease resistance and the RC delay. The drawbacks of the strategies are also obvious. Fat wires can not always be achieved since the space for global wires are very limited at the chip layout level. Shield wires will add routing redundancy. Besides the efforts on dimension and layout, new materials that have lower resistivity and dielectric constant are also promising in interconnect applications. Aluminum has been replaced with copper in the top metal layers since it has a lower resistivity (2.2 mOhm/cm) than aluminum (3.9 mOhm/cm). It means that for the same wire length, Cu will have lower resistance than Aluminum. Low dielectric (or Low K) materials are used for silicon insulators which are between the

metal layers. As K goes down from 3.0 to 1.5 or even lower, the overall coupling capacitance is expected to shrink. These two technologies will lower the RC delay by decreasing R and C . Tuning strategies attempt to maintain the signal integrity and performance at the same time. While new fabrication methods and new materials are promising, concise and careful circuit design is essential to achieve the success of the signaling on interconnect. Without boosting and restoration, the signal can not travel through the long wire properly and efficiently.

Preliminary works in this proposal focus on design based on a distributed RC network model. A lumped RC model is pessimistic for a modern resistive-capacitive wire. Distributed RC is more accurate estimation of delay and power. A 5-pi distributed wire is used for the wire segment since it has higher accuracy while it is still relatively simple for simulation.

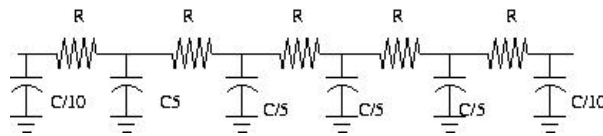


Figure 2.1 5-pi distributed RC wire

It is not trivial to model interconnect inductance including mutual and self inductance because the complex mutual magnetic flux metrics and the current return path. As it has been analyzed in [8], the inductive effects (i.e. ringing, overshoot, undershoot etc.) are not observed for differential current sensing in 180nm technology. This immunity is expected to retain in lower technology node. The reason is that the ac current in the two differential wires is always opposite in direction and hence the magnetic fields generated are opposite in nature, resulting in a very small effective inductance. Also,

the reflection coefficients at the receiver end and at the driver end are very small. The two wires in DCS acts as a return path, so the effect of return path impedance is almost negligible. Nevertheless, the edge rates become faster in 65nm and beyond than in 180nm, so the edge rates deterioration is expected to be more significant than in 180nm. Furthermore, since performance in repeated line is always prone to inductance, a simplified, yet, accurate way to model inductance is considered for future work. Effective inductance is calculated in PTM and inserted into the previously used RC model to give a RLC model. The effective inductance considers self inductance and neighbor mutual inductance and distributed into a 5-pi model.

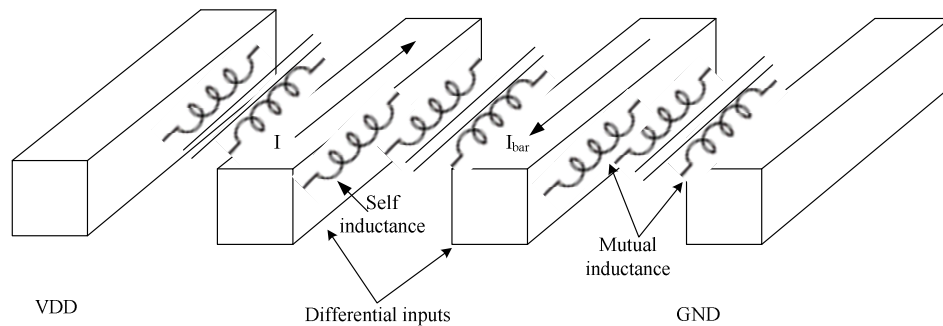


Figure 2.2 Inductance model for simulation

Wire parameters are from PTM (Predictive Technology Model) for 65nm technology [10] and ITRS(International Technology Roadmap for Semiconductor) for 45nm and 32nm respectively. For a top metal layer, the dimensions are shown in table 2.1.

Table 2.1 Top Metal Layer Dimensions from PTM

	width (um)	Space (um)	thickness (um)	height (um)	k _{ILD}
65nm	0.45	0.45	1.2	0.20	2.2
45nm	0.315	0.315	1.0	0.15	2.2
32nm	0.2205	0.2205	0.9	0.06	2.1

PTM calculate the resistance as:

$$R = \frac{\rho \cdot l}{w \cdot t} \quad (2.1)$$

And total capacitance as:

$$C_t = C_g + 2C_c \quad (2.2)$$

Where C_g is the area and fringe flux to the underlying plane and C_c is the coupling

capacitance that can be represented as:

$$C_g = g \left[\frac{w}{h} + 2.22 \left(\frac{s}{s + 0.70h} \right)^{3.19} + 1.17 \left(\frac{s}{s + 1.51h} \right)^{0.76} \cdot \left(\frac{t}{t + 4.53h} \right)^{0.12} \right] \quad (2.3)$$

And

$$C_c = \varepsilon \left[1.14 \frac{t}{s} \left(\frac{h}{h + 2.06s} \right)^{0.09} + 0.74 \left(\frac{w}{w + 1.59s} \right)^{1.14} + 1.16 \left(\frac{w}{w + 1.87s} \right)^{0.16} \cdot \left(\frac{h}{h + 0.98s} \right)^{1.18} \right] \quad (2.4)$$

The self-inductance is calculated by:

$$L_z = \frac{\mu_0 \cdot l}{2\pi} \left[\ln \left(\frac{2l}{w + t} \right) + \frac{1}{2} + \frac{0.22(w + t)}{l} \right] \quad (2.5)$$

The calculated results are:

Table 2.2 Calculated Wire Parameters for 65nm, 45nm and 32nm

Technology	R (Ohm/mm)	Cground (fF/mm)	Ccoupling (fF/mm)	Ctotal(fF/mm)	L(nH/mm)
65nm	40.7404	82.031	73.222	228.475	1.7032
45nm	69.84	78.01	86.02	250.01	1.74
32nm	110.85	112.63	96.87	306.38	1.78

2.1.2 Device Model

To keep it relatively simple while still show the accuracy, a BSIM3 MOSFET model from Predictive Technology Model (PTM) is used as model card for SPICE simulations.

BSIM3 and BSIM4 models that are developed by University of California Berkeley are among the most popular SPICE compatible device model cards.

PTM BSIM 4 is modeled based on several facts. It is assumed that device design and process technologies throughout the semiconductor industry are similar for a certain technology node. They treat several parameters such as L_{eff} , T_{ox} , V_t and R_{dsw} as process variables rather than design variables (e.g. L_{gate} and V_{dd}). This gives the advantage for designers to have a degree of abstraction. Additionally, BSIM3 gives the circuit designers transparency in the parameter dependency. If T_{ox} is changed, the on state current I_{on} , leakage current I_{off} , etc also change accordingly [11]. Early work conducted by Shockley is far from accuracy. Sakurai introduced the n^{th} power law [12] that closes the gap between simplicity and accuracy. As discussed in [13], device prediction is not a simple geometry scaling which will be too simple to capture the

basic MOSFET behavior. In order to maintain the accuracy, BSIM models require over 100 parameters to model the device characteristics, while still keeping the merit of simplicity to use since it is relative easy and straight-forward for the user to change the parameters. Predictive Technology model (PTM) MOSFET models are used as HSPICE simulation models for all technology on level 54.

When we assume that the body is connected to the source node i.e. $V_{BS}=0$, the basic device parameters extracted from BSIM 4 can be represented as:

$$V_{th} = V_{T0} \quad (2.6)$$

$$V_{DSAT} = K(V_{GS} - V_{TH})^m \quad (2.7)$$

$$I_{DSAT} = \frac{W}{L_{EFF}} B(V_{GS} - V_{TH})^n \quad (2.8)$$

For saturation region ($V_{DS} \geq V_{DSAT}$):

$$I_D = I_{DSAT}(1 + \lambda V_{DS}), \quad \lambda = \lambda_0 \quad (2.9)$$

For linear region ($V_{DS} < V_{DSAT}$):

$$I_D = I_{DSAT}(1 + \lambda V_{DS}) \left(2 - \frac{V_{DS}}{V_{DSAT}}\right) \frac{V_{DS}}{V_{DSAT}} \quad (2.10)$$

where V_{GS} , V_{DS} , and V_{BS} are gate-source, drain-source, and body-source voltage, respectively. W is the channel width and L_{EFF} is the effective channel length. V_{TH} is the threshold voltage, V_{DSAT} is the drain saturation voltage, and I_{DSAT} is the drain saturation current. V_{T0} is a parameter which describes the threshold voltage. Parameters K and m control the linear region characteristics, and B and n control the saturation region characteristics.

The output resistance of the inverter and differential sensing circuit which will be discussed in section 2.3 is derived from these theoretical equations.

2.2 Circuit Simulation

2.2.1 Repeater Circuit

This section discusses the circuit simulated in this thesis. By using the wire and device models discussed in section 2.1 and 2.2, a repeater insertion line shown in figure 2.2 has been set up in HSPICE.

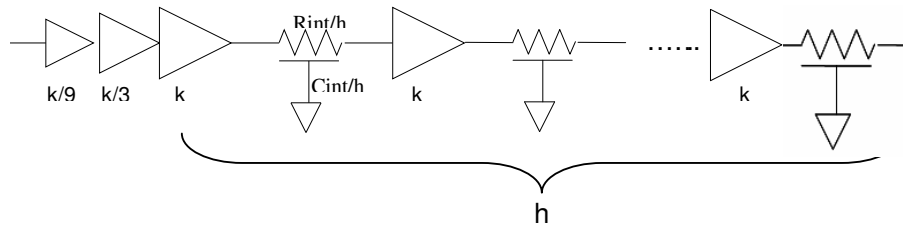


Figure 2.3 Repeater Insertion Line

Where h represents the repeater number along the wire, and k represents the size of the repeaters. If one interconnect with resistance R and capacitance C is divided into h segments, the resistance and capacitance will be R_{int}/h and C_{int}/h respectively. Each wire segment will have a 5-pi distribution model as discussed in section 2.1.

The output from a logic block is usually driven by a small or minimum logic drives, while the repeater size in the interconnect is much bigger. Also if the input needs to drive a long wire, it means the input needs to afford a huge load capacitance. Thus, it is not possible to have a sharp slope without buffers cascading at the beginning. These cascaded buffers do add an extra cost in the interconnect but it is required to drive the interconnect properly and efficiently. In [14], the design strategy of successive buffers has been discussed. It shows that the number of cascaded stages can be decided by the $\log h$, where h is the repeater size in the interconnect.

$$e^n = h \Rightarrow n = \log(h), \text{ where } \log e=1 \quad (2.11)$$

The total delay through the cascaded buffer will be the sum of delay in each buffer that constitutes the cascade. And the total delay through the wire will be the sum of the cascade buffer and each wire segment. Hence the 50% delay from then input to the output can be expressed as:

$$T_{50\%} = 0.7 \log(h) e R_0 C_0 + \frac{0.7 R_0 C_{\text{int}}}{h} + 0.7 k R_0 C_0 + \frac{0.4 R_{\text{int}} C_{\text{int}}}{k} + 0.7 h R_{\text{int}} C_0 \quad (2.12)$$

The optimal repeater size will always be around 300 of the minimum repeater size for all wire lengths as simulated in HSPICE. The approximate number of cascade stages is 2. We put two cascaded buffers to boost the input. The first-stage buffer is of size $k/9$ and the second stage buffer is of size $k/3$. It gives a steep rising and failing edge which is more close to reality.

2.2.2 Repeater Insertion Optimization

Among the different repeater insertion methods, Bakoglu's method is among the most basic and well-known.

In[14], the author presented a methodology for inserting repeaters in a long rc interconnect to break the quadratic delay dependency on the interconnect length. The conclusion was that the delay of a repeater should be equal to the delay of a wire segment in order to optimally drive the interconnect. Thus, the optimal number and size of repeater in a certain wire length can be derived. The relationship can be represented as:

$$k = \sqrt{\frac{0.4 R_{\text{int}} C_{\text{int}}}{0.7 R_0 C_0}} \quad (2.14)$$

$$h = \sqrt{\frac{R_0 C_{\text{int}}}{R_{\text{int}} C_0}} \quad (2.15)$$

Where,

k = number of repeaters in the repeater line

h = size of repeater

R_{int} = total resistance of the interconnect

C_{int} = total capacitance of the interconnect

R_0 = output resistance of a minimum size repeater

C_0 = output capacitance of a minimum size repeater

According to [14], the accurate size and number in a cascaded repeater interconnect can be re-written as:

$$k = \sqrt{\frac{0.4 R_{\text{int}} C_{\text{int}}}{0.7 R_0 C_0}} \quad (2.16)$$

$$h = \frac{\sqrt{4 R_{\text{int}} C_{\text{int}} R_0 C_0 + e^2 R_0^2 C_0^2} - e R_0 C_0}{2 R_{\text{int}} C_0} \quad (2.17)$$

Bakoglu's method sets a general boundary for interconnect circuit design. However, it is less accurate in the nanometer regime. The actual repeater size simulated in HSPICE is different from the theoretical results. In [8], three theoretical insertion methods have been compared. It gives the boundary of number and size repeater in an interconnect as:

$$100 < k \text{ (size of repeaters)} < 300$$

$$3 < h \text{ (number of repeaters)} < 9$$

We use a practical repeater insertion method which is simulation-based. By varying the size and number of repeaters in a certain wire length, the corresponding delay is

recorded. If the size and number of the repeaters achieving the optimal delay falls into the allowable range, it will be chosen as the optimal setting for that wire length.

2.2.3 Differential Current Sensing

The logic is presented by voltage levels referenced to the power supply voltage in the conventional VLSI digital design. The simplified representation of a voltage mode circuit can be found in [16]. Alternatively, the logic value could also be represented via current signals, since the voltage mode does not always have the best performance on delay, power and other design considerations such as reliability. A current sensing circuit allows the voltage at the output to change based on the input current, rather than based on the voltage level [16]. The difference between voltage mode and current mode circuits can be illustrated in figure 2.3 (a) and (b)

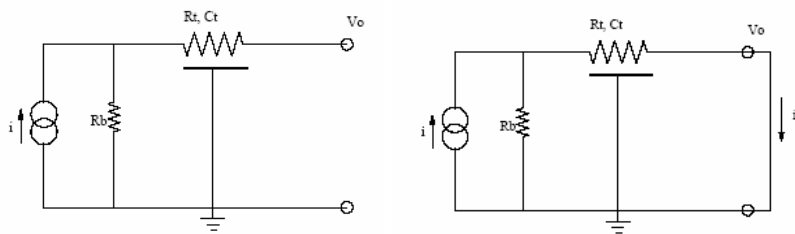


Figure 2.4 (a) left: Voltage Mode Configuration

Figure 2.4(b) right: Current Mode Configuration

Instead of the open end in the voltage mode, the current mode circuit has a shorted end. In the case of current-mode, the termination resistance is very small but in the

case of voltage-mode it is very large. Instead of sensing voltage, current is used as a mode of signaling in the current-mode. Ideally, there should be a path to ground from the driver in the interconnect application as shown in figure 2.4.

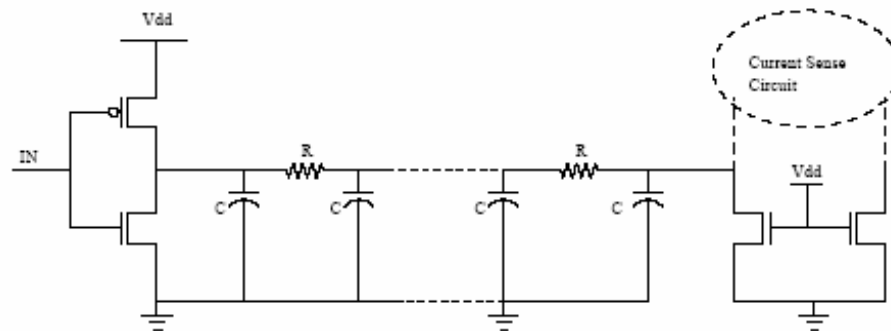


Figure 2.5 Current Sensing Circuit in interconnect

A current Sensing circuit is more complex than a voltage sensing circuit due to several reasons. The MOS transistors do not have a current threshold which means the current mode circuit has to set a current threshold for sensing. The capacitance of the interconnect is not charged to V_{dd} but to an intermediate value due to the low impedance path to the ground on the receiver side. In differential current sensing, a synchronizing signal is required to keep the synchronization between two inputs and two outputs.

In interconnects, current sensing circuits can minimize delay by reducing the terminating resistance [16]. Since CMOS devices are essentially voltage controlled devices with a threshold voltage but without a threshold current, the central part of current sensing circuits is focused on the sensor/amplifier design.

Several previous works have explored the differential current sensing amplifier in memory [17, 18, 19], FPGA crossbars [20] and interconnect [8].

Seevinck proposed a current-mode sense amplifier for an SRAM. It consists of four equal-sized PMOS transistors. The delay in the sensor amplifier is independent of bit line capacitance since the large capacitances of bit-lines have been clamped. It required the bit lines load to be low ohmic and some biasing voltages which is too complicated for interconnect signaling, so the application is limited to memory design.

Blalock and Jaeger developed a sense amplifier called Clamped Bit Line Sense Amplifier (CBLSA) for DRAM memory. It has 6 transistors: 2 PMOS and 2 NMOS transistors form a cross-coupled latch on the top, and 2 NMOS transistors form a low impedance path biased in the linear region. CBLSA employs the same mechanism to clamp the bit lines from swinging. The output of this amplifier is voltage so no extra stage of conversion is needed. This circuit also has limitations for interconnect application since it requires special biasing, thus involving a lot of precharging and sensing related synchronizing signals.

Another approach to the current sensing circuit is the single-ended sensor amplifier [19] such as Izumikawa and Yamashina's amplifier for multi-port SRAM and Shinha's sensor amplifier for FPGA crossbar.

In [4, 15], the authors showed that single-ended sensor amplifiers can work properly without an external signal for the functionality of the circuit and hence no routing overhead or generation of a timing pulse. A select signal can shut-off both the sensor circuit and the amplifier so it saves a lot of static power. Despite the advantages of

single-ended current sensing, it encounters several natural deficiencies for potential interconnect application. Process-related variations and coupling noise will be the two biggest concerns and hence the performance and reliability will be degraded. This is extremely important for interconnects due to the nature of wires. It is very common to have several wires in parallel and hence coupling noise is the least desirable aspects we will want to see in the interconnect. As global wires are more distributed than memory, process variation will also be a problem for single-ended current sensing circuit.

Figure 2.5 shows a differential current sensing circuit for interconnect proposed by Atul et al [8]. It simplified the complex biasing and synchronizing signals in Seevinck and Blalock's work while it also has less effect of orthogonal coupling and mutual inductance over a single-ended sensing circuit.

DCSA works very much like Blalock's amplifier. Initially the EQ signal is asserted thus equalizing the two outputs OUT and OUTBAR. The current flowing through the two paths is almost the same. The IN and INBAR are driven by a driver and due to the low impedance to ground a differential current develops and hence the current in one of the paths is more than in the other one. When EQ turns off M3, the cross-coupled latch (M1-M4) switches thus giving a voltage output determined by the differential current between IN and INBAR.

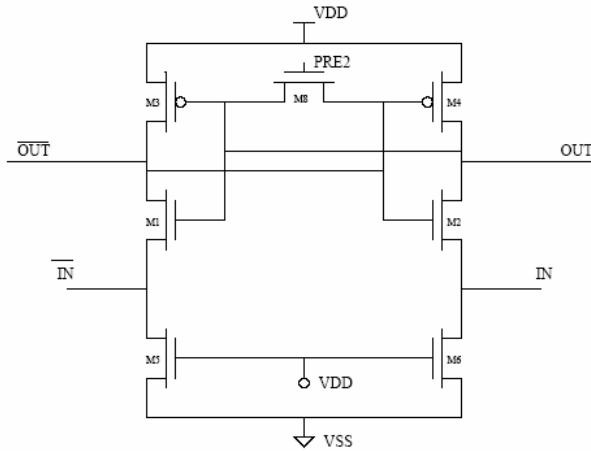


Figure 2.6 Maheshwari's Differential Current Sensing Amplifier

DCSA replaces two sensing signals in the original Blalock's CBLSA circuit with VDD and GND. Thus it makes the signaling much simpler than the original DRAM sensor amplifier. This is feasible since global interconnect doesn't need complex precharging and pre-equalizing signals as memory does. Interconnects only need a straightforward signal transfer while memory involves more functions such as read and write. It is also necessary to simplify the signal in CBLSA for interconnect since it is too expensive and not realistic to have so many equalization and synchronizing signals throughout the whole interconnect network. In chapter 3, we will discuss more about the existing problems of DCSA and propose improved solutions. In order to model the differential current sensing circuit, a circuit as shown in figure 2.6 has been set up in HSPICE for simulation. The drivers of the DCS circuit are two buffers that send the complementary signal to the receiver. The receiver, DCSA, works on the input signal and amplifies it to a pair of full swing voltage outputs. Since logic devices

are usually small, a minimum size of device load will be considered at the output as a logic block. The same condition will be applied on repeater insertion line for fair comparison.

The two NMOS that forms a low impedance at receiver are the same size as the driver in order to realize the output voltage match. The cross-coupled latch are sized accordingly as well. The size may increase as the wire length increases.

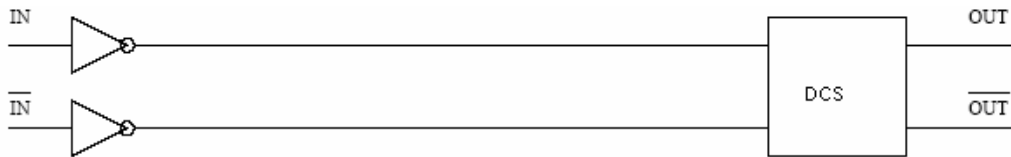


Figure 2.7 Differential Current Sensing Circuit

The simulated waveform of DCS has been shown in figure 2.7 the upper two signals are the input current measured on the driver side. The full swing equalization voltage has been overlapped with two current outputs in the middle. The bottom signals show the full swing voltage output.

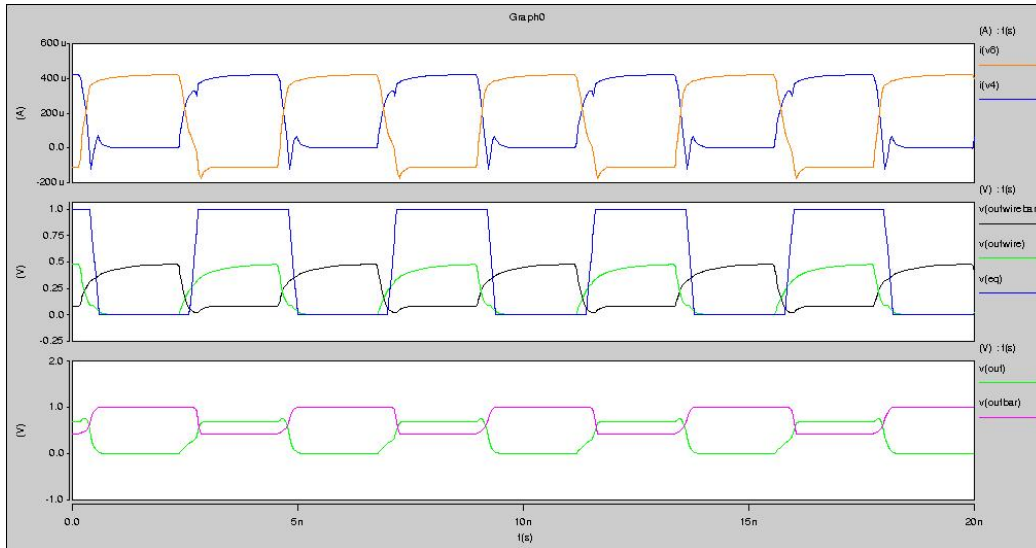


Figure 2.8 Simulated waveform of DCS

2.3 Experiment Setup

In order to ease the simulation automation and retrieve data, a PERL script has been employed to generate HSPICE script, initiate the simulation and collect the results from a results file.

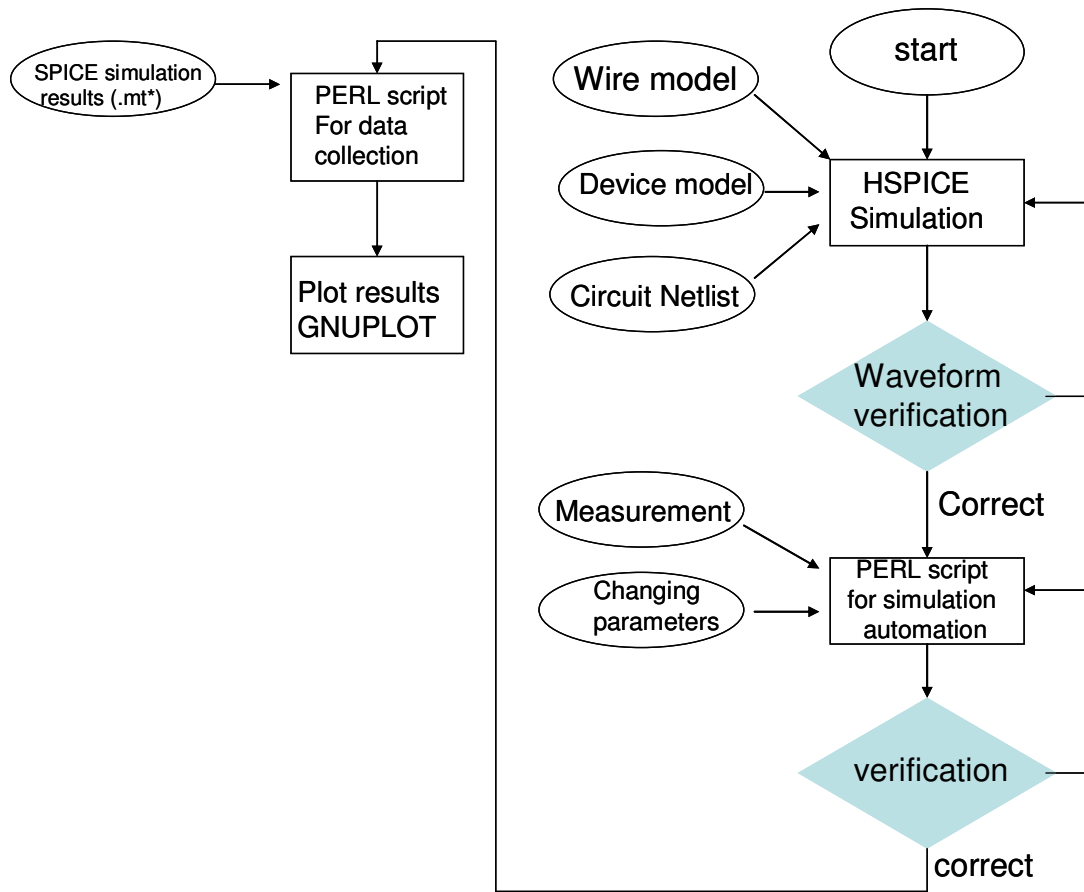


Figure 2.9 Experimental setup and flow

CHAPTER 3

ENERGY-AWARE DIFFERENTIAL CURRENT SENSING

An Energy-aware Differential Current Sensing Amplifier (DLASA) has been proposed in this chapter. The energy-saving method that utilizes power gating technique has been explained in section 3.1. In section 3.2 and 3.3, DLASA simulation setup and comparison method have been discussed. Results have been presented in section 3.4. In 3.4, a first order comparison between DLASA, DCS and repeated line has been explored first. Then, several secondary aspects such as activity factor, wire length, driver size, technology scaling and area efficiency has also been addressed. A summary is drawn in 3.5.

3. 1 Energy-aware Differential Current Sensing

Some preliminary simulation results of using DCS and repeater lines are shown in Fig. 3.1. The differential current sensing circuit demonstrates a more efficient signal transmission than a repeated line. DCS has less delay than the repeater insertion method for an interconnect longer than 2 mm. The corresponding energy at the

optimal delay for each wire length is also shown in Fig. 3.1. Since static and leakage current are dominant in DCS, DCS may consume more energy than the repeater method does. To be specific, DCS consumes more energy than a repeated line on interconnects from 1 mm to 6 mm. In general, interconnect circuits need more energy to drive as the wire gets longer. However, it is noted that DCS consumes more energy as the wires shorten from 4 mm to 1 mm by using a large driver size. This is due to the inherent design of DCS circuits as shown in [8].

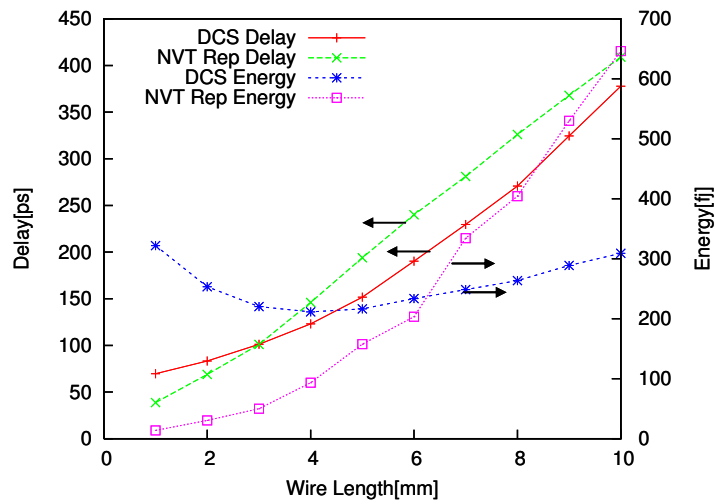


Figure 3.1 Delay and Energy comparison between DCS and repeater

Three major sources of power dissipation in the original DCS circuit are given by Equation 3.1:

$$E_{DCS} = E_{dynamic} + E_{static} + E_{leakage} \quad (3.1)$$

Energy consumption is associated with current. There are several ways to define static and leakage currents. To be clear and consistent, static current flows through the direct path from Vdd to ground in Fig. 3.2. In other words, static current is the current

when a transistor is on without signal transition. In [21], Roy et al. discussed six sources of leakage currents. They justified that four sources of leakage currents occurs in off-state, except pn junction Reverse-Bias current and Narrow-Width effect that occur in both ON and OFF states. Also, the off-state leakage currents consist most part of the leakage current amount. Since the most leakage current is in off state, we consider only off-state leakage for simplicity.

The proposed energy-aware Differential Leakage-Aware Sense Amplifier (DLASA) to replace the DCS receiver is proposed in Fig. 3.2. The DLASA circuit requires the same differential input signals IN and INBAR as in DCS. It consists of a pair of low impedance terminations (M5 and M6 in Fig. 3.2) and a cross-coupled latch (M1, M2, M3, M4 in Fig. 3.2). The latch is controlled by an equalization signal (EQ) through a NMOS transistor (M7 in Fig. 3.2). M8 and M9 are sized according to the low impedance path and cross coupled latch transistor sizes. Synchronizing signal SE controls the two transistors and the low impedance path.

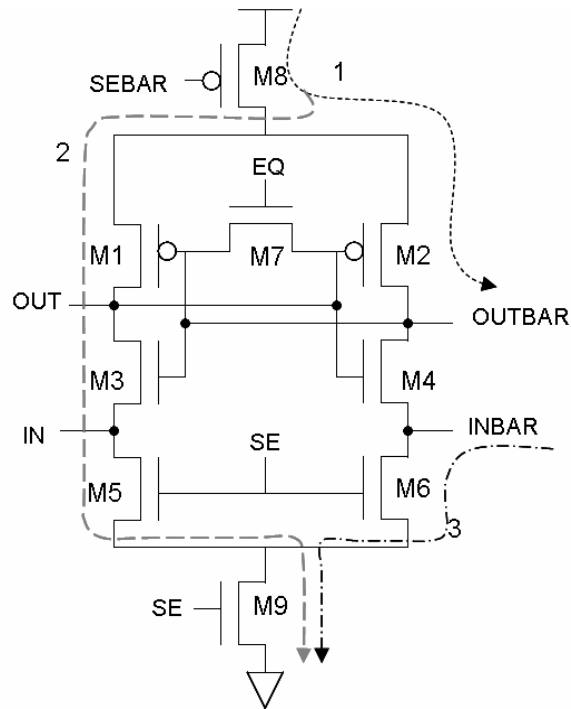


Figure 3.2 Differential Leakage-Aware Sense Amplifier (DLASA)

During the equalization phase, M7 is turned on. The M1, M4 and M2, M3 pairs work in linear region and cutoff region, respectively. This metastable state is broken in the evaluation phase after M7 is turned off. Finally, two pairs of inverters then operate either in saturated or cutoff regions in a stable state and the output (OUT and OUTBAR) is formed.

The signal waveform is illustrated in Fig. 3.3. DLASA has only less than 1/3 of the input current (i.e., 1.0 mA) of DCS. The reduced input current has clear ramification in reducing energy consumption. DLASA reduces around 2.0 mA input current by stopping the current sources in three ways. During the equalization phase, EQ is off and Sense Enable(SE) signal is on and the circuit will work in the same manner as the

original DCS. During the evaluation phase, after the output differential current is formed, SEBAR turns off M8 to prevent the static current through M1 and M2 as shown in path 1 in Fig. 3.2. In the evaluation phase, the four transistors in a cross-coupled latch are in either saturated or cutoff region. M9 prevents the direct path from the cross-coupled latch to the ground as shown in path 2 in Fig. 3.2. Low impedance path M5 and M6 are in linear region, therefore M9 also prevents static current from going through these two transistors as shown in path 3 in Fig. 3.2.

Power gating effectively saves energy in current sensing circuits, but it is not feasible in repeated lines as repeaters are distributed along the wire. Due to the transmission latency, each repeater sequentially experiences the same signal. Hence, the power gating signal needs a complicated timing to control the repeaters accurately. Furthermore, routing area for the separated control signals is another problem in applying power gating in repeater line.

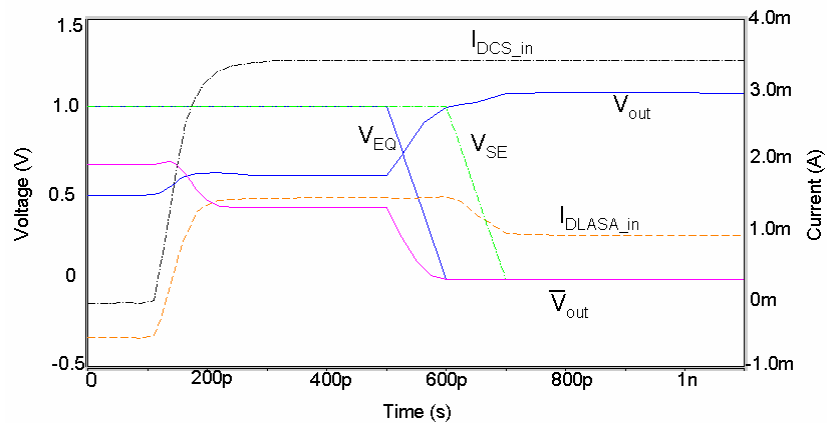


Figure 3.3 Simulation waveforms of DCS and DLASA

3.2 Experimental Setup

Table 3.1 shows the device and interconnect parameters that are used throughout this study. Wirelengths from 1mm to 10mm were used. 65nm Technology models were obtained from PTM [10]. Wire parasites for the dimensions given in Table 3.1 were also from PTM [10]. Global interconnects are considered to be shielded between supply and ground lines. Interconnects are modeled as a 5-pi distributed RC network.

Table 3.1 Interconnect And device parameters

Technology	Interconnect			Device	
	Dimensions(um)	R(Ω /mm)	C(fF/mm)	Threshold Voltage (V)	
				NMOS	PMOS
65nm	W = 4.5 μ m, S = 4.5 μ m T = 1.2 μ m, H = 0.2 μ m	40.7	Cg = 82.03 Cc = 73.22	HVT = 0.22 NVT = 0.19	HVT=-0.23 NVT=-0.21
45nm	W = 315 nm, S = 315 nm T=100 nm,H = 150 nm	69.84	Cg = 78.01 Cc = 80.02	HVT=0.26 NVT=0.0.24	HVT=-0.23 NVT=-0.21
32nm	W = 220.5 nm, S = 220.5 nm T =0.9 um, H = 60 nm	110.85	Cg = 112.63 Cc = 96.87	HVT=0.26 NVT=0.24	HVT=-0.22 NVT=-0.21

3.3. Repeater Optimization

There are several analytical repeater insertion methods that have been well explored in [22] and [23]. However, analytical optimal sizes and number of repeaters may not result in minimum delay since the analytical models do not consider every design aspect. Simulation provides the most accurate repeater optimization results. Fig. 3.4 shows the setup that was used to optimize repeaters. As was discussed in chapter 2,

two cascaded buffers are used to provide inputs to the repeater chain in order to mimic realistic input signals for repeater line. Repeater size is varied from 54 to 350 times of minimum size and the total number of repeaters is varied from 1 to 11.

Low leakage HVT repeaters were also considered and optimized along with nominal V_t (NVT) repeaters using the same methodology. Several methods in material or process technology could lower leakage current in CMOS devices, such as high-K gate materials, dual gate structures and SOI (Silicon on Insulator). But none of these strategies are easy to realize. In addition, these methods have a lot of side effects. For example, changing doping concentration to control threshold voltage can result in lower subthreshold leakage. Meanwhile, high V_t will slow down the device. Several efforts on high threshold voltage circuits design to achieve lower leakage power have been proposed in [21] [24]. Here, we set all repeaters to a 15% higher threshold voltage to get the optimal leakage power in a high V_t repeater circuit. The threshold value of HVT repeaters are 15% higher than NVT

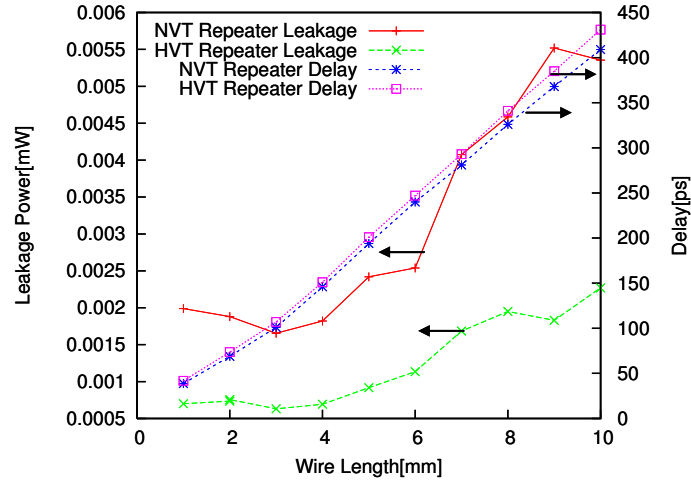


Figure 3.4 Delay and leakage power for HVT and NVT repeaters

Fig. 3.4 shows the leakage power and delay for HVT and NVT repeaters for different wirelengths. It can be seen from the plot that there is a maximum of 22ps delay difference between HVT and NVT repeaters for a 10 mm wire. It can also be seen from the plot, as expected, that HVT has significantly lower leakage power than NVT. For a 10mm wire, leakage power in HVT is 34% lower than NVT. In short, Fig. 5 clearly shows that the HVT repeaters have a clear advantage in lowering leakage while incurring a penalty in delay for longer wires. This is due to the fact that there are more repeaters for longer wires and hence the delay through each repeater adds up and results in a longer delay.

3.4. Results and Discussion

3.4.1 First Order Comparison

In order to provide a worst case comparison, HVT repeaters are compared with DLASA for leakage and NVT repeaters are compared with DLASA for speed.

Fig. 3.5 shows the delay and energy from 1mm to 10mm wire using DLASA, HVT repeaters and NVT repeaters. It can be seen from the figure that DLASA is faster than NVT in 4 mm and longer wire. It can be deduced from Fig. 3.4 and Fig. 3.5 that DLASA retains the performance advantage that DCS offers while reducing power compared to NVT and HVT repeaters. DLASA improves delay for interconnects longer than 4mm by a maximum of 18% as compared to NVT repeaters.

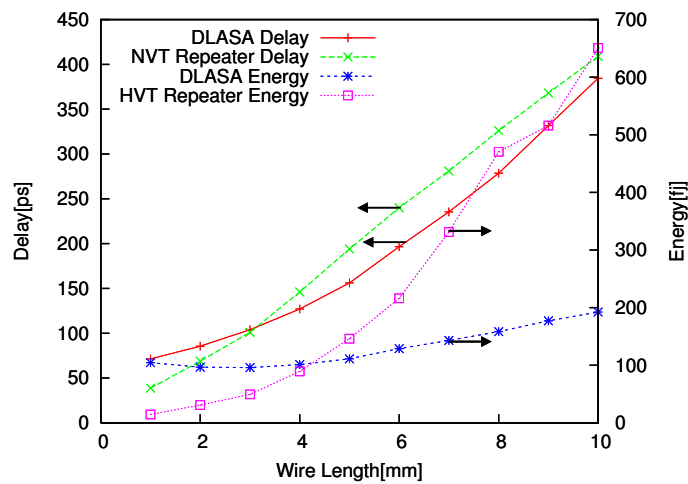


Figure 3.5 Delay and energy of repeaters and DLASA

3.4.2 Activity Factor Impact

Fig. 3.6 shows the impact of activity factor on total energy consumption for DCS, DLASA and NVT repeaters for a 5mm wire. As activity factor increases, total energy for a NVT repeater increases due to an increase in dynamic power. It can also be seen from Fig. 6 that for DCS and DLASA energy is constant across varying activity factors. The reason for non-varying energy is due to the fact that for both DCS and DLASA, static power dominates total power. Fig. 3.6 also shows that DLASA is more energy efficient than DCS by 59% due to the shut-off system that shuts off the static power after sensing. When compared with NVT repeaters, DLASA performs better for activity factor greater than 45%. This shows that DLASA is suitable for high activity buses. DLASA decreases energy consumption to less than one third of that of DCS. It remedies the current sensing technique on energy saving so that the current sensing circuit wins over repeater lines on 55% activity factor and greater. A better result of DLASA is expected in future technology, where leakage power continues to be more significant, and the DLASA circuit will save more energy.

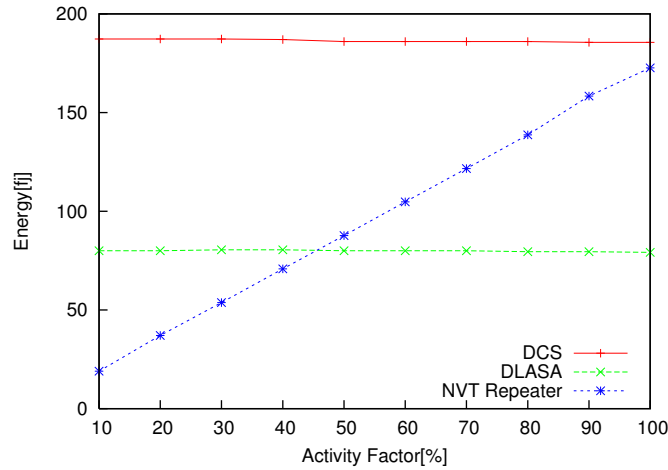


Figure 3.6 Energy Comparison under different Activity Factors

3.4.3 Wire Length Impact

Fig. 3.7 shows the impact of wirelength on leakage power for DLASA, NVT and HVT repeaters. Leakage power increases for both HVT and NVT repeaters with increasing wirelength due to the increase in the number of repeaters. As expected, leakage for HVT repeaters are lower than for NVT repeaters. It can be seen from Fig. 3.7 that DLASA leakage is lower than that of HVT repeaters for all wirelength. This is due to the fact that DLASA requires less area than repeaters and each interconnect requires only one driver and receiver. Overall the maximum reduction in leakage power by DLASA over HVT repeaters is 82% at 10mm. Similarly maximum reduction in leakage power by DLASA over NVT repeaters is 92% at 10mm.

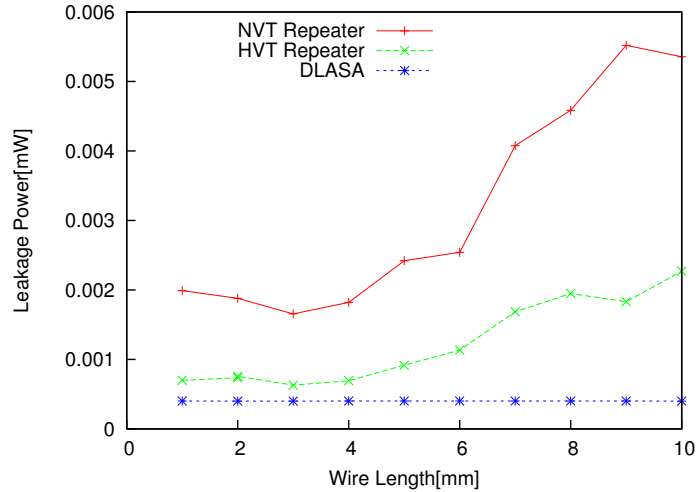


Figure 3.7 Leakage power of HVT, NVT repeaters and DLASA

Further analysis on static and leakage current reduction can be seen in figure 3.8. These comparisons include both static and leakage power, because the original DCS circuit lacks of mechanism to turn off the receiver during the off-state, which turns off the static current. Thus, it is unrealistic to separate leakage and static current in the original differential current sensing circuit. Energy-aware DCS also cuts down the static power by turning off the two switch transistors. Figure 3.8 shows that the reduction in power in DLASA over DCS is more obvious in short wires than long wires.

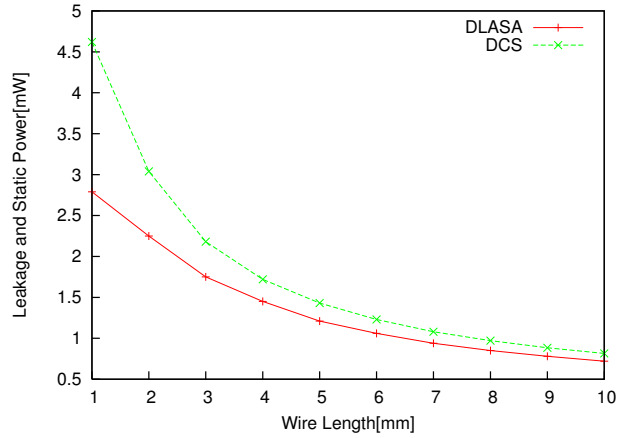


Figure 3.8 Static and leakage power in DCS and Energy-aware DCS from 1mm to 10mm

Energy saving on DLASA over DCS can be seen in all wire lengths. Figure 3.9 also shows that the DLASA has lower energy consumption and less propagation delay over both HVT and NVT repeated line when the interconnect is longer than 5 mm.

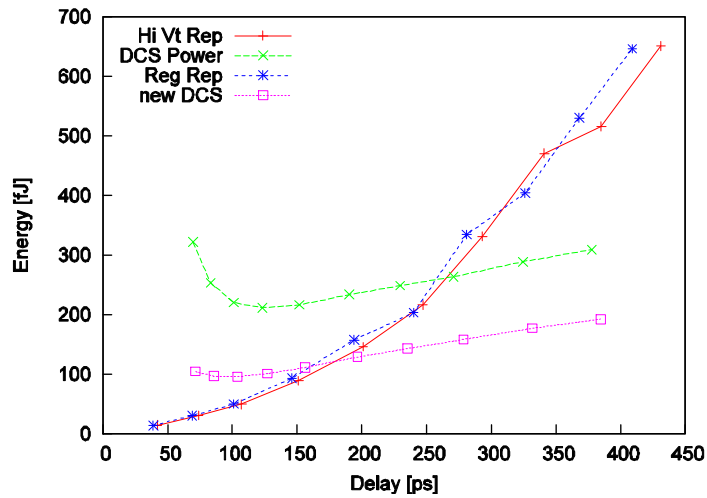


Figure 3.9 1mm-10mm wire Energy Versus Delay

3.4.4 Driver Size Impact

Three circuits are also simulated for leakage power for different driver size on a 5mm wire in figure 3.8. The leakage improvements of Energy-aware DCS over high Vt repeaters are seen in all sizes.

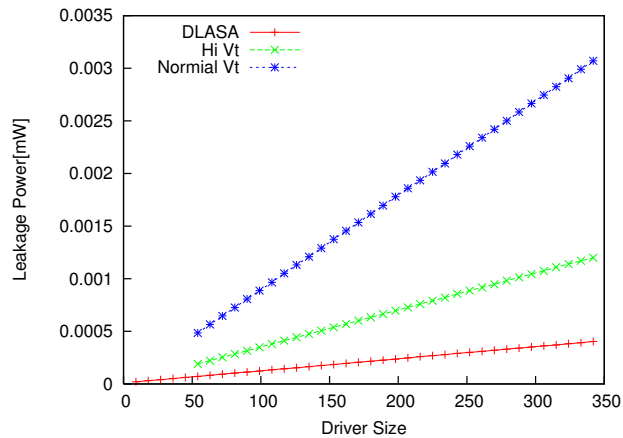


Figure 3.10 Leakage power of high Vt repeater, normal Vt repeater, and Energy-aware DCS varying driver size on 5mm wire

Figure 3.11 shows the energy delay plot for DLASA HVT and NVT repeaters for a 5mm wire. DLASA, HVT and NVT repeater sizes are varied in order to show the different energy delay optimization corners. At NVT repeater lowest delay DLASA provides an energy savings of 42%. At HVT repeater lowest energy DLASA provides a delay savings of 33%.

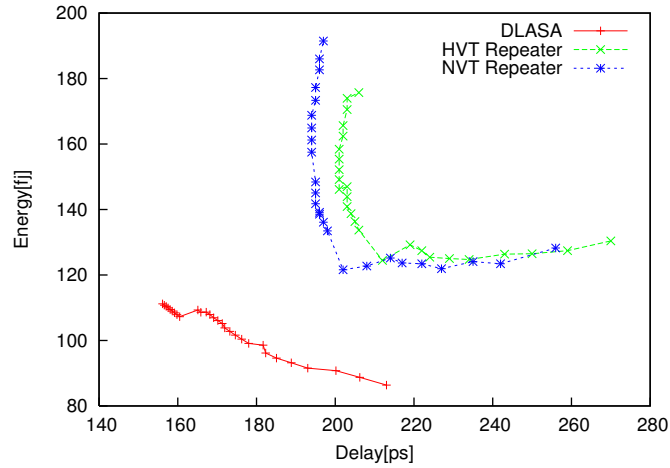


Figure 3.11 5 mm wire Energy Versus Delay on Driver Size Varying

Figure 3.12 shows the static plus leakage power saving on Energy-aware DCS compared to the original DCS on a 5 mm wire by varying the driver size from 5 to 350. As we can see, the smaller size DLASA circuit tends to have greater reduction in power. The reduction over DCS shows in every driver size as design expected. The reduction is 11% in the worst case.

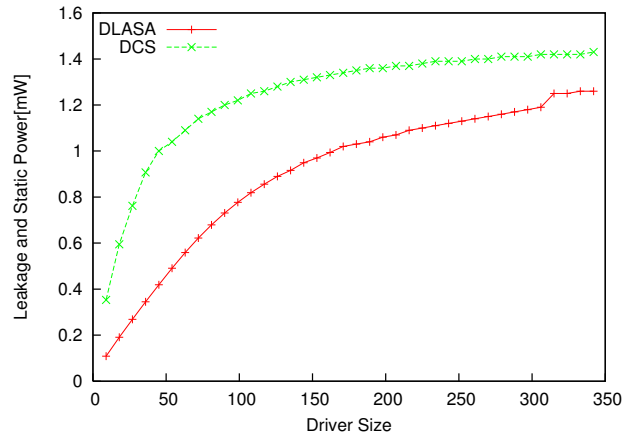


Figure 3.12 5mm wire Static+leakage power in DCS and Energy-aware DCS on varying driver sizes

3.4.5 Technology Scaling Impact

This section discusses the scaling impact on the DCS and DLASA circuits. DLASA still shows the advantage in terms of energy saving in all technologies. For a 3mm wire, the percentage of energy saving on DLASA is 81% comparing to DCS. It is clear that DLASA saves energy in lower technology and would be a better choice in the low power design. Meanwhile, propagation delay for DLASA and DCS in 32nm are very close to each other. For shorter wire, the difference is less than 2ns. And the difference increases as the wire gets longer. The worst case difference is 4.8ns on a 5mm wire.

In 45nm and 32nm technology, DLASA remain its merit in energy saving and has less delay penalty on propagation delay. This clearly shows that DLASA would be a better choice is lower technology VLSI design.

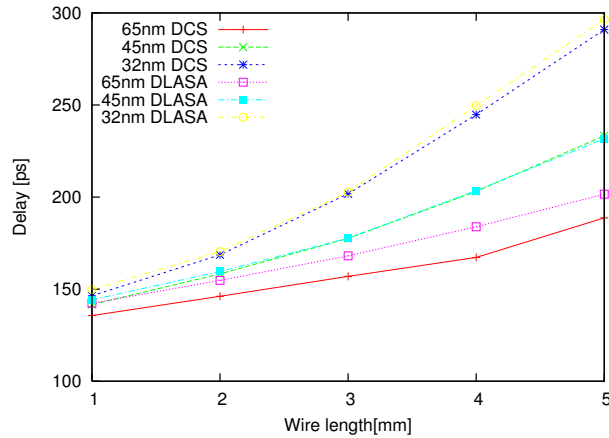


Figure 3.13 Technology scaling impact on DCS and DLASA respect to propagation

delay from 1mm to 5mm

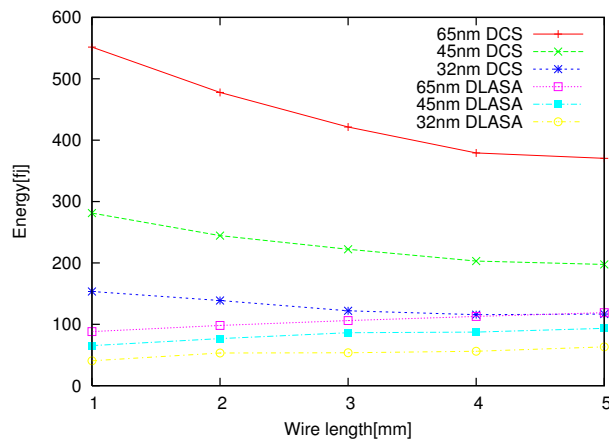


Figure 3.14 Technology scaling impact on DCS and DLASA respect to propagation

energy from 1mm to 5mm

Figure 3.15 shows the technology impact on DCS and DLASA on a 5 mm wire

Energy and delay are plotting with different driver size. DLASA has less energy consumption in all technologies comparing to DCS. In 32nm, the propagation of DLASA are very close to DCS while the energy saving are seen in all driver sizes. It

means that as DLASA can be used with different driver in 65nm to achieve other design goal except delay constraint, DLASA can continue to be used in lower technologies for similar design consideration.

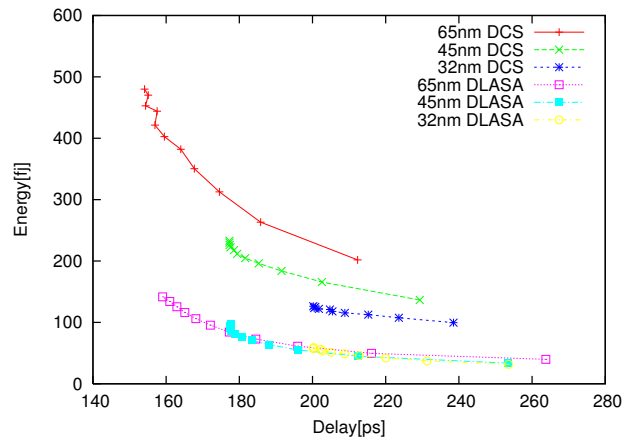


Figure 3.15 Technology impact on DCS and DLASA, 5 mm wire Energy Versus Delay on Driver Size Varying

3.4.6 Signaling Complexity and Area Efficiency

In Figure 3.17, circuit active area is compared among the designs based on sizes that result in optimal delay. Both DCS and DLASA are normalized to repeaters according to device width. It shows that, except in 1 mm wire, DLASA has a smaller area on wire length over DCS. Furthermore, the ratio of DLASA area to the normalized repeater insertion line area are always less than 1, which means that the total area of DLASA is always the smallest among the three circuit. As the wire gets longer, the area efficiency of DLASA is more improved compared to repeater. This is expected

since the size and number of transistors does not change much with wire length, while the repeater line will need more transistors along the wire to maintain the performance.

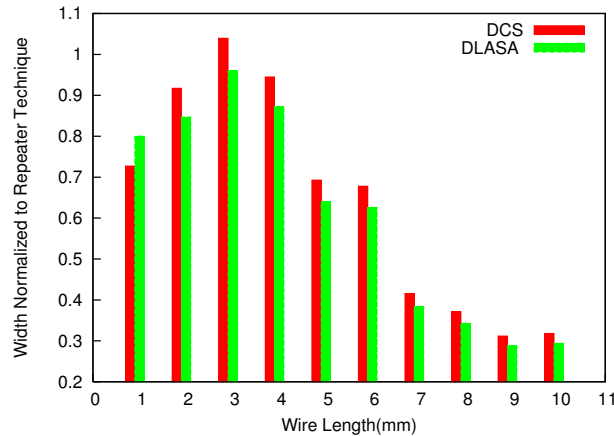


Figure 3.17 Circuit area comparison among DCS, DLASA and repeater

3.5 Conclusions

This chapter proposes a novel energy-aware differential sensing system for on-chip interconnects. A power gating technique is discussed and analyzed to reduce static and leakage power. Simulation results show that DLASA effectively reduces static and leakage power up to 39.6% compared to conventional DCS. This current sensing technique does not require complicated control signals and huge routing area for power gating, so power gating technique is feasible. The control signals that are required are locally derived from the clock. Simulation results also show that this energy-aware differential current sensing technique could be applied under various design considerations besides delay and power optimization.

Nominal V_t and High V_t repeaters were simulated and compared with the proposed system. Due to the nature of repeaters, it is impractical to apply power gating technique to reduce leakage power. Simulation results show that DLASA provides an energy savings of 42% at NVT repeater lowest delay and 33% delay savings at HVT repeater lowest energy. For a 5mm wire DLASA is 18% faster and than NVT repeaters and reduces leakage power by 58.1% compared to HVT repeaters. Though differential current sensing techniques use two input signals which consume more channel routing area than repeaters, the one driver-one receiver circuit saves 49.5% active area on average compared to repeaters. Since the application of current-sensing circuits is not limited in interconnect, the power gating technique and DLASA is expected to be applicable in other circuits such as memory sensing logic design.

Technology scaling impact need to be considered in the continuing scale shrinking design trend. It has been shown that DLASA has less delay penalty to DCS while it still keeps the advantage in terms of energy saving in lower technology. DLASA is a better choice for lower power application in 45nm and 32nm technology.

Area efficiency and signaling complexity has also been discussed. It shows that DLASA are very competitive to DCS on area for all wire lengths longer than 1mm. DLASA dose need one more clock signal to be involved and hence increase the signaling complexity. But this additional clock can be resolved locally with careful timing closure. It is applicable once the size of the device is fixed.

CHAPTER 4

INTERCONNECT CIRCUITS UNDER THERMAL CHALLENGE

This chapter discusses the thermal impacts on interconnect. A review of thermal challenges in DSM circuit design has been discussed. In 4.2, temporal thermal variation and its impact on the interconnect will be presented. Spatial thermal variation and its impact will be discussed in 4.3. Theoretical delay model for repeated line under thermal variation has been discussed in 4.4. Summary can be found in 4.5

4. 1 Thermal Challenge in DSM Integrated Circuits:

Attention on semiconductor device temperature and various cooling techniques have been significantly increased as the technology goes further in deep-sub micron regime. And the uneven heat distribution across temporal and spatial domains has been more attention in contemporary processors than ever before.

The most well known result of heat damage is physical devastation. But it is far from the only result. Temperature fluctuation can cause timing error by changing the delay time. Signal integrity can vary because the temperature surging can induce noise. A

hot environment will cause more power consumption which becomes a positive feedback between temperature and power. Temperature limits the power delivery and dissipation which is the primary design concern in future high-end processors [22].

In [23], the authors discuss the potential circuit risk in an excessive heat environment.

Thermal affects need to be considered during the circuit design stage since it will affect circuit performance in various aspects, including:

1. Circuit Reliability
2. Propagation Delays and Signal Integrity
3. Power Dissipation
4. Power/Ground Integrity

Chips become hotter because of the speed mismatch between integration density increase and power density increase. Static thermal control becomes inefficient when the thermal surging is largely dependent on the computation pattern. Leakage power becomes dominant in chips at 65nm and below, which makes the thermal problem more complicated. It makes sections such as cache blocks which are usually dense and inactive become hot [24]. Self-heating is also a concern in bipolar transistor since it is sensitive to temperature varying and in SOI device because of its poor thermal conductivity. Multilevel interconnects, which is a key component in a VLSI dice, face a changing of temperature variation due to the increasing number of metal layers, higher thermal conductivity of Low-k dielectrics and thermal intervention due to the effects of via, substrate and package. It is hence very important to quantify the

performance sensitivity of different interconnect circuits under thermal variation by using a proper thermal model and an accurate simulation approach.

There are a number of existing thermal models for different parts of a microelectronic design. For example, previous work [25] [26] presented a dynamic compact thermal model, HotSpot, at the micro architecture level. [27] presented a chip-level thermal model based on full-chip layout. In [28], the authors presented a thermal modeling approach based on analytical solutions of heat transfer equations, and the model was mainly focused at device level. A methodology for deriving more or less ‘standardized’ compact models is presented in [29]. In [30], Huang et. al proposed a compact thermal model for temperature-aware design.

In [31], no uniform substrate impact on interconnect was analyzed. In [32], the authors investigated the thermal coupling effects between interconnects. The authors in [33] analyzed the temperature scaling of multilevel interconnect in high-performance ICs from 90 nm to 22 nm technology node.

4. 2 Temporal Temperature Variation on Interconnect

4.2.1 Impact on Wire Segment and Single Transistor

This section will discuss the impact of temperature variation on individual transistor and inverter. The interconnect is modeled as 5-pi RLC segments. Repeater insertion has been optimized for delay by simulating 1mm-5mm wires and varying the repeater sizes uniformly to obtain delay-optimal data. A 50ps slew rate constraint has been set in the selection process, such that only signals with a reasonable rise time are

considered. For the optimization, a cascade of 2 buffers drives the repeated line as we have discussed in chapter 2.

A simplified analysis on the effect of temperature variations on devices and interconnect is summarized by Table 4.1. From this table, one can see the general trend of the temperature impact on delay of each component of a repeated interconnect. An inverter with a lumped capacitive load of 1fF and a uniform temperature profile of 25°C, or ambient temperature, is studied first. Then, a temperature profile which assigns 125°C to the PMOS device of the inverter, while keeping the NMOS temperature at 25°C is applied. For this profile, there is no significant impact on delay observed across technologies. However, when the opposite profile is applied (125°C to the NMOS while keeping the PMOS at ambient temperature), a more significant impact is observed on delay for all technology nodes. This is expected since the output delay will depend more strongly on the NMOS device in this case, because there is just one buffer on the line. The NMOS will operate slower at such a high temperature, thus producing the negative impact on delay that is observed at this scenario. The next experiment consisted of observing the delay on a segment of repeated interconnect (i.e. a buffer followed by a 5-pi wire segment) for different temperature profiles. 125°C is applied to the device while keeping the wire at room temperature, and then the device and the buffer are simulated at a uniform temperature of 125°C. Temperature effects on the wire dominate the impact on delay in the smaller two technology nodes for this model.

Table 4.1 Temperature Variation Effects on Delay

Technology Node[nm]	65	45	32
Inverter @ 25°C[ps]	16.75	19.3	22.3
PMOS @ 125°C[ps]	16.78	19.49	22.43
NMOS @ 125°C[ps]	20.51	24.43	27.64
Device and wire @ 25°C[ps]	51.7	51.6	50.7
Device @ 125°C, Wire @ 25°C[ps]	52.6	71	66.7
Device and wire @ 125°C[ps]	55.6	107.3	100.8

4.2.2 Impact on Repeated Line

In uniform temperature profiles, the temperature is assumed to be constant along the length of the interconnect at a given time. A temporal thermal variation analysis has been conducted to characterize the impact of thermal variations in interconnects, in the presence of a uniform temperature profile. Figures 4.1 and 4.2 show the delay and energy variation, respectively, due to temporal temperature variation for a 3mm wire in 65nm, 45nm and 32nm technology nodes.

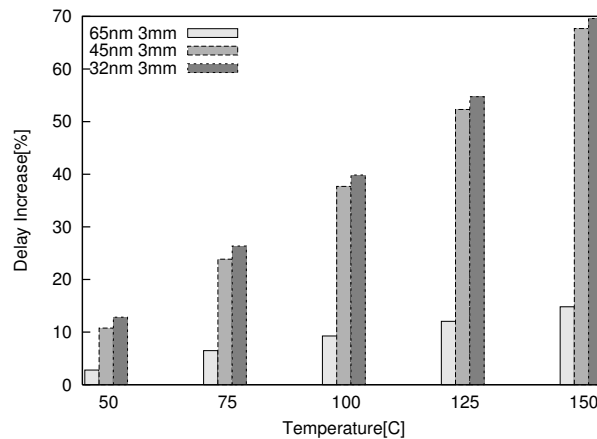


Figure 4.1 Percentage of delay increase for temporal thermal variation in 65nm,45nm and 32nm repeated interconnects.

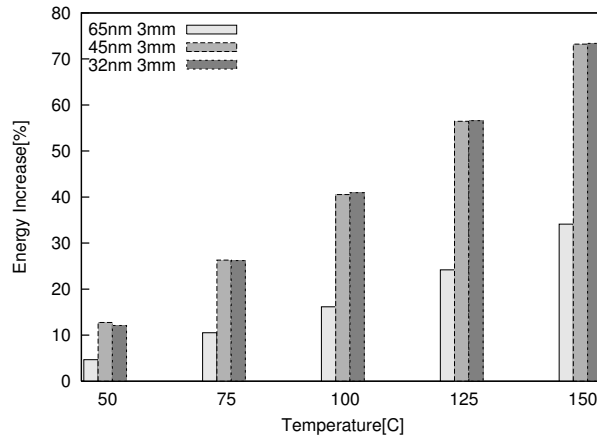


Figure 4.2 Percentage of energy increase for temporal thermal variation in 65nm,45nm and 32nm repeated interconnects.

Data is shown in terms of the percentage increase from the nominal case, which is the same interconnect at ambient temperature (25°C). Each temperature value noted on the x-axis corresponds to the uniform temperature the interconnect is subjected to at a given time. As expected, the figures show the delay and energy percentage increase are proportional to the temperature. Delay and energy show more percentage increase in the two lower technology nodes (45nm and 32nm) due to the uneven scaling of wires and devices in DSM VLSI circuits. As technologies scale down, timing budgets will be much tighter. Delay variation factors such as the ones just shown must be taken into consideration in the timing budget. The energy consumption of the interconnect circuits affects the temperature in the form of self-heating and thermal coupling. Excessive energy consumption due to operation in high-temperature environments may lead to harsh temperature increases. Since the interconnect circuits have less frequent activity than logic blocks, the temperature rise due to the

interconnect energy consumption may be trivial when compared to the delay increase. Figures 4.3 and 4.4 show the delay and energy increase for 1mm-5mm wires at temperatures from 50°C to 150°C in 45nm. As shown in Section 4.2.1, both the wire and the gate will contribute to the overall delay and energy increase due to higher temperatures. The propagation delay for a 5mm wire in 32nm technology can be as high as 160ps. It is expected that long, repeated wires are more vulnerable to thermal variations than short wires, even in a uniform temperature environment. Both delay and energy increase with increased temperature, due to thermal variation accumulation along the wire, which results in a significant overhead. It is likely the wirelength scaling will not be proportional to the power density increase. Thus, long interconnect design will become more challenging, even as the absolute length of the wire shrinks for 45nm and beyond.

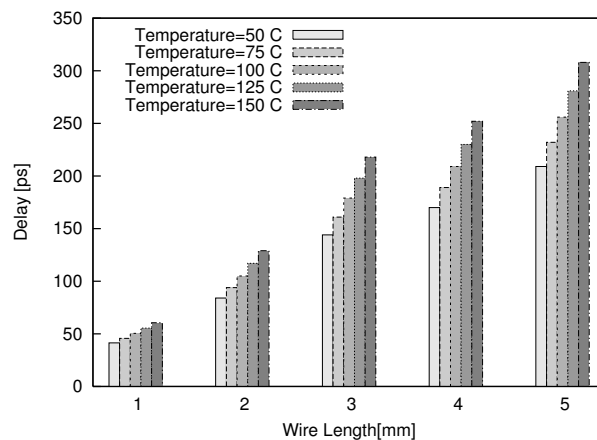


Figure 4.3 Temporal thermal variation impact on delay for 1mm-5mm repeated interconnects.

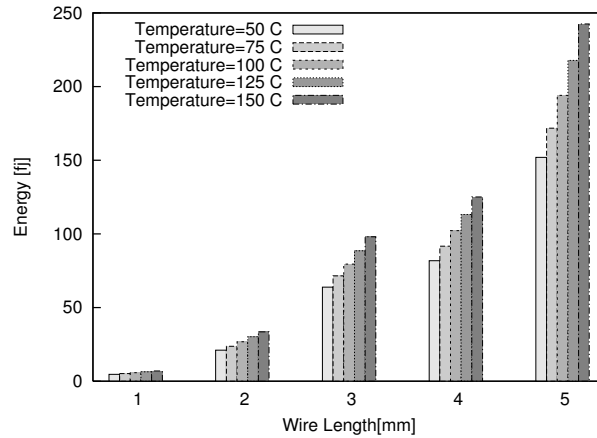


Figure 4.4 Temporal thermal variation impact on energy for 1mm-5mm repeated interconnects.

Figure 4.5 shows the delay variation due to temporal temperature variation in a 3mm wire for different numbers of repeaters. Once again, the percentage increase is with respect to the results at room temperature, all other conditions the same. By adding more repeaters into the wire, the delay of shorter wire segments will become linear. Since the relationship between the wire resistance and the temperature is close to linear, short wire segments are expected to experience less impact on the delay and energy. However, repeaters along the wire will contribute delay and energy overhead to the total delay and energy. There is only a very small change in delay and energy percentages among different repeater numbers for all three technologies. The wires with more repeaters have slightly more delay percentage increase than the ones with fewer repeaters because of the delay overhead introduced by the devices.

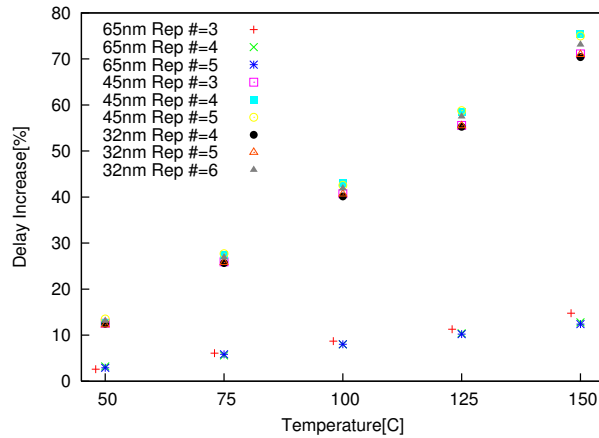


Figure 4.5 Temporal thermal variation impact on delay for different repeater numbers in 65nm, 45nm and 32nm repeated interconnects.

4.2.3 Impact on DCS and DLASA Circuit

To provide an alternative interconnect technique comparison, Figure 4.6 shows the delay percentage increase with respect to room temperature as the temperature increases from 50°C to 150°C on a current-sensed interconnect (DCS). Compared to the repeated line results shown in Figure 4.1, DCS has less delay percentage increase than repeaters in the presence of temporal thermal variations. Since the circuit has a low impedance path at the amplifier, resistance change is expected to be less in terms of temperature variation. Hence, the propagation delay of the circuit will be less sensitive to the temperature than a traditional repeated interconnect.

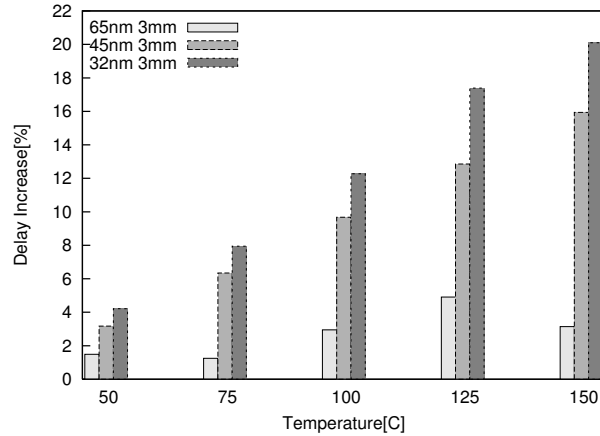


Figure 4.6 Percentage of delay increase for temporal thermal variation on a 3mm DCS wire.

Figures 4.7 and 4.8 show the delay and energy trend for DCS from 1mm to 5mm under temporal thermal variations from 50°C to 150°C. It can be concluded from these figures that DCS has less delay percentage increase than repeated lines by as much as 10ps. Furthermore, DCS is less sensitive to temperature variations in longer wires. As discussed in chapter 3, DCS senses the current instead of voltage which results in less sensitivity in delay overhead in longer wires than repeated lines and hence results in less increase in delay in the presence of thermal variations. Since the static power dissipation through the current path in the amplifier is dominant in DCS, the energy consumption of DCS does not vary significantly for different wirelengths. As the wirelength increases, the wire resistance increases and hence less amount of current is driven in the wire. As the temperature increases, the sensing capability of the circuit decreases and this results in a longer delay over the wire. This also decreases the average power and reduces the energy consumption as shown in Figure 4.8.

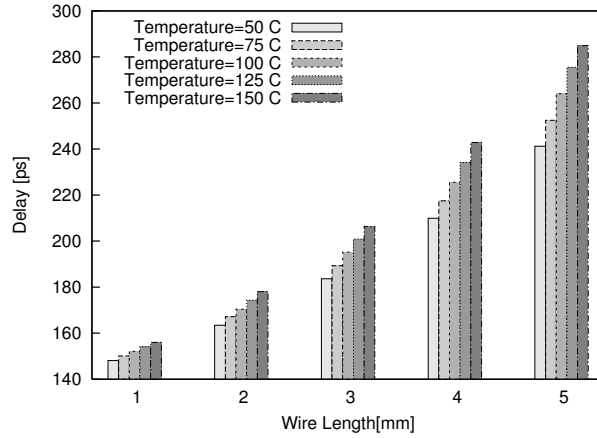


Figure 4.7 Temporal thermal variation impact on delay for 65nm, 45nm and 32nm for 1mm-5mm DCS.

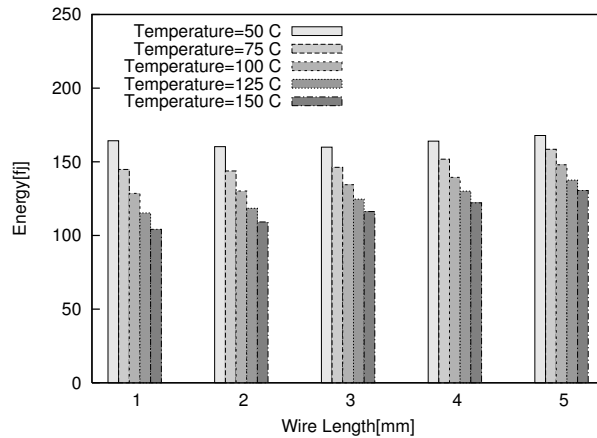


Figure 4.8 Temporal thermal variation impact on energy for 45nm, 1mm-5mm repeated interconnects.

Figure 4.9 shows the delay and energy comparison between both techniques on a 3mm wire in 45nm technology for temperatures from 50°C to 150°C. DCS shows better performance than repeaters in terms of delay and comparable energy consumption for temperatures above 125°C. This leads to the conclusion that DCS is less sensitive

under temporal thermal variation in terms of delay and shows a more favorable downward trend in the energy consumption when compared with repeaters.

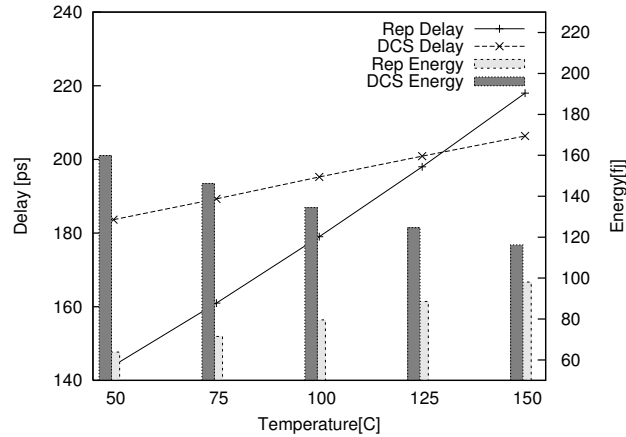


Figure 4.9 Impact on delay and energy due to temporal thermal variation on a repeated interconnect compared to DCS for a 45nm, 3mm wire.

In summary of the section, figure 4.10 and 4.11 compares the temporal thermal variation impact on DCA and DLASA in terms of delay and energy dissipation. DLASA has advantage in lower technology. First of all, propagation delay of DLASA increases at the same magnitude as DCS in each technology. Secondly, the delay overhead of DLASA is decreasing in lower technology nodes. In 32nm, the maximum delay difference between DCS and DLASA 3ps on a 3mm wire. However, DLASA still saves energy for all technology nodes under every temperature. Energy consumption on DLASA is only one third of DCS in worst case as shown in figure 4.11. It is also important to note that the performance of DLASA in terms of energy

saving keeps the same rate in all temperature which means the possible application of DLASA is suitable for high temperature and high power density environment.

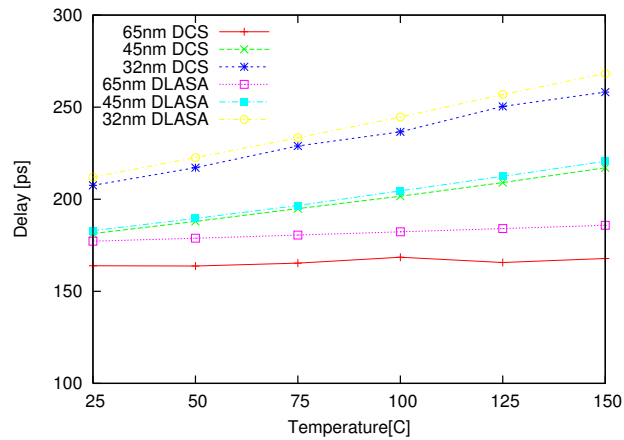


Figure 4.10 Impact on delay due to temporal thermal variation on a DCS and DLASA for 3mm wire.

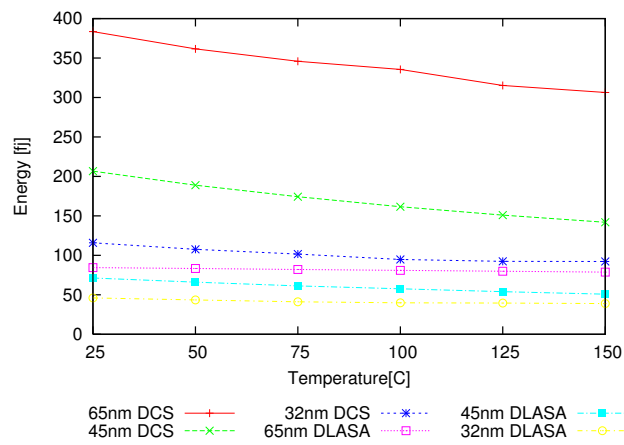


Figure 4.11 Impact on delay due to temporal thermal variation on a DCS and DLASA for 3mm wire.

4.3 Spatial Temperature Variation on Interconnect

Even though uniform temperature profiles give a general idea of the delay and energy trends on repeated interconnects and alternative circuit techniques, in the real world, nonuniform profiles may occur. A interconnect may be in an environment where the interconnect is segmented into temperature regions, and this in turn, impacts the performance in a different way that what we have seen in 4.2. A study of the impact of spatial temperature variations on interconnects follows.

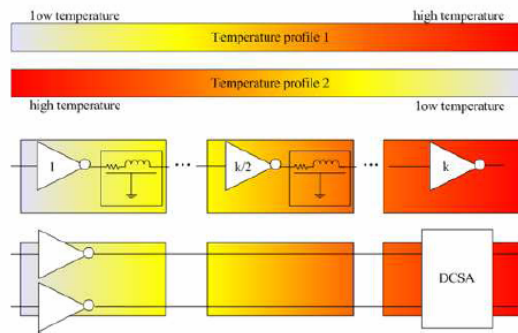


Figure 4.12 Spatial distribution profiles applied on a repeated interconnect and a current-sensed interconnect.

To proceed in a similar manner as we have done in 4.2, we have studied the spatial thermal variation impact on a 3mm wire in 65nm, 45nm and 32nm technologies and the delay and energy percentage increase results with respect to the interconnect performance at room temperature is presented in figures 4.13 and 4.14. The nonuniform temperature distribution profile applied to the interconnect for this analysis is shown in figure 4.12. For simplicity, the profile applied to the interconnect has been divided into three temperature regions, where the regions are divided by an

equal temperature gradient.

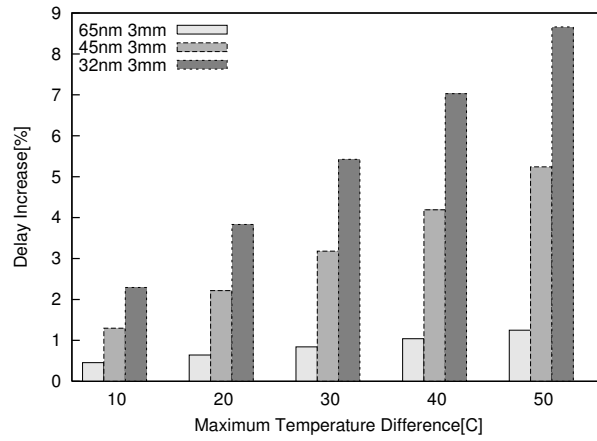


Figure 4.13 Impact of spatial thermal variation on delay for 65nm, 45nm and 32nm repeated interconnects.

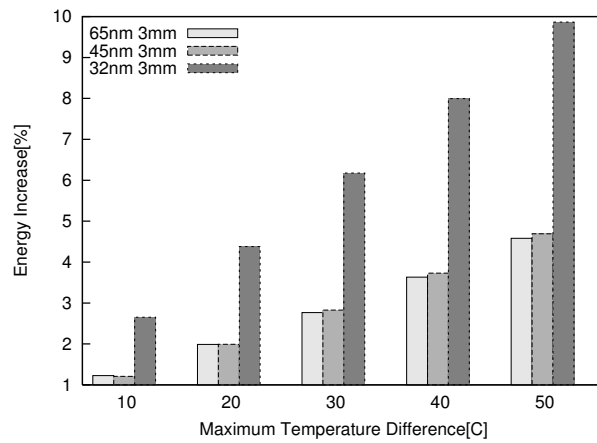


Figure 4.14 Impact of spatial thermal variation on energy for 65nm, 45nm and 32nm repeated interconnects.

Figures 4.15 and 4.16 show the delay and energy percentage increase, respectively, for temperature gradients from 10°C to 50°C. It can be seen that the delay percentage increase is higher in lower technology nodes. The difference could be as much as

8.5% for 32nm. It is expected that longer wires will experience more significant variation in delay since the possibility of crossing large temperature regions increases as the wirelength increases. On the other hand, the average wirelength is shrinking as technologies scale, which implies it is less likely to have many wires longer than 3mm in 32nm. If thermal considerations can be well incorporated into chip design, the delay and energy overhead is expected to be minimal.

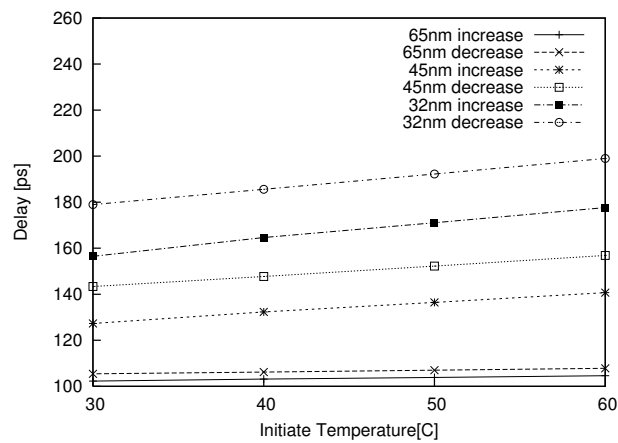


Figure 4.15 Impact of two nonuniform thermal distribution profiles on the delay of

65nm, 45nm and 32nm repeated interconnects.

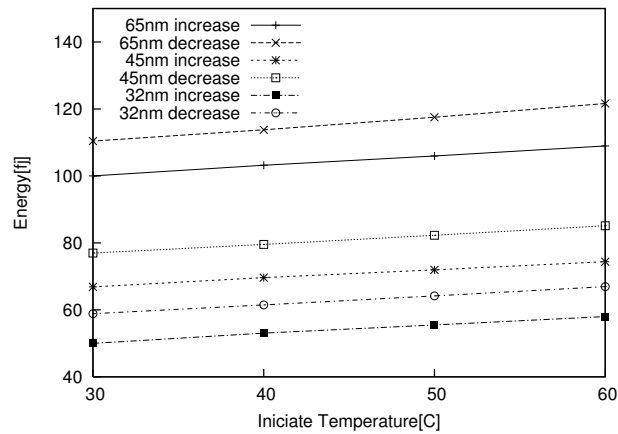


Figure 4.16 Impact of two nonuniform thermal distribution profiles on the energy of

65nm, 45nm and 32nm repeated interconnects.

Figure 4.17 and 4.18 shows the delay and energy variation under two different spatial thermal distribution profiles. As discussed in [31], on a wire analysis, a decreasing temperature profile tends to have more impact on propagation delay than an increasing temperature profile. Considering wire and repeaters, the impact of the temperature profiles is illustrated in Figure 4.17. The wire is modeled with 5 different temperature regions. Depending on the temperature profile, the lowest temperature is at the beginning or at the end of the repeated line. This lowest temperature is swept from 30°C to 60°C for both temperature profiles. There is a 15°C difference between two consecutive temperature regions. A temperature profile that decreases along the wire will have more adverse impact on the delay for all technologies. The difference that two different temperatures can cause on delay could be as much as 12.4ps on a 3mm wire in the worst case. The temperature profile impact is expected to be more significant in the lower technologies, i.e. 45nm and 32nm. Furthermore, simulation results in Figure 4.18 show that energy consumption follows the same trend as delay under these two distribution profiles.

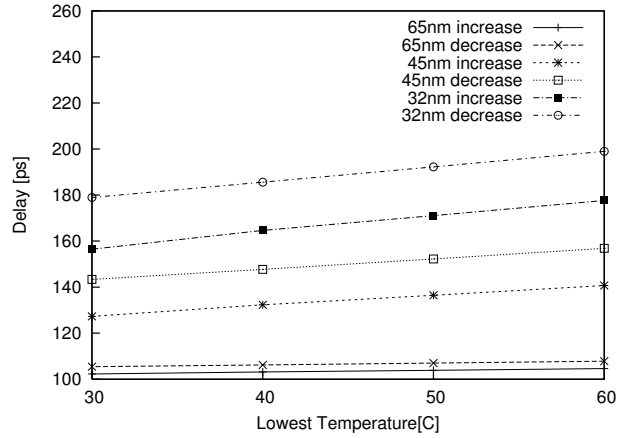


Figure 4.17 Impact of two nonuniform thermal distribution profiles on the delay of 65nm, 45nm and 32nm repeated interconnects.

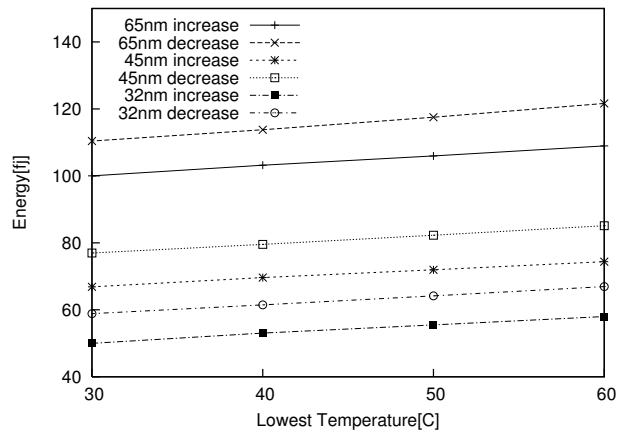


Figure 4.18 Impact of two nonuniform thermal distribution profiles on the energy of 65nm, 45nm and 32nm repeated interconnects.

Once more, to provide an alternative circuit technique for comparison with repeater insertion, Figures 4.19 and 4.20 show the delay and energy trend of DCS in the presence of the same two temperature profiles. DCS does not have the distributed

nature that repeated lines do, and the signal sensed in DCS is current rather than voltage. Thus, the most significant component in DCS is the amplifier, since the low impedance path located in the amplifier will be highly influenced by the temperature. This variation will further change the load resistance and the propagation delay. The performance degradation of DCS circuits is expected to be more significant if the amplifier is in the higher temperature region. The reversed performance trend in DCS gives designers an alternative option. If the repeated line will have a worst case thermal profile, DCS may be the choice to mitigate the degradation.

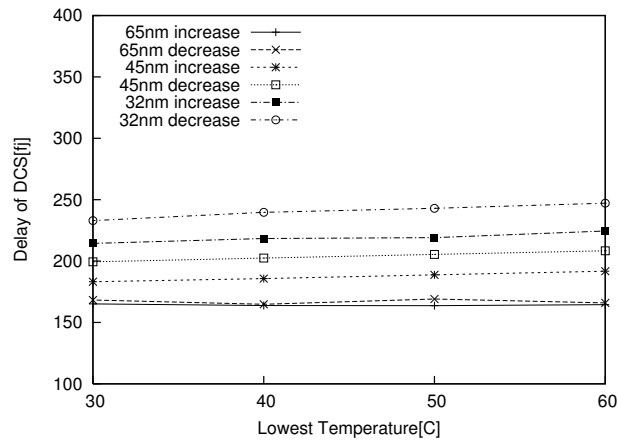


Figure 4.19 Impact of two nonuniform thermal distribution profiles on the delay of 65nm, 45nm and 32nm DCS.

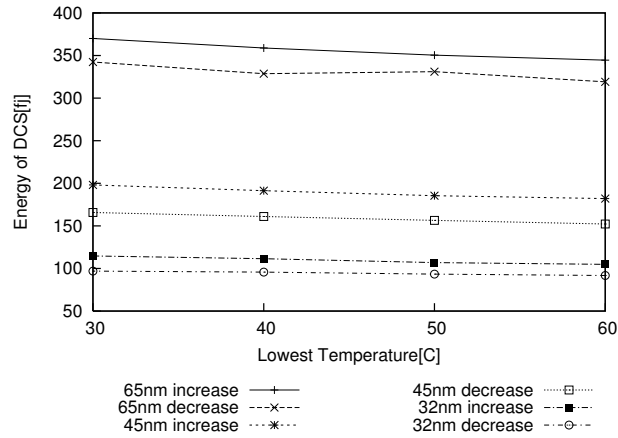


Figure 4.20 Impact of two nonuniform thermal distribution profiles on the energy of 65nm, 45nm and 32nm repeated interconnects.

Figure 4.21 shows the delay percentage increase on a repeated line in 65nm, 45nm and 32nm. 3 repeaters and 5 repeaters have been inserted into a 3mm wire that experiences the same temperature profile. The results are normalized to the delay resulting from the uniform 25°C temperature condition. An increasing delay percentage increase has been observed in all technologies for these conditions. Smaller technologies are more influenced and more sensitive to a higher average temperature environment. This observation can be explained by the fact that the gate delay variation contributes more to the overall delay under a nonuniform spatial temperature distribution.

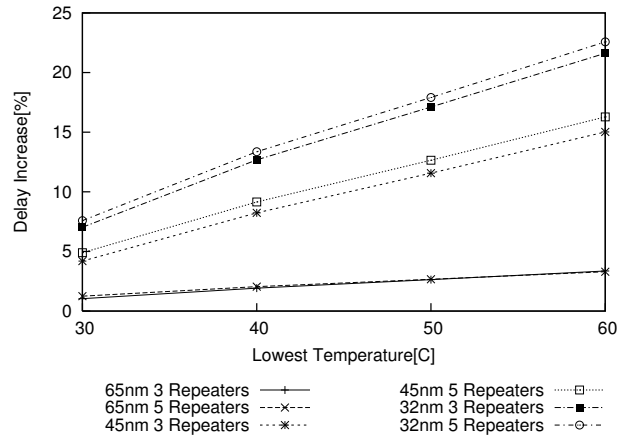


Figure 4.21 Impact of spatial thermal variations on delay and energy for varying number of repeaters on 65nm, 45nm and 32nm repeated interconnects.

Figure 4.22 shows the energy and delay on a 3mm wire implemented as DCS and repeated line. Both of these circuit techniques are subjected to a spatial temperature profile with 3 temperature regions. There is a 25°C difference between neighboring regions. As shown in the figure, repeated lines show better performance in terms of both speed and energy. Furthermore, they are expected to keep these merits as the average temperature increases. However, the advantage of repeaters over DCS for nonuniform spatial temperature profiles is not guaranteed if they experience a decreasing temperature profile, as previously shown in Figures 4.19 and 4.20.

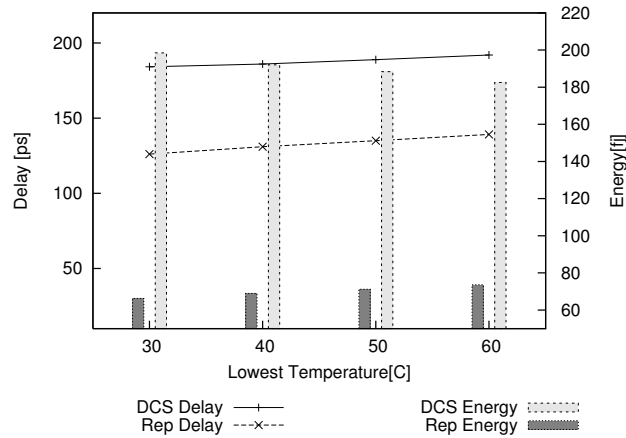


Figure 4.22 Comparison between the impact on delay and energy due to spatial thermal variation on a repeated interconnect and differential current sensing.

Temperature profiles that have been analyzed for repeater insertion line and DCS have also been used for DLASA simulation. Figure 4.23 shows a delay comparison between DCS and DLASA under decreasing and increasing temperature profiles on a 3mm wire in 45nm technology. It can be observed that DLASA also suffers more on a decrease temperature profile comparing to an increasing profile as DCS and repeater insertion line. Meanwhile, delay variation dependence on the temperature profile is less severe than DCS. It means that DLASA is less sensitive to the temperature profile and hence can be used in a design that has a fixed temperature and time constraint. This observation can be seen for all technologies.

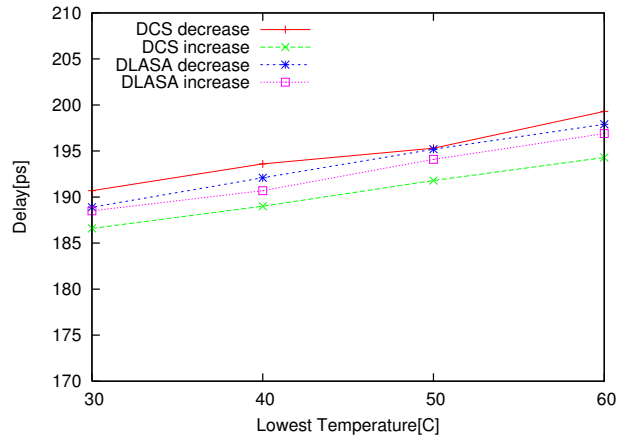


Figure 4.23 DLASA/DCS delay under different temperature profiles

Similar to the results in figure 4.23, energy dissipation of DLASA also has less dependency on different temperature profile than DCS circuit. And this advantage can be seen in all technologies. Figure 4.24 shows the results of DLASA and DCS circuits in decreasing and increasing temperature profiles for a 3mm wire in 45nm technology. The maximum difference of DLASA under the difference temperature is 4.2 fj, while DCS has 22.5fj.

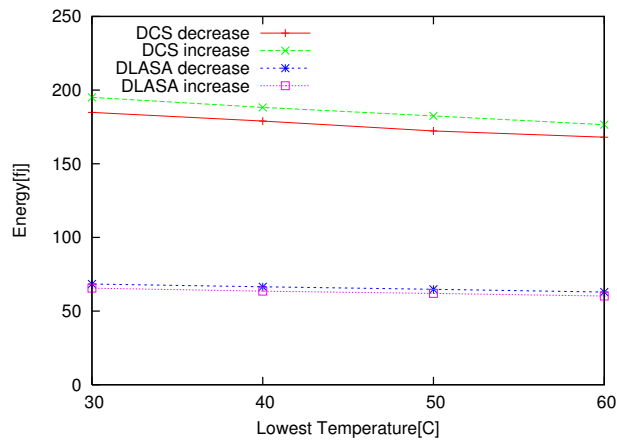


Figure 4.24 DLASA/DCS energy under different temperature profiles

Sensitivity to temperature variation over a spatial domain of DCS and DLASA has been shown in figure 4.25 and 4.26. The delay dependence on temperature variation does not change dramatically for lowest temperature at 30°C and the variation from 10°C to 50°C for both DCS and DLASA. The delay difference between DCS and DLASA due to temperature variation does not increase either. But the difference in lower technology nodes is smaller. It means that DLASA only has slightly more overhead on delay than DCS in lower technology under the same spatial variation profile. Meanwhile, figure 4.26 shows the advantage of energy saving by using DLASA. Under the same spatial variation temperature profile as shown in figure 4.25, the energy consumption variation on DLASA is only 1.2% comparing to DCS in 45nm and 1.3% in 32nm in worst case. Figure 4.25 and 4.26 show clearly that DLASA has the advantage of energy saving with very limited delay overhead and should be considered in a low power design.

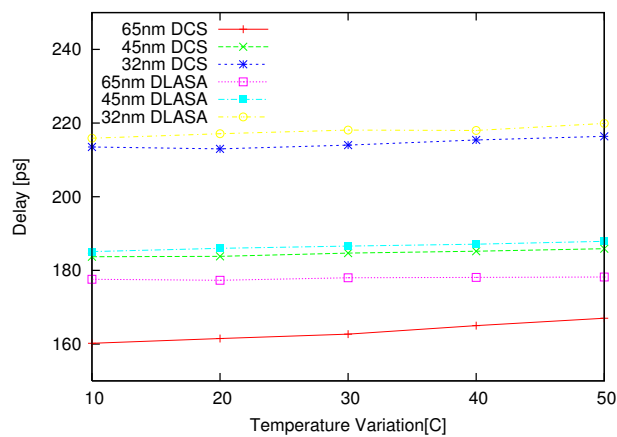


Figure 4.25 DLASA/DCS delay with different temperature variation

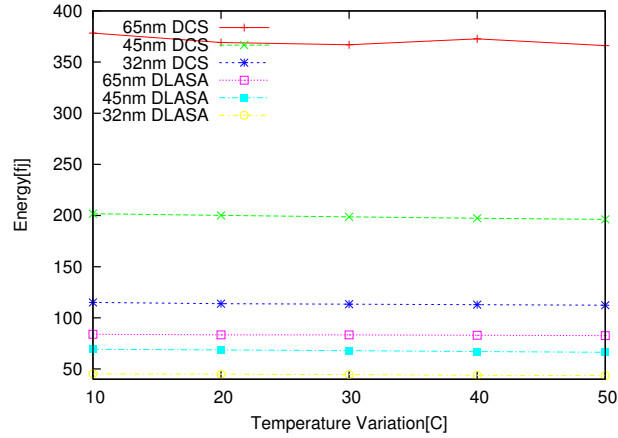


Figure 4.26 DLASA/DCS energy with different base temp

4.4 Analytical Model for Repeated Line

This section will discuss an analytical model of temperature variation for repeated line in a qualitative approach. An accurate device physics behavior under temperature variation involves a lot of quantum physics theory such as scattering which is beyond the topic of this thesis. Instead, a general discussion about the temperature variation on a repeated line is beneficial to understand the overhead contribution by device and wire.

To understand the impact of temporal and spatial thermal variations on delay, an analytical model must be developed. In the case of a repeated interconnect; the traditional delay expression consists of the Elmore delay of the wire plus the device propagation delay. Beginning by considering a wire of length l , divided by N repeaters into N segments, the total delay of the interconnect can be calculated as:

$$\sum_{n=1}^N (t_{p_gate,n} + t_{p_wire,n}) = \sum_{n=1}^N t_{p_gate,n} + \sum_{n=1}^N t_{p_wire,n} \quad (4.1)$$

where $t_{p_gate,n}$ is the gate delay of the nth gate and $t_{p_wire,n}$ is the wire delay of the nth segment. First, the wire delay is modeled in terms of temperature and will consider spatial thermal variation for both wire and gate in this analysis. The wire parameter that is most sensitive to temperature variations is the resistance R. We will assume inductance and capacitance do not change with temperature for this analysis. The Elmore delay of a wire segment is given in Equation 4.2 [31].

$$\sum_{n=1}^N t_{p_wire,n} = D_w = D_0 + (C_0 l + C_L) \rho_0 \beta \int_0^l T(x) dx - c_0 \rho_0 \beta \int_0^l x T(x) dx \quad (4.2)$$

where D_0 is given in Equation 4.3 and is the Elmore delay of the interconnect corresponding to the unit length resistance at 0°C.

$$D_0 = R_d (C_L + c_0 l) + (c_0 \rho_0 \frac{l^2}{2} + \rho_0 l C_L) \quad (4.3)$$

If we assume the thermal profile to be exponential along the interconnect as represented by Equation 4.4 and as assumed by [31], the delay of the nth segment in the wire can be represented as shown in Equation 4.5.

$$T(x) = a \exp(-bx) \quad (4.4)$$

$$D = D_0 + (C_0 L + C_L) \rho_0 \beta \int_{\frac{N}{(n-1)l}}^{\frac{n}{N}} a \cdot \exp(-bx) dx - C_0 \rho_0 \beta \int_{\frac{N}{(n-1)l}}^{\frac{n}{N}} x a \cdot \exp(-bx) dx \quad (4.5)$$

By integrating Equation 4.5, we obtain the total wire delays:

$$D = D_0 + (c_0 L + C_>) \rho_0 \beta (-a/b)(e^{-bl} - 1) + c_0 \rho_0 a / (b^2) [e^{-bl} (bl + 1) - 1] \quad (4.6)$$

The following step is to obtain a gate delay expression. In an inverter chain, for the j-th inverter stage, the propagation delay can be represented as:

$$t_{p,j} = t_{p0} \quad (4.7)$$

where t_{p0} is the intrinsic gate delay given by Equation 4.8. Since we have made the

assumption that capacitance does not vary significantly with temperature, only t_{p0} in Equation 4.7 is temperature-dependent. It, in turn, is caused by drain current variation. The drain current variation can be modeled by the mobility and the threshold voltage in Equations 4.9 and 4.10, respectively [15].

$$t_{p0} = 0.69R_{eq}C_{int} \quad (4.8)$$

$$\mu_n(T) = \mu_n(T_0)\left(\frac{T}{T_0}\right)^{\alpha_\mu} \quad (4.9)$$

$$V_{GS} = V_{GSF} = V_T(T_0) - \alpha V_T T_0 \quad (4.10)$$

From [34], an expression for the drain current can be obtained as shown in Equation 4.11.

$$I_D = \frac{\mu_n(T_0)\left(\frac{T_0}{T}\right)^{\alpha_\mu} C_{ox} W}{2} \alpha_{VT}^2 \quad (4.11)$$

From [11] an expression for R_{eq} is given, as shown in Equation 4.12.

$$R_{eq} = \frac{1}{V_{DD}/2} \int_{V_{DD}/2}^{V_{DD}} \frac{V}{I_{DSAT}(1 + \lambda V)} dV \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{7}{9} \lambda V_{DD}\right) \quad (4.12)$$

Plugging Equations 4.11 and 4.12 into Equation 4.8 and substituting for the constant α_{VT}^2 using Equation 4.13, a final expression for the gate delay is obtained and is shown in Equation 14.

$$\alpha_{VT}^2 = \left(\frac{\partial V_T}{\partial T}\right)^2 = \left(\frac{\partial V_T}{\partial(a \exp(\frac{n}{N} l(-b)))}\right)^2 \quad (4.13)$$

$$\sum_{n=1}^N t_{p_gate,n} = D_g = 0.69 \frac{3V_{DD}C_{int}}{2a^2 C_{ox} \frac{W}{L}} \left(1 - \frac{7}{9} \lambda V_{DD}\right) \frac{1}{\mu(T_0)\left(\frac{T_0}{T}\right)^{\alpha_\mu}} \cdot \sum_{n=1}^N \frac{1}{\frac{\partial V_T^2}{\partial \exp(\frac{-bnl}{N})}} \quad (4.14)$$

With this, an expression for the total delay considering both gate and interconnect can be developed and is shown in Equation 4.15 where D_w is given by Equation 4.6 and D_g is given by Equation 4.14.

$$D_{total} = D_w + D_g \quad (4.15)$$

A similar analysis can be done to develop an expression for the total delay in presence of temporal thermal variations. In that case, position x is constant and the temperature at any given time is the same for the whole device and interconnect structure.

It has been pointed that in equation 4.11, a compensation point for threshold and mobility can be set at $a_v = -2$. This point means that the decreasing of mobility compensates the decreasing of threshold in terms of delay. Thus the modeling of temperature variation impact is truly depending only on the local temperature of the wire. This is an ideal model that may not be true in short channel device. According to equation 4.15, we can conclude that temperature variation impact on repeated line is due to following factors:

1. For RC wire, resistance contributes the delay overhead and largely depends on the temperature distribution profile because of the Elmore delay.
2. For repeated line, the degree of mutual compensation between threshold and mobility is essential for accurate delay overhead predication. Since the total compensation point ($a_v = -2$) where device is independent of temperature is not realistic for short channel device, the quadratic decreasing of mobility due to temperature increasing will dominate the transistor delay.
3. The percentage of delay overhead contribution from wire and transistor in the

repeated line is largely depend on the transistor size, number, and supply voltage under same temperature profile. Since interconnect circuit usually has large transistors than logic circuit, transistors will dominant the delay overhead.

Table 4.2 shows SPICE simulation result of repeated line under a linear increasing temperature profile. The linearity is $0.1(\mu m) / \Delta T(^{\circ}C)$. It can be seen that where temperature variation is smaller than 10 ° C it will not have significant impact on a repeated line. A wire with more repeaters is more sensitive to the temperature variation.

Table 4.2 Spatial Temperature Variation Impacton Repeated line

Temperature Variation	65nm		45nm		32nm	
	1mm, 1 repeater	2mm, 4 repeaters	1mm, 1 repeater	2mm, 4 repeaters	1mm, 1 repeater	2mm, 4 repeaters
1°C	35.9ps	89.6ps	33.8ps	76.2ps	31.5ps	86.9ps
5°C	35.9ps	89.6ps	33.8ps	76.2ps	31.5ps	86.9ps
10°C	36.0ps	89.6ps	33.8ps	76.3ps	31.5ps	87.0ps
30°C	36.0ps	89.7ps	33.9ps	76.4ps	31.5ps	87.2ps
60°C	36.0ps	89.8ps	33.9ps	76.7ps	31.6ps	87.6ps

4.5 Summary and Conclusion

This chapter addressed the impacts on interconnect circuits under harsh uniform temperature changes and nonuniform spatial temperature distribution profiles. Temporal and spatial thermal variations were addressed in 65nm, 45nm and 32nm interconnect circuits. An analytical discussion has been provided, to consider temperature variation impact on both gate and wire delay. Standard repeater insertion and differential current sensing techniques have been implemented and their performance was compared under different thermal profiles. The circuits were analyzed in temperatures as high as 150°C for the temporal variations, with a

maximum temperature difference through wire of up to 50°C. High temperature caused more delay and power overhead in smaller technologies, i.e. 45nm and 32nm, by as much as 71% at 150°C for a given wirelength of 3mm in 32nm. Spatial temperature distribution profiles influenced the propagation delay by 14.7% for a maximum thermal gradient of 50°C in the worst case for a 32nm, 3 mm repeated wires. The repeated line is affected more by a decreasing spatial temperature profile than by an increasing profile. However, the delay degradation of an alternative differential current sensing (DCS) technique will be largely determined by the amplifier temperature. DLASA circuits have also been simulated and compared. It shows that DLASA has the same trend of delay and energy consumption comparing to DCS in the same temporal and spatial temperature profile. However, for the same temperature profile, DLASA is less sensitive than DCS. It supports the conclusion in chapter 3 that DLASA has the advantage in energy saving. Furthermore, the delay overhead will become smaller in the non-uniform temperature.

From these observations, we can conclude that as designs scale down in future technologies, shorter wires will be preferable from a thermal standpoint. Design for balanced core temperatures becomes extremely important, to avoid hotspots that may cause performance degradation. As an alternative to the traditionally used repeater insertion techniques, designers may consider the use of advanced circuit techniques such as DCS and DLASA.

CHAPTER 5

SUMMARY

This thesis explores several aspects in VLSI interconnect circuit design. First, it introduces the background and motivation about the necessity of this work. Current mode circuit application in interconnect has not been widely accepted. One of the reasons is that the DCS circuit consumes considerable amount of static and leakage power compared to traditional repeater insertion. Also, there is less study of interconnect circuit especially differential current sensing in terms of temperature variation tolerance.

An energy-aware differential current sensing amplifier (DLASA) has been proposed and analyzed. This amplifier utilizes two sleep transistors to mitigate the energy dissipation due to static and leakage in the original DCS circuit. Energy in the DCS is minimized because of the power gating and transistor stacking effects. DLASA has been simulated in 65nm, 45nm and 32nm and compared with DCS and repeated line. Results has been discussed and shown that DLASA can significantly reduce the energy consumption with very limited delay and signaling overhead.

Temperature impact on interconnect circuits due to temporal and spatial variation have also been analyzed. Repeated line, DCS and DLASA has been simulated and compared under different temperature profiles. Result shows that delay of repeated line are more sensitive to the temperature compared to DCS especially in lower technology node. The direction of thermal gradient will have different impact on interconnect circuits. DLASA has the same trend in terms of delay and energy comparing to DCS under same temperature profile but the sensitivity is lower.

BIBLIOGRAPHY

- [1] G. Chen and E. G. Friedman, " Low Power Repeaters Driving RC and RLC Interconnects with Delay and Bandwidth Constraints," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 14, No. 2, pp. 161-172, February 2006.
- [2] J. Lillis, C. Cheng and T. Y. Lin "Optimal wire sizing and buffer insertion for low power and a generalized delay model" International Conference on Computer Aided Design, 1995 pp.138
- [3] International Technology Roadmap for Semiconductor May, 2005
- [4] A. Maheshwari, Srividya Srinivasaraghavan and Wayne Burleson, Quantifying the Impact of Current-Sensing on Interconnect Delays Trends, IEEE ASIC SOC Conference, 2002
- [5] A. Maheshwari, Circuit and Signaling Methods for On-chip Interconnects, PH.D Dissertation chapter 9, University of Massachusetts Amherst, 2004
- [6] Vishak Venkatraman and Wayne Burleson, Robust Multi-Level Current-Mode On-Chip Interconnect Signaling in the Presence of Process Variations, Sixth International Symposium on Quality of Electronic Design, 2005
- [7] S. Xu, V. Venkatraman and W. Burleson, Energy-Aware Differential Current Sensing for Global On-Chip Interconnects, 49th IEEE International Midwest Symposium on Circuits and Systems, 2006
- [8] A. Maheshwari, W. Burleson, "Differential current-sensing for on-chip interconnects" IEEE Transactions on VLSI Systems VOL.12, 2004 pp.1321

- [9] Jinwook Jang, Sheng Xu, Wayne Burleson, Jitter in Deep Sub-micron Interconnect, IEEE Computer Society Annual Symposium on VLSI, 2005.
- [10] Predictive Technology Model <http://www.eas.asu.edu/~ptm/>
- [11] Rabaey, Jan M., Chandrakasan, Anantha, and Nikolic, Borivoje. Digital Integrated Circuits: A design perspective. Prentice Hall, 2003.
- [12] Sunter, S., and Roy, A. Bist for phase-locked loops in digital applications. In Proceedings of the IEEE International Symposium on Circuits and Systems (1999), pp. 532-540.
- [13] W. Zhao, Y. Cao, "New generation of Predictive Technology Model for sub-45nm design exploration," pp. 585-590, ISQED, 2006
- [14] Bakoglu, H.B. Circuits, Interconnections, and Packaging for VLSI. Addison Wesley, 1990
- [15] Srinivasan, Sriram, Circuit & Signaling Strategies for On-Chip Global Interconnects in DSM CMOS Master Thesis, University of Massachusetts Amherst, 2002
- [16] E. Seevinck, P. J. van Beers, and H. Ontrop, Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for cmos sram's, IEEE Journal of Solid-State Circuits, pp. 525-536, 1991.
- [17] T. N. Blalock and R. C. Jaeger, A high speed sensing scheme for 1t dynamic ram's utilizing the clamped bit-line sense amplifier, IEEE Journal of Solid-State Circuits, pp. 618-625, 1992.

- [18] Jinn-Shyan Wang, Wayne Tseng, and Li Hung-Yu, Low-power embedded sram with the current-mode write technique, *IEEE journal of Solid-State Circuits*, pp. 119-124, 2000.
- [19] M. Izumikawa and M. Yamashina, A current direction sense technique for multiport sram's, *IEEE Journal of Solid-State Circuits*, pp. 546-551, 1996.
- [20] Manoj Sinha and Wayne Burleson, Current-Sensing for Crossbars, *IEEE ASIC SOC Conference*, 2001
- [21] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits" *Proceedings of the IEEE*, VOL. 91, NO. 2, 2003
- [22] J. Donald and M. Martonosi, "Temperature-aware design issues for smt and cmp architectures," In *Proceedings of the Workshop on Complexity-Effective Design (WCED)*. ACM Press, 2004.
- [23] M. Pedram and S. Nazarian, Thermal Modeling, Analysis and Management in VLSI Circuits: Principles and Methods, *Proc. of IEEE*, special issue on Thermal Analysis of ULSI, 2006
- [24] V.De and S. Borkar, "Technology and design challenges for low power and high performance," in *Proc. ISLPED*, pp. 163-168, 1999
- [25] Y. Li, K. Skadron, Z. Hu, and D. Brooks. "Evaluating the thermal efficiency of SMT and CMP architectures". In *IBM T. J. Watson Conference on Interaction between Architecture, Circuits, and Compilers*, October 2004

- [26] K. Skadron, M. R. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan. "Temperature-aware microarchitecture: Modeling and implementation", ACM Trans. Archit. Code Optim., 1(1):94–125, 2004
- [27] Y-K Cheng; P. Raha, C-C Teng, E. Rosenbaum, and S-M Kang, "ILLIADS-T: an electrothermal timing simulator for temperature-sensitive reliability diagnosis of CMOS VLSI chips", IEEE trans Computer-Aided Design of Integrated Circuits and Systems, Volume 17, No 8 Pages:668 – 681, Aug. 1998
- [28] W. Batty et al. Global coupled EM-electrical-thermal simulation and experimental validation for a spatial power combining MMIC array. IEEE Transactions on Microwave Theory and Techniques, pages 2820–33, Dec. 2002.
- [29] C. J. M. L. H. Vinke, "Recent Achievement in the Thermal Characterization of Electronic Devices by Means of Boundary Condition Independent Compact Models," 13th IEEE SEMITHERM Symposium, Austin, Texas, 1997
- [30] W. Huang, M. R. Stan, K. Skadron, K. Sankaranarayanan, S. Ghosh, and S. Velusam. "Compact thermal modeling for temperature-aware design", In Proceedings of the 41st annual conference on Design Automation, pages 878–883, 2004.
- [31] A.H. Ajami, K. Banerjee, and M. Pedram, "Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects", IEEE Trans Computer-aided Design of Integrated Circuit and Systems, VOL. 24, NO. 6 Jun. 2005
- [32] D. Chen, E. Li, E. Rosenbaum, and S.-M. Kang, "Interconnect thermal modeling for accurate simulation of circuit timing and reliability," IEEE Trans. Computer-Aided Design, vol. 19, Feb. 2000, pp. 197–205

[33] S. Im, N. Srivastava, K. Banerjee, K.E. Goodson, “ Scaling Analysis of Multilevel Interconnect Temperature for High-Performance ICs”, IEEE Trans. on Electron Devices, Vol. 52, No. 12, 2710-2719, Dec. 2005

[34] I.M. Filanovsky, A. Allam, .Mutual Compensation of Mobility and Threshold Voltage Temperature Effects with Applications in CMOS Circuits., Proceedings of the IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, Vol. 48, No. 7, pp. 876-884, 2001.