

A Low-Power Analog Bus for On-Chip Digital Communication

by

Farah Naz Taher

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 3, 2013

Keywords: System-on-Chip Design, Low Power Design, On-chip Communication, Bus Architectures, Power Management, Analog Bus

Copyright 2013 by Farah Naz Taher

Approved by

Vishwani D. Agrawal, Chair, James J. Danaher Professor of Electrical Engineering
Victor P. Nelson, Professor of Electrical and Computer Engineering
Adit D. Singh, James B. Davis Professor of Electrical and Computer Engineering

Abstract

At present, performance and efficiency of a system-on-chip (SoC) design depends significantly on the on-chip global communication across various modules on the chip. On-chip communication is mostly implemented using a bus architecture that runs long distances, covering significant area of the integrated circuit. Difficult challenges in designing of a large SoC, e.g., one containing many processor cores, include hardware area, power dissipation, routing complexity, congestion and latency of the communication network. In this work, we propose an analog bus for digital data. In our scheme we replace n wires of an n -bit digital bus carrying data between cores with just one (or few) wire(s) carrying analog signal(s) encoding 2^n levels of voltage. This analog bus uses digital-to-analog converter (DAC) drivers and analog-to-digital converter (ADC) receivers. Such on-chip communication scheme can potentially save hardware area and power. Reduction in number of wires saves chip area and the reduction in total intrinsic wire capacitance consequently reduces bus power consumption. The scheme should also reduce signal interference and crosstalk by eliminating the need for multiple line drivers and buffers. In spite of overheads of the DACs and ADCs, savings in power consumption from our scheme is significant. We have carried out simulated experiments that serve as a proof-of-concept by evaluating power consumption of a single wire with DAC/ADC encoding in comparison to an n -bit digital bus of a large system. SPICE simulation for an ideal case shows that the ratio of bus power consumed by the proposed analog scheme to a typical digital scheme (without bus encoding or differential signaling) is given by $P_{analog}/P_{digital} = 1/(3n)$. For $500MHz$ frequency and $1mm$ intermediate wire line, 4-bit replacement analog bus consumes $16\mu W$ over $219\mu W$ in parallel bus. Whereas, the 8-bit replacement bus consumes $18\mu W$ over the $470\mu W$ power consumption in the 8-bit parallel bus.

Acknowledgments

‘The more you learn, the more you realize how little you know’– has become the code I live by, especially in the last two years. In the process of attaining this MS degree, I have realized at every step that whatever little I have achieved was not possible without the amazing people I have in my life as mentors, friends and family.

First and foremost I want to thank my advisor, Dr. Vishwani Agrawal, for being there for me from my first day in Auburn. He is a great mentor, guide, and teacher. He has always been very supportive, and guided me with encouragement, patience and judicious advice.

I would like to thank Dr. Adit Singh, not only for being in my thesis committee, but also for the two wonderful courses I had the privilege of taking with him. He has always been helpful and kind. I also thank Dr. Victor Nelson for agreeing to be in my thesis committee, and for giving his detailed feedback on the thesis.

I express my sincere appreciation and gratitude to Mr. Charles Ellis for giving me the opportunity to work in the Alabama Microelectronics Science and Technology Center. He helped me out in a difficult time by providing me with a research assistantship. I thank Dr. Suraj Sindia for all his help and suggestions. He has always been selflessly helping everyone in need. I would take the opportunity to thank all my teachers from my school, to North South University, to here in Auburn University.

I thank Mustafa Munawar Shihab for being my brother and my best friend. All the hard work we did together now feels worth it as we complete our Masters, and achieve a goal together yet again. Thank you Brother.

No words are adequate to express my gratefulness to my family for their unconditional love and support. I am more than lucky to have such a family who understands and supports my goal. I thank my mother Shahnaz Sultana and my father Abu Taher Chowdhury for all

the sacrifices they have made, and for all their encouragement that made me what I am today. I thank my sister Mayesha Naz Taher for all the love and courage she gave me.

I thank my husband Muhammad Asaduzzaman Shanto for all his love, patience, and support. I am lucky to have such a supportive and patient life partner. I dedicate my work to my awesome family.

Finally, I thank the Almighty for this wonderful life, and for the wonderful people He has filled it up with.

Table of Contents

Abstract	ii
Acknowledgments	iii
List of Figures	viii
List of Tables	xi
1 Introduction	1
1.1 Motivation	2
1.2 Contribution	4
1.3 Problem Statement	5
1.4 Organization	5
2 Overview: Power	7
2.1 Power Dissipation	7
2.1.1 Dynamic Power	8
2.1.2 Static Power	10
2.2 Power Reduction Techniques	11
2.2.1 Technology Scaling	13
2.2.2 Transistor Sizing	13
2.2.3 Interconnect Optimization	14
2.2.4 Clock Gating	14
2.2.5 Power Gating	14
2.2.6 Supply Voltage and Threshold Voltage Scaling	15
2.2.7 Multi-Voltage Design	15
2.2.8 Variable Supply and Threshold Voltages	15
2.2.9 Technology Mapping	15

2.2.10	Floorplanning, Cell Placement and Wire Routing	16
3	On-Chip Communication	17
3.1	Bus Architecture	18
3.2	Bus Topology	18
3.3	Issues With Parallel Bus	22
3.3.1	Routing Complexity	22
3.3.2	Area	22
3.3.3	Power Dissipation	23
3.3.4	Performance	23
3.3.5	Signal Integrity and Crosstalk	23
3.4	Possible Solutions	24
4	Previous Work	25
4.1	NOC	25
4.2	SerDes	27
4.2.1	Construction of SerDes	27
4.2.2	SERDES Approaches	29
5	Analog Bus	31
5.1	Concept	32
5.2	Structure	33
5.3	Proposition	34
5.4	V_{swing}	36
5.5	Theoretical Analysis	37
5.5.1	Voltage	37
5.5.2	Power	38
6	Data Conversion	40
6.1	Analog to Digital Converter	40
6.2	Digital to Analog Converter	43

6.3	Design considerations	45
7	Evaluation	47
7.1	Experiment Setup	47
7.2	Power Analysis: Replacement of 4-Bit Parallel Bus	48
7.3	Power Analysis: Replacement of 8-Bit Parallel Bus	52
7.4	Discussion of Results	53
8	Conclusion	55
8.1	Challenges and Future Work	55
8.1.1	Design suitable converters	56
8.1.2	Encoding Scheme	56
8.1.3	Combination of Analog Bus with other schemes	56
8.1.4	Mixed-Signal Compression of Digital Test Data	57
	Bibliography	58

List of Figures

1.1	A brief chronology of the major milestones in the development of VLSI [65]. . .	1
1.2	Four dimensions of optimization in VLSI design.	3
2.1	Switching Power [27].	8
2.2	Short-Circuit Power [27].	9
2.3	Static power [27].	11
2.4	Clock gating.	14
3.1	IBM Cell ring bus communication architecture [53].	17
3.2	Bus structure [53].	18
3.3	Shared bus [53].	19
3.4	Hierarchical bus [53].	19
3.5	Ring bus [53].	20
3.6	Split bus [53].	20
3.7	Crossbar bus [53].	21
3.8	Partial crossbar/matrix bus [53].	21
3.9	Tristate buffer bus [53].	21

4.1	Various communication architectures [8].	26
4.2	SerDes [28].	27
4.3	SerDes structure [28].	28
4.4	Example of serialization [37].	29
4.5	The <u>Silent</u> scheme [38].	30
5.1	Total interconnect length (m/cm ²) - Metal 1 and five intermediate levels, active wiring only [60].	31
5.2	Parallel bus and analog bus.	34
5.3	V_{swing}	37
6.1	Signals resulting from A/D and D/A conversion in a mixed-signal system [5].	41
6.2	Basic ADC structure [30].	42
6.3	Basic DAC structure [30].	44
7.1	4-bit parallel bus.	48
7.2	Analog bus replacing 4-bit parallel bus of Figure 7.1.	48
7.3	Experimental setup for analog bus replacing a 4-Bit parallel bus.	49
7.4	4-Bit input patterns.	49
7.5	4-Bit digital input converted to analog data.	50
7.6	Parallel bus vs. analog bus (bus width = 4, frequency = 1GHz).	51
7.7	Parallel bus vs. analog bus (bus width = 4, frequency = 500MHz).	51

7.8	An analog bus to replace an 8-bit parallel bus.	52
7.9	8-bit input patterns.	53
7.10	8-bit digital input converted to analog data.	54
7.11	Parallel bus vs. analog bus (bus Width = 8, frequency = 500MHz).	54

List of Tables

2.1	Strategies for low power designs [26].	12
2.2	Trade off associated with power management techniques [26]	13
4.1	Comparison overview of advantages/disadvantages of SerDes architectures [39].	29
5.1	Bit-wise noise tolerance of analog bus.	36
5.2	Random data patterns and transition analysis.	38
5.3	Comparison of parallel, serial and analog buses.	39
7.1	Setup	47
7.2	Comparison of power consumption of 4-bit parallel bus and analog bus for frequency = 1GHz.	50
7.3	Comparison of power consumption of 4-bit parallel bus and analog bus for frequency = 500MHz.	51
7.4	Comparison of power consumption of 8-bit parallel bus and analog bus for frequency = 500MHz.	52
7.5	Power consumption of 4-bit and 8-bit buses.	53
7.6	Converter design survey [52].	54

Chapter 1

Introduction

The transistor, one of the most important discoveries of 20th century and the heart of electronics, was invented at Bell Labs in New Jersey in 1947 by John Bardeen, Walter Brattain, and William Shockley. The second gigantic step, the invention of the integrated circuit, took place simultaneously at Fairchild and Texas Instruments from 1957 to 1959. So, it has been more than sixty years since the invention of the bipolar transistor, more than fifty years since the invention of the Integrated Circuit (IC) technology and there has been an extraordinary escalation of the electronics industry, with a massive impact on the way people live and work. In the last thirty years or so, by far the area of the industry with most developments has been in the VLSI of silicon chips. A brief chronology of the major milestones in the development of VLSI industry is depicted in Figure 1.1.

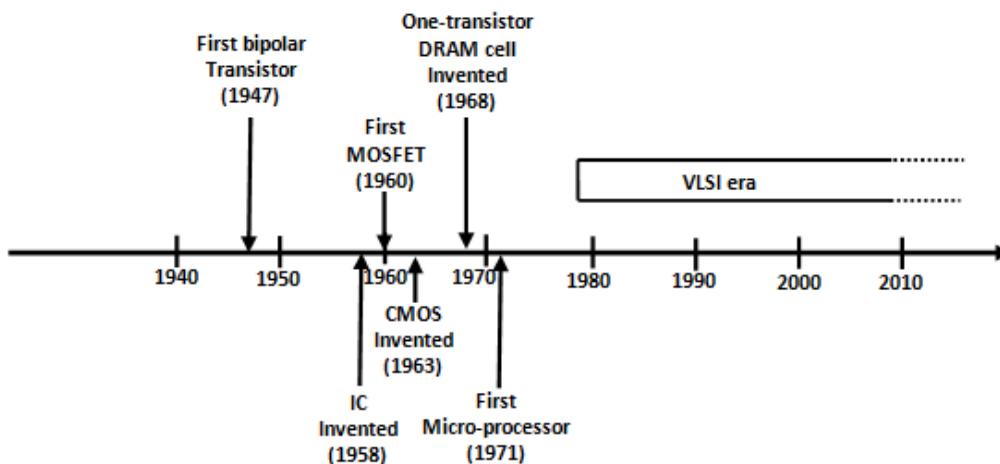


Figure 1.1: A brief chronology of the major milestones in the development of VLSI [65].

In 1965, Gordon Moore observed that Integrated Circuit (IC) complexity evolved exponentially, and manufacturers has been doubling the density of components per Integrated

Circuit at regular intervals, and they would carry on doing so as far as the eye could perceive [47–49]. As an outcome of these observations, in the 1970s a scaling algorithm known as Moore’s Law was developed [46]. It stated that device feature sizes would decrease by a factor of 0.7 every three years. The accuracy of the Moore’s Law in predicting growth in IC complexity had been a reliable method to calculate future trends, as well as settling the pace of innovation and competition. But in the latest technology nodes it appears that Moore’s Law and semiconductor industry are in the middle of a perfect storm [42].

Semiconductor growth is presently limited by overall electronics growth and the ‘smaller the better’ situation is no longer viable. Innovation will surely go on, and go on strong, but not with the traditional scaling of feature sizes; as it is reaching its saturation or close to that.

1.1 Motivation

Until recent years, power has been a second order optimization issue in chip design, only to follow the first order concerns of area, timing and testability. But now, for most System-on-Chip (SoC) designs, power budget is one of the most significant design objectives of a project. Reliability issues are getting increasingly vital for SoC design because of the use of nanometer technology. Exceeding a power budget can be fatal, causing poor reliability, reduced battery life, and increased temperature. Increased temperature decreases mean time to failure exponentially, and increases timing and leakage. It also introduces packaging and cooling challenges. Chip design has four distinctive features:

- I. Computation
- II. Memory
- III. Communication and
- IV. Input/Output

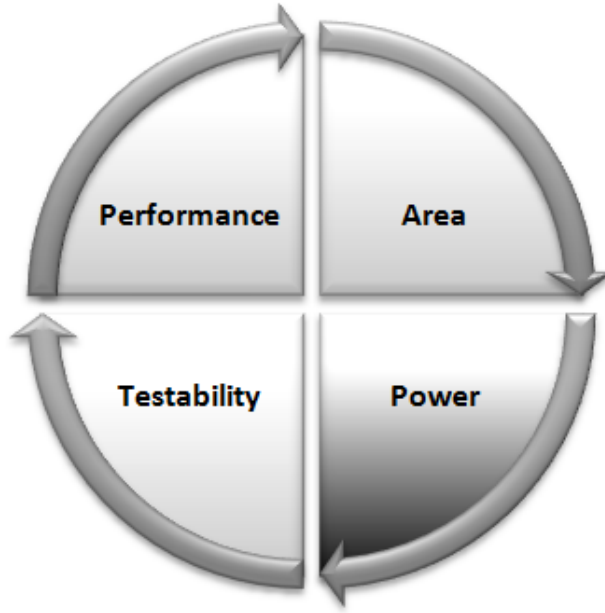


Figure 1.2: Four dimensions of optimization in VLSI design.

For continuing the performance growth, the microprocessor industry has shifted to multi-core scaling by increasing the number of cores per die each generation. Many researchers believe this core scaling will continue into hundreds or maybe thousands of cores per chip [9, 17]. Increased processing power and data intensive applications have attracted attention to the communication aspect of the system. Continuous voltage scaling has decreased the noise margin, making interconnects susceptible to cross talk, power supply noise, process variation, and radiation defects. The design of SoCs is turning out to be increasingly difficult, as adding more and more functionalities are worsening the already complex size, performance, and power consumption constraints. On-chip global communication is required for data and control transfer across various modules on the chip, and it significantly determines the performance of the integrated circuits in current technology. Global and intermediate bus architecture does not follow transistor scaling, and as a result makes long range on-chip data communication challenging in terms of latency, throughput, and power. A difficult challenge at present is the routing complexity and congestion of parallel buses that span over large distances on the chip, connecting various modules placed all around the chip.

Buses not only have to compete with power grid, clocks and other global signals for global resources, the process of boosting their performance by inserting drivers, repeaters and registers makes it considerably area-hungry. The performance enhancing techniques increases power dissipation due to increased capacitances. The power consumed by the interconnect for on-chip global communication now account for a significant fraction of the total power of a system, and this fraction is expected to grow as technology scales further. To address this issue of increased energy consumption, circuit techniques such as low-swing signaling and bit encoding can be used. As switching activity determines the dynamic power dissipation, some methods attempt to reduce the number of transitions on the bus. Techniques like Adaptive Supply Voltage Links are deployed at the system level for energy-efficient on-chip global communication.

1.2 Contribution

Improvement of the overall performance cannot be achieved by a single technology improvement. It is a product of all the technologies from semiconductor to system design. This work focuses on methods for possible reductions of power consumption and area of the bus architecture for on-chip communication. Modern SoC devices need significant amount of data transfer and computing power, which implies the number of on-chip modules will increase, as will the number of on-chip buses connecting them. Due to technology scaling, the delay and power dissipation of the on-chip communication is becoming on the major bottleneck in the current SoC designs. This thesis proposes the design of an on-chip analog bus for replacing the current parallel bus. Reduction in the number of wires saves chip area, and the reduction in total intrinsic wire capacitance consequently reduces bus power consumption. The scheme should also reduce signal interference and crosstalk by eliminating the need of multiple line drivers and buffers. Analog bus can even be useful for short chip-to-chip interconnections in order to reduce pin and trace counts. This analog bus uses digital-to-analog converter (DAC) drivers and analog-to-digital converter (ADC) receivers.

We replace n wires of an n -bit digital bus carrying data between cores with just one (or few) wire(s) carrying analog signal(s) encoding 2^n levels of voltage. Such on-chip communication scheme can potentially save area and power in spite of the additional the DACs and the ADCs used. Appropriate theoretical and experimental work has been done to validate the significant power saving that can be achieved by implementing this method. We have carried out simulated experiments by evaluating power consumption of a single wire with DAC/ADC encoding in comparison to an n -bit digital bus of a large system. SPICE simulation for an ideal case shows that, the ratio of bus power consumed by the proposed analog scheme to a typical digital scheme (without bus encoding or differential signaling) can be given by $P_{\text{analog}}/P_{\text{digital}} = 1/(3n)$. For 500MHz frequency and 1mm intermediate wire line, 4-bit replacement analog bus consumes $16\mu\text{W}$ over $219\mu\text{W}$ in parallel bus. Whereas, the 8-bit replacement bus consumes $18\mu\text{W}$ over the $470\mu\text{W}$ power consumption in the 8-bit parallel bus.

1.3 Problem Statement

The objective of this work is to develop a low power analog bus for on-chip communication to replace existing parallel digital bus.

1.4 Organization

The thesis is organized as follows:

Chapter 2 introduces the reader to sources of power consumption in CMOS design and various existing low power design techniques.

Chapter 3 explains on-chip communication and bottleneck of the area in detail.

Chapter 4 discusses previous contributions in the field of low power on-chip communication. The main focus is on the SerDes approach.

Chapter 5 introduces the concept of analog bus for digital on-chip communication.

Chapter 6 explains the proposed scheme with the results obtained during the experimental implementation.

Chapter 7 discusses the theory of analog-to-digital and digital-to-analog converters.

Chapter 8 concludes the thesis with challenges of the proposition and suggestions for future research.

Chapter 2

Overview: Power

Power dissipation is one of the most important factors for the choice of technology in VLSI design. According to Pollack's Rule, which states each technology generation doubles the number of transistor on a chip which enables that performance increase is roughly proportional to square root of increase in complexity [9, 10], the scale of integration depends on the increased device and power density. Designers pay special attention to apply power reduction techniques as the maximum power can limit the scale of integration. Power reduction techniques focus on the total power for both active and standby modes of the circuit. The total power in a design consists of dynamic power and static power. The components of power consumption in integrated circuits, consisting of registers, control, data path logic, clock tree, memory etc., are design and application dependent [12, 25, 27, 59].

2.1 Power Dissipation

The total power consumption of a CMOS circuit is

$$\mathbf{P}_{\text{Total}} = \mathbf{P}_{\text{Dynamic}} + \mathbf{P}_{\text{Leakage}} + \mathbf{P}_{\text{Short-circuit}}$$

Where,

P_{Dynamic} = Dynamic switching power dissipated while charging or discharging the parasitic capacitances during a node voltage transition.

P_{Leakage} = Combination of all the sub-threshold leakage power due to the non-ideal off-state characteristics of the MOSFET switches, and the gate leakage power caused by carrier tunneling through the thin gate oxides.

$P_{\text{Short-circuit}}$ = Transitory power dissipated during an input signal transition when both the pull-up and the pull-down networks of a CMOS gate are simultaneously on.

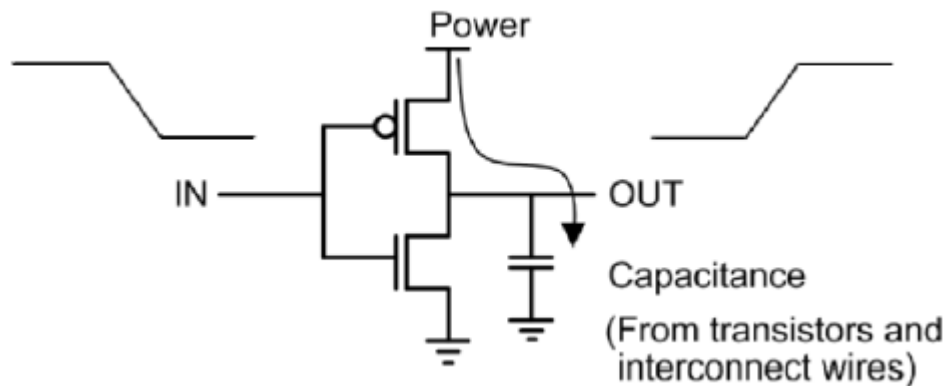


Figure 2.1: Switching Power [27].

2.1.1 Dynamic Power

Dynamic power is primarily due to switching capacitances and short circuit power. The primary source of dynamic power consumption is switching power, which is the power required to charge and discharge the output capacitance on a gate (Figure 2.1). Glitches present in the signals increase switching activity by 15% to 20% [12]. The switching power of a single gate can be expressed as

$$P_D = \alpha f_s C_L V_{DD} V_{swing}$$

Where, α is the switching activity, f_s is the operation frequency, C_L is the load capacitance, V_{DD} is the supply voltage, and V_{swing} is the voltage swing.

Internal power also contributes to dynamic power. Internal power consists of the short circuit current that arises when both the NMOS and PMOS transistors are on, and also the current required for charging the internal capacitances [12, 27, 56].

The short circuit power occurs for a short time during each transition, so the overall dynamic power is dominated by switching power.

Switching power is not a function of transistor size, but rather a function of switching activity and load capacitance, thus it is data dependent. Methods of reducing active power often focus on reducing V_{DD} , as dynamic power depends on V_{DD} quadratically. Measures are also taken to reduce the capacitance and the wire lengths [12, 27, 56].

Short-Circuit Power

In static CMOS circuits, between transitions of the input signals, due to non-zero rise and fall times of the input signals, for a certain small period of time both the pull-up and pull-down network transistors are simultaneously on, thereby forming a DC current path between the power supply and ground (Figure 2.2). The DC current in the circuit during this input signal transient is called the short-circuit current. Short-circuit current is a function of the rise/fall times of the input and output signals and the output load. The short-circuit current is significant if the rise and fall times of the input signals are considerably larger than the output rise and fall times, because the short circuit current path has the opportunity to exist for a longer period of time [35].

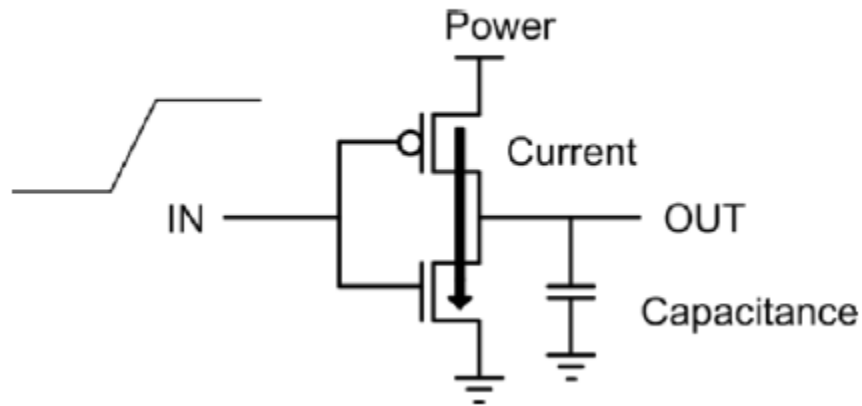


Figure 2.2: Short-Circuit Power [27].

Short-circuit power which is due to the nonzero rise and fall time of input waveforms, which contributes to less than 10% of the total dynamic power. Short-circuit power can be reduced by matching input and output rise and fall times.

2.1.2 Static Power

Static power, also known as leakage power, is related to the requirement of sustaining the logic values of circuit nodes between switching events. Static power dissipation is generally due to current leakage mechanisms (even in off state) within the circuit and does not contribute to any computation. A transistor switch is fundamentally a resistive/capacitive network between the power supply and ground. Current is drawn from the power supply, even when a transistor operates in the cut-off region, due to the non-ideal off-state characteristics (a finite resistance) of a transistor. The leakage currents are dominated by weak inversion and reverse biased pn junction diode currents in long channel devices [35].

Leakage can contribute a large portion of the average power consumption for low performance applications, particularly when a chip has long idle modes without being fully off [12, 27, 57].

In today's technology, leakage can account for 10% to 30% of the total power when a chip is active. Unfortunately, as CMOS technology scaling proceeds, mechanisms that cause leakage are becoming worse. Static power dissipation plays a vital role in determining how long and far Moore's Law can continue unabated.

There are four main sources of leakage currents in a CMOS gate: Sub-threshold Leakage, Gate Leakage, Gate Induced Drain Leakage and Reverse Bias Junction Leakage (Figure 2.3). Sub-threshold Leakage is the current that flows from the drain to the source current of a transistor operating in the weak inversion region when a CMOS gate is not turned completely off. The equation can be given as follows:

$$I_{SUB} = \mu C_{ox} V_{th}^2 \frac{W}{L} e^{\frac{V_{GS} - V_T}{nV_{th}}}$$

Where, W and L are width and length of the transistor, V_T = thermal voltage, and n is a function of the device fabrication process that ranges from 1.0 to 2.5. This equation tells us that sub-threshold leakage depends exponentially on the difference between V_{GS} and V_T . Subthreshold leakage increases exponentially with decreasing V_{th} and increasing

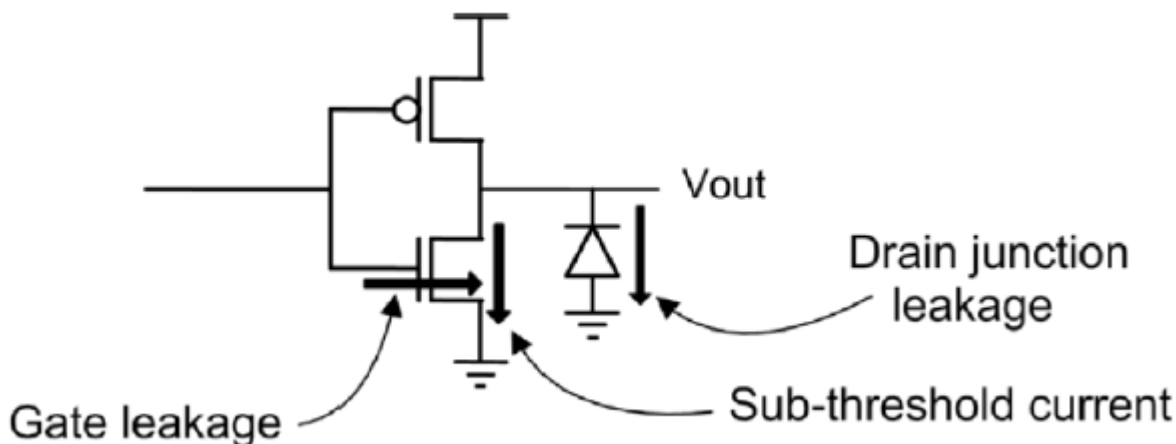


Figure 2.3: Static power [27].

temperature, which complicates the problem of designing low power systems. It is also dependent on transistor channel length in short channel devices [27]. Gate Leakage is the current which flows directly from the gate through the oxide to the substrate due to gate oxide tunneling and hot carrier injection. Leakage current has increased exponentially with reduction in gate oxide thickness. The gate oxide thickness (T_{OX}), which is only a few atoms thick, makes tunneling current substantial. Starting with 90nm, gate leakage can be nearly one-third as much as sub-threshold leakage. High-k dielectric materials are required to keep gate leakage in control [12, 27, 57].

Gate Induced Drain Leakage is the current which flows from the drain to the substrate induced by a high field effect in the MOSFET drain caused by a high V_{DG} .

Reverse Bias Junction Leakage is caused by generation of electron/hole pairs in the depletion regions and minority carrier drift [12, 27, 57].

2.2 Power Reduction Techniques

Until recent times, power was a second order problem in chip design, the first order considerations being cost, area, and timing. Now, for most System-on-Chip (SoC) designs,

Table 2.1: Strategies for low power designs [26].

Design Level	Strategies
Operating System Level	Portioning, Power down
Software level	Regularity, locality, concurrency
Architecture level	Pipelining, Redundancy, data encoding
Circuit /Logic level	Logic styles, transistor sizing and energy recovery
Technology Level	Threshold reduction, multi threshold devices

the power budget is extremely significant. Issues like thermal limits, packaging constraints, battery life, and cooling options are now key factors in the success of a product. Today some of the most powerful microprocessor chips can dissipate an average power density of 50-75 Watts per square centimeter. The power density creates problems with not only packaging and cooling, but also with decreased reliability. Exceeding the power budget is critical to the scheme, as it can cause an unacceptably poor reliability due to excessive power density and make the design fail before the required time.

There is a conflict between reduction and balance of dynamic and static power. Supply voltage is reduced to lower dynamic power and threshold voltage is reduced to sustain performance. But this process raises the leakage current. Technology has moved to a point where both static and dynamic power reduction is important and a balance needs to be struck between the techniques [27]. To optimize the power consumption in VLSI design, designers take various approaches for power management, using diverse strategies at various levels of the design process (Table 2.1).

Some of these power reduction techniques are discussed below in Table 2.2:

Table 2.2: Trade off associated with power management techniques [26]

Power Reduction Technique	Power Benefit	Timing Penalty	Area Penalty	Methodology Impact			
				Architecture	Design	Verification	Implementation
Multi Vt optimization	Medium	Little	Little	Low	Low	None	Low
Clock Gating	Medium	Little	Little	Low	Low	None	Low
Multi supply voltage	Large	Some	Little	High	Medium	Low	Medium
Power Shut off	Huge	Some	Some	High	High	High	High
Dynamic and Adaptive voltage Frequency scaling	Large	Some	Some	High	High	High	High
Substrate Biasing	Large	Some	Some	Medium	None	None	High

2.2.1 Technology Scaling

Technology scaling is the most common optimization method used. If the dimensions, voltages and doping are scaled by a factor α , the electric field configuration in the scaled device will be exactly the same as it was in the larger device but speed increases by the scale factor and the power density remains constant.

In recent technologies the supply voltage has reached 1V. It has imposed physical limitations to scaling as the silicon band-gap energy and built-in potentials of the device remains same with scaling. Threshold voltage scaling with manageable leakage is not further possible due to thermodynamic limitations. To accommodate the slower voltage scaling, electric field is increased by an additional factor, $\epsilon > 1$. As this method reduces reliability and increases power consumption, alternative methods should be chosen to overcome the issue [31, 57].

2.2.2 Transistor Sizing

To reduce junction capacitance and overall gate capacitance, transistor sizing is a significant method. There are several methods to minimize the area of the circuit that reduces power while maintaining performance [56].

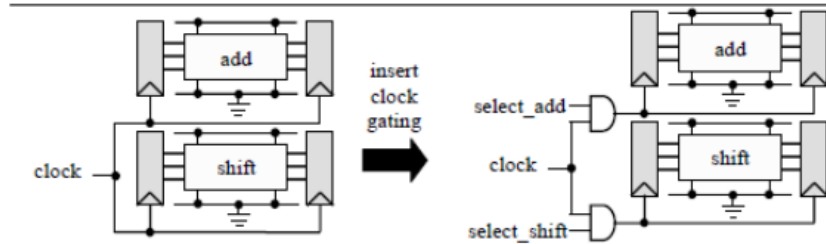


Figure 2.4: Clock gating.

2.2.3 Interconnect Optimization

In every technology scaling, the local interconnect capacitance reduces, but the global interconnect capacitance increases. The increasing die size increases the global interconnect length, as well as the capacitance and delay. For optimizing interconnect power, optimum width, height and spacing of wires are used. The research done in this thesis contributes to this issue also. A significant amount of power can be saved by interconnect optimization [56].

2.2.4 Clock Gating

For any general purpose microprocessor, only a small portion of the circuit is active at a certain time. Turning off the idle portion of the circuit is an effective way to save dynamic power consumption (Figure 2.4). The clock has the highest toggle rate and consumes a significant portion of the total dynamic power. The clock gating approach, where the clock is turned off when not required, can save a significant amount of power without changing any logic function of the circuit [12, 27, 56].

2.2.5 Power Gating

While clock gating reduces dynamic power, power gating reduces static leakage. Here, power rails are disconnected when transistors are in idle mode. This method consumes power, so it is worthy only when a unit is idle for a sufficient number of clock cycles [12].

2.2.6 Supply Voltage and Threshold Voltage Scaling

Reducing supply voltage reduces dynamic power as well as short circuit power. Delay increases with reduced supply voltage; as a result, threshold voltage also has to be reduced. But reducing threshold voltage increases leakage current. A tradeoff has to be made among performance, dynamic power and static power [12,27].

2.2.7 Multi-Voltage Design

Voltage scaling also increases the delay of the gates in the design. For System on Chip design, different blocks have different constraints and performance objective. The block which does not need to run particularly fast can have a lower supply voltage than the speed critical block. This method is called multi-voltage design [20,27]. Some methods of multi-voltage design are: Static Voltage Scaling (SVS), Multi-level Voltage Scaling (MVS), Dynamic Voltage and Frequency Scaling (DVFS), and Adaptive Voltage Scaling (AVS) [27].

2.2.8 Variable Supply and Threshold Voltages

To meet circuit timing constraints, high supply voltage and low threshold voltage are necessary. But low supply voltage reduces dynamic power, and high threshold voltage reduces leakage power. To reduce overall power and meeting timing constraints, high V_{DD} /low V_{th} is used in critical paths, and low V_{DD} /high V_{th} is used where sufficient timing slack is available [12].

2.2.9 Technology Mapping

Logic can be implemented by different combinations of cells. In technology mapping, a logic netlist is mapped to a standard cell library within a given technology. Nets with high activity can be assigned with lower input capacitance pins. Swinging activity can be reduced by refactoring, whereas balancing path delay can reduce glitches [12].

2.2.10 Floorplanning, Cell Placement and Wire Routing

A significant portion of total capacitance in a design is made of wire capacitances. Capacitance of a wire depends on its length, and wire lengths in a chip greatly depend on quality of global wire routing, floor planning and cell placement. Additional buffers to drive long wires also contribute to extra power consumption. Several techniques are applied to reduce the power consumption due to long global wire length [12]. *The technique discussed in this thesis also reduces wire routing.*

Chapter 3

On-Chip Communication

There is no turning back from the era of multi-million gate chips that the semiconductor industry has entered. Traditionally, the design and development of the System on Chip (SoC) technology focused on the computational aspects of the problem. But as the number of elements on a single chip and their performance requirements continued to increase, computation-based design shifted to communication-based design. Now-a-days, the communication architecture plays a key role in the area, performance, and energy consumption of the overall system [36, 44].

The System-on-chip (SoC) approach enables an increasing number of IP cores to be integrated on a single chip. A large number of different kinds of blocks of the size of a few hundred thousand gates comprise the computational resources. For such a complex design, the communication architecture is vital and has to be efficient [34, 44]. Conventionally, on-chip communication schemes are of two types - point-to-point (P2P) and bus-based communication architecture. An SoC bus architecture is shown in Figure 3.1.

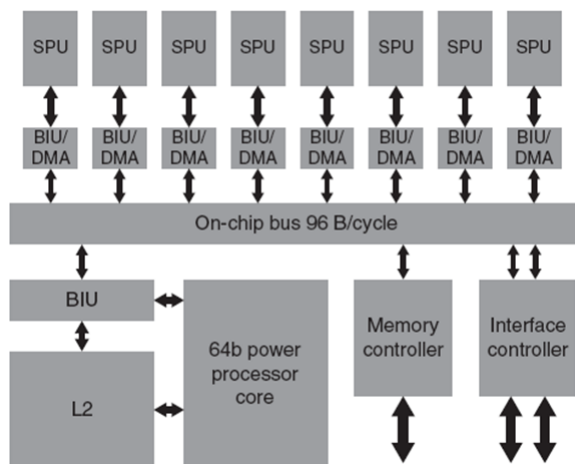


Figure 3.1: IBM Cell ring bus communication architecture [53].

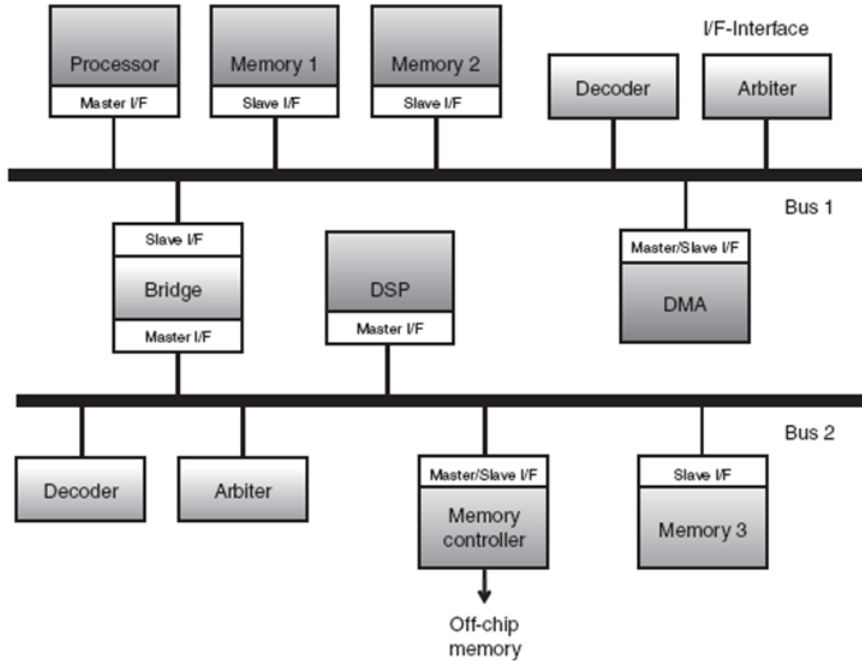


Figure 3.2: Bus structure [53].

3.1 Bus Architecture

A bus is a collection of signals (wires) that connects two or more IP components for the purpose of data communication. On-chip communication is mostly implemented using bus architecture in SoC designs. Figure 3.2 shows a typical bus system, where a variety of devices are tied to the bus for communicating between each other. Use of standard internal bus design around particular modules facilitates design reuse. The performance of the SoC design depends greatly on the efficiency of the bus structure [53].

3.2 Bus Topology

The bus architecture topologies can be classified as:

Shared bus. The simplest bus architecture commonly found in SoCs is shared bus, where several master and slave devices can be connected. Bus arbiter examines requests from the master interfaces periodically and grants access to an arbiter master according to bus protocol specification. The bus bandwidth can be limited by increased load on global bus lines.

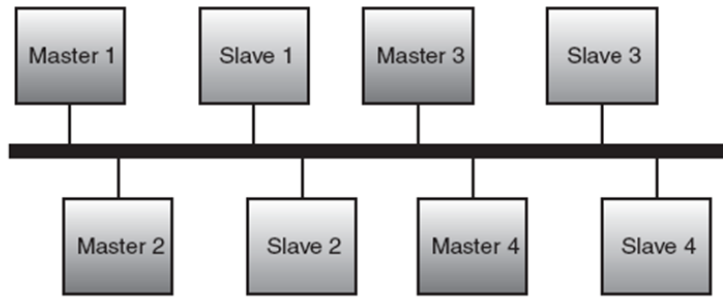


Figure 3.3: Shared bus [53].

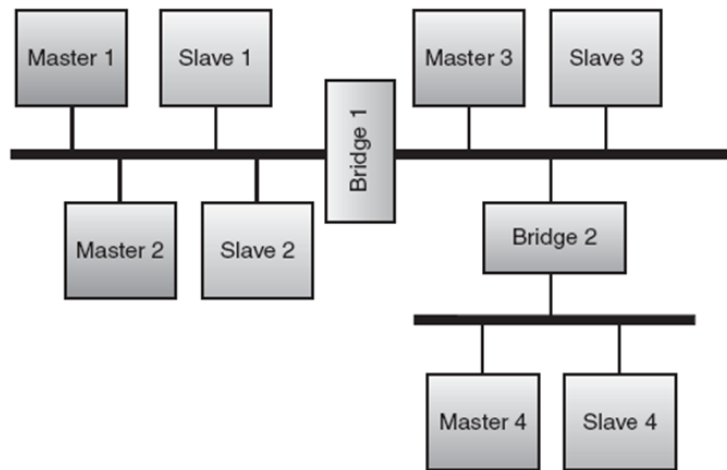


Figure 3.4: Hierarchical bus [53].

Advantages of Shared bus are simple topology along with low area cost, efficient implementation. Large load per data line, delay, and energy consumptions are the disadvantages of shared bus that limits its bandwidth. Low-voltage swing signaling techniques can overcome these disadvantages [44]. Figure 3.3 illustrated a shared bus.

Hierarchical Bus. In a hierarchical bus, several shared buses are connected by bridges to form a hierarchy. Components are placed in the hierarchy according to their performance level.. Hence, low and high performance components are placed in low and high performance bus. AMBA bus and CoreConnect bus are examples of this bus architecture. Hierarchical bus architecture offer larger throughput than shared bus, as it has decreased load per bus and the potential of transactions proceeding in parallel on different busses. Communications can proceed in a pipelined manner. However, additional overhead of transactions across

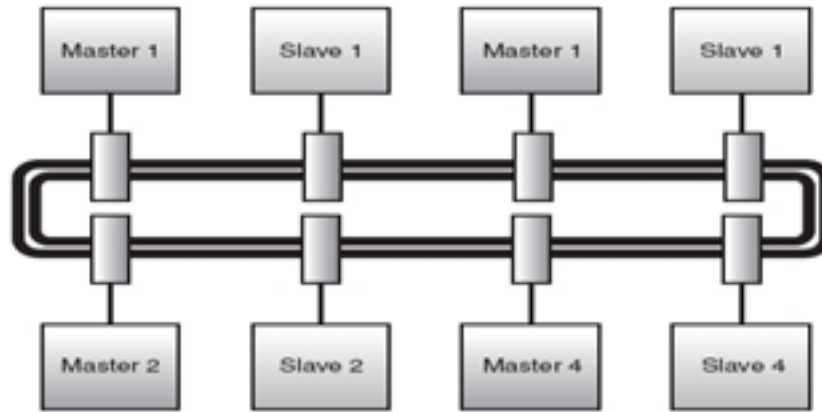


Figure 3.5: Ring bus [53].

the bridge during the transfer may make the bus inaccessible to other components [44]. Figure 3.4 illustrated a hierarchical bus.

Ring Bus. In the Ring Bus architecture each node component communicates using a ring interface implemented by a token pass protocol. Ring based bus is widely used in numerous architectures like network processor and ATM switches [44]. Figure 3.5 illustrates a ring bus.

Other Architectures. Some other bus architectures are Split Bus, Full Crossbar bus, Partial Crossbar Bus, tri-state buffer based bus, etc., as illustrated in Figures 3.6 through 3.9.

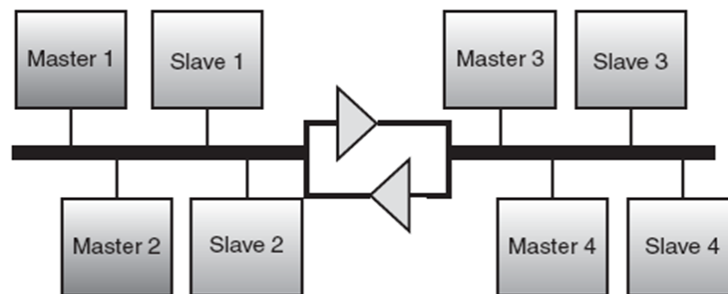


Figure 3.6: Split bus [53].

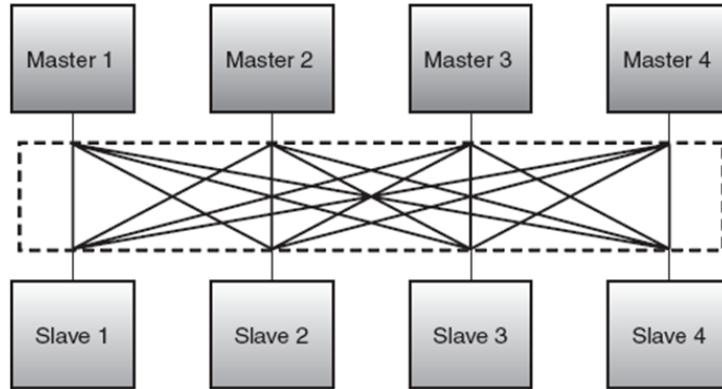


Figure 3.7: Crossbar bus [53].

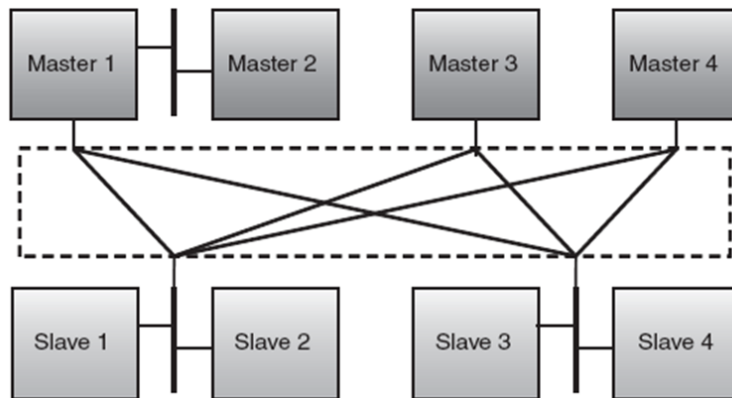


Figure 3.8: Partial crossbar/matrix bus [53].

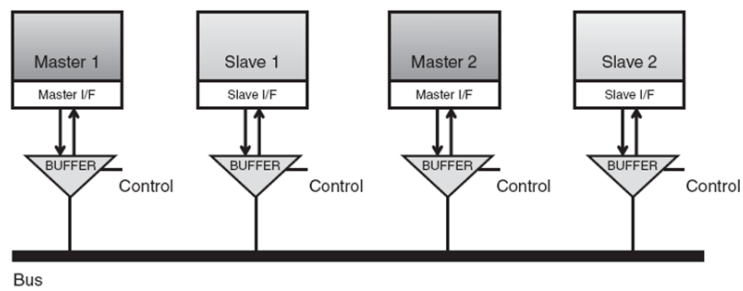


Figure 3.9: Tristate buffer bus [53].

3.3 Issues With Parallel Bus

Computation-based design shifted to communication-based design as communication has become the most critical aspect of system performance and cost. Whenever a system is imagined, it includes a bus system including various devices coupled with it. Communication architecture consisting of wires, repeaters, bus components can consume up to 50% of the total chip power [53]. Design, customization, exploration, verification and implementation of the communication architecture take up a significant portion of the system design cycle. A number of trends have enforced evolutions of systems architectures resulting in evolutions of the required buses. These trends consist of application convergence, integration of IP blocks in single chip, process evolution, time to market pressure, etc. [2, 14]. Parallel buses are a large number of wires bundled together that enable data to be transmitted in parallel [53]. Key issues of the bus architecture design are power consumption, performance, design time reduction, ease-of-use, and silicon efficiency. Complexities of parallel bus architecture are explained below [2, 14, 28, 53].

3.3.1 Routing Complexity

Bus architecture has to compete for global resources with clock, power grid and other global signals. The length of interconnect is increasing due to increasing number of modules that span large distances on the system on chip. The number of buses required is also increasing as the number of IP cores are increasing. As a result, routing on-chip parallel bus is getting complicated due to increasing congestion [2, 8, 14, 28].

3.3.2 Area

Besides routing complexity, a parallel bus also occupies large silicon area, as a number of drivers, repeaters and registers are inserted along with the interconnect. The use of wider metal pitch and protective shield to reduce coupling are also area consuming [2, 8, 14, 28].

3.3.3 Power Dissipation

Integrated circuits designed with battery constraints in mind makes energy efficient global communication techniques necessary. Every attached additional element in the circuit to construct a bus architecture adds to the overall capacitance. The power consumed by the bus architecture is a significant fraction of the total power consumption of the integrated circuit. Increasing number of cores creates increased number of bus lines, which correspond to increased capacitance. Furthermore fringe capacitance increases as interconnects are getting closer. The repeaters, buffers, etc. inserted to improve performance and throughput also consume lots of energy [28].

3.3.4 Performance

Bandwidth is limited, but shared by all elements. Skew and Jitter on the parallel bus make synchronization complicated, and therefore leads to bandwidth limitations. As technology scales, the RC delay of the interconnects gets worse. To counter this, more repeaters and buffers are inserted, which on other hand increases power consumption due to additional elements. Another method to reduce delay is to increase the pitch. This method reduces the delay, but the raise in area is significant [8, 28].

3.3.5 Signal Integrity and Crosstalk

Increased package density and feature size reduction causes complexity in on-chip communication. The most important signal integrity problems are crosstalk, signal skew, overshoot, and reflection. The crosstalk created in a parallel bus not only serves as a conductor of electrons but also introduces additional resistance, capacitance, and inductance. Crosstalk induces delay and noise too. Crosstalk between neighboring lines in a parallel bus creates data-dependent signal delay worse limiting the transmission bandwidth [28, 53].

3.4 Possible Solutions

Bus architectures cannot directly trail process and system architecture evolution. The architectures have to balance among the various driving forces. A prominent technique to reduce parallel bus issues in an inter-core bus communication is reducing the number of transitions occurring on each of the bus lines by bus encoding procedures [7]. This reduces the effective activity on the lines, and the number of lines that need to be run between two cores. Alternate schemes for power reduction include low voltage and differential signaling [23]; all of which try to limit the signal swing on the bit lines, thereby reducing power. Another solution is replacement of parallel buses with an on-chip serial link [29].

Chapter 4

Previous Work

Point-to-point (P2P) and bus-based communication architectures are the two types of on-chip communication schemes widely considered. Intellectual property (IP) cores communicate with each other through dedicated channels in P2P communication, providing utmost performance. This architecture however experiences scalability issues because of complexity, design effort and cost. Bus architecture connects multiple IP cores, reducing the complexity of dedicated communication. Still, bus based architecture also suffers from requirements of scalability in terms of performance and power efficiency [36].

A prominent technique to reduce power in an inter-core bus communication is reducing the number of transitions occurring on each of the bus lines by bus encoding procedures [7]. This reduces the effective activity on the lines, and the number of lines that need to be run between two cores. Alternate schemes for power reduction include low voltage and differential signaling [23]; all of which try to limit the signal swing on the bit lines, thereby reduce power. Techniques such as Adaptive Supply Voltage Links are employed at the system level for energy-efficient on-chip global communication. Another solution to the problems of parallel buses is to replace it with an on-chip serial link [29].

4.1 NOC

The network-on-chip (NOC) methodology is a solution to the design productivity problems in communication centric on-chip communication. The NOC architecture is an $m \times n$ mesh of switches and resources, placed on the slots formed by the switches [22]. NOC communication infrastructure connects the resources via a network of switches which communicate with each other using addressed data packets routed to their destination by the

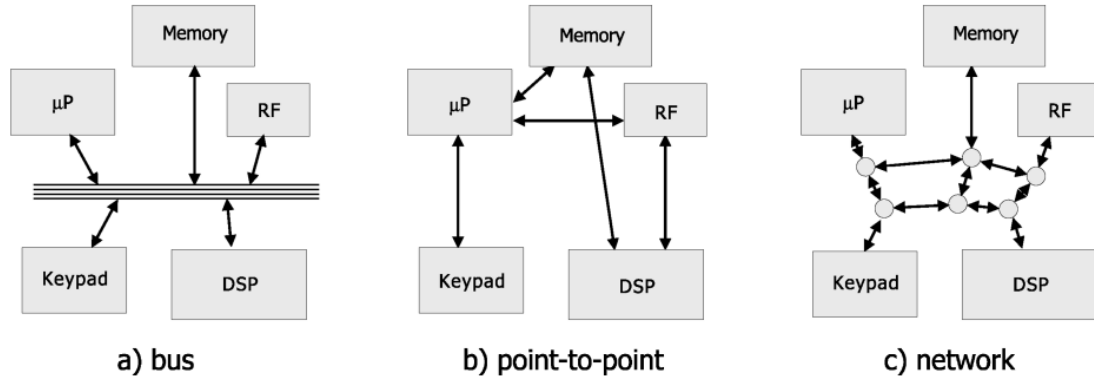


Figure 4.1: Various communication architectures [8].

switch fabric. Communication among IP cores is carried out by generating and forwarding packets through the network structure [8, 36]. Here, the hardware resources are developed independently as standalone blocks, and the NOC is created by connecting the blocks in the network. The configurable network, being a flexible platform, can be modified as per need of the workload, while maintaining the generality of the application. [8, 36]

Figure 4.1 shows the structures of bus, P2P and network on chip architectures [36]. NOC architecture has various advantages of scalability, design reuse and predictability factor. A large number of IP cores can be connected without using global wires, as communication can be achieved by routing packets. The approach provides highly scalable communication architecture. NOC offers great potential for reuse of network and IP cores complying with the network that can be reused in various applications. The architecture is structured, which facilitates controlled and optimized electrical parameters [8]. Multi-route and redundancy is possible in this architecture. Disadvantages of NOC are area and speed overheads. There is an area overhead because of the switches used and because the fixed wire layout is not always optimal. Internal network in the architecture with packaging, routing and switching may add latency in the system. Synchronization is imperative in this system [34].

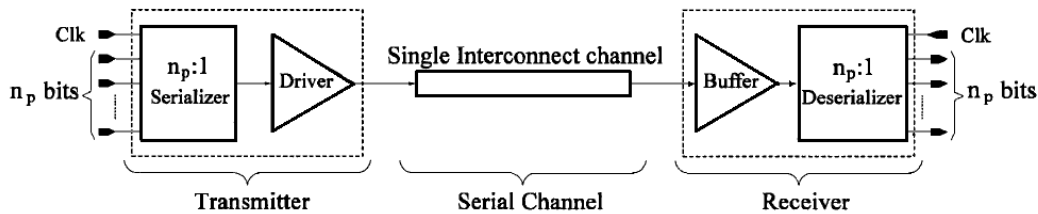


Figure 4.2: SerDes [28].

4.2 SerDes

A promising solution for on-chip communication that may replace parallel buses is an on-chip serial link. A parallel link comprises n wires that can carry n bits of data simultaneously through the link. Serializer/De-serializer (SerDes) is a widely used technique for replacing multiple lines of an on-chip bus with a single on-chip line to achieve high speed serial communication. It is illustrated in Figures 4.2 and 4.3. In this architecture, n parallel data bits are serialized on the transmitter side. The data transfer takes place at a speed which is n -times higher than the data rate of the parallel data. On the receiver side, the data have to be de-serialized to reproduce the n -bit parallel word. In general, n wires can be compressed into m wires where $m < n$.

Serial link can overcome various problems of parallel buses, especially wiring and routing complexity. Serial links are area efficient because of the reduction in numbers of line drivers and repeaters. This becomes possible because of the reduction in the number of interconnects in the on-chip communication [16, 21, 28, 29, 32, 38, 50].

4.2.1 Construction of SerDes

The structure of SerDes consists of three primary components:

1. Transmitter
2. Transport channel
3. Receiver

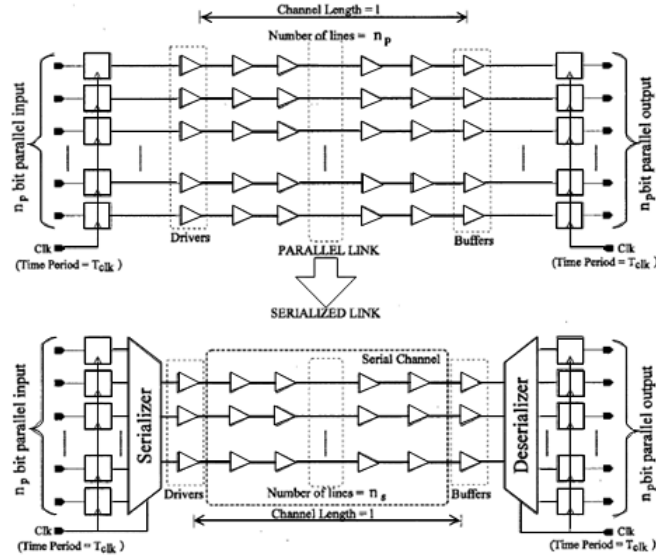


Figure 4.3: SerDes structure [28].

The transmitter transforms the low speed parallel data to high speed serial data. The signal is then transmitted through a serial channel. The receiver transforms the signal back to parallel data by de-multiplexing the data. The function of the transmitter is to recognize a data word of a specified width, serialize it and drive the data onto a channel. The width of the word is a function of the bandwidth of the input and the output. The receiver extracts a clock signal from the incoming signal in order to accurately sample the data from the signal. Though the serial link has several benefits, it has a more complex design than the parallel bus. Issues arise as serial data has to be shifted from and to parallel data for on-chip global communication. If a single interconnect is insufficient to convey the parallel data then multiple interconnects are needed [21, 28, 29, 32, 37, 50].

It is required to find a method for serializing the parallel bus signals in such a way that, increase of signal transition frequency is prevented to suppress an increase in power consumption. If the transition frequency is not controlled by some method, the power consumption of a serial bus becomes much higher than a parallel bus [21, 37]

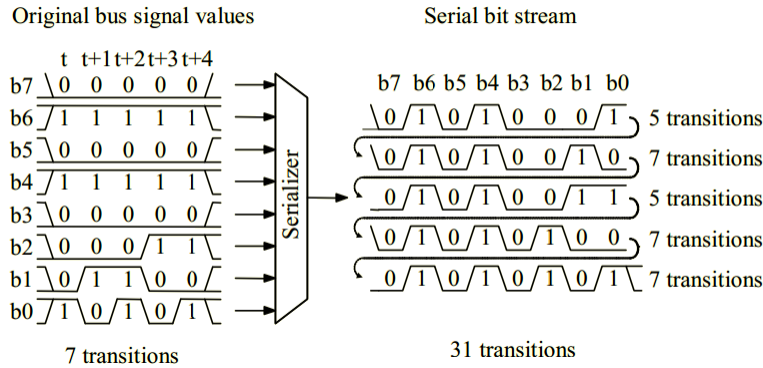


Figure 4.4: Example of serialization [37].

The serial channel has to simultaneously reduce number of interconnects and provide required bandwidth. To compensate the loss of data rate due to serialization, high throughput-signaling scheme is needed [29, 55].

4.2.2 SERDES Approaches

SerDes devices conform to several basic architectures, namely, Parallel Clock SerDes, Embedded Bit SerDes, 8b/10b SerDes and Bit Interleaving SerDes. Table 4.1 shows the pros

Table 4.1: Comparison overview of advantages/disadvantages of SerDes architectures [39].

Technology	Advantages	Disadvantages
Parallel Clock SerDes	Serializes wide buses Low cost Automatic transmitter/receiver sync	More pairs/wires needed Tight pair-to-pair skew requirements
Embedded Bit SerDes	10- and 18- bit widths available Lock to random data capability Relaxed clocking requirements	No inherent DC balance Not well suited for AC coupled or fiber applications
8b/10b SerDes	DC balance coding Works well in AC-coupled and fiber Environments Widely available	Byte-oriented Tight clocking requirements Requires comma for sync
Bit Interleaving SerDes	Aggregates existing slower serial Streams SONET/SDH-compliant versions	High speed design challenges Higher cost

and cons of several SerDes applications. The preference in selecting serializer/deserializer (SerDes) techniques has a big impact on cost and performance of the design.

There are some approaches proposed by researchers to compensate the power consumption of SerDes Technique. *Silent* is a serialized low-energy transmission coding technique to minimize the transmission energy. This approach is effective only when the traces are uniform. This coding technique, working by the means of the data correlation between successive data words, reduces the number of transitions on serial wires [38].

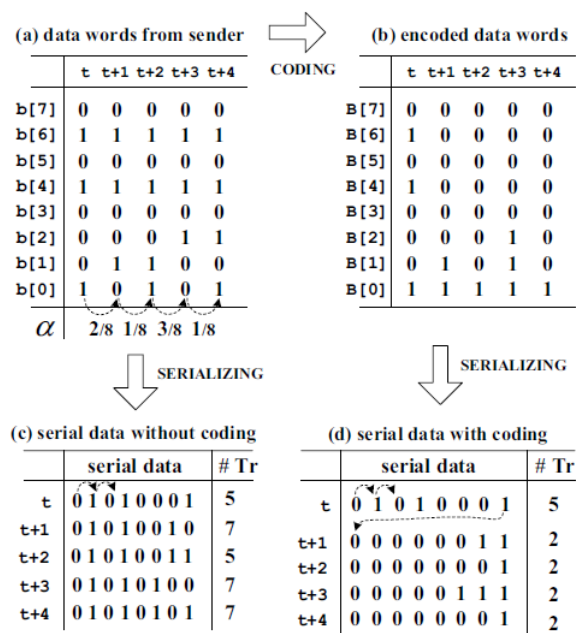


Figure 4.5: The *Silent* scheme [38].

Another technique presented in [28] is a serialized technique based on bit ordering on a serial link for switching activity reduction, called *LOUD*, to perform bit ordering using known data traces by building a graph and solving it using a branch and bound technique.

A technique for reducing bus power consumption without decreasing throughput focuses on reducing coupling capacitance of the on-chip serial bus [21].

Chapter 5

Analog Bus

As mentioned earlier, until recently, power was a second order issue in chip design, following the first order concerns of: cost, area, timing and testability. However, for most System-on-Chip (SoC) designs, the power budget is now one of the most significant design objectives of a project. But power reduction is not achieved through a single technological improvement; it is a product of the overall improvement of the technology. When power consumption is decomposed between the functional blocks and the communication paths between them, the second has become a principal component, as the feature size is reduced down to the deep sub-micron region.

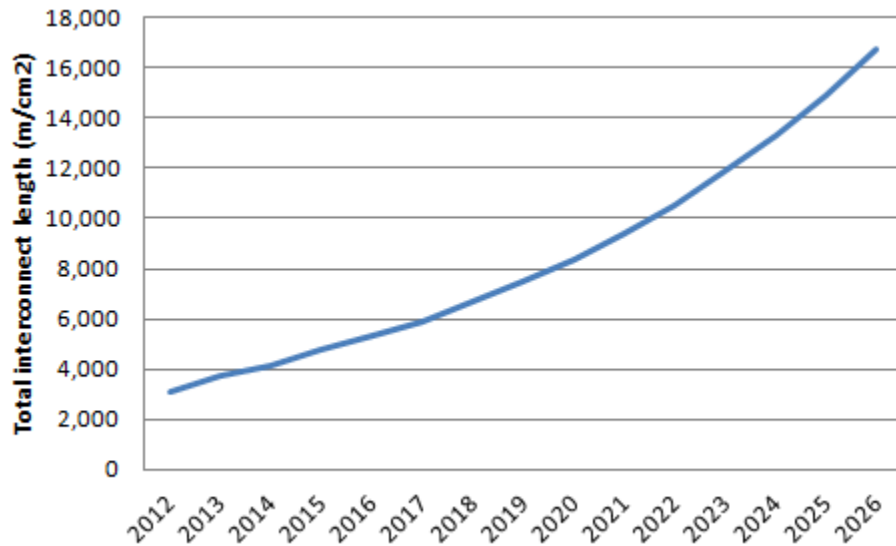


Figure 5.1: Total interconnect length (m/cm²) - Metal 1 and five intermediate levels, active wiring only [60].

There is lack of literature on designing interconnect framework in relation with the multiple core in the die [33]. The conventional design flow was mostly logic based which

emphasized on the design and optimization of logic and design where interconnect layout was done very late in the overall design. But now as technology has moved to nanometer dimension and gigahertz clock frequency; interconnect design plays a dominating role in determining performance, power, cost and reliability [13]. There are difficult challenges in designing a large SoC, e.g., one containing many processor cores, include hardware area, power dissipation, routing complexity, congestion and latency of the communication network.

Figure 5.1 shows the ITRS prediction of total interconnect length from 2012 to 2026 [60]. At present, performance and efficiency of SoC designs depend significantly on the on-chip global communication across various modules on the chip. On-chip communication is mostly implemented using a bus architecture that runs long distances, covering significant area of the integrated circuit.

5.1 Concept

Analysis shows that interconnect power can be over 50% of the dynamic power, over 90% of the interconnect power is consumed by only 10% of the interconnections [43,45]. Often these interconnects tend to be multiple bit lines, also known as bus, running between two cores.

Optimization of interconnect power is an important VLSI design challenge. Because, the RC delay driving long wires makes the chip slow, and large switching capacitance makes power consumption large. The power consumption for low-swing signaling depends both on voltage supply V_{DD} and voltage swing V_{swing} . Rather than waiting for a full swing, low-swing signaling improves performance by sensing when a wire swing through some small V_{swing} [68].

Every time the wire is charged and discharged, it transfers charge, $Q = CV_{swing}$. In a case where the effective switching frequency of the wire is αf , the average current is

$$\mathbf{I}_{avg} = \frac{1}{T} \int_0^T \mathbf{i}_{drive}(t) dt = \alpha C \mathbf{V}_{swing}$$

Here, C is the capacitance, V_{swing} is the voltage swing, f is the frequency, α is the activity factor. If there are n -lines in the bus and each of them has similar activity, then the total power consumed by such a bus will be n -times that of a single bit line. Total bus power is the sum of all n lines of the bus. So, power in the bus architecture can be expressed by

$$P_{\text{ParallelBus}} = \sum_{i=1}^n C_i V_{DD} V_{swing,i} f \alpha_i$$

$$P_{\text{ParallelBus}} = V_{DD} f \sum_{i=1}^n C_i V_{swing,i} \alpha_i$$

Power reduction techniques are based on architectural, logic or circuit design methods, decreasing all or some parameters among f , α_i , n or $V_{swing,i}$ [11, 21, 35, 38, 45, 68]. The proposed work focused on possible methods for reduction of power consumption in the VLSI bus system by reducing the number of wires and voltage swing through the use of an analog bus for on-chip digital communication.

5.2 Structure

We evaluate a digital-to-analog and analog-to-digital converter based inter-core communication scheme to significantly reduce the power consumption of multiple bit-line wide buses in multi-core processors and networks-on-chip. The proposed scheme replaces an n -bit wide bus running between cores with a single line, by encoding the information (that was to be carried on the n -bit bus) into 2^n levels of voltages on a single wire. Such a scheme offers the best of the two most prominent low power inter-core communication schemes - bus encoding and differential-low-voltage signaling, by encoding n lines into 1 and keeping the low average signal swing. Reduction in number of wires and in total intrinsic wire capacitance consequently reduces chip area and power consumption. Additional advantages might include the elimination of skew uncertainty due to removal of multiple signal wires, layout and timing verification simplicity, blockage reduction due to reduced number of vias and repeaters. Such bus encoding can also be gainfully employed in test access mechanism

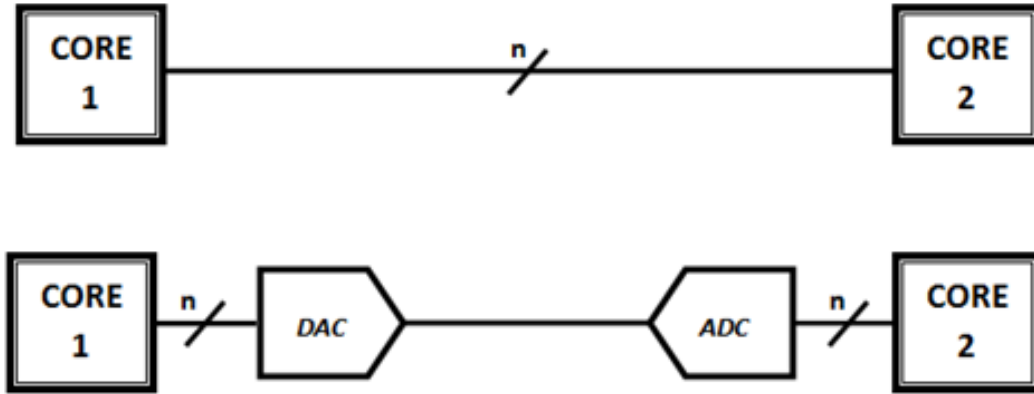


Figure 5.2: Parallel bus and analog bus.

for digital circuits, as it can compress the amount of data to be communicated between the test head and chip, thereby reducing test time.

In our scheme we replace n wires of an n -bit digital bus carrying data between cores with just one (or few) wire(s) carrying analog signal(s) encoding 2^n levels of voltage. For this, the analog bus utilizes digital-to-analog converter (DAC) drivers and analog-to-digital converter (ADC) receivers [63, 64]. Figure 5.2 shows this transformation from n wires (top) to a single wire (bottom).

As mentioned, such a scheme offers the best of both the prominent low power inter-core communication schemes - bus encoding and differential and low-voltage signaling, in that, it offers the ultimate encoding n lines to 1, and average signal swing will be about $V_{DD}/2$. Power consumption of the analog bus will be,

$$P_{\text{AnalogBus}} = V_{DD} f V_{\text{swing}} C \alpha$$

The capacitance and supply voltage remain same but we are reducing number of wires and voltage swing [63, 64].

5.3 Proposition

The analog bus can be used in cases where:

- i. Power consumed by analog bus architecture \leq Power consumed by parallel bus

- ii. The signal can be reproduced without any error

The choice of resolution for substituting the number of lines in digital buses with proposed analog bus depends on two criteria.

- i. Power consumed by ADC and DAC
- ii. Noise margin of the signal line

Corollary 1:

Analog bus is effective only if the power consumed by the analog bus architecture is less than the power consumption of the digital bus.

Explanation: The ratio of power consumed of the typical scheme (without bus encoding or differential signaling) to the proposed scheme can be given by,

$$\frac{P_{\text{ParallelBus}}}{P_{\text{AnalogBus}}} = \frac{V_{DD} f \sum_{i=1}^n C_i V_{\text{swing},i}^{\alpha_i}}{V_{DD} f V_{\text{swing}} C \alpha}$$

For equal supply voltage, frequency and activity factor and capacitance in each line, the ratio will be,

$$\frac{P_{\text{ParallelBus}}}{P_{\text{AnalogBus}}} = \frac{n V_{\text{swing,ParallelBus}}}{V_{\text{swing,AnalogBus}}}$$

Besides saving the wire area we also save power as long as,

$$(P_{\text{DigitalBus}} - P_{\text{AnalogBus}}) \geq (P_{\text{ADC}} + P_{\text{DAC}})$$

Corollary 2

To reproduce the signal in the digital bus without any error, the noise level should be less than half of the resolution of the ADC.

Explanation: Since a single wire will now be carrying a voltage of 2^n levels, ambient noise

Table 5.1: Bit-wise noise tolerance of analog bus.

Number of Bits	Noise Tolerance
4	62.5mV
8	3.9mV
12	0.24mV
16	0.02mV

levels can limit the successful communication between the cores. The noise tolerance of the ADC is a major design consideration, which determines how many digital wires can be replaced by a single analog wire. Table 5.1 shows bitwise representation for how much noise an ADC can tolerate for the device to reproduce the signal back to original data for a supply voltage of 1V.

The power consumption of the ADC/DAC must be cited as a design challenge in the implementation of the analog bus.

5.4 V_{swing}

Assume,

At time t_i , voltage in an analog bus is V_i and at time t_{i+1} the voltage in a analog bus is V_{i+1} . So, the voltage swing will be $= (V_{i+1} - V_i)$. The range of the voltage can be 0 to V_{DD} (Figure: 5.3).

For only two possible cases among all the possible swing variations, $V_{swing} = V_{DD}$. There can be 2^n possible cases where, $V_{i+1} = V_i$.

The total number of possible variations is

$$\begin{aligned}
 &= 1 + 2 + 3 + \dots + (2^n - 1) + 2^n + (2^n - 1) + \dots + 3 + 2 + 1 \\
 &= 2(1 + 2 + 3 + \dots + (2^n - 1)) + 2^n \\
 &= 2\left(\frac{(2^n - 1)(2^n - 1 + 1)}{2}\right) + 2^n \quad \left[\text{Using the formula, } \sum_{k=1}^n k = \frac{n(n+1)}{2}\right] \\
 &= 2^{2n}
 \end{aligned}$$

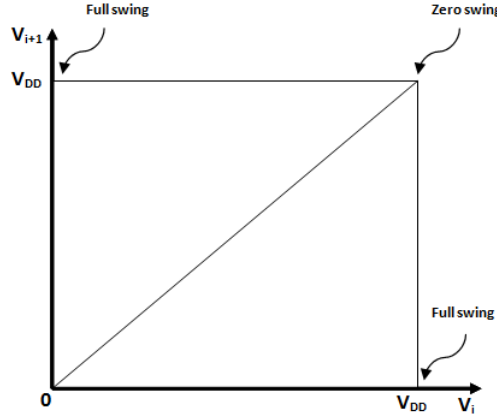


Figure 5.3: V_{swing}

The total possible voltage swing is

$$\begin{aligned}
 &= 1.V_{DD} + 2(V_{DD} - \frac{V_{DD}}{2^{n-1}}) + 3(V_{DD} - \frac{2.V_{DD}}{2^{n-1}}) + \dots + (2^n - 1)(V_{DD} - \frac{(2^n-2).V_{DD}}{2^{n-1}}) + 2^n(0) \\
 &\quad + (2^n - 1)(V_{DD} - \frac{(2^n-2).V_{DD}}{2^{n-1}}) + \dots + 2(V_{DD} - \frac{V_{DD}}{2^{n-1}}) + 1.V_{DD} \\
 &= 2[V_{DD}(1 + 2 + 3 + \dots + (2^n - 1)) - \frac{V_{DD}}{2^{n-1}}(2.1 + 3.2 + 4.3 + \dots + ((2^n - 1).(2^n - 2))] \\
 &= 2.V_{DD}[(1 + 2 + 3 + \dots + (2^n - 1)) - ((\frac{1}{2^{n-1}}).((1^2 + 1) + (2^2 + 2) + \dots + ((2^2 - 2)^2 + (2^2 - 2)))] \\
 &= \frac{1}{3}2^n(2^n + 1)V_{DD} \text{ [Using the formula, } \sum_{k=1}^n k = \frac{n(n+1)}{2} \text{ and } \sum_{k=1}^n k^2 = \frac{n(n+1)(2n-1)}{6}]
 \end{aligned}$$

So, average voltage swing is given by,

$$\frac{1}{3}2^n(2^n + 1)V_{DD}/2^{2n} = \frac{2^n+1}{3.2^n} \cdot V_{DD}$$

5.5 Theoretical Analysis

5.5.1 Voltage

Let us assume a situation where supply voltage is 1V. For a 4-bit data bus, analog bus quantization levels are $2^4 - 1 = 15$ and voltage resolution is $\approx 0.067V = 67mV$. Table 5.2 shows how a random set of digital data is converted into analog representation. The total number of transitions in the parallel bus is 32, each having a voltage-swing of 1V. The analog bus experienced an average voltage swing of 472mV, which is close to the average swing.

t

Table 5.2: Random data patterns and transition analysis.

Parallel Bus Digital Data (Volt)															
1	0	1	0	0	1	0	0	1	1	0	1	1	0	0	0
1	0	0	0	0	1	1	0	0	1	1	0	1	1	0	0
1	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1
1	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0
Converted Analog Bus (Volt)															
1	0.067	0.67	0	0.067	0.933	0.267	0.133	0.6	0.867	0.267	0.67	0.867	0.33	0.13	0.13

5.5.2 Power

Parallel Digital Bus:

Number of bit lines, $n = 4$, frequency, $f = 1GHz$, capacitance, $C = 0.2pF$, activity factor, $\alpha = 0.5$, supply voltage, $V_{DD} = 1V$, Swing Voltage, $V_{swing} = 1V$

$$P_{\text{ParallelBus}} = \sum_{i=1}^n C_i V_{DD} V_{\text{swing},i} f \alpha_i$$

So, the average power consumed by the 4-bit bus will be $400\mu W$.

Serial Digital Bus:

Frequency, $f = 4GHz$, capacitance, $C = 0.2pF$, activity factor, $\alpha = 0.5$, supply voltage, $V_{DD} = 1V$, Swing Voltage, $V_{swing} = 1V$

$$P_{\text{SerialBus}} = V_{DD} f V_{\text{swing}} C \alpha$$

So, the average power consumed by the serial bus (without considering the serializer and deserializer) will be $400\mu W$.

Analog Bus:

Frequency, $f = 1GHz$, capacitance, $C = 0.2pF$, activity factor, $\alpha = 0.5$, supply voltage, $V_{DD} = 1V$, $V_{swing} = 354mV$ (from Vswing calculation)

$$P_{\text{AnalogBus}} = V_{DD} f V_{\text{swing}} C \alpha$$

Table 5.3: Comparison of parallel, serial and analog buses.

Bus type	Number of lines	Number of transitions	Average power consumption
Parallel Bus	4	32	$400\mu W$
Serial Bus	1	34	$400\mu W$
Analog Bus	1	16	$35.4\mu W$

So, the average power consumed by the analog bus (without considering the power used by the DAC and ADC) will be $35.4\mu W$. There is a margin of $364.6\mu W$ for the additional power consumption of the DAC and ADC.

A comparison of the three bus structures in this example of a 4-bit bus is given in Table 5.3.

Chapter 6

Data Conversion

Analog-to-Digital Converter (ADC) and Digital-to-Analog Converter (DAC) are core components of modern signal processing systems. ADC and DAC are the link between analog and digital worlds (Figure 6.1). Digital Signal Processing (DSP) integrated circuits are constantly attaining higher speeds and more processing functions. Sub-micron CMOS technologies now allow Gigahertz range conversion speed. Televisions, digital receiver applications, local area networks, oscilloscopes, medical devices, etc., use different variations of data converters.

The majority of real life signals are continuous in both time and amplitude. The analog-to-digital converters (ADC) convert analog signals to discrete time digitally coded form for digital processing and transmission. The DACs generate an analog signal that represents the same signal as the digital input [66,71]. In most digital signal processing systems, an analog input is taken, which passes through an ADC and is converted some n -bit digital data. On the receiver side, a DAC converts the digital signal back to the original analog signal which is a 2^n level representation of the digital data. With advances of time and technology, new approaches must be adopted. Due to the routing complexity in the communication system in VLSI design, reduction in the number of interconnect wires is essential. To serve that purpose, in this thesis, the common signal conversion method is used in a reverse manner.

6.1 Analog to Digital Converter

There is an enormous demand for low-power, low-voltage ADCs that can be realized in a mainstream deep-submicron CMOS technology. An ADC has two major process blocks for sampling and quantization. ADC bandwidth depends on the Nyquist frequency. The signal is

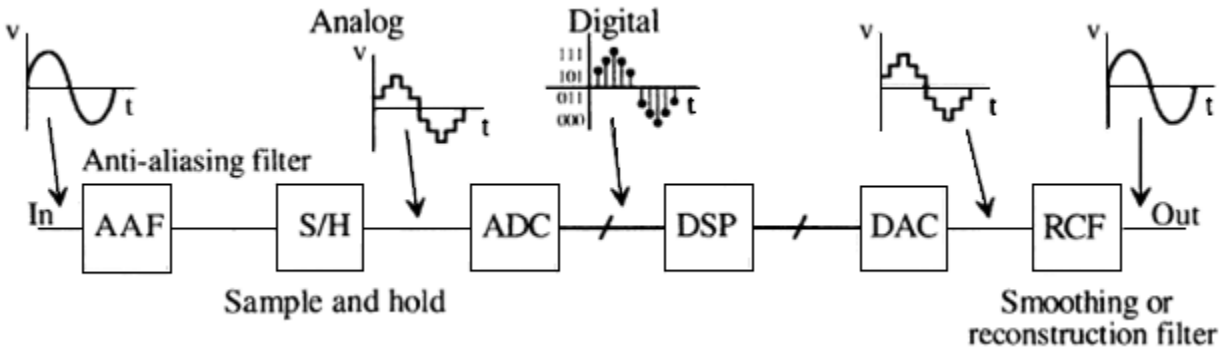


Figure 6.1: Signals resulting from A/D and D/A conversion in a mixed-signal system [5].

first fed to a sample and hold stage to convert the continuous signal into discrete-time signal but keeping the same amplitudes. Most current ADCs have the sample-and-hold function on-chip with a requirement of external sampling clock which initiates the conversion. The quantization is then done by the ADC without any loss, provided the Nyquist rate is met. The continuous signal is mapped to a finite number of discrete values [66, 71].

The relationship between input and output of an ADC depends on the reference value, and the accuracy of the reference is always the limiting factor on the absolute accuracy of an ADC. Most low-power ADCs now have power-saving modes of operation, such as, standby, power-down, and sleep modes [30]. Several ADC architectures [54] are described next:

Flash:

Architecture: Ultra-High Speed (used when power consumption not a primary concern).

Comparators used: $(2n - 1)$ for n -bits.

Conversion Method: increases by a factor of 2 for each bit.

Encoding Method: Thermometer Code Encoding.

Disadvantage: Sparkle codes/metastability, high power consumption, large size, expensive.

Conversion Time: Complexity increase by a factor of 2 for each bit.

Size: $(2^n - 1)$ comparators, die size and power increases exponentially with resolution.

Resolution: Component matching typically limits resolution to 8 bits.

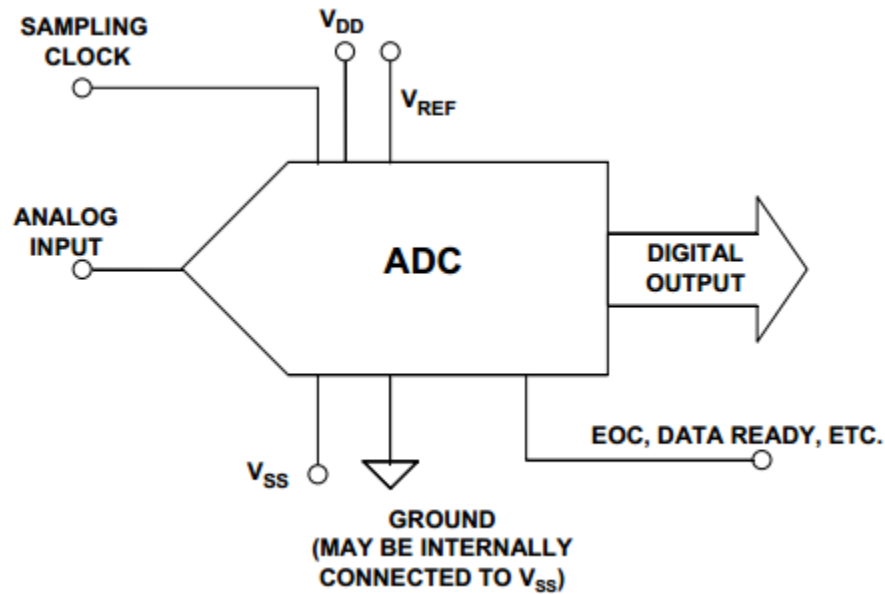


Figure 6.2: Basic ADC structure [30].

Pipeline:

Architecture: High speeds - from a few Msps (Million samples per second) to 100+ Msps, 8 bits to 16 bits, lower power consumption.

Conversion Method: Small parallel structure, each stage works on one to a few bits.

Encoding Method: Thermometer Code Encoding.

Disadvantage: Parallelism increases throughput at the expense of power and latency.

Conversion Time: Increases linearly with increased resolution.

Size: Die size increases linearly with increase in resolution.

Resolution: Component matching requirements double with every bit increase in resolution.

Sigma-Delta:

Architecture: High resolution, low to medium speed, no external precision components.

Conversion Method: Oversampling ADC, 5-Hz - 60Hz rejection programmable data output.

Encoding Method: Over-Sampling Modulator, Digital Decimation Filter.

Disadvantages: Higher order (4th order or higher) multi-bit ADC and multi-bit feedback

DAC Conversion.

Time: A tradeoff between data output rate and noise free resolution.

Size: Core die size does not change substantially with increase in resolution.

Resolution: Component matching requirements double with every bit increase in resolution.

Successive Approximation:

Architecture: Medium to high resolution (8 to 16bit), 5Msps and under, low power, small size.

Conversion Method: Binary search algorithm, internal circuitry runs at higher speed.

Encoding Method: Successive Approximation.

Disadvantages: Speed limited to 5Msps. May require Anti-Aliasing Filter (AAF).

Conversion Time: Increases linearly with increased resolution.

Size: Die size increases linearly with increase in resolution.

Resolution: Component matching requirements double with every bit increase in resolution.

6.2 Digital to Analog Converter

A DAC converts digital data, often a binary code, to an analog domain signal. The output of a DAC can be a voltage or a current. The DAC generates an analog output (signal) that represents the digital input (signal). It can go through a number of non-idealities like component matching, limited output impedance, noise etc. The input is specified by the n-bit words and the output analog representation is converted into 2^n levels. Because of the limited word length or number of bits, the digital input has a limited amplitude resolution. A variety of codes e.g 2's complement, offset binary, grey code, walking one, thermometer, etc., is used for digital to analog conversion [30,58,69]. Some DAC architectures based on [30,69] are described below.

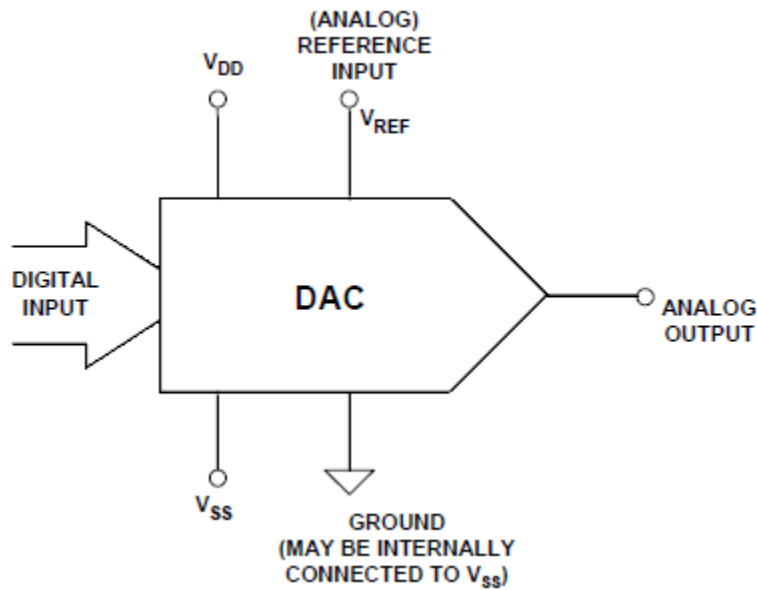


Figure 6.3: Basic DAC structure [30].

Binary-Weighted DAC

The binary-weighted DAC utilizes a number of binary weighted elements like current sources, resistors, or capacitors. The advantage of the binary-weighted DAC is that minimum number of switches and digital encoding circuits are used. The disadvantage of the architecture is, when the number of bits is large, MSB and LSB weight are larger making it prone to mismatch errors and glitches. This makes it hard to manufacture for higher resolutions.

Thermometer-Coded DAC

The thermometer-coded DAC architecture uses a number of equal-size elements with an input that is encoded utilizing thermometer code. It consists of $2n - 1$ switchable current sources connected to an output terminal, which must be close to ground. As the binary code needs to be converted to thermometer code, code converting circuits become large for higher resolutions. Thus the architecture is used for low resolutions of less than or equal to 8 bits.

Direct Encoded DAC

In direct encoded architecture, different amplitude levels are generated directly instead of creating weights. The data bits control the level representation at the DAC output. Each level requires one element and one switch making it element hungry.

R-2R DAC

R-2R resistor ladder is one of the most common DAC structures. It uses resistors of two distinct different values with a ratio of 2:1. For an n-bit DAC, $2n$ resistors are needed. This architecture can be used in two modes - voltage mode and current mode.

6.3 Design considerations

For high accuracy conversions, the accuracy with which the binary weighting of the bit weights is performed can be an important design criterion. The methods for accuracy use combination of matched elements and dynamic methods to improve the passive accuracy. The resolution is limited to 12bits, if resistors and capacitors are employed for matching. Expensive trimming methods can be used to overcome this problem. However, due to time and temperature variation, additional trimming elements can also destroy accuracy. For absolute accuracy, a special system is needed where digital value is converted into an accurate value. This system requires very high frequency. Combinations of passive and active matching components are used to overcome these problems. The key to achieve high speed, high accuracy and high resolution ADCs is the sample-and-hold amplifier. So the analog signal can be sampled and kept constant perfectly during the time of conversion. DAC performance limits as a result of parasitic resistance and capacitance, circuit noise, mismatch between internal references or weights, nonlinear analog circuits, and delay skew between switches [66]. The trade-off between the number of bits and the power consumption is very vital. Among the DAC choices, one particular architecture seems to pave the way of very

high sampling frequency - the current-steering technique [6]. Using survey data from the past years [52], it is observed that power efficiency in ADCs has improved at a rate of 2x every 2 years. This development is partly based on intelligently exploiting the strong points of current technology [51]. Overall, future development in the reduction of power dissipation by the converters will come by utilizing a combination of features involving reduction of complexity in analog sub-circuit, improved system embedding and raw precision [51].

Chapter 7

Evaluation

In order to evaluate the power reduction with the proposed analog bus scheme over a parallel bus, we first examine the power consumed in a case shown in Figure 7.1, without any DAC/ADC elimination, using typical bus capacitances of large chips. Next, we will examine the second case, shown in Figure 7.2 where we replace the parallel lines with a single line using ideal DAC and ADC from [5]. The simulations are done using simulation tool LTspice. LTspice is a high performance SPICE simulation tool with enhancements and models for easing the simulation provided by Linear Technology [4].

7.1 Experiment Setup

Simulations have been done for two cases. First, a 4 line parallel bus has been replaced by a 1-wire analog bus, where both drive the same load circuit, a 2-bit adder. In the second case, an 8-line parallel bus has been replaced by a 1-wire analog bus, where both the setups drive a 4-bit adder.

Table 7.1: Setup

Technology Node	22nm
Metal Layer	4
Intermediate Wire Capacitance	2pF/cm [60]
Supply Voltage	1V
Simulation Tool used	LTspice [4]
Spice models used	Ideal DAC and ADC [5]
Activity Factor	0.5
Frequency	500MHz and 1GHz
Input Data Pattern	Random
Wire length	1mm-5mm

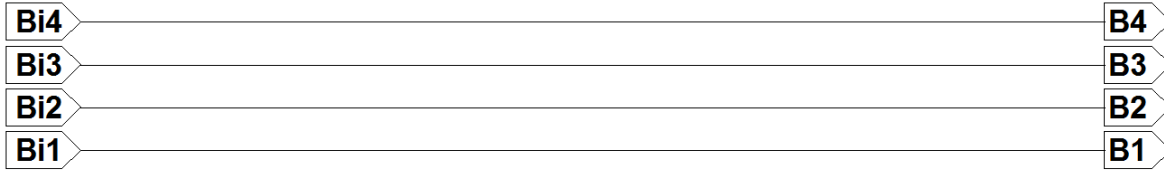


Figure 7.1: 4-bit parallel bus.



Figure 7.2: Analog bus replacing 4-bit parallel bus of Figure 7.1.

7.2 Power Analysis: Replacement of 4-Bit Parallel Bus

For simulation, a 4-bit parallel bus has been replaced by a 1 line digital bus. This setup is shown in Figure 7.3. Here, the analysis has been done for bus lengths of 1mm to 5mm. Capacitance is calculated using the intermediate wire value given in the ITRS roadmap 2012 interconnect manual [24, 60]. The digital input of the DAC is shown in Figure 7.4 and Figure 7.5 shows the DAC output, which is transmitted to the ADC.

Comparison of power consumption for a frequency of 1GHz with bus lengths of 1mm to 5mm (without addition of ADC/DAC power consumption) is given in Table 7.2 and Figure 7.6. The average power consumption per mm for the analog bus is around $33\mu W$.

The average power consumption per mm for each parallel line is $115.8\mu W$ and for a 4-bit bus it is $463\mu W$.

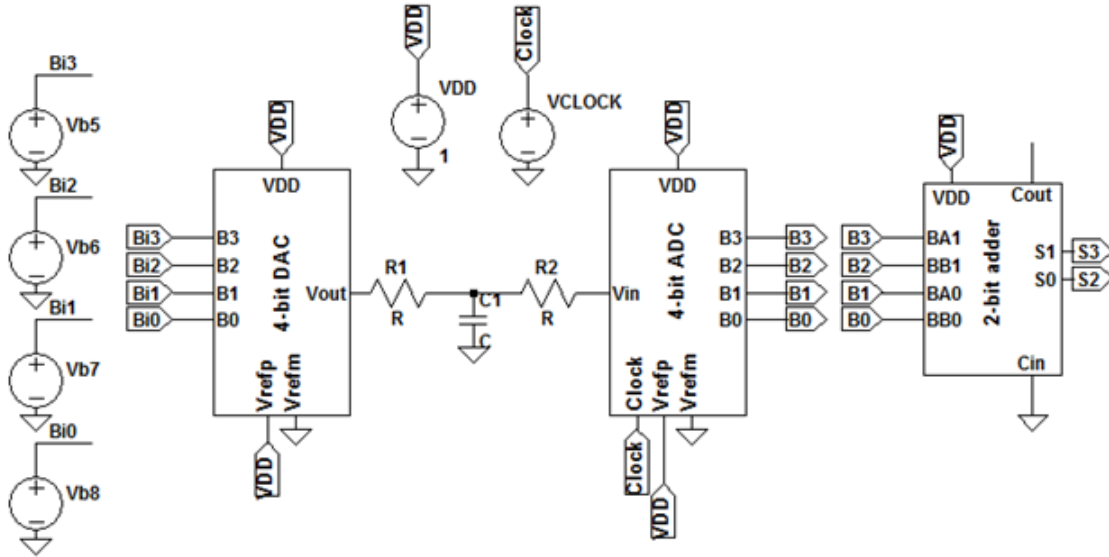


Figure 7.3: Experimental setup for analog bus replacing a 4-Bit parallel bus.

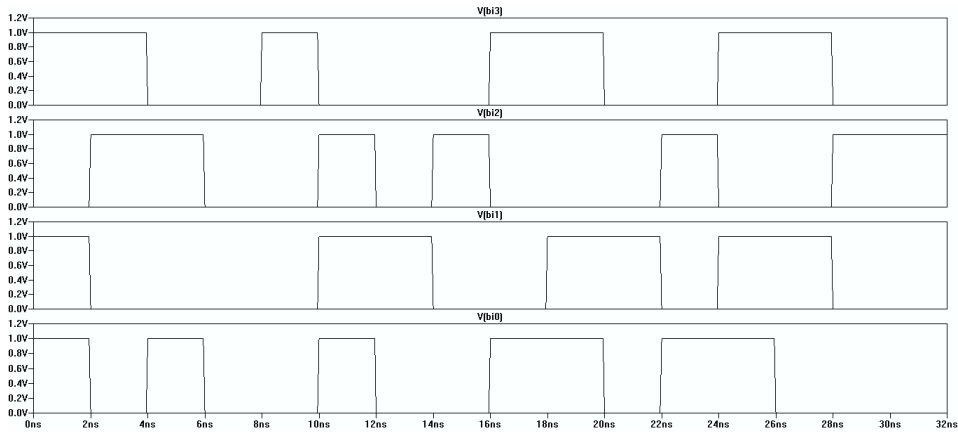


Figure 7.4: 4-Bit input patterns.

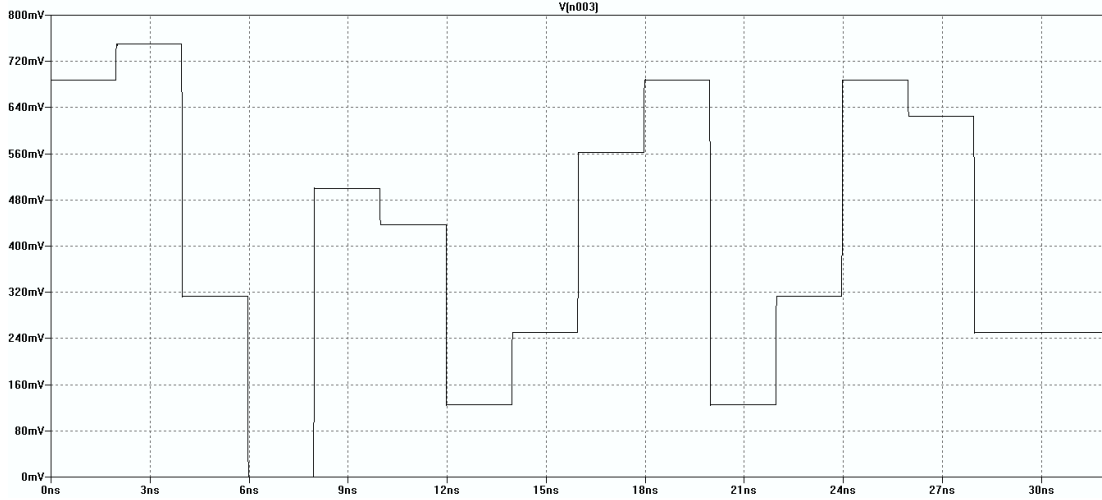


Figure 7.5: 4-Bit digital input converted to analog data.

Table 7.2: Comparison of power consumption of 4-bit parallel bus and analog bus for frequency = 1GHz.

Bus Length	Parallel Bus	Analog Bus
1mm	$464.23\mu W$	$36.7\mu W$
2mm	$928.3\mu W$	$67.2\mu W$
3mm	$1.39mW$	$97.1\mu W$
4mm	$1.85mW$	$126.5\mu W$
5mm	$2.31mW$	$155.9\mu W$

Comparison of power consumptions for a frequency of 500MHz with bus lengths of 1mm to 5mm (without addition of ADC/DAC power consumption) is given in Figure 7.7 and Table 7.3. The average power consumption per mm for the analog bus is around $16.17\mu W$. The average power consumption per mm for each parallel line is $54.8\mu W$ and for a 4-bit bus it is $219\mu W$.

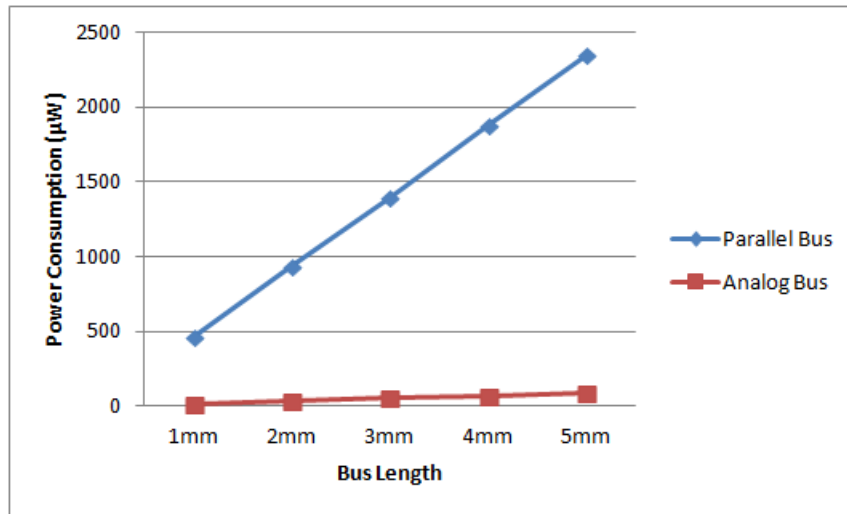


Figure 7.6: Parallel bus vs. analog bus (bus width = 4, frequency = 1GHz).

Table 7.3: Comparison of power consumption of 4-bit parallel bus and analog bus for frequency = 500MHz.

Bus Length	Parallel Bus	Analog Bus
1mm	219.22 μW	19.3 μW
2mm	438.95 μW	33.73 μW
3mm	658.13 μW	46.87 μW
4mm	875.34 μW	59.28 μW
5mm	1.095mW	71.44 μW

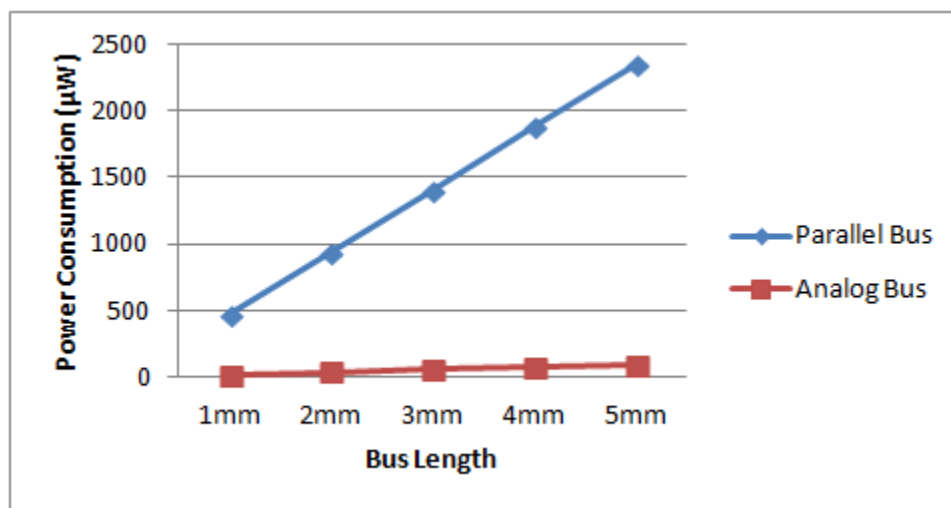


Figure 7.7: Parallel bus vs. analog bus (bus width = 4, frequency = 500MHz).

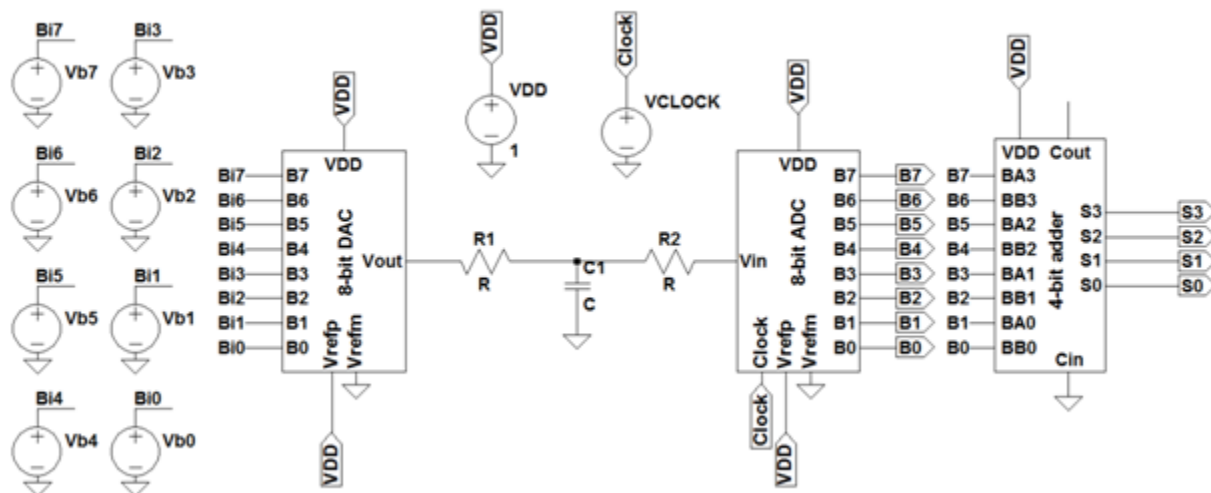


Figure 7.8: An analog bus to replace an 8-bit parallel bus.

Table 7.4: Comparison of power consumption of 8-bit parallel bus and analog bus for frequency = 500MHz.

Bus Length	Parallel Bus	Analog Bus
1mm	$469.8\mu W$	$19.2\mu W$
2mm	$939\mu W$	$36.82\mu W$
3mm	$1.4mW$	$54.4\mu W$
4mm	$1.88mW$	$71.84\mu W$
5mm	$2.35mW$	$89.2\mu W$

7.3 Power Analysis: Replacement of 8-Bit Parallel Bus

For simulation, we replaced an 8-line parallel bus by a 1 line analog bus (Figure 7.8). The digital and analog signals are shown in Figures 7.9 and 7.10, respectively. Here again the analysis has been done for bus lengths of 1mm to 5mm. Capacitance is calculated using the intermediate wire value given in the ITRS roadmap 2012 interconnect manual [24, 60].

Comparison of power consumption for a frequency of 500MHz with bus lengths of 1mm to 5mm (without addition of ADC/DAC power consumption) is given in Table 7.4 and Figure 7.11. The average power consumption per mm for the analog bus is around $18.3\mu W$. The average power consumption per mm for each parallel line is $58.65\mu W$ and for an 8-bit bus it is $469.2\mu W$.

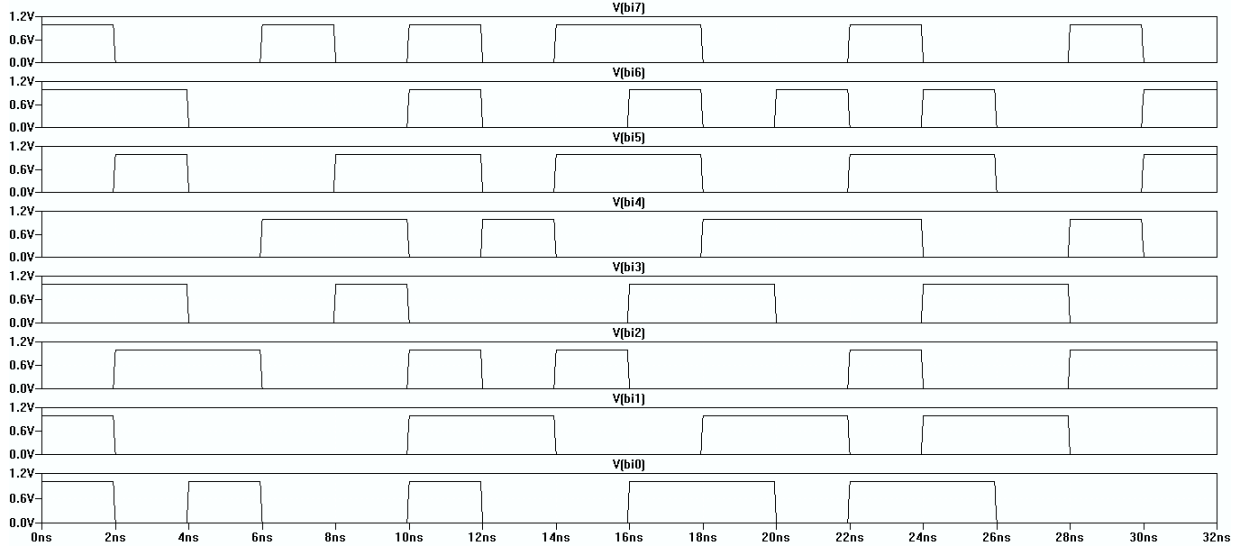


Figure 7.9: 8-bit input patterns.

Table 7.5: Power consumption of 4-bit and 8-bit buses.

Bus	4-bit bus power consumption			8-bit bus power consumption		
	Parallel	Analog	Power margin	Parallel	Analog	Power margin
1mm	$219.22\mu W$	$18.3\mu W$	$200.92\mu W$	$469.8\mu W$	$19.2\mu W$	$450.6\mu W$
2mm	$438.95\mu W$	$33.73\mu W$	$405.22\mu W$	$939\mu W$	$36.82\mu W$	$902.18\mu W$
3mm	$658.13\mu W$	$46.87\mu W$	$611.26\mu W$	$1.4mW$	$54.4\mu W$	$1.345mW$
4mm	$875.34\mu W$	$59.28\mu W$	$816.06\mu W$	$1.88mW$	$71.84\mu W$	$1.808mW$
5mm	$1.095mW$	$71.44\mu W$	$1.023mW$	$2.35mW$	$89.2\mu W$	$2.261mW$

7.4 Discussion of Results

Table 7.5 gives the results for 4-bit and 8-bit buses. It is observed that, the power consumption in the parallel bus has an exponential increase with respect to the bus length whereas the power consumption in the analog bus is increasing slowly. The power consumption of the ADC/DAC can be a design challenge for analog bus. From [3] and [1], ADC (ADS7924 from Texas Instruments) and DAC (LTC1591 from Linear Technology) it is observed that the power consumption of these devices is $5.5\mu W$ and $10\mu W$, respectively. But the converters are in kilohertz frequency range. But in the literature, there are converters which are in megahertz range (Table: 7.6). It can be said that gigahertz range converters do not seem to be too far.

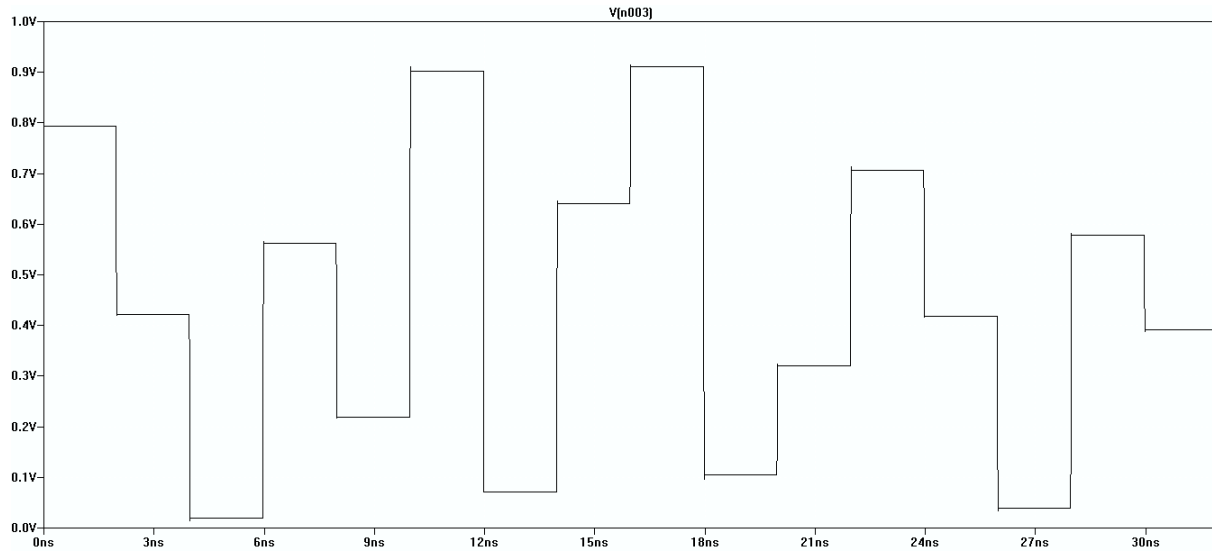


Figure 7.10: 8-bit digital input converted to analog data.

Table 7.6: Converter design survey [52].

Technology	Reference	Power (μW)	Frequency (MHz)
90nm	[15]	290	20
130nm	[41]	460	22
0.5 μm	[61]	550	10
65nm	[18]	806	88
65nm	[70]	820	50
90nm	[19]	820	40
65nm	[67]	950	150

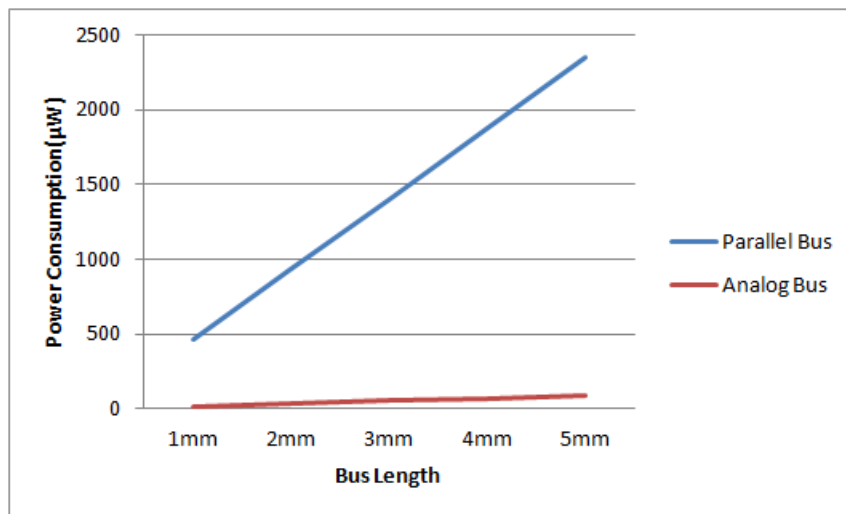


Figure 7.11: Parallel bus vs. analog bus (bus Width = 8, frequency = 500MHz).

Chapter 8

Conclusion

Technological development is enabling improved device density on a fixed chip area, and thousand cores do not look impossible anymore. This higher chip density in the design is making on-chip communication increasingly more important. In this thesis, first the importance of reducing power consumption for on-chip communication has been explained. Difficult challenges in designing a system with many cores include - hardware area, power dissipation, routing complexity, congestion and latency of the communication network. A unique concept of replacing parallel digital bus with an analog bus has been proposed here. A series of simulated experiments have been carried out to serve as proof-of-concept by evaluating power consumption of a single wire with DAC/ADC encoding in comparison to an n -bit parallel digital bus. Main advantages of this scheme are reduced power consumption and reduced bus area, along with reduction of routing complexity, and congestion. SPICE simulation for an ideal case shows that, the ratio of bus power consumed by the proposed analog scheme to a typical parallel digital scheme (without bus encoding or differential signaling) is given by $P_{\text{analog}}/P_{\text{digital}} = 1/(3n)$. Finally, though this thesis examined the feasibility of the scheme, much work remains to be done.

8.1 Challenges and Future Work

The efficiency of the proposed design depends upon the reduction in the number of bus wires, which, in turn, depends upon the design of DAC/ADC. Since a single wire will now be carrying a multitude of 2^n levels when it replaces n wires, the ambient noise levels can limit the successful communication between the cores. The noise tolerance of the ADC is a major design consideration, which determines how many digital wires can be replaced by a

single analog wire. Intended future work also includes design of encoding scheme for noise reduction techniques and data verification.

8.1.1 Design suitable converters

For implementation of the scheme, suitable compact and low power DACs and ADCs are needed. The preferable DACs should be able to convert the digital data to exact analog voltage value. On the other hand, the ADCs should be able to reconstruct the analog data into digital data without error. The converters should be of low power design to make this scheme advantageous.

8.1.2 Encoding Scheme

The noise present in interconnects is a major design consideration. Encoding schemes need to be explored to minimize the error rate. The least significant bit in the ADC is more prone to error as the resolution of the design is getting smaller with technology scaling. Careful measures need to be taken to ensure that the least significant bit is reconstructed properly. In cases where the scheme will be used in digital testing, dont care bits can be sent through the least significant bits. This may ensure utilization of the scheme without potential errors.

8.1.3 Combination of Analog Bus with other schemes

The author of [62] suggested a off-chip interconnect scheme that can be used to encode and decode binary signals into a 4-valued logic to reduce complexity. The four values used in the scheme are V_{DD} , $(V_{DD} - V_{thn})$, V_{thp} and 0. The potential of this scheme alone and with combination of analog bus can be analyzed for on-chip communication system.

8.1.4 Mixed-Signal Compression of Digital Test Data

As the complexity in the IC design is growing with scaling, longer test vectors are needed to detect the defects in the device. This is causing an increasing demand for reducing test time, cost and test power. A mixed-signal test compression method can be proposed based on data converters, which can have various benefits over the traditional methods [40].

Bibliography

- [1] “14-Bit and 16-Bit Parallel Low Glitch Multiplying DACs with 4-Quadrant Resistors,” *White Paper, Linear Technology Corporation*, Feb. 1999. <http://cds.linear.com/docs/en/datasheet/15917fa.pdf>.
- [2] “A Comparison of Network-on-chip and Busses,” *White Paper, Arteris, SA*, 2005.
- [3] “2.2V, 12-Bit, 4-Channel, microPOWER Analog-to-Digital Converter With I^2C Interface,” *White Paper, Texas Instruments Incorporated*, Jan. 2012. <http://www.ti.com/lit/ds/symlink/ads7924.pdf>.
- [4] “LTspice IV (Version 4.18b),” 2013. Linear Technology Corporation, <http://www.linear.com/designtools/software/#LTspice>.
- [5] R. J. Baker, *CMOS Mixed-signal Circuit Design*. John Wiley & Sons, 2008.
- [6] J.-B. Begueret, A. Mariano, and D. Dallet, “High-Speed A/D & D/A Conversion: A Survey,” in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2008, pp. 260–264.
- [7] L. Benini, G. De Micheli, E. Macii, M. Poncino, and S. Quer, “Power Optimization of Core-Based Systems by Address Bus Encoding,” *IEEE Trans. Very Large Scale Integration Systems*, vol. 6, no. 4, pp. 554–562, 1998.
- [8] T. Bjerregaard and S. Mahadevan, “A Survey of Research and Practices of Network-on-Chip,” *ACM Computing Surveys*, vol. 38, no. 1, p. 1, 2006.
- [9] S. Borkar, “Thousand Core Chips: A Technology Perspective,” in *Proc. 44th Design Automation Conference*, 2007, pp. 746–749.
- [10] S. Borkar and A. A. Chien, “The Future of Microprocessors,” *Comm. ACM*, vol. 54, no. 5, pp. 67–77, 2011.
- [11] T. D. Burd and R. W. Brodersen, “Energy Efficient CMOS Microprocessor Design,” in *Proc. Twenty-Eighth IEEE Hawaii International Conference on System Sciences*, volume 1, 1995, pp. 288–297.
- [12] D. Chinnery and K. Keutzer, *Closing the Power Gap Between ASIC and Custom: Tools and Techniques for Low Power Design*. Springer, 2007.
- [13] J. Cong, “An Interconnect-centric Design Flow for Nanometer Technologies,” *Proc. IEEE*, vol. 89, no. 4, pp. 505–528, 2001.
- [14] B. Cordan, “An Efficient Bus Architecture for System-on-Chip Design,” in *Proc. IEEE Custom Integrated Circuits Conf.*, 1999, pp. 623–626.
- [15] J. Craninckx and G. Van der Plas, “A 65fJ/conversion-step 0-to-50MS/s 0-to-0.7 mW 9b charge-sharing SAR ADC in 90nm digital CMOS,” in *Proc. IEEE International Solid-State Circuits Conf. Digest*, 2007, pp. 246–600.

- [16] R. R. Dobkin, A. Morgenshtein, A. Kolodny, and R. Ginosar, "Parallel vs. Serial On-chip Communication," in *Proc. ACM International Workshop on System Level Interconnect Prediction*, 2008, pp. 43–50.
- [17] H. Esmailzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Power Challenges May End the Multicore Era," *Comm. ACM*, vol. 56, no. 2, pp. 93–102, 2013.
- [18] J. Fredenburg and M. Flynn, "A 90MS/s 11MHz Bandwidth 62dB SNDR Noise-shaping SAR ADC," in *Proc. IEEE International Solid-State Circuits Conf. Digest*, 2012, pp. 468–470.
- [19] V. Giannini, P. Nuzzo, V. Chironi, A. Baschirotto, G. Van der Plas, and J. Cranineckx, "An 820 μ W 9b 40MS/s Noise-Tolerant Dynamic-SAR ADC in 90nm Digital CMOS," in *Proc. IEEE International Solid-State Circuits Conf. Digest*, 2008, pp. 238–610.
- [20] R. Gonzalez, B. M. Gordon, and M. A. Horowitz, "Supply and Threshold Voltage Scaling for Low Power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, no. 8, pp. 1210–1216, 1997.
- [21] N. Hatta, N. D. Barli, C. Iwama, L. D. Hung, D. Tashiro, S. Sakai, and H. Tanaka, "Bus Serialization for Reducing Power Consumption," *ISPJ Trans. Advanced Computing Systems*, vol. 47, no. SIG-3, pp. 686–694, Mar. 2006.
- [22] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Millberg, and D. Lindqvist, "Network on Chip: An Architecture for Billion Transistor Era," in *Proc. IEEE NorChip Conf.*, volume 31, 2000.
- [23] R. Ho, K. Mai, and M. Horowitz, "Efficient on-chip global interconnects," in *IEEE Symp. on VLSI Circuits*, 2003, pp. 271–274.
- [24] D. Ingerly, A. Agrawal, R. Ascazubi, A. Blattner, M. Buehler, V. Chikarmane, B. Choudhury, F. Cinnor, C. Ege, C. Ganpule, *et al.*, "Low-k Interconnect Stack with Metal-Insulator-Metal Capacitors for 22nm High Volume Manufacturing," in *Proc. IEEE International Interconnect Technology Conf.*, 2012, pp. 1–3.
- [25] S. M. Kang, "Accurate Simulation of Power Dissipation in VLSI Circuits," *IEEE J. Solid-State Circuits*, vol. 21, no. 5, pp. 889–891, 1986.
- [26] K. Kaur and A. Noor, "Strategies & Methodologies for Low Power VLSI Designs: A Review," *International J. Advances in Engineering & Technology*, vol. 1, pp. 159–165, 2011.
- [27] M. Keating, D. Flynn, R. Aitken, A. Gibbons, and K. Shi, *Low Power Methodology Manual: For System-on-Chip Design*. Springer, 2007.
- [28] A. Kedia, "Design of a Serialized Link for On-chip Global Communication," Master's thesis, University of British Columbia, Canada, 2006.
- [29] A. Kedia and R. Saleh, "Power Reduction of On-Chip Serial Links," in *IEEE International Symp. Circuits and Systems*, 2007, pp. 865–868.
- [30] W. A. Kester, *Data Conversion Handbook*. Newnes, 2005.
- [31] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage Current: Moore's Law Meets Static Power," *Computer*, vol. 36, no. 12, pp. 68–75, 2003.
- [32] T. W. Krawczyk Jr, *Circuits for the Design of a Serial Communication System Utilizing SiGe HBT Technology*. PhD thesis, Rensselaer Polytechnic Institute, Troy, New York, 2000.

- [33] R. Kumar, V. Zyuban, and D. M. Tullsen, “Interconnections in Multi-core Architectures: Understanding Mechanisms, Overheads and Scaling,” in *Proc. of 32nd IEEE International Symposium on Computer Architecture*, 2005, pp. 408–419.
- [34] S. Kumar, A. Jantsch, J.-P. Soininen, M. Forsell, M. Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, “A Network on Chip Architecture and Design Methodology,” in *Proc. IEEE Computer Society Annual Symp. on VLSI*, 2002, pp. 105–112.
- [35] V. Kursun and E. G. Friedman, *Multi-Voltage CMOS Circuit Design*. John Wiley, 2006.
- [36] H. G. Lee, N. Chang, U. Y. Ogras, and R. Marculescu, “On-Chip Communication Architecture Exploration: A Quantitative Evaluation of Point-to-Point, Bus, and Network-on-Chip Approaches,” *ACM Trans. Design Automation of Electronic Systems*, vol. 12, no. 3, p. 23, 2007.
- [37] J. Lee, “On-Chip Bus Serialization Method for Low-power Communications,” *Proc. Electronics and Telecommunications Research Institute*, vol. 32, no. 4, pp. 540–547, 2010.
- [38] K. Lee, S.-J. Lee, and H.-J. Yoo, “SILENT: Serialized Low Energy Transmission Coding for On-chip Interconnection Networks,” in *Proc. IEEE/ACM International Conf. Computer-Aided Design*, 2004, pp. 448–451.
- [39] D. Lewis, “SerDes Architectures and Application,” in *Proc. of the DesignCon*, 2004.
- [40] B. Li, V. D. Agrawal, and B. Zhang, “Mixed-Signal Compression of Digital Test Data.” Personal Communication, June 2013.
- [41] J. Lin and B. Haroun, “An Embedded 0.8 V/480 μ W 6b/22 MHz Flash ADC in 0.13- μ m digital CMOS Process Using a Nonlinear Double Interpolation Technique,” *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1610–1617, 2002.
- [42] C. A. Mack, “Fifty Years of Moore’s Law,” *IEEE Trans. Semiconductor Manufacturing*, vol. 24, no. 2, pp. 202–207, 2011.
- [43] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, “Interconnect-Power Dissipation in a Microprocessor,” in *Proc. ACM International Workshop on System Level Interconnect Prediction*, 2004, pp. 7–13.
- [44] M. Mitić and M. Stojčev, “An Overview of On-chip Buses,” *Proceedings of Facta Universitatis Series: Electronics and Energetics*, vol. 19, no. 3, pp. 405–428, 2006.
- [45] K. Moiseev, A. Kolodny, and S. Wimer, “Timing-Aware Power-Optimal Ordering of Signals,” *ACM Trans. Design Automation of Electronic Systems*, vol. 13, no. 4, p. 65, 2008.
- [46] E. Mollick, “Establishing Moore’s Law,” *Annals of the History of Computing*, vol. 28, no. 3, pp. 62–75, 2006.
- [47] G. E. Moore, “Cramming More Components onto Integrated Circuits,” *Electronics*, vol. 38, no. 8, Apr. 1965.
- [48] G. E. Moore, “Progress in Digital Integrated Electronics,” in *IEEE International Electron Devices Meeting Digest*, 1975, pp. 11–13.
- [49] G. E. Moore, “Lithography and the Future of Moore’s Law,” *Proc. SPIE*, vol. 2437, May 1995.
- [50] A. Morgenshtein, I. Cidon, A. Kolodny, and R. Ginosar, “Comparative Analysis of Serial vs Parallel Links in NoC,” in *Proc. IEEE International Symp. System-on-Chip*, 2004, pp. 185–188.

- [51] B. Murmann, "A/D Converter Trends: Power dissipation, Scaling and Digitally Assisted Architectures," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2008, pp. 105–112.
- [52] B. Murmann, "ADC Performance Survey 1997-2013, ISSCC & VLSI Symposium," 2013. <http://www.stanford.edu/~murmman/adcsurvey.html>.
- [53] S. Pasricha and N. Dutt, *On-Chip Communication Architectures: System on Chip Interconnect*. Morgan Kaufmann, 2010.
- [54] J. V. Patel and H. Bhatt, "Performance Evaluation Of Different Types Of Analog To Digital Converter Architecture," *International J. Engineering*, vol. 1, no. 10, 2012.
- [55] J. Patil, L. He, and M. Jones, "Clock and Data Recovery for a 6 Gbps SerDes Receiver," in *Proc. 3rd IEEE International Conf. Computer Science and Information Technology*, volume 5, 2010, pp. 217–221.
- [56] B. C. Paul, A. Agarwal, and K. Roy, "Low-Power Design Techniques for Scaled Technologies," *Integration: The VLSI J.*, vol. 39, no. 2, pp. 64–89, 2006.
- [57] M. Pedram and J. M. Rabaey, *Power Aware Design Methodologies*. Springer, 2002.
- [58] B. Razavi, *Principles of Data Conversion System Design*. New York: IEEE Press, 1995.
- [59] R. R. Schaller, "Moore's Law: Past, Present and Future," *IEEE Spectrum*, vol. 34, no. 6, pp. 52–59, 1997.
- [60] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors," 2012. <http://www.itrs.net/Links/2012ITRS/Home2012.htm>.
- [61] D. Senderowicz, G. Nicollini, S. Pernici, A. Nagari, P. Confalonieri, and C. Dallavalle, "Low-Voltage Double-Sampled $\Sigma\Delta$ Converters," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1907–1919, 1997.
- [62] A. D. Singh, "Four-valued Interface Circuits for NMOS VLSI," *International J. Electronics*, vol. 63, no. 2, pp. 269–279, 1987.
- [63] F. N. Taher and V. D. Agrawal, "A Low-Power Analog Bus Approach for On-Chip Digital Communication," in *31st IEEE International Conf. Computer Design*, 2013. Submitted.
- [64] F. N. Taher, S. Sindia, and V. D. Agrawal, "An Analog Bus for Low Power On-Chip Digital Communication," in *Work-in-Progress Poster Session, Design Automation Conference*, (Austin, Texas), June 2013.
- [65] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 2009.
- [66] R. J. Van de Plassche, *CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters*, volume 2. Springer, 2003.
- [67] R. H. van Veldhoven, R. Rutten, and L. J. Breems, "An Inverter-Based Hybrid $\Delta\Sigma$ Modulator," in *Proc. IEEE International Solid-State Circuits Conf. Digest*, 2008, pp. 492–630.
- [68] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*. Addison-Wesley, 2010.
- [69] J. Wikner, *Studies on CMOS Digital-to-Analog Converters*. PhD thesis, Linköping University, Sweden, 2000.

- [70] M. Yoshioka, K. Ishikawa, T. Takayama, and S. Tsukamoto, “A 10b 50MS/s 820 μ W SAR ADC with on-chip digital calibration,” in *Proc. IEEE International Solid-State Circuits Conf. Digest*, 2010, pp. 384–385.
- [71] A. Zjajo, *Low-power High-resolution Analog to Digital Converters: Design, Test and Calibration*. Springer, 2011.