

Improving Robustness of Ultra-low Power Systems through Power Reduction in Wired Signaling Circuits

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Electrical Engineering)

by

Christopher J. Lukas

March 2017

Abstract

The Internet of Things (IoT) will offer new levels of services and rich data sets. Services like health monitoring will be improved through more personalized care. Rich data sets provided by applications like infrastructure management will allow a better understanding of wear and tear on buildings and bridges, creating a safer environment for us to live in. In order to achieve this, IoT systems must be placed everywhere from inside clothing to monitor body functions, to inside concrete bridge structures to monitor vibrations. There will be no dedicated supply of power to nodes in these places, meaning that energy must be harvested from the environment. Due to the nature of energy harvesting circuits, the energy storage component should last through many charge cycles, especially in smaller form-factor applications. Existing commercial technologies last hundreds to few thousands of cycles and often require regular maintenance and periodic full discharge cycles to avoid damage, limiting their use in self-powered, IoT applications. Electrical energy storage in capacitors and super-capacitors provide a more robust solution, having a cycle life greater than 500000 cycles [1]. As a result, the energy efficiency and power consumption of IoT systems must be reduced to the power levels a capacitor or super-capacitor can deliver.

Power is reduced by decreasing supply voltage and optimizing circuits, architectures, and system level knobs. Circuit level optimizations have a greater effect if they were previously contributing to a large portion of the ULP system budget. Circuits within this category are mainly analog, memory, and chip Input and Output (I/O). Analog and memory circuits are heavily researched, and lower power solutions are being developed. On the other hand,

ULP communication circuits have not been thoroughly researched, and optimizations can provide great reductions in both circuit and system power. Off-chip power reduction will allow for regular communication between nodes without needing to manage how often data is transmitted. ULP On-package communication will allow integration of different process technologies on the same chip, leveraging the best aspects of each process. On-chip communication can be improved to allow more fine-grained system level tuning to configure a platform for more desired operating modes.

This dissertation presents techniques at both the circuits and systems level to reduce the power consumption and increase the level of integration of a proof of concept battery-less, self-powered, System in Package (SiP). The heart of these improvements are in communication. Techniques are introduced to improve power and energy of off-chip communication, and a transceiver for robust, wired, on-body networks is demonstrated. An ULP on-package cold-boot bus is presented, enabling integration of an ULP non-volatile memory for loading instructions when the system powers up. A flexible on-chip bus is implemented to allow fine-grained power and energy SoC optimization at the logic block level. Also presented are applications and improved test methods to facilitate commercialization of self-powered, battery-less platforms. A proof of concept relative positioning system is implemented to reduce reliance on high-power GPS radios, and provide positioning when a GPS signal cannot be acquired. Lastly a test methodology is presented for subthreshold SoCs to reduce functional test time by testing at a higher optimal voltage and predicting delay and power of the system at the lower operational voltage.

Lower power communication, improved application space, and a faster test process will enable commercialization of systems. Having ubiquitous commercial battery-less platforms that are more robust will provide large sets of data to give insight to new areas of medical, industrial, and environmental research and will improve quality of life. The improved quality of life provided by these devices through an increase in personalized medical care and an abundance of sensor data will progress how our society works and interacts.

Approval Sheet

This dissertation is submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (Electrical Engineering)

Christopher J. Lukas

This dissertation has been read and approved by the Examining Committee:

Benton Calhoun, Adviser

Joanne Dugan, Committee Chair

Stephen Wilson

Steven Patek

Steven Bowers

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, Dean, School of Engineering and Applied Science

March 2017

To everyone who's helped me succeed

Acknowledgments

I would like to begin with thanking my family for giving me the support for going to graduate school. From an early age my parents supported creativity and learning. They have always been there when I wanted to try new things and use that experience to become a better person. I would like to thank my uncle James Communale for introducing me to the more technical aspects of electronics, as this brought me in the direction of studying computer and electrical engineering.

During my time at the University of Pittsburgh I was still trying to find my way in engineering. Fortunately there were a handful of professors passionate about their work who were also great teachers. Dr. Stanchina introduced me to microelectronic circuits and solid state physics, leading me to become more interested in computer hardware and circuits. Dr. Levitan introduced me to VLSI, building larger systems of circuits to solve many problems.

I would like to thank my adviser Benton Calhoun for giving me the opportunity to work here, as well as supporting my ideas and giving me freedom to research topics that I found most interesting. The removal of unnecessary stress was wonderful. I never had to worry about my funding or how we would pay to fab so many chips. Ben always made sure I had the proper materials and actively mentored management and communication skills. I would like to thank my dissertation committee for guiding my work in a way that will allow it to be more impactful, as well as providing perspective from their discipline.

Robust Low Power VLSI group members have been very helpful in ramping me up to researching circuits. Particularly, Kyle Craig and Alicia Klinefelter. Kyle and Alicia made

me comfortable as soon as I joined the group, and taught me to become more confident in my work. Kyle has continued with excellent technical discussions while working at Psikick over Scotch and cigars and has been a resource of knowledge that I am thankful for. Alicia has always had a balanced viewpoint and a great attention to detail, which has helped me become more precise in my work. Yousef Shakhsheer and Yanqing Zhang were also important in helping me feel at home in Bengroup, but unfortunately for me graduated before I could sponge up as much knowledge from them as I would have liked. Ben Boudaoud helped improve my critical thinking with circuit design and how it relates to systems and also motivated my research into I/O and signaling. Seyi and Jim helped keep my sanity in check by pulling me out of the office and keeping me social, and Seyi was always a good person to bounce the crazy circuit ideas off of. Aatmesh helped begin my interest in analog circuits. Farah has been an excellent colleague to work with throughout my time at UVA. She taught me to be more thorough and rigorous in my work and has helped improve my work ethic while churning through the countless tapeouts we've done together. Abhishek helped me become more comfortable with designing new analog circuits, teaching me that just because no one uses a design for a particular application doesn't mean you can't or shouldn't. I would also like to thank Jacob Breiholz for the crazy amount of time he spent testing our SoC revisions to tighten up the design. Terry Tigner has been incredible in getting me the resources I needed. She made it possible to achieve all of the important deadline submissions. I also spent time working with Divya, Arijit, He, Harsh, and NingXi. Their group discussions and brainstorming helped shape my work and for that I am grateful.

During my time in grad school I interned at NVidia research in Durham, North Carolina. John Poulton was an amazing mentor, helping me become more methodical with modeling. He has always been there to support my academic research as well, giving much needed industry perspective on my work at UVA. Matt Fojtik provided insight to tool flow design, and has also been a great brewing mentor. John Wilson helped me to think outside the box and find simple solutions to complicated problems.

Contents

Contents	vii
List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Motivation	1
1.2 Thesis	6
1.2.1 Reducing Power	6
1.2.2 Improving robustness	6
1.2.3 Reducing time to market	7
1.3 Approach	7
1.3.1 Reducing Communication Circuit Power	7
1.3.2 Introducing a System in Package with Improved Integration and Ro- bustness	8
1.3.3 Applications and Test Time Reduction in Ultra-low Power Systems .	9
1.4 Contributions and Organization	9
1.4.1 Background	10
1.4.2 Ultra-low Power Off-chip Communication	10
1.4.3 Ultra-low Power On-package Communication	10
1.4.4 Ultra-low Power On-chip Communication	11
1.4.5 Enabling Ubiquitous Self-powered Systems	11
1.4.6 Conclusion	11
2 Background	12
2.1 Ultra-low Power Circuits	12
2.2 Wired Communication	13
2.2.1 Wired Off-chip Communication	14
2.2.2 Wired On-chip and On-package Communication	14
2.2.3 Commercial Wired Communication	16
2.3 Systems On Chip	16
2.3.1 Systems in Package	17
2.4 Chip Testing Process	17

3	Ultra-low Power Off-chip Communication	19
3.1	Introduction	19
3.2	Motivation	20
3.3	Prior Art	23
3.4	Chip-to-Chip Serial Link for Ultra-Low Power Systems	25
3.4.1	Energy and Power Co-optimization	26
3.4.2	Error Correction Coding	30
3.4.3	Results	30
3.5	Differential Link for Reliable Wireline Communication in ULP On-Body Sensor Networks	33
3.5.1	Design	34
3.5.2	Results	39
3.5.3	Conclusion	44
4	Ultra-low Power On-package Communication	46
4.1	Introduction	46
4.2	Motivation	48
4.3	Prior Art	50
4.4	Asymmetric Bus Enabling Cold-Boot Self-Powered Systems in Package	50
4.4.1	Design	51
4.4.2	Results	54
4.4.3	Conclusion	55
5	Ultra-low Power On-chip Communication	57
5.1	Introduction	57
5.2	Motivation	58
5.3	Prior Art	59
5.4	ULP Flexible On-Chip Bus Enabling GALS and DVFS in Self-Powered Systems	60
5.4.1	Design	61
5.4.2	Results	64
5.4.3	Conclusion	66
6	Reducing Self-powered System Time to Market	68
6.1	Introduction	68
6.2	Motivation	70
6.3	Prior Art	73
6.4	Sub μ W System on Chip	74
6.4.1	System Overview	74
6.4.2	System on Chip Design Process	76
6.4.3	System in Package Design Process	80
6.4.4	Energy Harvesting Platform Power Management	81
6.4.5	LPC-Memory Integration	82
6.4.6	On-package Cold-Boot Management System	83
6.4.7	Non-Volatile Boot Memory	84
6.4.8	Boot Sequence	85

6.4.9	Backup Sequence	86
6.4.10	Radio	87
6.4.11	Custom Radio Interface	87
6.4.12	Application: Wakeup and Fall Detection	88
6.4.13	Measurements and Comparison	89
6.5	Self-powered, Battery-less Positioning System	90
6.5.1	Algorithm Flow	91
6.5.2	Accumulating Error	91
6.5.3	Results	92
6.6	Modeling Trans-threshold Correlations for Reducing Functional Test Time in ULP Systems	93
6.6.1	Scope of this work	94
6.6.2	Correlation Between Voltages of Operation	95
6.6.3	Implications of Voltage Correlations	96
6.6.4	Faults Containing High Correlation	96
6.6.5	Faults Containing Varying Correlation	97
6.6.6	Methodology	98
6.6.7	Applied Methodology to Simulated Ring Oscillator	101
6.6.8	Implications of Predictive Analysis	107
6.6.9	Measured Results	107
6.6.10	Digital Combinational and Sequential Circuit	107
6.6.11	High Level Implications of Circuit Level Results	113
6.6.12	Trans-threshold Correlations in Other Testing Areas	114
6.6.13	Contributions	114
6.6.14	Conclusion	117
7	Conclusion	119
7.1	Contributions	120
7.2	Impact	123
7.3	Future Work	124
7.3.1	Off-chip Communication	124
7.3.2	On-package Communication	125
7.3.3	On-chip Communication	125
7.3.4	Ultra-low Power Systems on Chip	125
7.3.5	Applications of Self-powered Internet of Things Systems	126
7.3.6	Test Methodologies of subthreshold SoCs	126
A	Publications	128
A.1	Completed	128
A.2	Submitted	128
A.3	Planned	129
B	Acronyms	131
	Bibliography	137

List of Tables

3.1	A comparison of a typical SPI transmission at ULP SoC voltage verses the optimized transmission of our fabricated circuit.	31
3.2	On-body link comparison to existing wireline work	43
3.3	Measured on-body link power breakdown	43
4.1	Power breakdown of On-package Cold-Boot Bus	54
4.2	On-package bus comparison to state of the art off-chip buses (ULP on left, HPC on right)	55
5.1	Measured Accelerator Active Power Breakdown and Comparison	65
6.1	The LPC stall conditions and interrupt sources.	83
6.2	Comparison to state-of-the-art SoCs	90
6.3	Sequential logic test failure windows, correlation, and test time % data . . .	113

List of Figures

1.1	Design space of communication. The focus of this dissertation, ultra-low power, low throughput wired signaling, shown in green (darker), has not yet been researched.	5
2.1	Diagram showing types of wired communication.	15
3.1	Pie chart of typical ULP SoC power breakdown. I/O is a large portion of the power budget, motivating research into lowering power.	21
3.2	De Facto standards and state of the art communication circuits.	24
3.3	Model of single ended transmission line. Series inductors and resistors are ignored at low frequency operation. Clock and data lines were both designed for capacitance and in turn delay matching.	25
3.4	Plot of Minimum energy points for a range of activity factors. Each activity factor has an associated optimal operating voltage, creating a new trend of minimum power. Black diamonds show simulation values, while the dashed line shows the model.	28
3.5	Micrograph of the fabricated test chip for serial off-chip communication. Because of low visibility due to the metal fill layers above the circuit, the layout has been overlaid.	29
3.6	Plot showing measured error rate as a function of voltage and bit rate for the transceiver. Due to the steep slope ECC is not energy effective.	31
3.7	Eye diagram of typical transmission measured at the receiver.	32
3.8	A comparison of power and energy per bit with current state of the art	33
3.9	Block diagram of transceiver circuit.	35
3.10	Transistor diagram of transmitter circuit.	36
3.11	Transistor diagram of clock receiver and bias generator. Bias generator allows operation over a wide range of voltages.	36
3.12	Transistor diagram of edge detector. Edge detector turns received clock edges into pulses, allowing the data receiver to latch data on both edges.	37
3.13	Transistor diagram of data receiver, a latched comparator designed for low voltage operation.	37
3.14	Transistor diagram of contention free SR latch. Contention free SR latch receives output data from the data receiver, latching it to provide a steady output for each half clock cycle.	38

3.15	DC simulation of bias generator, showing VDD (dashed) and each output configuration (solid).	39
3.16	Transient simulation of transmitter shows reduced swing centered at VDD/2.	40
3.17	Die photo of differential link test chip.	41
3.18	Measured bit error rate plot of transceiver.	41
3.19	Measured eye diagram of transceiver over an Ethernet channel. Transceiver's supply operating voltage was 240 mV, while the channel showed a reduced 120 mV peak to peak, with an eye height of 55 mV. Bit error rate $< 10^{-9}$ (no shown errors)	42
3.20	Measured eye diagram of transceiver over the flexible textile interconnect. Transceiver's supply operating voltage was 300 mV, while the channel showed a reduced 160 mV peak to peak, with an eye height of 110 mV. Bit error rate $< 10^{-6}$ (no shown errors)	42
3.21	Photograph of transceiver snapped onto a shirt. Flexible textile interconnect shown on the right side of the image electrically connects the button snaps through the inside of the shirt.	43
3.22	Transceiver using 3 meter Ethernet cable throughout operating voltages compared to state-of-the-art-designs. Minimum Energy Point (MEP) shows improved co-optimization of energy and power.	44
4.1	Micrograph of a die showing bond pads around the proximity of the die. Wires extending off the bottom of the image have been bonded to the pads.	47
4.2	Example diagram of potential application of on-package bus. On-package buses enable integration of multiple dice in the same package.	47
4.3	Image of a dual inline package showing pins used to interface with off-chip components.	48
4.4	Block diagram of the on-package cold-boot bus.	51
4.5	Transistor level diagram of the transceiver. Binary weighted drivers allow for reduced swing on the transmission line and sense amplifier receivers handle the reduced signal.	53
4.6	Die photos of battery-less SoC in 130nm low power technology and Non-volatile boot memory in 130nm FeRAM technology (dice not to scale).	55
5.1	Example diagram of an on-chip SoC bus. The main controller uses the on-chip bus to communicate send and receive data to other logic blocks on the same chip.	58
5.2	Transistor diagram of an SR latch. SR latches are used in the on-chip bus to allow block level power optimization through flexible level shifting.	63
5.3	Architecture of the blocks in the SoC using SR latches for level shifting and memory. Input SR latches (in gray) are optional for storing memory mapped I/O while the block is power gated.	64
5.4	Micrograph of the fabricated test chip for optimization of on-chip accelerator power.	66

5.5	Simulated schmoo plot of SR latch used as a level shifter. Green (light) shows success and red (dark) shows failures when shifting between the given input and output voltage. SR latch shifts both upwards and downwards, creating an ideal solution for post-fabrication block level voltage optimization.	67
6.1	Overview of the System in Package showing integration of dice with their data interfaces and power supply. The System in Package leverages the best features of all three technologies. Dice drawn to scale.	71
6.2	Die photo of SoC with floorplan of block placement. Die dimensions are 3.6 mm by 3.8 mm.	75
6.3	Layout of SoC control block with overlay showing breakdown of logic.	78
6.4	Layout of SoC peripheral block with overlay showing breakdown of logic. Wishbone bus runs through the center for lower power, higher max chip frequency, and increased modularity.	79
6.5	Block diagram for System in Package. Diagram shows interfaces between SoC, radio, and NVM, as well as supplied power and main logic blocks in both systems. Accelerometer (ADXL) is shown interfacing with SoC SPI and GPIO.	81
6.6	Block diagram for the Energy Harvesting Platform Power Manager. This sub-system harvests energy and provides power to the SoC, NVM, radio, and COTS components.	82
6.7	Chart showing power modes for the System in Package. System power is greatly reduced while stalled and waiting for interrupts.	84
6.8	Flow chart for SoC bootup sequence from NVM. This feature allows the system to recover after long periods without harvesting.	86
6.9	Flow chart for SoC backup sequence to NVM. This feature improves quality of service for applications which the system must keep track of events over a long period of time.	88
6.10	Application example for SoC. The system utilizes its reduced power sleep mode while waiting for an interrupt from the COTS accelerometer. This allows the system to reach an average 507 nW power consumption.	89
6.11	Flow used for continuous position measurements in self-powered relative positioning system.	92
6.12	Measured delay correlation in a logic block containing combinational and sequential circuits. r^2 is used to determine how well the data fits the regression, and is generally within the range of 0 to 1. An r^2 value of 1 implies a perfect fit, while 0 implies no correlation between the line of best fit and the data.	97
6.13	The characterization flow chart describing the different steps required before mass testing can be started.	99
6.14	Flow chart showing test flow after characterization has been completed. Test flow reduces time by predicting low operating voltage delay.	101
6.15	Simulated Monte Carlo delay in a ring oscillator operating at 0.5V and 0.6V.	103
6.16	Simulated Monte Carlo delay in a ring oscillator operating at 0.5V and 0.8V.	103
6.17	Simulated delay correlation in a ring oscillator operating at a range of voltages.	104
6.18	Simulated Monte Carlo delay regression for a ring oscillator. The dashed blue lines show the window of prediction for testing at the higher voltage.	105

6.19	Simulated total test time for ring oscillator. Simulated minimum test time is at 0.6V, taking 19.8% of test time at 0.5V.	106
6.20	Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.4V delay.	108
6.21	Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.5V delay.	110
6.22	Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.6V delay.	111
6.23	Measured total test time for digital logic containing combinational and sequential circuits. Measured minimum test time is at 0.4V, taking 18.6% of test time at 0.3V.	112
6.24	Modeled total test time for digital logic containing combinational and sequential circuits. Calculated minimum test time is at 0.398V, taking 18.57% of test time at 0.3V.	113
6.25	Measured power correlations for digital logic containing combinational and sequential circuits. Predictions allow binning of ULP components simply by testing delay at a higher voltage.	115
7.1	Figure showing the impact of adding presented off-chip serial link to SoC. The first chart shows I/O being a large portion of the system power budget. As the system power scales downward, the second figure shows the I/O dominating power consumption. The third figure shows the presented circuits being used for communication between the SoC and a research based, low voltage sensor.	124

Chapter 1

Introduction

1.1 Motivation

The first step towards ubiquitous computing was the development of the internet. Almost every computer is now connected to the network that allows it to communicate in real time with any other computers on earth. This ability to share data between devices provides the foundation for the Internet of Things (IoT). The target of IoT is to embed computing devices in everything around us to provide real-time access to previously unimaginable quantities of data. Devices can be placed during construction of buildings and bridges on load bearing beams to measure structural integrity, in cities to sense ozone and pollution, and throughout the world for more accurate climate data sensing. Work is already being done to bring sensing to the body, providing real time health data to improve quality of life. Such systems will also provide previously unattainable sensor information to solve engineering problems. The IoT will work in conjunction with the ever growing computational power of desktop computers and servers capable of doing intelligent analysis from the newly accessible data. The results will improve and redefine our knowledge of the world.

Unlike computers, IoT devices have a strict set of requirements dictated by the application space they are targeting. These requirements restrict the form factor, the lifetime, the

processing power, the communication interfaces, and many other features of the device. For example, health monitoring devices must have a small form factor and a long lifetime so they can be wearable and unobtrusive. Infrastructure management will require sensors placed inside and around building structures to improve our understanding of wear in structures and inform decisions involving safety. Thus, such a system will need to communicate its data without having access to a continuous power source.

The form factor requirement motivates the use of a System on Chip (SoC) solution, as SoCs integrate multiple functions into a small, single chip. These SoCs usually contain a main controller, sensors to collect information, accelerators for processing data, memory for storing processed data, and a communication interface to relay this data. To improve lifetime of the device, its power consumption must be reduced and/or its battery capacity and number of recharge cycles must be improved. However, increasing the number of recharge cycles of a battery is a significant challenge and increasing battery capacity results in a larger form factor. Thus, researchers have proposed replacing batteries with supercapacitors, using energy harvesting devices to enable self-powered operation, and significantly reducing the systems power consumption.

Self-powered operation also addresses the lack of power source concern for many applications including the infrastructure management system. This means energy will need to be harvested from the environment. Harvesting modalities like motion, solar, and thermoelectric harvesting will allow IoT systems to operate without a constant source of power. However, energy harvesting is intermittent, meaning that energy needs to be stored for times when motion, solar, or heat energy are not available. Enough energy needs to be stored so that the system does not lose power during these lapses in harvesting.

Due to the cyclical nature of harvesting, the energy storage component must last through many cycles of charging and discharging. If this occurs tens of times per day, it means the storage component will go through tens or hundreds of thousands of cycles in its lifetime. Batteries tend to last many hundreds of cycles, and in some cases last one to two thousand. In

contrast, capacitors store charge electrically instead of chemically, and as a result last through millions or more cycles. As a result, these systems will need to be capable of operating with a capacitor or supercapacitor as their energy source and survive lapses in harvesting for extended periods of time.

This chain of thinking leads to the inevitable conclusion that if we are to create an IoT based on mobile, wearable, and small form factor devices, the energy and power consumption of the devices must be reduced to the power levels a capacitor or supercapacitor can deliver. In small form factor devices, harvesting is limited to μWs of power, meaning that if the system should survive long lapses in harvesting, its power budget should be well below the expected harvesting power, potentially even below $1 \mu\text{W}$. While ULP battery-less systems has been demonstrated in research, they still face a number of challenges that hinder their use in commercial products.

What is required to move from a proof of concept to producing a usable self-powered IoT product? First and foremost, devices will need to be battery-less and sustainable using harvested energy. Second, they must embed energy harvesting circuits capable of harvesting from one or more harvesting elements as well as storage components (e.g. super-capacitors) capable of a high number of discharge and recharge cycles. Third, these devices must operate reliably with an ultra-low power (ULP) budget due to the limited power provided by harvesting elements. In order to stay within this budget, many components like the SoC and sensors must be redesigned, often involving extensive research. Fourth, due to the varying nature of harvested power, these devices must be able to recover from complete power loss, either through non-volatile storage or through over-the-air reprogramming. However, both approaches consume a large amount of power and thus must be redesigned for ULP budgets. Finally, after the design process is completed, the manufactured devices must be tested efficiently to speed the time to market.

This dissertation addresses three main concerns facing self-powered IoT product commercialization: reducing the power consumption of IoT devices through the redesign of

off-chip and on-chip communication circuits, improving robustness of IoT devices through the increased integration enabled by on-package communication, and reducing the time to market through reduced post-manufacturing test time and improved application space.

To enable self-powered operation of IoT devices, the SoC within them must meet certain requirements. It must be energy efficient, meaning it must often or always operate in the near or sub-threshold region of transistor operation. This is the region of low voltage operation where the SoC balances active energy, or energy used to perform a task, and leakage energy, which is due to a constant power consumption by the system while it is powered [2]. It must also integrate an energy-harvesting power management sub-system that harvests energy onto the super-capacitor then regulates it before supplying the remainder of the SoC. The SoC power budget must be strict and predetermined during the definition of the SoC specifications because the power management designer must design for a specific load range for improved power efficiency. The SoCs building blocks must also be designed for ULP operation and their power consumption must be reduced. Three types of blocks are the main contributors to power: SRAMs, analog circuits, and off-chip communication. SRAMs and analog circuits are widely researched in the ULP space, but wired off-chip communication is not.

Wired communication allows for IoT devices to communicate with the outside world. This may include sensors used for collecting information about the device's environment, peripheral devices such as memory or accelerators that the IoT device does not have on chip, other IoT devices for sharing of information which will allow them to work as a group, and even aggregators which will further process collected data and provide access through the internet. Figure 1.1 shows the proposed space of wired communication research in comparison with existing technology. Much time has been spent researching high bandwidth wired communication and high power low bandwidth communication, however minimal time has been spent studying new designs of low power, low throughput communication, as there has been no motivation for exploration until the recent interest in self-powered systems for the IoT. Low power, low throughput wired communication not only allows data transfer

between sensors and the SoC, but also enables higher level of integration when used to connect multiple systems within a compact system in package (SiP). Thus, the field of self-powered, ubiquitous computing will benefit immensely from improved wired communication.

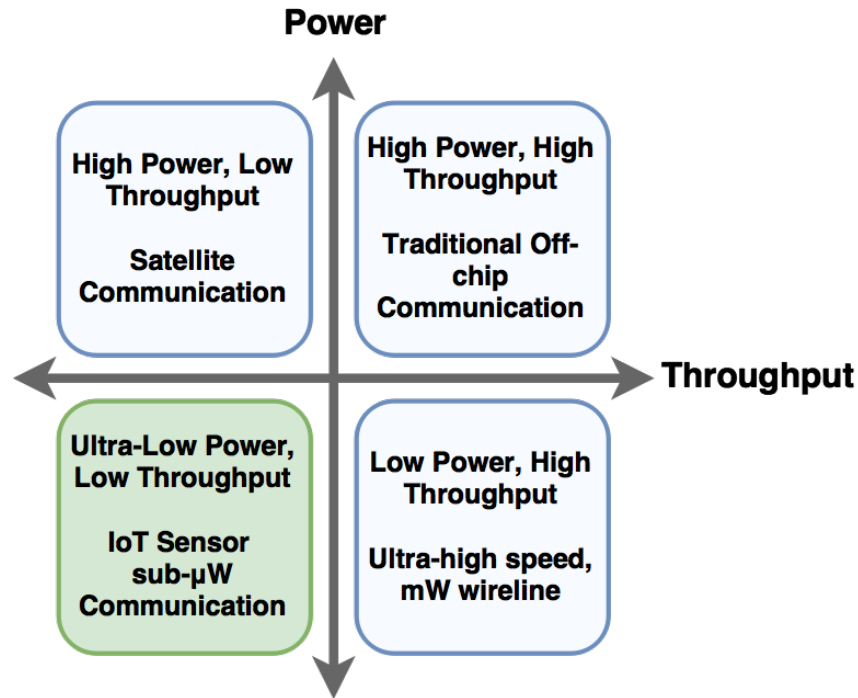


Figure 1.1: Design space of communication. The focus of this dissertation, ultra-low power, low throughput wired signaling, shown in green (darker), has not yet been researched.

Reducing the power consumption and improving the robustness of SoCs are the first steps to enabling widespread adoption of self-powered IoT devices. Currently, many new applications have not yet been realized as a consequence of no longer needing a battery. Because of this, there is still much potential to better our knowledge of how these devices will advance the quality of our every day lives. Thus, these SoCs must prove their usability in providing useful data and solving complex problems. Creating new application spaces, improving circuit and system designs, and establishing methods for manufacturing and testing will help enable self-powered ubiquitous computing.

1.2 Thesis

In order to promote the adoption of self-powered IoT systems, SoC power must be reduced to levels that can be provided by harvested energy, the system must be able to operate in a variety of environments with inconsistent harvesting conditions, and the time to market must be reduced to enable commercialization. As a result, new circuits must be designed for lower power operation. Sub-system components must be designed to handle varying conditions and recover when conditions remain poor for long periods of time. Then, an approach to overcome hurdles in time to market must be developed. Finally, demonstrating applications will motivate more widespread use of products.

1.2.1 Reducing Power

Reducing the SoC power budget allows for the system to operate solely on harvested energy. Further reductions in power improve SoC lifetime and reliability. Communication is a large portion of an ULP SoC power budget. Thus, designing communication circuits to operate at lower voltages and frequencies will improve circuit power, and as a result, reduce the SoC power budget.

1.2.2 Improving robustness

Communication circuits and interfaces allow higher levels of integration between components in the system. Designing these interfaces from the ground up along with the components will allow new system features to improve the sophistication and robustness of the platform. Off-chip communication will enable multi-node systems, and on-package communication with an NVM will allow the system to recover from power loss.

1.2.3 Reducing time to market

Demonstrating applications for self-powered IoT systems and improving chip test time will reduce the time to market. Showing IoT applications that typically would require a battery powered system will display benefits of self-powered systems in that space. Reducing functional test time in subthreshold SoCs will reduce expenses and time required in bringing the product to market.

1.3 Approach

These challenges will be addressed through the following: reducing power in communication circuits, creating a more robust SoC and SiP through higher levels of integration, and demonstrating new applications and test methodologies to reduce time to market.

1.3.1 Reducing Communication Circuit Power

Off-chip communication is a fundamental requirement in any computing system. In ULP systems, they consume a large portion of the system power budget. Methods for allowing communication circuits to operate at a low voltage and handle small analog signals will be explored. Off-chip communication for short distances is studied to replace high power solutions like SPI in ULP SoCs. A transceiver is fabricated in 130nm CMOS technology, and communication is verified through traces on a printed circuit board (PCB). The resulting power consumption is 1.24 nW and the energy cost for transmitting each bit is 0.38 pJ.

We then explore communication for longer distances and higher quantities of noise. We develop a transmitter with no static current, and avoid static current on all but one receiver component, for which we have a sleep mode. The transceiver is fabricated in 130nm CMOS technology and demonstrated using conductive textiles and Ethernet cables as interconnects. The measured power consumption is 3.77 nW and the energy per bit transmitted per mm is 11.4 fJ/b/mm.

We then look into reducing the power consumption of the complete SoC through efficient on-chip communication between its building blocks. An on-chip bus is created to improve power at the block level of the SoC, allowing each block to operate at either its minimum power point or minimum energy point. SR latches are utilized to allow each block to run at its own independent voltages while still reliably communicating with the rest of the system. The resulting fabricated SoC saves an average of 45.2% power per logic block.

1.3.2 Introducing a System in Package with Improved Integration and Robustness

In this approach, we introduce a ULP on-package interface to improve the systems robustness to power failures. This interface allows the SoC to recover critical data after power loss. The on-package interface uses voltage reduction features as well as a parallel interconnect to improve energy. The master to slave transmitter is a single serial line, while the slave transmitter is an 8 bit parallel line. The transmitter uses binary weighted CMOS drivers with an additional pull-down transistor to create a reduced swing digital signal. A sense amp at the receiver collects the reduced swing signal and returns it to full swing digital logic. The on-package bus is fabricated in a CMOS 130nm and an FeRAM 130nm technologies. The on-package bus consumes 8.49 nW and has a normalized energy of 1.09 fJ/b/mm.

We then explore the system level effects of the circuit improvements through the design of an SoC and later a System In Package (SiP). The improved off-chip communication allows for considerable decrease in average system power, and on-package communication allows for improved system integration by enabling booting the platform from a non-volatile memory (NVM). The SoC also shows reduced power through co-optimization of sub-system components. The Power Monitor (PM) and Cold-boot Management System (CBMS) are tightly integrated, the main controller is tightly integrated with the instruction memory, and the radio interface is tightly integrated with the data compression algorithm. These components work together to improve integration and functionality without increasing the

system power budget. The SoC is integrated with the NVM and radio to create an entire platform in a single package. The system's numerous features enable it to have a total power of 507 nW while sensing for fall detection.

1.3.3 Applications and Test Time Reduction in Ultra-low Power Systems

To address issues hindering IoT commercialization, we demonstrate potential commercial applications for their use. We demonstrate free-fall detection, a solution for tracking and quantifying shipping integrity. The SoC interfaces with an accelerometer to detect when a free-fall event occurs, collects accelerometer data at the time of the event, and transmits this information compressed over the radio. During this application, the peak power remains below $2\ \mu\text{W}$, and while sleeping consumes 490 nW. Another application demonstration shows inertial positioning as a supplement to a radio based GPS. The SoC collects continuous acceleration data through the accelerometer, numerically integrates twice over a window, and achieves a new position based on these measurements.

To address manufacturing issues in commercialization, we explore methods for reducing functional test time in the systems presented in this dissertation. Predictions are made using statistical analysis and will allow test time at increased voltages while predicting delay and other metrics at the lower subthreshold operational voltage. Test time is reduced to 18.6% of the standard at-speed time it would take to test at the deployed operating voltage.

1.4 Contributions and Organization

The goal of this dissertation is to further enable ubiquitous IoT computing. The contributions leading to this goal are developing robust ULP wired communication for IoT devices to reduce total system power, improving robustness of the system as a whole, and enabling

commercial acceptance of such devices through proven applications and improved testing for commercialization and manufacturability. The remainder of the dissertation is as follows:

1.4.1 Background

Chapter 2 gives more insight into the background of communication circuits for ULP systems and provides the necessary information to understand the contributions of this dissertation.

1.4.2 Ultra-low Power Off-chip Communication

Chapter 3 presents two transceivers for off chip wired communication. It explains the disadvantages of existing off-chip solutions and the need for new contributions in this area for IoT applications. The first transceiver targets short distance and low noise communication between chips placed on a PCB. We also present a design methodology that aids in the design of I/O circuits that minimize energy and power. The second transceiver targets longer distance communication with improved noise immunity. This transceiver is not only designed to consume low power but also to enable multi-node systems to improve robustness against failures.

1.4.3 Ultra-low Power On-package Communication

Chapter 4 proposes the use of more efficient communication between chips on a package, showing how existing research is still too inefficient for continued self-powered operation. This chapter introduces an on-package bus to enable the SoC to communicate with an NVM to recover critical data after power loss. This bus also allows designers to leverage the benefits of different technologies without compromising form factor. We also introduce some of the main design considerations to take into account when designing an SiP. These include the bus width, pad budget, integration technologies, and others.

1.4.4 Ultra-low Power On-chip Communication

Chapter 5 explores on-chip communication to reduce system-level power consumption. An on-chip bus is introduced to enable communication between blocks that run at different voltages, while providing level shifting and isolation between different voltage domains. The proposed bus relies on a ULP SR latch for synchronization, level shifting, and isolation. The on-chip bus was fabricated in 130nm CMOS technology.

1.4.5 Enabling Ubiquitous Self-powered Systems

Chapter 6 introduces other methods which will contribute to enabling ubiquitous commercial self-powered IoT devices. This will mainly include additional applications of self-powered systems and overcoming obstacles in the manufacturing and test process. This chapter starts by presenting a ULP sub- μ W SiP and a few potential applications, including a shipping-integrity tracking application and ULP positioning application. Next, a method to reduce the test time of subthreshold circuits is presented. The required time, and as a result, cost of testing these new systems is high due to the low operating frequency of subthreshold circuits. In this section, we introduce a trans-threshold correlation algorithm which still retains a reliable pass or fail result while reducing the test time of these circuits.

1.4.6 Conclusion

The final chapter concludes this dissertation and highlights the high level impact of this work.

Chapter 2

Background

2.1 Ultra-low Power Circuits

Process technology transitions have regularly been implemented since the 1960s to reduce power for computer processors [3]. Heat limitations were met as processors consumed more power due to increased complexity. As processes moved from BJT to NMOS to CMOS logic, power was reduced to reduce heat and enable larger integration and more functionality. CMOS technology eventually became most common, as the circuits consumed very low leakage between operations.

Reducing CMOS circuit power began with power supply reduction. As process technology scaled, threshold voltages scaled as well. Research explored how transistor threshold voltages and supply voltages can be tuned for power reduction [4]. This work points to reduced voltage operation, and reducing the threshold voltage to maintain performance. Later work explored finding the minimum possible energy a circuit can consume by tuning the same knobs [2]. The conclusion was that the minimum energy point in circuits is generally below the increased threshold voltage, resulting in further research into subthreshold circuit design. New challenges arise in this mode of operation, leading to new types of circuits designed to avoid the pitfalls of increased variation. As the benefits to reducing power became more understood,

techniques like dynamic voltage scaling (DVS) and dynamic frequency scaling (DFS) have been employed to manage the trade-off between power and performance. Additionally, at the circuit level, techniques utilizing higher threshold devices and non-CMOS logic families further reduced power.

Entire subthreshold systems have been built using techniques specifically for ultra-low power consumption [5]. New work continues to increase the complexity and reduce the power of these types of systems. Although many techniques have been used for power reduction in digital and analog circuits, little work has been done towards utilizing these techniques to improve wired communication.

2.2 Wired Communication

Wired communication is the transmission of data from one block of logic to another block of logic through wires. This may be between two blocks of logic on a chip [6] [7], between two blocks of logic on different chips within a package [8], or between two blocks of logic in different packages. This is useful for systems to relay information from the processor to accelerators, collect information from off-chip sensors, or allow larger scale integration through stacked chips in a package.

Wired communication consists of a transmitter, a medium for the signal to travel across, and a receiver to read the signal. The transmitter sends the data across the medium in the form of a signal or pattern, and the receiver detects the electrical changes in the medium according to a predefined set of rules. The receiver then provides readable data to be processed by logic. A clock is needed on the receiver to synchronize the incoming data. In some cases the clock signal is sent in parallel with the data, while in other cases the data is encoded in such a way to allow for clock and data recovery (CDR) [9].

2.2.1 Wired Off-chip Communication

When an SoC communicates with another device not contained on that chip or package to relay information or complete a task, this is called off chip communication. Often an SoC will not contain all of the components needed to complete a task for a given application. Some example off-chip peripheral components are off-chip memory [10], environmental sensors [11], or a base station for collecting processed data. An SoC may not have non-volatile memory or may want to collect and store large quantities of data. In this case, off-chip memory will be used to store instructions for when the SoC powers up or for storage of extra data. An SoC cannot have a sensor for every possible application, so an off-chip sensor is often used which is best suited for the task. All of these examples require improvements in off-chip communication. In Figure 2.1, off-chip communication is shown in green.

Off-chip wired communication is performed in many different protocols for many purposes. Common protocols used by SoCs and microcontrollers are Serial Peripheral Interface (SPI) [12], Inter-Integrated Circuit (I²C) [13], and Universal Asynchronous Receiver / Transmitter (UART) [14]. SPI is often the most energy efficient because it does not consume static current and is simple by design. Other protocols like Peripheral Component Interconnect (PCIe) [15], Ethernet (IEEE 802.3) [16], and Transition-Minimized Differential Signaling (TMDS) - used in Digital Visual Interface (DVI) [17] and High Definition Multimedia Interface (HDMI) [18] are high throughput serial protocols used by computers to communicate with peripherals or other computers, and transmit audio and video information to computers designed to display that information.

2.2.2 Wired On-chip and On-package Communication

SoCs have multiple logic building blocks that work together as a system. In order for those building blocks to work together, they must be able to communicate over a common medium. For on-chip circuits this medium is most commonly an electrical bus of wires in parallel. This bus is managed by a logic block acting as the bus master, organizing the flow of information

from the output of one block to the input of the next. Buses can contain more than one master, but only one will manage the flow of information at a time. In Figure 2.1, on-chip communication is shown in red.

SoCs can be made from multiple groups of building blocks on different chips created using different process technologies. This allows leveraging the benefits of each technology and integrating them into a single system containing the most efficient circuits. A self-powered SoC may contain energy harvesting circuits which need to harvest from a high voltage source. Designing the entire system in a process technology for operating at these voltages would be very inefficient for the rest of the logic. This same SoC may require non-volatile memory which is not available in most process technologies. Having a system that utilizes high voltage technology for harvesting, non-volatile memory technology for storing instructions, and energy efficient technology for the remaining logic is ideal, but requires communication between these chips. On-package communication circuits solve this problem, allowing these technologies to work together. In Figure 2.1, on-package communication is shown in blue.

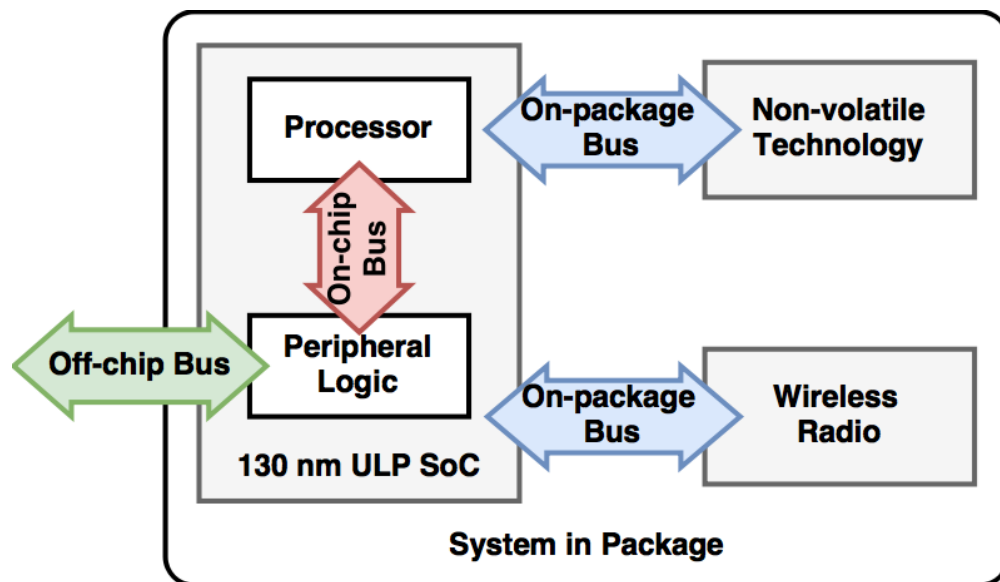


Figure 2.1: Diagram showing types of wired communication.

2.2.3 Commercial Wired Communication

Wired communication is pervasive in commercial devices. Universal Serial Bus (USB) [19] is used to connect peripherals to computers, headphones are used to transmit audio information to portable speakers, and high bandwidth fiber optics are used to transmit data to allow the internet speeds we have today. While wireless communication is becoming more popular, wired communication provides a more efficient means of transmitting data and at a higher throughput, making it irreplaceable.

Commercially, wired communication has not targeted energy efficient, low power communication because historically devices using wired communication have had unlimited continuous power from the power grid. With the advent of laptops, wired communication standards have not changed, as this is still a small percentage of the laptop power budget. As commercial self-powered Internet of Things (IoT) devices emerge, new standards will need to be developed. On-chip wired communication has seen some developments in buses for low power SoCs with the ARM Advanced Microcontroller Bus Architecture (AMBA) Advanced Peripheral Bus (APB) [20]. However, these are architectural level bus protocols, leaving space for circuit level optimization.

2.3 Systems On Chip

Systems on Chip (SoCs) consist of a main controller with various peripheral blocks for improving the speed and energy of solving a given problem. They are often used in mobile platforms as they contain almost all the components for building a computer on their own [21]. In addition, they are generally designed for lower power applications, and as a result are generally lower performance than a typical desktop system. SoCs are becoming more sophisticated as they integrate more components onto a single chip.

2.3.1 Systems in Package

Systems in Package (SiP) seek to further integration by placing and interfacing multiple chips in a single package. Interfacing can happen between wire bonds or flip chips using solder balls. These techniques also allow stacking of chips for an even smaller footprint. SiPs often integrate different technologies in a single package to leverage the benefits of each. They also improve delay and energy between components.

2.4 Chip Testing Process

In order for these chips to be sold they must be tested to ensure they work properly [22]. This testing process can be separated into two categories. Pre-manufacture testing, and post-manufacture testing. Pre-manufacture testing involves ensuring that the design has the correct functionality, as well as ensuring that this functionality solves the right problem. To ensure correct functionality, the design is rigorously tested for bugs. This process happens again once the chip has been fabricated, as not all functionality can be simulated. Once the checks have been made to verify functionality, each and every chip is tested before being shipped out as a product [22]. Transistor threshold variation causes some chips to work slower than others, and as a result causes failures. In addition, pieces of metal can short wires on chip, or wires on chip can contain voids, causing unintended changes in how the logic functions. Each chip needs to be tested because of the potential for these variation and manufacturing defects.

Chip defects are often modeled to determine their effect on the system. This is known as fault modeling and often involves simulating specific circuit failures [23]. Defects in the chip can be broken down into a handful of fault models [24]. A stuck-at fault happens when the output of a gate is either stuck at 0 or 1. A bridging fault happens when two logical values remain the same when they should be independent. Delay faults happen when the critical path of logic is transitioning to the proper value, but not in the expected amount of time.

Delay faults are more common in subthreshold circuit design due to the log-normal delay variation distribution.¹ Because testing for delay faults is done at-speed, conventional testing methods used for subthreshold circuit testing will consume exponentially higher time. This dissertation will explore solving this problem.

¹Delay is equal to $\frac{CV}{I}$. Subthreshold current (I) is proportional to $e^{-V_{th}}$ where V_{th} is the threshold voltage. Threshold voltage has a normal distribution, thus subthreshold circuits display a log-normal distribution in delay.

Chapter 3

Ultra-low Power Off-chip Communication

3.1 Introduction

Self-powered SoCs must operate at low clock frequencies and low voltages to provide the necessary efficiency to operate from harvested energy. The low voltage, low frequency operation is possible due to the low sampling rate of the sensors interfacing to the SoC and the limited data processing required on-chip. Because of this, the required bandwidth of the data communication to and from these SoCs is less than that of other systems like cell phones and desktop computers. This low bandwidth requirement allows for an increase in energy efficiency primarily through lower voltage transmissions. ULP communication will be used for communicating status and configuration between SoCs, as well as data that may have been collected from sensors to be processed by another SoC or transmitted to an aggregator. These communications can be wireless, wired through a PCB trace, or even through textiles worn on the body (for body sensor networks).

Wireless transmissions consume high power and have low energy efficiency when compared

to wired communication due to the inverse-square law.¹ Heavy duty cycling is often needed to keep its power within the system budget [25], thus wireless communication should be avoided when possible. On the other hand, little work has been previously done in ULP wired communication, however large system level power improvements can be achieved through wired communication with few trade-offs. In this chapter, we present two transceiver designs that allow ULP wired communication between self-powered SoCs. The circuits presented are orthogonal to many existing energy reduction methods (encoding and bus topologies), meaning they can be expanded to include these methods for necessary applications. Off-chip wired communication is signaled in two major forms: single ended and differential. CMOS circuits can often be used in single ended signaling when operating at a low enough bandwidth, resulting in very energy efficient data transfer. Differential signaling allows for common mode noise rejection, allowing for signaling over a longer distance, but this comes at the cost of a more complicated design, more wires for transmission, and often more static power. This chapter presents a single ended transceiver for short distance communication on PCBs (up to approximately ten centimeters) and with little noise, and a differential transceiver for longer distances (meters) and noisier environments.

3.2 Motivation

ULP communication enables self-powered systems to relay collected information, providing rich data sets. However, these systems have a strict power budget in order to continually operate on harvested energy. Figure 3.1 shows the power breakdown of a self-powered SoC [26]. Most of the power budget is consumed by few components such as SRAMs, analog circuits, and I/O [25]. While extensive research is being done to reduce the power consumption of SRAMs [27] and analog circuits [28], ULP I/O on SoCs is not an area commonly researched,

¹The inverse-square law shows energy seen at the receiver to be $E = \frac{Power}{4\pi r^2}$ where r is distance. Low throughput wired communication follows $E = \frac{1}{2}CV^2$ at the transmitter where C is linearly proportional to distance.

thus large potential gains are possible in terms of energy efficiency and power reduction by redesigning I/O circuits for the IoT ULP space.

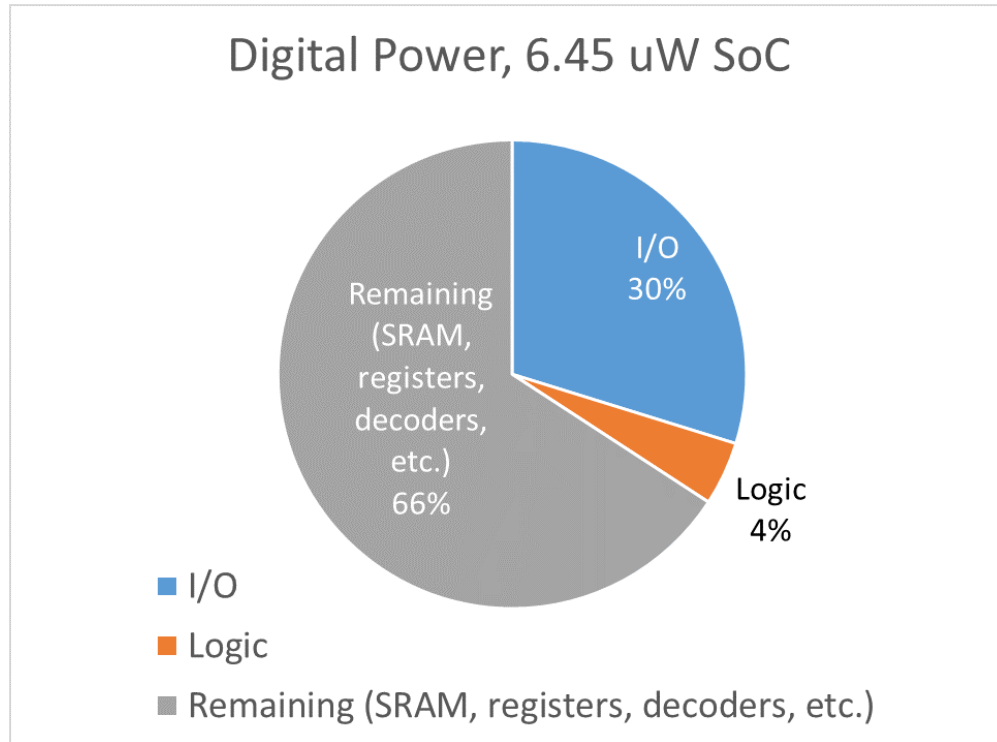


Figure 3.1: Pie chart of typical ULP SoC power breakdown. I/O is a large portion of the power budget, motivating research into lowering power.

ULP communication also enables the development of many-node systems. A many-node system improves robustness by allowing a single node failure without loss of status, and each node can be optimized to perform a task specific to its location. Health monitoring systems, for example, can have heart monitoring sensors, temperature sensors, respiration monitor and activity sensors placed around the body. However, the problem of reliably communicating between such on-body ULP $< 10 \mu\text{W}$ sensors placed in different locations remains unsolved. Existing off-chip communication consumes over 10% of the $< 10 \mu\text{W}$ digital system budget in ULP SoCs even after heavy duty cycling [25] and will not reach long distances. Wireless communication fits many needs, but on-body applications have limitations on power, form factor, or environment that discourage wireless solutions. Thus, ULP communication circuits

are required to enable transitioning from a single purpose body monitoring sensor node to an advanced on-body sensor network by significantly reducing the communication cost.

Other applications can also benefit from this type of communication. For example, a self-powered SoC can be attached to load bearing beams on a building or bridge to monitor vibrations. In this case, a status representing a fault or failure can be transmitted instead of raw data, creating a long distance, low bandwidth requirement. In a similar application, these systems could be used to monitor geological events. If multiple vibration sensing SoCs were placed quite some distance apart, they could use long distance wired communication to determine attributes like time, distance, strength, and direction of an earthquake in real time.

Tsunami detection systems will also benefit from long distance, energy efficient communication by allowing them to become self-powered. Energy harvesting would happen on the detection buoy from wave energy and solar power, while a pressure sensor on the sea floor communicates with the buoy through a wire connected through the anchor. Improved energy efficient, long distance communication will allow this system to remain self-powered while detecting Tsunamis long before they reach land, potentially saving lives.

To enable this wide application space, we optimize the design of low throughput communication transceivers by dividing them into two categories: a short distance, low noise transceivers and long distance, noise immune transceivers. The former is mostly used when multiple sensors and processing elements need to communicate over low-noise PCB connections. The latter is used when sensors need to be distributed over long distances and noise sources might reduce the integrity of the transmitted signals. We optimize the design of these transceivers for different throughput requirements to make ULP I/O a viable solution in self-powered systems.

Applications involving sensing which do not require a high bandwidth can benefit from ULP I/O. One such application is health monitoring systems where multiple sensors are placed around the body to sense different information. This body sensor network will have communication between chips on a PCB as well as communication between those chips,

sensors, and other processing nodes spread throughout the body. Most PCB connections will be short distances and noise can be ignored, while others such as between nodes worn on different parts of the body will need more noise resistant circuits for longer distances traveled. The problem with existing protocols such as I²C and SPI is they target a higher throughput and were not designed for such low power operations. Heart sensors, temperature sensors, and respiration monitors, as with most bodily functions, can be sampled at low frequencies and will have low bandwidth requirements, and as such, can benefit from new designs with these relaxed constraints.

For low throughput I/O circuits to be utilized in many IoT devices, they also need to be researched for differing applications with different throughput requirements and transmitting long distances in noisy environments. Having a range of optimal designs for different self-powered applications will make ULP I/O a viable solution in self-powered systems.

3.3 Prior Art

There is much prior work involving energy efficiency in off-chip signaling, with recent work in the hundreds of fJ/bit range. However, as previously stated much of this work seeks to optimize for energy in conjunction with performance instead of power, leaving a void in ULP, energy efficient signaling. A ground referenced serial link [29] achieved 0.54 pJ/bit along a 4mm link between chips. This work operated at 20Gb/s and consumes milliwatts of power - orders of magnitude too high for an IoT device. Utilizing a similar design will not solve this power problem because static power (450 μ W) will not change with decreased frequency, resulting in a power floor still too high for self-powered, IoT budgets.

Current state-of-the-art energy efficient communication circuits use the energy-per-bit metric as a key comparison point. However, optimizing energy-per-bit alone is a problem for ULP systems if the solution is increasing frequency and activity factor (probability of switching in a given clock cycle). Because IoT systems have both power and energy

constraints, these changes can result in consumption levels above the power budget, making it an incomplete solution. In addition to this, circuits optimized for high frequency operation are designed with the assumption that a large amount of data is waiting to be transmitted. This is often not the case in the low-energy, low-performance domain.

Because of the need to amortize static and leakage energy, current research for chip-to-chip I/O is pushing into higher frequencies to reduce energy-per-bit. Figure 3.2 shows that the lowest E/b designs use very high frequencies. Solutions for lower data rate communications are lacking. For example, the I²C standard (up to 5 MHz) assumes a termination resistor of approximately 5k Ω that consumes significant static power which is amortized by increasing frequency. SPI has a similar typical max frequency but avoids the static current of I²C. Both standards assume a load capacitance of 5pF. Overall, the figure shows the need for lower energy I/O circuits at lower frequencies.

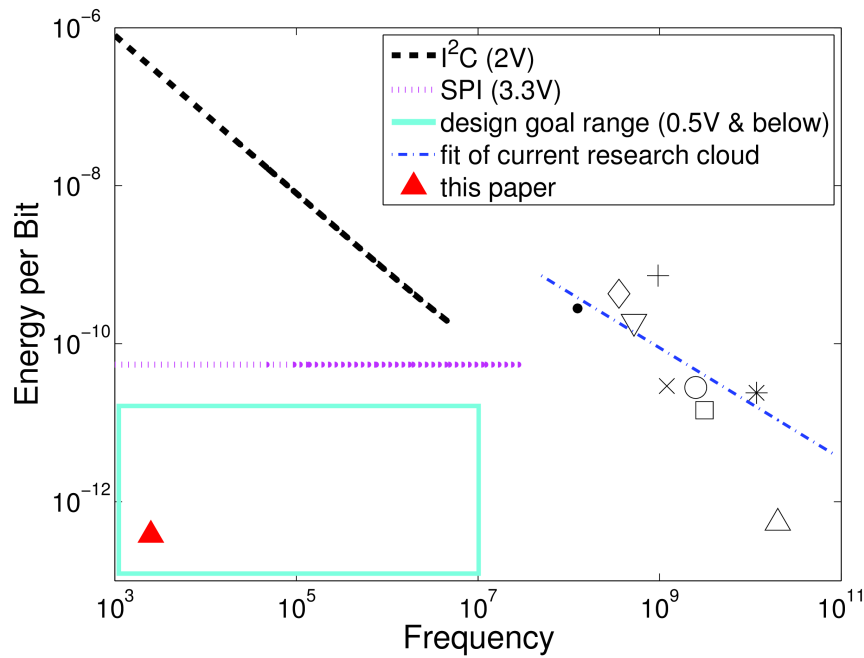


Figure 3.2: De Facto standards and state of the art communication circuits.

3.4 Chip-to-Chip Serial Link for Ultra-Low Power Systems

This work seeks to optimize an off-chip serial link in terms of energy and power, assuming noise on the transmission line is minimal. The equation for energy of an operation, $E_{op} = CV^2 + I_{leak}Vt$ shows a quadratic savings in energy with reduction in operating voltage. As a result, the main goal of this work is to create a circuit with low leakage, operating at the lowest voltage possible.

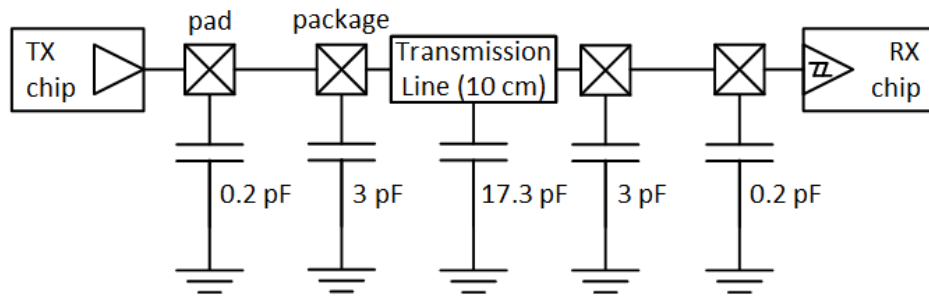


Figure 3.3: Model of single ended transmission line. Series inductors and resistors are ignored at low frequency operation. Clock and data lines were both designed for capacitance and in turn delay matching.

This type of design is acceptable for transmission over short distances such as PCB traces on a body sensor node, on the order of a few centimeters. Because CMOS circuits are energy efficient, an inverter based driver was used as the transmitter, allowing operation over a wide range of voltages and frequencies. A Schmitt trigger was used as the receiver to filter noise during transmission due to low frequency and slew rate at low voltages. Figure 3.3 shows the capacitive load which must be overcome by the driver. The clock signal was sent in parallel with the data signal, as the overhead of a CDR is a higher cost in power than the rest of the circuit. Once a minimum energy point was discovered, a trade-off between energy and power was explored. Figure 3.8 shows that lower voltage operation yields lower power and energy, below that of existing state of the art energy efficient designs.

3.4.1 Energy and Power Co-optimization

Energy constrained digital circuits minimize the energy-per-operation metric by determining the optimal operating voltage [2]. However, in transmission circuits where there is added control over the activity factor, and in turn, throughput of the circuit, determining the optimal operating voltage using this method is no longer enough. Activity factor provides a new knob that allows further reduction in energy-per-cycle.

Considering both metrics will allow us to create a communication circuit more suitable for ULP systems. This new design comes at the cost of lower throughput. However, IoT or body sensor nodes often will be transmitting data with a low throughput requirement, so reducing power at the cost of throughput is an acceptable trade-off.

Energy-per-cycle is defined as the average energy consumed by a circuit during one clock cycle. This energy includes both sleeping and active cycles. Energy-per-bit (3.1) differs from energy-per-cycle (3.2) in that it has an opposite relationship with activity factor (α). Lowering activity factor in (3.2) results in a lower energy-per-cycle and therefore a lower average power at a cost of lower throughput. The same change in activity factor results in an increased energy-per-bit due to more time spent leaking for each bit transmitted. Equation (3.1) assumes transmission of pseudorandom data across the transmission line.

$$E_{bit} = \frac{CV_{DD}^2}{\frac{1}{\alpha} - \frac{2}{1-2\alpha}} + \frac{W_{eff}L_{DP}KC_gV_{DD}^2e^{-\left(\frac{V_{DD}}{nV_{th}}\right)}}{2\alpha} \quad (3.1)$$

$$E_{cycle} = \alpha E_{dyn} + E_{leak} = \alpha CV_{DD}^2 + I_{leak}V_{DD}t \quad (3.2)$$

In ULP sensor nodes, off-chip communication (wireless or wired) has often been the largest energy consumer when operating at the nodes maximum data rate [25]. By reducing the activity factor, the energy consumption for a given cycle of these components will fall. This knob allows for optimization that is unique to communication circuits. In addition, by varying the operating voltage of a circuit, it is possible to calculate the minimum energy-

per-cycle for a given activity factor in order to find an optimal operating voltage (3.4) [2]. [2] assumes that the circuit is operating at its maximum frequency and calls the metric minimum energy-per-operation. Equation (3.5) takes into account that the throughput given by the maximum possible frequency for a voltage of operation may not be required. The optimal voltage (3.4) is a function of β , where n is the subthreshold slope and V_{th} is the thermal voltage. The *LambertW* function is defined as $x = W(x)e^{W(x)}$ and for our purposes is constrained to the branch $W - 1(x)$. In (3.5), C is the total switching capacitance of the circuit, W_{eff} is the average transistor width relative to the characteristic inverter, L_{DP} is the logic depth, K is the delay fitting parameter, and C_g is the load capacitance. As α decreases, β approaches zero and V_{DDopt} increases, as shown in Figure 3.4. This becomes a special case where increasing voltage will decrease the total E/b due to low activity factor, decreased active energy, and less leakage over shorter cycle times.

$$E_{minat} \frac{\partial(CV_{DD}^2 + I_{leak}V_{DD}t)}{\partial V} = 0 \quad (3.3)$$

$$V_{DDopt} = nV_{th}(2 - \text{lambertW}(\beta)) \quad (3.4)$$

$$\beta = \frac{-2\alpha C}{W_{eff}L_{DP}KC_g} \geq \frac{-1}{e} \quad (3.5)$$

The minimum energy-per-cycle can be found for many activity factors to find a trend of minimum energy points. The final curve is minimum energy-per-cycle as a function of activity factor. Solving analytically results in (3.6) [2].

$$E_{Copt} = [nV_{th}(2 - \text{lambertW}(\beta))]^2 \times [\alpha C + W_{eff}L_{DP}KC_g e^{-(2 - \text{lambertW}(\beta))}] \quad (3.6)$$

If a power budget is defined for a communication circuit, this optimization results in a

maximum activity factor and voltage that can be used to operate within a given systems budget. Both energy and power are important to minimize in the design of an ultra-low power transmission line circuit and can be used together to find an optimal point of operation.

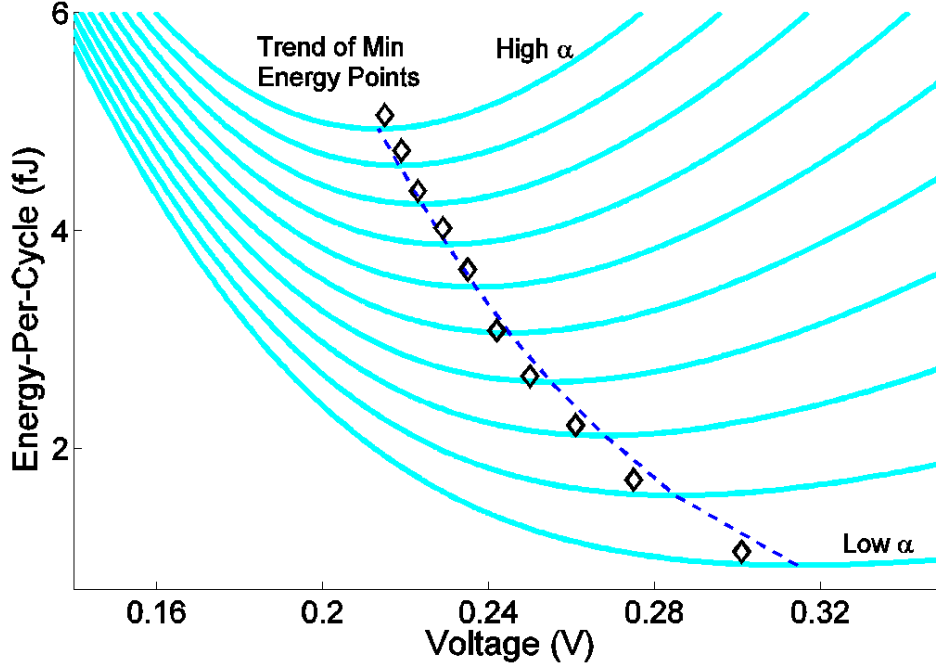


Figure 3.4: Plot of Minimum energy points for a range of activity factors. Each activity factor has an associated optimal operating voltage, creating a new trend of minimum power. Black diamonds show simulation values, while the dashed line shows the model.

The maximum activity factor in (3.4) corresponds to the minimum energy-per-operation from [2]. However, the minimum energy point in [2] may result in a higher throughput than needed for a given application. The circuit's power can be further reduced through lowering activity factor. This produces a new curve with a new minimum energy point. As fewer cycles are spent transmitting data, the leakage energy dominates the curve and pushes the optimal point to the right, increasing its voltage.

Using equations (3.1) and (3.6) we designed a circuit that would minimize energy-per-cycle while still retaining an acceptable throughput. We designed the circuit to have as low leakage as possible, reducing the minimum energy point. Because of the low leakage, the minimum energy point at high α is below the circuits minimum operating voltage. This means we

want to reduce the voltage until it matches our throughput requirements. We fabricated a single-ended SPI-compatible chip-to-chip link in a commercial 130 nm process with one pin for clock and one for data. While a differential transmission is preferred in higher throughput circuits, it requires two drivers and transmission lines, as well as a differential amplifier on the receiving end, which generally require a constant current. This is not ideal for this design space, as we will not be amortizing this energy with higher throughput.

The transmission line is not terminated, as the circuit operates at a low enough frequency where termination is not required. This is a benefit of low throughput communication, as a termination resistor will result in a non-amortizable constant current. Higher voltages will result in faster transition slopes, creating more noise in the transmission line. This design decision limits the maximum voltage of the circuit. Impedance matching isn't necessary, as we are not worried about reflections at low frequencies or maximum power transfer. On the contrary, the highest power efficiency is desirable, so a low source resistance is chosen, but not so low as to be high leakage. Figure 3.5 shows a chip micrograph of the fabricated circuit.

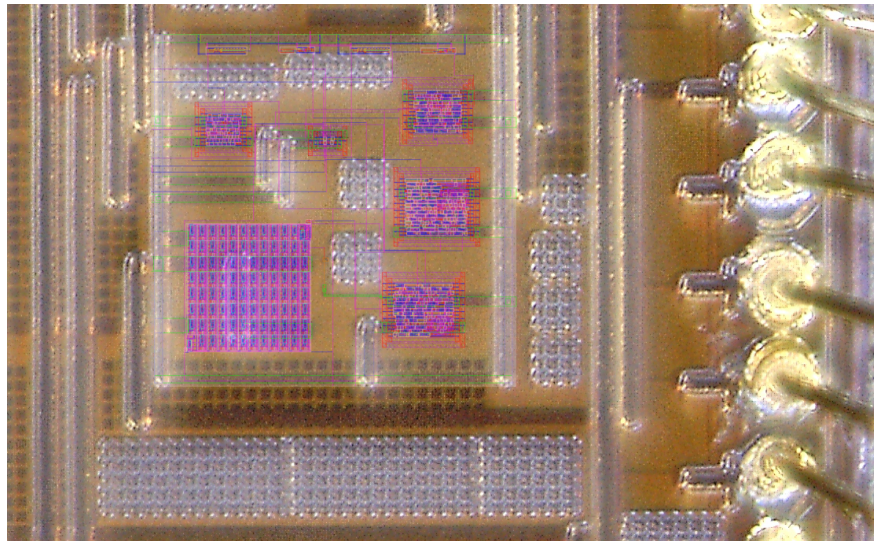


Figure 3.5: Micrograph of the fabricated test chip for serial off-chip communication. Because of low visibility due to the metal fill layers above the circuit, the layout has been overlaid.

The fabricated chips block diagram (3.3) shows the major sources of capacitance, which decrease the energy efficiency of the circuit. We chose to use a 10 cm trace between the

transmitter and receiver to get a pessimistic capacitance on the PCB transmission line. In a non-test environment, the transmission line capacitance can be greatly reduced by both optimizing the distance between the transmitter and receiver, and moving from a PGA package to one with lower inherent capacitance.

3.4.2 Error Correction Coding

To handle noise while operating at low voltages and improve energy efficiency by correcting errors when the transceiver approaches its maximum operating frequency and begins to see errors, Error Correction Coding (ECC) was implemented. Hamming and 8b/10b can be selected for encoding data at the transmitter and then detecting errors and correcting them if possible at the receiver. The problem that arose with this method is that errors did not appear at a gradual rate when the frequency was increased as shown in Figure 3.6. Because the error correction code only can correct a single error per every 7 bits (hamming 7,4) or detect an error every 10 bits (8b/10b), the steep error rate increase did not offer energy or correction benefits for the implemented transceiver.

3.4.3 Results

The proposed transceiver can operate across a wide range of voltages (200mV to above 1.2V). A reference point of operation (voltage and frequency) will be used to compare the energy efficiency and power of the design across the operating range. A typical operating voltage for ULP body sensor nodes is assumed (500 mV) [25]. The accepted methodology is to use the maximum working frequency for the circuit at its operating voltage to reduce energy-per-bit by amortizing leakage, so a frequency of 2.1 MHz is used as a reference. This operating point was found to have energy-per-bit of 2.60 pJ for the fabricated circuit while consuming 5.46 μ W of power.

When using our transceiver design, we found the minimum operating voltage to be 200 mV (3.25 kbps), below which the circuit became unreliable. In a situation where the throughput

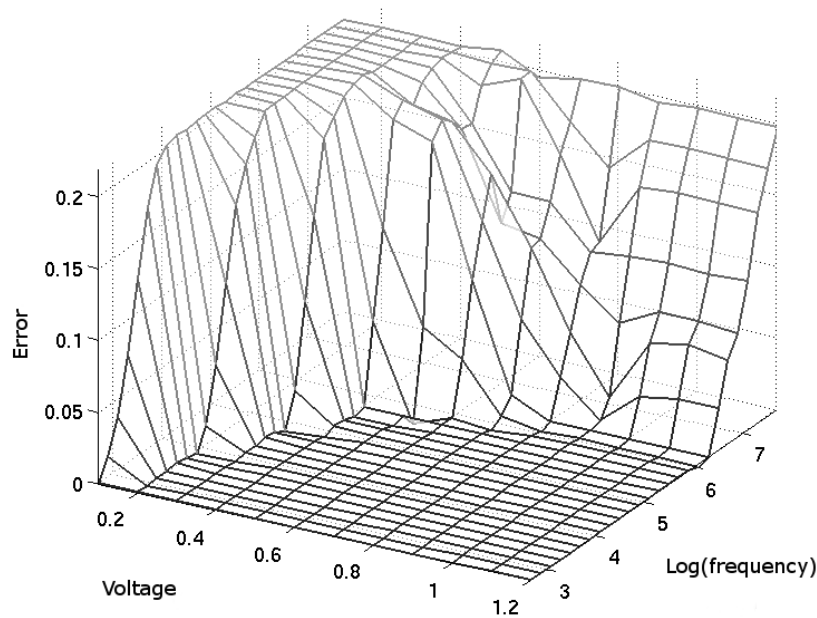


Figure 3.6: Plot showing measured error rate as a function of voltage and bit rate for the transceiver. Due to the steep slope ECC is not energy effective.

Table 3.1: A comparison of a typical SPI transmission at ULP SoC voltage verses the optimized transmission of our fabricated circuit.

	Reference	Optimized E/b
Voltage	500 mV	200 mV
Energy-per-bit	2.6 pJ	0.38 pJ
Power	5.46 μ W	1.24 nW

is not high enough and optimal energy-per-bit is preferred, the voltage can be raised until the requirement is satisfied. If optimal energy-per-cycle is preferred due to lower throughput requirements, the minimum energy-per-cycle curve can be followed by lowering activity factor only enough to satisfy the throughput requirements. Table 3.1 shows a comparison between the reference point and the optimized energy-per-bit.

Using the optimal point of operation along with the ULP circuit design strategies, we were able to create a circuit that operated with 30% lower energy-per-bit than the state of the art but with almost seven orders of magnitude less power than that same design. The flexibility of the circuit allows it to operate into the above threshold range to 700mV before

reflections in the transmission line affect the reliability of the circuit. At this point, a high Ohm terminating resistor will allow higher voltage operation.

The black filled square (leftmost) in Figure 3.8 shows the point of minimum energy-per-bit. The curves show the flexibility of the fabricated chip. The black line shows increasing V_{DD} while retaining maximum frequency to increase throughput at the cost of energy-per-bit and power. The dashed blue lines show the use of the activity factor knob to reduce power. The green lines show varying voltage and activity factor to retain constant throughput.

Ideally a circuit will be at the bottom left of Figure 3.8, having both a power consumption and energy-per-bit as close to zero as possible. Pushing to the lower left of the figure is limited by the physical constraints of the circuit, mainly its ability to reduce its operating voltage. However, operating this circuit at 126 mV is possible and results in reduced power consumption (approximately 800 pW) at the cost of severely degraded throughput (60 bps). Once the minimum operating voltage is found, the maximum throughput for that voltage can be determined. The maximum throughput gives the minimum energy-per-bit due to amortization of leakage energy. If this throughput is above the system's requirement, the best course of action is to travel down the blue line in the figure by reducing activity factor until the throughput matches the budget. If the throughput is below the system's requirement the voltage can be increased by traveling up the black line, increasing the maximum frequency of the circuit until the necessary throughput is met.

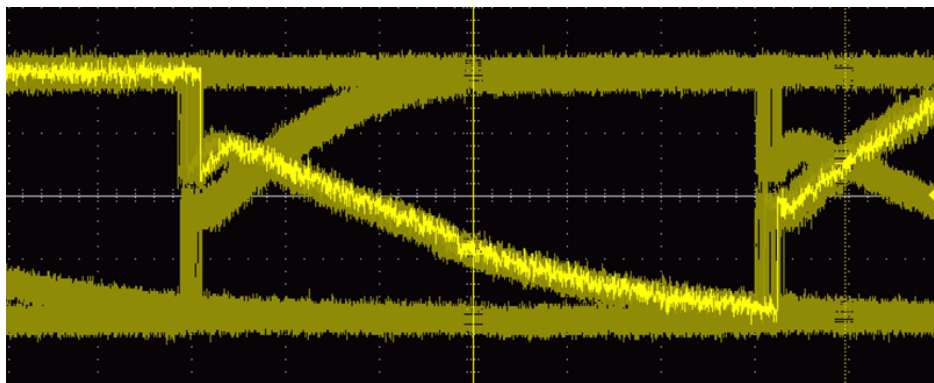


Figure 3.7: Eye diagram of typical transmission measured at the receiver.

Figure 3.7 shows an eye diagram for the transmission line operating at its lowest energy-per-bit at around 200 mV peak-to-peak. Even at the low voltage, the receiver can clearly distinguish between high and low bits. Using this design in an ultra-low power SoC will reduce energy-per-bit spent communicating with other devices by over two orders of magnitude when compared to existing SPI (Figure 3.2). Because this saved energy is often a large part of the systems budget, it can be used to implement other logic blocks and stay within the same energy budget.

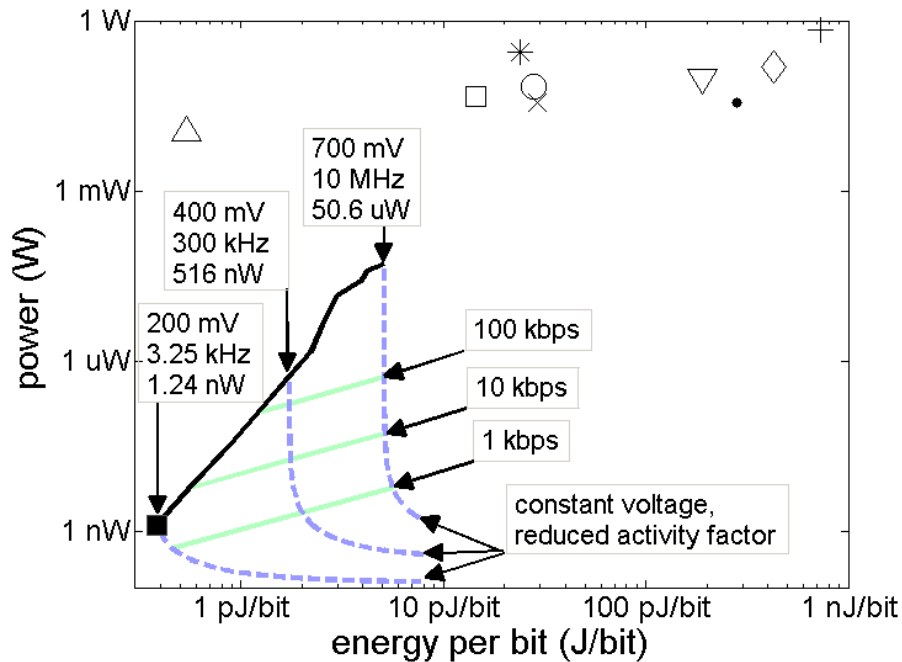


Figure 3.8: A comparison of power and energy per bit with current state of the art

3.5 Differential Link for Reliable Wireline Communication in ULP On-Body Sensor Networks

Minimal energy single-ended wireline transceivers address the power problem but are limited to short distances due to ground drift, cross-talk, radio interference, and IR drop. Differential signaling seeks to solve these problems but has not yet been optimized for minimum energy

or power in the self-powered design space with low required data rates. Since most existing signaling circuit designs optimize E/b using high power solutions at very high data rates, they have limited use in ULP systems in which sampling rates are typically less than 1kHz. Infrequent, high-speed burst transmissions reduce energy efficiency, have high peak current, and will not scale. Thus, in this section, a differential wireline transceiver is introduced to address the need for robust communication that minimizes both energy and power for such systems.

Differential signaling is the use of two transmission lines to send a signal and its logical inverse. The difference between these lines is compared, creating a logical output which ignores signals common to both transmission lines, such as radio frequency noise. As the distance in the line increases, the ability of it to act as an antenna and collect noise increases, increasing the need for noise filtering.

Communicating over long distances (>10 centimeters) will be necessary in IoT devices to carry processed data to a base station. Current wireless technology for communicating over such distances consumes milliwatts of power [30]. Low power wireless technology reduces this power but either transmits at extremely low bitrates or has a limited distance of communication. ULP wired signaling will allow for a more efficient means of communication over long distances. This contribution will also allow for sensor nodes to be embedded in solid objects like asphalt in roads or concrete in buildings and bridges while relaying sensor information about such structures in a way that cannot be done with wireless technology.

3.5.1 Design

We present an ULP, low energy-per-bit, differential wireline transceiver for use in on-body networks and communication over multiple meters. We demonstrate the system using a flexible textile-based interconnect as well as a 3 meter Ethernet cable.

Figure 3.9 shows the transceiver design implemented on a 130nm chip. Compared to most low E/b, higher data rate solutions, this design saves power by removing components with

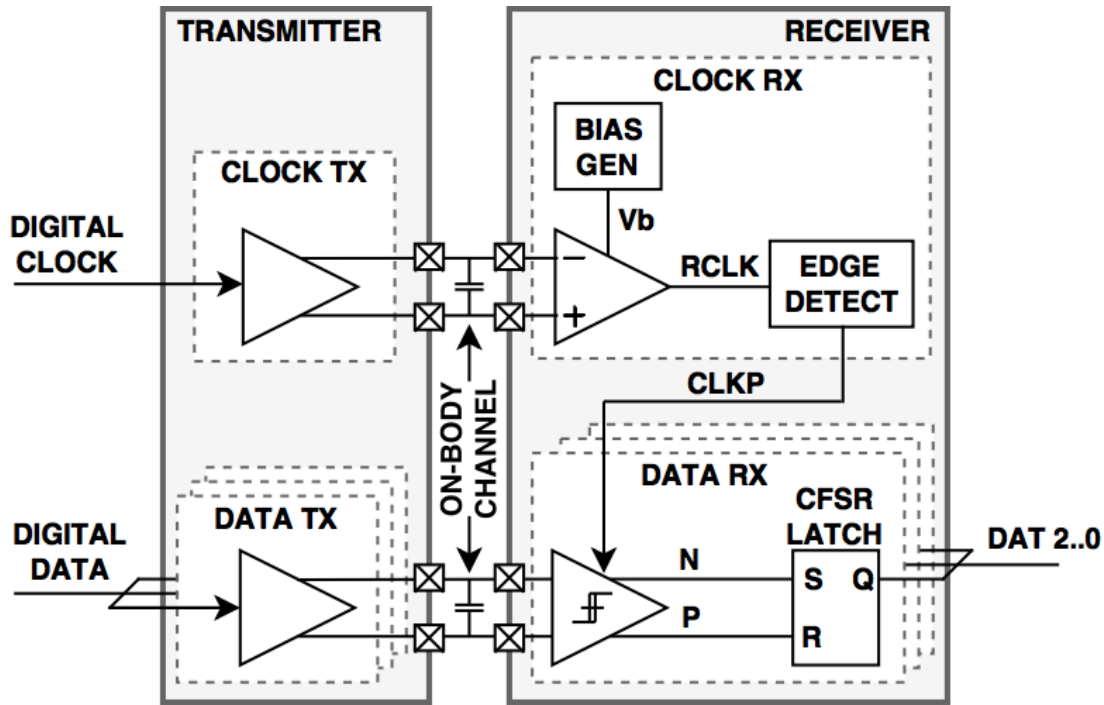


Figure 3.9: Block diagram of transceiver circuit.

static current and replacing them with low power, low leakage designs to benefit from noise cancellation and efficient wired communication. The system uses one differential line to send the clock and parallel differential lines for data.

Figure 3.9 shows the transmitter and receiver circuits and their support structures. The transmitter (Figure 3.10) contains both NMOS and PMOS source followers in parallel that operate with a full swing digital input, creating a CMOS-like buffer driver with a reduced swing output. This provides the signal centered on $V_{DD}/2$ without the static current of a conventional differential amplifier (Figure 3.16). To make the circuit differential, the single-ended transmitter is copied with an inverter on the data input, as shown in Figure 3.10. The clock and data bus lines use the same full transmitter.

The clock receiver circuit (Figure 3.11) is a continuous time comparator using a PMOS current mirror and an NMOS tail to allow for a configurable voltage bias. The bias generator (Figure 3.11) consists of four stacked long channel NMOS devices: two reverse biased diodes, and two with configurable digital inputs. Due to differences in transistor source to body

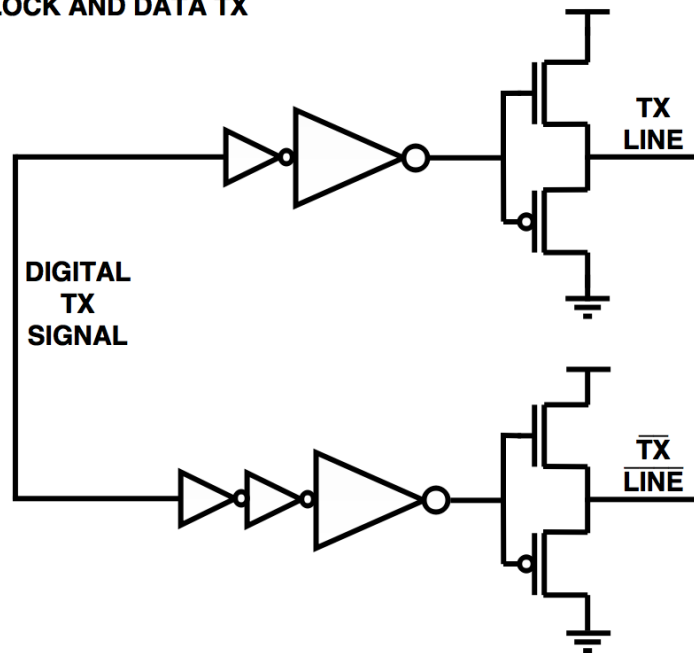
CLOCK AND DATA TX

Figure 3.10: Transistor diagram of transmitter circuit.

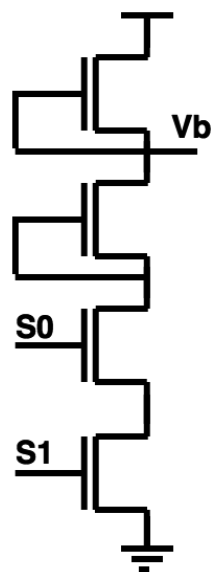
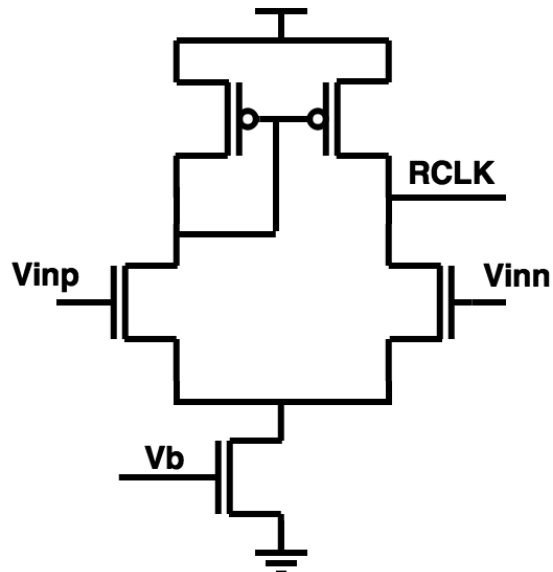
BIAS GEN**CLOCK RX**

Figure 3.11: Transistor diagram of clock receiver and bias generator. Bias generator allows operation over a wide range of voltages.

potential, the generated bias voltage tracks VDD at voltages below 0.3V and flattens out at higher voltages (Figure 3.15). This allows the comparator to operate flexibly at a wide range

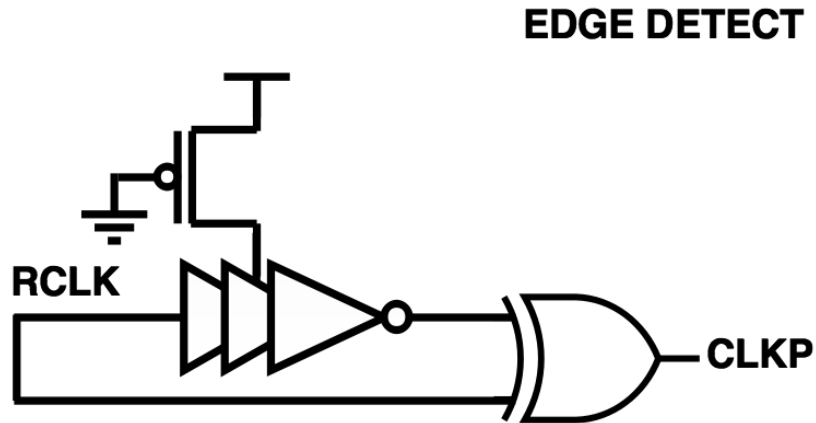


Figure 3.12: Transistor diagram of edge detector. Edge detector turns received clock edges into pulses, allowing the data receiver to latch data on both edges.

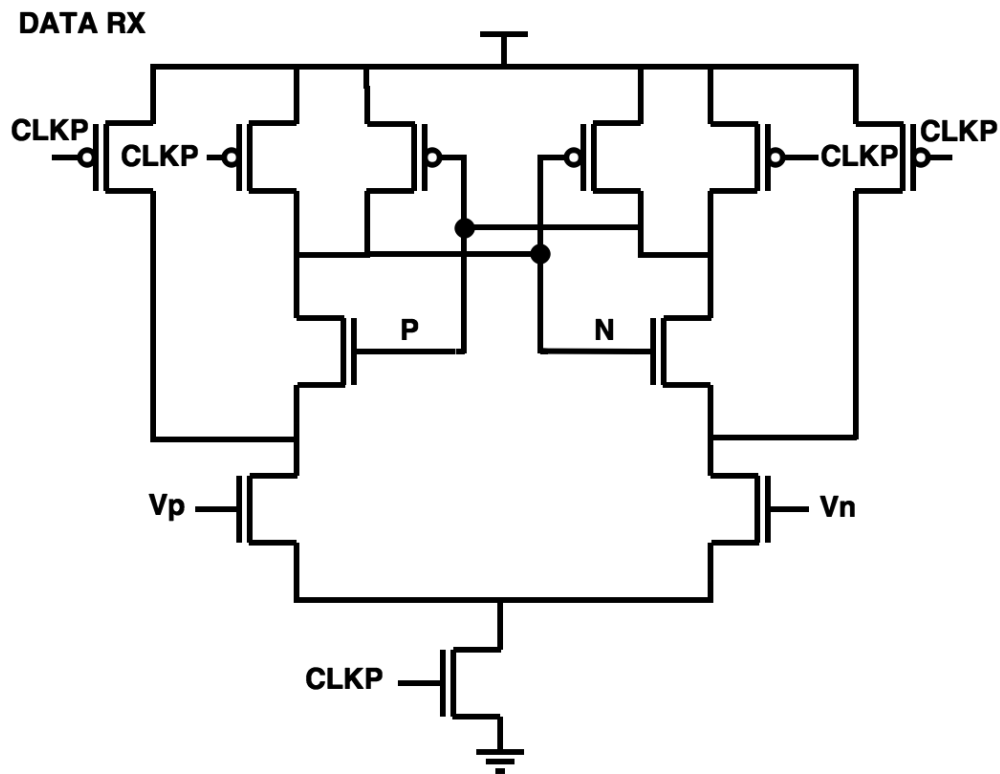


Figure 3.13: Transistor diagram of data receiver, a latched comparator designed for low voltage operation.

of voltages (from 0.24 to 1.0V) and frequencies (from 1.5kHz to 3MHz respectively). Thus, the proposed transceiver not only can be used in many ULP applications but also in systems

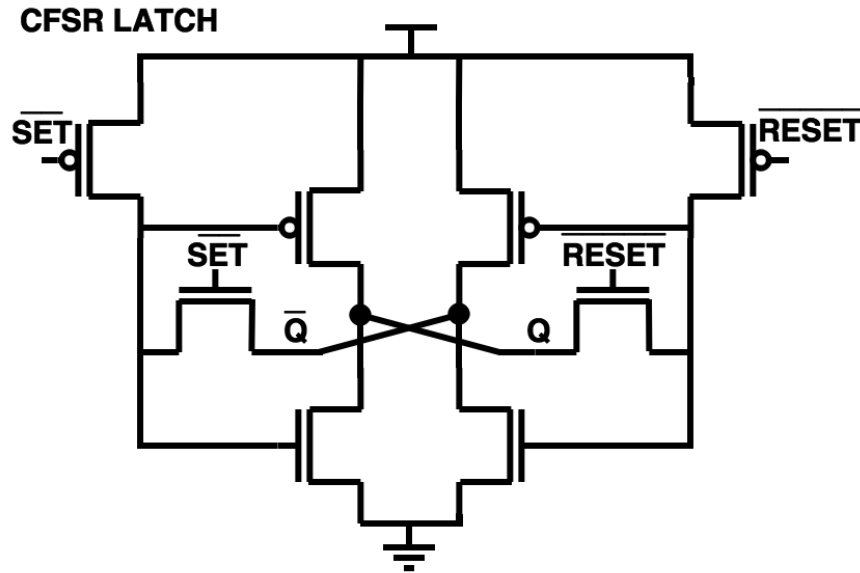


Figure 3.14: Transistor diagram of contention free SR latch. Contention free SR latch receives output data from the data receiver, latching it to provide a steady output for each half clock cycle.

with loosely regulated supplies (e.g. operating solely on harvested energy). The received clock signal travels through an edge detector (Figure 3.12) to generate a pulse (CLKP) to enable dual-edge data sampling. The edge detector contains a delay chain feeding the clock into an XOR gate with a current starving, long channel PMOS header, which widens the generated pulse. The PMOS header gate connects to ground in a forward biased diode configuration and is current limited by the channel length. This method creates an appropriate pulse for the data samplers across a wide range of voltages.

The data receiver (Figure 3.13) is a clocked comparator for minimal energy consumption with additional transistors for low voltage operation. The N and P outputs connected to a contention-free SR latch (CFSR) (Figure 3.14). The CFSR latch design uses a cross-coupled inverter with one pull up transistor each for set and reset, and removes contention by blocking the appropriate feedback while switching to further reduce receiver power.

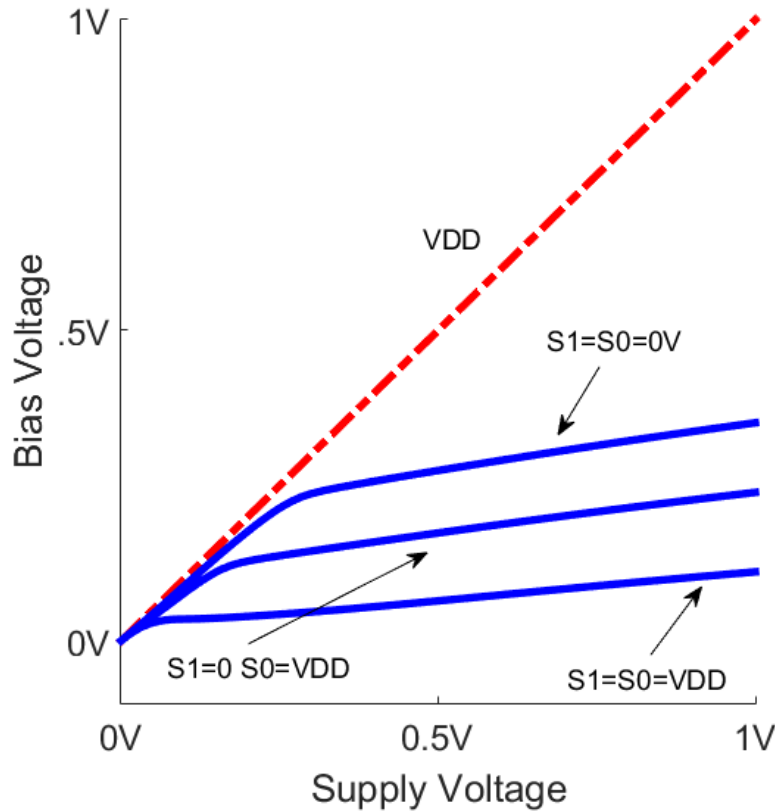


Figure 3.15: DC simulation of bias generator, showing VDD (dashed) and each output configuration (solid).

3.5.2 Results

We demonstrate this scheme using a 130nm CMOS test chip (Figure 3.17). We create a 3-bit-wide bus by copying the data transceiver to amortize the cost of clock TX. The channel for wireline testing is made up of four unshielded twisted pairs (one for clocking, three for data) tested at a distance of 3 meters. Figure 3.18 presents the Bit Error Rate (BER) bathtub curve for wireline transmission, showing an error rate below 10^{-9} over more than 40% of the phase. Figure 3.19 shows a typical eye diagram for an Ethernet channel, having a 120 mV peak to peak signal at a supply voltage of 240 mV.

We also measured the test chip with flexible interconnects in a shirt system (Figure 3.21) using thermoplastic polyurethane (TPU) films and screen-printed conductive inks to demonstrate use in BASNs. The conductive textile links are electromechanically stable

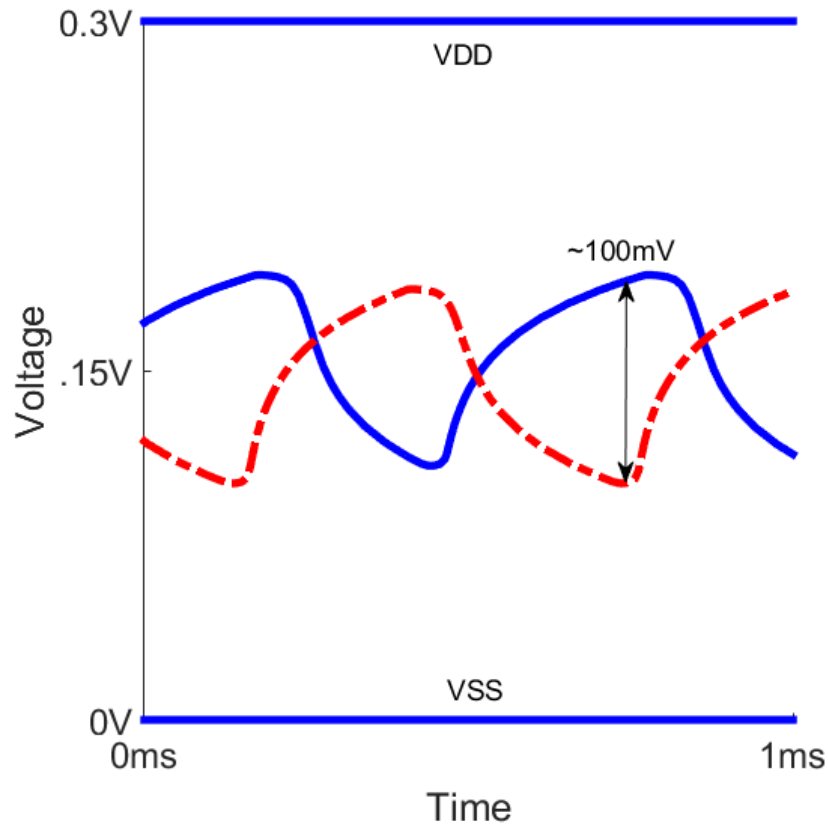


Figure 3.16: Transient simulation of transmitter shows reduced swing centered at $V_{DD}/2$.

(self-healing with 0.7Ω resistance change over 1000 cycles at 10% strain), flexible, durable, washable up to 100 washes, low-profile, and low-cost [31]. The meandering interconnect lines are screen-printed onto the film, oven cured at 50 degC, laser cut to allow for flexibility and stretchability, encapsulated with another TPU layer of film, and heat pressed onto the textile. Button snaps connect the chip's test PCB to the flexible interconnect and allow for removal of the PCB from the clothing for washing. Figure 3.20 shows a typical eye diagram for the textile interconnect channel, having a 160 mV peak to peak for an operating voltage of 300 mV.

Figure 3.22 shows E/b/mm and throughput for state-of-the-art signaling designs, as well as our design over its VDD operating range. Even though the high throughput designs may have low E/b/mm at high power and throughput, their energy does not scale when duty cycling

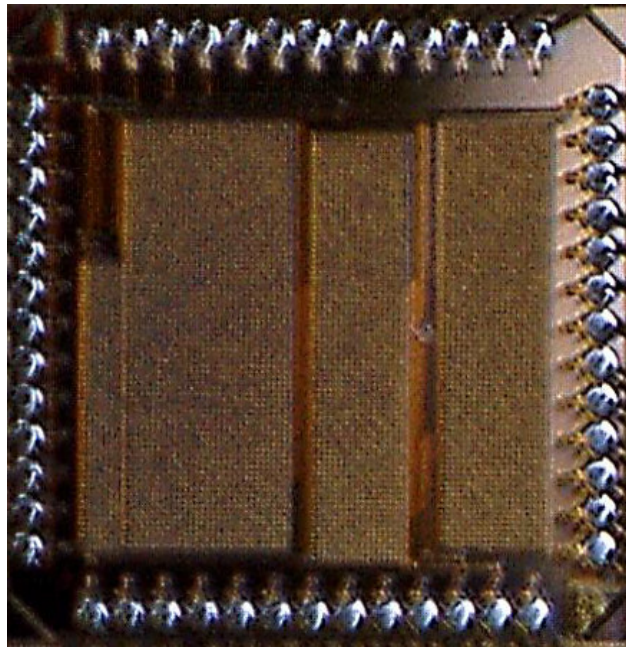


Figure 3.17: Die photo of differential link test chip.

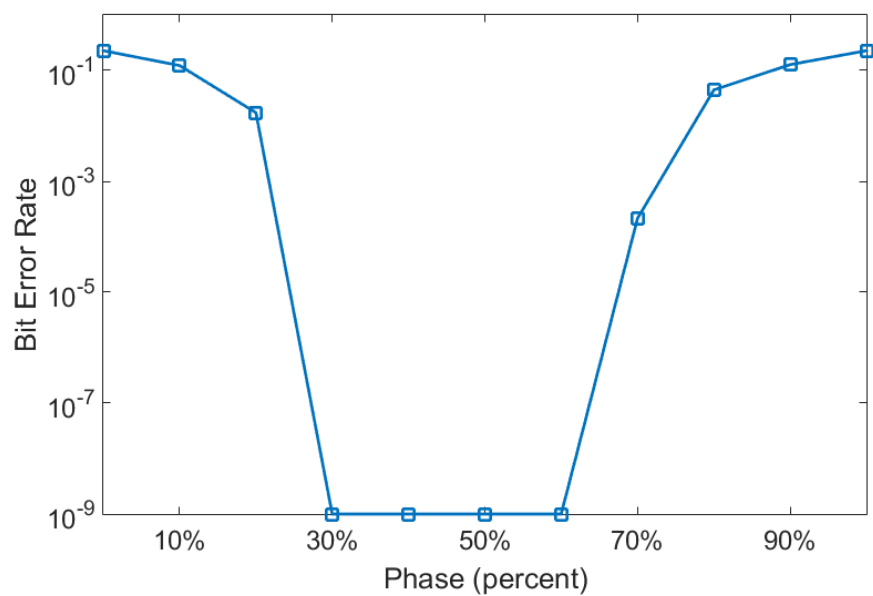


Figure 3.18: Measured bit error rate plot of transceiver.

due to their intrinsic high-leakage nature, making them unsuitable for ULP networks. Table 3.2 shows a comparison with state-of-the-art designs. Table 3.3 shows a power breakdown of the test chip at its minimum energy per bit and max frequency through Ethernet and flexible textile interconnects. Other designs are used to compare metrics, but none solve the problem

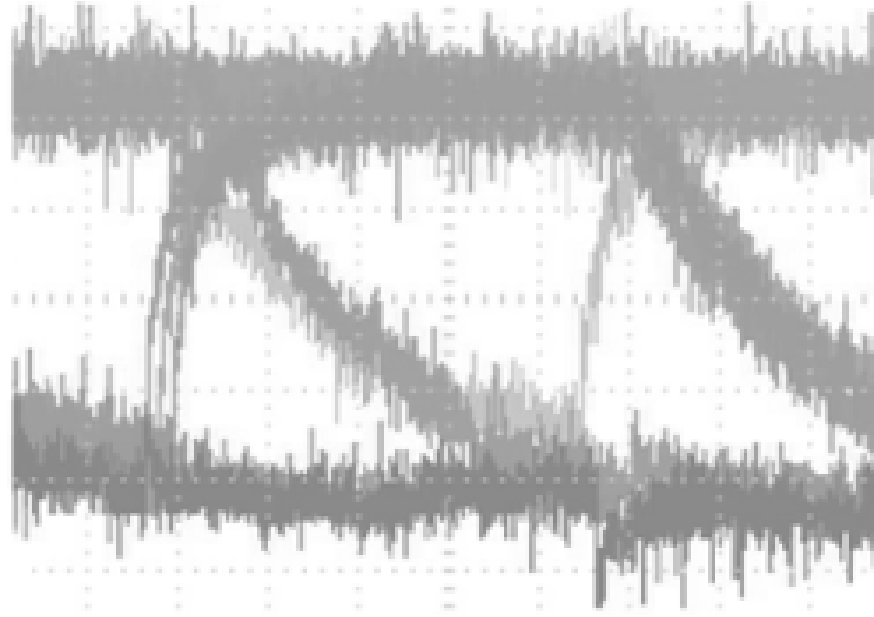


Figure 3.19: Measured eye diagram of transceiver over an Ethernet channel. Transceiver's supply operating voltage was 240 mV, while the channel showed a reduced 120 mV peak to peak, with an eye height of 55 mV. Bit error rate $< 10^{-9}$ (no shown errors)

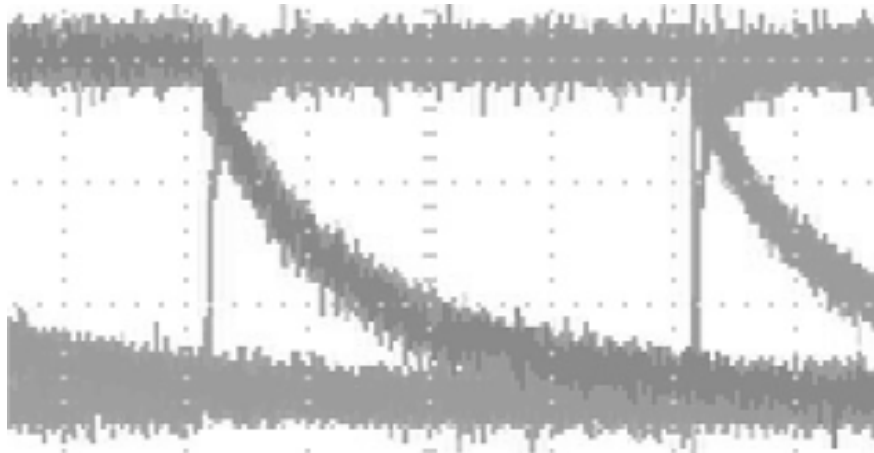


Figure 3.20: Measured eye diagram of transceiver over the flexible textile interconnect. Transceiver's supply operating voltage was 300 mV, while the channel showed a reduced 160 mV peak to peak, with an eye height of 110 mV. Bit error rate $< 10^{-6}$ (no shown errors)

of sub 10 μW ULP on-body networks. Our design shows over 3X lower E/b/mm than the single ended design presented in the previous section and 5 orders of magnitude lower power than [32], enabling efficient wireline connectivity for BASN applications.

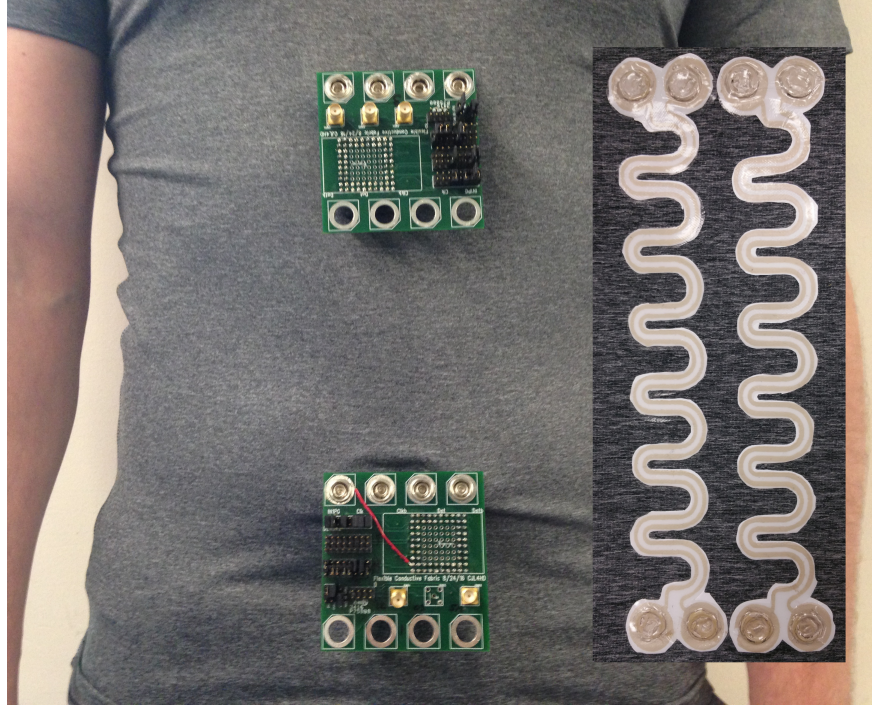


Figure 3.21: Photograph of transceiver snapped onto a shirt. Flexible textile interconnect shown on the right side of the image electrically connects the button snaps through the inside of the shirt.

Table 3.2: On-body link comparison to existing wireline work

	3m RJ45 at MEP	3m RJ45 at max f	Textile at MEP	JSSC '16	ISCAS '15	ISSCC '15	ISSCC '16
Process (nm)	130			65	130	65	28
VDD (V)	0.240	1.0	0.3	1.2	0.2	0.45-0.7	0.9/1.5
Avg Pwr (uW)	0.00542	132.3	0.00377	6300	0.00124	670	403
E/b (pJ)	3.61	44.1	3.77	79	0.38	0.29	14.4
E/b/mm (fJ)	1.2	14.5	11.4	39.5	3.8	0.305	2.88
Bitrate	1.5 kb/s	3 Mb/s	1 kb/s	80 Mb/s	3.25 kb/s	1.3 Gb/s	28 Gb/s

Table 3.3: Measured on-body link power breakdown

Power Breakdown	3m RJ45	Textile	Sleep (leakage)
Voltage	240 mV (1.5 kbps)	300 mV (1 kbps)	240 mV
TX Clock	1.56 nW	2.40 nW	0.094 nW
TX Data (per lane)	1.23 nW	1.32 nW	0.118 nW
RX Clock	0.016 nW	0.027 nW	0.005 nW
RX Data (per lane)	0.013 nW	0.018 nW	0.011 nW
Bias generator	0.105 nW	0.000 nW	0.100 nW
Total	5.42 nW	3.77 nW	0.585 nW

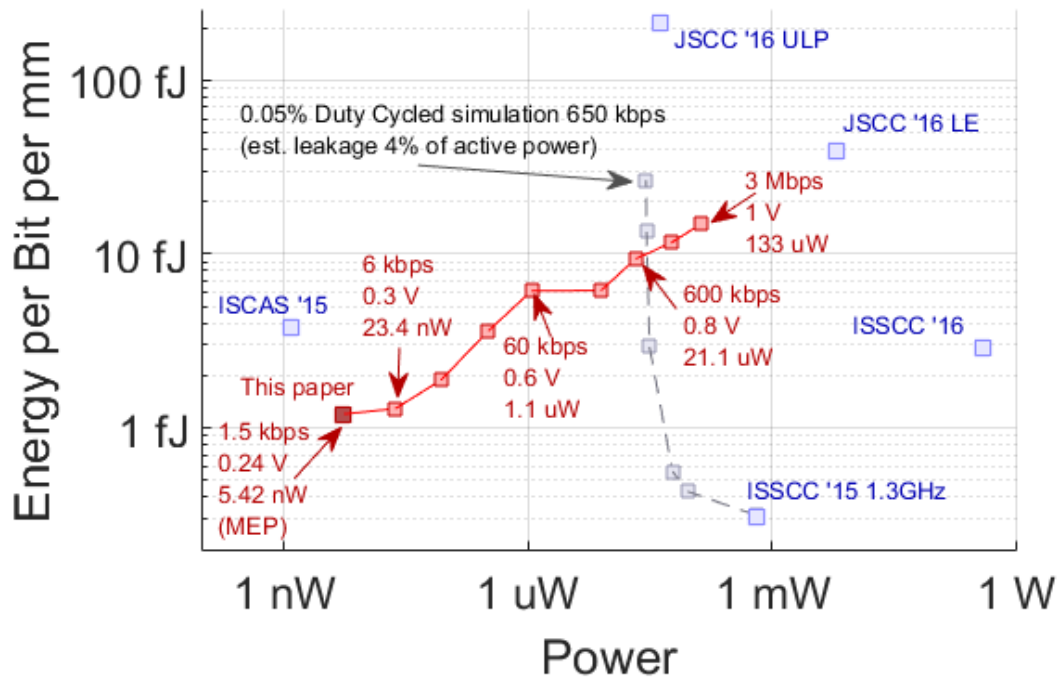


Figure 3.22: Transceiver using 3 meter Ethernet cable throughout operating voltages compared to state-of-the-art-designs. Minimum Energy Point (MEP) shows improved co-optimization of energy and power.

3.5.3 Conclusion

This chapter presented two transceiver designs to enable communication between ULP devices in IoT design space. The first transceiver design was a single-ended transceiver for short distance communication over a PCB. We also described a method for finding the minimum energy-per-cycle for that transceiver by varying activity factor, and then using this as an energy budgeting technique for designing a communication circuit with energy constraints. We then went on to relate this with the energy-per-bit metric and create a trade-off between the two in which an optimal point is found, resulting in an operating voltage and activity factor. A fabricated circuit was then used to show the benefits of using this optimization, having an energy-per-bit of 0.38 pJ/bit and power of 1.24 nW. This methodology will allow designers to have a more accurate energy model that can be used for designing communications circuits more suited for ULP systems.

The second transceiver design was a differential transceiver for long distance, noise immune communication. We optimized these circuits for low leakage to create a minimum energy point at low voltage, resulting in both low energy and power. Low leakage was attained through replacing designs which would normally consume static current with low leakage alternatives. The transceiver also was designed for operation over a wide range of voltages by removing the need for bias generation in all but the clock receiver. In this case a bias generator was used to allow a wide range of operation. We then demonstrated the transceiver through Ethernet and in a flexible textile based interconnect for use in on-body networks. This design achieved a power consumption of 3.77 nW and energy of 11.4 fJ/b/mm.

Chapter 4

Ultra-low Power On-package Communication

4.1 Introduction

In the previous chapter, we re-designed off-chip communication circuits to better match the requirements of ULP systems and enable networks of these systems to communicate data efficiently. However, off-chip communication is not the only type of communication that needs to occur within ULP systems. At the heart of these systems is an SoC that consists of many building blocks that gather, process, and transmit data. The SoC needs to communicate with off-chip memories and peripherals, and those can be integrated in the same package to reduce form factor, energy, and power. In this chapter, we propose an on-package transceiver that enables ULP systems to leverage multiple technologies without compromising form factor.

On-package signaling uses pads and bond wires (shown in Figure 4.1) to allow communication between two or more silicon dice placed in the same package. On-package communication can be implemented in multiple ways. One way to handle this type of communication is to treat it similar to off-package communication (shown in Figure 4.3) with data being transferred serially. Another option would be to leverage the fact that on-die pads do not

contribute to the limited pin count if they are connected to another die within the same package. Pads are metal area on the silicon used to bond wires to the die, whereas pins are the electrical components on the package which allow communication outside the package. Off-chip communication uses bond wires between a pad and a pin, whereas on-package communication uses wires bonded between pads. This results in a much higher wire limit for use in communication between dice on a package. An example application of on-package communication is shown in Figure 4.2. This chapter will cover contributions to on-package buses for use in ULP systems.

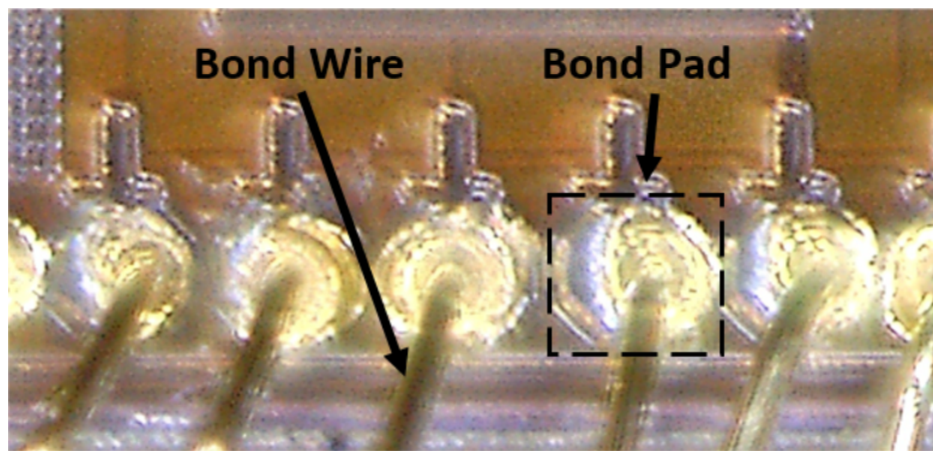


Figure 4.1: Micrograph of a die showing bond pads around the proximity of the die. Wires extending off the bottom of the image have been bonded to the pads.

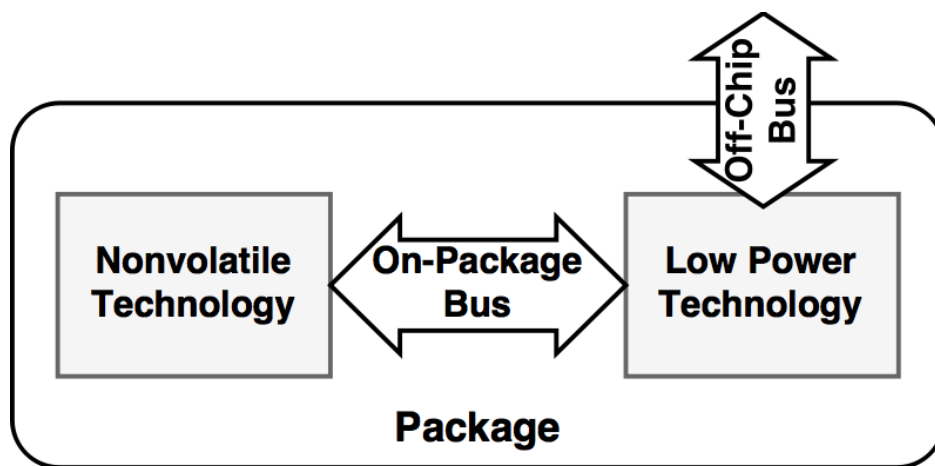


Figure 4.2: Example diagram of potential application of on-package bus. On-package buses enable integration of multiple dice in the same package.

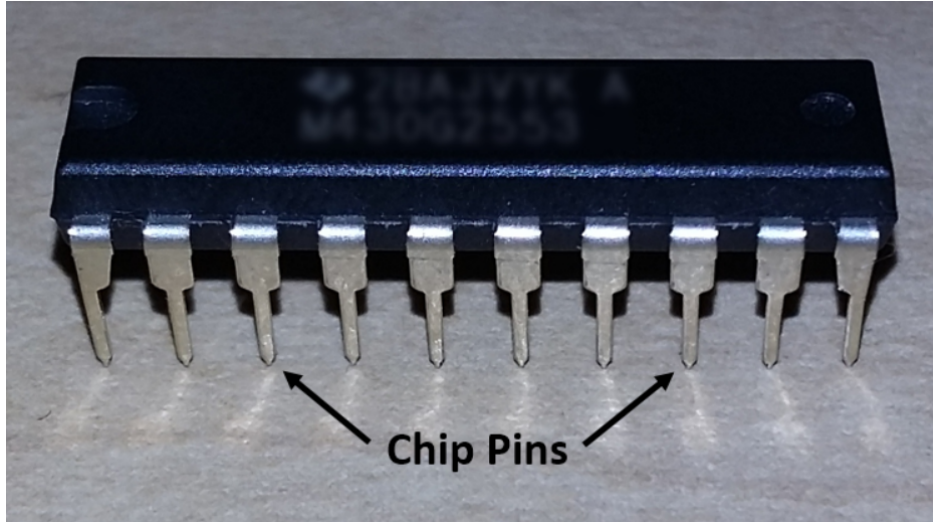


Figure 4.3: Image of a dual in-line package showing pins used to interface with off-chip components.

4.2 Motivation

On-package communication has historically been motivated by the need for higher performance, lower latency communication between an SoC (or processor) and high capacity off-chip memory (DRAM) [33]. This is because moving the peripheral chip's logic closer to the processor reduces attenuation in transmission line, allowing for higher speed circuits on the interface. More recently, on-package communication is being researched for integrating large multi-core systems to improve yield [8].

ULP systems can benefit from integrating chips into a single package and using on-package communication, however the motivation comes from different factors. First, integrating multiple chips on a PCB limits their usefulness in applications where form factor is key due to unnecessary packaging and off-chip interconnect; whereas, SiPs improve form factor. Second, throughput on PCBs is restrictive in ULP systems due to the data transmission frequency decreasing linearly with an increase in PCB trace length.¹ This effect is especially pronounced when operating at low, subthreshold voltages due to the capacitance remaining the same while current decreases exponentially with decreased supply voltage. Subthreshold

¹Because $Delay = \frac{C \times V}{I}$ for low throughput communication, a linear increase in PCB length results in a linear increase in delay.

operating voltages are becoming increasingly common with a desire to decrease SoC power consumption. Integrating SiPs improve the maximum throughput by reducing the distance between dice, thereby reducing capacitance. Third, energy and power are a costly part of the system budget, and the reduced capacitance of SiP integration improves both. Lastly, SiPs enable the leveraging of multiple technologies into a single, compact package. Each technology has its own benefits so this will enable self-powered SiPs to increase flexibility, perform more advanced tasks, and be suitable for more applications.

However, existing low power and low energy on-package communication is designed for higher throughput than needed for self-powered SoCs (GHz and above). As a result, the power consumption of these transceivers will be orders of magnitude higher than the power budget of an entire ULP system. Even though these circuits can be duty cycled to reduce power, they will still have low energy efficiency. In addition, higher throughput circuits will often consume relatively high static power due to overhead in designing the circuits to operate at such a high throughput. Self-powered SoCs are an emerging technology, so no existing work is low enough power for continuous self-powered on-package communication.

Unlike existing work for general use, the on-package bus we present in this chapter is designed to improve power, specifically for cold-booting an SoC. Existing self-powered SoCs rely on volatile SRAM memory to hold program instructions due to their reduced power requirement. SRAMs are usually programmed through a dedicated programming port on the bench. When the system loses power the instructions are lost, meaning the system must be rebooted through the programmer. Because these systems are designed to operate in conditions with intermittent harvesting, they are prone to power loss and should be capable of restoring themselves into an acceptable and known operating condition. Designing a specialized on-package bus to boot the SoC from program memory allows for application specific optimization.

4.3 Prior Art

Existing work on bus integration is too high power, too high energy, and too generic to allow proper control over the NVM chip from the SoC during the boot phase. Utilizing existing work risks the system entering a state where it cannot harvest enough energy to complete the boot cycle before losing power, resulting in a continuous loop of failure to load instructions.

Most on-package communication research is done on high performance systems. [33] uses Through Silicon Vias (TSVs) for high-bandwidth DRAM access. [8] seeks to improve the density of interconnect and remove limits on die size using a silicon bridge across on an organic package. Both designs seek to address problems different than those arising in ULP systems, and as a result would need to be adapted for the design space.

Low power on-package work exists for IoT devices [34] which defines a flexible architecture but consumes high active power. It is designed to be modular, allowing communication between many dice and allowing multiple bus masters. Due to the generic nature of this solution, it is also energy inefficient and as a result is unsuitable for use in self-powered battery-less systems.

4.4 Asymmetric Bus Enabling Cold-Boot Self-Powered Systems in Package

At the circuit level, on-package buses consist of a single clock and a varying number of data lines. For ultra-low power systems seeking low energy, low bandwidth operation, this can be a simple CMOS inverter based driver and Schmitt trigger. These designs have low leakage, which reduces the power while the bus is not being used for communication. Sharing the clock between multiple data lines amortizes the energy consumed by the clock, but must be traded off with the number of pads used by additional lines. Drivers can be made smaller than those used in off-chip communication while retaining the same bandwidth.

The on-package bus in this dissertation has been designed to interface our self-powered SoC with a non-volatile memory (NVM) in order to enable a battery-less system that can boot from the NVM to recover from a power loss event. The first hurdle in creating a cold-booting, self-powered system is in the technology. Process technologies do not contain all of the necessary components, namely low power transistors and NVM, in a single fabrication process. A ULP bus for communication between the self-powered SoC and non-volatile instruction memory solves this problem. Placing all dice in a single package and designing for on-package communication further improves the system.

Figure 4.4 shows a block diagram of the interface between the SoC and NVM. The SoC acts as the master, requesting and transmitting information as necessary. The SoC also contains a control signal, powering down the NVM chip after the instructions and data are loaded.

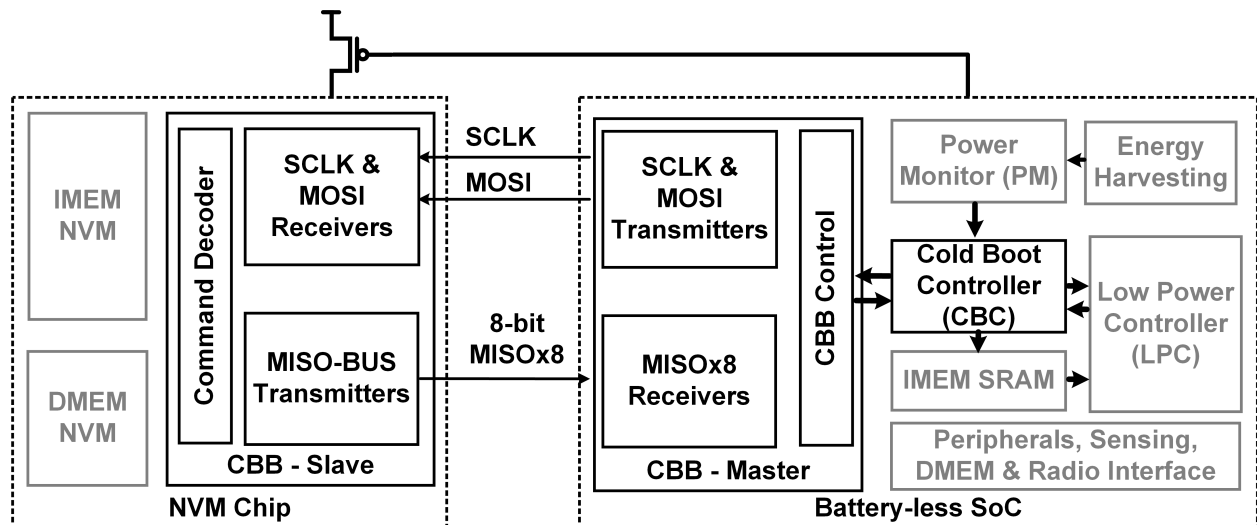


Figure 4.4: Block diagram of the on-package cold-boot bus.

4.4.1 Design

The proposed bus interface is co-designed with both the SoC and the NVM chip to maximize its efficiency and reduce the overall system power. On the SoC side, a power monitor (PM) and a cold-boot controller (CBC) are implemented to facilitate back-up and recovery. The

PM keeps track of the available energy, and notifies the main controller on the SoC when the energy drops below a critical level. Then, the main controller collects all the critical data into the CBC which then transmits it to the NVM chip through the bus. Upon a power-on reset, the PM tracks the available energy and keeps the SoC in standby mode with the NVM chip in off mode until the available energy exceeds a boot-up threshold. Once enough energy is available the PM instructs the CBC to recover the programming and critical data from the NVM chip. After a bootup command is issued, the CBC keeps track of the incoming data and programs the SRAM memory.

Existing NVM consumes a large amount of power in proportion to a self-powered SoC power budget, and as a result, cannot be dropped in to a system as is. The circuits proposed in this section reduce the time an NVM must remain on with an increased throughput parallel bus and low power design techniques. To reduce the NVM read power, one row is read and saved into a low power register file for transmission. The size of the row is chosen carefully to minimize the overall power of the system by balancing the peak power of a row read operation with the average power across the transmission cycles. In addition, after all the data is transferred to the SoC, the PM enables a power gate to the NVM chip, saving precious power. This reduction further optimizes power at the System in Package (SiP) level as it will allow retention without requiring an on-chip non-volatile instruction memory.

Operating at a higher throughput despite a low system clock rate is desired to minimize time spent with the NVM powered on, however utilizing a PLL to increase data rate will increase area overhead, system complexity, and reduce the overall energy efficiency of the interface. The parallel nature of the bus interface removes the need for clock generation and amortizes the cost of the clock. Integrating into an SiP also has the added benefit of not consuming package pins, giving the freedom to further widen the bus width.

The transmitters contain a binary weighted CMOS driver to provide control over drive strength. This allows the bus to adapt to the desired SoC clock frequency. The transmitters are tuned in order to trip the sense amp based receivers in just under the clock period. This

creates a reduced swing transmission for additional power savings without the overhead of an additional supply. Figure 4.5 shows a low level diagram of the transceiver.

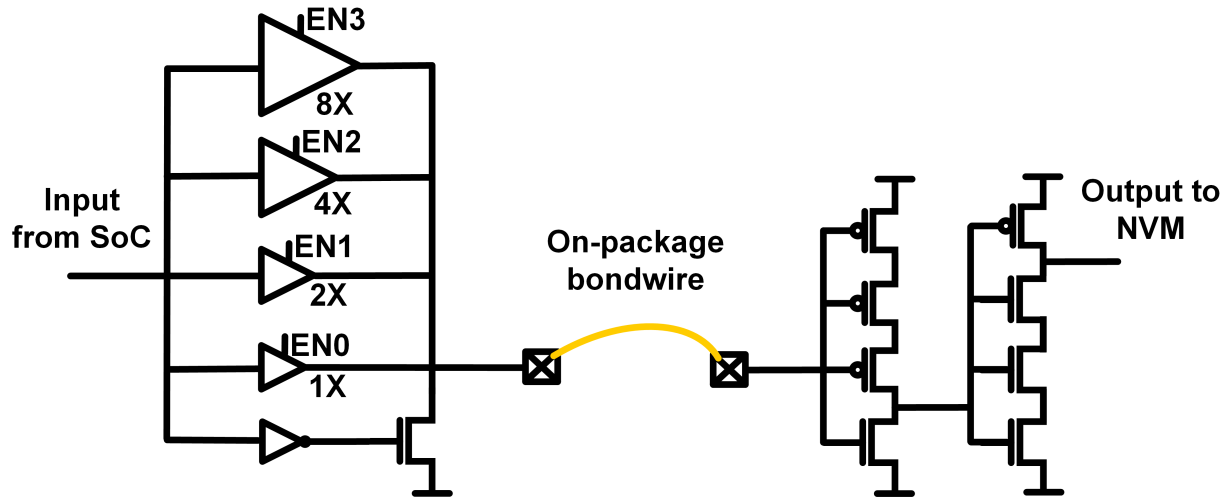


Figure 4.5: Transistor level diagram of the transceiver. Binary weighted drivers allow for reduced swing on the transmission line and sense amplifier receivers handle the reduced signal.

Boot Process

The on-package bus also contains control circuits to handle booting the system when it begins harvesting energy. The Cold Boot Circuit signals the bus controller when its time to boot, and the controller requests instructions through the bus. During the boot process, the SoC transmits a clock signal through SCLK (Slave Clock), while the boot request is transmitted serially through MOSI (Master Out Slave In). The SoC continues to transmit a clock until an EOF (End Of File) is received. During this phase, the NVM transmits the program data, followed by an end of program code. The NVM then transmits a byte describing how many critical bytes of data have been backed up and followed by those bytes. The program data is transferred to the SoC's program memory, while the critical data is held in the CBC.

Table 4.1: Power breakdown of On-package Cold-Boot Bus

Power Breakdown	Active	Leakage
Transmitter	8.33 nW	0.012 nW
Receiver	0.158 nW	0.141 nW
Total	8.49 nW	0.153 nW

Backup Process

During the backup process, the SoC transmits clock and data similarly to the boot process, followed by a backup request. The SoC then transmits the backup bits serially and the NVM stores them in the data memory.

4.4.2 Results

To demonstrate the design, we fabricate in the transceiver in two different 130nm technologies. The master contains bus control logic, two transmitters, and eight receivers. The slave contains the memory interface, two receivers, and eight transmitters. The NVM receivers consist of two full sets of receiver logic. One set is a Schmitt trigger for full swing operation, and the other is a sense amp design (Figure 4.5). For functionality testing, each chip has its own package mounted on a test PCB, with SCLK, MOSI, and MISOx8 jumpers between boards. Once functionality is verified, the chips are integrated into a single QFN package for reduced energy and power. Table 4.2 shows comparison to existing state of the art. Energy is less than high performance computing (HPC) designs, while power is less than ULP designs. A power breakdown is shown in table 4.1.

The NVM used with the SoC consumes an average of approximately 4 μ W. Because of the NVM architecture, the NVM read energy does not change if SPI is used instead of the ULP bus. As a result, the bus reduces power by reducing the time the NVM chip is powered on by 8 \times . The digital and leakage power over this time period will quantify energy savings for the system.

Table 4.2: On-package bus comparison to state of the art off-chip buses (ULP on left, HPC on right)

	This Work	CICC 2014 [34]	ISSCC 2013 [35]
Process	130 nm	180 nm	32 nm
Voltage	0.35V	-	0.6-1.08V
Active Power	8.49 nW	20.1 uW	101 mW
Sleep Power	0.153 nW	8 nW	-
Bitrate	256 kbps	400 kbps	256 Gb/s
E/b	0.0332 pJ	50.2 pJ	0.79 pJ
E/b/mm	1.09 fJ/b/mm	-	1.58 fJ/b/mm

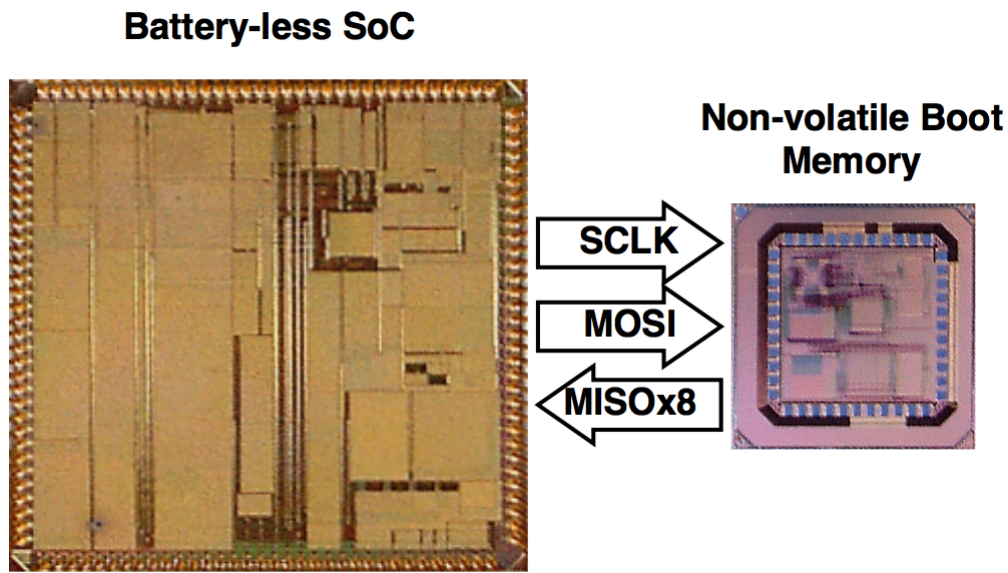


Figure 4.6: Die photos of battery-less SoC in 130nm low power technology and Non-volatile boot memory in 130nm FeRAM technology (dice not to scale).

4.4.3 Conclusion

This chapter presents an ULP cold-boot bus for increasing system integration in self-powered systems. We demonstrate this through the integration of a ULP NVM and a self-powered SoC. During the boot process (on-board integration) the transceiver consumed 8.49 nW, with an energy of 1.09 fJ/b/mm. This work shows $1512\times$ energy improvement and $2367\times$ power improvement over existing ULP designs, and $23.8\times$ energy improvement over existing HPC designs, with 7 orders of magnitude power improvement. With this improved power and energy efficiency, the interface will improve reliability and quality of service of self-powered

systems by enabling them to boot from NVM while remaining within a sub- μ W system power budget.

Chapter 5

Ultra-low Power On-chip Communication

5.1 Introduction

In the previous chapter, we discussed on-package communication and its use in ULP systems. We used the on-package bus to improve sophistication through increased levels of integration as well as to minimize the power and energy of the system through low power design techniques. In this chapter, we propose an on-chip bus architecture that allows each component on the SoC to operate at its minimum energy or power point, furthering the power and energy optimization of the system.

On-chip signaling uses buses and wire routing to allow communication between multiple logic blocks on the same chip. Unlike off-chip buses, on-chip bus lengths are on the order of microns. The shorter lengths of on-chip buses result in lower energy-per-bit and lower power consumption. On-chip signaling also has the benefit of allowing additional supportive logic along the length of the wire such as repeaters or routing logic to improve immunity against noise sources. On-chip communication often uses a wide parallel bus to allow for high bandwidth communication between logic blocks on a chip as shown in Figure 5.1. This is

unique to on-chip communication because additional wires in parallel have very little cost in area. Lastly, because on-chip buses are integrated with the digital logic blocks surrounding them, they can be co-optimized to improve energy and power at the system level.

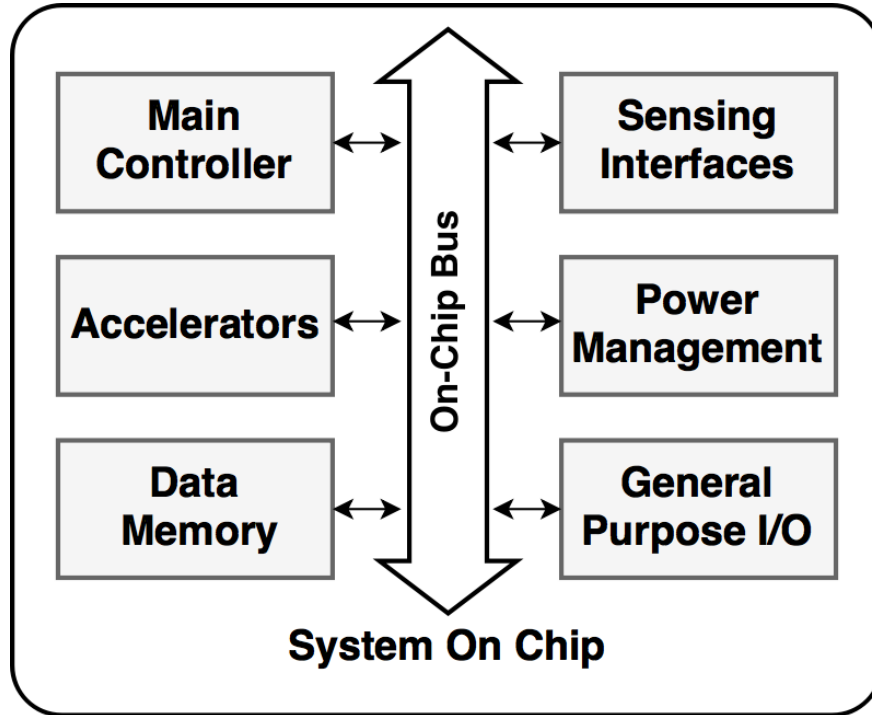


Figure 5.1: Example diagram of an on-chip SoC bus. The main controller uses the on-chip bus to communicate send and receive data to other logic blocks on the same chip.

5.2 Motivation

As mentioned earlier, SoCs are being designed with continually lower power in mind in order to increase battery life or remove the battery altogether. Power budgets are very strict, and adding new subsystems while remaining within the budget is a challenge. Existing ULP SoCs are optimized at the system level by choosing a V_{DD} optimal for the average of all digital blocks. This results in a suboptimal design. Using finer grained voltage domains for each block allows further reduction in system power and energy by operating each block at its minimum functional voltage. However, generating a voltage for each block could have significant overhead. One low-overhead method for fine grained voltage control is to use a

resistive power grid [36]. Previous work explored the power savings of tuning a resistive power grid in a multi-core processor, but did not explore lower-accelerator-level optimization for use in ULP systems. Resistive power grids rely on a set of power gates with different strengths that could be turned on or off to generate intermediate voltages or to completely power off a block when it is not needed. To allow safe communication between blocks running at different supplies, isolation and level shifting are required. However the use of standard level shifters can consume a large power relative to the budget of the SoC. As a result, a new low cost method must be developed to allow communication between multiple voltage domains.

The next problem that arises in ULP SoC optimization is the question of how to divide the system into voltage domains. The trade-off is between the granularity of voltage control and the overhead of additional level shifting for each voltage domain. Having a voltage domain for each accelerator and logic block is ideal, but will not improve the system if the overhead of additional level shifters on every interface outweighs the benefits. The solution is a function of the number of logic blocks used in the system and the necessary performance, as well as the area and power overhead of using additional level shifting. In this chapter, we will focus on using bus logic to allow power and energy optimization throughout a ULP SoC.

5.3 Prior Art

Standard bus logic between blocks and the controller use a CMOS inverter based driver with inverters along the length of the bus to minimize RC delay. These repeaters improve performance at the cost of energy. Much research exists for on-chip bus circuits and architecture in an effort to improve upon CMOS inverter based communication. Relevant work on reducing bus power includes bus invert coding [37], which adds an invert signal to the bus for data with high hamming distances from the previous word sent, reducing the average power by up to 25%, where Hamming distance is the number of bits that need to flip from one clock cycle to the next. Other work [38] seeks to improve power and energy efficiency by stacking

the bus signaling scheme. In ULP systems, bus power does not contribute to a large portion of the system budget, but can be modified instead to allow further optimization in other portions of the system.

5.4 ULP Flexible On-Chip Bus Enabling GALS and DVFS in Self-Powered Systems

On-chip buses often consist of a large driver on the master and a CMOS inverter or buffer on the slave. Tri-state drivers are used on the slave to only drive the bus if that slave is being accessed by the master. Repeaters can be used to improve the speed of the bus by reducing propagation delay. In smaller designs, multiplexers can be used in place of tri-states, however they are not as energy efficient, requiring more wires and increasing the total capacitance of the bus. ULP SoCs need to have an on-chip bus to allow data flow between the main controller and the sensors or peripheral logic. In order to be self-powered, these SoCs must be aggressively power-gated, and thus the SoC usually has many voltage domains. Unfortunately, varying the voltage at each domain to save power makes communication between the blocks challenging. Operating the bus at the highest logic voltage will minimize delay along the bus, but will result in a considerable increase to bus power and energy. In addition this requires level shifters on each bit of the bus for each slave. Level shifters remove the possibility of static current when shifting from a lower to a higher voltage by removing potential short circuit paths from a power source to ground, or from a higher source to a lower one. Using a conventional level shifter on each bit of a bus and for each block on the bus will consume a considerable amount of the system's power budget. Operating at the lowest logic voltage will allow for minimum power, however this may cause the bus to become the bottleneck of the entire SoC. In this case the maximum clock frequency will be lowered, increasing slack and reducing the energy efficiency of the rest of the system's logic.

This chapter presents a bus circuit and structure which enables Globally Asynchronous, Locally Synchronous (GALS) logic and fine grained Dynamic Voltage and Frequency Scaling (DVFS) optimization on ULP systems. Traditional GALS is used in high performance systems to avoid the challenge of high speed global synchronization. GALS allows each logic block to operate at a different frequency or phase, allowing for finer grained optimization. In ULP systems, GALS can be used to enable fine grained DVFS. In both high performance and low power systems GALS reduces voltage droop by scattering the phase of the clock, helping with setup time at high speeds and keeping gates above the logic failure point in low voltage systems. DVFS allows flexible voltages and frequencies for targeting different operating modes at different times. With DVFS, a system can switch from low power mode to high performance mode by increasing voltage and clock frequency. Fine grained DVFS allows each block in the system to vary its voltage and frequency independent of other blocks. GALS and DVFS have been explored for high performance systems trying to reduce power, but standard methods do not scale into the ULP domain. For example, a power supply creating the different voltages for DVFS requires more overhead than the savings it provides in the ULP domain. Instead, a resistive power grid could be used to create a set of voltages for each block with little overhead. Resistive power grids are created using binary weighted power gates to give fine grained control over the virtual voltage within the desired logic block.

5.4.1 Design

Using a block level resistive power grid control allows digital a power monitor to control each block's operating voltage and frequency based on the available harvested energy. For self-powered systems, power is not always the most important figure of merit at a given time. As a result, we propose three modes of operation for the fine grained controller, each dependent on the harvesting status. A ULP mode for the case of low stored energy and the harvester is not providing much or any power for the storage node. A minimum energy point (MEP) mode for the case where the energy stored is abundant, and the system is not near

loss of power. Finally a high performance mode for when the storage node is full and energy can still be harvested. A ULP mode is necessary for improving the probability the node stays alive when energy storage is low, as the ULP mode will maximize lifetime. However, harvested energy is wasted when the SoC is operating in ULP mode. This means that when stored energy is not critically low the system should operate in a mode to minimize energy, maximizing the number of operations that can be performed with the current level of energy storage. Lastly, power management circuits often contain a shunt to short circuit the storage node to ground if the voltage becomes too high for the CMOS circuits to safely operate. This wasted power can instead be 'shunted' through the logic in a higher power, high performance mode using the fine grained controller to keep the storage at a safe voltage while using the power to do work.

Each power mode will have different settings for the resistive grid on each logic block, tuning that block for the desired metric. As a result, each block will operate at its own voltage, requiring level shifting between each block and the bus. A challenge arises in the level shifting between blocks and buses and managing the overhead, especially because using independent voltage domains with multiple modes means level shifting may happen upwards or downwards, but it is not known before the time of operation. The use of an SR latch as a level shifter allows shifting in both directions, provided that the set and reset transistors are properly sized. It also has the benefit of acting as an asynchronous memory that can replace memory mapped registers on the bus, removing the power and area overhead of traditional level shifting. If a full flip-flop is necessary, a clock enable signal can be added to SR latch input, and a D latch can be added on the output for edge sensitive data flow.

Figure 5.2 shows the SR latch. It consists of two cross coupled inverters acting as a memory storage, as well as two transistors to pull the inverters to V_{DD} or V_{SS} . In the case of the figure, the set and reset transistors must overcome the pull down transistor of the opposing inverter to change the state of the cell. If these transistors are made strong enough, The SR latch will operate at low input voltages (in our case down to 0.2V), allowing for a

shift upward to a higher voltage domain. Shifting downward works as well, as providing a voltage above V_{DD} to set or reset will change the state of the cell in the same manner. The use of SR latches in combination with resistive power grids present a solution to the many voltage domain problem.

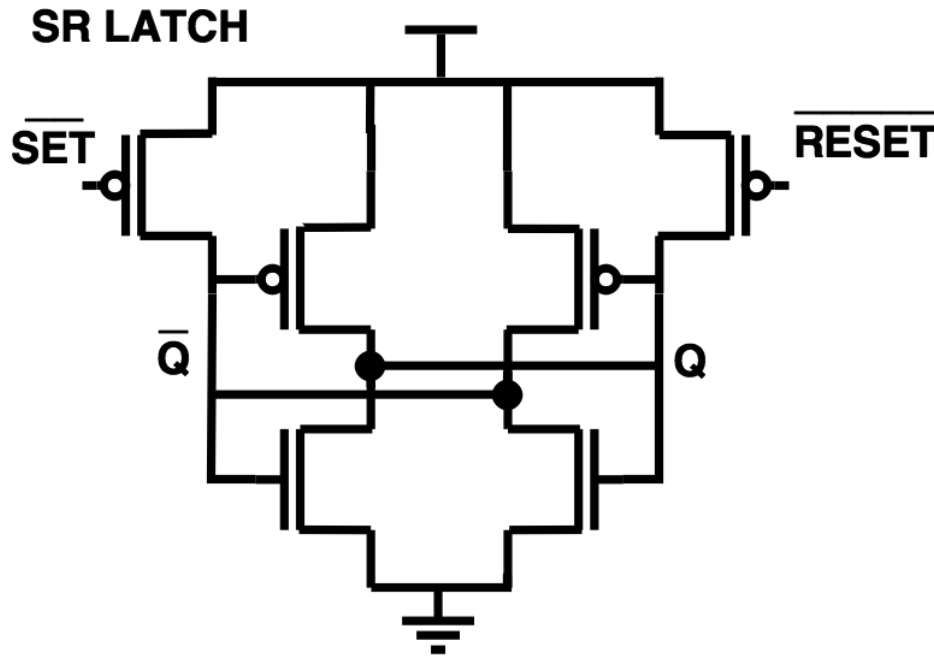


Figure 5.2: Transistor diagram of an SR latch. SR latches are used in the on-chip bus to allow block level power optimization through flexible level shifting.

Because there is no overhead in swapping out D Flip-flops for SR latches, designing for ULP means using a new voltage domain for each block and reducing voltage as much as possible. On the other hand, designing for MEP operation in a higher performance domain may include removal of level shifters in some places in order to have fewer voltage domains, with a level shifter splitting the bus into two or more operating voltages.

The SR latches were added to the slave wrapper around each block to provide data storage, level shifting, and isolation. One set was added at the outputs from the block to the bus and used for storage, level shifting and isolation. Another set was added for the inputs from the bus to the block for level shifting. The main configuration register of the block was kept as a regular D-flip-flop running on the bus power domain, since these inputs are critical and

are used to determine the operating mode of the block. In our SoC, we also included an additional set of SR latches to hold the inputs of the block in case it is power gated. This set could be optional but was added to keep the block in a deterministic state after it is powered up. Figure 5.3 shows the distribution of SR latches at the interface between a block and the bus.

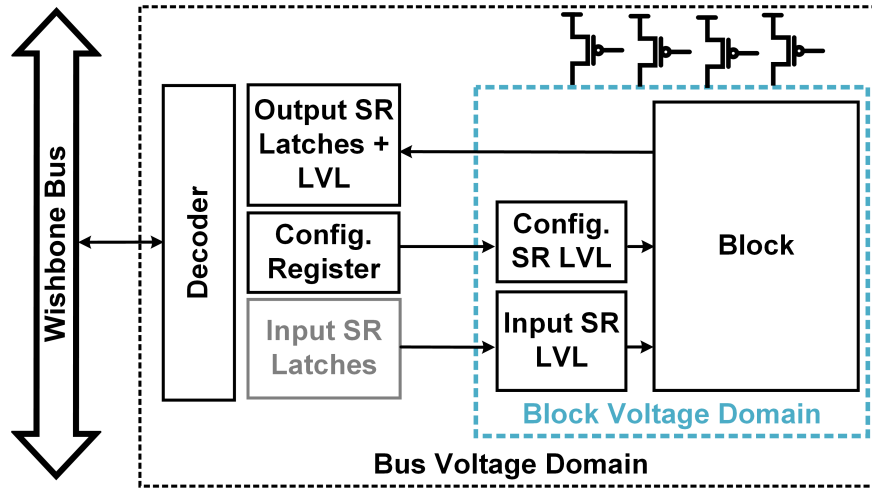


Figure 5.3: Architecture of the blocks in the SoC using SR latches for level shifting and memory. Input SR latches (in gray) are optional for storing memory mapped I/O while the block is power gated.

Each block's range of operation is a function of whether that block fails and whether the level shifter is capable of shifting between the block level and bus level. Figure 5.3 shows the wide range of level shifting voltages the SR latch is capable of, often leaving the logic block as the limitation in voltage range of operation.

5.4.2 Results

In the 130nm test chip (Figure 5.4), we present an ULP SoC with three operating modes having the combined goal of maximizing system lifetime in an uncertain harvesting environment. The SoC will enter the most appropriate mode for the given harvesting status. The ULP non-deterministic mode is the default safe state the system enters upon cold-boot. Each block operates at its minimum safe operating voltage in a GALS state. It consumes the least

Table 5.1: Measured Accelerator Active Power Breakdown and Comparison

Logic Block	Active Power			Leakge Power		
	This Work	Ref. SoC	Savings	This Work	Ref. SoC	Savings
MAC	7.03 nW	13.92 nW	49.5%	1.97 nW	3.56 nW	44.7%
Timer	12.55 nW	29.61 nW	57.6%	3.14 nW	6.49 nW	51.6%
SPI Logic	105.2 nW	166.4 nW	36.8%	4.22 nW	5.71 nW	26.1%
FIR	163.8 nW	258.65 nW	36.7%	8.40 nW	12.6 nW	33.3%
Average Savings			45.2%			38.9%

power, having the goal of increasing the total energy stored. However, in GALS systems, operation is non-deterministic, thus real-time circuits must be either redesigned for a GALS system or completely power gated. The MEP mode is entered once a threshold of energy has been harvested, and the system is in a stable state. It is a higher power consuming state, but is deterministic, allowing real-time circuits to enable, and is more energy efficient. The low power, high performance overharvesting boost mode is entered when the energy source (supercapacitor in test case) is approaching its limit in stored energy. In this case, any additional energy harvested will be otherwise wasted so it is best used to perform more computation in times of surplus.

Figure 5.5 shows the simulated operating range of the SR latches used to level shift to and from the logic blocks. The SR latch enables each block to operate at voltages below 0.3V. Table 5.1 shows a comparison between the SoC presented in this chapter and the SoC presented in the next chapter. In this SoC, each block's voltage can be controlled independently to determine its minimum operating voltage. In Table 5.1, the chip is assumed to operate from the crystal (32 kHz), and the minimum operating voltage (and thus power) of each block was determined. By using the proposed SR latch an average of 45.2% active power savings can be achieved for each block.

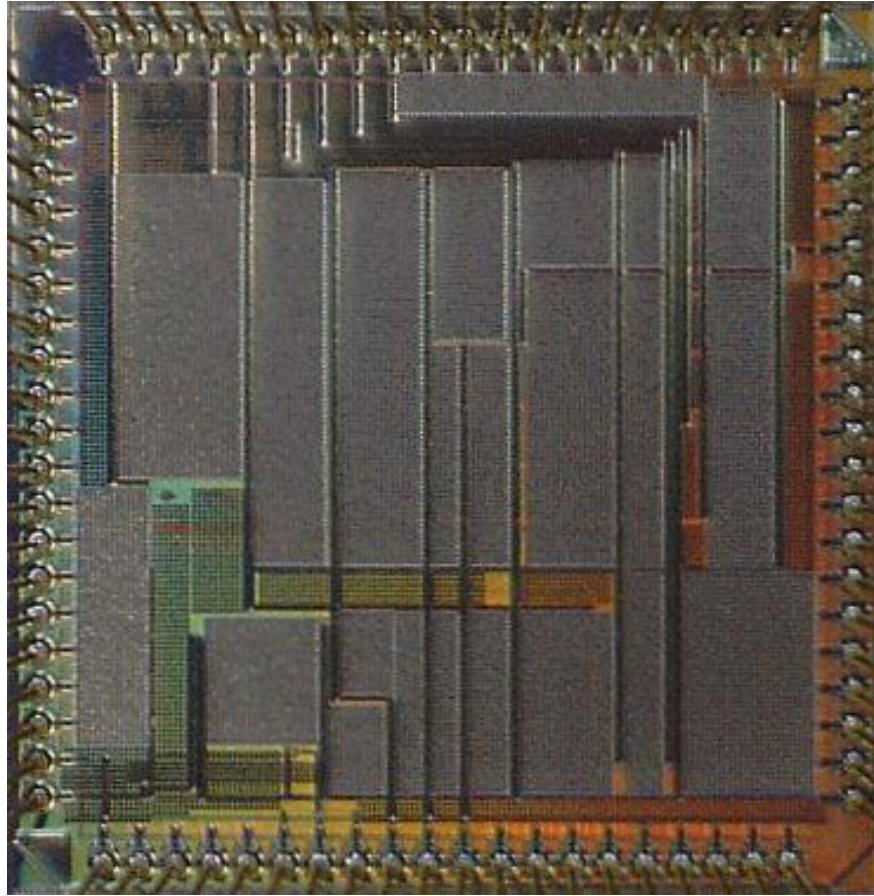


Figure 5.4: Micrograph of the fabricated test chip for optimization of on-chip accelerator power.

5.4.3 Conclusion

In this chapter we presented an ULP SoC designed for block level energy and power optimization, providing a more robust self-powered system based on varying harvesting conditions. The optimization is done through adding the ability to independently control each block's operating voltage, which is made possible through level shifters able to store memory mapped register data and shift both upwards and downwards in voltage. An SR latch was selected to solve these problems, and was synthesized in the SoC digital tool-flow. The resulting system is able to save an average of 45.2% power per accelerator over a comparable existing ULP SoC. For these measurements, the same clock frequency (32 kHz) was used in both the reference SoC and this work.

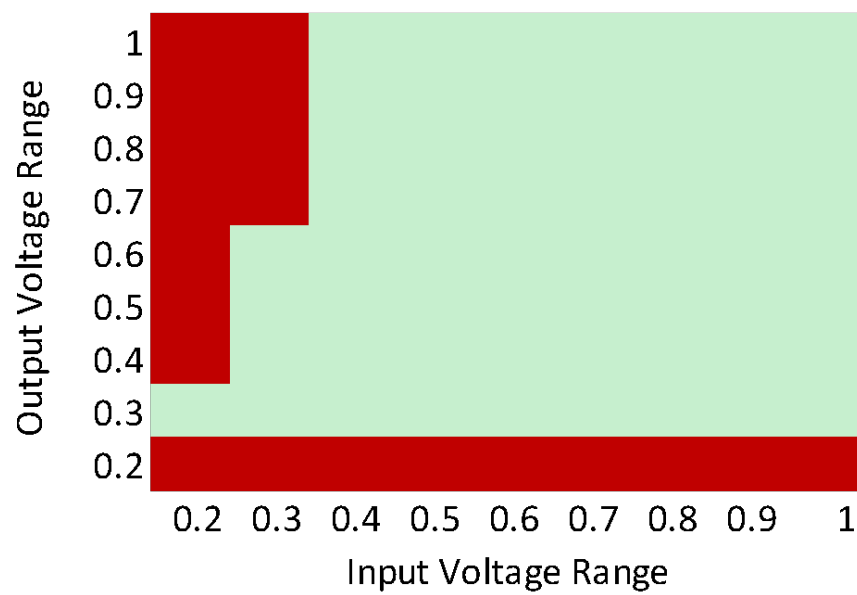


Figure 5.5: Simulated schmoop plot of SR latch used as a level shifter. Green (light) shows success and red (dark) shows failures when shifting between the given input and output voltage. SR latch shifts both upwards and downwards, creating an ideal solution for post-fabrication block level voltage optimization.

Chapter 6

Reducing Self-powered System Time to Market

6.1 Introduction

Ubiquitous self-powered systems will change our everyday lives by providing a new depth of information about the world around us. It has been predicted that the number of these IoT nodes will be above 1 trillion [25], however there are not yet commercial products within the self-powered, battery-less space. This is because new technological developments must still be made to make these products possible. To address the challenges in self-powered SoC design, we need improved robustness through higher levels of system integration, demonstrations of potential applications to show the feasibility of ubiquitous self-powered systems, and developments in manufacturing verification and test.

Improving sophistication and reliability of ULP SoCs improves the application space, allowing them to solve real-world problems and as a result motivates commercialization. Reducing power is one method of improving reliability in self-powered systems, as it decreases the probability of a power loss during times when energy is not being harvested. Reliability is also improved when a solution to handling the loss of power events is found. One solution

would be to use a non-volatile processor with registers that do not lose their value, however commercial non-volatile processors consume over 1 mW, making them unsuitable for self-powered systems. Another solution is to boot a ULP SoC from an NVM and backup critical data to the NVM as the SoC is actively operating. In previous chapters, we proposed techniques to reduce the power consumption of wired communication circuits to levels that make them feasible to use in self-powered IoT systems. We also introduced a ULP on-package bus to reduce the communication cost of transferring data between a ULP SoC and its backup NVM. In this chapter, we will use the ULP bus to interface a self-powered SoC and a ULP NVM sub-system to build a reliable SiP. The SoC is designed to consume below $1\mu\text{W}$ of power without compromising its flexibility. The NVM is optimized for ULP applications and is used for booting and storing critical data, greatly improving the usefulness and reliability of the system.

In order to be commercially feasible, self-powered SoCs must also have potential applications that will be useful for improving everyday life. One potential application is health monitoring where sensors record heart data, patient movement, breathing rates, and body temperature to enable a continuous stream of data for doctors to manage patient health and improve accuracy of diagnoses. Self-powered systems can also help with deployment of environmental sensors placed far away from continuous power supplies or in difficult to reach places. In this chapter, we expand the application space of the proposed self-powered SoC by implementing continuous positioning using an accelerometer as a potential low power replacement for GPS, a high power wireless solution.

To improve the time-to-market of these self-powered SoCs, the test and verification process must also be optimized. The verification and test process begins early in the product cycle, as the design specification must be verified against the requirements. The test process comes later in the cycle, after fabrication. The test of the products before packaging is an important yet time-intensive part of the process, especially in the case of subthreshold systems. Every die on every wafer must be tested for manufacturing defects due to transistor variation, shorts

between wires, and other physical failures [22]. Because critical paths in the design must be tested for delay, the testing process is done at the normal operating clock speed (at-speed testing) to ensure sequential and combinational logic operates properly. Self-powered SoCs typically have a much lower operating frequency due to the lower voltage operation and as a result will take much longer to complete these tests. In this chapter, we introduce the concept of trans-threshold correlation to reduce the test time of self-powered SoCs.

6.2 Motivation

Self-powered SoCs work with off-chip sensors and components to solve an application level problem. In order to be autonomous, self-powered SoCs need the ability to boot from NVM. Intermittent harvesting may result in a loss of power. Without an NVM, the system will not recover. A self-powered SoC with standard integrated power management cannot support off-chip components like a cold-boot NVM. As a result, its necessary to create a platform level power management system to handle both the SoC and other components.

Research systems are often a combination of individual research glued together to create an SoC. Improved co-optimization of SoC building blocks allows for improved data flow and power management, reducing total platform power for the same operation. This co-optimization involves planning and integrating SoC building blocks to better handle a set of applications. Figure 6.1 shows the SoC and other system components integrated into a single package, all co-optimized for data flow and minimal power consumption.

Self-powered systems are still an emerging technology and require a proven application space as well as manufacturability in order to become ubiquitous in the consumer market. Many potential applications have not been proven as viable under a self-powered system budget. Proving their ability to provide long term solutions to an array of applications will help with entry into the market as well as acceptance. One such potential application is positioning.

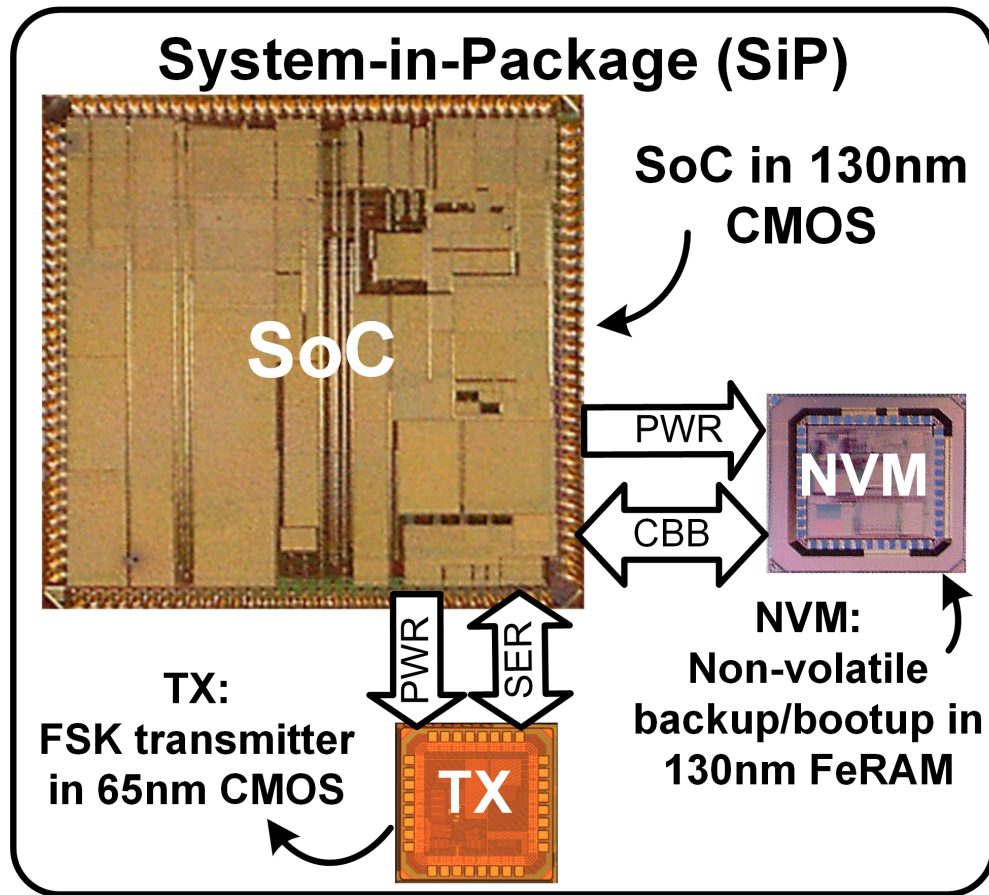


Figure 6.1: Overview of the System in Package showing integration of dice with their data interfaces and power supply. The System in Package leverages the best features of all three technologies. Dice drawn to scale.

Global positioning is a technology that tracks the location of an individual in terms of their coordinates on earth. It is used for navigation and also has recreational and emergency service applications. Typically these systems use satellite communication to determine the position of a device. This radio operation has a high cost in power (on the order of 100 mW) [39] [40] and would be infeasible for continual positioning in a self-powered system where the complete system consumes less than $1\mu\text{W}$. Creating a positioning system which does not require continuous radio communication would allow self-powered systems to gain the benefits of positioning. If such a system measured its relative position to an aggregator with a known position through the use of an accelerometer, a new absolute position can be found without the need for satellites. Such a system will provide a position in places like caves

and underground levels of buildings where satellite signals cannot be acquired. As a result, this system will be more robust than a traditional GPS alone, with a potentially unlimited lifetime due to energy harvesting, and the ability to measure position where traditional GPS cannot.

After expanding the application space of self-powered systems, the next step is to look into the manufacturing challenges of these systems. The current challenge in productization of such systems is managing yield at low voltages. When the supply voltage of the system is scaled down to the subthreshold region, the impact of threshold voltage variation increases significantly due to the exponential dependence of current on threshold voltage in the subthreshold region. This results in higher failure rates and reduced yield. Thus, more work is necessary in designing ULP systems with emphasis on high yields, and an important and unexplored part of this work is testing the designs' functionality once fabricated. Challenges with chip yield are being addressed, but issues with testing time unique to subthreshold circuits have not yet been researched.

The problem with conventional testing of subthreshold systems is that functional chip testing is becoming more time intensive due to the increasing complexity of SoCs. This, coupled with the fact that testing SoCs in mass quantities is a time intensive process as a result of increased delay during subthreshold operation, increases the test time required for self-powered systems. With maximum frequencies 10,000X lower than nominal voltage systems, the testing time will be the same multiple higher than traditional testing, resulting in prohibitively high test costs. This will be a financial burden on emerging companies trying to develop the new technology. Reducing the time required to test self-powered SoCs will reduce manufacturing costs, allowing them to enter the consumer market at a higher rate and lower cost to the end user.

6.3 Prior Art

Much work exists in which IoT SoCs are proven feasible, often through an example application. In [41], a low power SoC is designed which measures ECG data and consumes $32\ \mu\text{W}$. In [42], a low power SoC is designed which measures EEG data for seizure detection, consuming $9\ \mu\text{J}$ to derive a feature vector. [43] presents a $500\ \mu\text{W}$ neural interface that streams neural waveforms over 15m. [44] nearly achieves continuous self-powered operation while collecting temperature and capacitance measurements. [45] presents an SoC with ECG, bio-impedance, accelerometer sensing and motion artifact reduction with a multi-sensor system power consumption of $374\ \mu\text{W}$. Most of these SoCs target health monitoring and some almost achieve fully self-powered operation.

More recently, SoCs have been shown which are in the power consumption range of fully continuous self-powered operation. [46] describes a $64\ \text{nW}$ ECG SoC, proving that ECG measurements are within the possibilities of a self-powered IoT SoC. This number was later reduced to $3\ \text{nW}$ and proved fully self-powered operation through the use of three on-board photodiodes in [28]. The first self-powered wireless body sensor node (BSN) SoC [5] proved that sensing of ExG with feature detection and additional processing is possible entirely through energy harvesting and storage on a capacitor instead of a battery, showing that self-powered IoT devices are possible. The next generation of this device was shown in [25], reducing the system power to $6.45\ \mu\text{W}$ and showing motion monitoring and fall detection. This system was tested end-to-end from solar harvesting to radio transmission of the data, proving another application of self-powered systems.

Research is active in inertial based positioning, with solutions like sensor fusion to reduce error [47]. The other accepted solution is to integrate an accelerometer with multiple gyroscopes and other sensors to increase accuracy [48]. This comes at the cost of high power, increased weight, and results in an expensive sensor system. Utilizing sensor fusion is a lower cost, simpler solution, and motivates the creation of a self-powered, low cost inertial system that can be part of an aggregate solution.

Manufacturing and testing of SoCs is an area that has been thoroughly researched for years. However, because self-powered SoCs are new in the area of circuits and systems research, there has not yet been exploration in manufacturing and testing of self-powered systems due to their fundamentally different operation than existing commercial SoCs. Work like [49] and [50] study reducing operating voltage during test, but do not cover systems designed for operating at these reduced voltages. In addition, no work exists which attempts to reduce test time of emerging IoT devices.

6.4 Sub μ W System on Chip

Recent self-powered SoCs have shown sustained battery-less operation, having power management, harvesting, and energy storage. The challenge in designing next generation systems is continued operation in times of intermittent harvesting. To this end, the research goal is to continually improve energy efficiency and reduce power. Many ULP SoCs have been presented, however one main limitation is the inability to cold-boot from a reprogrammable NVM. Until now, self-powered systems either required booting from a test port or non-programmable ROM. This section presents a self-powered SiP containing a 507nW health monitoring SoC, a 1Mb/s FSK radio transmitter, and a non-volatile FeRAM based cold-boot and backup memory. The die photo is shown in Figure 6.2.

6.4.1 System Overview

We designed this system with the goal of decreasing average power while retaining features of previous generation SoCs. Additionally we introduced new features to improve SoC robustness. We built the SoC from the ground up with the goal of blocks working together instead of independently, improving software development time, data flow, and system test and verification time.

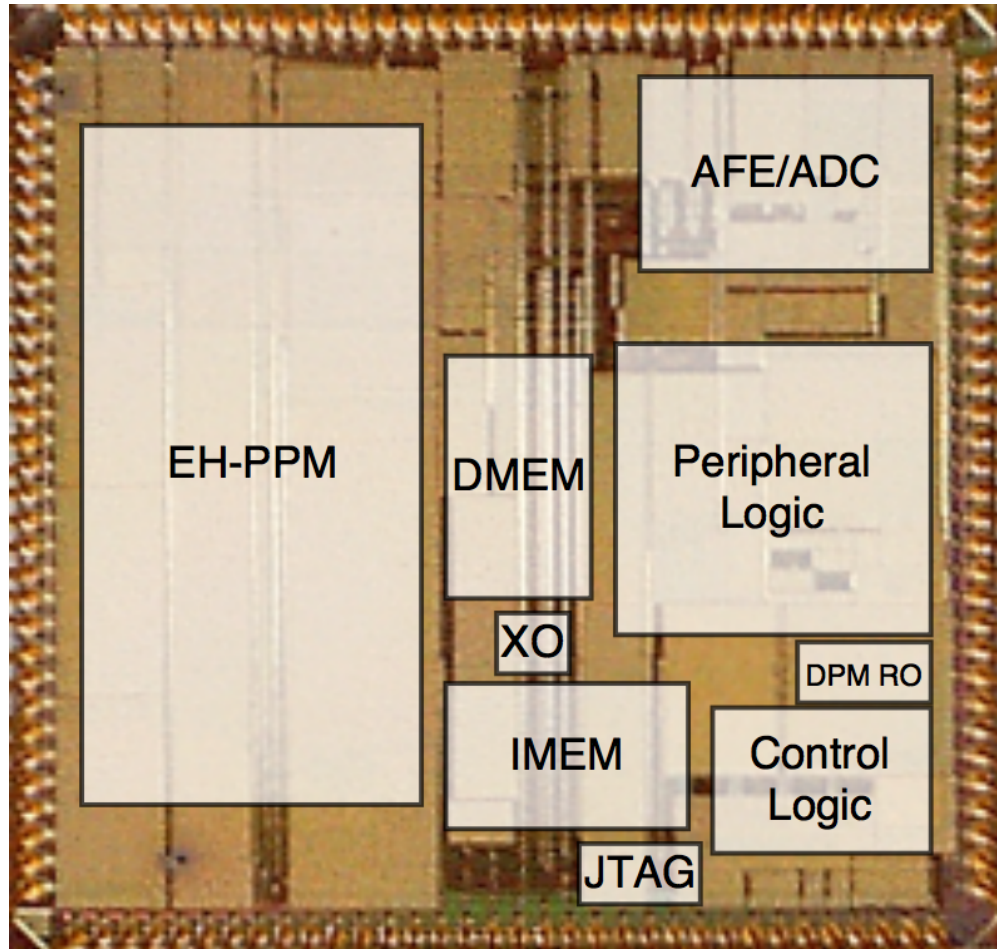


Figure 6.2: Die photo of SoC with floorplan of block placement. Die dimensions are 3.6 mm by 3.8 mm.

The SoC consists of many components for processing and making decisions from sensor data¹. The SoC contains an energy harvesting platform power manager (EH-PPM) that can harvest energy either from solar or TEG and store it on a supercapacitor. This block also includes three regulators to provide stable power rails to the rest of the SoC. The EH-PPM also contains a power up logic that generate the POR signal that starts the rest of the system. After POR, the Cold-Boot Management System (CBMS) boots the system when enough energy has been stored on the supercapacitor. It also manages the SoC based on the amount of power stored, with three power modes. The low power controller (LPC) runs 16 bit instructions from the instruction memory (IMEM) and has a built in arithmetic unit.

¹This system was a group effort by many members of the research group.

The SoC IMEM can store up to 2KB of instructions, and its data memory (DMEM) can store 2KB of data. The memories are optimized to power down banks when not in use, reducing leakage. The analog front-end (AFE) is designed for ECG measurements, and directly forwards data to the ADC. Once digitized, the ECG signal is sent to the RR interval accelerator, which is integrated with the AFIB detection accelerator, to create an entire data path that does not need management from the LPC. Because radio transmission is expensive, the data flow to the system radio has been optimized for power. The SoC contains an integrated custom serial interface with compression block, enabling data compression before transmission over the radio, reducing the regularity of transmissions. The system also contains a generic SPI interface for COTS component communication. In addition, there are general purpose accelerators, such as an FIR filter, multiplier, and timers, to handle further processing of data and control. Eight GPIO pads allow for interfaces to be created after fabrication of the SoC, and contains multiple pins that can be designated as wake-up interrupts.

This SoC benefits from many communication circuit contributions adding improvement at the system level. We designed communication circuits for a simple on-chip bus, an on-package cold boot bus, and a SPI bus to communicate with off-chip COTS components and off-chip research based systems. These contributions allow for a higher level integration than seen before through reducing total system power and enabling recovery after a loss of system power.

6.4.2 System on Chip Design Process

System in Package Floorplanning

We began the SoC design process with planning for the SiP. This involved multiple iterations of defining interfaces between the chips that will be in the package, using that information to inform what needs to be included in each chip's pad ring, and where the locations of those pads will need to be. This also required defining what other pads will be used in the pad

ring, which motivates a clear design specification early in the SoC design process. Chip pad count is limited, so budgeting which signals to bring to pads and the risk analysis of not bringing other signals to pads is needed. Once a list of pads was created, their positions in the ring was defined to ensure proper interfaces to other chips can be made while analog blocks in each chip can remain near to their respective pins to minimize parasitic capacitance and resistance.

System on Chip Floorplanning

Once the pad ring has been defined for the SoC, the internal floorplanning began. Area estimates were made for each block, and blocks were visually laid out to ensure everything will fit within the pad ring and near to each block's respective pads. For the digital portion of the SoC, the logic was separated into two main blocks: one consisting of the control, power monitor circuits, and radio interface (Figure 6.3), and the other consisting of the peripherals. This allowed minimization of power between blocks when compared to each block interfacing at the system level, while still maintaining a manageable synthesis and place and route time. Each block was separated from the rest and given its own power ring, enabling fine grained power gating of logic blocks. The blocks were laid out on both sides of the on-chip bus, minimizing wire lengths and in turn improving the maximum system clock rate at a given voltage.

Because the bus ran through the center of each block, the control block can be integrated with the peripheral block through the management of block I/O pins after compiling each separately (Figure 6.4). Ensuring that these pins are position matched for both blocks allows them to be placed right up against each other, minimizing the wire length overhead of compiling the blocks separately. This also allows for modular replacement of peripheral blocks or multiple peripheral blocks to be connected around each other and the control block in future revisions of the SoC.

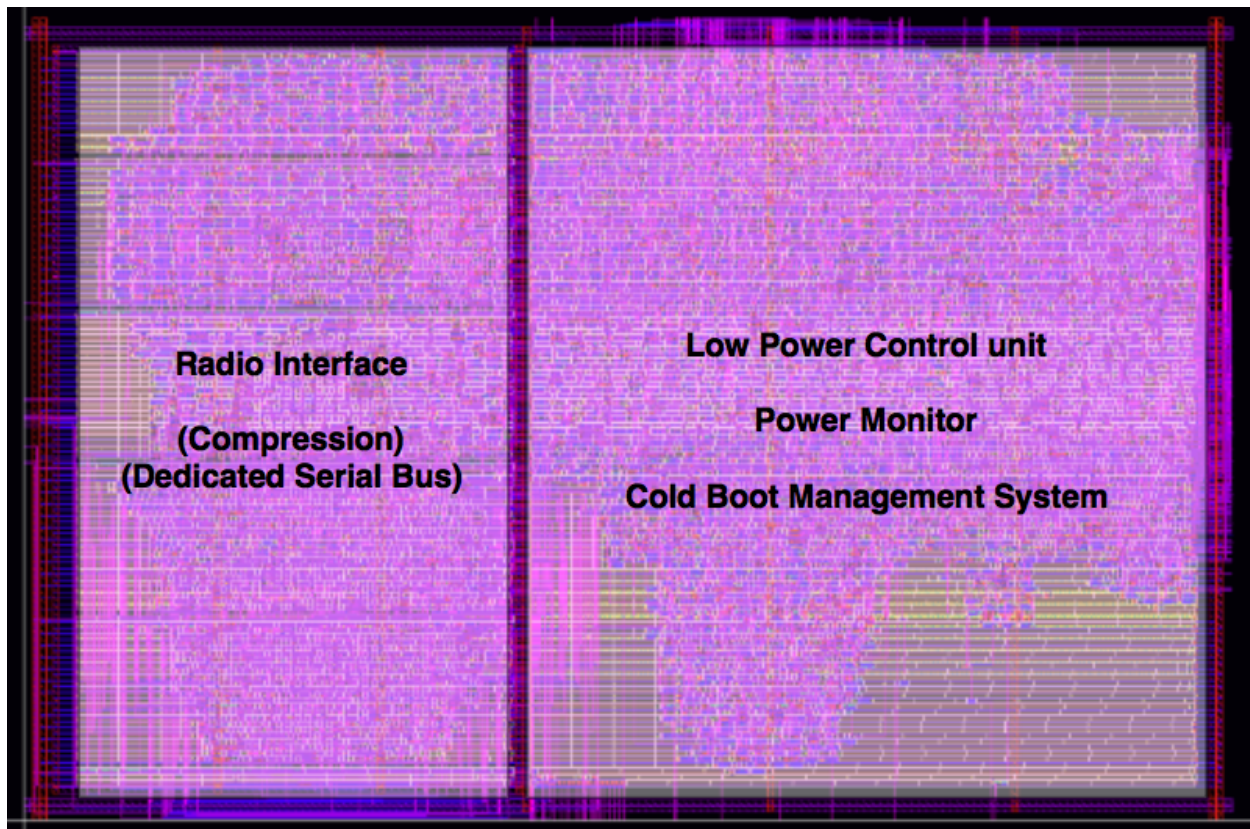


Figure 6.3: Layout of SoC control block with overlay showing breakdown of logic.

To ensure smooth start-up, the pads were designed to be controlled by the POR signal, staying in a known and disabled state until the EH-PPM can provide enough power to the pad voltage rails. This keeps the system from crashing during power up, as the EH-PPM is designed for low loads while the pads powering up can cause a large transient current consumption.

Top Level Integration

Top level integration is made easier by integrating blocks at the lower level. The main challenge during integration is minimizing wire length between blocks while also managing congestion. The instruction memory is kept as close to the control block as possible to ensure timing is met, and the analog blocks are kept against the pad ring. Once this has been completed, the entire chip layout must be verified against the schematic to ensure the

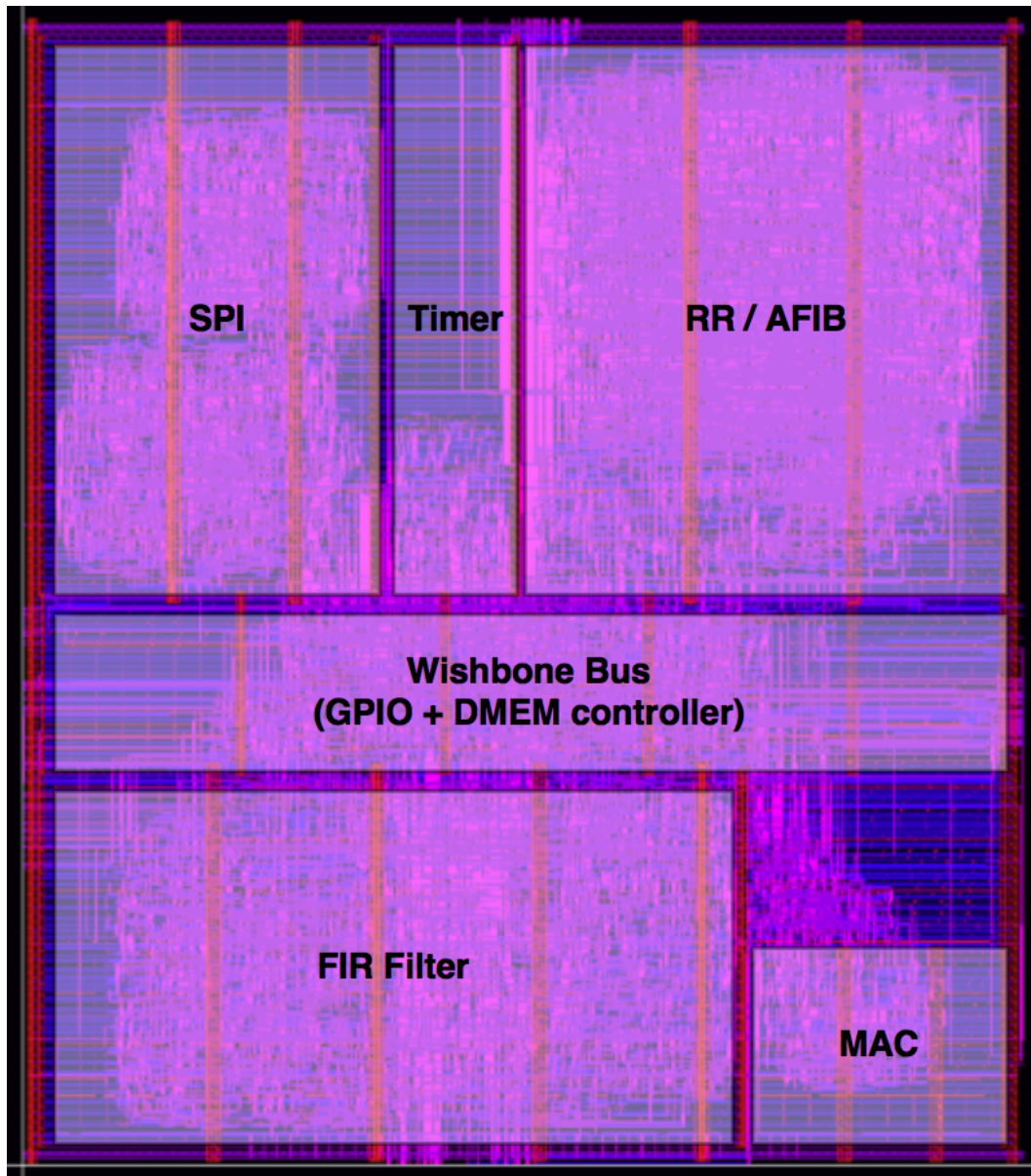


Figure 6.4: Layout of SoC peripheral block with overlay showing breakdown of logic. Wishbone bus runs through the center for lower power, higher max chip frequency, and increased modularity.

fabricated chip functions as desired. Next, the entire chip layout must be checked against design rules in the process technology, checking for geometric process limitations which might cause the fabricated system to fail. Once this is completed, a final simulation is run through the chip's pads to verify basic functionality and the chip is sent for fabrication.

6.4.3 System in Package Design Process

The SoC is designed for integration with other chips within a single package. This integration must be planned before the fabrication of any of the chips, as pads must have parallel interfaces to avoid bond wires crossing. Floorplanning of the SiP begins with estimating the area of each die and creating interfaces to define each pad ring. The number of pads on each die must be managed in order to keep the total pad count within the limit of the package. The digital logic for the interfaces is generated and verified before fabrication of either chip. The CBMS is designed for integrating the non-volatile memory with the SoC. A second custom bus is created for integrating the SoC with the radio. Because the radio packets are serial in nature, this bus interface contains a serial register file to reduce latency. Figure 6.5 shows a block diagram for each component in the SoC as well as interfaces and main components in the NVM. The brown blocks are sensing interfaces for measuring body functions and collecting peripheral sensor data. The green blocks are accelerators designed for processing sensed data, detecting results, and interrupting the LPC to handle these triggers. The yellow areas show blocks powered by the $VDDH$ rail on the fully integrated EH-PPM. The blue areas show blocks powered by the $VDDL$ rail on the EH-PPM. All digital and sensing circuits on the SoC are powered by $VDDL$.

Once chips have been fabricated, each is packaged independently and verified for functionality on its own test PCB. This step allows more control and visibility on each chip and also allows more accurate power measurements. Once each chip is verified, jumpers are placed between PCBs to test communication and power delivery between the chips. This step results in pessimistic measurements, as jumpers inherently have large capacitance, so power and energy will be higher than the fully integrated SiP. Once all of the critical components and interfaces are verified, the system can be integrated into a single package.

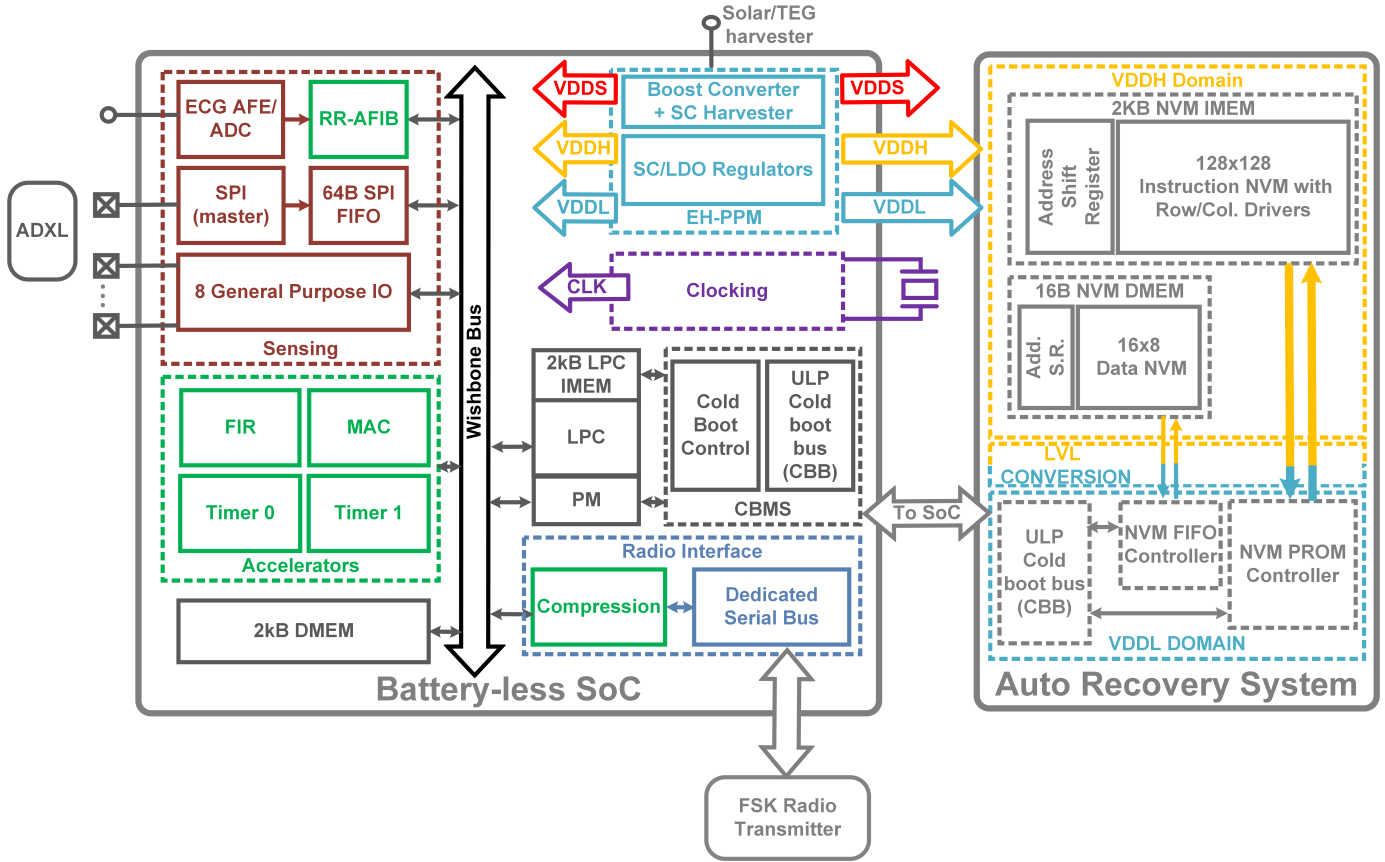


Figure 6.5: Block diagram for System in Package. Diagram shows interfaces between SoC, radio, and NVM, as well as supplied power and main logic blocks in both systems. Accelerometer (ADXL) is shown interfacing with SoC SPI and GPIO.

6.4.4 Energy Harvesting Platform Power Management

The EH-PPM consists of a boost converter and voltage multiplier for harvesting energy.² The boost converter is used in cases where the voltage harvested is low, as is often the case with thermoelectric generators. The voltage multiplier is used in cases where the harvested voltage still must be increased, but by a lesser amount. This is the case for small photo-voltaic cells. The EH-PPM also contains three regulators for use in different parts of the system. A 1.8V rail is generated for powering off-chip COTS components, and in our example system, powers a COTS accelerometer. A 1.0V rail is generated for powering the radio and NVM. The 0.5V rail is designed for powering the digital and analog on-chip components. Figure 6.6 shows a

²The EH-PPM was designed by Abhishek Roy.

block diagram of the EH-PPM. Harvesting, regulation, and power up controls are all handled on-chip.

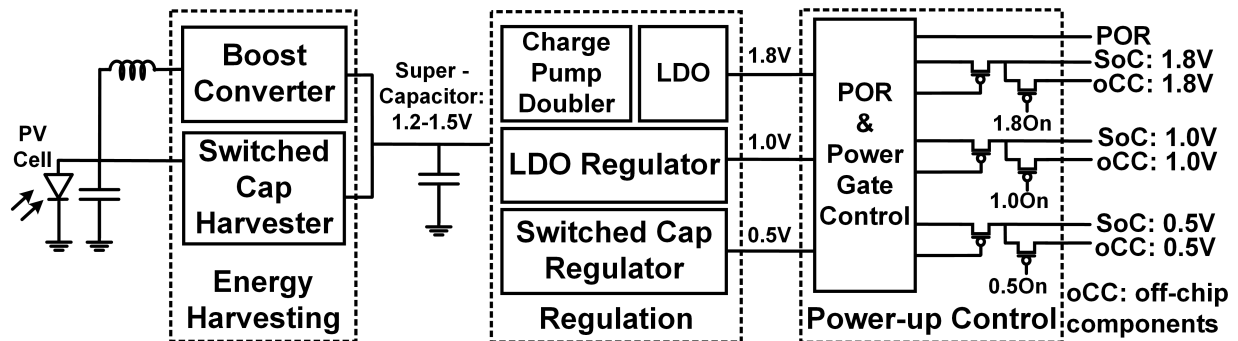


Figure 6.6: Block diagram for the Energy Harvesting Platform Power Manager. This subsystem harvests energy and provides power to the SoC, NVM, radio, and COTS components.

6.4.5 LPC-Memory Integration

At the heart of the SoC is a custom general purpose low power controller (LPC) that runs the application. The LPC instruction set is designed to be compact and to make use of the different features of the SoC. It consists of move, stall, jump (conditional and unconditional), arithmetic, and logic instructions. The LPC has eight stall conditions and nine interrupt sources that are linked to the different building blocks of the SoC. For example, the first GPIO pin can be configured as an interrupt source to force the LPC into the interrupt service routine. The programmer can also choose to stall until that GPIO pin is set externally. Table 6.1 shows the sources of stalls and interrupts.

The LPC is also tightly integrated to its 2KB SRAM instruction memory (IMEM). The IMEM is built from eight high-threshold voltage transistors to reduce leakage power during standby state. It also has four power saving modes: read burst, hold, standby, and shutdown. In read burst mode, the IMEM saves the contents of a row in latches to reduce the number of memory reads required and thus the power consumed. In hold mode, the clock signal is gated to reduce switching activity. In standby mode, the peripherals within the IMEM are power gated to reduce their power consumption. In shutdown mode, one or both of the

Table 6.1: The LPC stall conditions and interrupt sources.

Source	Stall Condition	Interrupt Condition
Timer	until Timer 0 expires	when Timer 0 expires
	until Timer 1 expires	when Timer 1 expires
GPIO	until GPIO 0 is set	when GPIO 0 is set
	until GPIO 1 is set	when GPIO 1 is set
Radio Interface	until serial transmission is done	when serial transmission is done
	until radio transmission is done	when radio transmission is done
RR-AFIB	until RR interval is captured	when RR interval is captured
	until AFIB is detected	when AFIB is detected
PM	until PM goes out of red mode	when PM signals backup is required

IMEM banks are completely power gated and their data is lost. The LPC takes advantage of these power saving techniques within its IMEM to keep the system within the power budget. The stall condition of the LPC is directly linked to the standby mode of the IMEM, meaning that when the LPC stalls its IMEM goes into standby mode. The LPC can also disable one bank of the IMEM if the program size is small enough to fit in one bank, and enable read burst mode to reduce the system power consumption. These features (one bank disabled and standby) allow up to 65% reduction in the power consumed by the SoCs control unit. Figure 6.7 shows the measured power savings due to the introduced features.

6.4.6 On-package Cold-Boot Management System

The CBMS consists of the Cold-Boot Controller (CBC) and the Cold Boot Bus (CBB), and is responsible for communicating with the NVM to retrieve data after complete power loss. The CBMS is also responsible for programing the IMEM and signaling the LPC when the cold boot is completed. The PM initiates the startup sequence when the EH-PPM has stored enough energy on the supercapacitor to power the system and the NVM. The CBMS signals the PM when instructions have been successfully loaded, and the PM powers down the NVM. The CBMS boot process reduces the boot time through multiple optimizations, making a system boot from NVM possible.

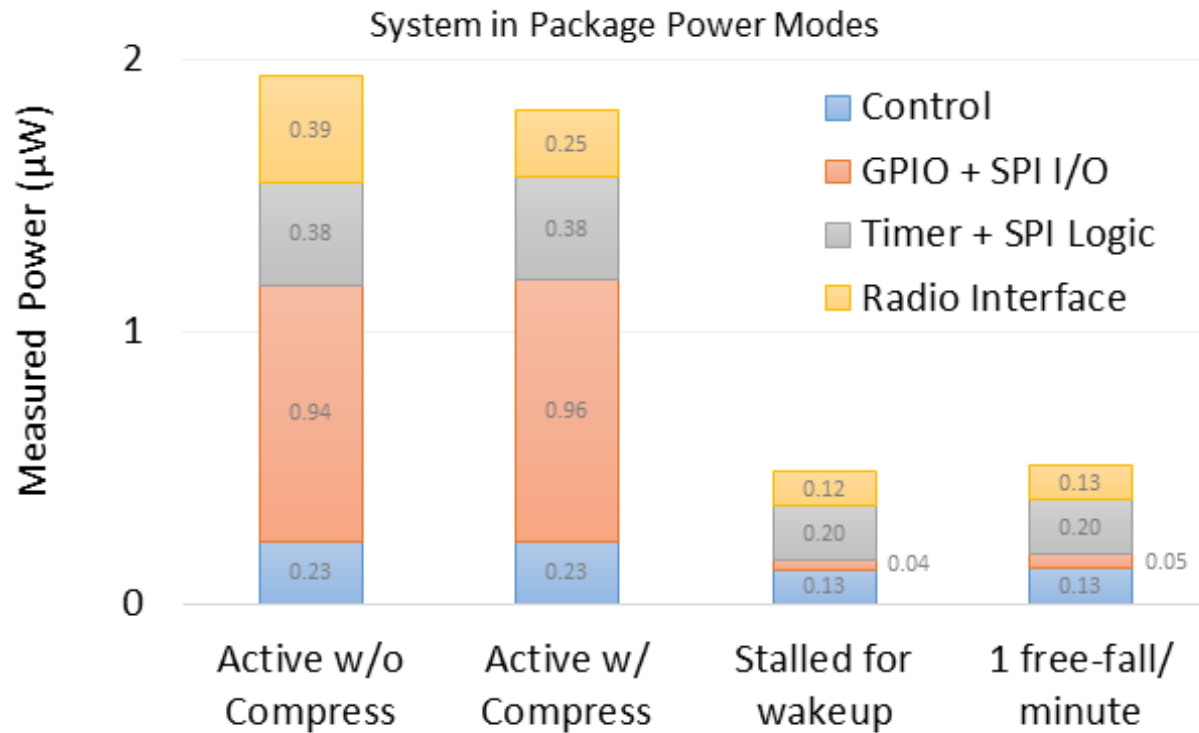


Figure 6.7: Chart showing power modes for the System in Package. System power is greatly reduced while stalled and waiting for interrupts.

During the SoC's boot phase, the SoC sends a bootup request command and instructions are loaded from the NVM. During the backup phase, the SoC sends the appropriate command and critical data can be backed up to the memory. The NVM will backup 16 bytes for recovery. This can be necessary when the system has not been harvesting and is soon to lose power.

6.4.7 Non-Volatile Boot Memory

The NVM³ holds the program instructions as well as any critical data the SoC produces. It consists of a 2KB instruction memory (IMEM), a 16-byte data memory (DMEM), IMEM and DMEM controllers, and the CBB for interfacing with the SoC. The IMEM is optimized

³The NVM was designed by Farah Yahya, and the NVM to bus interface and CBB were designed by myself.

for reduced read energy. It is read one row at a time and the contents of the row are held in low power latches until the CBC finishes transmitting them to the SoC. Thus, a read operation is performed to the IMEM every $R/8$ cycles where R is the number of bit per row. This reduced read rate amortizes the active read energy of the NVM and keeps it within the power budget of the harvester. To further reduce the system energy, the design of the array is simplified. Since the NVM IMEM acts as a programmable read-only memory, random access is not required. This allows us to simplify the design of the memory access decoder into a simple shift register.

The IMEM and DMEM controllers read the commands from the CBB and manage the control signals of the IMEM and DMEM. To reduce their power consumption, these controllers run at a near/sub-threshold voltage provided by the SoC. Level conversion is added for the signals going into the IMEM and DMEM. To keep level conversion power at a minimum, only the control signals (read, write, shift, etc.) are converted. To synchronize between the CBB and the rest of the system, SR latches are used for command handshaking between the CBB and the controllers.

The CBB (Chapter 5) is a parallel interface allowing communication between the NVM and the SoC. The bus is a two direction interface, with one pin going from the SoC to NVM, and eight pins going from NVM to the SoC. This shortens the boot phase without the need for generating a higher frequency clock. The CBB also uses reduced swing transmitters and receivers for energy and power savings.

6.4.8 Boot Sequence

The SoC is capable of booting without any additional programmer equipment or benchtop support. The SoC's EH-PPM and PM begin the boot sequence when enough energy has been harvested to the supercapacitor. The EH-PPM provides a power-on reset once it is able to provide a stable voltage at each rail. This allows the crystal oscillator to begin oscillating. Once the crystal has stabilized, a digital reset signal is triggered, a voltage controlled oscillator

(VCO) begins measuring the voltage on V_{cap} . The voltage threshold for booting is configurable off-chip and will depend on the application and harvesting modality. Once the measured V_{cap} exceeds the bootup threshold, the PM signals the CBC to begin the request of program data from the NVM chip. Once the program data is received, it is loaded into the IMEM and the LPC begins executing instructions. The non-volatile data memory is then loaded for applications needing higher levels of quality of service, even when loss of power events occur.

Figure 6.8 shows this process.

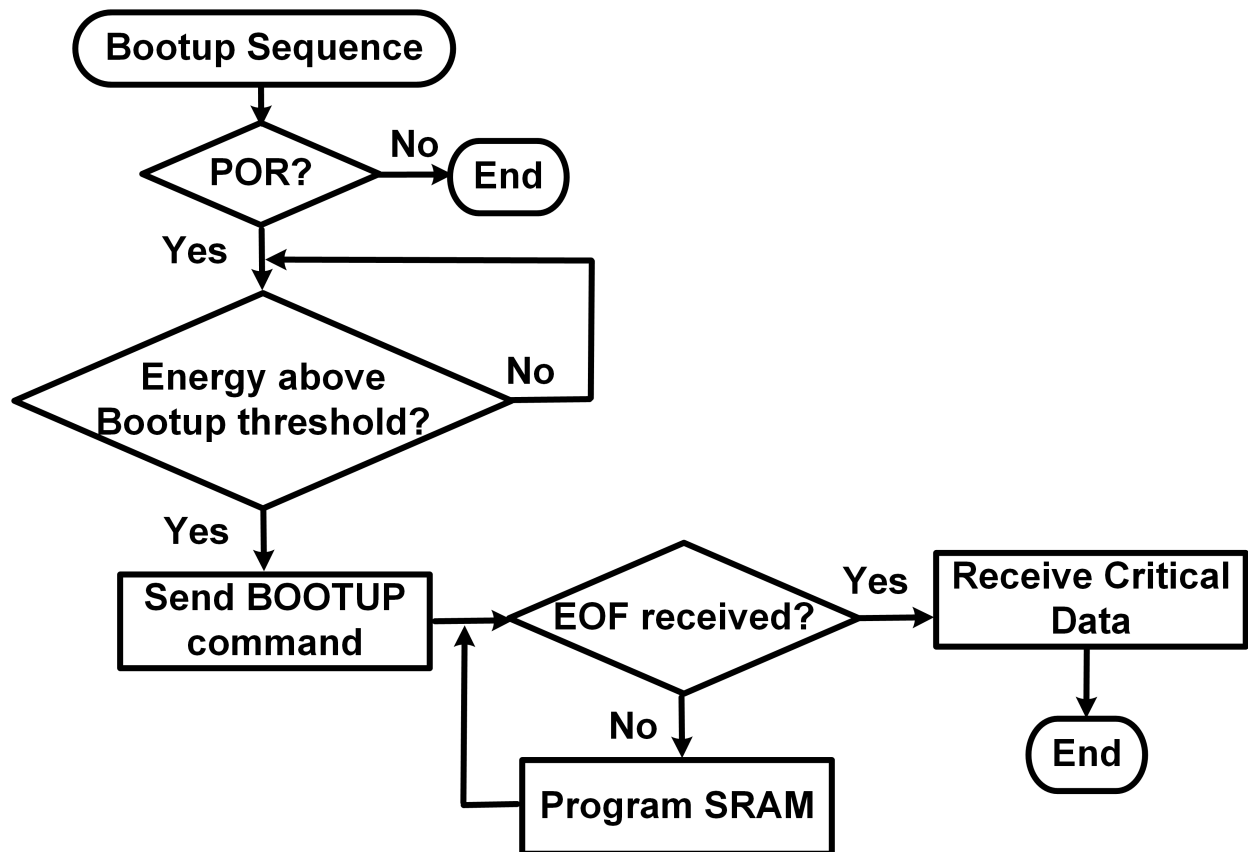


Figure 6.8: Flow chart for SoC bootup sequence from NVM. This feature allows the system to recover after long periods without harvesting.

6.4.9 Backup Sequence

In the event of power loss, the system still needs to retain critical information to ensure that it can continue in the same state it was when power was lost. To address this, the CBMS is

also capable of backing up 16 bytes of data for recovery when the system power is restored. An example of use would be counting the number of specific free-fall events in a shipping integrity system. If the system will soon lose power, it can backup the count to the NVM and load it when harvesting resumes.

The backup process, shown in Figure 6.9, begins when the PM detects a low voltage on V_{cap} . The PM sends an interrupt to the LPC, giving the user control over what data is backed up. The user may also choose not to back up any data. In this case, the process ends and normal operation continues. If the user chooses to back up critical data, the program will write to the CBMS memory mapped address and notify the PM when this has completed. The PM then tells the CBMS to begin the backup transfer. The CBMS sends a request for backup over the CBB, followed by the number of bytes and the data itself. Once this has completed, the LPC continues operation.

6.4.10 Radio

The SiP includes a 1 Mb/s 724 μ W frequency shift key (FSK) transmitter for wireless beacon applications.⁴ The FSK transmitter is powered by the EH-PPM, and includes a custom serial interface to the SoC. After configuration from the SoC, the transmitter calibrates frequency, and the power amplifier transmits a packet over a 400 μ s period. The radio can be powered down by the SoC when not in use.

6.4.11 Custom Radio Interface

The SoC contains a custom radio interface containing a compression accelerator and a Dedicated Serial Bus (DSB).⁵ The compression block stores the compressed data in a FIFO, which is then forwarded to the DSB. The DSB has a memory mapped interface to the system bus, waiting for a transmit command. The DSB also contains a state machine to handle the

⁴The FSK transmitter was designed by Xing Chen, and the baseband processor and SPI interface were designed by Farah Yahya and myself.

⁵The compression accelerator was implemented by Jacob Breiholz.

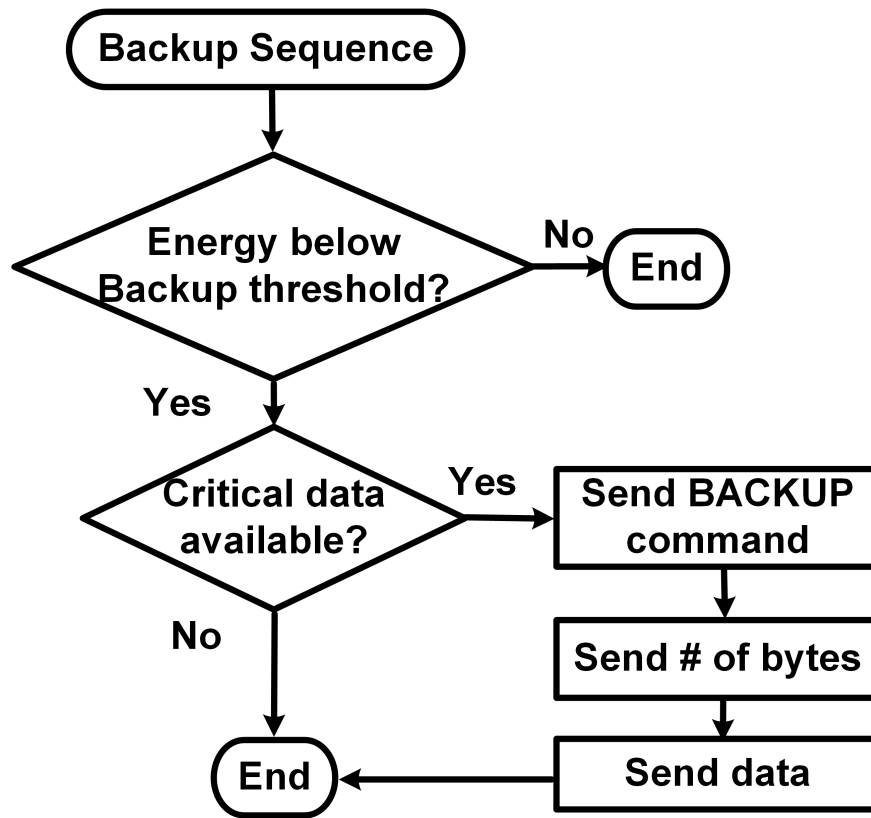


Figure 6.9: Flow chart for SoC backup sequence to NVM. This feature improves quality of service for applications which the system must keep track of events over a long period of time.

transmission of each command differently, transmitting only the number of bits necessary, with no byte or word alignment. This keeps the transmission time to a minimum, removing the need for the padded zeros a conventional SPI would have. As a result, the latency is reduced, and the radio power on time is minimized.

6.4.12 Application: Wakeup and Fall Detection

The SoC is demonstrated through a shipping-integrity tracking application that uses the SPI, GPIO, LPC, timer, compression, and radio interface. The SoC begins with configuring an accelerometer[11] for this application. The accelerometer is configured to log acceleration data and generate an interrupt when a free-fall event occurs. After configuration occurs, the SoC enters a ULP sleep mode while waiting for an event to occur, consuming 485nW. Once an

event occurs, the SoC powers on SPI, collecting and logging the previous half second before the event occurred. The LPC is turned on to process this data. The LPC then sends the data through the compression block and powers on the radio. Once the data is compressed, it is transmitted through the custom serial interface and transmitted. Because the radio has a latency, it responds to the SoC with an interrupt when transmission is complete. At this point the SoC powers down the radio transmitter, radio interface (with compression), LPC, SPI, and timer. The SoC then returns to the interrupt driven sleep mode, waiting for another fall event. The process and relevant accelerators used in each step are shown in Figure 6.10.

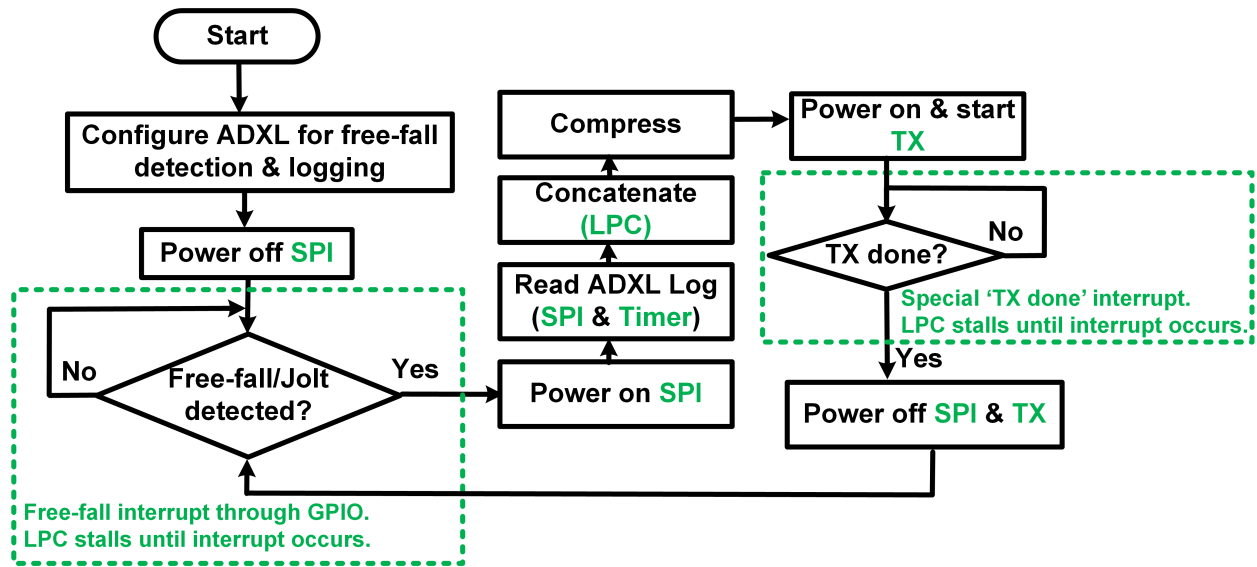


Figure 6.10: Application example for SoC. The system utilizes its reduced power sleep mode while waiting for an interrupt from the COTS accelerometer. This allows the system to reach an average 507 nW power consumption.

6.4.13 Measurements and Comparison

Measurements were made at each step in the application to find typical power numbers. While the SoC is active, its average power is 1.94 μ W. This includes communication with the accelerometer, compression, and radio communication. While active and not communicating, the SoC consumes 610 nW. When in low power mode, waiting for an interrupt, the total consumption is 485 nW. Averaging to assume one fall event every minute, this power becomes

507 nW. Table 6.2 shows comparisons to state-of-the-art ULP systems.⁶ The proposed system shows higher levels of integration with a lower total power consumption.

Table 6.2: Comparison to state-of-the-art SoCs

	This Work	[51]	[46]	[52]	[25]
Battery-less	Yes	No	No	Yes	Yes
Harvests Power	Yes	No	Yes	Yes	Yes
Integrated EH-PPM	Yes	No	No	No	No
Powers off-chip sensors	Yes	No	No	No	No
Regulated Voltages	1.8, 1.0, 0.5V	0.25-1.2V	-	unregulated	1.2, 0.5V, var.
Interface to NVM	Yes	No	No	No	No
On-chip SRAM	4KB	24KB	3.7KB	256B	12KB
Accelerators	5	2	3	-	7
Sensing Interfaces	3	2	1	-	2
Total Power	507 nW	850 nW	45 nW	295 pW	2.3 μ W
Logic in Total Power	MCU+SPI+ IO+Timer+ GPIO+RI	MCU	AFE+ DSP	MCU	MCU+IO+ SPI+FIR

6.5 Self-powered, Battery-less Positioning System

Global positioning is a tool used to find the location of a device, giving the user an accurate measure of their position on Earth. The most common application of positioning is navigation by cars, aircraft, and ships. It is also used for emergency purposes and recreational sports. A Global Positioning System (GPS) generally consists of a radio receiver and an Application Specific Integrated Circuit (ASIC), and requires three or more satellites orbiting in the same region of the Earth as the user. The satellites' positions and the devices' distances from each of the satellites are needed to triangulate location. The radio receiver acquires transmissions from satellites, and the ASIC uses the received data to calculate how long it took the radio signals to travel from the satellite to the receiver. The receiver continues to collect transmitted data from the satellite over time to retain a real-time position.

⁶In table 6.2, MCU is defined as the main controller and its instruction memory

Another method of maintaining a real-time position is through the use of an accelerometer. This method is generally less accurate, but does not require a radio or active satellites to acquire a position and is usually used in inertial navigation systems (INS). The accelerometer starts from a known position and updates that position using sensed acceleration. Numeric integration is performed twice on a window of acceleration data and a real-time position is acquired. This is known as the Euler method. Other more accurate methods such as Verlet integration and the Runge-Kutta methods have been presented. Because optimizing the positioning algorithm is beyond the scope of this work, the Euler method will be used to create an accessible proof-of-concept.

6.5.1 Algorithm Flow

Figure 6.11 shows the program flow for measuring position. First, the accelerometer must be configured to operate in the correct mode and at the desired frequency. The SoC will then collect a window of acceleration vectors from the accelerometer and integrate them to find a position step. As each position step is calculated from its window of acceleration, the sum is found to achieve the current position.

6.5.2 Accumulating Error

INS suffers from high growth in error over time due to errors in numerical double integration. Because of the double numerical integration, the units of error are proportional to time squared, meaning that while error may be low in the short term, it becomes poor after a long period with no correction.

Many applications do not need a relative position measurement for a long period of time. Applications involving measurements of cyclical movement such as measuring lengths of a person's step or keeping track of relative position in rotating machinery allow the system to start with an initial position with every new cycle. Research has been done in reducing INS

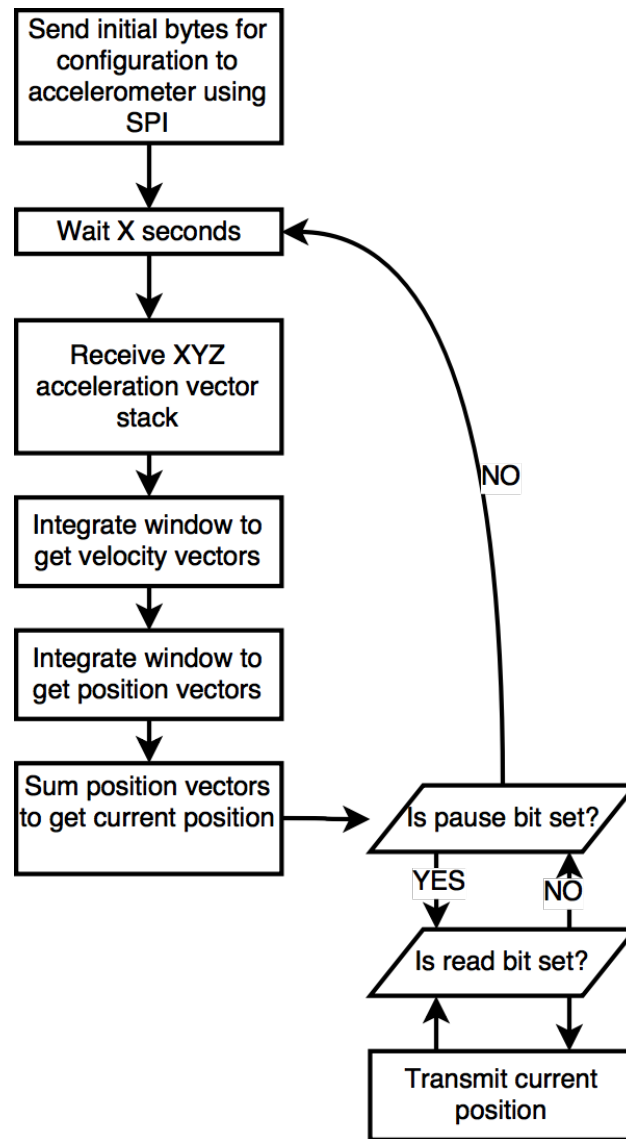


Figure 6.11: Flow used for continuous position measurements in self-powered relative positioning system.

error through sensor fusion [47] to achieve an accuracy of one to two meters in a 600 second test period, further motivating the use of many nodes in an IoT for problem solving.

6.5.3 Results

We tested the positioning system using the SoC in section 6.4 with the same accelerometer used in the free-fall detection demonstration [11]. The SoC consumes an average of approximately $1.8 \mu\text{W}$ while performing the calculations for positioning.

6.6 Modeling Trans-threshold Correlations for Reducing Functional Test Time in ULP Systems

New developments in circuit technology are allowing systems to be designed with continually improved energy and power efficiency. Energy and power improvements mean that systems can operate for a longer period of time on a single battery charge, or even indefinitely through extreme energy reduction, power management, and use of high energy density super-capacitors [26]. Improvements to circuit designs and operation at lower voltages make this possible. Because these systems have been designed to operate differently than conventional circuits, a testing methodology must be created in order to properly ensure functioning systems.

Chip testing is a multi-step process that ensures the end product functions as expected. One of the first steps in testing is logic verification. Logic can be simulated before fabrication to ensure it functions as specified. After chips are manufactured, the first fabricated chips are verified for logic bugs and electrical failures. Certain physical attributes of a circuit are challenging to simulate, and might result in unexpected failures after manufacturing. Thus, in the first stage, the manufactured design is verified against the design specifications. After verification is completed, chips are ready for mass production. At this stage, every chip must be tested for manufacturing errors. Faults generally occur when defects in the process cause bridging and open circuits in wires. Often these are modeled as stuck-at faults. Delay faults can also be induced by process variations. Chips must be tested for these faults, as they are a random process that cannot be predicted. This test process happens after fabrication to catch errors that happened in manufacturing. Functional (delay) testing is one of the most expensive steps in chip design. This is because all chips must be tested before delivery of parts to ensure products are free of errors [22]. In addition, test equipment is very expensive. As a result, chip test time must be minimized to manage chip costs.

In order for commercial self-powered systems to provide application level solutions, more research must be done in testing near and subthreshold SoCs. One such area is optimizing

post-manufacturing functional test time. In this section, we propose increasing the voltage of a ULP SoC during test to decrease overall test time by increasing the frequency, as the maximum frequency of operation of a circuit is a function of that circuit's supply voltage. We then find correlations between the circuit delay distributions at high voltages and low voltages to allow testing at higher voltages and accurately calculate expected operation at lower voltage. This will allow for faster testing and accurate models of how the circuit will fail at lower voltages for chip yield in manufacturing.

6.6.1 Scope of this work

Functional test time in the scope of this dissertation is defined as the time needed to test digital circuits, mainly in SoCs, for proper logical results at the desired full-speed clock rate. Subthreshold systems have an increased challenge with digital timing due to increased effects of each device's threshold voltage variation. Because subthreshold current is an exponential function of the threshold voltage, a normal distribution in V_T results in a log-normal distribution in current, having a long tail in the slow side of devices. To make matters worse, on-current to off-current (I_{on} to I_{off}) ratios are decreased, meaning that smaller variations in current will cause failures in logic gates. This physical side effect of low voltage operation means clock uncertainty and yield become problematic.

SoCs contain 3 main categories of circuits. Digital circuits perform most of the computation, memories store instructions and the results of computations, and analog/mixed signal circuits collect sensor information from the outside world and pass it on to the digital circuits for analysis. The technique described in this section covers testing of digital and memory circuits, however it can be scaled to analog and mixed signal circuits, assuming they are designed for a range of operating voltages. This section will be dedicated to proving the method for digital logic circuits.

This work uses trans-threshold correlations (correlations between subthreshold and nominal operation) to rigorously predict low-voltage delay during high-voltage, high-speed testing,

reducing functional test time in subthreshold SoCs. This improves the process of measuring functional failures and binning chips into groups based on power consumption or maximum frequency. The process is done with no additional on-chip circuits, resulting in a post-fabrication solution with no area overhead. This method is orthogonal to existing test methods, meaning that decisions in test coverage and methods of conducting a certain test remain unchanged, with the exception that they occur at a different voltage. The main limitation of this work is in Test Access Mechanisms (TAMs). Logic block wrappers and built-in self test structures (BISTs) must be able to operate at a higher frequency than the maximum frequency of the Device Under Test (DUT). In digital terms this means the TAM must have a higher setup slack. One potential solution to mitigate this limitation is to operate TAMs at a higher voltage around the DUT, giving it a higher maximum operating frequency. Existing work has been done on low voltage testing and SoC testing, but none of them handle the increased test time problem of systems designed for lower voltage and subthreshold operation.

6.6.2 Correlation Between Voltages of Operation

The maximum operating frequency of a circuit is a function of that circuit's voltage. Increasing the voltage of an SoC during test can decrease overall test time by increasing the maximum operating frequency. Because of variations, not every chip will increase its maximum frequency by the same multiple for a given voltage change. However, correlations can be found between the distributions in higher voltages and lower voltages to allow functional testing at higher voltages while accurately predicting expected frequencies at lower voltages. This will allow for characterization of how the circuit will fail at lower voltages for chip yield in manufacturing.

Fault models are used to determine the cause of errors in a circuit. Different faults behave differently throughout voltages of operation, and will have different correlations between those voltages. Stuck-at faults and bridging faults result in logical errors and will present

themselves the same way at many voltages. Such faults will have a high correlation. As a result, testing SoCs at nominal voltages and high frequencies will illuminate these faults.

Delay faults, on the other hand, do not always have high correlations between a large change in operating voltage, as each transistor has a different threshold voltage and therefore different delay curve as a function of voltage. If paths of slow transistors (high threshold voltage caused by variation) cause a propagation delay longer than the clock period, a delay fault (setup violation) is present. Functional tests are applied at the operational speed of the SoC to test for delay faults. Using a subset of chips to be tested for characterization, delay correlations can be found between chips operating at a low voltage and nominal voltage. These correlations will give further insight into how the systems' delays change over voltages, enabling high speed testing of future chips of the same design.

6.6.3 Implications of Voltage Correlations

Circuits with a high enough delay correlation between high and low voltages can be tested at high voltages. Predictions can then be made about how they operate at the lower voltage, resulting in binning and pass/fail results based on testing requirements. Higher ΔV between allowable test and operating voltages means exponentially less test time is required, due to the exponential change in delay as a function of V_{DD} in subthreshold SoCs. However, for some faults, the correlations decrease as test voltage increases. Cases must be broken down and analyzed separately.

6.6.4 Faults Containing High Correlation

Many faults contain a correlation approaching 1 when comparing test results at low and high operating voltages. This is because some faults are not a function of voltages. Faults like opens, shorts, and stuck-at faults do not change with operating voltage and can be tested at any voltage, yielding the same result. An SoC that normally operates at 300 or 400 mV can

be tested for these faults at the process technology's nominal voltage (approximately 1V) resulting in over 1000X speed up in testing for these cases.

6.6.5 Faults Containing Varying Correlation

Faults with varying correlation as a function of operating voltage offer the most interesting analysis and perspective opportunities for test time reduction. Delay faults, for example, will appear at higher system frequencies with increased voltage. Coupling faults will appear at higher voltages while not existing at lower voltages, due to higher transition speeds increasing coupling effects.

Delay has been shown to decrease in correlation with an increase in testing ΔV as shown in Figure 6.12. In other words, as the test voltage increases, less is known about the system delay for a given chip at a low operational voltage.

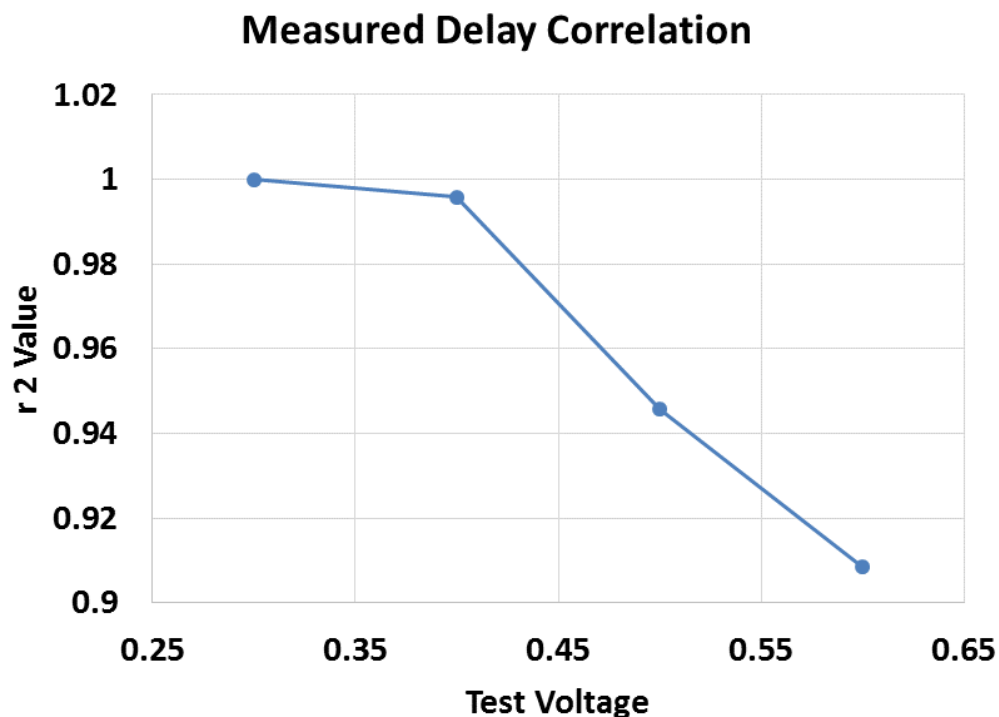


Figure 6.12: Measured delay correlation in a logic block containing combinational and sequential circuits. r^2 is used to determine how well the data fits the regression, and is generally within the range of 0 to 1. An r^2 value of 1 implies a perfect fit, while 0 implies no correlation between the line of best fit and the data.

6.6.6 Methodology

In order to properly and confidently test subthreshold chips at a high voltage, a rigorous statistical analysis must be done. The process starts by creating a model of the correlations between different ΔV s. This model will define a point of uncertain failure based on the point of defined failure. The point of defined failure is the defined point of failure at the specified (low) operating voltage. The point of uncertain failure is the high voltage threshold below which the success or failure at the low operating voltage cannot be determined with certainty. This point is determined by correlating the point of defined failure to the high voltage while taking into account the prediction interval. Next, chips are tested and the metric of each is compared to the point of uncertain failure. If the metric exceeds it, then the chip is functional. Otherwise, the chip must be retested at the low operating voltage. Since retesting at low voltage is time consuming, it is necessary to keep the number of chips that need retested to a minimum. For this reason, an optimal high voltage for testing will be determined that balances the speedup gained with the penalty of retesting.

Before describing the methodology in details, we start by defining a few statistical terms that will be used in the rest of the section. The present population is the population of chips manufactured to date. Sample size is the number of chips chosen to create a statistical sample set. This set of chips are used to create a characterization of the present population. Statistical power is a measure of desired accuracy when a sample set is used for predictions. A prediction interval is an estimate range of future observations based on the existing sample set. The prediction interval can be used on a single new sample to predict the interval of a correlating metric. Stratified sampling is to take samples from known groups to potentially decrease the necessary sample size. In the case of chips, this would be to sample from all process corners to achieve a sample set more representative of the entire population.

The details of this methodology are described in the following steps and in Figure 6.13:

1. **Determine Sample Size** - The sample size should have a strong statistical power, and should include samples of the present population. Power analysis can be used to

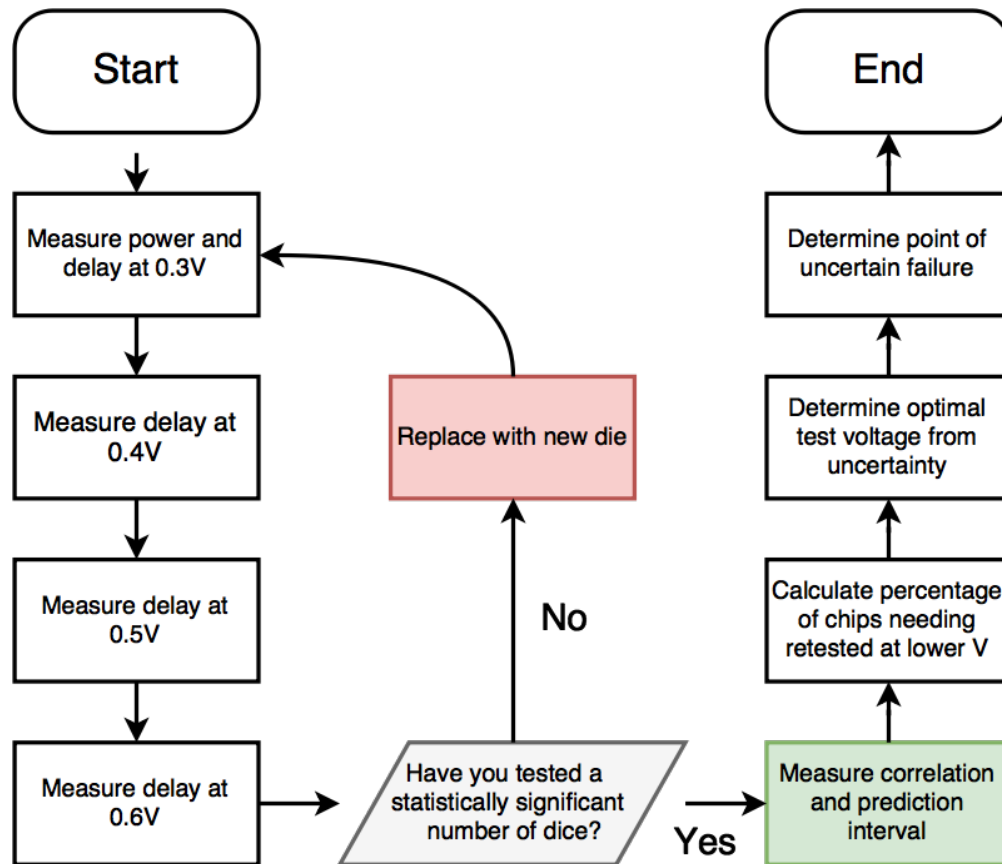


Figure 6.13: The characterization flow chart describing the different steps required before mass testing can be started.

determine the minimum sample size.

2. **Determine Sampling Method** - The sampling method will depend on the availability of chips from different wafers, however simple random sampling is preferred if possible. Including samples from wafers as they are fabricated ensures samples are of the present population. Test time of characterization is considered insignificant due to the small relative number of chips needed for characterization in comparison to the number of chips generally produced commercially. Commercial chip fabrication is in the millions of chips per product, while the desired statistical significance level can be found in tens or few hundreds of chips.

3. **Test Sample Chips** - The delay and other relevant metrics are measured at multiple operating voltages for the sample set to create a model of the population.
4. **Calculate Correlation as Function of Voltage** - Each ΔV will result in a different delay correlation, decreasing as ΔV increases. At this step, the point of uncertain failure is determined for the high voltage to aid in the next step. This point is defined based on the prediction interval.
5. **Optimize Test Time** - Higher voltage testing is faster but will result in wider prediction intervals, resulting in more chips needing retested at a lower voltage. A minimum total test time can be achieved by minimizing the sum of initial test time and retests at low voltage.
6. **Calculate Predicted Delay of New Sample** - During the test process, each chip is tested at the chosen high voltage, predicting its true subthreshold operating delay. If the predicted delay is below the point of uncertain failure, it must be retested.
7. **Retest Samples on Tail of Prediction Interval** - Samples with predicted delay below the point of uncertain failures will be retested at a slower, more accurate, lower operating voltage to ensure correct delay calculations. Because test speed-up is so significant in this method, low voltage retesting is undesirable and should be kept to a minimum.

The method above creates a model using a sample set of chips to optimize test time through properly tuning the test voltage which impacts both the prediction interval width and number of samples needing to be retested at a lower voltage. Testing at a higher voltage drastically reduces test time because delay decreases exponentially with increased operating voltage. However, due to lower correlations in delay between test voltages, the prediction interval width increases. The uncertainty in predicted delay is higher with a wider prediction interval, meaning more chips near the failure point need to be retested. This trade-off will

create a minimum test time at a voltage between the subthreshold operating voltage and nominal process voltage.

Once the optimal test voltage is determined, large scale testing can be started. The test procedure for this stage is defined in the flow chart shown in Figure 6.14.

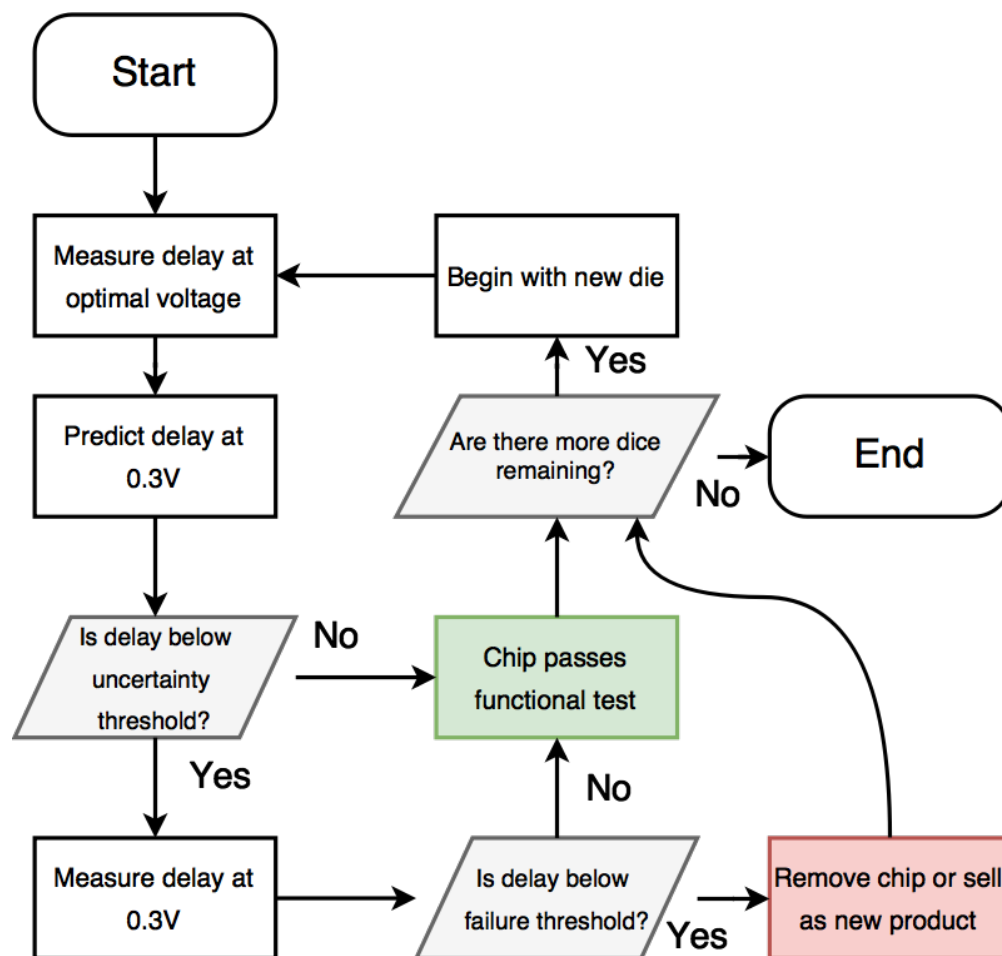


Figure 6.14: Flow chart showing test flow after characterization has been completed. Test flow reduces time by predicting low operating voltage delay.

6.6.7 Applied Methodology to Simulated Ring Oscillator

To test this methodology, a small ring oscillator (RO) was created and simulated in SPICE using Monte Carlo to model the effects of variation. In the rest of this section, we go through the test methodology one step at a time and explain it through the RO example.

Sampling and Characterization

Ideally, a true simple random sampling will be used to produce an accurate characterization of the population, however in many cases this may not be possible, as not all chips might be fabricated before characterization. In this case, stratified sampling can be used by collecting random samples from each process corner in an attempt to get a complete view of all process corners. This will yield a wide multimodal distribution in delays covering all ranges before having access to the complete population of fabricated chips. Once the samples are selected, each chip is tested for delay (or other desirable metric) at a range of operating voltages, first at the subthreshold operational voltage, and last at the nominal voltage of the process technology.

In our RO example, 1000 samples were randomly selected from a global corner (which includes samples from all process corners) and used as the sample population. Next, these samples were simulated at different voltages and their maximum frequencies were calculated. Figures 6.15 and 6.16 compare delays between the subthreshold operating voltage (0.5 V) and a voltage showing high correlation (0.6 V) and a voltage showing low correlation (0.8 V).

Statistical Analysis

The correlation between each test voltage and the subthreshold operating voltage is calculated. A line of best fit is used to visualize how well the delays correlate. In the case of subthreshold circuits, taking the logarithm of the delay is necessary to allow correlation calculations on a linear, Gaussian sample set. The line of best fit has a metric calculating error called the r^2 value. Ensuring this value is high (above 90%) is important to keep the prediction interval tight, resulting in fewer chips to retest. To ensure the line of best fit is as close as possible, residuals are inspected for patterns. If a pattern is visible then the line of best fit does not completely represent the data set. Residuals showing no pattern and a high r^2 mean that the line of best fit is now usable for prediction.

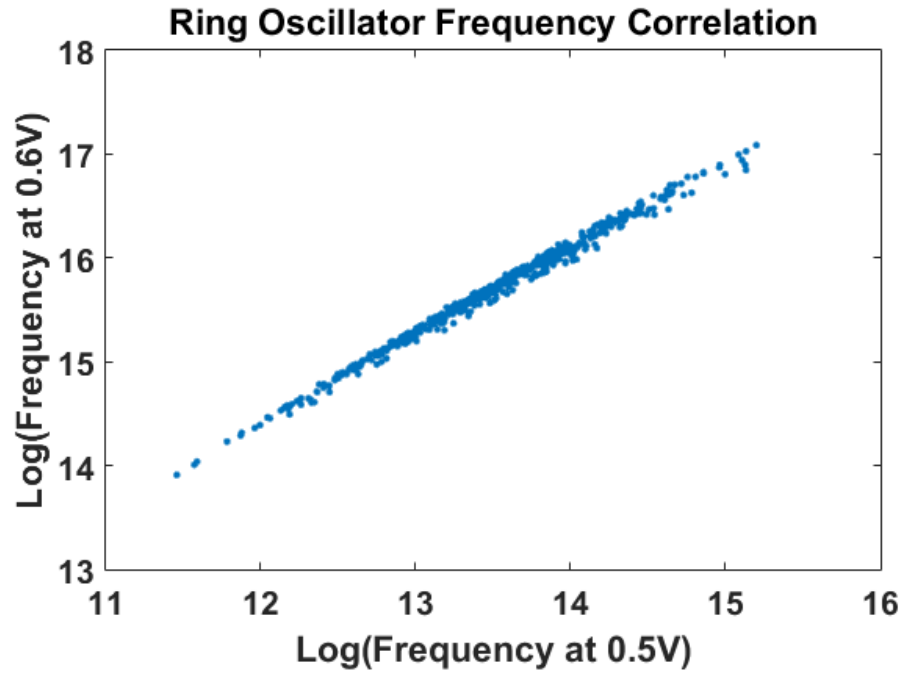


Figure 6.15: Simulated Monte Carlo delay in a ring oscillator operating at 0.5V and 0.6V.

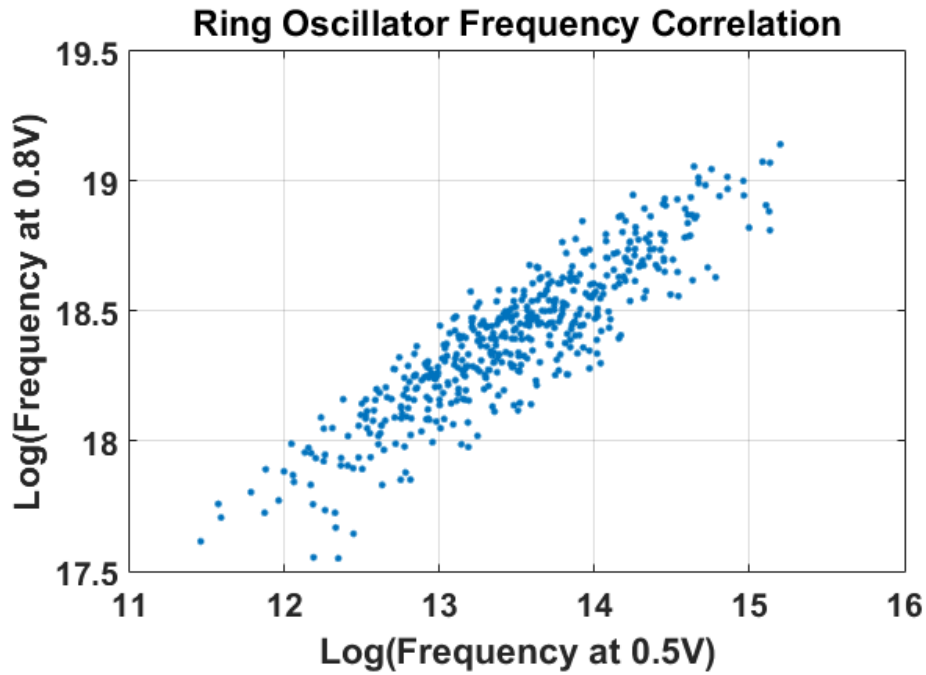


Figure 6.16: Simulated Monte Carlo delay in a ring oscillator operating at 0.5V and 0.8V.

Referring back to the RO example, Figure 6.17 shows the correlation between delays at the subthreshold operating voltage (500 mV) and higher voltages as measured by the r^2 value.

This figure shows the decreasing correlation as the test voltage is increased.

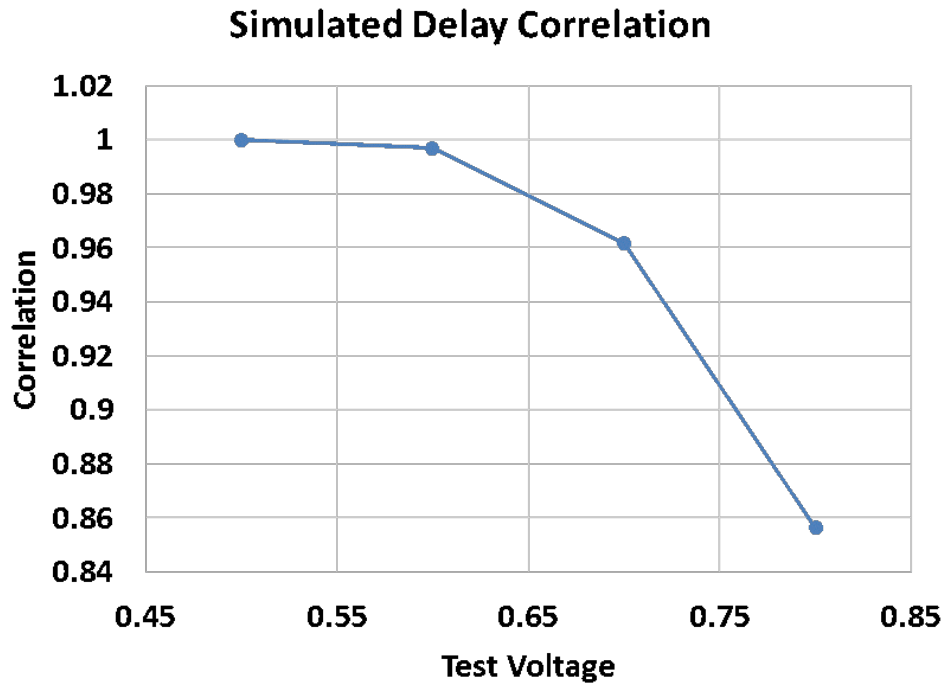


Figure 6.17: Simulated delay correlation in a ring oscillator operating at a range of voltages.

Predictions Based on Voltage Correlations

For each test voltage, a prediction interval is calculated using the test voltage sample delays.

The equation [53]:

$$y \pm t \sqrt{s^2 + xSx^T}$$

Calculates the bounds of the prediction interval and is a function of s^2 - the mean squared error between delays, t - confidence level (99% used as an example in this section), S - covariance matrix, and x - a row vector of the design matrix or Jacobian evaluated at a specified predictor value. Calculating the prediction interval near the point of defined failure creates a window of uncertain failure. In the case of high voltage, high speed testing, these potential failures must be retested. Chips having the prediction interval below the failure

point pass the high speed test and do not need retested. Figure 6.18 shows 99% prediction intervals of the ring oscillator simulation.

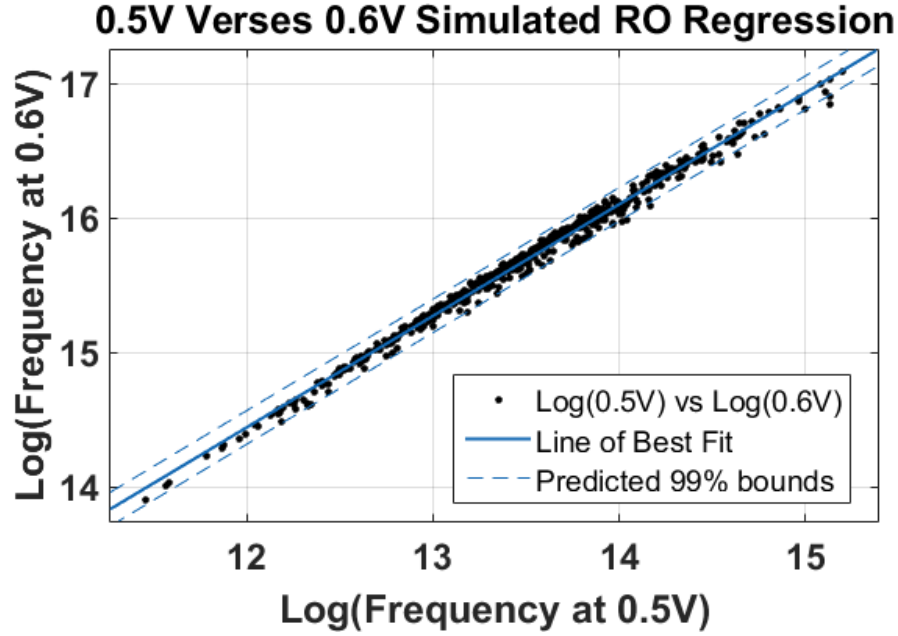


Figure 6.18: Simulated Monte Carlo delay regression for a ring oscillator. The dashed blue lines show the window of prediction for testing at the higher voltage.

Optimization of Test Time

Next, the percentage of chips needing to be retested for a given voltage is found by calculating the z score of the boundary using the mean and standard deviation of that voltages delay. Total test time is calculated using the retest percentage and the average low voltage delay (D_{Vlow}). A line of best fit can be used to interpolate and estimate the optimal test voltage to be used on the population. This can then be verified if necessary by retesting the characterization population at the optimal voltage. Total test time as a percentage of the test time at the designed low voltage $Vlow$ follows the equation

$$T_{Total}\% = \frac{T_{HighSpeed} + T_{Retest}}{D_{Vlow}} \times 100 \quad (6.1)$$

where high speed test time is defined as

$$T_{HighSpeed} = N_{chips} \times D_{Vhigh} \quad (6.2)$$

retest time is defined as

$$T_{Retest} = N_{failed} \times D_{Vlow} \quad (6.3)$$

where D_{Vhigh} and D_{Vlow} are found using the sample set means. T_{Total} is optimized through the trade-offs between $T_{HighSpeed}$ and T_{Retest} . The certainty of the prediction interval (PI) window (shown as 99% above), the number of samples, and strength of correlation determines window width, and as a result, what percentage of chips must be retested. In the case of the ring oscillator, high speed testing at 0.6V was 5× faster, and a 99% PI means 8.5% must be retested at 0.5V, resulting in a total test time of 19.8% when compared to the 0.5V benchmark. Recalculating for all ranges of test voltages results in the plot shown below (Figure 6.19).

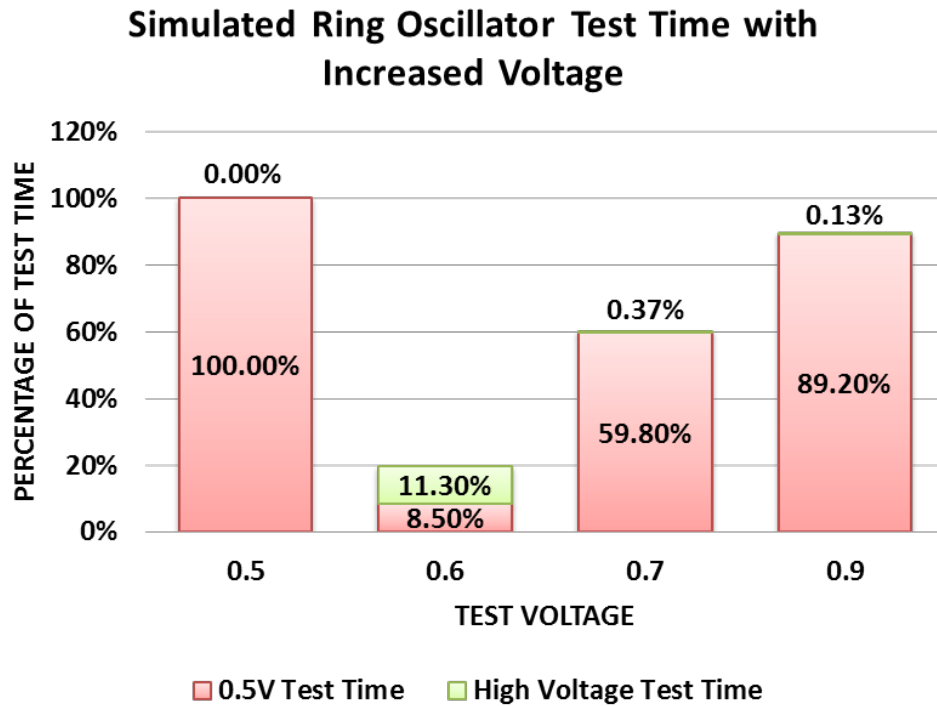


Figure 6.19: Simulated total test time for ring oscillator. Simulated minimum test time is at 0.6V, taking 19.8% of test time at 0.5V.

Using this figure, we can determine the optimal test voltage and its corresponding prediction interval window size. Interpolating between simulated delay voltages allows for an analytical minimization of time. The ring oscillator showed an optimal test time at 0.59V (having approximately the same test time as 0.6V), a $5\times$ savings in subthreshold SoC delay test time.

6.6.8 Implications of Predictive Analysis

This methodology of using trans-threshold correlations to reduce test time can also be used when measuring average power during an operation, or to find correlations in delay changes as a function of temperature change, removing the need to test all chips in a temperature chamber.

6.6.9 Measured Results

Multiple test chips have been fabricated in a 130nm commercial CMOS technology in support of this work. A circuit with combinational and sequential logic consisting of a counter was tested and a low power SRAM was tested. These circuits are representative of the principal circuits in a ULP SoC, and measured results showing the working trans-threshold prediction methodology are conclusive evidence of test time improvement for a commercial subthreshold SoC.

6.6.10 Digital Combinational and Sequential Circuit

The digital circuit is a counter consisting of D flip-flops as well as supporting combinational logic to increment the count each clock cycle. The circuit was tested from 0.3V to 0.6V, along the typical range of subthreshold SoCs. A regression was done at each set of voltages, showing prediction intervals and residuals of the data to analyze the strength of the regression. Figure 6.20 shows a tight window of prediction due to the high correlation in delay time

between 0.3V and 0.4V. The Point of Defined Failure is a user defined failure point (in this case 50kHz), below which chips do not meet the product specification. The Point of Uncertain Failure is the point where predictions may fail at the lower voltage, but due to imperfect correlation, the true value is unknown. Chips tested at 0.4V showing a higher delay must be retested at the operating voltage (0.3V) to ensure they will pass the 50kHz requirement. Residuals are fairly random in nature and do not show any obvious pattern with the exception of higher frequencies being more common.

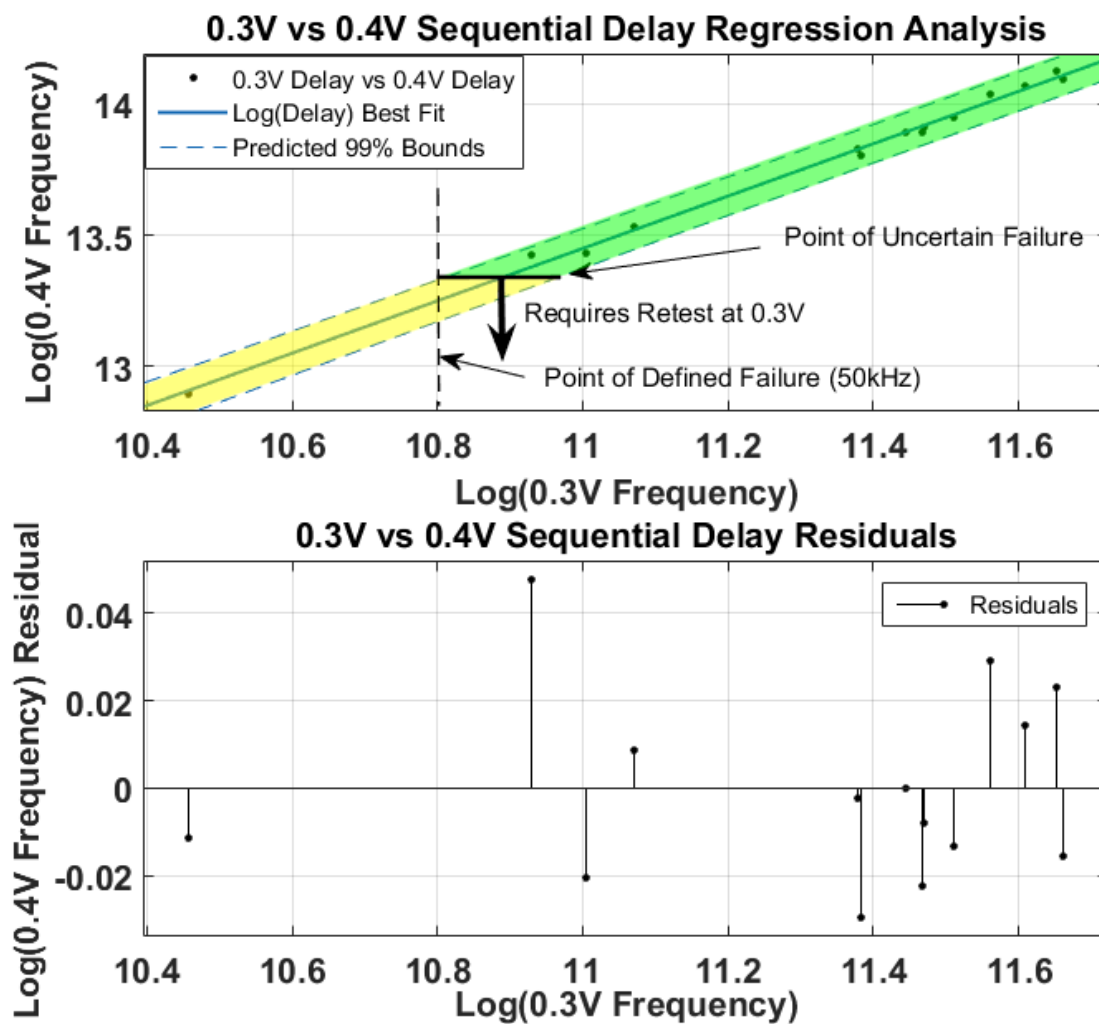


Figure 6.20: Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.4V delay.

Figure 6.21 shows an expanded window of uncertainty, as the correlations between delay

weaken. The Point of Uncertain Failure becomes much farther from the 0.3V, 50 kHz point on the graph. Residuals begin to show less randomness, as error from the mean increases with higher frequency. This is caused by some chips having an average threshold voltage at or below 0.5V, and now super-threshold effects on failure are starting to change the failure mode.

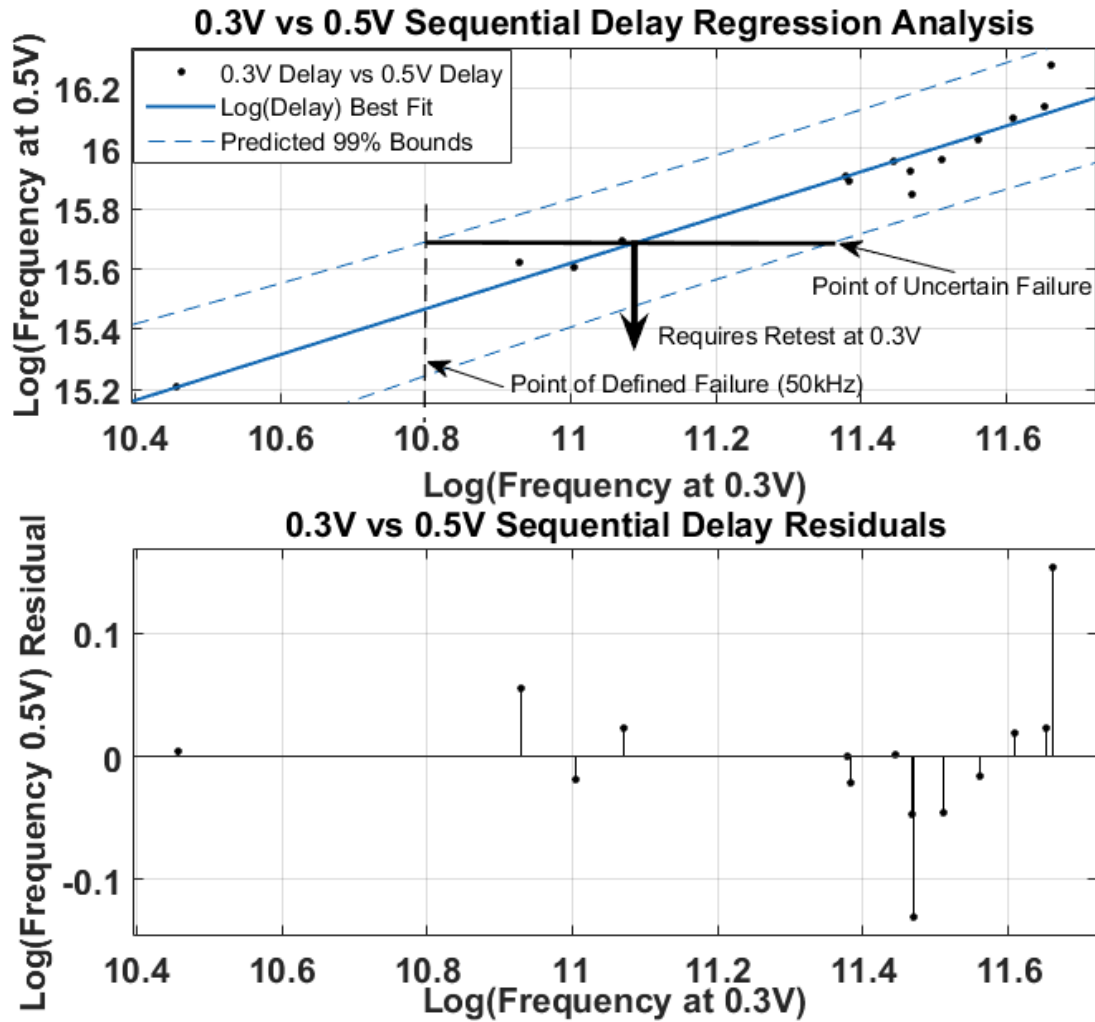


Figure 6.21: Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.5V delay.

Figure 6.22 shows an extreme window of uncertainty, almost to the point where all chips must be retested. The Point of Uncertain Failure is now well beyond the average delay at 0.3V. The residuals show similar patterns to the previous figure, having increased error as frequency increases, due to super-threshold effects.

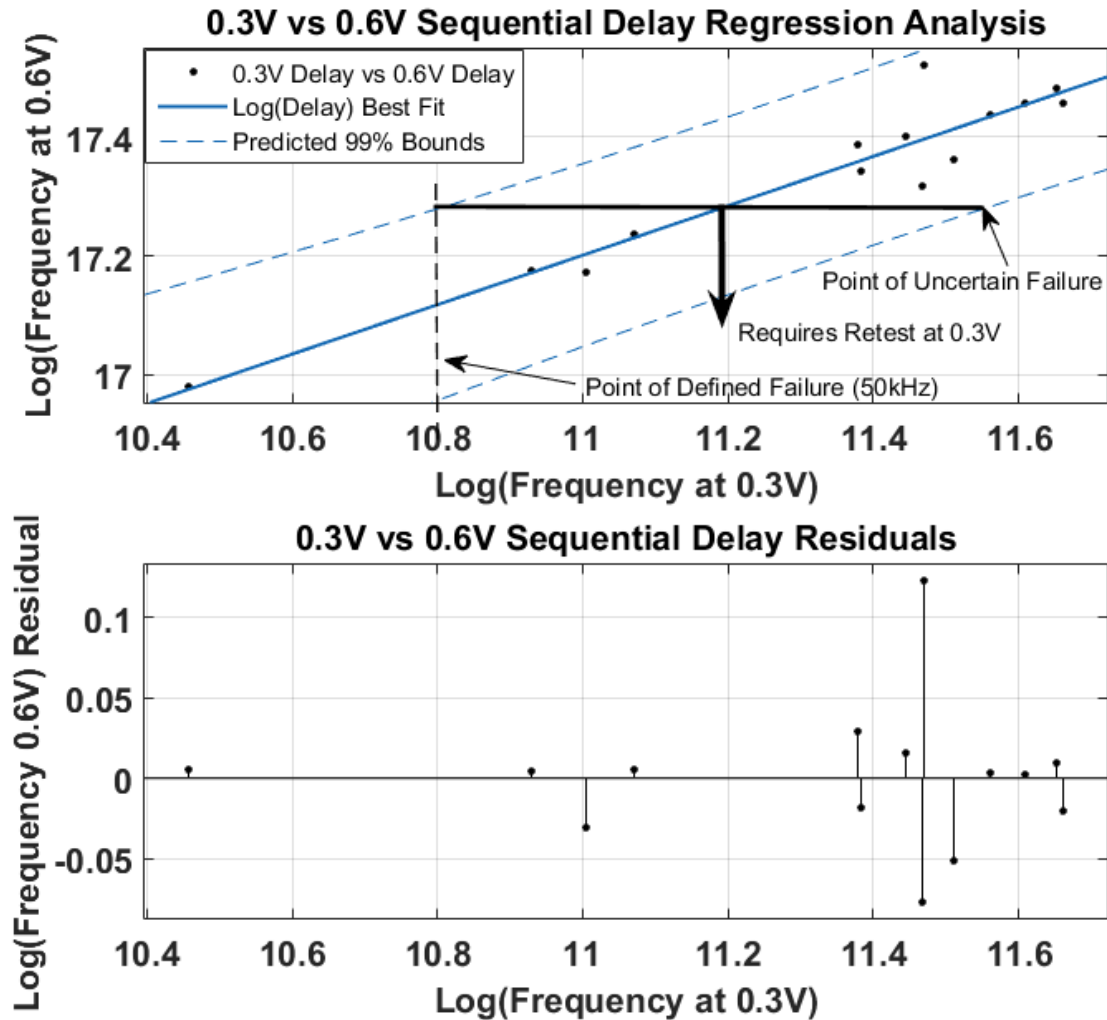


Figure 6.22: Regression analysis and residuals of measured delay in the counter circuit. Comparison is between 0.3V and 0.6V delay.

Figure 6.23 shows the measured results of the test chip correlations in terms of predicted chips needing to be retested at the operating voltage. 0.3V is the base case and will take 100% test time. 0.4V shows a balance of test time between chips tested at the 0.4V and retested chips. As voltage continues to increase, retest time dominates total test time, and the benefits of the statistical analysis no longer continue to improve test time.

Even though Figure 6.23 shows the minimum test time in the plot, it is not yet certain whether total test time has been minimized. Because transistor circuit delay has a first order model [2], expected delay can be interpolated by fitting the model to measurements. Secondly,

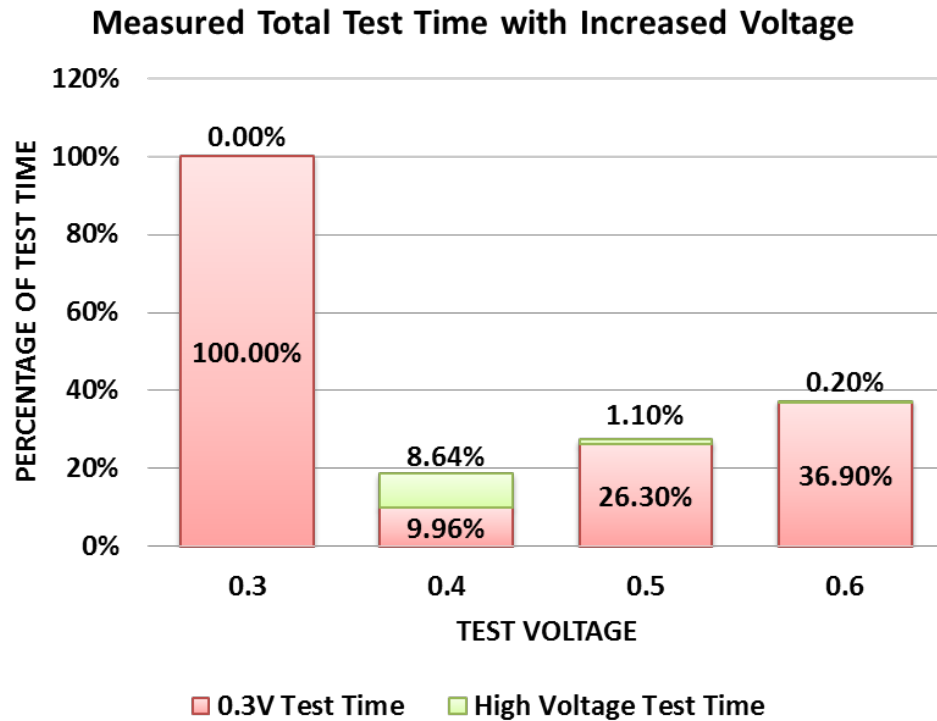


Figure 6.23: Measured total test time for digital logic containing combinational and sequential circuits. Measured minimum test time is at 0.4V, taking 18.6% of test time at 0.3V.

due to a clear pattern in correlation, a model can be created to estimate r^2 at arbitrary voltages near those measured. Because the measured 0.4V delay showed the minimum total test time, values were taken above and below to find a new voltage with potentially lower test time. Figure 6.24 shows test time as a result of the models created from measurement data. The minimum modeled time was found at approximately 0.398V and resulted in a test time of an estimated 18.57%.

Table 6.3 shows the measured failure frequencies corresponding to 50 kHz at 0.3V. r^2 values are shown to decrease as voltage increases. The percentage of chips needing retested increases as the test voltage increases as well due to the increased uncertainty. The total time is calculated using equation 6.1.

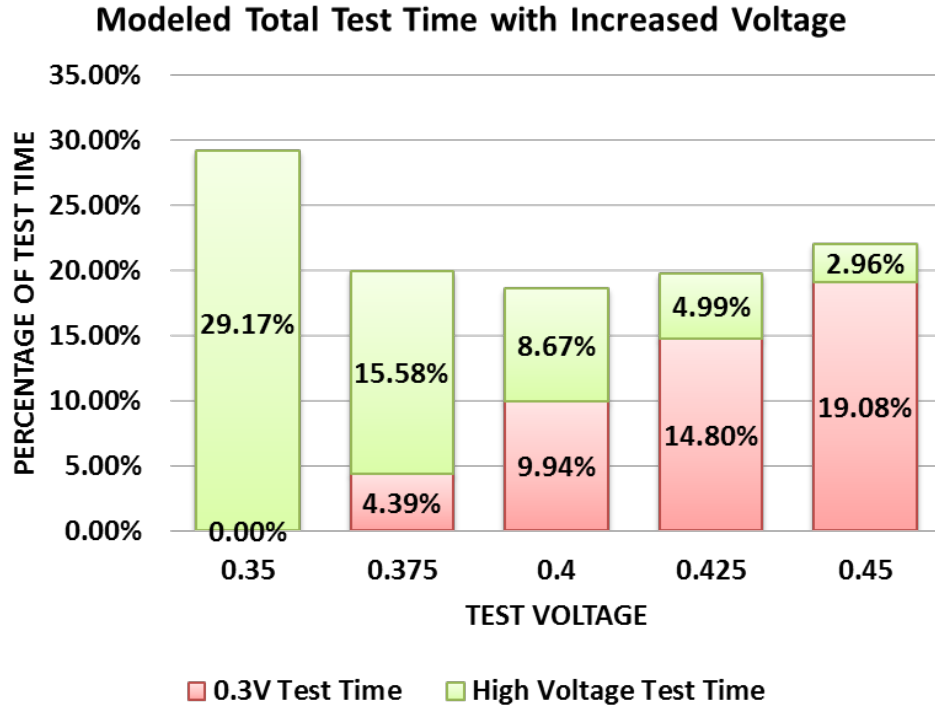


Figure 6.24: Modeled total test time for digital logic containing combinational and sequential circuits. Calculated minimum test time is at 0.398V, taking 18.57% of test time at 0.3V.

Table 6.3: Sequential logic test failure windows, correlation, and test time % data

V	0.3V Point of Uncertain Failure	V_{high} Point of Uncertain Failure	r^2	$\log(\hat{\mu})$	Retest%	V_{high} Imprv.	Total Time
0.3	10.82	-	1	11.32	0	-	100%
0.4	10.89	13.34	0.9958	13.78	9.96%	8.64%	18.6%
0.5	11.10	15.7	0.9457	15.87	26.3%	1.10%	27.4%
0.6	11.20	17.29	0.9084	17.34	36.9%	0.2%	37.1%

6.6.11 High Level Implications of Circuit Level Results

Because combinational and sequential circuits have been verified through this method, the entire digital portion of a subthreshold SoC can be tested using this method as well. For testing at the system level, characterization will be done block by block. Each block can then be tested at its optimal test voltage. Preliminary data shows similar patterns in other types of circuits such as SRAMs, however that is reserved as future work.

Because statistical analysis studies real-world uncertainty, there are certain assumptions

being made during its use. We assume the characterization is statistically strong (contains a large enough sample set), and that some small percentage of chips will fall outside the prediction interval and will falsely pass the test. This uncertainty (in a percentage of false passing chips) can be quantified and is traded off with total test time. Tuning from predicted 99% bounds to a higher number will result in fewer incorrect categorizations, but will take an increased test time.

6.6.12 Trans-threshold Correlations in Other Testing Areas

Similar testing can be done using other metrics as well. Correlations can also be made between operating temperatures. With a high enough correlation, predictions from high voltage testing at room temperature can be made for low voltage delay as a function of an increase or decrease in temperature. These predictions remove the need for temperature chamber testing on all chips designed for use in extreme environments and in automotive applications. Measuring power consumption during characterization at each voltage will create a model allowing similar predictions to be made of the power of a chip operating in subthreshold. Delay measurements can be done during the high voltage tests and the insight can be acquired with no additional test time. Figure 6.25 shows measured correlations for the digital combinational circuit with measured power correlations overlaid and 99% prediction intervals in dotted lines.

6.6.13 Contributions

The SoC project was the result of joint effort by many members of the group. The participants were Farah Yahya, myself, Jacob Breiholz, Abhishek Roy, Harsh Patel, NingXi Liu, Xing Chen, Avish Kosari, Shuo Li, Divya Akella, and Oluseyi Ayorinde. Farah Yahya and I co-managed the project from beginning to end, from deciding the research direction of our next generation system to submitting the GDS CAD file, describing the circuit to the manufacturer. My contributions consisted of:

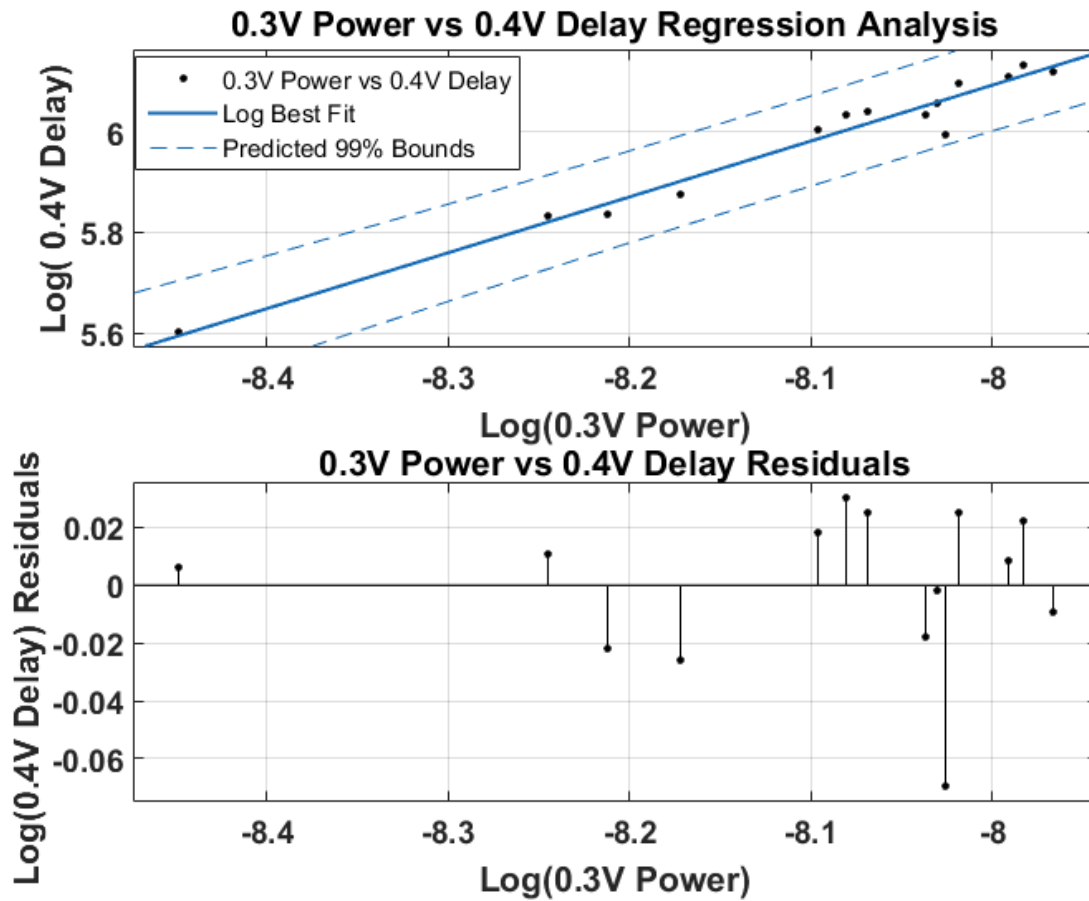


Figure 6.25: Measured power correlations for digital logic containing combinational and sequential circuits. Predictions allow binning of ULP components simply by testing delay at a higher voltage.

1. Defining the specifications of the SoC (with Farah Yahya)
2. Defining the architecture of the SoC (with Farah Yahya)
3. Defining and floorplanning the System in Package (with Farah Yahya)
4. Defining the interfaces between the digital blocks (with Farah Yahya)
5. Defining the power on sequence (with Abishek Roy and Farah Yahya)
6. Designing the NVM custom interface
7. Defining the crystal startup sequence

8. Implementing the ADC (with Farah Yahya)
9. Designing digital, SPI, and GPIO pad designs
10. Implementing wishbone bus, master and slave logic wrappers
11. Designing the ULP Cold Boot Bus
12. Designing Dedicated Serial Bus
13. Defining the radio interface (With Farah Yahya and Xing Chen)
14. Verifying the digital blocks and SoC (With Farah Yahya and Jacob)
15. Designing backend toolflow for digital blocks and bus
16. Integrating the analog and digital blocks in the SoC (with Farah Yahya)
17. Testing the fabricated system (with the entire team)

During the SoC design process, two tapeouts took place to fabricate circuits:

1. November 2015: Digital SoC fabrication to verify toolflow and functionality at low voltages
2. August 2016: Full SoC fabrication

The positioning project was a joint effort with Farah Yahya and led by myself. My contributions consisted of:

1. Defining the project goals
2. Defining the software specification
3. Defining the algorithm for finding position
4. Writing the software (With Farah Yahya)

5. Defining the test process
6. Testing the system

My contributions to the trans-threshold correlation modeling project consisted of:

1. Defining the project goals
2. Designing a methodology for characterizing chips
3. Simulating sample circuit for proof-of-concept
4. Optimizing test time from predictions
5. Quantifying uncertainty in statistical analysis
6. Finding statistically significant patterns in metrics beyond delay
7. Characterizing the fabricated circuits

During the trans-threshold correlation modeling project, a tapeout took place to fabricate circuits:

1. February 2016: Basic logic circuits fabricated to test correlations in delay over a voltage range

6.6.14 Conclusion

This chapter presented a battery-less, self-powered SoC with integrated platform level power management. The sub-system components have been co-optimized to improve data-flow, reducing the active time of the blocks and increasing time spent in low power and stall modes. This SoC was integrated with a cold-boot NVM through the custom CBMS and a wireless radio through its dedicated serial bus to create a self-sufficient SiP. To the best of the authors knowledge, this is the first sub- μ W SoC with this level of integration, and the first

self-powered system with a programmable, non-volatile memory for cold-booting instructions and critical data backup.

Using the system created in this chapter, we present a proof-of-concept $1.8 \mu\text{W}$ self-powered positioning system for navigation in applications where use of a radio based system is not possible. This system can be used in addition to a GPS system to reduce average power or in cases where accuracy is less important or position measurements are cyclical.

We then presented a methodology for reducing test time in subthreshold systems like the one presented in this chapter. The method models trans-threshold correlations to allow high, nominal voltage testing at high speeds and accurately predicting delay and power at the low, subthreshold operational voltage. Using this process has resulted in over 80% measured savings in time for sequential and combinational test circuits with no overhead to area or prefabrication design time.

Chapter 7

Conclusion

The work presented in this document will contribute to the field of electrical engineering by enabling smarter next generation computer systems, which will have an improved ability to communicate with the outside world while remaining self-powered. Reducing the power of these systems will enable a higher quality of service in more varying harvesting conditions. Managing the system power budget will also allow for addition of other components and communication interfaces to enable higher levels of system integration. This more sophisticated platform will be more robust as a result. The demonstration of applications for this platform and improvement in the post-manufacturing test process for subthreshold circuits will reduce time to market and motivate use in the commercial space.

Having lower power, more robust systems with a clear application space will push widespread expansion of self-powered IoT devices into the commercial space. The improved quality of life given by these devices through an increase in personalized medical care and an abundance of sensor data will change how our society works and interacts. It will enable academic research into fields previously unrelated with circuit and computer system design by providing new tools and resources. The IoT will have a broad and profound impact on the economy and how businesses operate.

7.1 Contributions

This dissertation presents many contributions to self-powered systems. We categorize and summarize them in this section.

Chip-to-chip Serial Link for ULP Systems

1. A methodology was presented to Co-optimize energy and power
2. A ULP low-throughput transceiver was designed to operate between 0.2V and 0.7V with reduced power and energy consumption following the proposed methodology
3. Error correction coding was evaluated for potentially improving energy
4. A chip-to-chip link was fabricated in 130 nm technology to demonstrate ULP energy efficient circuits in the low throughput space
5. The proposed transceiver can communicate data at a minimum of 1.24 nW or 0.38 pJ/bit at a throughput of 3.25 kbps. It can also communicate data at a throughput of 10 Mbps but will consume higher power (50.6 μ W)

Differential Link for Reliable Wireline Communication in ULP On-Body Sensor Networks

1. A differential transceiver was introduced to enable ULP systems to communicate over long distance and in noisy environments
2. Low leakage techniques were introduced to reduce the circuit's minimum energy point voltage, resulting in a lower power, more energy efficient design
3. A low swing transmitter with no static current was introduced
4. An ULP receiver was introduced, consuming less than 1 nW active power
5. A contention free SR latch was introduced for latching data on the receiver
6. The transceiver was demonstrated in a flexible textile interconnect

7. Circuits were co-designed for use over a wide range of voltages
8. The transceiver was fabricated in 130 nm technology to demonstrate a long distance link for IoT networks
9. The proposed transceiver consumed 3.77 nW active power and its efficiency was 11.4 fJ/bit/mm

Asymmetric Bus Enabling Cold-Boot Self-Powered Systems in Package

1. An on-package bus interface is introduced that is specialized for booting an SoC from NVM
2. The proposed bus allows the SoC and NVM to be integrated into a single package reducing the power needed to transmit data between the two chips.
3. A binary weighted driver was introduced for reduced swing transmissions to reduce energy and power
4. A sense amp based receiver was introduced to amplify the reduced voltage swing transmission for use in CMOS logic
5. Bootup time is reduced through parallelization of program data transmission, reducing on-time of NVM
6. Transceivers were fabricated in two 130 nm technologies to demonstrate a bootable system
7. The proposed transceiver consumed 17.96 nW active power and its efficiency was 1.27 fJ/b/mm

ULP Flexible On-chip Bus Enabling GALS and DVFS in Self-Powered Systems

1. An architecture is presented to enable finer tuned power and energy optimization in ULP SoCs

2. An SR latch is used in place of D flip flops to enable storage and level shifting up and down in voltage
3. The proposed architecture enables the SoC to operate in one of three power modes are presented: a ULP mode, an MEP mode, and a HP-LP mode
4. An SoC was fabricated in a 130 nm technology to demonstrate the architecture's use in larger systems

Sub μ W System On Chip

1. A battery-less system on chip was introduced for ULP sensing applications
2. The system used co-optimization of sub-system components to improve system power
3. The system was designed for integration with an NVM and radio
4. The system was fabricated in a 130 nm process to demonstrate a proof-of-concept
5. The system contained an on-package cold boot management system for recovery upon power loss events
6. System was integrated into a System in Package to demonstrate higher levels of integration
7. The SoC consumes 507nW when running a shipping-integrity tracking algorithm. This power breakdown includes communication interfaces (SPI, GPIO, and Radio), processing units (timer, compression, and LPC), and storage (IMEM).

Self-powered, Battery-less Positioning System

1. A proof-of-concept self-powered positioning system was developed
2. A low computational cost algorithm was implemented to track position from acceleration data

Modeling Trans-threshold Correlations for Reducing Functional Test Time in ULP Systems

1. A test methodology was presented for reducing time to market in subthreshold SoCs
2. A correlation in delay was presented between varying voltages
3. A prediction interval was calculated using measured logic delay
4. An optimal voltage was found for reducing test time over 80%
5. Implications of correlations on power measurements were explored
6. A test chip was fabricated in 130 nm to demonstrate measurements of delay

7.2 Impact

The off-chip communication circuit contributions in this thesis will completely change the system power breakdown in future systems utilizing these methods. Lower power will enable larger and smarter networks of self-powered systems. They will also enable wired on-body networks without the typical power overhead of wireless communication. Figure 7.1 shows the effect of the off-chip serial link being utilized for communication in our presented SoC. Off-chip I/O moves from 53% of the system budget to 4%. Like the off-chip communication contributions, the on-package interface will enable a higher level of integration, resulting in a more robust platform. The on-chip flexible bus will improve the system's power budget, potentially allowing even more chips integrated into a single package. The SoC and SiP we developed show how sub-system co-optimization can improve both the level of integration and platform reliability. The SoC application demonstrations and test time reduction methods, along with the power reduction techniques and improvements to robustness will reduce the system's time to market, enabling commercialization of these ULP platforms.

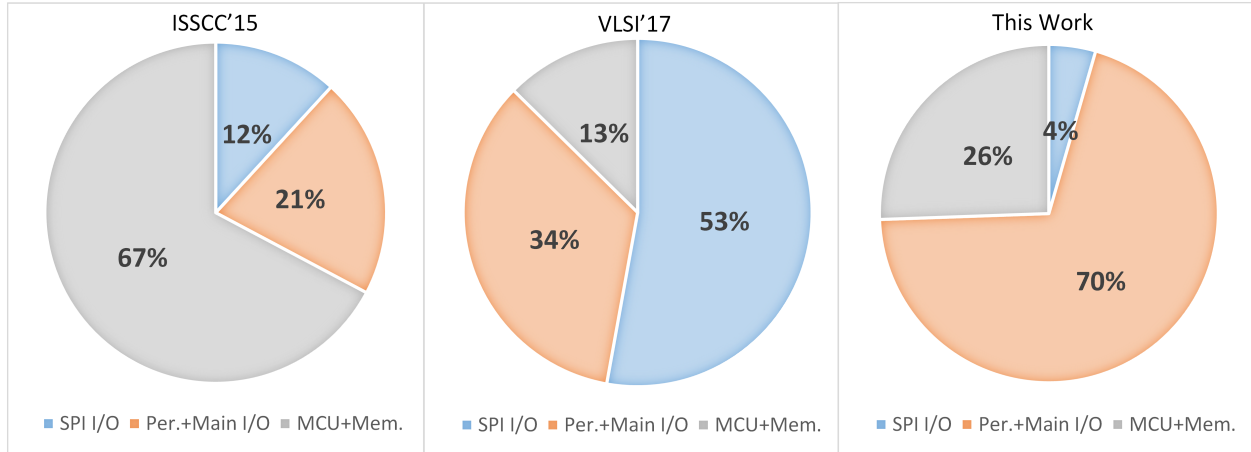


Figure 7.1: Figure showing the impact of adding presented off-chip serial link to SoC. The first chart shows I/O being a large portion of the system power budget. As the system power scales downward, the second figure shows the I/O dominating power consumption. The third figure shows the presented circuits being used for communication between the SoC and a research based, low voltage sensor.

7.3 Future Work

In this section, we will present some of the research opportunities that arise from the work proposed in this dissertation.

7.3.1 Off-chip Communication

Now that ULP, energy efficient designs have been demonstrated, a wide space for new research has opened. Explorations into improving throughput without consuming much more power will enable more data to be delivered in a self-powered network. Network modeling will inform the best trade-offs between power, energy, and throughput for a set of applications. Additionally, creating circuits that are compliant with existing network protocols and standards will be helpful in integrating self-powered systems with existing networks. Integrating media access control with off-chip circuits will enable better identification of individual nodes in a network.

7.3.2 On-package Communication

Similar to off-chip communication, ULP on-package communication has a wide open space for future work. On-package communication can be designed for each specific application, creating an optimal data flow for a given application. A collection of on-package interfaces for IoT will allow modular SoCs to drop in desired peripheral chips in a package. On-package transceiver analog circuits can remain simple, as the capacitance between dice is not currently the bottleneck in energy and power consumption. Having an extender to enable access to the on-chip bus through an on-package interface will allow the SoC controller to interface with other chips in the package using the memory mapped address space, putting control in the hands of the user.

7.3.3 On-chip Communication

On-chip communication circuits can be further integrated with their systems to produce a smarter, more optimal SoC. Having a modular on-chip bus which integrates with on-package communication will allow user-level interfacing between the main controller and peripherals in optimal technologies for the applications while remaining self-powered.

7.3.4 Ultra-low Power Systems on Chip

Future work in ULP SoCs will involve integrating the improved communication and using that to inform the system's data flow for new applications. Having a model of the system's data flow as well as harvesting and power management will better inform the designer of the bottlenecks, creating a more balanced system for a larger design space.

An SoC toolkit can be developed by aggregating many accelerators into a library and reworking them to interface with the on-chip bus and will speed the development flow of many specialized SoCs. If accelerators are grouped together for optimal dataflow, the digital portion of the SoC can be auto-generated without modification when fabricating a chip. A

library of harvesting and power management circuits can be collected and the interfaces can be standardized. This will allow for different analog circuits to be dropped into the system for a desired application. These infrastructure improvements will minimize the overhead in circuits and systems related research.

7.3.5 Applications of Self-powered Internet of Things Systems

The potential design spaces for self-powered systems are numerous, and the exploration has only begun. As new sensors enter the ULP space, SoCs will be able to leverage them to solve new problems. The proof-of-concept self-powered INS creates an open space for new research. Research into improving the sophistication of the algorithm for accuracy and how that relates to ULP systems with lower processing throughput is an important first step. Demonstrating more advanced self-powered sensor fusion techniques using [47] and custom ULP digital accelerators to find position will create a robust positioning network without the need for a radio. Lastly, studying the trade-off between creating more compact code on the SoC and accuracy will present a range of solutions that allow the user to optimize accuracy and power for their desired application.

7.3.6 Test Methodologies of subthreshold SoCs

Functional test is only one component of the test process in subthreshold SoCs. Exploration into methods for reducing test time in other steps of the testing process will further reduce the time to market. The results of testing can also be used as a feedback mechanism. Using functionally failed chips to inform the design process can improve yield and improve the weakest link in the system. Additionally, correlations can be found in other, non-digital components of the system to improve test time. Many analog circuits have a wide range of operating voltages, especially communication circuits. Correlations can be found in these circuits as well, speeding up analog test time. Memory circuits (SRAMs) are in most SoCs,

and can benefit from the same method. Read and write delay can be tested at higher voltages, verifying proper functionality in a shorter time.

Appendix A

Publications

A.1 Completed

- [CJL1] **Lukas, C.J.**; Calhoun, B.H., "A 0.38 pJ/bit 1.24 nW Chip-to-Chip Serial Link for Ultra-Low Power Systems", *2015 International Symposium on Circuits and Systems (ISCAS)*, Lisbon, 2015, pp. 2860-2863.
- [CJL2] Roy, A.; Klinefelter, A.; Yahya, F.B.; **Lukas, C.J.**; et al., "A 6.45W Self-Powered IoT SoC with Integrated Energy-Harvesting Power Management and ULP Asymmetric Radios for Portable Biomedical Systems" *IEEE Transactions on Biomedical Circuits and Systems* (TBioCaS), vol. 9, no. 6, pp. 862-874, Dec. 2015.
- [CJL3] Yahya, F.B.; **Lukas, C.J.**; et al. "A Battery-less 507nW SoC with Integrated Platform Power Manager and SiP Interfaces", **accepted to VLSI 2017**.

A.2 Submitted

- [CJL4] Yahya, F.B.; **Lukas, C.J.**; et al. "FAR: A 4.12W Ferro-electric Auto-Recovery Subsystem for Battery-less SoCs", **submitted to ESSCIRC 2017**.

- [CJL5] **Lukas, C.J.**; et al. "A Cold-Boot Management System Featuring an 8.49 nW, 1.09 fJ/b/mm Asymmetric Bus Enabling Self-Powered Systems in Package", **submitted to ESSCIRC 2017**.
- [CJL6] **Lukas, C.J.**; et al. "Modeling Trans-threshold Correlations for Reducing Functional Test Time in Ultra-low Power Systems", **submitted to ITC 2017**.
- [CJL7] **Lukas, C.J.**; et al. "A 3.77 nW, 11.4 fJ/b/mm Link for Reliable Wireline Communication in Ultra-Low Power On-Body Sensor Networks", **submitted to ESSCIRC 2017**.
- [CJL8] Breiholz, J; Yahya; F.B.; **Lukas, C.J.**; et al. "A 4.4 nW Lossless Sensor Data Compression Accelerator for 2.9x System Power Reduction in Wireless Body Sensors" **invited to MWSCAS 2017**.

A.3 Planned

- [CJL9] Journal expansion of battery-less SoC paper - planned for JSSC.
- [CJL10] Paper on BLE transmitter with University of Michigan titled "A 724uW BLE Compatible FSK Transmitter for Wireless Beacon Applications".
- [CJL11] Paper on ultra low power voice activity detection - planned for BioCaS 2017.
- [CJL12] Paper on Integration of SoC, FeAR and BLE into one package.
- [CJL13] Paper on flexible on-chip bus enabling GALS and DVFS in self-powered Systems - planned for ISSCC 2018.
- [CJL14] Paper on self-powered positioning system - planned for BioCaS 2017.
- [CJL15] Paper on modeling and sizing for energy efficient ultra-low power CMOS I/O circuits.
- [CJL16] Paper on ULP SAR ADC - planned for JLPEA.

[CJL17] Paper on ULP BLE receiver.

Appendix B

Acronyms

ADC analog-to-digital-converter

AFE analog front end

AFIB atrial fibrillation

ALU arithmetic logic unit

CBB cold-boot bus

CBC cold-boot controller

CBMS cold-boot management system

CMOS complementary metal oxide semiconductor

COTS commercial off-the-shelf

CU control unit

D2D die-to-die

DMA direct memory access

DMEM data memory

DMU data management unit

DVFS dynamic voltage and frequency scaling

ECC error correction coding

ECG electrocardiography

EOF End Of File

EH-PPM energy harvesting platform power manager

FeAR ferroelectric-based auto-recovery system

Fe-RAM ferroelectric random access memory

FIFO first in first out

FIR finite impulse response

FSK frequency shift keying

GPP general purpose processor

GPS global positioning system

GPIO general purpose input output

I²C inter integrated circuit

IMEM instruction memory

I/O input / output

IoT internet of things

LPC low power controller

MAC multiply accumulate

MC Monte Carlo

MEP minimum energy point

PCB printed circuit board

PLL phase-locked loop

PM power monitor

POR power on reset

PV photovoltaic

RAM random access memory

ROM read-only memory

RR heart rate interval between two R peaks

SAR successive approximation

SiP system-in-package

SoC system-on-chip

SOR system power on reset

SPI serial peripheral interface

SRAM static random access memory

TEG thermoelectric generator

UART Universal asynchronous receiver/transmitter

ULP ultra-low power

ULP-BUS ultra-low power bus

V_{CAP} super-capacitor voltage

VCO voltage controlled oscillator

V_{DD} supply voltage

V_{MAX} maximum voltage of technology

V_{MIN} minimum operating voltage

V_{REF} reference voltage

V_T transistor threshold voltage

V_{SS} ground voltage

VVSS virtual ground voltage

XOR crystal power on reset

Bibliography

- [1] A.G. Pandolfo and A.F. Hollenkamp. Carbon properties and their role in supercapacitors. *Journal of Power Sources*, 157(1):11 – 27, 2006.
- [2] B.H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *Solid-State Circuits, IEEE Journal of*, 40(9):1778–1786, Sept 2005.
- [3] W. M. Holt. 1.1 moore’s law: A path going forward. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 8–13, Jan 2016.
- [4] R. Gonzalez, B. M. Gordon, and M. A. Horowitz. Supply and threshold voltage scaling for low power cmos. *IEEE Journal of Solid-State Circuits*, 32(8):1210–1216, Aug 1997.
- [5] Yanqing Zhang, Fan Zhang, Y. Shakhshier, J.D. Silver, A. Klinefelter, M. Nagaraju, J. Boley, J. Pandey, A. Shrivastava, E.J. Carlson, A. Wood, B.H. Calhoun, and B.P. Otis. A batteryless 19 uw mics/ism-band energy harvesting body sensor node soc for exg applications. *Solid-State Circuits, IEEE Journal of*, 48(1):199–213, Jan 2013.
- [6] *AMBA Specification*. ARM, 2 edition, 1999.
- [7] *WISHBONE System-on-Chip (SoC) Interconnection Architecture for Portable IP Cores*. Open Cores, b.3 edition, 2002.
- [8] R. Mahajan, R. Sankman, N. Patel, D. W. Kim, K. Aygun, Z. Qian, Y. Mekonnen, I. Salama, S. Sharan, D. Iyengar, and D. Mallik. Embedded multi-die interconnect bridge (emib) – a high density, high bandwidth packaging interconnect. In *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, pages 557–565, May 2016.
- [9] A. Shrivastava, J. Pandey, B. Otis, and B. H. Calhoun. A 50nw, 100kbps clock/data recovery circuit in an fsk rf receiver on a body sensor node. In *2013 26th International Conference on VLSI Design and 2013 12th International Conference on Embedded Systems*, pages 72–75, Jan 2013.
- [10] Atmel. *AT25M02 Datasheet*, 2017. Rev. 8832C.
- [11] Analog Devices. *ADXL362: Micropower, 3-Axis, 2 g/4 g/8 g Digital Output MEMS Accelerometer*, 2012. Rev. F.

- [12] Motorola Inc. *SPI Block Guide*, 2003. V03.06.
- [13] NXP Semiconductors. *I2C-bus specification and user manual*, 2014. Rev. 6.
- [14] Texas Instruments. *KeyStone Architecture Universal Asynchronous Receiver/Transmitter (UART)*, 2010. Rev. 1.0.
- [15] PCI SIG. *PCI Express Base Specification*, 2010. Rev. 3.0.
- [16] IEEE. *IEEE Standard for Ethernet*, 2015. Rev. IEEE Std 802.3bm-2015.
- [17] Digital Display Working Group. *Digital Visual Interface (DVI)*, 1999. Rev. 1.0.
- [18] HDMI Licensing, LLC. *High-Definition Multimedia Interface*, 2006. Rev. 1.3a.
- [19] *Universal Serial Bus 3.1 Specification*, 2013. Rev. 1.0.
- [20] ARM Limited. *AMBA APB Protocol Specification*, 2008. Rev. 2.
- [21] Qualcomm Snapdragon 600E Processor APQ8064E Data Sheet. *Qualcomm Snapdragon 600E Processor*, 2016. Rev. C.
- [22] Weste and Harris. *CMOS VLSI Design, A Circuits and Systems Perspective*. Addison Wesley, 2011.
- [23] S. Al-Arian and D. Agrawal. Physical failures and fault models of cmos circuits. *IEEE Transactions on Circuits and Systems*, 34(3):269–279, Mar 1987.
- [24] D. G. Edwards. Testing for mos ic failure modes. *IEEE Transactions on Reliability*, R-31(1):9–18, April 1982.
- [25] A. Klinefelter, N.E. Roberts, Y. Shakhsher, P. Gonzalez, A. Shrivastava, A. Roy, K. Craig, M. Faisal, J. Boley, Seunghyun Oh, Yanqing Zhang, D. Akella, D.D. Wentzloff, and B.H. Calhoun. 21.3 a 6.45 uw self-powered iot soc with integrated energy-harvesting power management and ulp asymmetric radios. In *Solid-State Circuits Conference - (ISSCC), 2015 IEEE International*, pages 1–3, Feb 2015.
- [26] A. Roy, A. Klinefelter, F. B. Yahya, X. Chen, L. P. Gonzalez-Guerrero, C. J. Lukas, D. A. Kamakshi, J. Boley, K. Craig, M. Faisal, S. Oh, N. E. Roberts, Y. Shakhsher, A. Shrivastava, D. P. Vasudevan, D. D. Wentzloff, and B. H. Calhoun. A 6.45 uW self-powered SoC with integrated energy-harvesting power management and ULP asymmetric radios for portable biomedical systems. *IEEE Transactions on Biomedical Circuits and Systems*, 9(6):862–874, Dec 2015.
- [27] F. B. Yahya, H. N. Patel, V. Chandra, and B. H. Calhoun. Combined sram read/write assist techniques for near/sub-threshold voltage operation. In *2015 6th Asia Symposium on Quality Electronic Design (ASQED)*, pages 1–6, Aug 2015.

- [28] P. Harpe, Hao Gao, R. van Dommele, E. Cantatore, and A. van Roermund. 21.2 a 3nw signal-acquisition ic integrating an amplifier with 2.1 nef and a 1.5fj/conv-step adc. In *Solid-State Circuits Conference - (ISSCC), 2015 IEEE International*, pages 1–3, Feb 2015.
- [29] J.W. Poulton, W.J. Dally, X. Chen, J.G. Eyles, T.H. Greer, S.G. Tell, J.M. Wilson, and C.T. Gray. A 0.54 pj/b 20 gb/s ground-referenced single-ended short-reach serial link in 28 nm cmos for advanced packaging applications. *Solid-State Circuits, IEEE Journal of*, 48(12):3206–3218, Dec 2013.
- [30] M. Babaie, F. W. Kuo, H. N. R. Chen, L. C. Cho, C. P. Jou, F. L. Hsueh, M. Shahmohammadi, and R. B. Staszewski. A fully integrated bluetooth low-energy transmitter in 28 nm cmos with 36 *IEEE Journal of Solid-State Circuits*, 51(7):1547–1565, July 2016.
- [31] M. A. Yokus, R. Foote, and J. S. Jur. Printed stretchable interconnects for smart garments: Design, fabrication, and characterization. *IEEE Sensors Journal*, 16(22):7967–7976, Nov 2016.
- [32] W. S. Choi, G. Shu, M. Talegaonkar, Y. Liu, D. Wei, L. Benini, and P. K. Hanumolu. 3.8 a 0.45-to-0.7v 1-to-6gb/s 0.29-to-0.58pj/b source-synchronous transceiver using automatic phase calibration in 65nm cmos. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, Feb 2015.
- [33] T. Zhang, K. Wang, Y. Feng, Y. Chen, Q. Li, B. Shao, J. Xie, X. Song, L. Duan, Y. Xie, X. Cheng, and Y. L. Lin. A 3d soc design for h.264 application with on-chip dram stacking. In *2010 IEEE International 3D Systems Integration Conference (3DIC)*, pages 1–6, Nov 2010.
- [34] Ye-Sheng Kuo, P. Pannuto, Gyouho Kim, ZhiYoong Foo, Inhee Lee, B. Kempke, P. Dutta, D. Blaauw, and Yoonmyung Lee. Mbus: A 17.5 pj/bit/chip portable interconnect bus for millimeter-scale sensor systems with 8 nw standby power. In *Custom Integrated Circuits Conference (CICC), 2014 IEEE Proceedings of the*, pages 1–4, Sept 2014.
- [35] M. Mansuri, J. E. Jaussi, J. T. Kennedy, T. C. Hsueh, S. Shekhar, G. Balamurugan, F. O’Mahony, C. Roberts, R. Mooney, and B. Casper. A scalable 0.128-to-1tb/s 0.8-to-2.6pj/b 64-lane parallel i/o in 32nm cmos. In *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pages 402–403, Feb 2013.
- [36] K. Craig, Y. Shakhsheer, S. Arrabi, S. Khanna, J. Lach, and B. H. Calhoun. A 32 b 90 nm processor implementing panoptic dvs achieving energy efficient operation from sub-threshold to high performance. *IEEE Journal of Solid-State Circuits*, 49(2):545–552, Feb 2014.
- [37] M.R. Stan and W.P. Burleson. Bus-invert coding for low-power i/o. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 3(1):49–58, March 1995.

- [38] J. M. Wilson, M. R. Fojtik, J. W. Poulton, X. Chen, S. G. Tell, T. H. Greer, C. T. Gray, and W. J. Dally. 8.6 a 6.5-to-23.3fj/b/mm balanced charge-recycling bus in 16nm finfet cmos at 1.7-to-2.6gb/s/wire with clock forwarding and low-crosstalk contraflow wiring. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 156–157, Jan 2016.
- [39] Skytraq venusflpx product brief. "http://www.skytraq.com.tw/products/Venus838FLPx_PB_v1.pdf", Feb 2017.
- [40] Copernicus ii gps receiver reference manual. "http://trl.trimble.com/dscgi/ds.py/Get/File-499737/CopernicusIIRefMan_1A_68340-06-ENGa.pdf", July 2011.
- [41] Hyejung Kim, Sunyoung Kim, N. Van Helleputte, A. Artes, M. Konijnenburg, J. Huisken, C. Van Hoof, and R.F. Yazicioglu. A configurable and low-power mixed signal soc for portable ecg monitoring applications. *Biomedical Circuits and Systems, IEEE Transactions on*, 8(2):257–267, April 2014.
- [42] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Gutttag, and A.P. Chandrakasan. A micro-power eeg acquisition soc with integrated feature extraction processor for a chronic seizure detection system. *Solid-State Circuits, IEEE Journal of*, 45(4):804–816, April 2010.
- [43] S. Rai, J. Holleman, J.N. Pandey, F. Zhang, and B. Otis. A 500 uw neural tag with 2 uvrms afe and frequency-multiplying mics/ism fsk transmitter. In *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 212–213, 213a, Feb 2009.
- [44] G. Chen, M. Fojtik, Daeyeon Kim, D. Fick, Junsun Park, Mingoo Seok, Mao-Ter Chen, Zhiyoong Foo, D. Sylvester, and D. Blaauw. Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pages 288–289, Feb 2010.
- [45] N. Van Helleputte, M. Konijnenburg, Hyejung Kim, J. Pettine, Dong-Woo Jee, A. Breeschoten, A. Morgado, T. Torfs, H. de Groot, C. Van Hoof, and R.F. Yazicioglu. 18.3 a multi-parameter signal-acquisition soc for connected personal health applications. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pages 314–315, Feb 2014.
- [46] Dongsuk Jeon, Yen-Po Chen, Yoonmyung Lee, Yejoong Kim, Zhiyoong Foo, G. Kruger, H. Oral, O. Berenfeld, Zhengya Zhang, D. Blaauw, and D. Sylvester. 24.3 an implantable 64nw ecg-monitoring mixed-signal soc for arrhythmia diagnosis. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pages 416–417, Feb 2014.
- [47] D.A. Grejner-Brzezinska, C.K. Toth, L. Li, J. Park, X. Wang, H. Sun, I.J. Gupta, K. Huggins, and Y.F. Zheng. Positioning in gps-challenged environments: Dynamic

- sensor network with distributed gps aperture and inter-nodal ranging signals. In *Proceedings of the 22nd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2009)*, pages 111–123, September 2009.
- [48] Mohinder S. Grewal, Lawrence R. Weill, and Angus P. Andrews. *Global Positioning Systems, Inertial Navigation, and Integration*. Wiley-Interscience, 2007.
 - [49] Hong Hao and E.J. McCluskey. Very-low-voltage testing for weak cmos logic ics. In *Test Conference, 1993. Proceedings., International*, pages 275–284, Oct 1993.
 - [50] S.A. Bota, M. Rosales, J.L. Rossello, and J. Segura. Low v_{dd} vs. delay: is it really a good correlation metric for nanometer ics? In *VLSI Test Symposium, 2006. Proceedings. 24th IEEE*, pages 6 pp.–363, April 2006.
 - [51] J. Myers, A. Savanth, D. Howard, R. Gaddh, P. Prabhat, and D. Flynn. 8.1 an 80nw retention 11.7pj/cycle active subthreshold arm cortex-m0+ subsystem in 65nm cmos for wsn applications. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, Feb 2015.
 - [52] W. Lim, I. Lee, D. Sylvester, and D. Blaauw. 8.2 batteryless sub-nw cortex-m0+ processor with dynamic leakage-suppression logic. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, Feb 2015.
 - [53] Confidence and prediction bounds - matlab & simulink. ”<https://www.mathworks.com/help/curvefit/confidence-and-prediction-bounds.html>”, Feb 2017.
 - [54] A. Burdett. Ultra-low-power wireless systems: Energy-efficient radios for the internet of things. *Solid-State Circuits Magazine, IEEE*, 7(2):18–28, Spring 2015.
 - [55] N.E. Roberts and D.D. Wentzloff. A 98nw wake-up radio for wireless body area networks. In *Radio Frequency Integrated Circuits Symposium (RFIC), 2012 IEEE*, pages 373–376, June 2012.
 - [56] H. Cho, H. Kim, M. Kim, J. Jang, Y. Lee, K. J. Lee, J. Bae, and H. J. Yoo. A 79 pj/b 80 mb/s full-duplex transceiver and a 42.5 uW 100 kb/s super-regenerative transceiver for body channel communication. *IEEE Journal of Solid-State Circuits*, 51(1):310–317, Jan 2016.
 - [57] T. Norimatsu, T. Kawamoto, K. Kogo, N. Kohmu, F. Yuki, N. Nakajima, T. Muto, J. Nasu, T. Komori, H. Koba, T. Usugi, T. Hokari, T. Kawamata, Y. Ito, S. Umai, M. Tsuge, T. Yamashita, M. Hasegawa, and K. Higeta. 3.3 a 25gb/s multistandard serial link transceiver for 50db-loss copper cable in 28nm cmos. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 60–61, Jan 2016.
 - [58] D. Jeon, Y. P. Chen, Y. Lee, Y. Kim, Z. Foo, G. Kruger, H. Oral, O. Berenfeld, Z. Zhang, D. Blaauw, and D. Sylvester. 24.3 an implantable 64nw ecg-monitoring mixed-signal soc for arrhythmia diagnosis. In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 416–417, Feb 2014.