Digital Baseband Techniques and System Modeling Considerations for Low-Power Duty-Cycled Wake-up Receivers

---

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

by

Henry Lansford Bishop

August 2021

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Henry Lansford Bishop

This Dissertation has been read and approved by the examing committee:

Advisor: Benton H. Calhoun

Advisor:

Committee Member: Steven M. Bowers

Committee Member: N. Scott Barker

Committee Member: Barry W. Johnson

Committee Member: Bradford Campbell

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

August 2021

ABSTRACT

The Internet-of-Things (IoT) has the potential to revolutionize our understanding and real-time awareness about the world around us in order to make our lives simpler, safer, and more efficient. However, IoT growth depends strongly on how cost effective the solutions are. Battery-based systems may suffer from limited lifetimes, which incur regular maintenance, replacement, and downtime costs. Energy-harvesting based systems can suffer from poor dependability when environmental sources supply insufficient power, which leads to poor application quality. Generally, lifetime and dependability both increase when system power consumption is lowered. Thus, to achieve global IoT scale beyond the 10s of billions of devices, dominating factors of system power consumption must be addressed.

Many edge IoT devices perform events with relatively low-activity factors, such that their dominant source of power consumption stems from components performing essential functions. Given that these systems communicate wirelessly, they must keep their receivers active in order to hear wireless requests even in these low-activity factor scenarios. This energy inefficient idle listening can easily account for a significant portion of the floor power. Wake-up receivers (WuRXs) are a candidate solution to this challenge by acting as energy-efficient wireless notifiers that wake up a system upon reception of a specific wireless code. However, many of the reported WuRXs require excessive power consumption for desired levels of performance. Further, the real power improvement from wake-up receivers, and other components, in an integrated system is difficult to predict without the context of system architecture, design options, and operational constraints. This limits the ability of designers to realize energy-efficient IoT systems.

This dissertation addresses WuRX design challenges in three phases. An analysis and comparison are completed on two conventional WuRX duty-cycling techniques as well as the proposed within-packet duty-cycling mechanism, which was developed to further reduce WuRX power consumption. A variety of critical WuRX digital baseband techniques and architectures are presented

with accompanying modeled and measured results. Lastly, the discussed concepts are applied in commercial CMOS on four different wake-up receiver designs achieving state-of-the-art power at near and sub-μW levels, due to the presented duty-cycling methods, along with high-performance in terms of RF sensitivity, wake-up latency, and interference rejection.

This dissertation then addresses low-power IoT system design challenges by proposing a system-level modeling framework based on system architecture, design, and operation information to enable fast virtual prototyping and an understanding of trade-offs across a variety of scenarios. Several experiments concerning the integration of wake-up radios into the system context are conducted with this framework to demonstrate examples of the component's impact.

ACKNOWLEDGMENTS

At the time of this writing, I have been at the University of Virginia for nearly eight years as a student. I started as a first-year who felt anxious, yet excited, to start on this new adventure. Now all these years later I find that I have grown, learned, and experienced so much. Along with all those memories are the people that have crossed my path and who I am indebted to for a great many things.

I'd like to start by thanking my advisor, Ben Calhoun, for seeing potential in me as a fourth-year student, even when I wasn't searching for a Ph.D. at all. After I talked with him for the first time, I stepped away thinking that if I joined the program I'd learn a great deal, not just about engineering, but about leadership and management which where characteristics that I greatly cared about. I can say with confidence that this was one of the greatest take-aways for me from the program. Next, I'd like to thank Steve Bowers who acted almost like a co-advisor at times. I worked on more wake-up radios chips for him than I can seem to keep track of at this point. I learned a great deal from him and his group. I'd like to thank the rest of my committee, Scott Barker, Brad Campbell, and Barry Johnson, for their willingness to help me work through the uncertainty of developing the proposal and dissertation. And for a great may things, from admin help to great conversation, I'd like to thank Terry Tigner for all she gave to Bengroup. It was always a joy to stop by Terry's office, chat, and maybe get a piece of chocolate.

Now I'd like to thank the many colleagues I have had the pleasure of working with. The majority of my research work on wake-up radios was in collaboration with Anjana Dissanayake and Jesse Moody. I learned a great deal from both of you and will have fond memories of our time together. I'd also like to acknowledge Ningxi Liu who invited me to join the wake-up radio projects in the first place and taught me about designing digital. I greatly appreciate the fun side projects that I got to work on with Peng Wang and Rishika Agarwala. I want to thank Peng for her friendship over the years and for Thursday Qdoba bowls. My success couldn't be complete without

mentioning my many colleagues in Bengroup who I've worked with over the years as well: Shuo Li, Daniel Truesdell, Jacob Breiholz, Sumanth Kamineni, Shourya Gupta, Natalie Ownby, Katy Flynn, Xinjian Liu, Peter Le, and Suprio Bhattacharya. It takes teams to make tapeouts successful and I'm grateful for all I've learned from them. I'd like to acknowledge Anjana, Peng, and Rishika for the fun times over the years outside of work as well. There are also several people who I would like to acknowledge outside of Bengroup. From John Lach's old group at UVA, I'd like to thank Victor Leal Sobral and Luis Lopez Ruiz. From Steve Bowers' group, I'd like to especially thank Pouyan Bassirian, Divya Duvvuri, Linsheng Zhang, and Jay Sheth. And from the University of Michigan, I'd like to thank Omar Abdelatty. For supporting this research, I have many funders to thank including the National Science Foundation, DARPA, and the University of Virginia.

This process, and much more, wouldn't have been possible without the support of my friends and family. I'd specifically like to acknowledge Drew and David from the cozy house I've had the pleasure of calling home over the last four years. Their friendship, and the house, has been invaluable to me during this marathon of an educational experience. And lastly, I'd like to acknowledge my Dad, Mom, sister Claire, and the rest of my family for all of their support. They have supported me and my hobbies since I was young, which turned from basement science experiments to the career that I'm now pursuing. I couldn't thank them enough.

”Non est ad astra mollis e terris via”

”There is no easy way from the earth to the stars”

— Seneca the Younger —

Table of Contents

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Motivation

The Internet-of-Things (IoT) represents a technological movement to merge the digital realm with the physical one by connecting people with intelligence from the objects in the world around them [23]. The breadth of the IoT has expanded dramatically over time to encompass a wide variety of markets, applications, and technologies. The Industrial IoT, which covers sectors including agriculture, energy and utilities, logistics, manufacturing, mining, and oil and gas, is becoming critical to improve efficiency and reliability for operations in order to lead to reduced costs and minimize unplanned downtime (Fig. 1.1) [24]. It is projected that by 2025 there will be an overall growth of 107% in the number of connected industrial IoT sensors. The healthcare sector is expected to spend from $21 billion US in 2019 to $54 billion US in 2029 for end point electronics [25]. Even in the midst of a global pandemic, the IoT has proven to enable continuity of business without traditional physical presence through capabilities such as remote monitoring [26]. Underpinning all of these use-cases and benefits are the advanced technologies that enable such ubiquitous intelligence, communication, and processing. Information relevant to IoT applications is sensed by systems scattered across a potentially vast region of space and embedded into the environment itself. Many communication schemes, from 4G and 5G cellular technologies to global satellite backhauls, are implemented to effectively deliver critical data to servers performing machine learning to extract actionable information for the user. The explosive growth of the IoT evident from these numbers speaks to not only our collective desire for advanced levels of connection and intelligence, but a historical pattern of how electronic technology has lived up to the task

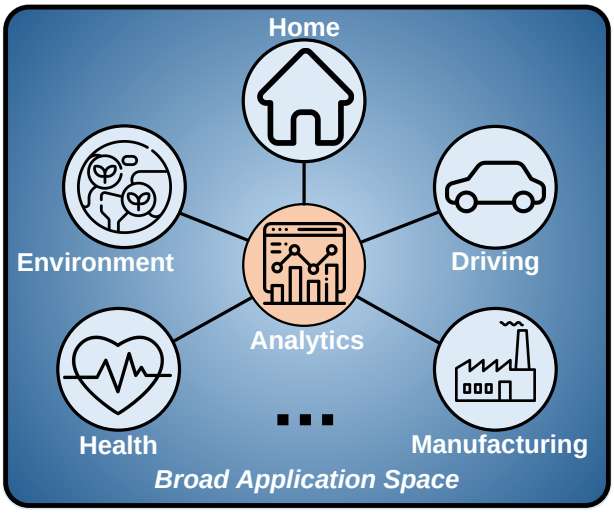


Figure 1.1: The Internet-of-Things supports a wide variety of possible application verticals to provide greater intelligence about the world around us.

in the past.

1.1.1 Scaling Electronic Computing

Even from millenia ago, humans engineered computing machines that reduced the complexity of a specific task, from the abacus to Blaise Pascal's mechanical calculator. However genius those inventions were for their time, the impact of the transistor and invention of the integrated circuit are arguably one of the most important discoveries for enabling the high-scale and wide-spread nature of these computing capabilities. But simply making a technological breakthrough wasn't enough to change the world. The key element that allows us to have mobile computers in our pockets today is through effective scaling while maintaining low cost. In other words, the economics matter just as much as the technology. A paper written in 1965 by Intel co-founder Gordon Moore noted an observation about the increasing number of transistors that would fit into a single chip, though more importantly how it would be done in a cost optimized fashion. This idea concerned the balance between the cost per transistor and the cost of failed chips due to defects (yield) when transistors became too closely packed together [27]. Now known as "Moore's Law", the idea that "the number

of transistors per chip should double every two years" turned from observation to self-fulfilling prophecy as it became the standard by which companies would set targets for future designs [28]. While this law enabled wide-spread hardware, another important technological explosion was seen only soon after in the form of digital connection.

1.1.2 Scaling Electronic Connection

Originally motivated by Cold War era pressures, the Advanced Research Projects Agency (ARPA) sought ways to ensure reliable and low-latency communication for defense related applications, primarily in light of defending the United States against Soviet ICBMs which could be launched at a moments notice [29]. This birthed the era of packet switching networks in comparison to more vulnerable single node systems used at the time, which if they failed could cripple any necessary defense response [30]. Later this "ARPANET" combined with standardized efforts (TCP/IP) and other public funding sources, including the National Science Foundation, to help develop what is now recognized as the World Wide Web. The number of Internet hosts in late 1990 was approximately 300,000 globally, but only a year later was 600,000 [31]. In 1992, this became 1.1 million followed by over 2 million. Coincidentally, this explosive rate of growth nearly matches the same as Moore's Law. Regardless, these technological trends were critical to creating the world we know today and were sparked by visionaries then driven by the expectations and desires of customers, governments, trends, and economics. Today, these same expectations are the reason why the number of cameras per phone increases every year, smart watches have over taken quartz watches almost overnight, and electric cars were a mere science experiment in 2010 but only a few years later mass-produced Tesla cars could drive themselves.

1.1.3 Scaling an Internet Revolution?

It follows then that the Internet-of-Things should follow a similar trend characterized by rapid growth. Back in the mid-2010's, initial predictions for the number of connected IoT devices

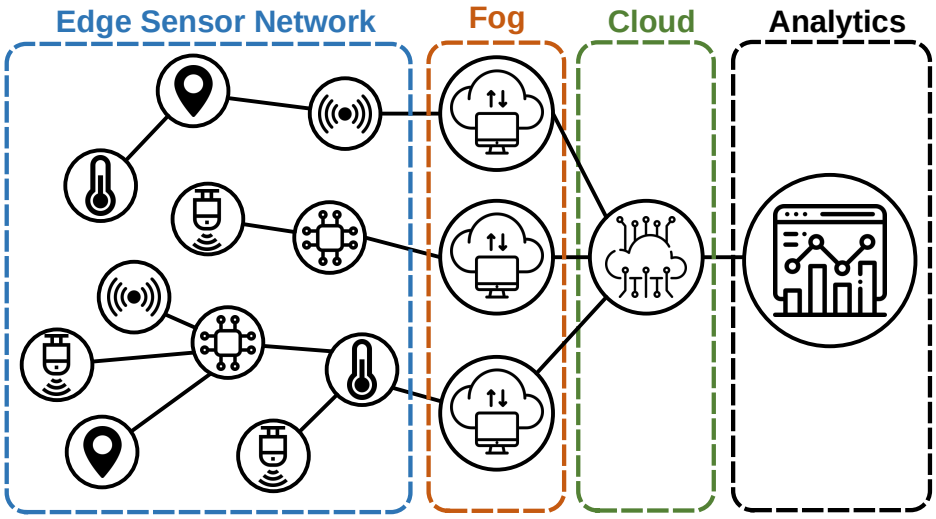


Figure 1.2: A common IoT architecture to connect large-scale edge sensing systems to users.

by 2020 ranged widely, from 18 to 50 billion and sometimes higher [32]. In a 2020 report, the predicted number of connected devices at the time was just over 20 billion falling on the low end of predictions demonstrating a low compound annual growth rate for IoT adoption and development of around 10% [33]. In comparison to the previous technological explosions, this one seems to pale in comparison. In reality, many high-level challenges exist between planning and realizing gains from IoT deployments, including security, standards, and deterministic return-on-investment (ROI). However, there is one key underlying problem that affects a significant amount of current, and desired, IoT applications: energy availability in sensing nodes [34].

The systems that actually recover information from the physical world are often referred to as ”edge” systems, due to their logical placement at the outer edges of an IoT networking construct (Fig. 1.2). The ”fog” systems then relay data and perform greater computational tasks followed by the cloud infrastructure which enables storage, analysis, and reporting of the information relevant to the user. Large-scale and high-performance computing are applied to cloud services, but aren't appropriate for ”lightweight” sensor nodes that are meant to be embedded into environment. These sensor nodes are also incredibly numerous and may be difficult to get to as a result of what they are intended to record and relay. Given this, these edge sensors require mobility and must run

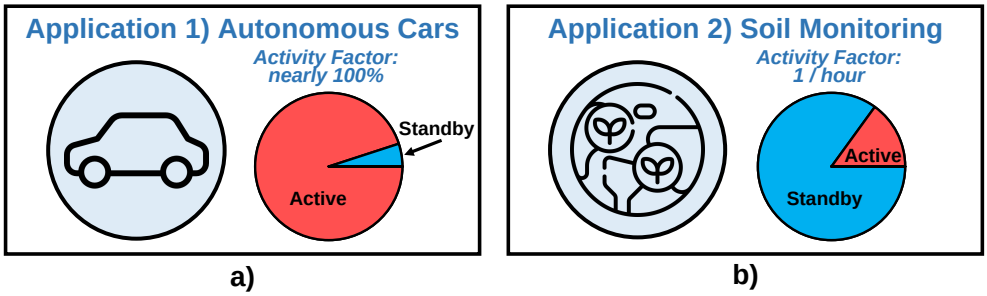
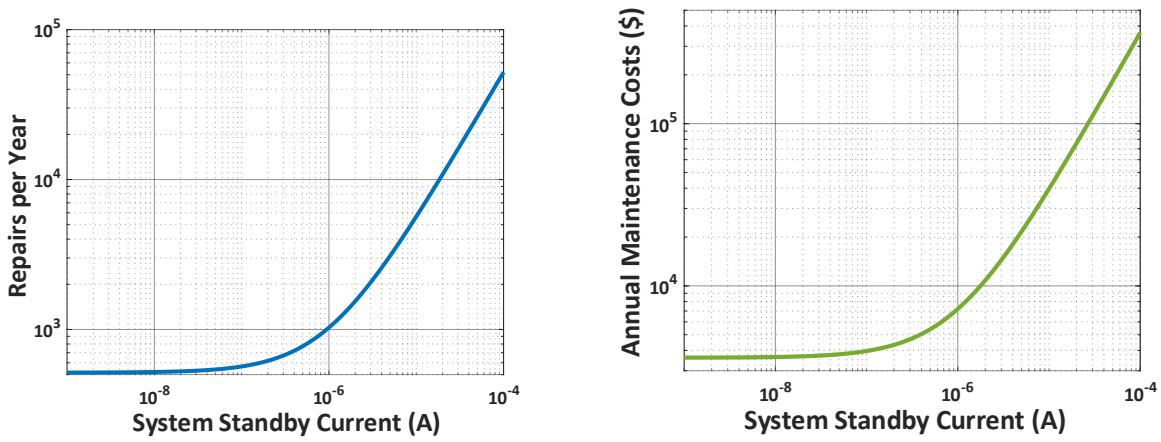


Figure 1.3: Two differentiating examples of IoT applications based on activity level which impacts dominating sources of power consumption.

on sources of energy not derived from readily available mains power, primarily batteries. Society is familiar with battery-powered systems because we recharge our cell phones and laptops every day. However, the influence of the IoT is predicated upon massive scale far beyond just several devices per person. As the IoT scales to tens and hundreds of billions of devices, it's crucial that these systems remain economically viable for the businesses the implement these solutions. Each IoT node in a deployment has an average lifetime, given that they each would have a fixed battery capacity (often quoted in mAh) and average level of power/current consumption based on the amount and type of activity performed. It's convenient to bin the power/current consumption of a node based on two categories: active and inactive. Active consumption occurs when the system is performing a specific task for the application itself, while inactive consumption is when the node remains idle.

An example of an active power dominated system, as shown in Figure 1.3, could be autonomous cars. In this situation, a camera could be streaming data to a processor for on-the-fly computation followed by communication to other systems. High-power components are vigilantly working to get the job done, which creates a high average power scenario. However the majority of the sensing nodes for the IoT aren't characterized by high activity factors. Rather, these more humble systems would perform simple "sense, compute, communicate" type actions, such as soil pH sampling. Inactive power dominates in these applications and is due to essential "always-on" components. Inactive power/current states are lower in magnitude compared to their active coun-



(a) (b)

Figure 1.4: Maintenance costs for an example IoT deployment with 10k battery-powered nodes (a) Number of nodes that need to be repaired on average per year. (b) A simplified calculation of expenses for labor and parts.

terparts and when considering the scale of the IoT it's important to consider the direct impact these values have on cost and the number of battery replacements.

To better understand this motivation, let's consider an example IoT scenario. Assume a given IoT deployment for a manufacturing facility requires 10,000 nodes to operate with a given level of activity factor for periodic sensing tasks (ex. temperature, humidity, air flow) that consume 1mA average active current for processes including sensing, processing, and transmission. Let's also assume these lightweight nodes are characterized by small form factor and therefore use coin cell batteries to run, in this case the CR2025 [35]. A node requires maintenance when it reaches the end of its lifetime, which creates several costs: 1) replacement batteries or recharging time, 2) work hours spent traveling to and maintaining the nodes, 3) accumulated downtime from the nodes which negatively affect business output. With this particular IoT deployment delivered to the company, let's say they find that it takes a rather modest 15 minutes to replace the batteries in nodes which have hit their lifetime where a simplified view of their costs are salaries of $20

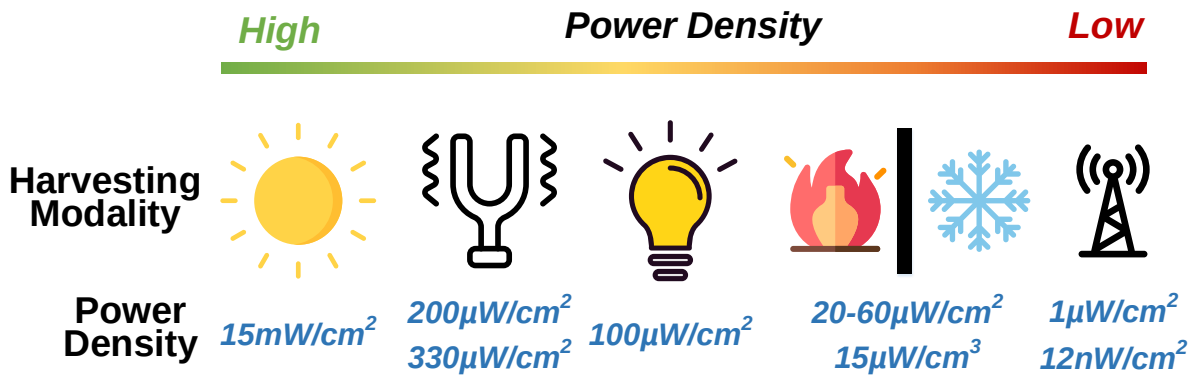


Figure 1.5: Demonstrated power densities for solar, vibration, indoor lighting, heat differential, and ambient RF environmental sources [1, 2]. These values are not guaranteed, but are meant to show what is generally possible from each harvesting modality.

per hour and where battery replacements are $2 a piece. A pessimistic calculation of the total number of replaced nodes per year and annual costs show that they are strongly dependent upon the inactive current consumption for the devices, stretching across orders of magnitude (Fig. 1.4). At least 500 systems must be replaced per year but easily climbs up to over 10,000 when systems consume significant idle power. In a similar trend, the projected maintenance costs range from thousands to hundreds of thousands of dollars. The simplicity of this approximation overlooks several factors that would in reality further increase costs. Two major factors are how difficult it is to access the devices embedded into the environment and the true cost of labor and parts. The battery assumed here is also ideal, where in reality yield, performance, and true capacity will be less than expected. As a result, these costs due to poor idle current could ruin any ROI benefit for this particular deployment if not at least significantly impact returns. This also only represents a single deployment. Consider then the impact of billions of edge systems with unoptimized standby states. For the sake of creating a cost-effective IoT, edge sensing systems must be designed for energy efficiency in their ”always-on” modes.

1.1.4 Increase Energy Availability or Reduce Power Consumption

In attempt to further reduce the challenge of finite system lifetime, alternative means of powering the system through energy harvesting (EH) can be employed (Fig. 1.5) [1, 2]. From sources such as light, heat differentials, and vibration, a node can harvest, store, and regulate that energy to run itself. Perpetual operation of a self-powered system requires that the following inequality, based on theory in [36], be satisfied which dictates relationships between the harvesting, consumption, and storage of energy over a defined time. The variables P_H and P_L are the instantaneous harvesting and effective load powers and E_{T_1} is the initial energy stored at the beginning of the time period. The harvesting efficiency factor, $\eta(x)$, is a function of numerous variables including P_H.

$$E_{T_2} = \int_{T_1}^{T_2} ([\eta(x)P_H(t)]^+ - P_L(t))dt - \int_{T_1}^{T_2} P_{E,Leak}(t)dt + E_{T_1} > 0 \tag{1.1}$$

where energy storage upper bound is given by

$$[x]^+ = \begin{cases} 0 & E(t) \geq E_{max} \\ x & E(t) < E_{max} \end{cases} \tag{1.2}$$

However, the dependability of a self-powered system can be tricky to predict and manage if the harvesting environment is not steady and predictable. Unfortunately, many harvesting environments can be unpredictable and scale over a range of several orders of magnitude in terms of instantaneous power consumption. Figure 1.6 shows that for an example $10\mu W$ system there are three approximate regions for dependable operation: dependable, undependable, and no operation. While some harvesting sources may at first glance seem sufficient, their power profiles are actually characterized by a statistical distribution, shown here as the color gradient for each line where dark represents high probability. When these deployments aren't dependable, the sensing mechanisms can't be trusted to vigilantly monitor and report back key information. In effect, they may not be of any help and lead to a false sense of trust. Another challenge is that most sources of local environmental energy aren't characterized by high power density. This means that large transducers (high area/volume) may be required to sufficiently power the device. Some systems opt to combine both

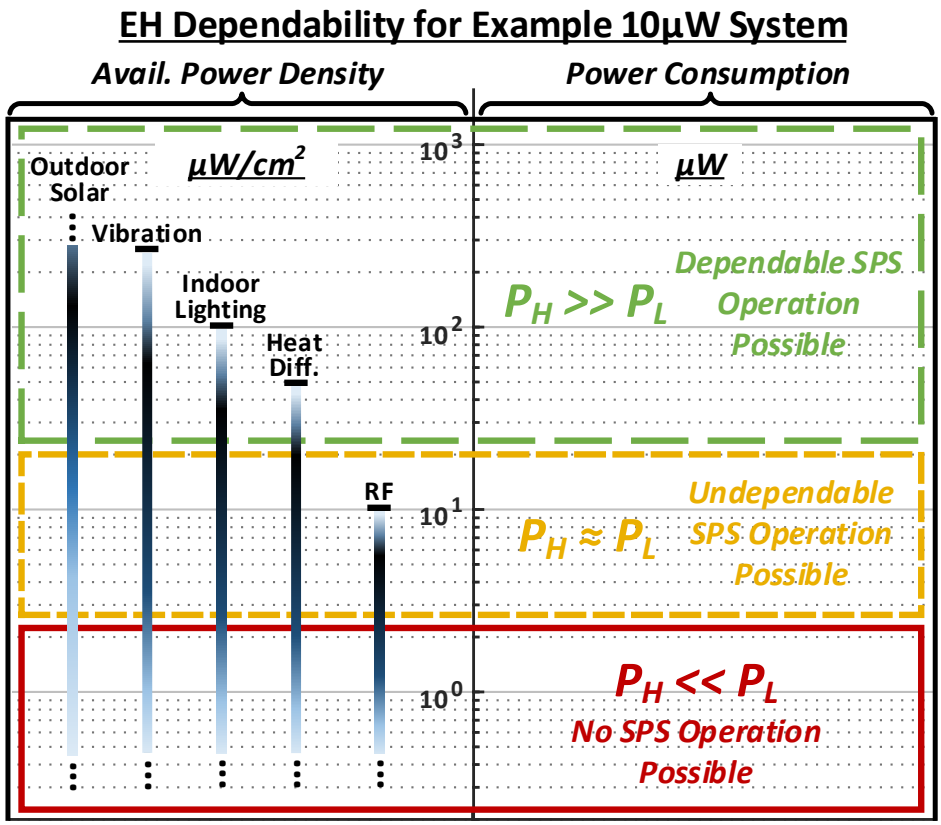


Figure 1.6: Harvesting sources are often not predictable and vary considerably, shown by the color gradient per source on the left. For a $10\mu W$ example system, this shows general regions by which this system may be reliable demonstrating the difficulty to achieve robust self-powered operation.

battery and energy harvesting techniques for improved performance. However, form factor matters significantly for these edge sensing systems, so battery and harvesting transducer sizes must be tolerable for the application which limits energy availability. There is also the added weight, financial cost, and environmental impact to consider. In summary, poorly planned battery-based IoT deployments suffer from excessive costs while unoptimized self-powered solutions are limited in dependability which affect coverage and awareness. A better solution is to improve the systems themselves by reducing power consumption.

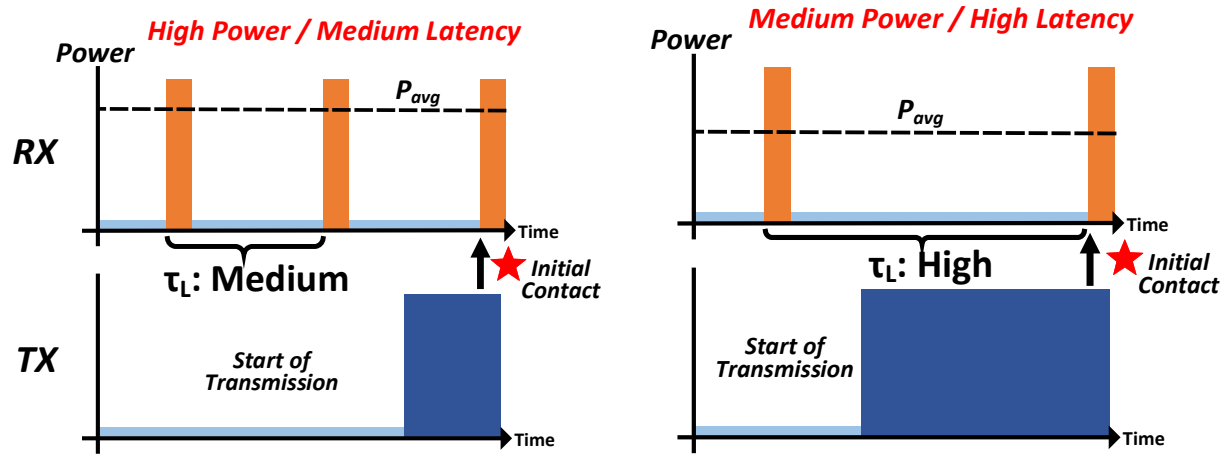
Numerous low-power systems-on-chip (SoC) have been proposed over time, generally focusing on specific IoT applications. Works presented in [37, 38] emphasize nW-to-μW sensing

capabilities for biomedical and ambient light applications. These lightweight nodes are both able to not only record, process, and transmit data, but also use the same modality as a source of energy harvesting. Other work, [39], developed a general purpose low-power (500nW to $1\mu W$) SoC which demonstrated several applications. Commercial low-power systems, including [40], are built to operate to operate robustly at low-power in industrial environments, while at the cost of nearly $100\mu W$. Achieving power on this scale often requires numerous techniques at the hardware, software, and application levels. Simple nodes can be optimized to do one very specific task, but general purpose systems will require a greater range of hardware making this task more difficult. Designing circuits to operate at lower voltages and lower duty-factors are common techniques to achieve these goals along with clock and power gating. For energy storage, many systems prefer super-capacitors over traditional batteries for a greater number of possible recharge cycles as well as lower internal resistance for high current loads, such as transmitters. Generally speaking, systems will contain at least several essential components in order to accomplish their edge IoT tasks. These include analog front ends for sensing, a lightweight processor, memory, and hardware accelerators for computation, a transmitter (TX) and receiver (RX) for wireless communication, and different types of regulators (LDOs, switching-based architectures) for power management. In order to improve lifetime by reducing power consumption, we should target the components that contribute most significantly to the power budget of the system.

1.1.5 Main Limiting Factor for Low-Activity System Power

Nearly every electronic device in the world today relies on wireless communication making transmitters and receivers an essential part of edge IoT devices. Mobility necessitates that these systems be wireless, but wireless communication comes at the cost of high power consumption. In fact, wireless communication generally dominates the average power consumption profile, especially for energy-limited IoT edge systems [41, 42]. Therefore, system lifetime is most affected in a large number of situations by improving how wireless communication is used and implemented.

As previously mentioned, many edge IoT sensing applications don't require high levels of



(a) (b)

Figure 1.7: IoT node with polling RX has a lose-lose trade-off situation where either (a) the receiver consumes far too much power with greater-than-desired level of latency or (b) the receiver would be slightly lower power but would take too long to respond to requests.

activity, such as basic environmental sensing. In this case, the transmitter for any particular node is rarely used, but the receiver must be active to some degree in order to hear requests that come through. Without any sense of scheduling, the receiver must poll the wireless channel searching for incoming data. Conventionally, power improvements are made at the application and networking layers. Often, systems are programmed to simply communicate less frequently, thereby reducing the average power contribution, but consequently increasing networking delay. This is observed in the comparison between scenarios (a) and (b) in Figure 1.7. The first shows a higher standby power due to the RX (P_{avg}), but is characterized by a medium amount of networking latency (τ_L) and as well as lower energy impact per TX event due to the shorter duration it takes to initiate contact. On the other hand to reduce standby power, the receiver can trade this off for worse latency and worse TX energy expenditure. Neither of these situations in this particular trade-off space are a win however.

The low-activity factor nature of edge IoT applications means that a great deal of energy is wasted in the RX in order to achieve reasonable networking latencies. This is known as the ”idle

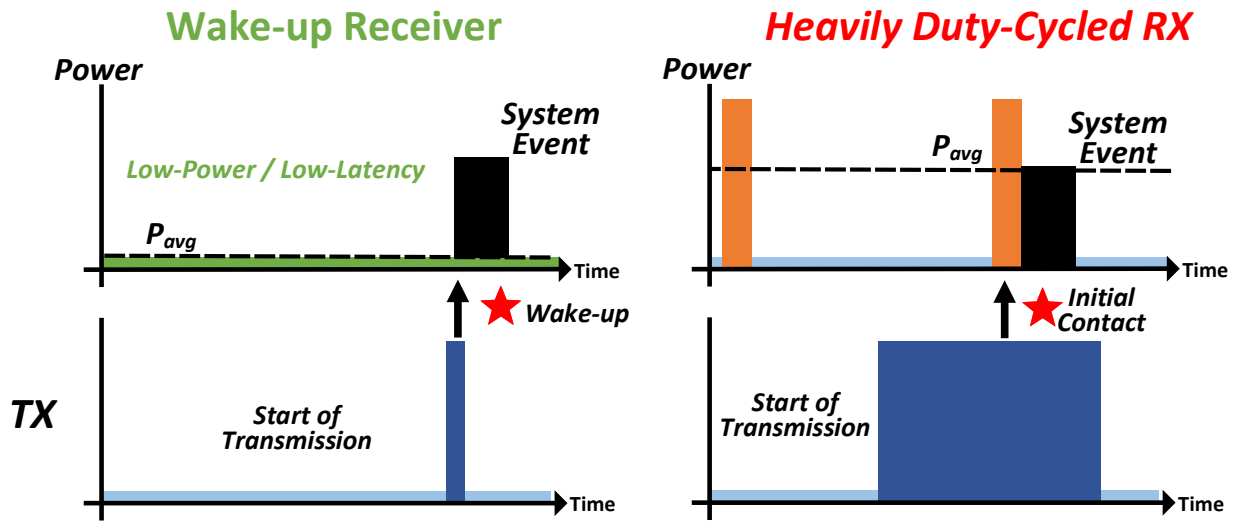


Figure 1.8: Wake-up receivers a solution to the idle RX listening problem and are designed to operate at low-power and low-latency, which subsequently reduces the power burden on the TX when generating an event on the listening system.

listening" problem [43]. Therefore, the most significant impact in energy savings for many low-activity and low-power sensing systems can be achieved by developing solutions to idle listening by the primary receiver. A variety of attempts have been made at the network level to improve RX operation and save energy and are generally implemented as different flavors of media access control (MAC) protocols. [43–45]. These approaches take the simplified cases in Figure 1.7 and apply adaptive networking techniques to help lower the power. However, software-based approaches are limited by the high-power and black-box nature of commercial-off-the-shelf (COTS) transceivers. Further power improvement can be made by innovating at the hardware level.

1.1.6 The Wake-up Receiver

A new category of receiver called the wake-up receiver (WuRX) is an alternative solution to the idle listening challenge by addressing it at the hardware layer (Fig. 1.8). The primary objective of a WuRX is to act as a low-power, "always-on" event-driven wireless notifier to activate the primary receiver upon the request from another system. This request comes in the form of a wake-up

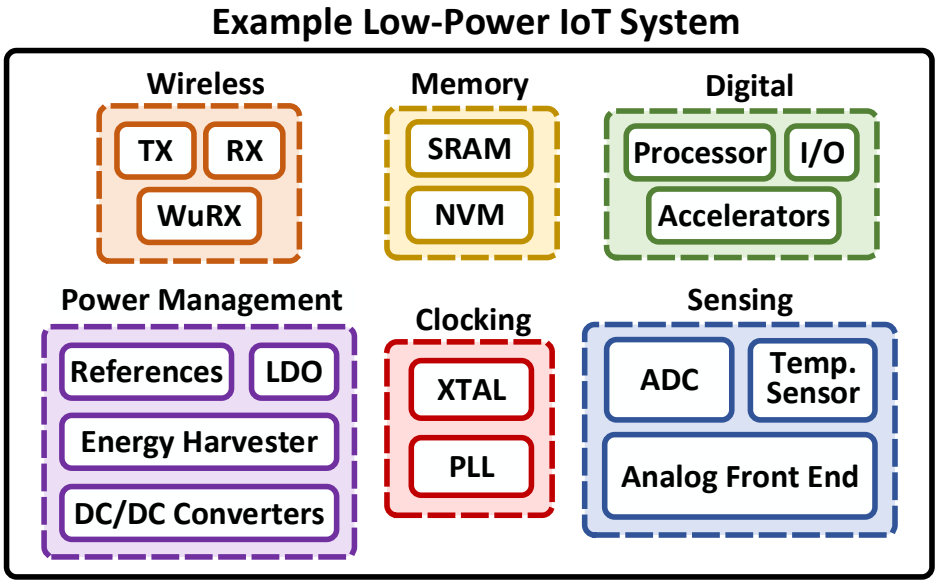


Figure 1.9: Example breakdown of the many components that make up a low-power IoT sensing system.

packet, which is effectively an address to uniquely separate one wake-up receiver from another. In this way, the RX is only used when needed, which can present a huge leap in both communication and system energy efficiency for single-hop and node-to-node networks. The primary challenge in developing wake-up receivers is to achieve sufficient performance depending on the application because this reduction in power always comes at a cost. By innovating in this new category of receiver, the WuRX can enable greater lifetime and availability for the plethora of IoT devices that will be needed in the future for cost-effective solutions.

1.1.7 Determining System Impact of the Wake-up Receiver

While the wake-up receiver as a component is anticipated to make a large impact in certain situations, it is tough to actually forecast when and by how much it can improve the lifetime and reliability for edge IoT systems. First consider the structure of an example low-power IoT system shown in Figure 1.9. It is comprised of many components that enable an array of sensing capabilities, generic and dedicated digital processing, volatile and non-volatile data storage, power

management, and of course wireless communication. The WuRX is just one component in a sea of other integrated circuits that make up an IoT system-on-chip (SoC), so how can we know if, for example, that scaling aspects of the wake-up radio would be any better than adjusting the transmitter or processor. The complexity of the system and IoT application are significant enough that the simple approximation that in all ”low-activity factor” scenarios the WuRX can help by ”a lot”. It's critical not only to design effective components, but to also plan ahead what the system will do and quantifiably model whether or not under the energy availability constraints how successful the system can be.

1.2 Thesis Statement

Wake-up receivers characterized by sub or near-μW power and high performance, especially in terms of RF sensitivity, wake-up latency, and interference mitigation, can be achieved by applying duty-cycling and certain digital methods to high-power wake-up radio architectures. Further, a generic system modeling framework capable of relating design and operating variables to power consumption allows designers to rapidly prototype current and future systems for power and performance optimization.

1.3 Approach

This dissertation addresses the challenges of designing effective wake-up receivers for the low-power IoT in three phases: 1) Addressing receiver-level power improvements through several duty-cycling schemes, 2) detailing critical digital baseband methods necessary to realize these wake-up receivers, and 3) demonstrating these concepts in custom wake-up receiver designs fabricated in commercial CMOS. This dissertation addresses the generic IoT system design challenge by creating a new system modeling framework to enable design, operational, and trade-off analyses for low-power devices which will be demonstrated in the use cases with a wake-up receiver.

1.3.1 Duty-Cycling for Wake-up Receivers

Low power is the most critical metric that distances a wake-up receiver from a primary receiver on an IoT system. There are many possible techniques to enable low-power operation, but simultaneously achieving sufficient performance is a challenge. We take the approach of implementing high active power receiver architectures, relative to conventional nW-level passive RF front end topologies, to achieve sensitivity and interference performance, then rely on several duty-cycling techniques to reduce average power consumption to nW-to-μW levels satisfying power and performance requirements. Three different duty-cycling techniques are analyzed to demonstrate relative advantages to one another: bit-level duty-cycling, packet-level duty-cycling, and a new proposed within-packet duty-cycling.

1.3.2 Digital Baseband Design and Implementation for Wake-up Receivers

The digital baseband plays a critical role in the success of wake-up receivers. There are a variety of architectures implemented based on the duty-cycling scheme employed. Lower-level digital techniques including binary correlators for generating wake-up signals, automatic gain and offset control loops, power management timing blocks, and data path logic are presented which all play a critical role in the successful development and operation of a wake-up receiver. We demonstrate the theory and discussions with measurements from four different wake-up receivers taped out in a 65nm LP CMOS process.

1.3.3 Developing and Demonstrating a System-Level Modeling Framework

While component-level innovation is critical to the path of achieving low-power systems, this represents only the first phase of the challenge. As systems become more complex and as more applications emerge, the design and operational complexity increases as well. Optimizing a system for a variety of factors and understanding key trade-offs becomes an intractable problem to manage alone. We present a system modeling framework that allows a user to connect the dots between

design, operational, and trade-off choices with power consumption throughout the system. We use this flexible software tool capable of describing any system, with varying levels of accuracy and detail, to investigate several important questions about the use of wake-up receivers in a system's context to show how useful this component is and under what situations.

1.4 Organization and Contributions

1.4.1 Background

Chapter 2 highlights important background information for material in this dissertation. It includes overviews of important receiver metrics, techniques to lower power consumption, and prior art wake-up radio methods.

1.4.2 Duty-Cycling Methods for Wake-up Receivers

Chapter 3 discusses conventional duty-cycling techniques, including bit-level and packet-level duty-cycling, used to reduce the average power of wake-up receivers. It also proposes a new scheme called within-packet duty-cycling for improved performance by utilizing a carrier-sense technique. Each method is quantified in terms of its effect on average power and wake-up latency. A mathematical comparison of the three techniques demonstrates the advantages of the proposed duty-cycling method.

1.4.3 Digital Baseband Techniques

Chapter 4 presents on the digital controllers, power management, and data path blocks critical to wake-up receiver operation. First, this chapter will lay out key building blocks necessary for the digital baseband and architectural differences based on the implemented duty-cycling method. An analysis on binary correlators presents on their theoretical behavior and measured results for

different implementations. The proposed within-packet duty-cycling algorithm is introduced and theoretically analyzed. We lay out a methodology for choosing all algorithm variables for an optimal baseband sensitivity. Measurement results are presented to demonstrate agreement with theory. Using the same carrier-sense logic, a TX carrier-sense multiple-access block is discussed which outlines how to enable collision-free node-to-node wake-ups. This chapter will end by discussing several implemented automatic gain and offset control loops and associated measurement results.

1.4.4 Implemented Duty-Cycled Wake-up Receivers

In Chapter 5, we collectively report four different wake-up receivers that were designed, fabricated, and tested for a variety of low-power IoT applications. The first three are intended for kilometer-range and sub-μW applications. The last wake-up receiver was motivated by the need for near-μW wake-up receivers with low latency in the popular 2.4GHz ISM band, given the extensive level of wireless infrastructure.

First, we present the development and testing of a bit-level duty-cycled Tuned-RF wake-up receiver that improved state-of-the-art sensitivity by 26dB (compared to other low-power wake-up receivers) and reduced power by over 1000x (compared to other receivers with similar sensitivity). Second, we present the development and testing of a second-generation bit-level duty-cycled wake-up receiver that enabled high digital reconfigurability in sensitivity, power, and latency by 11dB, 410x, and 672x respectively in order to satisfy a wide variety of IoT application requirements without requiring silicon redesign. Third, we present the development of a third-generation bit-level duty-cycled wake-up receiver that achieved high levels of integration and multi-channel wake-up capability all at nanowatt power for lower cost and area as well as denser wake-up networks. And lastly, we present the development of a 2.4GHz packet-level/within-packet duty-cycled wake-up receiver demonstrating a 9x and 10x improvement in power and wake-up latency compared to the packet-level duty-cycling mode and a 2.2x and 10x improvement in power and latency compared to prior art.

1.4.5 System-Level Modeling Framework

Chapter 6 presents a custom software modeling framework to inform engineers on the design, operational, and trade-off space for low-power systems. The framework architecture is discussed in three distinct parts: system structure and hierarchy, modeling context, and modeling experiments. In this framework, we developed widely applicable functions that allow the user to 1) sweep operating variables and report power consumption, 2) set a target system power and let the tool tune a given variable to achieve that power, 3) exchange two operating variables while maintaining a constant system power and review the trade-off, and 4) calculate lifetime as a function of operating variables. An example low-power IoT node model is used to demonstrate numerous modeling capabilities built into the framework and is based upon silicon results.

1.4.6 System Context for Within-Packet Duty-Cycled Wake-up Receivers

Chapter 7 utilizes the modeling framework from Chapter 6 to answer questions related to utilizing a wake-up receiver in a low-power system context. It focuses on three modeling experiments in particular. The first covers the topic of system level impact by moving from packet-level duty-cycling to within-packet duty-cycling. The second explores the trade-off of event latency versus communication activity, or in other words ”how quickly” versus ”how often” wake-up events are required. The final experiment discusses dependability concerns for using a wake-up receiver in a system through the lens of false carrier-sense detections and false-wakeups.

CHAPTER 2

BACKGROUND

2.1 Receiver Range and Sensitivity

Wireless communication based on electromagnetics is predicated upon the ability to transmit encoded data with modulated electromagnetic energy then receive it at another point in space. The range equation can be used to estimate how much power will be received (P_r) (or must be transmitted (P_t)) as a function of distance (d) and antenna gains (G_t, G_r) [46]. This is given by:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} \tag{2.1}$$

This equation assumes that the transmit and receive antennas are fixed with isotropic radiation patterns. One important take-away from this equation is the relationship between wavelength and received power, due to the antenna's effective area. Broadly speaking, lower wavelengths will "travel further" than shorter wavelengths and as a result play an important role in their choice for different IoT applications. And based upon a certain P_r, the receiver then needs to recover the original data with a required level of signal-to-noise ratio (SNR) to achieve a desired information error ratio. This process of recovery further adds noise and distortion to the signal and it is critical to budget for this during the design process. The sensitivity equation relates these concepts and is derived from the original formula of noise figure, which is defined as:

$$NF = \frac{SNR_{in}}{SNR_{out}} \tag{2.2}$$

The noise figure is important to a receiver because it is related to the tolerable error ratio for the encoded data, which depends on the modulation scheme and required bandwidth (B) [47]. This

equation can be written as

$$P_r = P_{RS} \times NF \times SNR_{out} \times B \tag{2.3}$$

or in the commonly used logarithmic measure for power, relative to a milliwatt (*dBm*):

$$P_{r,dBm} = P_{RS,dBm/Hz} + NF_{dB} + SNR_{min,dB} + 10log(B) \tag{2.4}$$

where $P_{RS,dBm/Hz}$ is the source impedance of the receiver, and when matched $P_{RS,dBm/Hz} = kT = -174dBm/Hz$ at room temperature, $SNR_{min,dB}$ is the minimum required SNR to detect the data based on the modulation scheme, and NF is the noise figure of the receiver chain. The use of more complex modulation schemes can provide higher data rates, but at the expense of greater bandwidth and resulting loss in sensitivity (requires a larger minimum input power to demodulate data). The error rate for data is often quantified in terms of "bit error rate" or "bit error ratio" (both denoted by *BER* in the literature), where the former describes the rate of bit errors in time while the latter quantifies a unitless measure of error performance at the bit level. For digital communications, the minimum required signal-to-noise ratio is often redefined in terms of energy per bit divided by the noise power spectral density, or E_b/N_0. This is because digital information is represented over a finite time, T, and therefore has zero average power and finite energy making this a more natural figure of merit. The sensitivity equation then becomes:

$$P_{r,dBm} = P_{RS,dBm/Hz} + NF_{dB} + \frac{E_b}{N_0}_{min,dB} + 10log(R_b) \tag{2.5}$$

where $E_b/N_0 = SNR \times B/R_b$. As a result, a better sensitivity means better range for the same transmission power, because the required input RF power to the receiver is lower to achieve a specified BER (often compared at values such as 10^{-3}). Sensitivity is a key metric, especially for low-power wireless devices, as reducing power consumption often means sacrificing sensitivity performance and therefore limiting the possible applications.

2.2 Wireless Interference

The range equation describes a rather ideal signal propagation scenario. In real life, there are objects and multiple paths the energy can take to go from a source transmitter to the intended re-

ceiver. Path loss, fading, and multi-path are characteristics of the channel that the electromagnetic energy radiates through. These phenomena attenuate signal power by factors greater, and more complex, than those predicted by the range equation and distort the signal creating inter-symbol interference (ISI), often requiring the use of equalization filtering to correct for the effects. Also, we want to enable many users to access the wireless resource for simultaneous communication, which can generate interference if signals aren't isolated properly. Different multiple access techniques and wireless standards exist to provide high-scale communication. These methods work by encoding data in orthogonal mediums such that User 1's data won't interrupt User 2's data and vice versa [46]. Orthogonality is defined as functions that satisfy the following criteria:

$$\int_0^T x_j(t)x_k(t)dt = K_j\delta_{jk} \qquad 0 \leq t \leq T \qquad j,k = 1,...,N \tag{2.6}$$

where

$$\delta_{jk} = \begin{cases} 1 & j = k \\ 0 & otherwise \end{cases} \tag{2.7}$$

The value T is the symbol duration and $\{x_j(t)\}$ are linearly independent basis functions. Several of the most popular multiple access techniques include using information in time (TDMA), frequency (FDMA), code (CDMA), and spatial (SDMA) domains to generate orthogonal data streams. Wireless standards are used to define essential functions and specifications which govern numerous aspects of a particular form of wireless communication, usually centered around a particular application, and help define receiver requirements. Examples of these are the ubiquitous IEEE 802.11 set of standards for Wi-Fi and the Bluetooth Special Interest Group for personal area networking common in small, mobile devices. Because standards require that transmissions occur within a set of predefined channels, it is critical that the associated receivers are able to block close-by interferers (i.e. other data) in the frequency domain while still receiving data with sufficient BER. Therefore, interference tolerance and rejection is an essential quality for receivers. This is defined differently depending on the standard, but often quoted in terms of the largest interferer the RX must tolerate while receiving a small desired signal [47]. The signal-to-interference ratio (SIR) is one figure of merit for this measure. A common SIR test backs off the intended signal by 3dB

from stated receiver sensitivity, positions the blocker at a certain offset, and tests how strong the interferer needs to be in order to achieve a maximum BER value (as a result of SNR degradation). This blocker is then swept across frequency to show a ”mask” of interference performance in the frequency domain. Achieving large SIR at low frequency offsets is highly desirable as this enables better isolation from unwanted corruption as well as more efficient packing in the frequency domain. Standards often define required adjacent channel interference requirements as well. Because channel spacing is known ahead of time in these situations, it becomes easier to prescribe the magnitude of blockers to be tolerated up to several channels away for any particular receiver. The level of interferer tolerance is highly dependent on the receiver architecture, often requiring exceptional filtering and compression characteristics. All of this leads to greater levels of power consumption.

2.3 Reducing Receiver Power with Protocols

Given the required performance of receivers just for sensitivity and interference, low-power wireless systems need to adopt methods to reduce the average power of their radios. Often, many power saving methods come from protocols. One common standard for low-power applications is the Bluetooth Low Energy (BLE) standard.

Bluetooth Low Energy is a special mode integrated into Bluetooth 4.0 and is found in many commercial-off-the-shelf (COTS) SoCs available today. To stay within the scope of this work, we will consider just the portion of BLE operation which matters for low-activity systems common place in the edge IoT. This BLE operation is the default ”standby” state that the receiver moves into in order to check for incoming communication requests. Keeping the receiver always-on in this discovery mode would not be acceptable, so the receiver operation is duty-cycled to reduce average power. This means it periodically goes from active to inactive states. The active duration is known as the scan window and the total period is known as the scan interval [48]. Because there are three possible advertising channels for possible transmitters who want to talk (Ch. 37, 38, 39), the receiver must check every channel one after another in a round-robin fashion with every new

scan window. The trade-off here is that the effective communication latency takes a hit when the receiver turns on less frequently. It takes longer to establish a communication link as a result.

The Bluetooth example represents the classic relationship between average power consumption and networking latency found in low-power systems, which was previously described in Figure 1.7. Generally for low-power systems, there are three classifications for duty-cycled communications: 1) synchronous, 2) asynchronous, and 3) semi-synchronous [49]. Out of these three options, asynchronous operation (like Bluetooth) is often highly preferable for low-power nodes because it doesn't require the extra effort it takes to achieve synchronization with, and maintaining synchronization of, entire networks of energy-limited systems. One popular asynchronous method to establish a link, utilized by [44, 45], is with the use of preamble sampling. Normally, a long train of "1"s and "0"s are sent before the data packet begins to align the receiver's clock with the transmitter's clock. By detecting the presence of this preamble or not, the receiver has the ability to turn off quicker which can help save some power consumption. In order to reduce the average power to a sufficient level for COTS receivers, the latency becomes extremely high and the transmitter simultaneously burns power for an even longer period of time taking over the local wireless medium.

2.4 Wake-up Receiver Prior Art

The primary utility of the wake-up receiver lies in enabling more energy-efficient wireless notification for low-activity systems, generally ones that have tight power consumption requirements. As a result, it is essential that any wake-up receiver operates at far lower power consumption than the standard receiver on the system. This is because the WuRX will contribute to the system floor power consumption and end up accounting for a large percentage of total power consumption. Any further savings from this component can dramatically increase battery lifetime or improved the reliability of energy-harvesting systems. The challenge then becomes determine how to design "low-power" wake-up receivers as well as ensuring it remains compatible with application require-

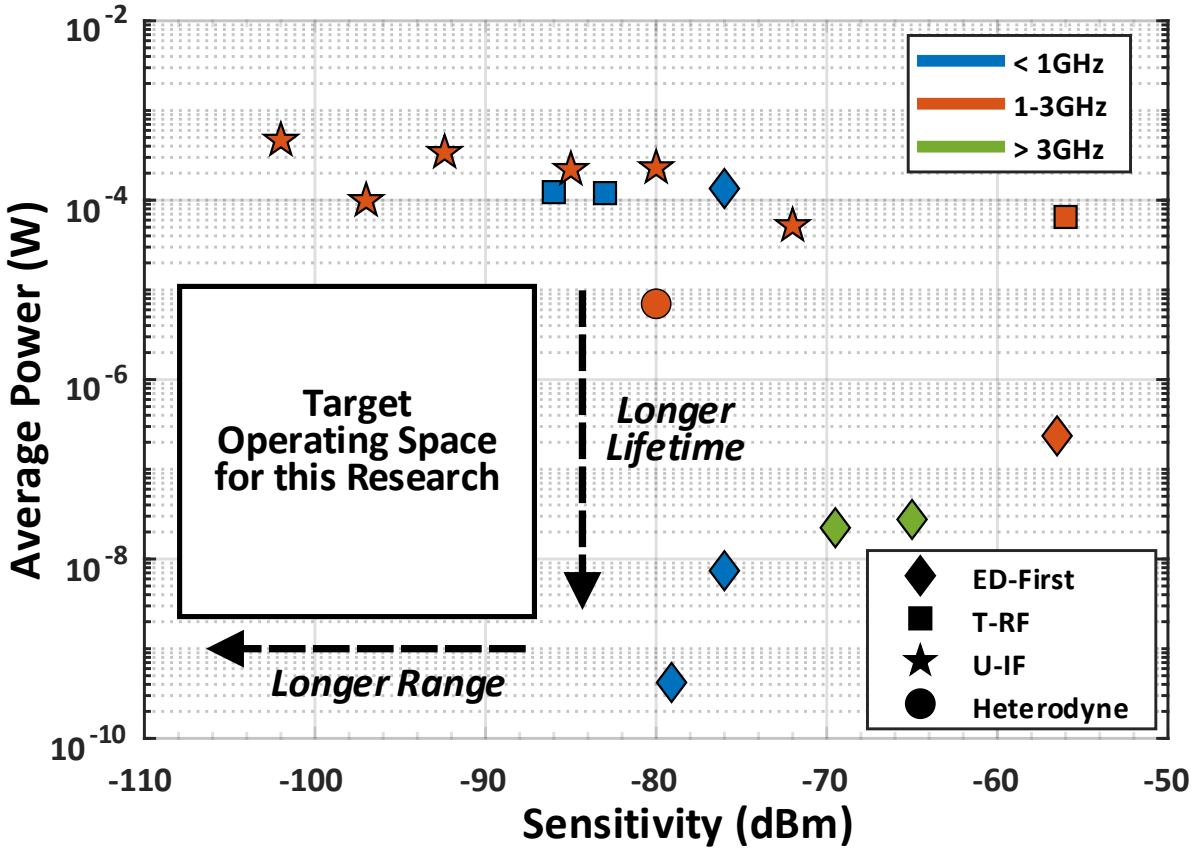


Figure 2.1: Scatter plot comparison of wake-up receivers for different architectures and operating frequencies as a function of average power and sensitivity [3–18].

ments. This research focuses on the target operating space shown in Figure 2.1 in comparison to other published wake-up receivers.

In the literature, there are primarily four main architectures used to implement wake-up receivers (Fig. 2.2). High-performance receivers often implement coherent detection, whereby they encode information at least in part in the phase component of the transmitted signal. These architectures may use numerous downconverting stages with high order filtering to achieve the stringent specifications required to operate in industry standards. For reconfigurability, high performance analog-to-digital (ADC) and digital-to-analog (DAC) converters are sometimes use to directly sample data before full demodulation to baseband enabling digital processing to take care of recovering the data. To save power, these WuRX architectures use non-coherent methods and employ energy

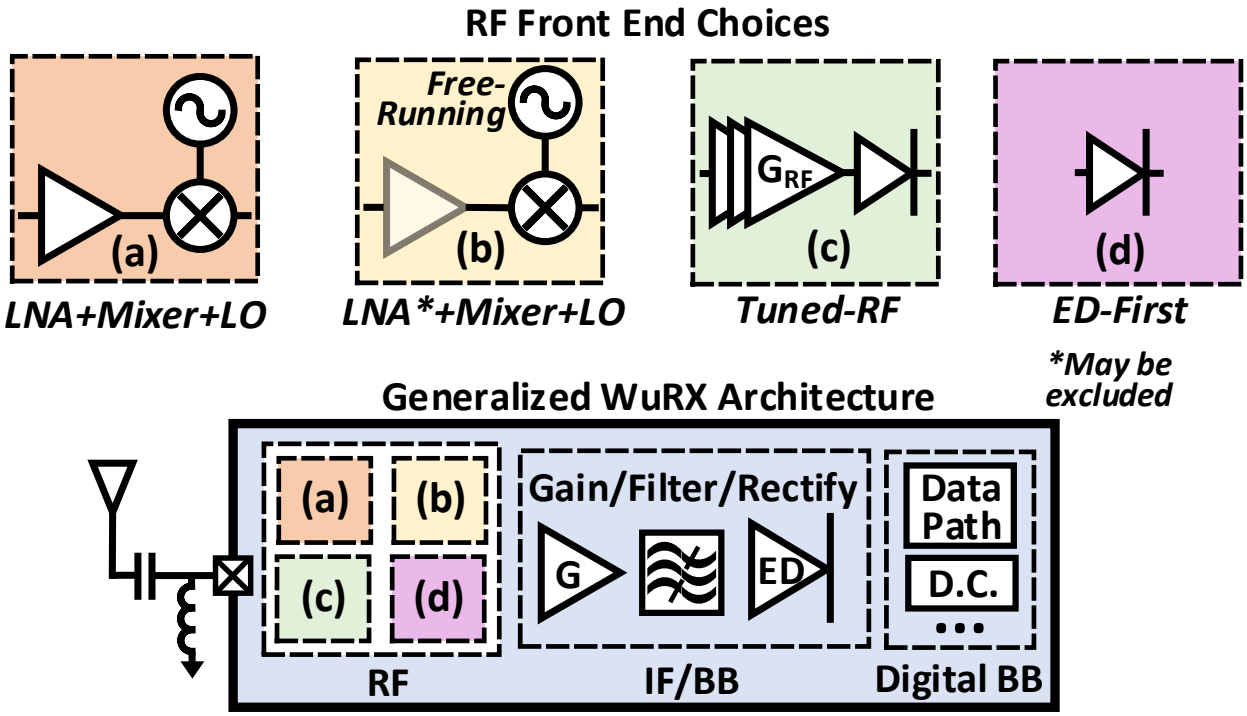


Figure 2.2: The four primary categories of wake-up receiver architectures differ mostly on the RF front end because it is the highest power consumption section of the component.

detection schemes through the use of rectifiers (envelope detectors (ED)) in the analog intermediate frequency (IF) or baseband (BB) stages along with filtering and gain stages in various configurations. Simple non-coherent modulation schemes, including on-off keying (OOK) and frequency shift keying (FSK), are used to encode data on the wireless carriers. In the digital baseband, the data path logic for bit decisioning can often look similar as well, utilizing correlation to identify wake-up signatures. Duty-cycling (D.C.) control loops may be important for the higher power architectures. Gain and offset controls are also implemented in some works to deal with temperature drift and interference. As a result, most of the ”right half” side of these architectures look similar. The diagram in Figure 2.2 generalizes the analog IF/BB and digital BB stages for simplicity. The primary difference lies in what is typically the most power-hungry part of the receiver: the RF front end.

The first WuRX type is based upon the classic heterodyne architecture. This generally em-

ploys a low-noise amplifier (LNA) followed by a mixer driven by a locked local oscillator (LO). The mixing result generates an IF stage which is followed by filtering, amplification, and energy detection stages. This particular architecture has the highest active power and must be aggressively duty-cycled in order to achieve the target WuRX power [3]. Few wake-up receivers in the literature utilize this topology because of the high-power consumption, with active currents $> 1mA$. Most wake-up receivers that use down-conversion techniques instead implement the next architecture style as it is more well suited for use in this context.

The second WuRX type makes a compromise with the first to reduce active power by loosening the restrictions on the RF front end, primarily the LO and LNA. The LO is designed to be unlocked or free-running, such that it isn't driven by a PLL critical for coherent detection. This "uncertain-IF" (U-IF) architecture gets its name from the randomly fluctuating IF signal down-converted from RF as a result of the unlocked LO. Ring oscillator and LC-based VCOs have been employed in this case, but come with key trade-offs including phase noise, temperature stability, and power [5, 6]. Even though LC VCOs are generally higher-power, they are used more often in the literature due to their improved performance. Sometimes this architecture employs a "mixer-first" topology removing the LNA for further power savings [5]. In order to reduce noise, filters such as BAWs and FBARs have been added to the RF input while other work implement sharp filtering at IF using N-path architectures [4, 5, 8]. Modulation and signalling schemes that encode information in the frequency domain, such as FSK, may require separate IF/BB paths after down-conversion to isolate symbols from each other [7]. Even with the power savings, duty-cycling is still important to achieve low power, though many don't employ it in the literature as seen by the reported numbers in the scatter plot hovering around $100\mu W$.

The third architecture further simplifies the RF front end to reduce power by removing classic heterodying components and primarily relying on high levels of RF gain followed by an RF envelope detector for bringing information down to IF or baseband. This is often referred to as the Tuned-RF, or LNA-First, architecture. It is also common in this architecture for there to be a high-Q RF filter at the RF input to reduce the high amount of noise that would be inevitably amplified and rectified throughout the chain, similar to some versions of the U-IF architecture [11, 12]. Other

signaling techniques, such as two-tone, help improve performance by moving the data away from dc post-rectification [12]. Also, by implementing variable RF gain stages a trade-off can be made between sensitivity and power [10]. However even these architectures demonstrate 10s to 100s of microwatts of power consumption, so further reducing the active power requires cutting all active power RF components.

The envelope detector-first (ED-First) architecture represents the lowest power category of all the architectures, while also suffering from the worst performance. These WuRXs can achieve nW-level power consumption primarily because they use a passive RF front end consisting of an input matching network to maximize signal swing into an envelope detector [13–18]. This would be followed by analog amplification, filtering, and comparison to determine encoded RF data. However, the detector-first architecture is limited in its sensitivity primarily due to the finite obtainable voltage-boost, non-linearity of the detector device, and output thermal noise of the detector [50]. Generally, this RF sensitivity limitation is around -80dBm which limits wireless range.

It's critical to consider the required application space before determining which wake-up receiver architecture to employ, based upon the wide span in performance and power. From Figure 2.1, the ED-First architecture achieves the desired levels of power consumption, but lacks in sensitivity significantly affecting range. To achieve improved distance for wireless communication, necessary for all IoT nodes, as well as sufficient interference rejection, latency, the move must be made to higher power architectures. And in order to achieve the desired power levels, duty-cycling must also be employed.

CHAPTER 3

DUTY-CYCLING METHODS FOR WAKE-UP RECEIVERS

3.1 Summary

This chapter covers the topic of duty-cycling in wake-up receivers. Two specific methods, bit-level duty-cycling and packet-level duty-cycling, are discussed and mathematically analyzed with respect to their impact on power consumption and wake-up latency. A new duty-cycling method, within-packet duty-cycling, is proposed which improves the power/latency trade-off compared to the traditional approaches. With this new duty-cycling scheme, a mathematical comparison between all three reveals the advantages of using within-packet duty-cycling in wake-up radios.

3.2 Standard Duty-Cycling Methods

Power scalability afforded by duty-cycling is in part due to the reduced percentage of time that high-power components remain active (τ_{act}) over a defined, repeating period (τ_{per}). The remaining utility from duty-cycling stems from how low a component's inactive power consumption is. If the inactive power (P_{in}) is nearly equal to the active power (P_{act}), then duty-cycling of any kind is of little use. The generic relationship between duty-cycling and average power (P_{avg}) can be expressed by:

$$P_{avg} = \alpha(P_{act} - P_{in}) + P_{in} \tag{3.1}$$

where the duty-factor is represented by $\alpha = \tau_{act}/\tau_{per}$. Each duty-cycling method follows this equation, but implements variations on α. For wake-up receivers, the primary trade-off to lower average power via duty-cycling is an increased wake-up latency. Because duty-cycling schemes rely upon periodic events or signals (based upon τ_{per}), there are two standout candidates to define τ_{per} for WuRXs. These two conditions are when τ_{per} equals 1) the bit period (τ_b) and 2) the wake-up packet period (τ_{wp}). The wake-up packet consists of all the bits in the wake-up code transmitted over a local network that the WuRX can listen to. The bit period and wake-up packet period are the smallest and largest unit of time-repeating information necessary for the WuRX to capture which is why they represent the lower and upper bound for useful duty-cycling techniques, in terms of active duration. These duty-cycling schemes are called bit-level duty-cycling (BLDC) and packet-level duty-cycling (PLDC), respectively. Each technique has their advantages and disadvantages which will be covered in the next section and throughout the rest of this dissertation.

3.2.1 Bit-Level Duty-Cycling (BLDC)

Bit-level duty-cycling (BLDC) is defined by setting the duty-cycle period equal to the bit period which is the inverse of the bit rate, as shown in Figure 3.1. The active time is long enough to sample a portion of the current value of the transmitted bit before the receiver turns off. This technique was employed in several previous wake-up receiver works in order to reduce the average power down to $7\mu W$ and $240nW$ [3, 7]. It will repeat this action every bit period and each sample will be used to help determine if the wake-up code was sent. The active duration can be approximated by a necessary start-up time (τ_{su}) for the receiver and RF integration time (τ_{int}). The start-up time represents both the time it takes for the internal bias points to reach steady state and for the propagation delay of the signal to go from input to output. A sufficient RF integration time is required to achieve a given sensitivity, assuming noise limitations are not yet reached. While the receiver is inactive, digital circuits, clocks, and references continue to operate and consume a relatively low amount of static power consumption setting the floor standby power of the component. Leakage from various inactive components is also a contributing factor to static power. For this

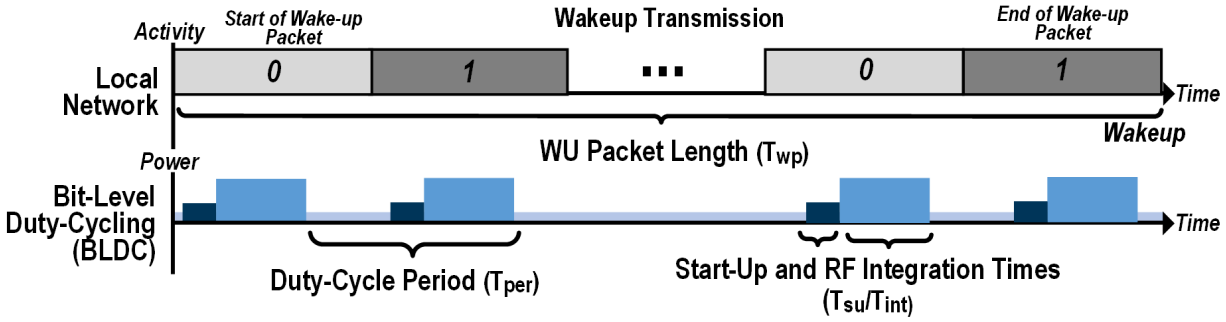


Figure 3.1: Bit-level duty cycling concept illustrated in time domain. A portion of each bit is sensed per bit period during the active duration.

scheme, α_{BLDC} is determined by:

$$\alpha_{BLDC} = \frac{(\tau_{su} + \tau_{int})}{\tau_{per}} \tag{3.2}$$

It is also noteworthy that BLDC would be the most backwards compatible duty-cycling scheme with ED-first wake-up transmissions.

Wake-up Latency Considerations. A transmitter in the local network needs to provide a constant transmitted bit over τ_{per} so the bit can be asynchronously received. This is done for each bit in the wake-up code, of length N_{code}. Once the last bit in the wake-up code is received by the BLDC WuRX, a wake-up interrupt signal is issued. This assumes that the resulting bit decision is latched into the correlator at the very end of the active duration. Wake-up latency is defined as the time delay between when a wake-up transmission is first sent and then fully received by a WuRX, which triggers a wake-up flag. In BLDC, wake-up latency varies as a function of the bit period. However, this delay isn't deterministic. A random phase offset between the transmitted data and WuRX active time generate uncertainty in the resulting latency. This allows us to calculate an average latency $L_{avg,BLDC}$ instead. The best-case latency (lowest latency) occurs when the WuRX active duration begins right when the transmitted bit period starts. After $(N_{code} - 1)\tau_{per} + \tau_{act}$ amount of time, the wake-up signal can be driven high. In the worst case, the WuRX active duration is shifted to the very end of the transmitted bit period, such that the WuRX turns off as the next bit is transmitted. Here, it takes $N_{code}\tau_{per}$ amount of time to receive a wake-up. The bounded

amount of uncertainty in time for the wake-up is then:

$$L_{uncertain,BLDC} = N_{code}\tau_{per} - ((N_{code} - 1)\tau_{per} + \tau_{act}) = \tau_{per} - \tau_{act} \tag{3.3}$$

In heavily duty-cycled scenarios where $\tau_{per} >> \tau_{act}$, the uncertainty in wake-up latency is approximately a single bit period. Assuming a uniform distribution for the relative phase between the transmitted bits and WuRX bit period, the average latency is:

$$L_{avg,BLDC} = N_{code}\tau_{per} + \frac{(\tau_{act} - \tau_{per})}{2} \tag{3.4}$$

These equations assume that right after the sample, a bit decision is created. In the case of the wake-up receivers designed and referenced in this dissertation, the baseband clock period is equal the desired bit period. This means that at the rising edge of the bit period the data officially latches into the correlator instead right at the falling edge of the active duration. As a result, there is an added latency in the time it takes to make a bit decision which changes Equation 3.4 to:

$$L_{avg,BLDC} = N_{code}\tau_{per} + \frac{(\tau_{per} - \tau_{act})}{2} \tag{3.5}$$

Now the BLDC duty-factor, α_{BLDC}, can be rewritten in terms of these variables, given by:

$$\alpha_{BLDC} = \frac{(N_{code} + 1/2)(\tau_{su} + \tau_{int})}{L_{avg,BLDC} + (\tau_{su} + \tau_{int})/2} \tag{3.6}$$

If the code length is assume to be $>> 1/2$ and $L_{avg,BLDC} >> \tau_{act}/2$, then α_{BLDC} approaches the original definition, given that $\tau_{act} = \tau_{su} + \tau_{int}$.

Sensitivity, Power, and Latency Trade-off Space. Based on the way α_{BLDC} has been defined here, both average latency and integration time can be used to affect the WuRX average power consumption. It is less reasonable to adjust the wake-up code length as this would generally be fixed by network requirements. The three-way trade-off between sensitivity, power, and latency is unique to BLDC and allows for more reconfigurability across applications. However, there are bounds on the effectiveness of each trade-off. If the objective is to lower P_{avg}, then the bit period, and therefore latency, can be increased to a large extent. If the WuRX is active power dominated,

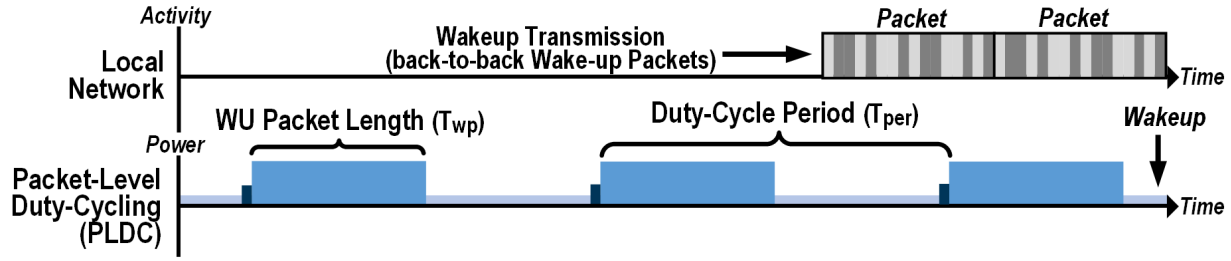


Figure 3.2: Packet-level duty cycling concept illustrated in time domain. All wake-up data is captured in a single-shot and a decision whether or not to raise a wake-up flag is made before deactivating.

then doubling the latency will approximately halve the average active power consumption. Eventually at low duty factors, P_{avg} will become dominated by the static floor power, P_{in}. The transmitter will also occupy a greater amount of the wireless resource.

Alternatively, the integration time can be reduced in order to lower power. Ideally, if the integration time is halved, then the sensitivity would be reduced by 3dB because integration bandwidth (therefore, noise) is increased by 2X. This trade-off requires that $\tau_{int} >> \tau_{su}$ but is bounded when τ_{int} approaches τ_{su}. In this case, the startup time limits the extent of lowering the duty factor. Furthermore, quantization limits from digital control methods also affect resolution and dynamic range. It is important that enough time resolution exists to tune τ_{su} while providing enough range to support maximum latency on the order of seconds. Similarly, resolution on analog IF and BB bandwidth tunability impacts sensitivity performance across the three-dimensional operating space.

3.2.2 Packet-Level Duty-Cycling (PLDC)

The packet-level duty-cycling method effectively stretches out the active and inactive time periods used in bit-level duty-cycling such that the receiver should remain on for at least a single packet's worth of time including start-up time, this sum equalling ($\tau_{su} + N_{code}\tau_{bit}$), and at most $\tau_{su} + 2N_{code}\tau_{bit}$, where τ_{bit} is the bit duration (Fig. 3.2). Packet-level duty-cycling is more common

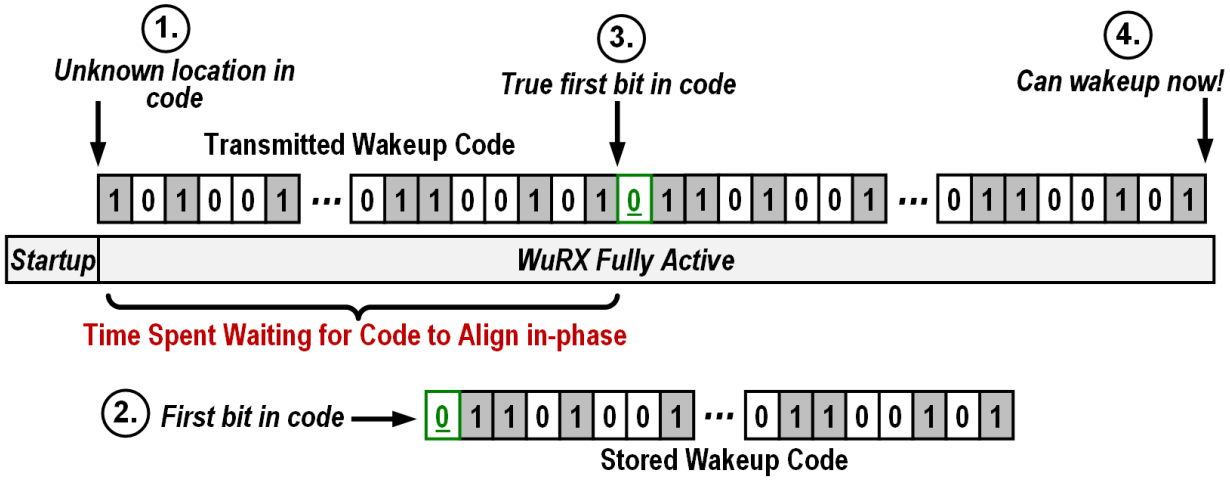


Figure 3.3: Because the WuRX asynchronously turns on and doesn't lock with the first bit in a wake-up code, it must wait for the code to first align before it can be properly detected.

than BLDC and variants are used in Bluetooth and the B-MAC and X-MAC protocols. A version of it was also demonstrated as a feature for the WuRX in [51] enabling wake-up from BLE and Wi-Fi transmitters for $4.4\mu W$. One reason for its popularity is because it is simpler to scale the number of packets transmitted by an arbitrary TX than scale the data rate, while also coordinating that change across a network.

An important point previously mentioned was that a PLDC WuRX might need to stay on up to nearly two times the length of the wake-up code. The reason for this variation in on-time stems from the asynchronous start time of the WuRX relative to the start of the transmitted wake-up code. Figure 3.3 depicts this problem by showing an example wake-up transmission and locally stored wake-up code. At Point (1), the WuRX has finally started up and receives the first bit from a transmitter. Point (2) is the first bit in the local wake-up code and is shown in green. In order for the wake-up signal to be driven high, the received data must match in phase with the stored code. But because the data at Point (1) is not in phase, the WuRX must wait until Point (3) to align with the first bit (again, in green to denote the first bit in the wake-up signature). Then at Point (4), the WuRX receives all the necessary data to signal the wake-up flag. As a result, the basic PLDC implementation must assume worst case condition and stay on for two packet time durations at

most. Because the unit duty-cycle period is based on the wake-up packet, the method to scale α_{PLDC} is by adding an integer number of repeated wake-up packets to the transmitted data. The result is then

$$\alpha_{PLDC} = \frac{(\tau_{su} + 2N_{code}\tau_{bit})}{(MN_{code}\tau_{bit})} \tag{3.7}$$

where M is the number of packets sent during the duty-cycle period and $M \geq 2$. The number of transmitted packets must be $M + 1$ due to the asynchronous behavior of the PLDC WuRX.

As the reader will find out in later sections, this extra on-time created by the asynchronous nature of the WuRX can be overcome such that the active time is reduced to exactly $\tau_{su} + N_{code}\tau_{bit}$. This case is depicted in Figure 3.2. For the remainder of this discussion, the reader should assume this is true until demonstrated in later sections. As a result, the updated version of Equation 3.7 is

$$\alpha_{PLDC} = \frac{(\tau_{su} + N_{code}\tau_{bit})}{(MN_{code}\tau_{bit})} \tag{3.8}$$

The numerators of Equations 3.2 and 3.8 both contain τ_{su} followed by the useful active duration of the WuRX. Interestingly, a PLDC WuRX has an efficiency gain here because τ_{su} occurs only once per period while a BLDC WuRX accumulates $N_{code}\tau_{su}$ over the course of enough time to demodulate a wake-up signal. This will be important in upcoming comparisons.

Wake-up Latency Considerations. Similar to the BLDC scheme, PLDC can also have a calculated average latency. However, duty cycling at the packet level requires that all the data must be captured at the same time, which significantly increases the amount of uncertainty in average latency relative to the BLDC case. The lowest possible latency in a PLDC scheme is when the WuRX activates simultaneously with the transmitter: $N_{code}\tau_{bit}$. The worse case latency is if the wake-up flag isn't raised until the very end of the transmission period given by $(M+1)N_{code}\tau_{bit} - \tau_{bit}$. The uncertainty for PLDC average latency is then

$$L_{uncertain,PLDC} = MN_{code}\tau_{bit} - \tau_{bit} \tag{3.9}$$

leaving the average latency to be approximately:

$$L_{avg,PLDC} = N_{code}\tau_{bit}(1 + \frac{M}{2}) - \frac{\tau_{bit}}{2} \tag{3.10}$$

Based on Equation 3.10, the latency of the WuRX is manipulated by varying M. As M grows, the ideal average latency can be approximated by $(MN_{code}\tau_{bit})/2$, which is half of the PLDC duty-cycle period. Compared to the bit-level alternative, this represents nearly 2x savings in average latency. However, there is greater uncertainty when the wake-up signal activates for a packet-level scheme (anytime during the long duty-cycle period) while in a bit-level scheme the variance is confined to that of a single bit. In short, using a PLDC WuRX provides approximately half the average latency as the same WuRX using BLDC, but significantly more uncertainty in when the wake-up signal would be driven high.

The previous explanation concerning the best case and worst case latencies for a PLDC WuRX is a close approximation. Depending on how the received data (which could be a valid wake-up code or invalid bit decisions from a source such as receiver noise) is correlated against the internal wake-up code, there is a finite amount of time for the PLDC WuRX to determine if a flag must be raised. As a result of this added latency due to computation time, these best/worst case estimates are increased slightly. This increase in average latency is equal, in seconds, to $N_{code}\tau_{bit}/2$. In turn, this changes Equation 3.10 to

$$L_{avg,PLDC} = N_{code}\tau_{bit}(\frac{M+3}{2}) - \frac{\tau_{bit}}{2} \tag{3.11}$$

3.3 Within-Packet Duty-Cycling (WPDC)

Previously, it was mentioned that BLDC and PLDC represent the two primary methods for duty-cycling wake-up receivers. This was based on the observation that asynchronous duty-cycling requires periodic bursts of high-power activity and longer durations of lower power operation. However, there is an inefficiency that can be exploited in a PLDC WuRX. Restating earlier points, wake-up receivers are useful for systems that require low-power consumption and are characterized by low wireless activity factors. In Figure 3.4, the depiction of a PLDC WuRX shows that for a large majority of the time the transmission channel remains idle, as the wake-up network activity is sparse. In this case, the PLDC WuRX wastes energy every time it demodulates data be-

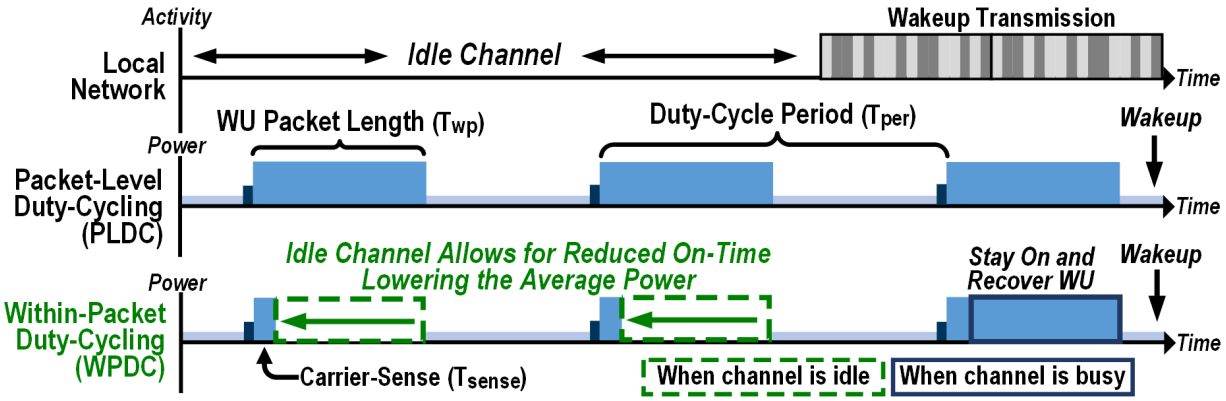


Figure 3.4: The within-packet duty-cycling method reduces the active time of the WuRX saving average power.

cause information is rarely sent. To improve upon the standard packet-level duty-cycling scheme, the within-packet duty-cycling (WPDC) method is proposed which employs a carrier-sense mechanism to help reduce the average power consumption without affecting latency. It does this by determining if the channel is ”idle” right when it first turns on, over a length of τ_{sense} seconds. If so, then it shuts the WuRX down early until the next duty-cycle period. If the channel is ”active” then it will remain on to collect any potential data that is being transmitted. All latency characteristics of PLDC transfer to WPDC so no further analysis is needed as $L_{avg,PLDC} = L_{avg,WPDC}$.

Traditionally, the carrier-sense function was used in ”carrier-sense multiple access” (CSMA) media access control (MAC) layer protocols to prevent contention between different system trying to access the wireless medium simultaneously [52]. If activity was heard on the local channel, then any node with data to send would ”backoff” for some period of time before attempting to retransmit. CSMA has been implemented in a wide variety wireless protocols and has been expanded upon for many years. Applying carrier-sense-like functionality to help reduce receiver power consumption was first applied in [44] followed by other works, such as [45]. However, this research applied only to COTS high-power receivers with limited abilities. It was also a rather inefficient implementation because the protocol was limited by the amount of control provided by the COTS receiver. The application to the wake-up receiver context is a far more effective method by which

to enable event-driven, low-power wireless capabilities. By controlling every aspect of the receiver in hardware, better power savings can be achieved.

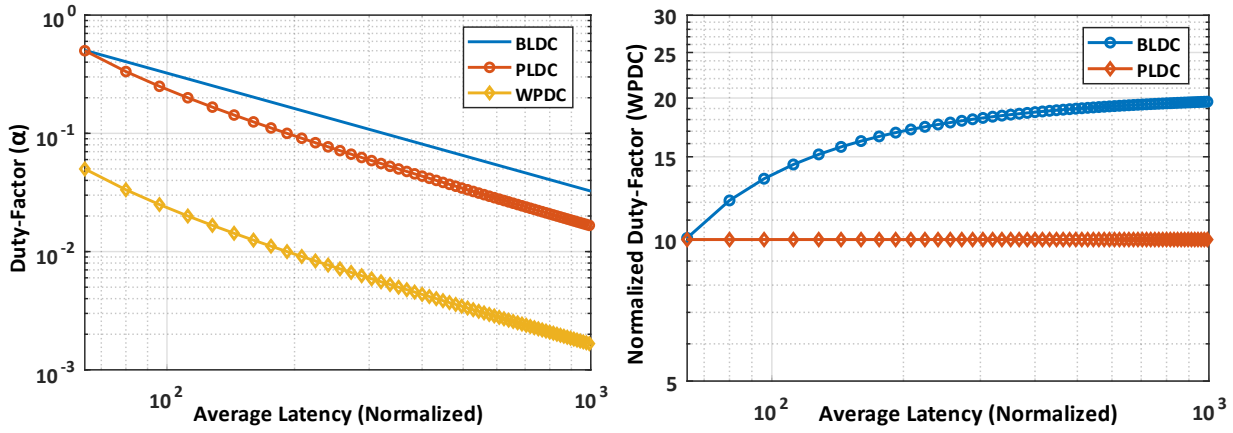
Because the carrier-sense mechanism adjusts the active time of the WuRX based on channel conditions, α_{WPDC} can be written as follows:

$$\alpha_{WPDC} = \frac{\tau_{su} + \beta N_{code}\tau_{bit} + (1-\beta)\tau_{sense}}{MN_{code}\tau_{bit}} \tag{3.12}$$

where β represents the effective network activity factor and τ_{sense} is the necessary time to sense the channel. When the channel experiences no activity, $\beta = 0$ which maximizes power savings. As β increases, the WuRX power profile begins to look more like PLDC operation, eventually hitting a ceiling power when totally active. To enable sufficient power savings, $\tau_{sense} << N_{code}\tau_{bit}$. The factor β can be further broken down into several distinct components given by

$$\beta = \beta_{WU} + \beta_{noise} + \beta_{int} + \beta_{WU,other} \tag{3.13}$$

Each component is based on either true positive or false positive sources. The only true positive source of data comes from valid wake-up transmissions, β_{WU}. The next term, β_{noise}, represents the contribution of false data due to receiver noise. This should generally be the dominating factor given the low-activity factor of the application. Similarly, β_{int} is due to interference from other sources that isn't valid wake-up data. Given the finite SIR of these WuRXs, nearby channel activity could falsely trigger the carrier-sense operation. Both of these nonidealities increase the overall β and negatively impact average power. Other sources of valid local network activity, $\beta_{WU,other}$, appear as valid data, but don't wake up most of the deployed wake-up radios. This factor will scale based on anticipated local network size and is effectively the cost of overhearing other wake-up messages. However, even though this overhearing effect by the wake-up receiver can cause an increase in its average power, it generally won't scale closely to that of a PLDC WuRX. It is important to model the impact each of these factors when implementing a WPDC WuRX.



(a) (b)

Figure 3.5: (a) The WPDC method outperforms BLDC and PLDC methods across all latencies; (b) The relative improvement in effective power savings when using WPDC for $\tau_{sense} = N_{code}\tau_{bit}/10$. Model based on the assumption that $\tau_{su} = 0$.

3.4 Comparison of Duty-Cycling Techniques

Each duty-cycling method discussed so far has been defined with a respective duty-factor, α, and average latency, L_{avg}. Given that the major trade-off for duty-cycled WuRXs exchanges these two variables, with the above information a comparison of this trade-off can be demonstrated. This will show the relative effectiveness of these duty-cycling methods in these two domains if they were demonstrated on the same receiver hardware in an active-power dominated operating condition. To sweep average latency, Equations 3.5 and 3.11 are solved for τ_{per} and M, respectively. The number of repeated packets, (M), must be an integer number because only whole packets are sent. To couple BLDC to PLDC and WPDC in this model, the variables $\tau_{int,BLDC}$, $\tau_{b,PLDC}$, $\tau_{b,WPDC}$ are set equal to one another to assume an equal amount of time that bit energy is integrated. This also allows us to approximate that the equivalent RF sensitivities are equal. In reality, this is not exactly the case but is done here to simplify the comparison. In this model, there is no additional start-up time required for the receiver in question (τ_{su}) and the wake-up code length is 32b. For WPDC mode, $\beta_{WPDC} = 0$ and $\tau_{sense} = N_{code}\tau_{bit}/10$. The results of this model are shown in Figure 3.5.

Average latency is represented in a normalized fashion with respect to $\tau_{b,PLDC}$ and $\tau_{b,WPDC}$, which as stated before, is set equal to $\tau_{int,BLDC}$. The valid latency points for PLDC and WPDC exist at the marked locations in Figure 3.5 due to the required integer number M repeated packets based on code length. Each sequential point represents one added wake-up packet. At first glance, there are a few primary takeaways. The WPDC method significantly outperforms both BLDC and PLDC across the entire operating space, as shown in Figure 3.5 (a). The relative improvement due to WPDC can be found in 3.5 (b). Simply from this figure, the improvement over PLDC appears as a constant. This is verified by the following equation:

$$\alpha_{PLDC,WPDC} = \frac{\alpha_{PLDC}}{\alpha_{WPDC}} = \frac{\tau_{su} + N_{code}\tau_{bit}}{\tau_{su} + \gamma N_{code}\tau_{bit}} \tag{3.14}$$

where

$$\gamma = \frac{(1-\beta)\tau_{sense}}{N_{code}\tau_{bit}} \tag{3.15}$$

which represents a combination of both the fraction of the total packet used to sense the channel and the fraction of time the channel is considered idle. As a result, $\gamma = 1/10$ in this example which means the relative improvement is 10 agreeing with the modeling result. The relative improvement over BLDC denoted by $\alpha_{BLDC,WPDC}$ is characterized by a more complex relationship. It is simpler, however, to calculate the normalized duty-factor under the conditions of high latency. The first step to simplify this result can be found to be:

$$\alpha_{BLDC,WPDC} = \frac{\alpha_{BLDC}}{\alpha_{WPDC}} \approx \frac{N_{code}M}{\gamma L_{avg}} \tag{3.16}$$

While the variable M requires a ceiling function to be applied, in order to assume integer number of packets, in this case $N_{code}\tau_{bit} << L_{avg}$ so the resolution on the average latency is sufficiently large to remove this condition in this approximation. This leaves the final result to be:

$$\frac{\alpha_{BLDC}}{\alpha_{WPDC}} \approx \frac{2}{\gamma} \tag{3.17}$$

In this example with $\gamma = 1/10$, the asymptotic advantage of WPDC relative to BLDC is determined to be 20. This agrees with the result in Figure 3.5 (b) at very high latencies. Therefore, the advantage at high latency of PLDC relative to BLDC is the ratio of Equations 3.17 and 3.14 which is 2.

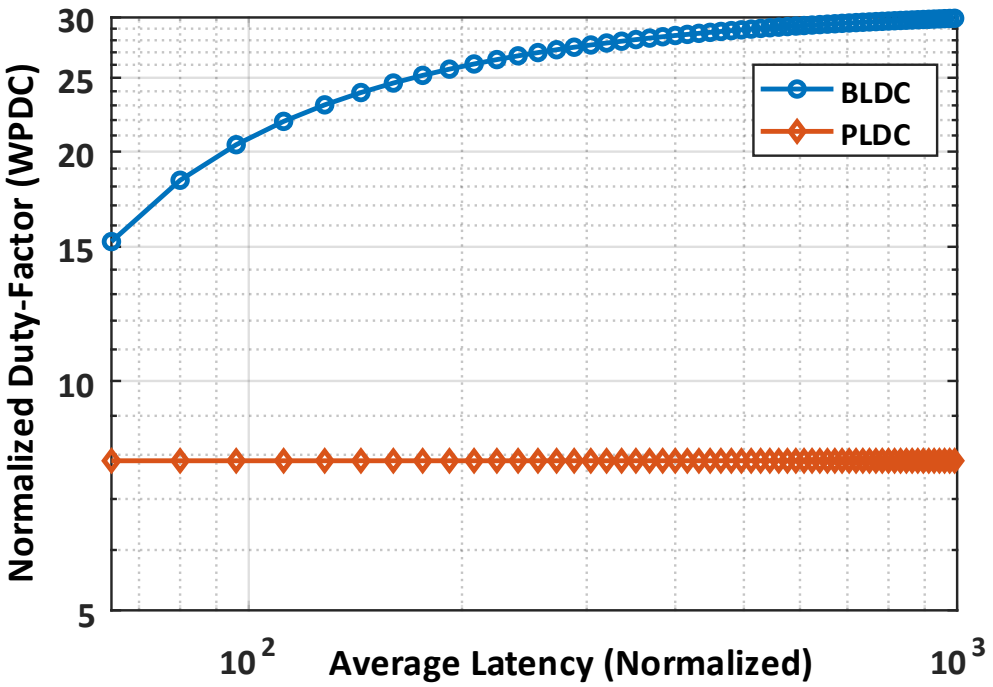


Figure 3.6: WPDC advantage under non-zero start-up time conditions ($\tau_{su} = \tau_{bit}$) reduces in comparison to PLDC but increases in comparison to BLDC.

This example did not include the realistic effects of the unavoidable start-up time required for WuRXs. In Figure 3.6, a revision is made to the model to show what happens when $\tau_{su} = \tau_{bit}$. As τ_{su} increases, several things happen. The separation between PLDC and WPDC begins to decrease as the magnitude of τ_{su} becomes more comparable to τ_{sense}. This occurs as γ in Equation 3.14 slowly becomes dominated by the increasing start-up time in both numerator and denominator terms. On the other hand, bit-level duty-cycling incurs the start-up penalty with every bit compared to once per packet. As a result, the advantage of BLDC distances itself from those of PLDC and WPDC.

3.5 Conclusions

This chapter discussed the importance of duty-cycling for low-power wake-up receivers. The two standard duty-cycling methods, PLDC and BLDC, were analyzed with respect to their impact on average power, via their respective duty-factors, and average latency. The proposed WPDC technique was similarly analyzed and then compared with a mathematical model to demonstrate power and latency advantages over the standard techniques. With this model, WPDC is able to always outperform both PLDC and BLDC. Irrespective of channel activity, BLDC has the least optimal trade-off amongst the three options and performs worse at high-latencies, which are often desirable in the application space to keep average power low.

An itemized list of the contributions in this chapter is as follows:

1. Proposed the within-packet duty-cycling (WPDC) method to significantly improve wake-up receiver power and latency performance compared to standard packet-level duty-cycling (PLDC) and bit-level duty-cycling (BLDC)

2. Calculated average power and average latency for the three duty-cycling techniques in order to quantify these key metrics and understand the relationships between them

3. Quantified the improvement of WPDC relative to the other two duty-cycling techniques by comparing power and latency performance

CHAPTER 4

DIGITAL BASEBAND TECHNIQUES

4.1 Overview

This chapter covers the theory and design of digital basebands for duty-cycled wake-up receivers. This chapter will begin with a quick introduction of digital baseband architectures and how they differ from bit-level to packet-level duty-cycled implementations. This will be followed by discussions and analyses concerning binary correlators and the proposal for a new type of correlator for use in packet-level and within-packet duty-cycling. False wake-up considerations are then mathematically modeled and compared against measurement results. Then, the impact of within-packet duty-cycling will be addressed by laying out how the associated carrier-sense algorithm functions as well as how to optimally achieve desired WuRX performance while optimizing baseband sensitivity. Next, an implemented carrier-sense multiple-access accelerator will be discussed. Lastly, gain and threshold control techniques for temperature and interference compensation will be addressed. The implementation of all these methods will be discussed in the following chapter that covers the design and measurements of four separate low-power wake-up receivers.

4.2 WuRX Digital Baseband Architectures

The digital baseband in a duty-cycled wake-up receiver is an essential part of the component and performs several critical functions. An important requirement for the digital is low static power consumption because it will contribute to the floor power and needs to still be running, in part, when the RF and analog blocks are turned off. The first essential function the digital

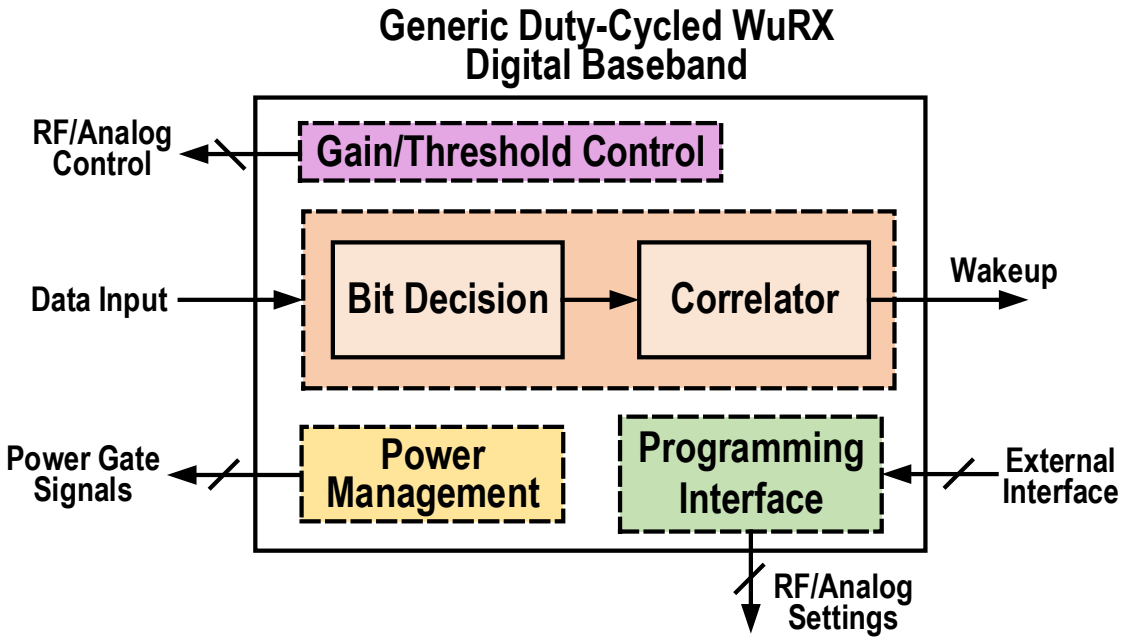


Figure 4.1: Generic layout of the digital baseband for a duty-cycled wake-up radio.

baseband performs is in the data path (Fig. 4.1). Information out of the analog baseband interfaces with a binary comparator or ADC to digitize voltage levels. This then enters a bit decision block which converts a multi-bit value to a logical '1' or '0' binary bit. Assuming a transmitted wake-up code with small synchronization error, the sampled and thresholded data represents the digital baseband's "best guess" of the original information. Because wake-up receivers are only ever searching for a unique address, this bit stream is then sent through a correlation block to determine how similar the incoming data appears relative to its local code. A programmable threshold can be applied to adjust the minimum required Hamming distance post-fabrication. When the two binary vectors match closely enough, then a wake-up flag is driven high. Duty-cycling each component in an adjustable fashion requires a programmable synchronous timing mechanism. This power management/timing block delivers power gating signals to RF and analog components across the chip to enable various forms of duty-cycling. To compensate for drift or saturation in the analog output, due to temperature or interference, gain and offset control can be implemented to re-tune the IF and BB gain settings and then reset the bit decision threshold to achieve a desired symbol distribution. Lastly, a programming interface is required in order to communicate with the chip,

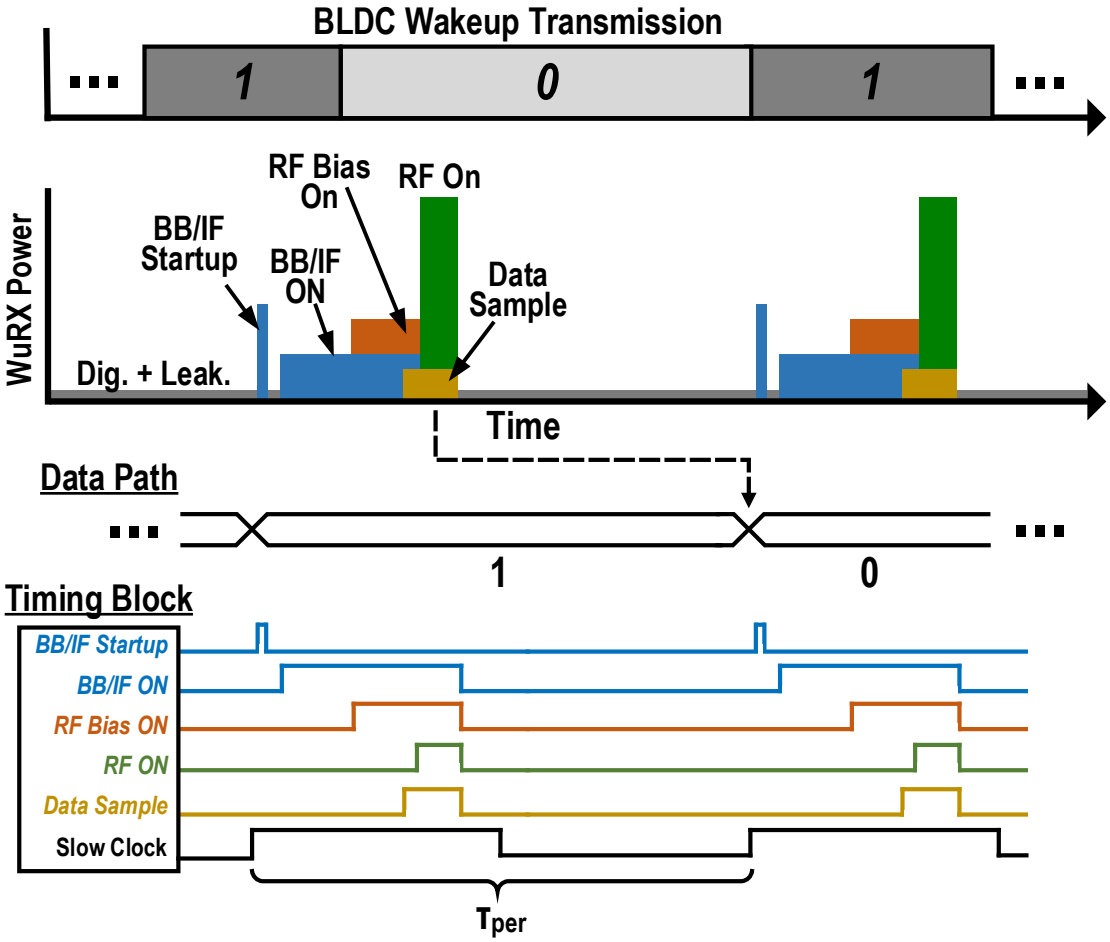


Figure 4.2: Time-domain view of power gating signals and data path operation for BLDC.

control and adjust settings, and read status information.

4.2.1 BLDC Timing

While a high-level block diagram view of bit-level and packet-level (inclusive of within-packet) duty-cycled wake-up radios appear similar, their time domain functions vary. The BLDC WuRX requires periodicity at the bit level and therefore requires a timer to count for as long as the bit period generating a ”slow clock” reference. At best, each component’s power gating signal has full controllability over the start and end points for when it should be enabled. To implement

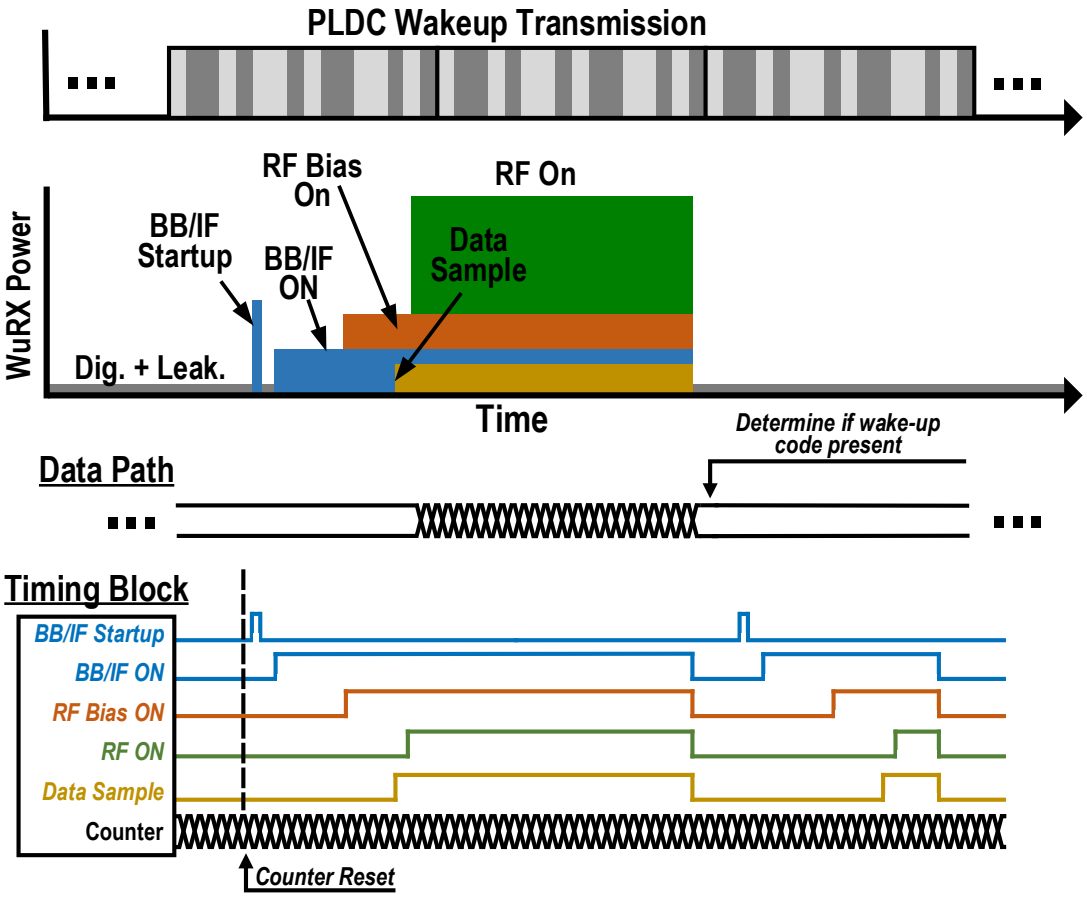


Figure 4.3: Time-domain view of power gating signals and data path operation for PLDC.

this, a central synchronous counter is used to set the slow clock period and then all power gating signals are assigned two thresholds to determine start and end count values. Staged startups are important in order to minimize the extra on-time required for blocks as seen in Figure 4.2. The next rising edge of slow clock latches in the new bit to the data path and the correlator reevaluates its calculation. These processes repeat with every slow clock cycle.

4.2.2 PLDC Timing

Packet-level duty-cycling changes the time domain operation of the digital baseband slightly (Fig. 4.3). A counter is still needed to define the periodicity, but should be set to an integer multiple

of the packet length and does not generate an actual slow clock because the digital clock itself is used to set the bit rate. It also uses a ”one-shot” type action by recovering all the data in a single stream. When enough data has been captured, the correlator then looks at the data to see if the information contains the wake-up code.

4.3 Digital Binary Correlators for Wake-up Receivers

4.3.1 Background

Traditional digital communication theory recognizes the matched filter as the optimal linear filter for maximizing SNR power, for a given transmitted symbol [46]. Interestingly, the cross correlation function produces the same output but only at symbol time T. Over this symbol time, the output $z(t)$ is defined as

$$z(t) = \int_0^T r(\tau)s(\tau)d\tau \tag{4.1}$$

where $r(\tau)$ is the received signal and $s(\tau)$ is the prototype signal (in this case reference code). Qualitatively, correlation provides a large output when both inputs are very similar, while providing a low output when they are dissimilar. Often, high-performance receivers will implement entire correlator banks in order to compare the incoming waveform with the set of allowable symbols to best determine which symbol was received. Correlator banks can also be used to help mitigate multipath effects through the use of rake receiver architectures. For low-power wake-up receivers, correlation is not used to identify the most likely symbol sent but rather to identify whether or not a wake-up signal was transmitted for ”this” node in particular. To quickly recap, the wake-up receiver uses the internal wake-up code to compare against the incoming bit stream in order to determine if they match closely enough. If they do, then the wake-up flag should be driven high in order to alert the main system. Otherwise, it should remain low. A certain level of error tolerance can be added to the code in order to prevent significant loss in sensitivity, as the probability of perfectly resolving every bit in a wake-up packet (also known as the packet error ratio) is far higher than the probability of incorrectly deciding one bit (bit error ratio).

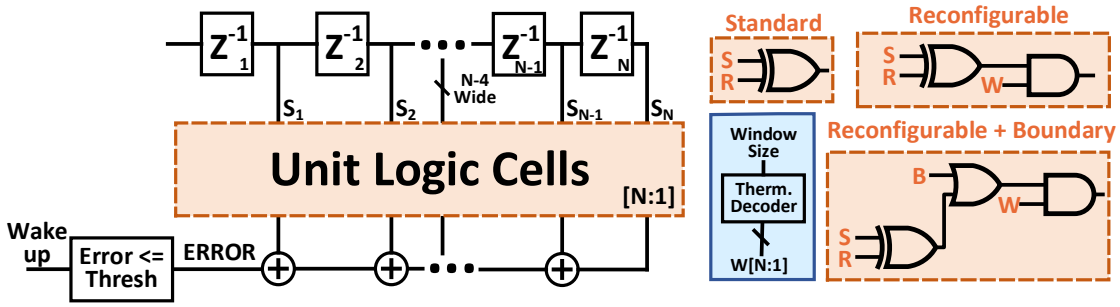


Figure 4.4: Generalized linear correlator architecture with three options for unit logic cells.

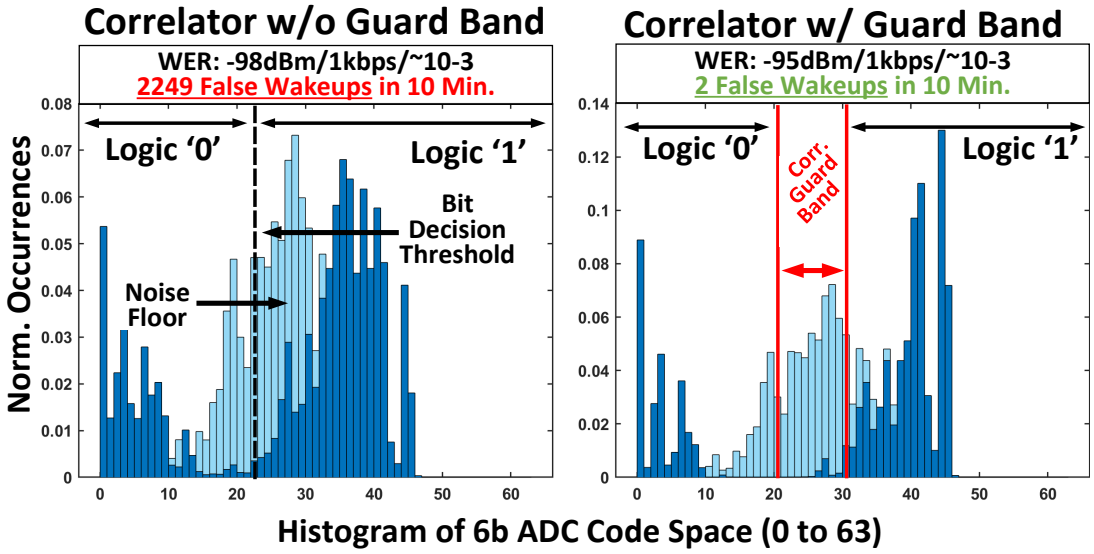


Figure 4.5: Measurement of correlator guard band impact on the false wake-up ratio.

4.3.2 Correlator Architectures

Linear Correlator. The correlator's primary function in the wake-up radio is to compare the binary reference code vector against the signal vector, R and S respectively each of length N (Fig. 4.4). Some set of similar single-bit logic cells then takes the two vectors and produces an N-bit output which represents the distance, or more simply error, between the two vectors. All error bits are summed and thresholded to determine if a wake-up flag should be issued with the current data input. As data flows through the correlator, the old data needs to move somewhere. A standard linear correlator performs a linear shifting operating which acts in a FIFO manner getting rid of the oldest data first. There are three specific types of unit logic cells implemented in these

works. The first type performs a standard error calculation by utilizing an logical XOR gate. If the input bits are similar then there is no error. The second type enables a reconfigurable length correlator by defining a window size variable which enters a thermometer decoder to gate the error contribution from a segment in the correlator. The XOR gate to determine error is connected to a logical AND gate which performs the gating function. When a 32b correlator is desired to be realized out of a synthesized 64b correlator, the final 32 unit logic cells have their window setting driven high by the decoder to effectively turn them off and not contribute any error result. The user then disregards the section of the wake-up code associated with the most significant 32b, in this case.

The last type is a reconfigurable correlator that also allows for boundary region detection. This requires the use of an ADC because boundary information is based upon whether or not the sample is within or outside of the defined boundary based on ADC code space. There are two data inputs in this unit cell. The first is the standard signal line and the other is the boundary bit for that bit in the received data, B. Both bitwise errors and guard band errors are taken into account to compute the final distance between received bits and wake-up code. The addition of a guard band allows for more precise tuning of the desired false wake-up ratio, not achievable with only a correlator, but at the expense of sensitivity. This technique helps reduce the false wake-up ratio as shown in Figure 4.5. The receiver noise histogram is shown in a light shade centered around ADC code “27”. The skewed receiver 0/1 symbol histogram is shown in a darker shade and offset with a mean approximately at ADC code “22”. This is primarily due to mismatched IF path gains. An unbiased 16b code and bit rate of 1kbps are used in this experiment, in the third WuRX mentioned in the next chapter. For an achieved sensitivity of -98dBm with 10^{-3} WER, the receiver falsely woke up to noise 2249 times in a ten-minute period. By applying the guard band from ADC codes “20” to “30”, a majority of the noise histogram is rejected as error bits reducing the false wake-up ratio to 2 per 10 minutes. As a result, the receiver requires 3dB more input power to achieve the same WER.

A bit-level duty-cycled WuRX uses a linear correlator such as this. With each slow clock cycle, it sends a new bit into the correlator and a comparison is performed to calculate whether or

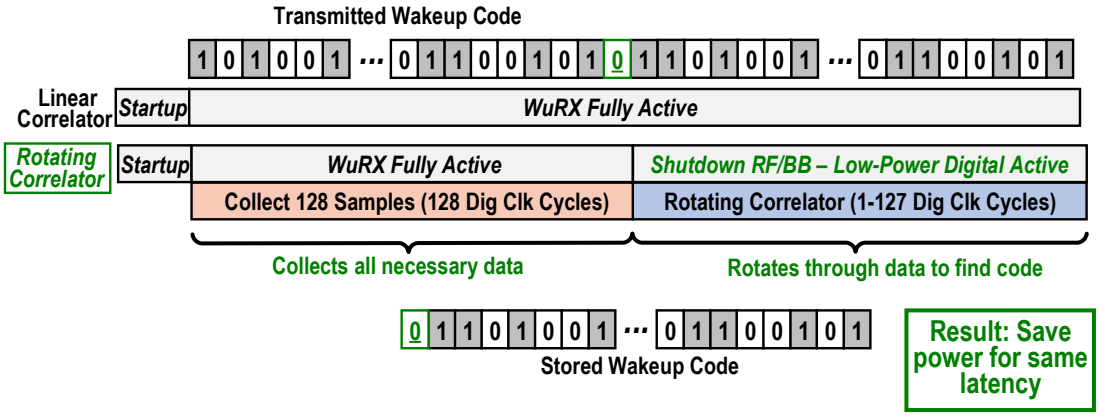


Figure 4.6: The rotating correlator architecture improves active power by up to 2x compared to the linear correlator.

not a wake-up should occur. A packet-level duty-cycled wake-up receiver can operate similarly, but with a slight twist. All the data is collected in a one-shot fashion, however because of the transmitter/wake-up receiver phase offset problem presented earlier the radio needs to remain on for up to two full packet periods. In this case, data should be loaded into the correlator for the first N bits and then the correlator logic should be turned on and kept running for the next N bits before finally turning off. If a wake-up transmission is occurring during this time period, then its possible to receive the wake-up in less than $2N\tau_{bit}$ seconds, but this happens infrequently. As a result of the low wake-up activity factor these kinds of IoT nodes will encounter, the wake-up radio will by default have an average active time twice of what was anticipated also increasing the average power.

Rotating Correlator. By reviewing Figure 3.3, we can see a specific behavior to take advantage of in the relationship between the transmitted code phase and the wake-up radio local code phase. Because the transmitted data is sent back-to-back in the same order, we can easily just record the exact number of bits required to fill up the correlator, N, turn off the high-power RF and analog components, then rotate through the saved data to search for the wake-up code as it would have the same effect as if we were listening to it live (Fig. 4.6). However in this fashion we reduce the active time from $\tau_{su} + 2N_{code}\tau_{bit}$ to $\tau_{su} + N_{code}\tau_{bit}$ allowing us to save up to 2x active power.

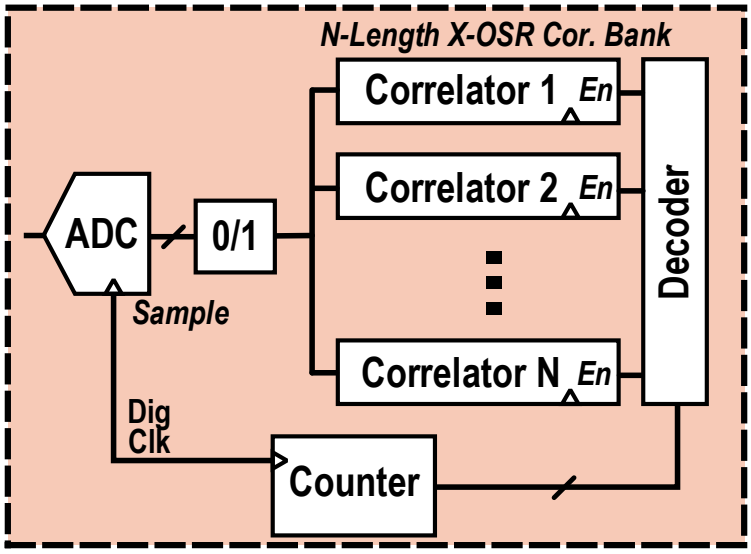


Figure 4.7: The oversampling correlator architecture allows one to get around improperly decided bits due to TX/WuRX clock synchronization and drift problems.

This is most impactful in low-latency applications when the WuRX is primarily active power dominated. The architecture for this correlator is the same as linear correlator, with the slight difference that when a counter matches the correlator length it should stop injecting new data in and rather circulate the data internally while performing the correlator calculations.

4.3.3 Oversampled Correlator

So far, we assume that one sample of the comparator or ADC leads to a single bit decision and therefore bit sent into the correlator. However, the asynchronous nature of the communication scheme to wake-up the receiver means that the TX and WuRX clocks won't be perfectly aligned in frequency or phase leading to drift. At some point there will be crossings between TX symbol times and when the WuRX samples that data. This unfortunately leads to symbol corruption as no clock recovery is utilized here. Instead, by oversampling the data by some rate, X, bounding the drift, and using a correlator bank the wake-up radio can get rid of this problem and not have to worry about extra drift (Fig. 4.7). This is a rather simple solution to implement for PLDC and WPDC

wake-up receivers, however for bit-level types this requires that they actually turn on X-times as often meaning increase the bit rate by the oversample rate. This incurs a power penalty, which could be more than tolerable in some circumstances. The BLDC wake-up radios demonstrated for this work use single-sampling while the PLDC/WPDC WuRX presented in future chapters is oversampled by a factor of four.

4.4 False Wake-up Considerations

The objective of the correlator is to enable individual wake-up addressing, while sufficiently suppressing false wake-ups (positives) to minimize their impact on system power. For binary correlators, received bits are latched serially through a shift register followed by parallel exclusive OR and summation operations to determine the distance, or error, between received data and unique wake-up address, of length N with M 1's in the code. If this error is low enough, then a wake-up flag will be raised. Often, an error tolerance, X_0, greater than zero is applied to allow for several incorrect bit decisions, resulting in a slight sensitivity increase due to coding gain. However, this is quickly limited by an increasing number of false wake-ups (positives). In low-activity networks where wake-up receivers are primarily used, the majority of bit decisions are generated by receiver noise, which can produce false wake-ups. For OOK-based WuRXs, the threshold is set such that the probability of observing a logic "1" is relatively low compared to observing a "0", which are P_1 and P_0 [53]. In the discussed architectures, energy is received in both symbols. In some cases, the false wake-up ratio is a desirable way to quantify false positives at is irrespective of time. In other situations, the false wake-up rate is better when quantifying how many per hour there are, for example. The conversion between one and the other is simply a factor of frequency, related to the duty-cycling method period.

4.4.1 Bit-Level Duty-Cycled Context

For a bit-level duty-cycled WuRX, the total number of possible wake-up decisions in an hour is $3600 f_{per}$, where f_{per} is the data rate and the inverse of τ_{per}. The probability that a false wake-up will occur is related to both the probability of a certain sequence of bits occurring, products of P_1 and P_0, and the number of ways that particular sequence can generate a false wake-up, based on the wake-up code and error tolerance. Because a wake-up is still allowed with less than X_0 error bits in the correlator, the probability of a false wake-up given X_0 tolerance includes the summation of the false wake-up probability for $X = X_0, X_0 - 1, \ldots, 0$. For an arbitrary code weight and bit probabilities, the resulting false wake-up ratio (FWR) per hour can be determined to be

$$FWR = 3600 f_{per} \sum_{X=0}^{X_0} \sum_{x=0}^{X} \binom{N-M}{X-x} \binom{M}{x} P_1^A P_0^B \tag{4.2}$$

where $A = M - 2x + X$ and $B = N - M + 2x - X$. The inner summation accounts for the number of ways that errors can be allocated to '0' and '1' based on how biased the code is. The outer summation sums across each total number of allowable errors so the false rate can be determined from 0 to X errors, which could all flow into the correlator.

In a simplified case, the decision threshold is set on the noise profile such that $P_1 = P_0 = 1/2$. For an unbiased wake-up code (M = N - M), the expected false wake-up ratio per hour is reduced to

$$FWR = 3600 f_{per} \sum_{x=0}^{X_0} \binom{N}{x} \frac{1}{2^N} \tag{4.3}$$

An experiment to compare calculated versus expected false wake-up rates for a BLDC WuRX is shown in Figure 4.8. Based on Equation 4.3, the expected false wake-up ratio was mathematically modeled as a function of error threshold for three different correlator code lengths, which the bit-level duty-cycled WuRX could be digitally tuned to. The WuRX was then set to a bit rate of 1kbps, the RF input was terminated, and ten minutes of ADC data was recorded. By simulating the correlator and using the recorded data, an extrapolated false wake-up ratio per ten minutes was determined due to receiver noise, assuming the correlator doesn't reset after wake-ups. Measured false wake-up rates match closely with anticipated results.

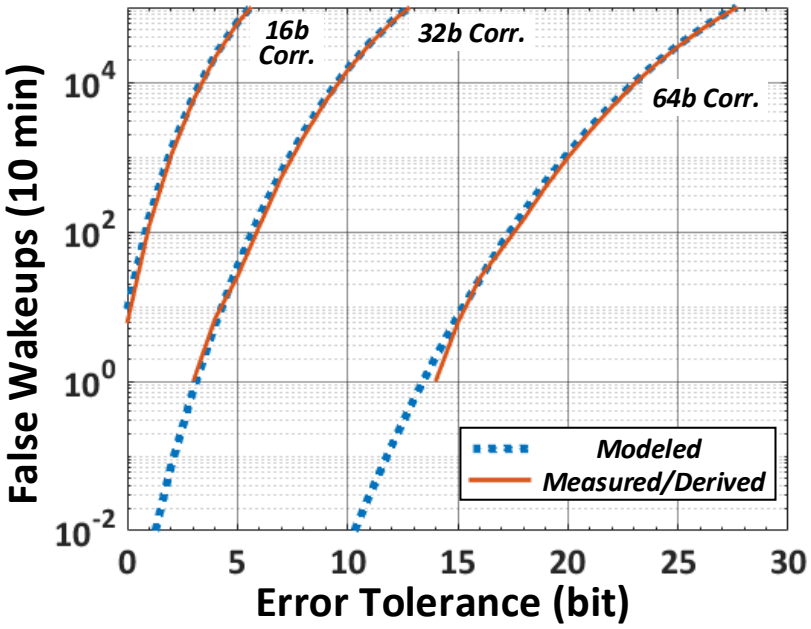


Figure 4.8: Math model vs. measured/derived false wake-up ratio for 16b, 32b, 64b correlator modes.

4.4.2 Packet-Level Duty-Cycled Context

For packet-level duty-cycled wake-up receivers, the false wake-up relationships are slightly different. In a packet-level duty-cycling scheme, the wake-up radio will need to observe N different correlation results per period and may also incur a higher false wake-up ratio if oversampling is used in the digital architecture with parallel correlator banks. For a PLDC WURX with an unbiased code, error tolerance of X_0, oversample ratio of OSR and decision threshold at the noise distribution mean, the false wake-up ratio is given by:

$$FWR_{PLDC} = OSR \times N \sum_{x=0}^{X_0} \binom{N}{x} \frac{1}{2^N} \tag{4.4}$$

The false wake-up relationships for the within-packet duty-cycled wake-up radio will be investigated several sections from now concerning how to choose optimal WPDC settings.

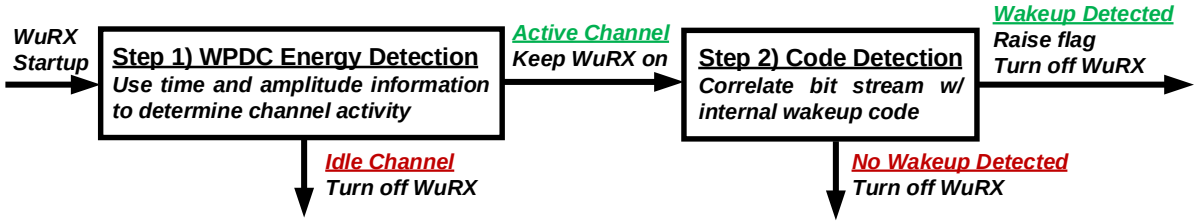


Figure 4.9: Two-step wake-up process for WPDC.

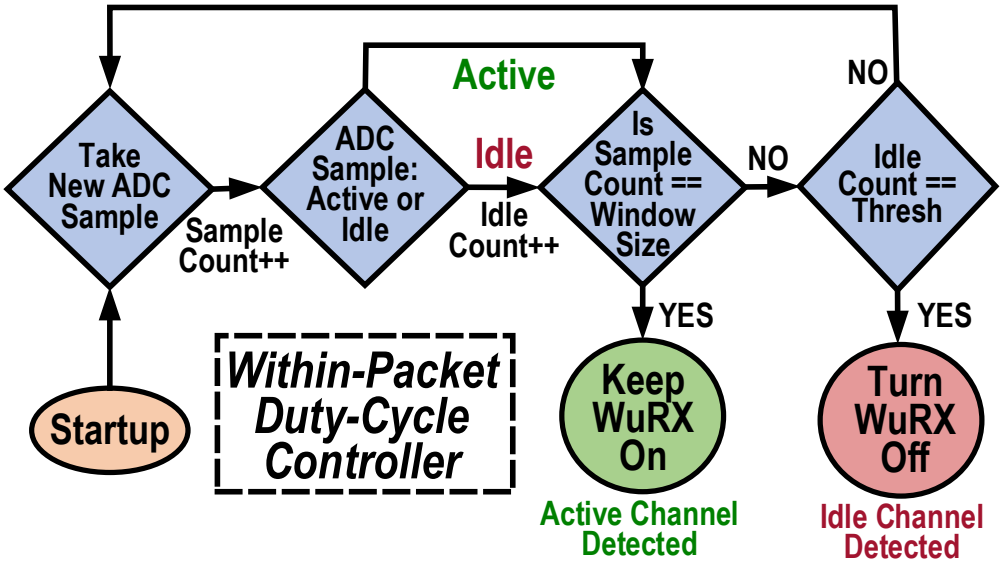


Figure 4.10: Within-packet duty-cycling state machine used to determine channel activity.

4.5 Within-Packet Duty-Cycling Algorithm

A within-packet duty-cycled wake-up receiver requires two separate steps in order to wake-up (Fig. 4.9). After startup, the radio goes through an energy detection phase. This step uses time and amplitude information in order to determine if the channel is active or idle. An idle channel condition turns off the WuRX early which reduces the average active time and therefore power. The active condition then allows the WuRX to move into the second phase which is based on code detection. Using the correlator, the recovered data is compared against the internal wake-up code. If there is a close enough match, then the wake-up flag goes high.

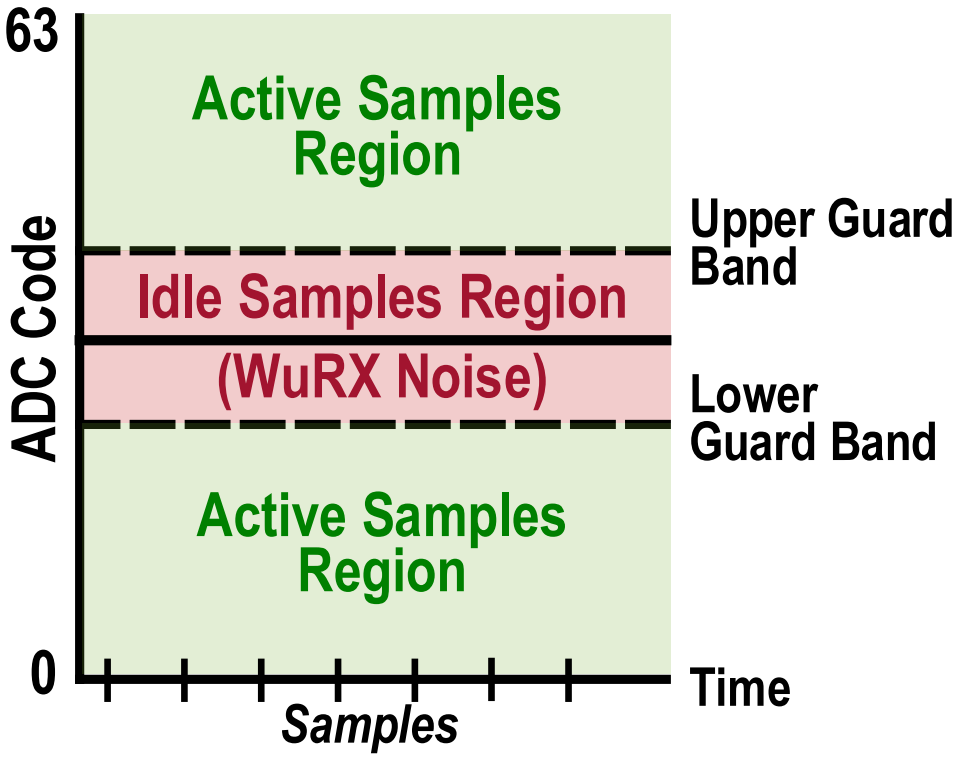


Figure 4.11: ADC code space split up into two regions based on guard band thresholds defining the ”active” and ”inactive” signal levels for carrier-sense functionality.

Let’s begin by going over the time-domain state machine that enables the within-packet duty-cycling. As shown in Figure 4.10, after radio startup the state machine begins by taking a sample from the ADC. A sample count variable is incremented to determine the total number of samples taken. Then the sample is determined to be either active or idle. If idle, then an idle count is incremented to keep track of this count. The first comparison checks whether the wake-up radio should stay on by seeing if the current sample count is equal to a ”window size” variable (W), which sets how long the carrier-sense operation should last. If an insufficient amount of time has elapsed, then the controller checks if the idle count is large enough to turn the WuRX off early in comparison to a threshold (T). These steps will repeat until one condition is met. In order to determine if samples are “idle” or “active”, the ADC code space is split in two using guard bands, similar to the situation with the earlier correlator architecture but instead implemented at the bit

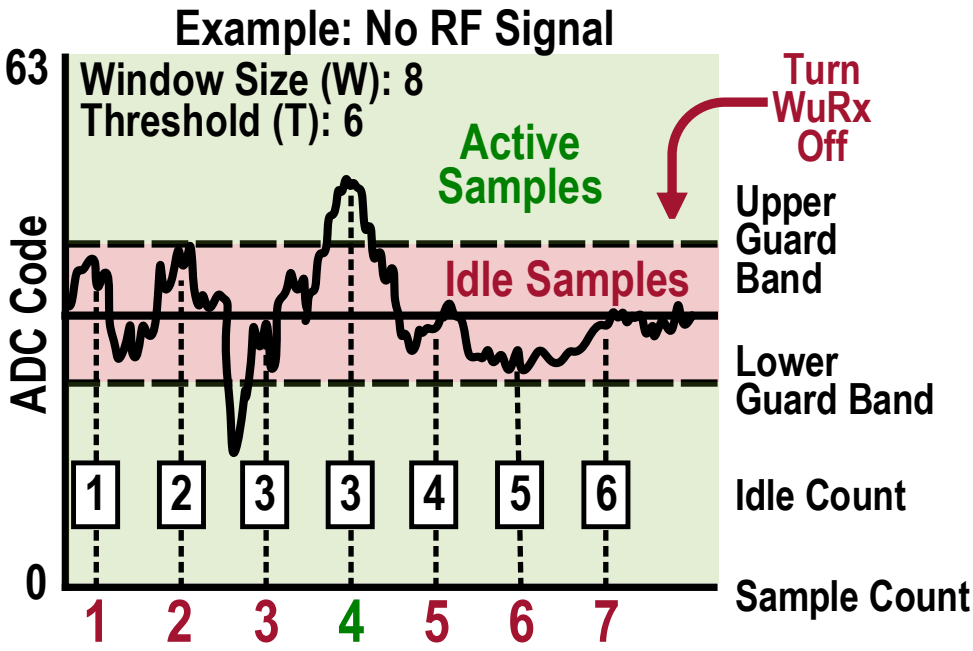


Figure 4.12: Example of the WPDC state machine operation under idle channel conditions.

decision stage (Fig. 4.11). These guard bands are two thresholds above and below the bit decision threshold (magnitude $\pm A$) which should be programmable in the receiver. This designates the region around the receiver baseband noise distribution as "idle" while the outside region is "active". In this figure, the space is for a 6b ADC which is the implemented size in the later described WuRX that uses this feature. The noise region exists in the center code space as this is the natural bias point for the output of the analog baseband. In order to lower the average power with WPDC, the window size needs to be reduced which in turn scales the carrier-sense time. Tuning WPDC variables requires careful attention and will be discussed in the next section.

Two examples for WPDC are shown in Figures 4.12 and 4.13 to further demonstrate the two primary conditions the carrier-sense mechanism is designed to detect. In the first example, the WuRX experiences a idle channel condition. Over the course of the programmed window size, equal to 8 here, most samples of the baseband output reside within the defined guard band region, with several noisy transient spikes occurring outside. Therefore, the idle count remains quite high. At the end of the window period, the idle count equals the required threshold in order

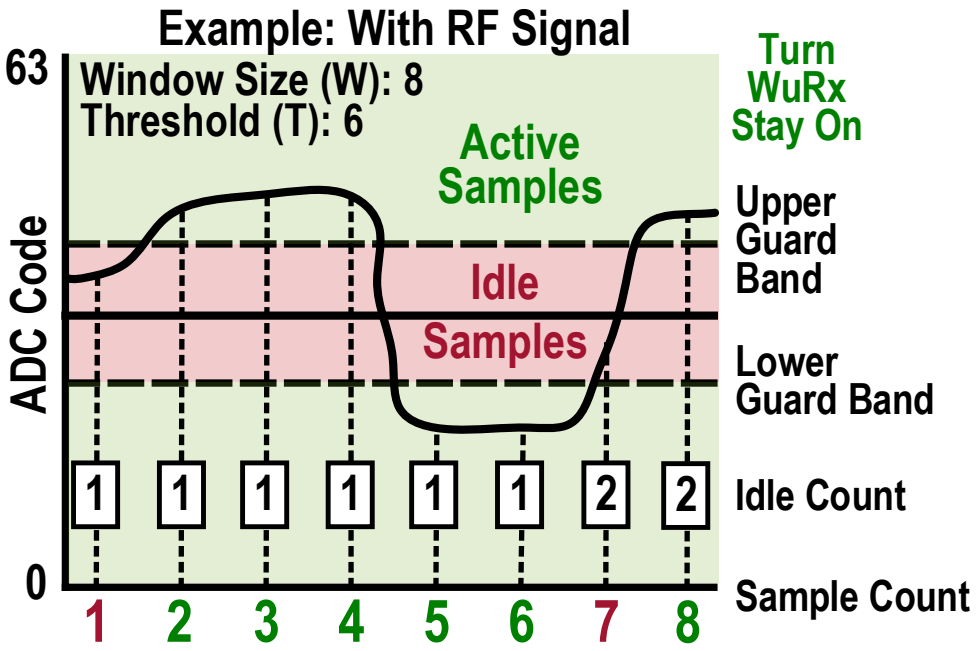


Figure 4.13: Example of the WPDC state machine operation under active channel conditions.

to turn the wake-up radio off. In the next example, the baseband output presents data to the digital. The magnitude of the samples are significant enough to reach over and under the guard bands to sufficiently suppress the idle count and therefore keep the WuRX on by the time the window size equals the sample count. At that point, the WuRX would move on to the code detection phase.

4.6 Choosing Optimal WPDC Settings

The WPDC mechanism introduces three new variables that impact key metrics: the window size, idle count threshold, and guard band level. These settings along with correlator length, (which is usually fixed) and error tolerance affect WuRX power consumption, latency, and the false wake-up ratio. These metrics are generally determined by the system requirements ahead of time, so the WuRX settings can be thoughtfully applied. As a note, the term “ratio” will be used instead of “rate” to normalize the given quantity against time in order to be more general. The discussed steps in this section include an example walk-through of the process of WPDC setup along with

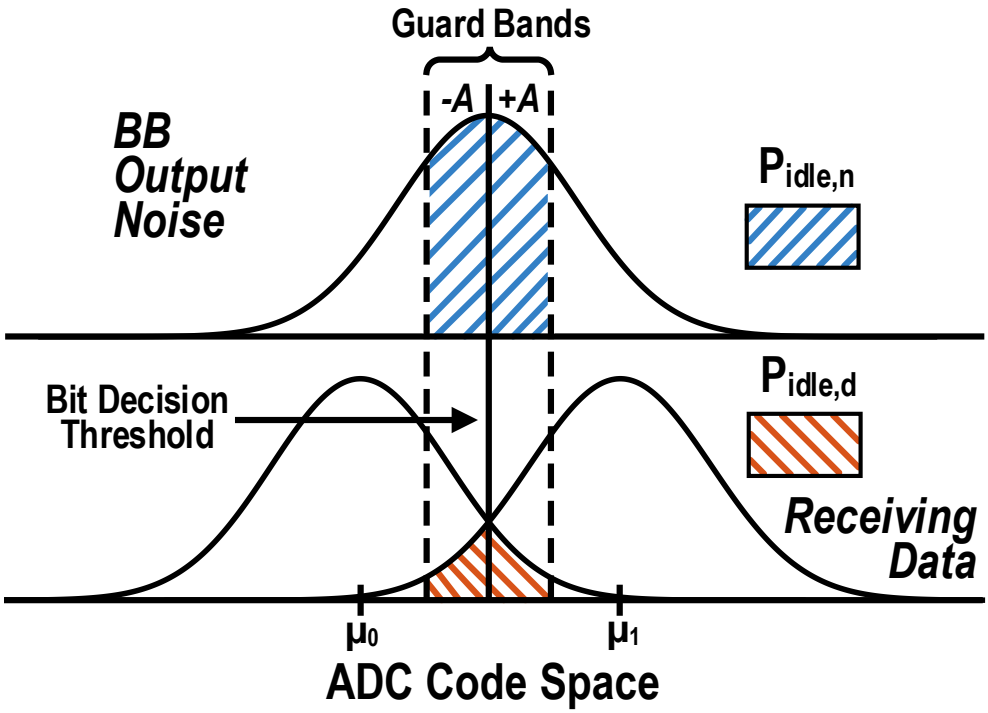


Figure 4.14: Representation of baseband output with and without data divided by WPDC guard band, which sets the idle region.

supporting measurement data.

4.6.1 WuRX Power and Latency

The trade-off between power and latency for WPDC must be modeled first as a function of the window size. This can be done combining Equations 3.1, 3.11, and 3.12, where $\tau_{sense} = W\tau_{bit}$ and assuming $\beta = 0$ for initial analyses. The value of W should be set to satisfy pre-determined power and latency requirements. An argument can be made to set W arbitrarily low because lower power is preferred, but this will come at the cost of further sensitivity degradation. The example used in this article will assume that a window size of 5 is desired based on these modeling results.

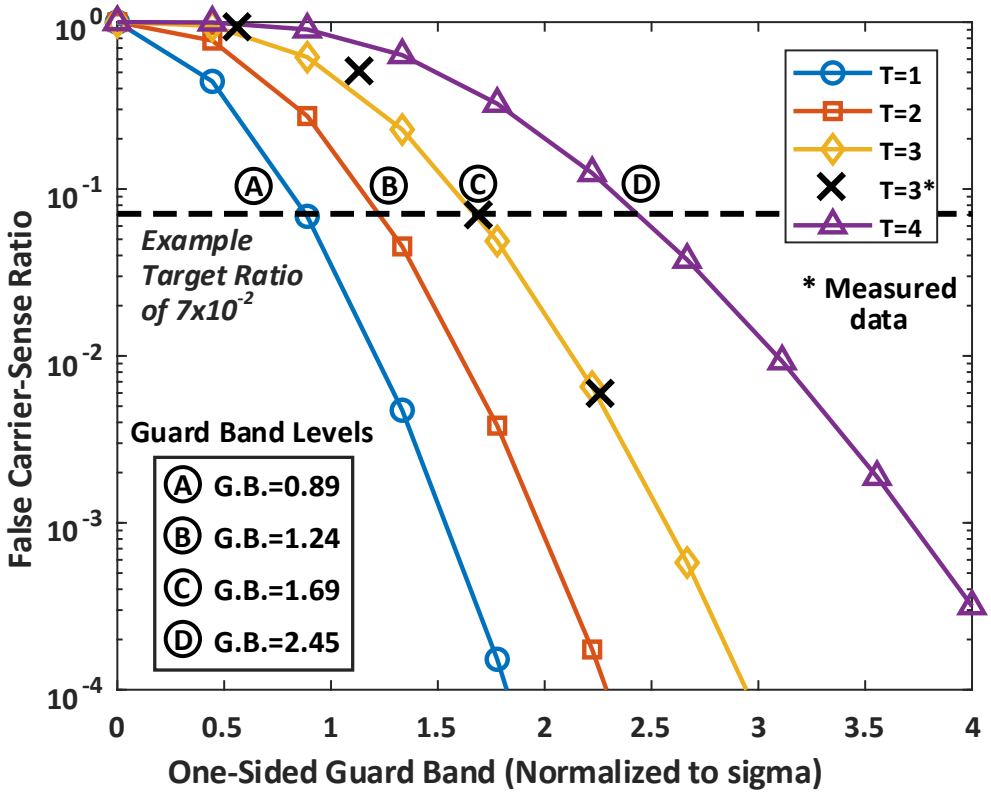


Figure 4.15: False carrier-sense ratio as a function of guard band extent and idle threshold. A target ratio of 7×10^{-2} creates several options for threshold/guard band pairs. (Measured data for $T = 3$).

4.6.2 False Carrier-Sense Ratio

Ideally, the carrier-sense mechanism correctly determines if the channel is idle or active every time the WuRX turns on. However, this decision is prone to false positives and negatives which impact the WuRX and system in different ways. False positive carrier-senses lead to an increase in dc-power as well as the false wake-up ratio, which in turn will lead to a greater system level dc-power. The false carrier-sense ratio (FCSR) can be calculated by:

$$FCSR = \sum_{x=0}^{T} \binom{W}{x} P_{idle,n}^{x} (1 - P_{idle,n})^{W-x} \tag{4.5}$$

where $P_{idle,n}$ is the probability that an ADC sample falls within the guard band idle region

due to noise when no data is sent. If the receiver's baseband output noise can be represented as a normal distribution (Fig. 4.14) and the guard bands are symmetric about the mean (also the decision threshold), then $P_{idle,n} = 1 - 2Q(A)$, where $Q(x)$ is the tail distribution function and guard band magnitude, A, is relative to the standard deviation of the noise profile (σ). The tolerable level of the false carrier-sense ratio should be determined by incorporating this factor into the same model as Step 1 and setting $\beta = FCSR$. A particular ratio should be chosen based on the result. Then, a second model needs to be developed for the false carrier-sense ratio as a function of guard band magnitude for different values of the idle threshold (from $T = 1$ to $W - 1$), which generates $W - 1$ new curves. For every line representing a different value of T, the point that corresponds to the desired false carrier-sense ratio should be marked. Then, the associated guard band magnitude for each curve can be determined. With a window size of 5, Fig. 4.5 shows modeled result for values of $T = 1 - 4$, where the associated guard band value is denoted by A through D. These are the four options to choose from for T and guard band magnitude. They are needed in upcoming steps. Several measured data points are presented on the curve where $T = 3$ to demonstrate agreement with expected results when applying a controlled noise source to the ADC input.

4.6.3 Desired False Wake-up Ratio

False wake-ups will impact a system differently based on the intended application, so the tolerable level of expected false wake-ups should be predetermined. The false wake-up ratio (FWR) for a PLDC WuRX was given in Equation 4.4. The WPDC carrier-sense mechanism however reduces this ratio by a factor proportional to the false carrier-sense ratio, significantly improving the false wake-up performance over PLDC. The resulting false wake-up ratio for WPDC is:

$$FWR_{WPDC} = FCSR \times FWR_{PLDC} \tag{4.6}$$

Using Equation 4.6, a plot can be made showing false wake-up ratio as a function guard band

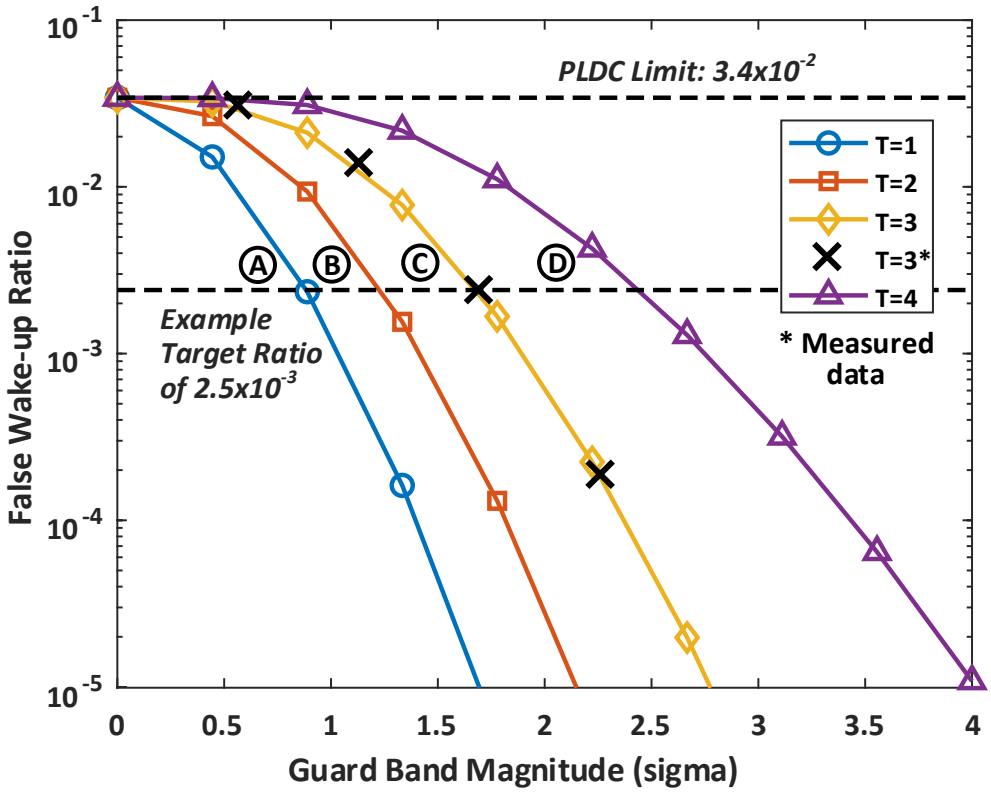


Figure 4.16: False wake-up ratio relative to guard band and idle threshold. (Measured data for $T = 3$).

for the same values of T as in Figure 4.5. The desired code size and error tolerance is also taken into account. This example considers the same correlator used in this work (32 bit) with an error tolerance of 6 bits. The guard band points saved from the plot in Step 2 can be marked on this new plot (Fig. 4.16). Because the false wake-up ratio is proportional to the false carrier-sense ratio, points A through D will have the same false wake-up performance as shown by the horizontal dashed line. It is important to check that the resulting FWR is less than the target value. The desired ratio can be achieved by manipulating the error tolerance if it isn't already satisfied. Wake-up performance for PLDC mode is static because guard bands aren't applicable. In WPDC mode as guard bands increase, it's harder for noise to appear as a valid carrier signal which improves the ratio. By this point, code size and error tolerance are fixed leaving the idle threshold and guard band left.

4.6.4 Idle Threshold and Guard Band

The final step is to determine the appropriate values for the idle threshold and guard band magnitude to maximize WER sensitivity for the worst allowable BER. As a result of using guard bands, a certain amount of signal energy is required to "reach" above threshold in order attain the same WER sensitivity. Generally, a target WER of 10^{-3} is chosen as reference. The wake-up error ratio is dependent on two factors: false negative carrier-sense ratio (referred to here as missed carrier-sense ratio, MCSR) and false negative wake-up detection ratio (traditional WER). A missed carrier-sense occurs when valid data is sent, but not detected by the WPDC algorithm. A false negative wake-up is when the carrier-sense trips, but the code isn't found in the correlator while data is sent. Thus, the missed carrier-sense ratio is determined by

$$MCSR = \sum_{x=T+1}^{W} \binom{W}{x} P_{idle,d}^{x} (1 - P_{idle,d})^{W-x} \tag{4.7}$$

Therefore, WER is approximated by:

$$WER = MCSR + (1 - MCSR) \times \sum_{x=X+1}^{N} \binom{N}{x} P_{b}^{x} (1 - P_{b})^{N-x} \tag{4.8}$$

The value $P_{idle,d}$ is the probability that when data is being sent the baseband waveform is sampled inside of the guard band region incorrectly decided as an idle sample (Fig. 4.14). When data is generated at the baseband as a polar NRZ-L signal, the receiver noise profile shifts positive and negative relative to the centered noise distribution. If we assume an even distribution of "1"s and "0"s, this probability can be calculated as:

$$\begin{aligned} P_{idle,d} &= \frac{1}{2}[\Phi(A - \mu_1) - \Phi(-A - \mu_1)] + \frac{1}{2}[Q(-A + \mu_0) - Q(A + \mu_0)] \\ &= \Phi(A - \mu) - \Phi(-A - \mu) \end{aligned} \tag{4.9}$$

where $\Phi(x)$ is the cumulative distribution function and μ_0 and μ_1 are the non-zero means due to signal energy in both symbols. When received signal energy for both symbols is equal

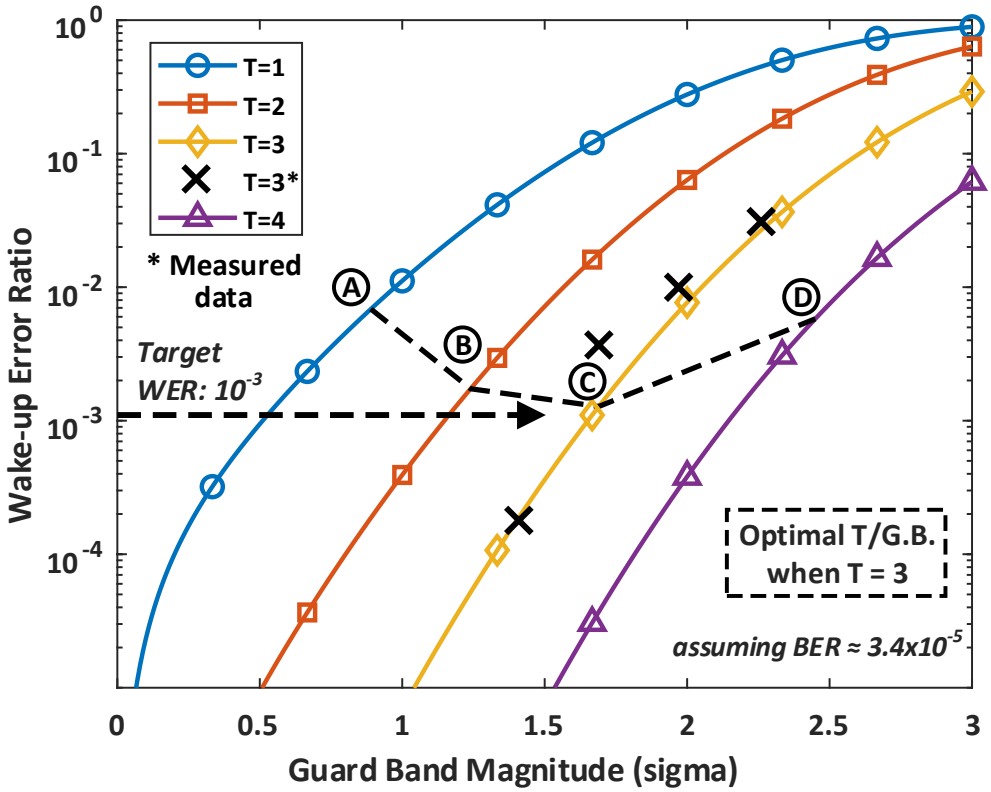


Figure 4.17: Model of wake-up error ratio assuming a BER of 3.4×10^{-5} where the most sensitivity is achieved by choosing $T = 3$ and G.B. of 1.69. Measured data presented for $T = 3$ curve.

$\mu_1 = \mu_0 = \mu$ which simplifies the result, as shown in Equation 4.9. The value P_b represents the bit error ratio, which is a function of the signaling scheme and E_b/N_0. With this information, a model can be made relating WER as a function of guard bands and idle threshold curve. The previously recorded idle threshold/guard band points can be plotted to determine what curve provides the optimal WER. By sweeping BER, the optimal idle threshold/guard band pair will settle on a WER of 10^{-3}. Figure 4.17 shows an example of this step, whereby a line connecting points 'A' through 'D' shows that point 'C' achieves the target WER for the worst tolerable BER. This chosen point is when $T = 3$ and $A = 1.69$.

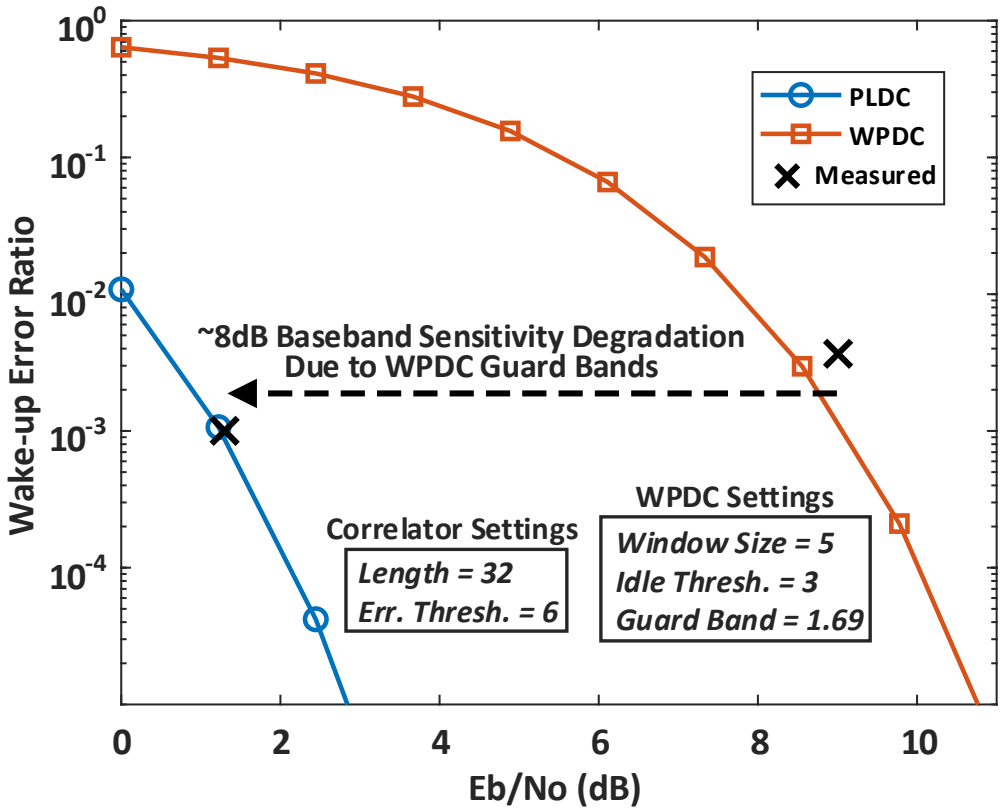


Figure 4.18: Model of baseband waveform E_b/N_0 versus wake-up error ratio demonstrating a theoretical loss of 8dB in baseband sensitivity due to WPDC operation at WER $\approx 10^{-3}$.

4.6.5 Sensitivity Degradation

We can evaluate the impact on baseband sensitivity due to WPDC now that all variables are fixed, as shown in Figure 4.18. There is an expected 8dB degradation in baseband sensitivity due to the use of WPDC when compared to PLDC mode. Two measurement points near WER $\approx 10^{-3}$ support this modeled result.

Implementation Considerations. It has been assumed up to this point that the receiver noise profile does not change and there is an infinite resolution to choosing guard band levels. A variety of factors in a real receiver will contribute to fluctuating noise distributions at the baseband output, e.g. temperature variance and gain adjustment, which will force the WPDC operating points to shift in an undesirable manner. As a result, robust WPDC-based wake-up receivers should

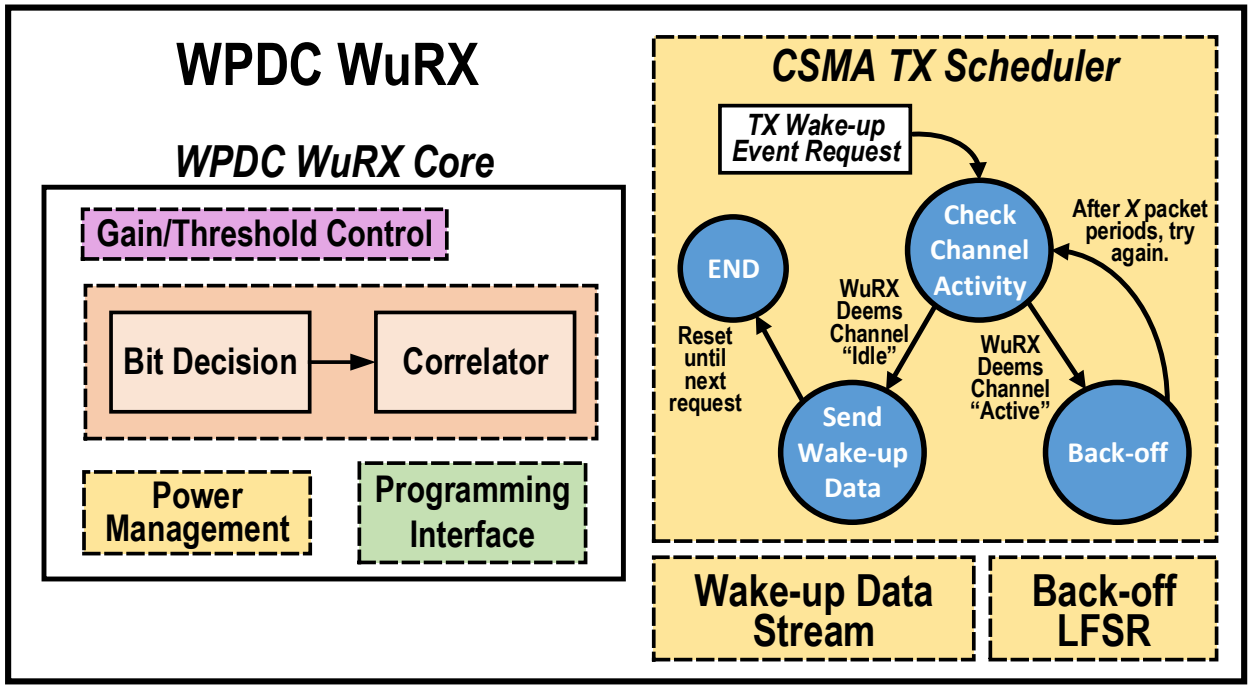


Figure 4.19: For node-to-node wake-ups in high-scale sensor networks, the same carrier-sense method in WPDC can be used to implement carrier-sense multiple-access schemes to help prevent wake-up collisions.

implement methods to compensate for potential fluctuations in the noise profile, e.g. fine automatic gain control, to mitigate the associated challenges. The combination of low ADC resolution and baseband gain also limits how finely WPDC guard band settings can be chosen and should be taken into account. Uneven IF path gains contribute to a distorted noise distribution which can further complicate the analysis.

4.7 Enabling Node-to-Node Wake-ups with Carrier-Sense Multiple Access

So far the discussions in this dissertation have focused on the wake-up receiver itself, but have not addressed what is sending the wake-up message. For simplicity, some IoT networks choose

star-based sensor networking topologies to allow a gateway with sufficient energy availability to control and command the array of nodes. However as the IoT scales, it becomes necessary to disseminate responsibility for wake-ups to other nodes themselves. In effect, node-to-node wake-ups is an important concept to begin looking in the direction of. Carrier-sense multiple-access is an accessible MAC-level method because it is rather lightweight and doesn't in and of itself require complex synchronization needed in other protocols. Wake-up radios are designed for asynchronous communication and the within-packet duty-cycling technique employs a similar carrier-sense technique to CSMA, so this MAC-layer method lends itself well to this context to help manage the transmission of wake-ups. A hardware CSMA with collision avoidance algorithm was designed and integrated into a PLDC/WPDC WuRX to demonstrate this concept.

When building upon a within-packet duty-cycled WuRX, the CSMA logic requires several additional blocks (Fig. 4.19). A scheduler is needed to interface between the user, who signals intent to wake-up another node, and the other WuRX and CSMA blocks. The role of the scheduler is primarily to take over control of the normal, periodic WuRX duty-cycling operation when a TX event is requested. Under idle channel conditions, the timing block will turn the WuRX on and deliver a "idle channel" flag to the scheduler. If a TX event was also requested, then the scheduler will tell the timing block to turn off the WuRX for the amount of time equivalent to $M+1$ packets, as required by PLDC and WPDC methods, in order to then transmit the encoded data necessary for waking up another system. However if the channel is deemed active, then the scheduler will back-off the transmit time as prescribed by the CSMA algorithm and allow the WuRX to continue its operation. The calculation for back-off, based in the units of a single packet period, is determined by a 7b LFSR with programmable seed values. This pseudo-random number generator enables back-off time diversity in a larger network with little overhead. If the channel continues to appear active, the back-off time will increase until it saturates to the length of the LFSR. To generate the TX data, a PHY layer block first receives an enable flag from the scheduler. Based on the wake-up code that should be transmitted, the data streaming block creates an IF-encoded bit stream of back-to-back wake-up messages that lasts for $M+1$ packets, as defined by the PLDC and WPDC standard. This data stream is intended to modulate a transmitter to generate the proper RF

waveform as no on-chip TX was designed.

4.8 Threshold and Gain Compensation

4.8.1 AGOC Methods

The objective of automatic gain and offset control (AGOC) is to bring the amplifier chain out of compression, realign the baseband output voltage within the usable range, and constantly adjust the bit decision threshold to achieve a desired rate of "1"s, all due to dynamic changes in RF interference as well as temperature drift. Target interferer profiles this control loop is designed to handle are characterized by long, drawn-out time domain behavior such that it significantly affects WuRX performance ($> 1s$). This is in comparison to short, transient bursts from other local transmitters. The discussed AGOC control loops were implemented for OOK-based signalling methods, but could be adapted to other circumstances. Two AGOC architectures are presented in order to determine the use of gain control or offset control (Fig. 4.20). A comparator bank based approach is a simple method that requires a bit decision comparator with accompanying boundary decision comparators. The bit decision comparator provides data to the digital baseband, where the output is 1 if $V_{in} > V_{bit}$, and is critical to offset control. The effective input voltage (V_{in}) region for what is managed by offset control is determined by the high ($V_{ref,H}$) and low ($V_{ref,L}$) voltage references on the boundary comparators. If V_{in} remains in either of these upper or lower regions, then gain control is used to bring the input voltage back to the manageable range for compensating offset. Alternatively, an ADC can be used where the code space is split up into similar regions, but instead now denoted by different ADC codes. The threshold resolution for the comparator architecture depends on the level of tunability built into the analog comparators themselves, but also suffer from mismatch, offsets, and non-linearity. Using an ADC is also beneficial because every V_{in} sample has N-bit resolution instead of a 1-bit resolution with the comparator without severe power penalty given the low-static current consumption. Considering these factors, it is generally preferred to use an ADC.

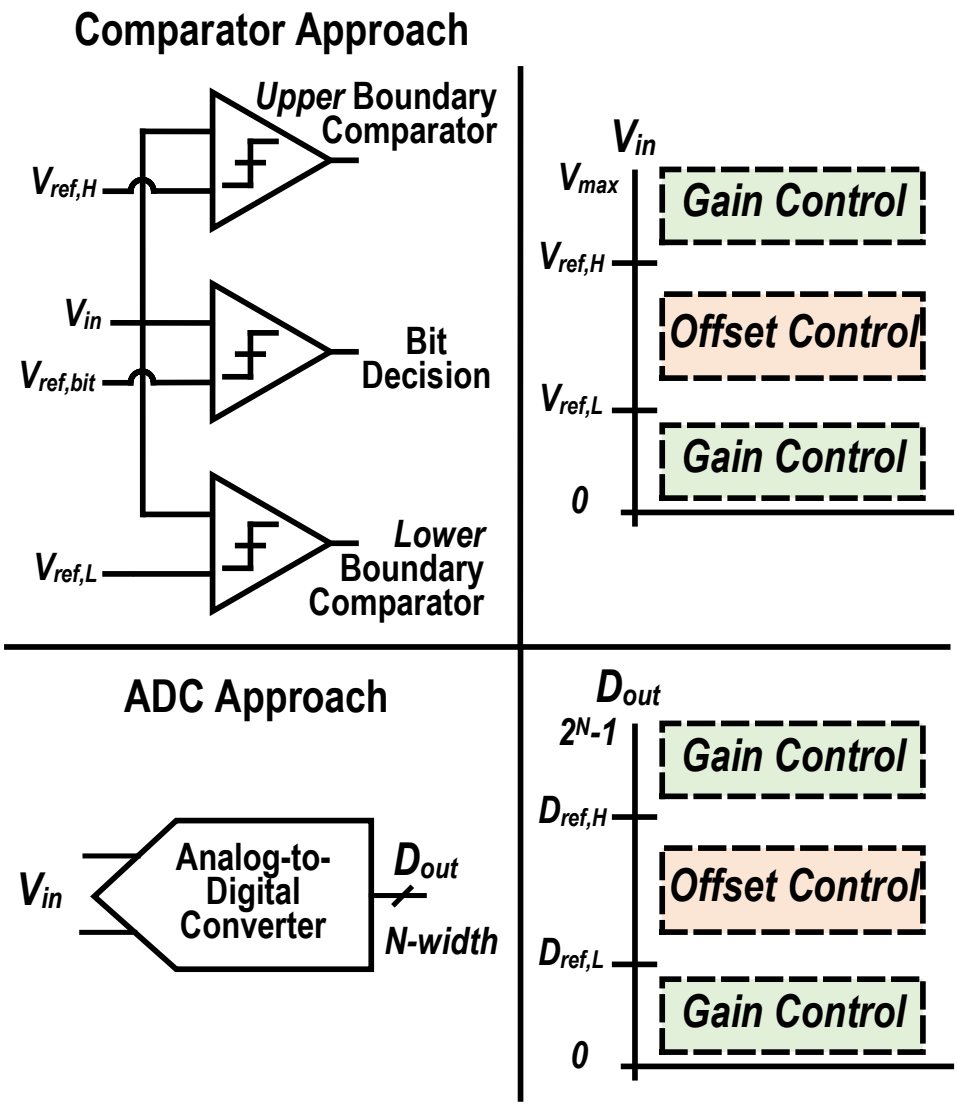


Figure 4.20: Comparator or ADC-based AGOC architectures can both provide the necessary information to switch between gain and offset control at the appropriate times.

By default, the AGOC algorithm begins in offset control mode (Fig. 4.21). It uses several coarse and fine levels of control over the offset for both quicker and slower adjustments when needed. The primary intention for OOK-based AGOC is to achieve a low rate of logical 1 bit decisions, as this is the only symbol in which data, and energy, is encoded in. Offset changes are made at slowest over the course of a defined window period of time in terms of a given number of samples. A ”fine 1s count threshold” is applied to slowly modulate the offset up and down by one

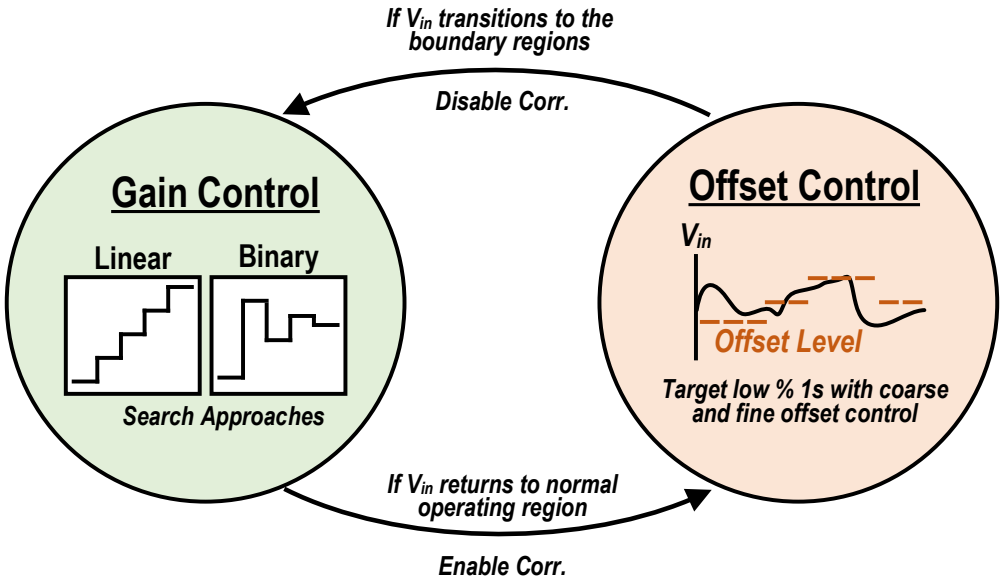


Figure 4.21: Comparator or ADC-based AGOC architectures can both provide the necessary information to switch between gain and offset control at the appropriate times.

point based on the number of 1s detected over the window period. Moderate and sequential count thresholds adjust the bit decision threshold more quickly allowing for faster convergence. Because energy is encoded only in the 1s symbol for OOK, it is faster to go from a low offset to high offset than vice versa. This is due to the rare probability of seeing a 1 symbol with an offset too high which makes it difficult to tell how further the real target threshold should go.

While temperature variation can be compensated primarily with offset control, actual RF interference may require changing gain stages in order to realign the baseband output with the necessary input range for the comparator or ADC. The input could rise too far, and possibly compress requiring less gain, or become too small requiring more gain. A search process is required in order to determine the proper gain stage and could be easily implemented with linear or binary search approaches. During this time, the correlator is disabled because all bit decisions are considered invalid at this time and could lead to false wake-ups. Once gain control properly resets the output to within the offset control range, the offset controller takes over and the correlator is enabled again.

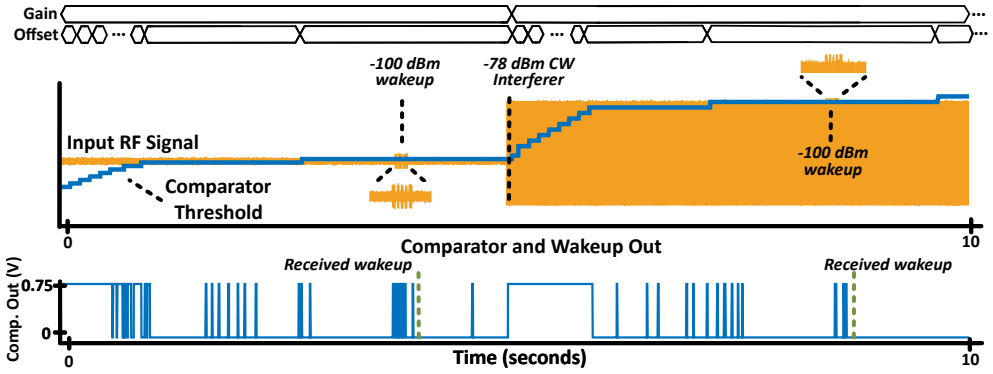


Figure 4.22: Demonstration of the AGOC algorithm dynamically adjusting WuRX gain and offset in the presence of strong in-band RF interference.

4.8.2 AGOC Performance

A demonstration of the implemented AGOC algorithm, based on a comparator architecture, is shown in Figure 4.22. This was achieved in the first generation bit-level duty-cycled wake-up receiver found in the next chapter. This figure shows several features: the baseband gain and offset settings, the corresponding bit decision comparator output, as well as the input RF power level all across time. This measurement begins by the offset controller ramping up to match the necessary '1's rate (approximately 10%) A wake-up transmission was first sent and properly decoded by the WuRX to demonstrate sensitivity at -100dBm input RF power. Over the course of these first few seconds, the offset control loop is attempting to achieve the desired '1's rate. At five seconds in, a -78dBm constant wave (CW) in-band interferer is applied to the RF input with a 500kHz offset from the data-tone. The comparator bank rails upwards and switches to gain control which quickly changes the gain step to bring the comparator back into range. The majority of the compensation time for this measurement was due to offset control trying to find the new operating level, which takes less than one second. After several more seconds of much more fine grained offset control, a second wake-up message is sent at -100dBm with the interferer still present.

4.9 Contributions

This chapter presented and analyzed numerous digital baseband techniques important for implementing low-power wake-up receivers. An itemized list of the contributions in this chapter is as follows:

1. Designed digital basebands for bit-level and packet-level/within-packet duty-cycled wake-up receivers and discussed implementation and operational details

2. Proposed the rotating correlator compatible with packet-level/within-packet duty-cycling to reduce WuRX active time and thus save power, in comparison to the conventional linear binary correlator

3. Created and compared math-based models, simulations, and measurements to quantify false wake-up ratios/rates for each duty-cycling method which is important to determine when utilizing a wake-up receiver to mitigate excessive system power consumption

4. Developed methodology for choosing within-packet duty-cycling operating variables to satisfy power, latency, and false wake-up requirements while minimizing degradation in sensitivity

5. Implemented several variants of automatic gain and offset control (AGOC) mechanisms for dynamically adjusting the bit decision threshold useful for compensating against temperature fluctuation and slow-changing interferers

CHAPTER 5

IMPLEMENTED DUTY-CYCLED WAKE-UP RECEIVERS

In this chapter, the previously discussed bit-level, packet-level, and within-packet duty-cycling schemes are demonstrated on four separate low-power wake-up receiver designs. Three BLDC wake-up receivers and a reconfigurable PLDC/WPDC wake-up receiver are discussed. These sections highlight motivation for the individual works and high-level architecture, design, operating choices as well as key measurement results. All of the CMOS wake-up receivers were designed in a 65nm LP process.

5.1 Tuned-RF Bit-Level Duty-Cycled Wake-up Receivers

As previously discussed, ED-first wake-up receivers readily achieve nW-level power consumption which makes them well fit for ultra-low-power IoT applications. A primary shortcoming is their poor RF sensitivity performance (generally limited to around -80dBm) limiting their effective range. For long distance applications, including agricultural and forest monitoring, improved WuRX sensitivity is a necessity for these systems. To accomplish this, a higher always-on power receiver architecture needs to be paired with a duty-cycling method in order to meet both the high sensitivity and low-power requirements. A series of three designed and tested wake-up receivers are discussed next that utilize the Tuned-RF (T-RF) architecture, characterized by always-on powers on the order of 10s of μW, and implement bit-level duty-cycling to achieve a flexible range of sensitivities better than -90dBm and pushing past -100dBm, scalable average power from nW-

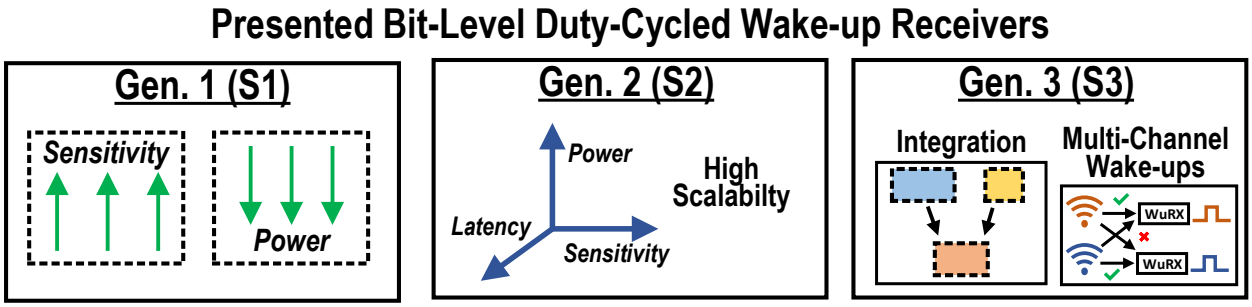


Figure 5.1: Each BLDC WuRX generation (colloquially named Samplifier or 'S' for short) was designed for different research questions.

to-μW, and programmable wake-up latencies from millisecond to seconds. These WuRXs are referred to here as "Samplifiers" and categorized as S1, S2, and S3 for short. They will be referenced as such throughout the rest of this document. Each generation of the BLDC T-RF WuRX explores a different research question (Fig. 5.1). S1 is the prototype samplifier WuRX design which was intended to maximize sensitivity and minimize power consumption [54] [55] through the use of high RF gain and sharp RF MEMS filtering to limit noise presented to the envelope detector. S2 was designed primarily to span the three dimensional trade-off space between power, latency, and sensitivity while also implementing a new signaling scheme to improve interference rejection and enable baseband FSK detection [56]. S3 focused on implementing multi-channel wake-up capability and improving the integration factor of the WuRX [57]. It is to be noted that latency measurements for each generation of samplifier is not given in terms of average latency, represented by Equation 3.5, but instead approximated by $L_{BLDC} = N_{code}\tau_{per}$.

5.1.1 BLDC T-RF: Generation 1 (S1)

The system architecture for the first generation bit-level duty-cycled WuRX is shown in Figure 5.2. At the RF input, single-tone OOK modulated data at 428MHz center frequency is sent through an input match to provide passive voltage boosting to the high-gain RF stage. The RF section is comprised of an LNA, common source amplifier, and an inductorless regenerative ring amplifier with a simulated voltage gain of 68dB and consumes $45.2\mu W$ when always-on. The ring

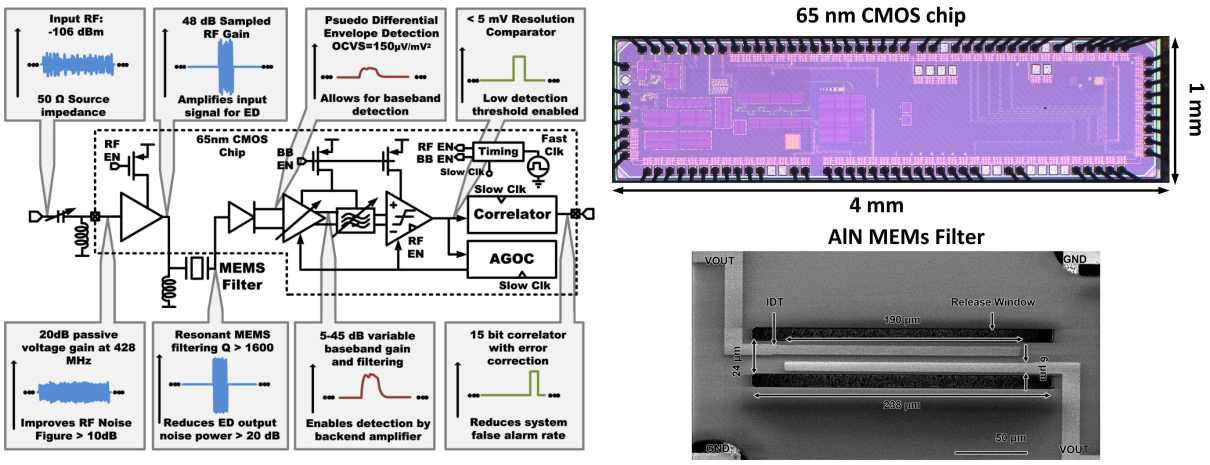


Figure 5.2: First generation BLDC T-RF architecture (left) with die micrograph of WuRX and AlN MEMS RF filter (right).

amplifier architecturally looks like a ring oscillator, but is backed off from its oscillation point to act as an high-gain amplifier. A high impedance ($5k\Omega$) 2-port Aluminum Nitride (AlN) micro-electrical-mechanical system (MEMS) filter is used here to reduce the RF output noise equivalent bandwidth (reducing significant noise contributed by the ring amplifier) to improve sensitivity, while as a byproduct improving out-of-band SIR performance. With a quality factor of approximately 1500, it has a fractional bandwidth of 0.53%.

After this stage, the baseband is similar to that of an ED-first topology. A passive pseudo-differential triode-mode Dickson-based envelope detector rectifies the OOK signal generating a short duration IF response, relative to the RF integration time. The detector has an open-circuit voltage-sensitivity (OCVS) of approximately $150\mu V/mV^2$. Without the RF MEMS filter, significant amounts of self-mixed noise from the RF front-end would significantly limit sensitivity at the output of the ED. The now rectified data signal is filtered and amplified after the RF ED with a digitally tunable programmable gain array (PGA) (0-45dB in 5dB steps) and band-pass filters (1kHz - 10kHz bandwidths). The baseband itself takes the longest to startup requiring around $390\mu s$ pre-charge time before RF integration begins. The baseband consumes an active power of $376nW$.

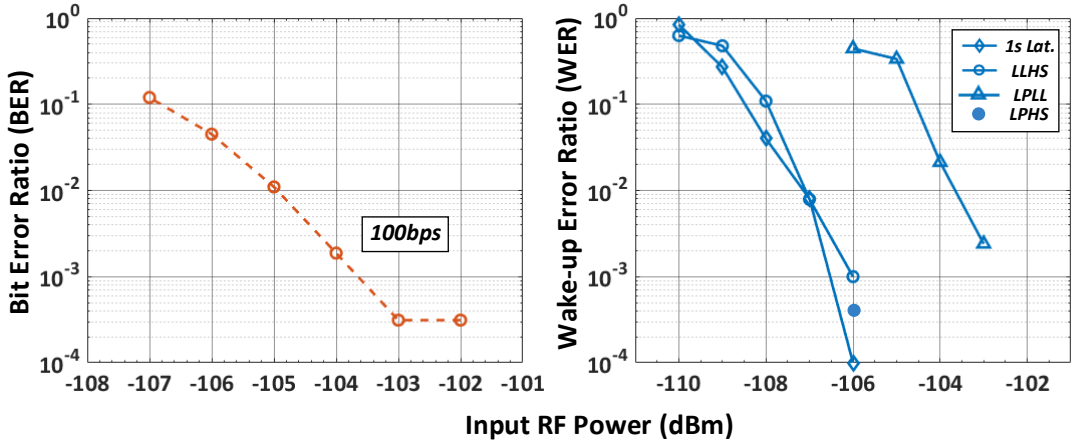


Figure 5.3: S1 BER (left) and WER (right) sensitivity measurements for specific operating points.

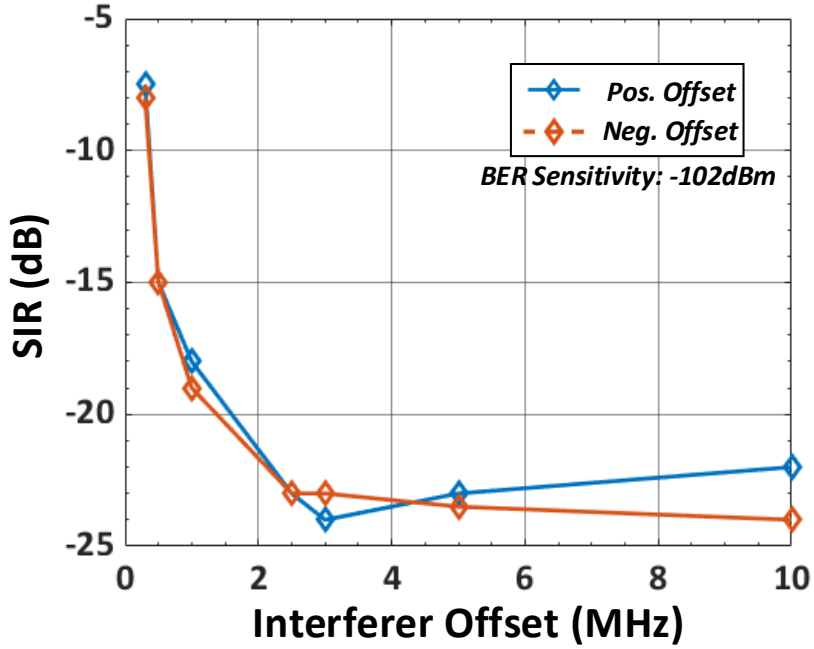


Figure 5.4: S1 SIR performance with constant wave interferer as a function of positive and negative offset from the signal carrier.

The signal is then sent to a digitally tunable, 6b-weighted analog comparator. The bit decisions generated by this comparator are fed into a 15b binary correlator with adjustable error tolerance that linearly shifts in a new bit at every bit period and in a binary fashion correlates the

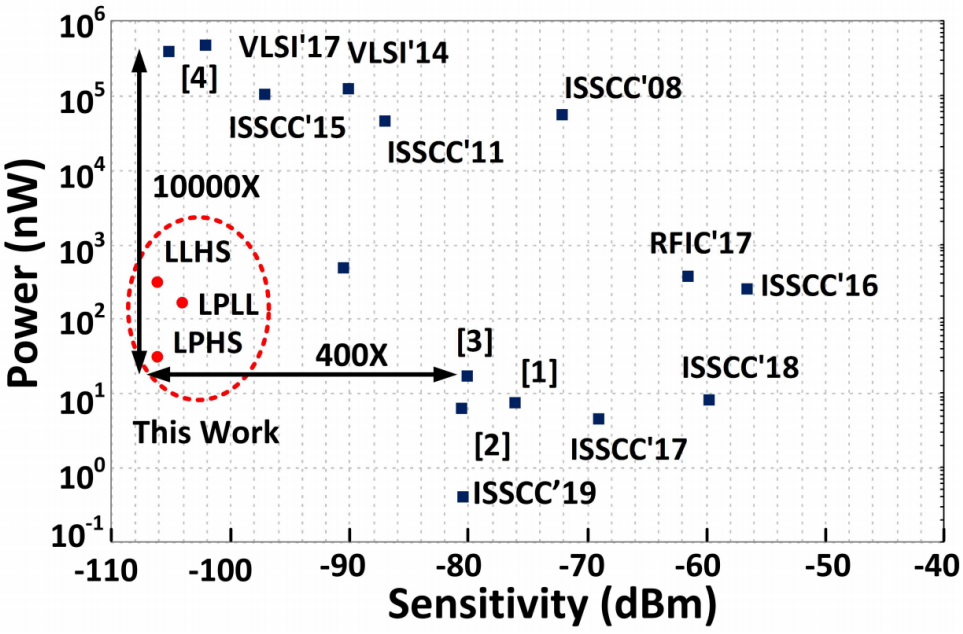


Figure 5.5: Sensitivity vs. power scatter plot with several points for S1 plotted against comparable low-power receivers at the time of publishing.

internal data against a reference code to generate wake-up flags. An automatic gain and offset control (AGOC) loop was designed to compensate for drift of the ideal threshold due to effects including temperature and slow-changing interference by adjusting both the comparator offset and PGA. The performance of this loop was presented in Figure 4.22. An on-chip dynamic leakage suppression (DLS)-based ring oscillator is used to minimize the power impact of the always-on clock source.

To enable the bit-level duty-cycling of the WuRX, a digital timing block creates synchronized power-gating signals to enable and disable each high-powered block. Blocks requiring longer start-up time are enabled/pre-charged ahead of sampling time. This staggered approach to enabling individual components helps minimize the total energy spent on start-up. Factors contributing to the inactive power of the receiver are the synthesized digital core, clock, and leakage from other power-gated components, which is just below 33nW. Die micrograph images of the 65nm CMOS and MEMS devices are presented in Figure 5.2.

In Figure 5.3, measurement results for both bit error ratio (BER) and wake-up error ratio (WER) are presented for the first Samplifier design. To divide up the BLDC trade-off space, three ”corner” WER points were measured: low-latency and high-sensitivity (LLHS), low-power and low-latency (LPLL), and low-power and high-sensitivity (LPHS). In addition, a one-second latency WER curve was measured. A single BER measurement was taken for reference which used the same integration time as other high-sensitivity WER points. These results demonstrate that at the LPHS point, the S1 WuRX achieves -106dBm WER sensitivity at 33nW average power consumption. To enable the LPLL mode, baseband filter bandwidths and RF integration time are decreased, which has the effect of reducing sensitivity. The bit period is also shortened in order to achieve the lower-latency setting. The LLHS mode is tuned, relative to the LPHS point, by only decreasing the bit period. Overall, this WuRX demonstrated a tunable range of 9x in power, 20x in latency, and 3dB in sensitivity. Given the objective of primarily targeting high-sensitivity and low-power operation, the golden WER point for this first prototype was -106dBm WER sensitivity at 33nW average power. The false wake-up rate was measured to be less than one per hour over the course of these WER measurements. The SIR performance for S1 is provided in Figure 5.4 for a constant wave interferer. Data was taken plotted with reference to the main signal carrier for both negative and positive frequency offsets. The WuRX achieved 22dB SIR at 3MHz offset and 16dB SIR at 500KHz offset.

A comparison scatter plot is shown in Figure 5.5 that shows several measured points for S1 against comparable low-power receivers at the time of publishing. The combined bit-level duty-cycling technique and T-RF WuRX enabled an improvement in sensitivity by over 20dB and reduction in power by several orders of magnitude.

5.1.2 BLDC T-RF: Generation 2 (S2)

The system architecture for the second generation BLDC T-RF WuRX operating at 430MHz is shown in Figure 5.6 which also includes the die micrographs for CMOS and MEMS devices. Upon comparison, the RF architecture is reminiscent of that from S1. This design utilizes ”two-

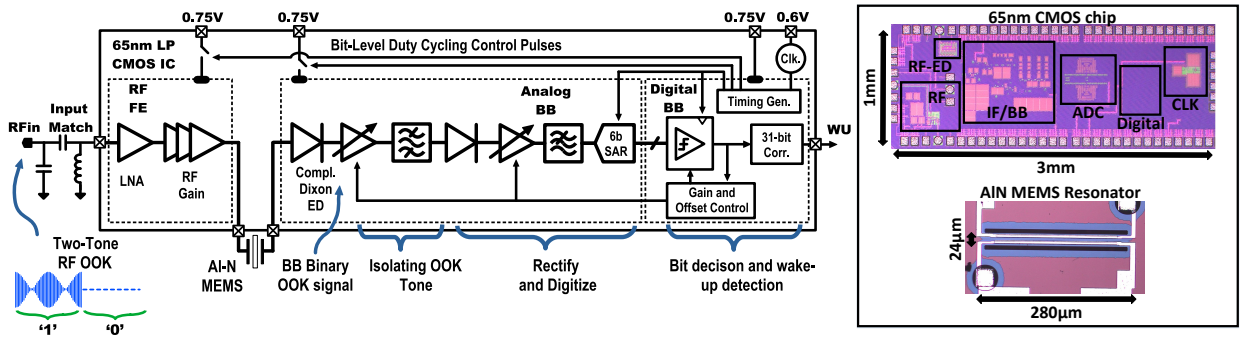


Figure 5.6: S2: Second generation BLDC T-RF architecture (left) with die micrograph of WuRX and AlN MEMS RF filter (right).

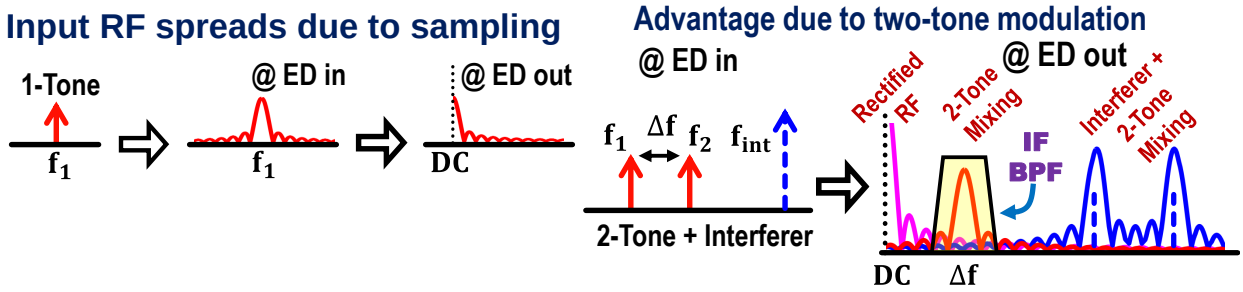


Figure 5.7: S2: The two-tone OOK signaling method ensures data will be better protected by pushing it to an IF frequency where it can be better filtered and isolated.

tone" RF OOK modulation, which encodes the symbol "1" on two carriers spaced Δf apart, as shown in Figure 5.7. For this work, Δf = 63kHz. By periodically sampling the RF spectrum, energy is rectified to DC where our signal is further distorted by noise and interferers. This signaling method enables better interference rejection by sending the information to an IF due to the two tones mixing together from the ED, rather than near-DC as before with "1-tone" in S1. At a higher IF, sharper filtering is easier to achieve. The RF input goes to a tapped capacitor input match and through RF gain stages. A RF AlN MEMS filter is used in a similar fashion to S1 to restrict the noise equivalent bandwidth. The envelope detector then rectifies the RF energy sending it to IF gain and bandpass filtering stages. Because the energy is currently at an IF frequency, it requires a second rectification stage by another ED to turn it into a baseband signal where it is further amplified and filtered. Instead of using a single bit comparator, this design opts for a 6b SAR-based

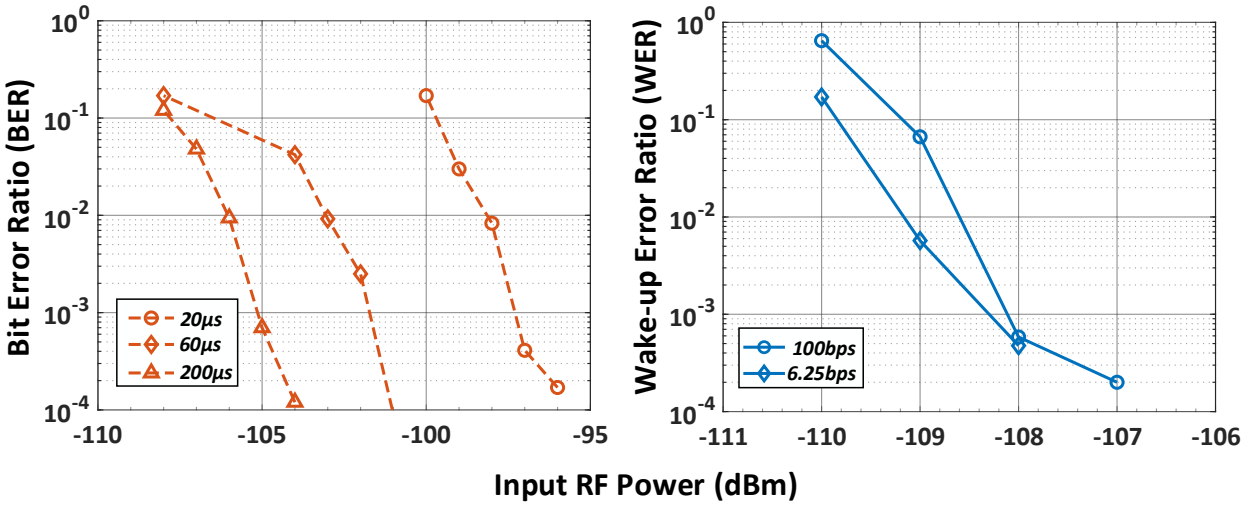


Figure 5.8: S2 RF sensitivity measurements for bit error ratio (left) and wake-up error ratio (right).

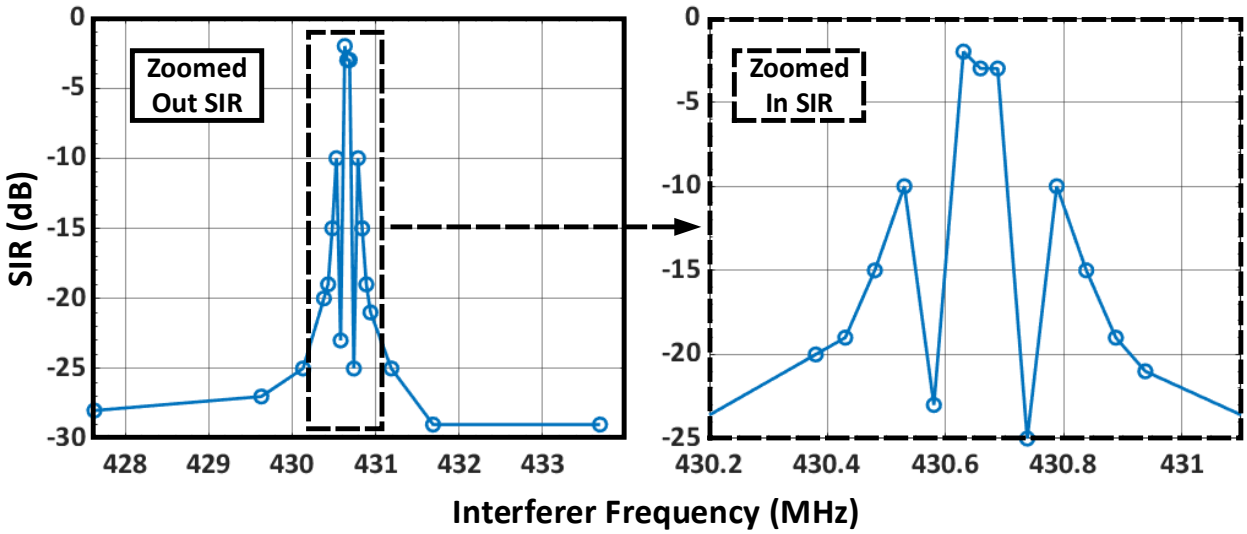


Figure 5.9: Near and far-out measurement of S2 SIR performance with constant wave 2-tone interferer.

analog-to-digital converter (ADC) to enable improved bit decision and AGOC performance. The digitized 6b result is then fed into a digital comparator which uses a 6b threshold to generate bit decisions. A longer 31b correlator is used in S2 for generating wake-up flags. A similar timing block mechanism is used to duty-cycle each component through individually addressed counter

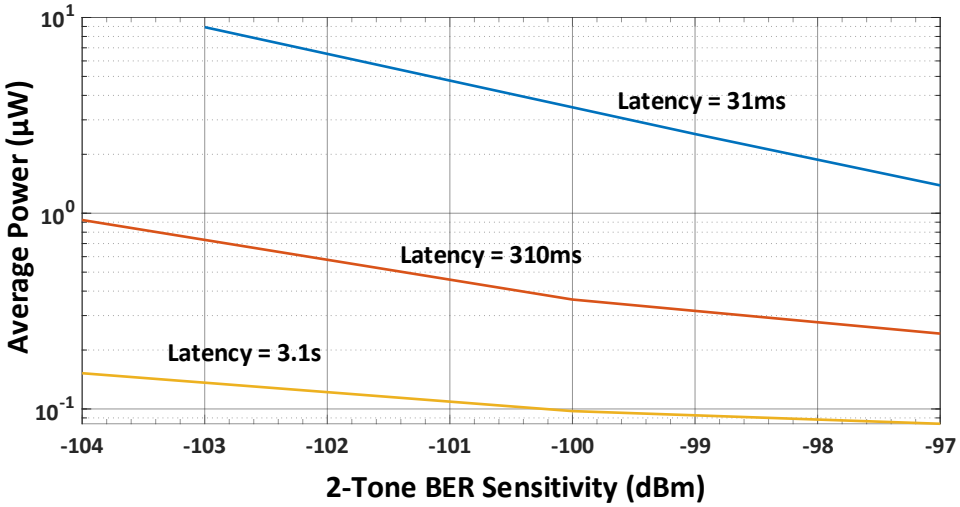


Figure 5.10: Measurement of S2 BLDC three-dimensional trade-off space in average power, latency, and RF BER sensitivity.

thresholds.

RF sensitivity measurements are presented in Figure 5.8. Three bit error ratio measurements, with bits generated from a PRBS9 pseudo-random bit sequence, are shown across integration times which demonstrate the span of achievable BER sensitivity for the WuRX, from approximately -97dBm to -105dBm. Two separate WER measurements are shown for the same RF integration times, but at differing latencies (function of bit rate). This is done to show that there isn't significant difference in sensitivity, as expected, across latency if the RF integration time doesn't change. The "6.25bps" point represents a -108dBm WER sensitivity at 130nW average power consumption. The measured false wake-up rate for these measurements was less than one per hour.

Two views for the S2 SIR performance are presented in Figure 5.9 for a constant wave interferer. In the "zoomed out" view, the MEMS filter enables the better-than -25dB SIR performance in the out-of-band region. For the in-band region, the SIR takes a significant hit except on frequencies that overlap with the 2-tone signal frequency.

A series of measurements were also taken to demonstrate the trade-off space for the BLDC method. For each latency setting, the RF integration time was digitally reprogrammed to exchange

sensitivity and average power (Fig. 5.10). The three digitally programmed RF integration times were 20μs, 60μs, and 200μs. Accounting for all measurements taken, S2 demonstrated scalability of 410x in average power, 672x in data ratio, and 11dB in sensitivity.

5.1.3 BLDC T-RF: Generation 3 (S3)

The third and final bit-level duty-cycled WuRX focused on improving integration as well as enabling multi-channel wake-up capabilities (Fig. 5.11). This integration effort primarily focused on removing the external MEMS device, at the cost of sensitivity. A new signalling scheme was applied to this WuRX called channel embedded OOK (CE-OOK) which embeds ”0” and ”1” symbols as different OOK data rates (Fig. 5.12). Two IF tones appear at the RF ED output which can be isolated from one another through bandpass filtering, then further rectified to translate to antipodal baseband signaling. This effectively ”channelizes” the data and relaxes constraints on the transmitter. With programmable filter settings, more IF channels can be obtained. Similar to S1 and S2, the RF input match leads directly to the RF gain stage comprised of an LNA, variable gain ring amplifier, and RF buffer which in total consume $50.3\mu W$ active power and deliver 61dB gain (simulated). Due to the desire for greater integration the MEMS device is removed and the RF output directly goes to a Dickson-based envelope detector for first stage rectification. The CE-OOK data then appears at the IF as two separate tones, where one is present when sending a ”0” and a different IF tone is present when sending a ”1”. Because these tones are fixed by the transmitter, separate filtering paths are used to isolate them from one another proceeded by another rectification stage. A measurement of the filter responses for parallel IF isolation paths centered at 61kHz and 84kHz is shown in Figure 5.12 (right). At this point, they are compared in an analog fashion to generate a baseband FSK waveform. This baseband signal is further low-pass filtered and amplified. A similar 6b SAR ADC used in S2 samples the baseband output and presents the result to the digital comparator for bit decision. In total, the IF, baseband, and ADC consume $2.1\mu W$ active power. A 64-bit reconfigurable length correlator with guard bands is implemented in this design to enable a broader range of wake-up functionality. Additionally, an offset correction

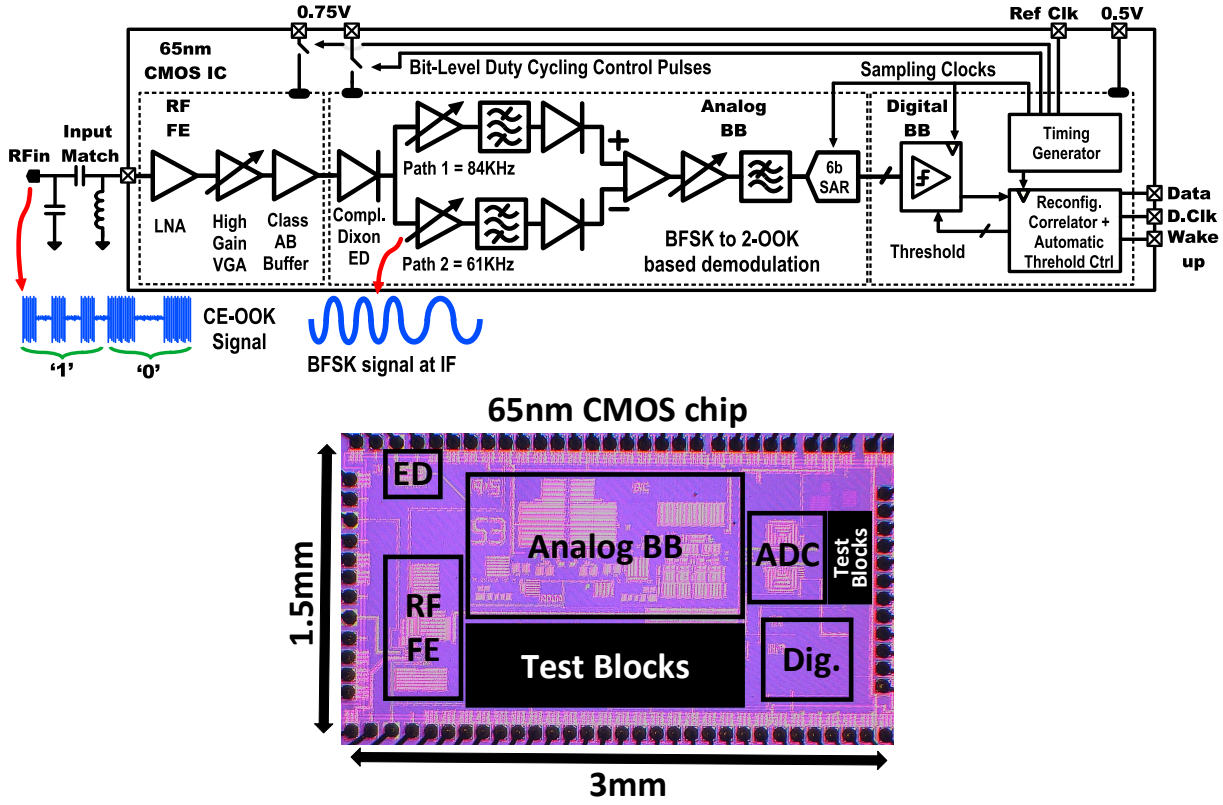


Figure 5.11: Third generation BLDC T-RF architecture (top) with die micrograph of WuRX (bottom).

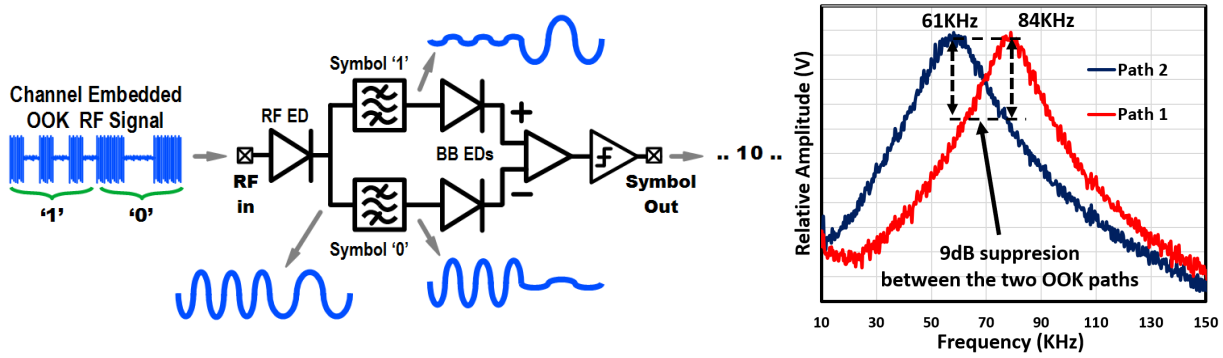


Figure 5.12: A single IF channel pair filter frequency response, which allows for channelized wake-ups if CE-OOK rates are tuned to IF channel pair center frequencies.

control loop can slowly tune the digital threshold to compensate for temperature drift. A standard BLDC timing block is used to power-gate all of the components in a synchronous fashion similar

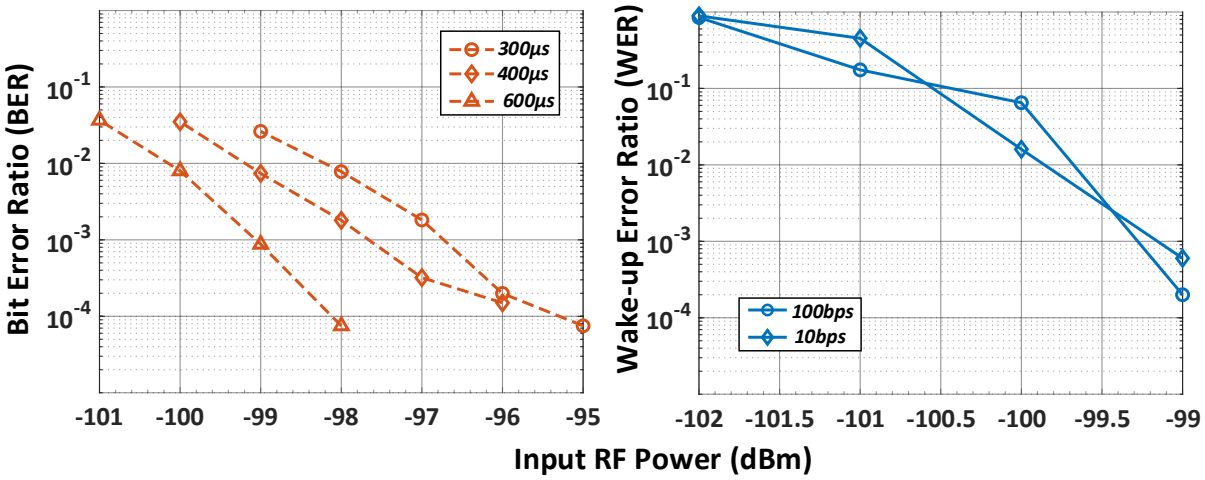


Figure 5.13: RF sensitivity measurements (bit error ratio and wake-up error ratio) for S3

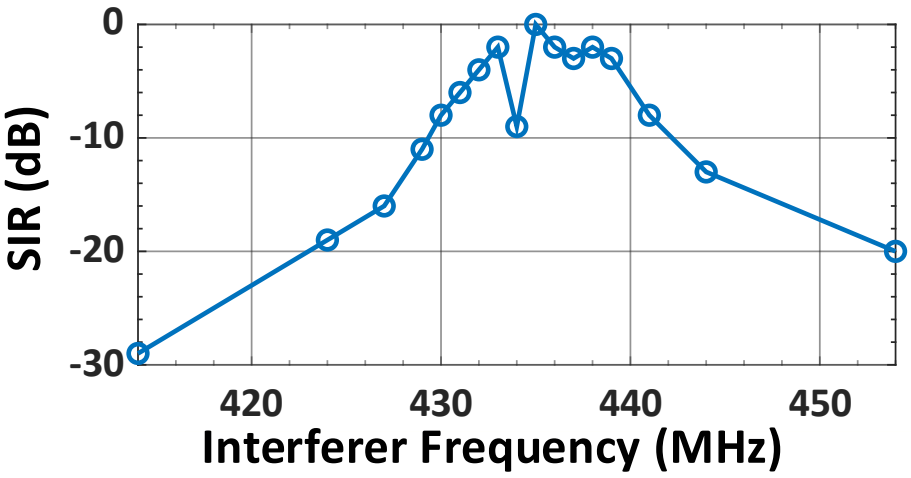


Figure 5.14: Measurement of S3 SIR performance with constant wave interferer.

to S1 and S2. Once per slow clock cycle, all duty-cycled blocks will turn on in their specified order ending with a single sample by the ADC. Each component is assigned 12b start and stop indices to raise and lower the respective power gate or startup signal when the counter crosses the programmed threshold values. The digital clock frequency and counter size determine the resolution and range for the timing block. For a digital baseband clock of 50kHz, the timing block provides 20μs step resolution and a maximum bit period of 164ms. This time step is small enough to provide sufficient tuning for staggered startup and finely-tuned RF integration times, while also

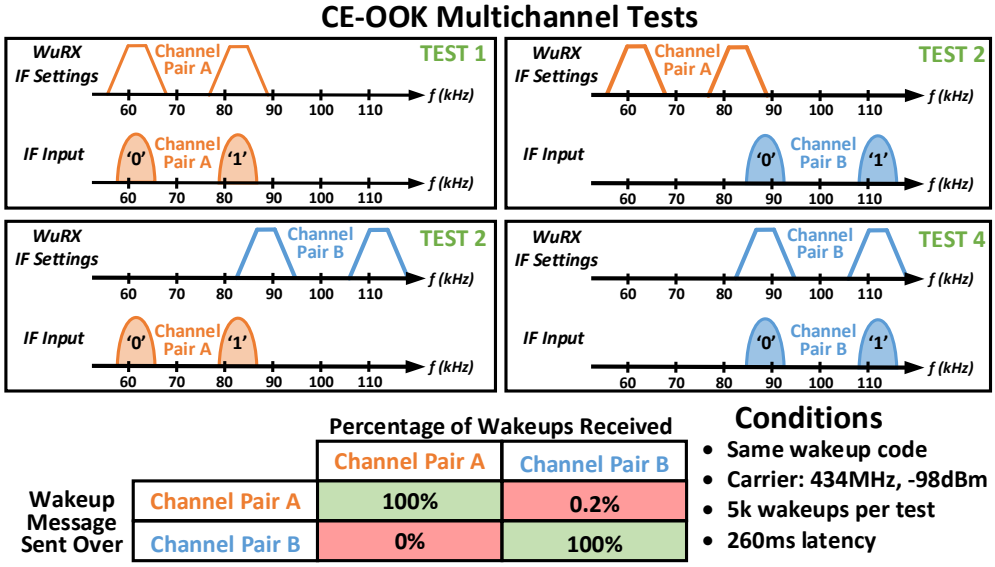


Figure 5.15: Measurement of S3 channelization using CE-OOK modulation reporting true positive and false negative rates for two separate channel pairs.

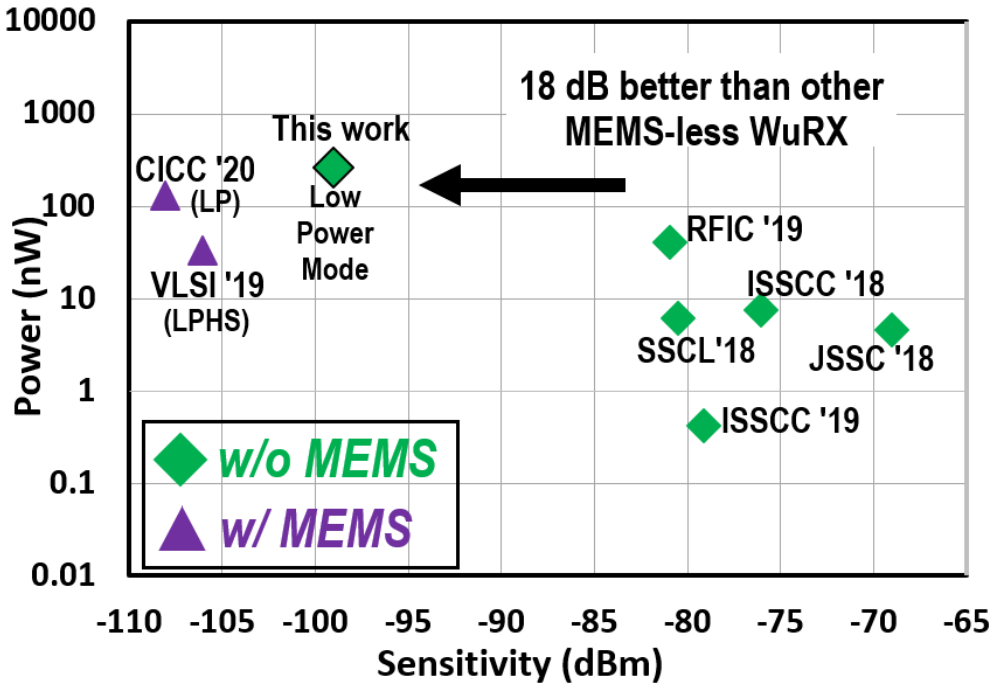


Figure 5.16: Sensitivity vs. power scatter plot with several points for S1, S2, and S3 plotted against comparable low-power receivers at the time of publishing.

supporting long wake-up latencies of up to 5.25s with a 32b code for low-power operation. At lower latencies, baseband startup times limit how small the bit period can be. An "always-on" mode can be applied to the analog baseband bypassing the standard timing block so that it remains constantly running providing a wider range of latencies and mitigating some start-up related offsets that can affect sensitivity.

BER and WER sensitivity measurements for S3 are presented in Figure 5.13 for several operating points. Similar to S2 results, three integration times are selected to measure across the trade-off space for S3 demonstrating BER sensitivities from approximately -97dBm to -99dBm. These BER curves were measured using data rates of 1kbps and PRBS9 sequence. These measurements assume a false wake-up rate of < 1 false positive per hour. The wake-up code used was two concatenated 13b barker codes, for a total of 26b. The WuRX achieved a WER sensitivity of -99dBm for 260ms and 2.6s latency while consuming 2.17μW and 260nW, respectively. Both WER curves were taken with 400μs integration time. Compared to S2, S3 experiences significant WER sensitivity degradation primarily due to the lack of RF MEMS filter.

The SIR performance for S3 is shown in Figure 5.14. The desired signal power is 3dB above the measured sensitivity at 500bps with 400μs integration time. Same carrier-frequency interferer yields a -9dB SIR due to the CE-OOK signaling method, but out-of-band SIR is comparably worse to S1 and S2 due to the lack of RF MEMS filtering.

The CE-OOK signaling scheme in S3 allows for the generation of IF channels enabling frequency division multiple access that can be used for simultaneous wake-up events lowering collisions in a wake-up network. As shown in Figure 5.15, four tests demonstrate the ability for two "channel pairs" to pass and isolate wake-ups, depending on what channel pair data was sent on. These channel pairs were centered at 61/84KHz and 88/113KHz for Channel Pair A and B, respectively. Each test consisted of 5000 wake-ups and alternated between what channel the data was sent over and what channel the WuRX was programmed to receive data on. Both channel pairs correctly passed 100% of the intended wake-ups, while in the worst case unintentionally passed 0.2% of the time.

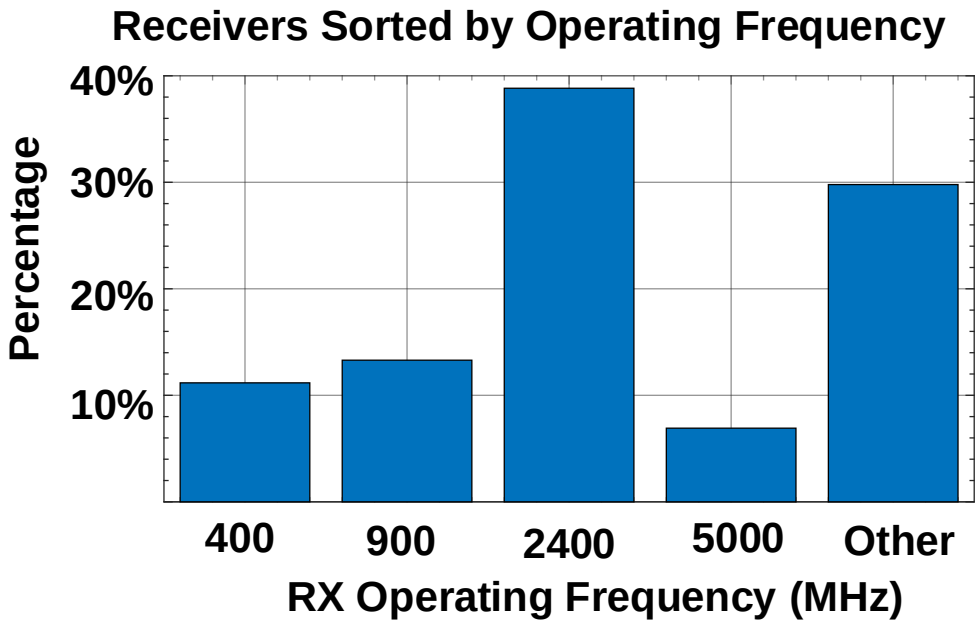


Figure 5.17: The 2.4GHz ISM band is a popular choice for receiver design given how many devices use this wireless band for operation, which makes it a compelling choice for future WuRXs. Data based on [19].

Figure 5.16 shows a scatter plot of sensitivity versus power consumption for notable S1, S2 (CICC 2020), and S3 (VLSI 2019) data points relative to other low-power receivers at the time of publishing.

5.2 Uncertain-IF Packet-Level and Within-Packet Duty-Cycled Wake-up Receiver

The Samplifier series of wake-up receivers primarily targeted long-range IoT applications, but there are many other application spaces where wake-up receivers can be of use. The wide majority of the wireless devices we use today have built-in transceivers to communicate over Wi-Fi and Bluetooth. These standards operate on the very popular 2.4GHz ISM band. From a recent low-power radio survey, the majority of published designs operate at and around 2.4GHz clearly

Example Commercial SoC Specs

Off Current w/ Mem. Retention	2.4GHz RX Sensitivity
~1µA*	-97dBm*
0.7µA**	-96dBm**
1µA***	-93dBm***

Figure 5.18: Several examples of sleep currents and RX sensitivities for commercially available IoT systems-on-chip. (*: [20], **: [21], ***: [22])

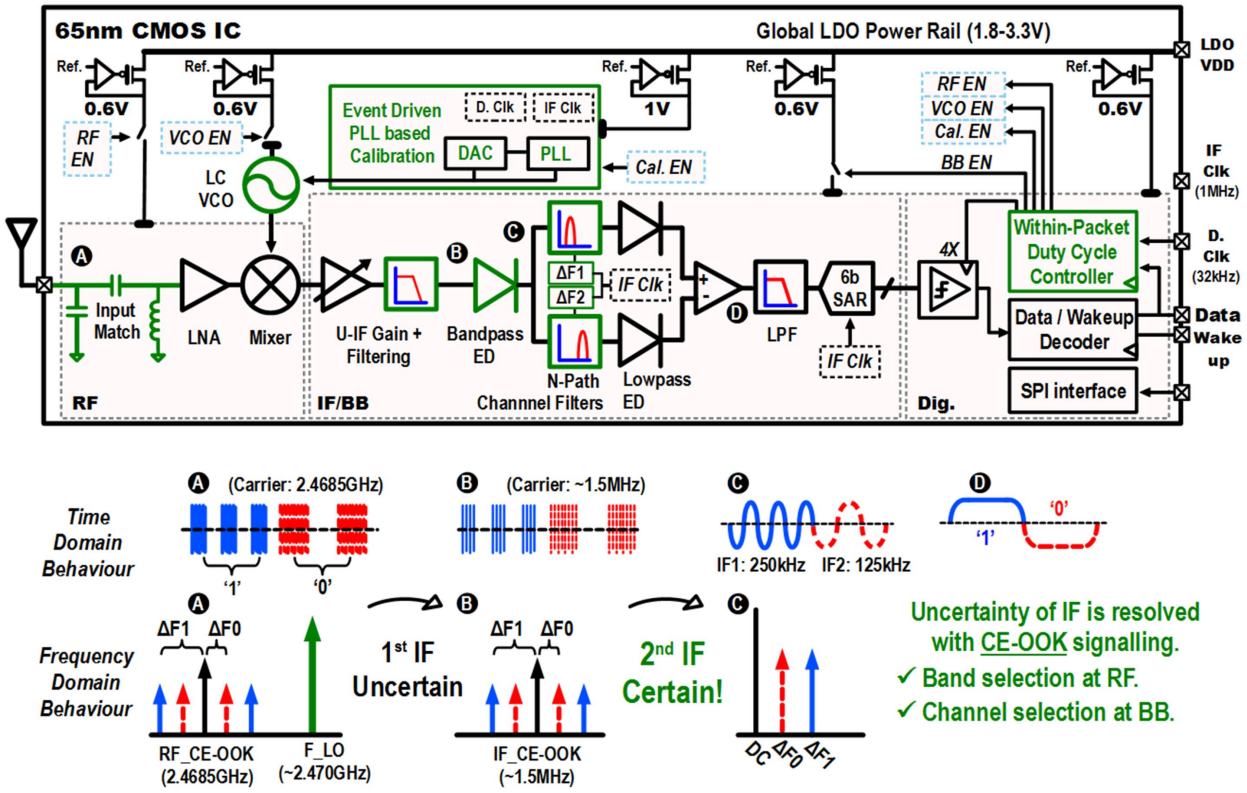


Figure 5.19: The Uncertain-IF packet-level/within-packet duty-cycled wake-up receiver targeted at near-μW operation for 2.4GHz ISM devices.

showing how relevant the band is (Fig. 5.17). Given how many of our consumer devices alone use this technology, a wake-up receiver could help significantly in these systems. The IEEE working group 802.11ba was founded in 2016 to begin developing a wake-up standard for the 802.11 stan-

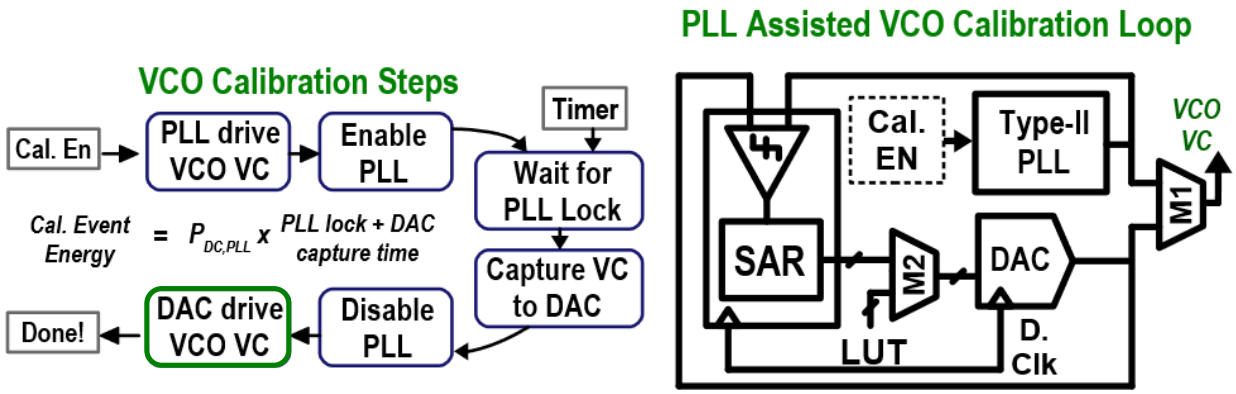


Figure 5.20: Free-running VCO calibration achieved through the use of an on-chip PLL and SAR-based search scheme to identify necessary control voltage.

dard. A functional wake-up receiver adhering to an early version of this standard and integrated with a transceiver was subsequently published [9,58]. In a networking context, they modeled that a wake-up radio could provide up to 70x in power savings compared to traditional low-power "target wake-up time" methods. Rather than complying with the upcoming WuRX standard, this work seeks to achieve a far-lower power WuRX compatible with 2.4GHz transceivers. There are several base needs that are identified which drive the design of this WuRX. The first is with respect to power consumption. Many of the IoT SoCs on the market today advertise low-power sleep modes to remain relevant in edge IoT application spaces. Generally these values hover around $1\mu A$ (Fig. 5.18). Another important specification to note is the RX sensitivity of the main transceiver, which typically achieves better than -90dBm. Based on this information, it becomes clear that a WuRX compatible with these low-power SoCs must have an average power of near-μW and RF sensitivity better than -90dBm. Consumers also expect low-latencies with sub-1s response times. Interference rejection is also important in this crowded RF spectrum.

The proposed WuRX solution shown in Figure 5.19 adopts a U-IF topology with the addition of event-driven frequency calibration through an on-chip PLL. The WPDC method and rotating correlators provide drastic power savings over conventional PLDC, where the WuRX needs to stay on for at least twice the duration of a packet. Since U-IF architecture inherently cannot achieve

precise baseband channelization, channel-embedded OOK (CE-OOK) signaling is adopted from S3 to provide deterministic sub-carrier channel filtering at the baseband. By emulating a multi-tone signal, CE-OOK generates IF content with a precise frequency at the output of an envelope detector (ED). The WuRX is implemented in a low-power (LP) process with access to multiple Vth devices to further minimize the leakage in both analog and digital blocks.

The CE-OOK RF input signal is sensed through an on-chip matching network, amplified by an LNA, and downconverted to an uncertain-IF using an active mixer driven with a free-running VCO. The U-IF signal is further amplified and filtered by a programmable 4th order Gm-C low pass filter (LPF). The bandwidth of this LPF limits the allowable frequency drift of the LO and limits the maximum sensitivity and SIR. The filtered signal is rectified by an ED with a bandpass output for dc-offset mitigation purposes. The output of rectifier produces the embedded channels in CE-OOK signal in BFSK format. Then, a tuned-filter based energy comparison is performed to demodulate and produce an analog bit stream. An n-path (n=4) based bandpass filters (BPF) followed by an ED and a difference amplifier implements the energy comparison for '0' and '1' symbol channels. The difference is amplified and lowpass filtered prior to digitizing with a 6-bit SAR ADC. The ADC output oversamples the baseband output by 4x for bit decision. The oversampled bit stream is used for detection of the internal wake-up code with four parallel correlator banks as well as the within-packet duty-cycle control loop. The local power supply rails are regulated by dedicated on-chip low-dropout regulators (LDO) and the reference clocks of 1MHz and 32KHz are provided externally. Although the LDOs do not provide any power benefit when referred to the battery voltage, it is meant to be used as a proof-of-concept demonstrating the effects of LDOs on WuRX metrics such as sensitivity and robustness under duty-cycling, and can be redesigned for a lower global supply voltage.

The block diagram of the VCO calibration circuitry and corresponding timing diagram is shown in Figure 5.20. Upon receiving the "Cal. EN" signal, the PLL VDD is enabled and the multiplexer M1 selects the PLL to drive the control voltage of the VCO. After a programmable amount of time required for the PLL to acquire the lock condition, the "Capture EN" signal is issued. This enables the clock signal of the SAR based logic to resolve the control voltage in to

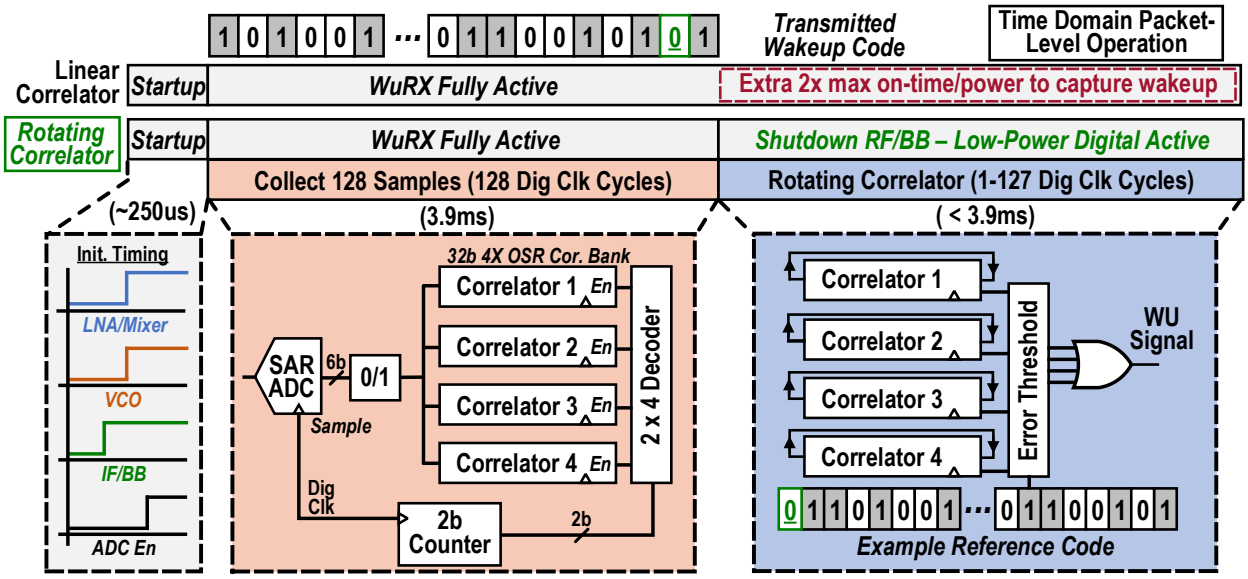


Figure 5.21: The digital baseband provides sufficient start-up time and enables the one-shot PLDC and WPDC duty-cycling schemes with a 4x-oversampling correlator bank.

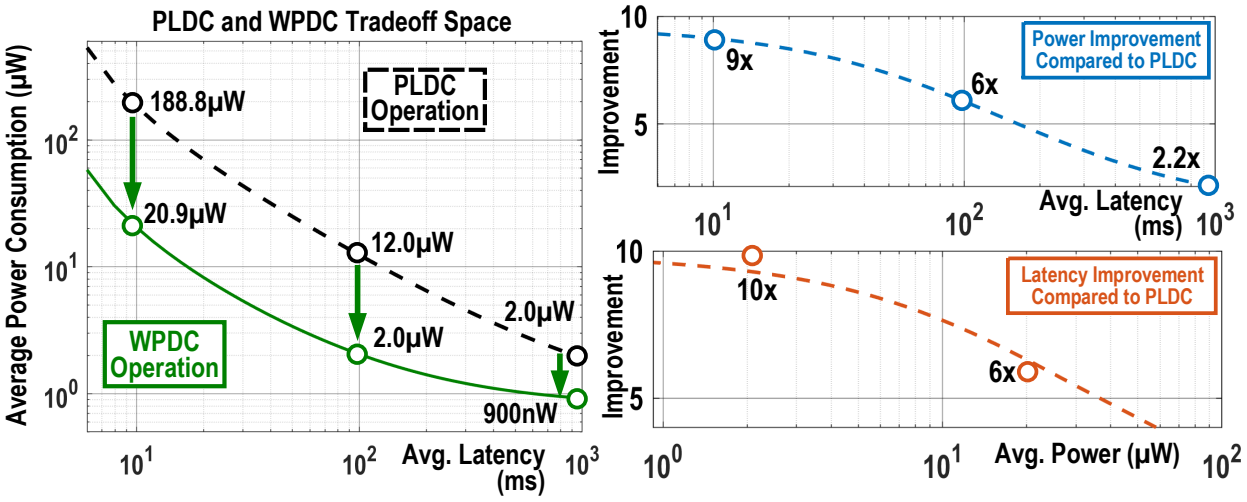


Figure 5.22: The latency vs. power trade-off space for PLDC and WPDC (left). WPDC advantage in power and latency (right). Curves are modeled and points are measured.

a 7-bit DAC. Finally, the multiplexer M1 sets the VCO to be driven by the output of the DAC and disables the PLL VDD completing the calibration event. The DAC is driven with the 32KHz clock and the SAR clock signal is generated by the capture logic block internally. Additionally,

the multiplexer M2 can be used to set an external DAC code word through the SPI interface as the initial condition or to be used with a look up table.

A 4x oversampling 32b rotating correlator bank is implemented in the digital baseband which captures the exact number of samples (128) in the wake-up code and then turns off high-power components and rotates through the data searching for the wake-up code with the rotating correlator architecture (Fig. 5.21). The within-packet duty-cycling carrier-sense functionality can be turned on or off based on digital programming. The timing block regulates duty-cycle behavior for all components and initiates the data collection process for the carrier-sense mechanism and the rotating correlator in a periodic behavior. When the data path is initiated by the timing block, the ADC sends the samples through a digital comparison block and then to the correlator with a sampling rate of 32.768ksps. Oversampling is achieved by using four parallel correlator blocks and a counter/decoder pair which toggles the enable lines for each correlator every fourth sample.

An accurate power versus latency model of the wake-up receiver was created to display the advantage of going from the PLDC to WPDC mode of operation. Three measured points per line are displayed as circles on these operating curves to verify that modeled results adhere to measurements (Fig. 5.22). The relative advantage, calculated as the ratio of power for a fixed latency or ratio of latency for a fixed power, helps show how much WPDC improves over PLDC and when it begins to falter. The greatest benefit, as discussed in the previous chapter covering duty-cycling methods, is seen at the lowest latencies when the WuRX is dominated by active power consumption. Here, up to a 9x improvement in power and 10x improvement in wake-up latency was measured. At lower duty-factors, the advantage drops as the floor power approaches the active power. The floor power is approximately 750nW, which is dominated by the digital and timing block. Here, there is only a 2.2x improvement in power. The die micrograph of the fully-integrated PLDC/WPDC wake-up receiver is shown in Figure 5.23 along with the RF input match measured in the 2.4GHz ISM band.

Wake-up error ratio measurements were made in both the WPDC and PLDC operating modes for the WuRX (Figures 5.24 and 5.25). Three WPDC points were taken for three different average

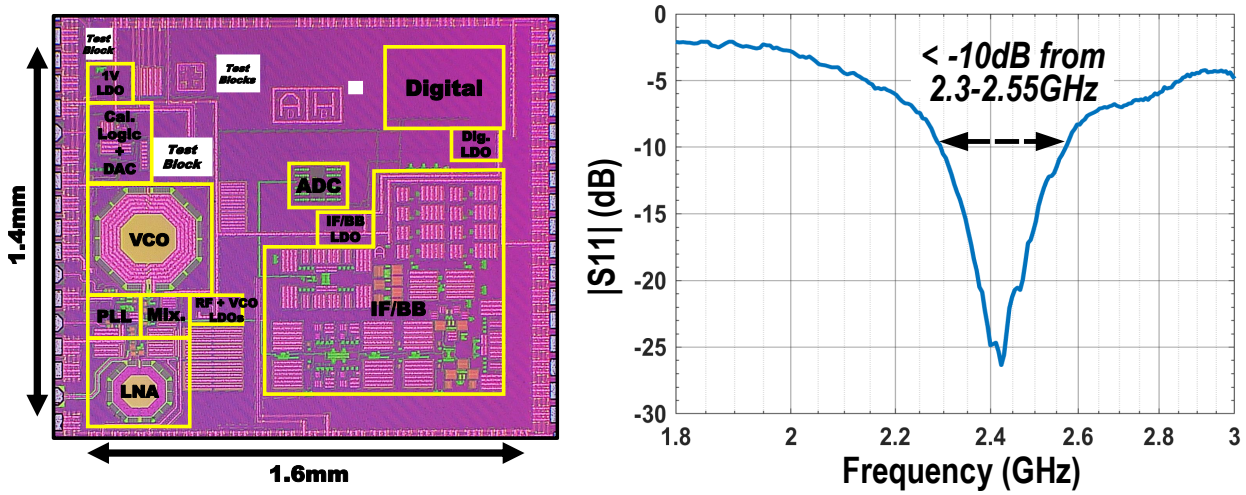


Figure 5.23: Die micrograph of WuRX and input match centered at 2.4GHz.

wake-up latencies: 10ms, 100ms, and 1s. These measured points showed decent agreement with each other, where at 100ms average latency the sensitivity was -91.5dBm with a false wake-up rate of less than one per ten hours. This exceptionally low rate was due to the nature of WPDC operation, which was discussed in the chapter covering digital baseband methods. For the same reasons, the PLDC mode demonstrated nearly a 2dB improvement in sensitivity as a result of not using ADC guard bands. This value differs from the 8dB degradation at baseband from the previous chapter done in the study to determine WPDC parameters because the applied settings were different.

The SIR performance of this wake-up receiver is presented in Figure 5.26. At 20MHz out, it achieves -47.5dB SIR for a CW tone interferer. With a 16-QAM interferer input, the SIR reduces to -19.5 due to the spread energy from the signal. However, the non-linear nature of the guard bands and WPDC algorithm makes it such that with an extra 3dB RF input signal power the SIR performance jumps up to -32dB.

As previously discussed, the WPDC method improves power in the assumption that the network activity remains relatively low. The results of a model for the impact activity factor versus average power consumption is shown in Figure 5.27 for the 10ms average latency point. Two points on this curve were measured on the chip to ensure agreement with the model. With a mod-

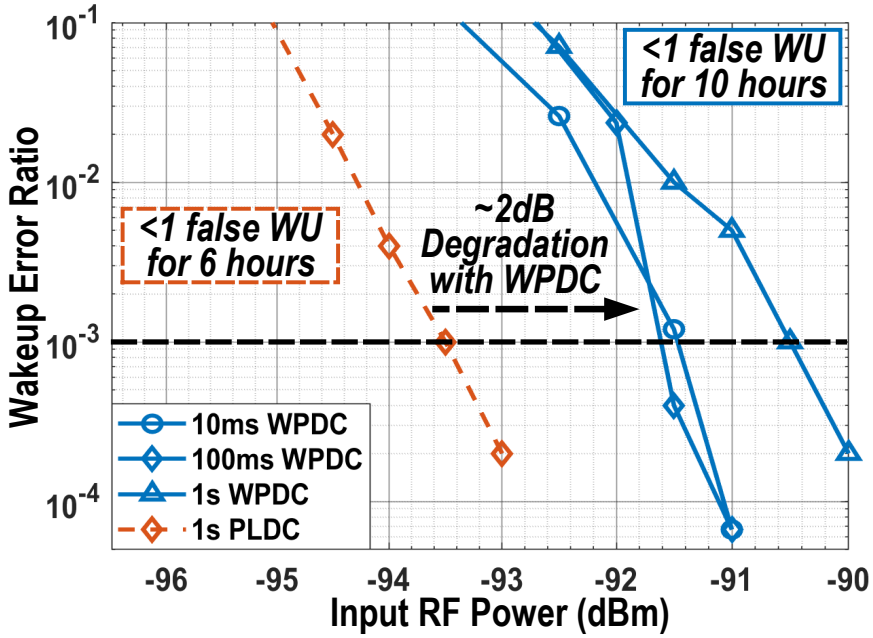


Figure 5.24: Wake-up error ratio for both WPDC and PLDC modes showing that the sensitivity target was reached (better than -90dBm).

Duty-Cycle Mode	Sensitivity (dBm)	Power (µW)	Latency (ms)
WPDC	-90.5	0.9	1000
WPDC	-91.5	2	100
WPDC	-91.5	20.9	10
PLDC	-93.5	2	1000

Figure 5.25: Table of measured WER points with associated average power and latency values.

est 1000 wake-ups per hour, the predicted and measured increase in power consumption was only 3.2%. In order to double the WuRX power, it took 36,700 wake-ups per hour. If the activity level keeps increasing beyond this point, then WPDC power will asymptotically approach PLDC power level.

A comparison table similar to the other presented wake-up radios is given here to compare

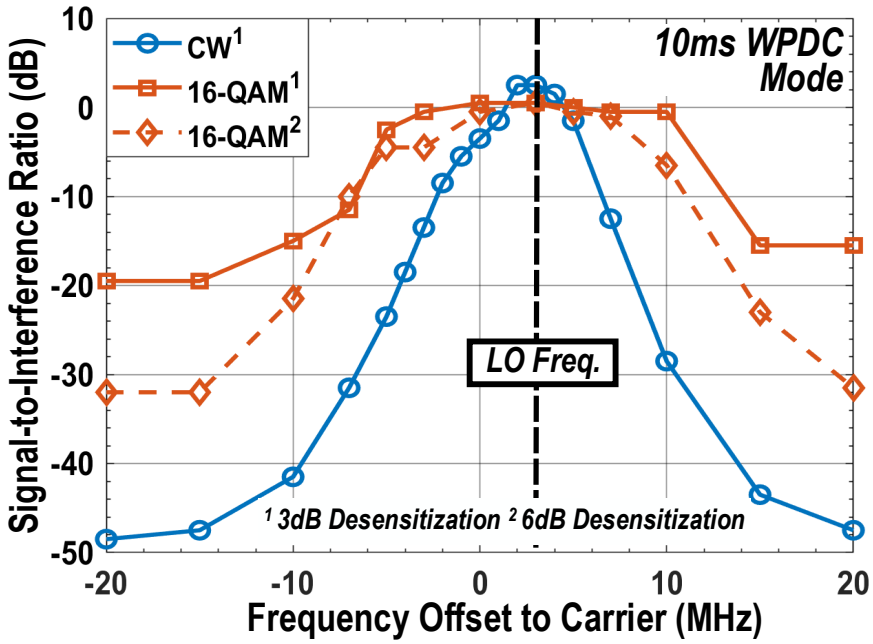


Figure 5.26: SIR performance in the WPDC mode with constant wave and 16-QAM interferers.

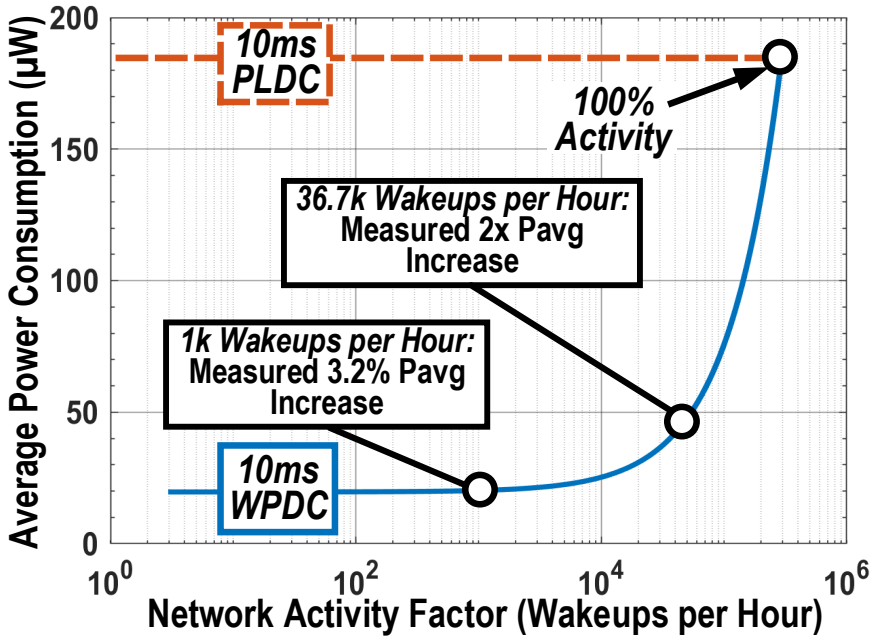


Figure 5.27: Wake-up radio average power in WPDC mode when varying the level of network activity (β) in terms of wake-ups per hour.

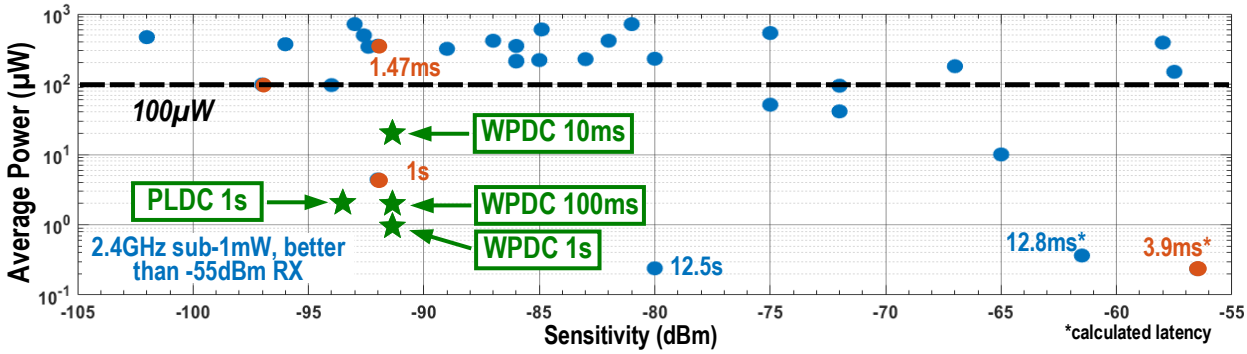


Figure 5.28: Scatter plot of this wake-up receiver in comparison to other state-of-the-art receivers at 2.4GHz at the time of publishing. Data taken from [19].

this WPDC and PLDC WuRX against current state-of-the-art works that operate below 1mW P_{avg} in Figure 5.28. The three WPDC points and one PLDC point are shown in green. Latency is also denoted to help further differentiate this work from other points that may achieve similar sensitivity and power, but different latency.

5.3 Contributions

Each of the four presented wake-up receivers broke new ground in order to further the capabilities for future low-power IoT systems. The Samplifier series of wake-up receivers enabled nW-level, long-range wake-up capabilities in a variety of flavors while the WPDC/PLDC WuRX showed that it demonstrated the performance necessary to pair well with a variety of low-power IoT SoCs to enable more energy-efficient communication in the extensive 2.4GHz ISM wireless infrastructure. A final cohesive scatter plot of the presented works is shown in Figure 5.29 to show how the limits of wake-up receiver sensitivity and power performance was pushed over the course of these works. The bit-level duty-cycled wake-up receivers demonstrated power reduction for similar sensitivity of better than 1000x and sensitivity improvement greater than 20dB for similar power consumption. The within-packet duty-cycled wake-up receiver demonstrated a 2.2x improvement in power consumption and 10x improvement in latency compared to the current

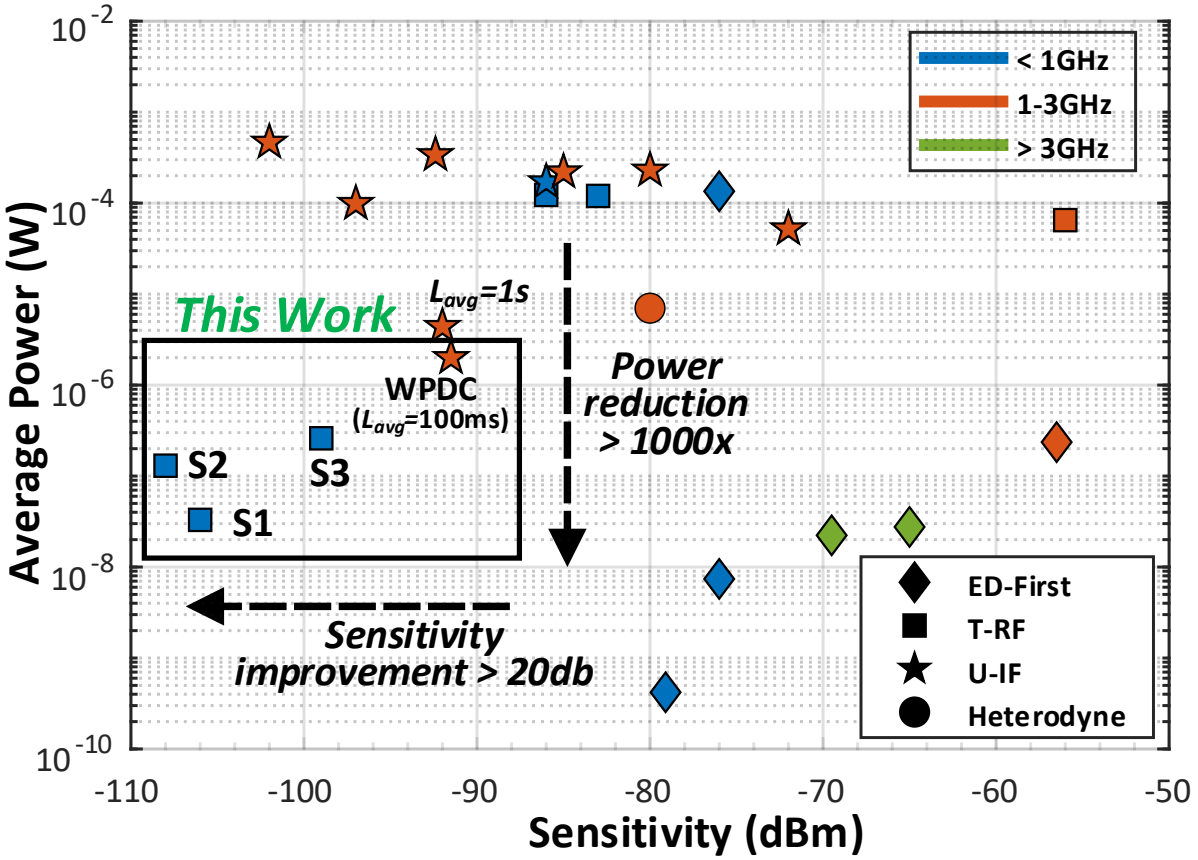


Figure 5.29: The four champion power and sensitivity points for each wake-up receiver presented in this chapter compared against other work.

state-of-the-art at 2.4GHz.

The following is a summary of the contributions of the work presented in this chapter:

1. Contributed to the development and testing of a bit-level duty-cycled wake-up receiver that improved state-of-the-art sensitivity by 26dB (compared to other low-power wake-up receivers) and reduced power by over 1000x (compared to other receivers with similar sensitivity)

2. Contributed to the development and testing of a second-generation bit-level duty-cycled wake-up receiver that enabled high digital reconfigurability in sensitivity, power, and latency by 11dB, 410x, and 672x respectively in order to satisfy a wide variety of IoT application

requirements without requiring silicon redesign

3. Contributed to the development of a third-generation bit-level duty-cycled wake-up receiver that achieved high levels of integration and multi-channel wake-up capability all at nanowatt power for lower cost and area as well as denser wake-up networks

4. Contributed to the development of a 2.4GHz packet-level/within-packet duty-cycled wake-up receiver demonstrating a 9x and 10x improvement in power and wake-up latency compared to the packet-level duty-cycling mode and a 2.2x and 10x improvement in power and latency compared to prior art

CHAPTER 6

SYSTEM-LEVEL MODELING FRAMEWORK

6.1 Background and Prior Art

IoT systems that have a readily available source of power are primarily limited in performance by the hardware of the device. In systems with low energy availability however, the full functionality of the system cannot be utilized because of power and energy constraints. In most cases, the power and energy cost of functionality is critical to assess and model before design and deployment. If this doesn't happen, then excessive power consumption can easily degrade the lifetime and dependability of the system. In both cases, it is up to the designer and user to allocate power and energy consumption. The decision process behind optimally spending energy for different types of functionality can be complex. A full integrated circuit system is comprised of many components with even more possible operating states, where each state has varying levels of power consumption over time. A system's hierarchy can expand at a given layer, horizontally, and in increasing levels of specificity, vertically.

System modeling and simulation can dramatically reduce the difficulty of analyzing a complex structure. Some of the most popular and widely adopted wireless sensor network (WSN) simulators found in the literature focus on modeling network functionality for energy-limited systems [59, 60]. Castalia is a discrete event simulator built on OMNeT++, where unit models that represent sensor nodes connect over a common wireless channel model. An extension on Castalia, called GreenCastalia, incorporates models more useful for energy harvesting systems. This includes definitions for energy harvesters, energy storage, and energy management algorithms. However, GreenCastalia doesn't expand on Castalia's simulation capabilities and can only report on

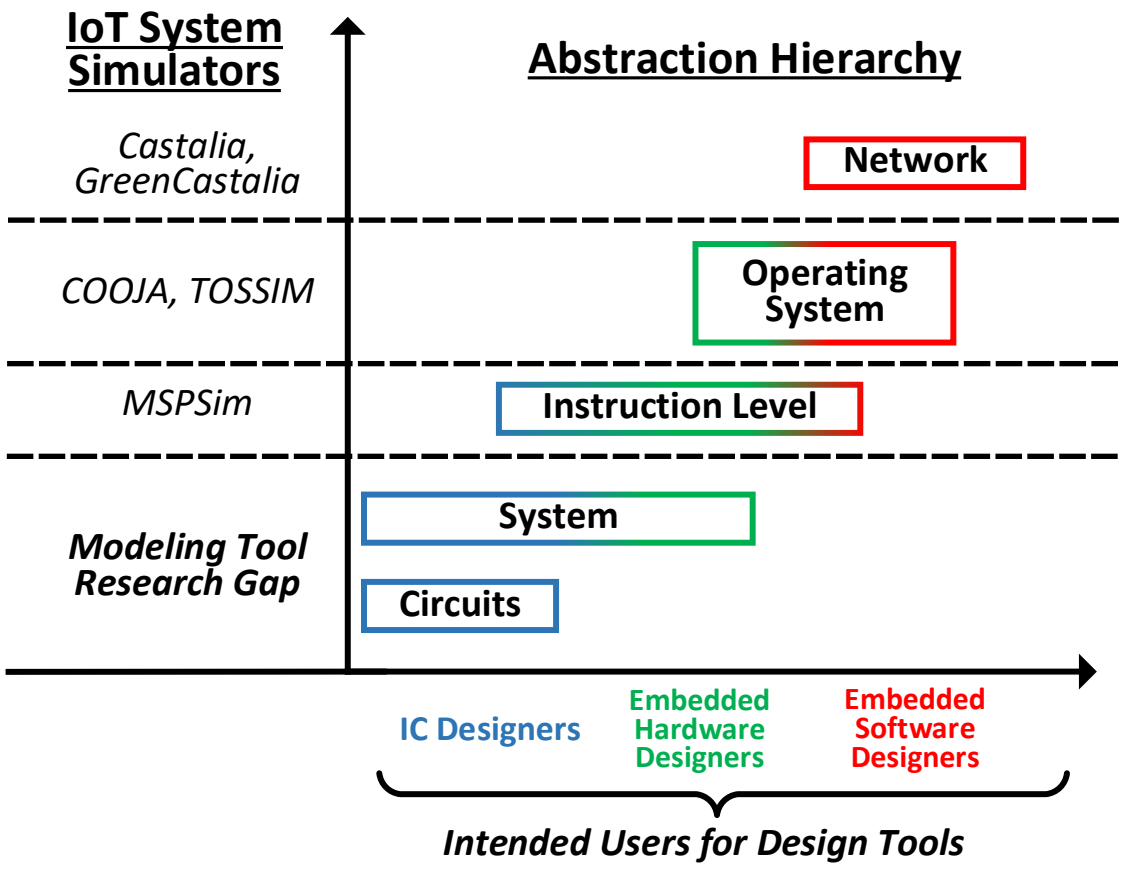


Figure 6.1: Many simulators for IoT systems and networks appear in the literature, but none are built for circuit designers.

network-related characteristics. Other WSN simulators model lower layers of abstraction hierarchy. Sensor network operating systems, particularly Contiki and TinyOS, can be emulated through the COOJA and TOSSIM simulation platforms respectively [61–63]. Both are able to simulate the coexistence of numerous nodes communicating together. In a limited capacity, they support representations of hardware power consumption as well. At a lower level of abstraction, an example of instruction-level emulation is supported for the popular TI MSP40 ISA with the MSPSim simulator. Each of these tools performs well in their respective areas: network, OS, and instruction-level considerations for wireless sensor networks. But in each case, weak support for circuit and hardware definition and modeling prevents these simulators from being useful to circuit and hardware

designers because they don't provide any effective insights into power consumption. The resulting design tool research gap for effective circuit-to-system design and analysis of energy-limited devices is depicted in Figure 6.1.

6.2 Proposed System Modeling Framework

The proposed system modeling framework allows the user to virtually prototype the development of low-power systems and their operation in applications to determine relationships and trade-offs between design, operation, and power consumption. The Python-based system power modeling framework described in this section consists of several core parts all of which allow the user to describe arbitrarily complex low-power systems, depending on the desired level of specificity. The first section discussed concerns modeling the structure of the system in question. This encompasses several degrees of freedom for the user concerning: 1) what components currently comprise the system, 2) how components are currently connected in the system, and 3) how supply rails are provided to components in the system. These relationships can be well expressed with a tree data structure. There are three primary categories (implemented as classes) of nodes in the tree: 1) components, 2) groups of components, and 3) voltage regulators. However, the user needs to do more than simply define the system structure; they need to define how power is calculated based on relevant information. In a standard low-power IoT system, there are a wide variety of components and sub-components which can be generally categorized by digital, analog, RF, and power management. Each has their own set of relevant design knobs and application variables that will impact their power consumption. For example, an analog front-end (AFE) can be used to interface with the human skin to perform electrocardiogram (ECG) monitoring and is characterized by a given sampling rate. On the other hand, a wireless receiver could have several possible architecture choices for downconversion from RF requiring differentiating models and variables for each option. As a result, this framework uses defined model and variable objects to handle the wide breadth of possible ways each component in the hierarchy can have its power calculated across experiments. In effect, variables can plug into any model or set of models. These models

in turn are then associated with specific component objects throughout the hierarchy. Systems aren't designed and programmed at the individual component level. Many times designers refer to different modes of operation, such as "low-power" mode or "high-performance" mode, referring to a particular way all the components act together with different operating points. This concept is applied in this framework by encapsulating a one-to-one relationship between components and models throughout the hierarchy as well as variables and static operating points. Finally, the system class encompasses both variability in hierarchy as well as modes to allow for the most flexible and least congested coding experience for modeling complex experiments. Each of these concepts is further explained in the following sections.

6.2.1 System Structure

Component-like Objects. In this modeling framework, a system's hierarchy is built upon four specific types of "component-like" objects, which are outlined in Figure 6.2. The first object of note is the "Component" object, which is the only type of component-like object that can exist as a leaf node, when considering the system as a tree structure. It is also the only type generally attributed with models because they represent the most atomic level of specificity in terms of power consumption. The class assumes a universal duty-cycling model of power consumption for every object. The power for a duty-cycled component is represented by

$$P_C^N = f_c(v_1, v_2, ...) \approx \alpha_C(P_{act,C} - P_{in,C}) + P_{in,C} \tag{6.1}$$

where v_i represents the relevant variables that affect the particular component's power consumption P_C^N with a depth in the tree structure of N. Each of the four values in the duty-cycle equation can be controlled by user-defined models. For example, the duty-factor is appropriate to change in many software contexts, but the designer could also see the impact of changing the inactive power of the component, which could be achieved by reducing leakage current in hardware. The user could also choose to completely override the built-in duty-cycle model to manually calculate average power based on a more complicated set of relationships. Component objects can be defined in terms of either power or current as well. When defining current, a voltage rail must be specified. Only

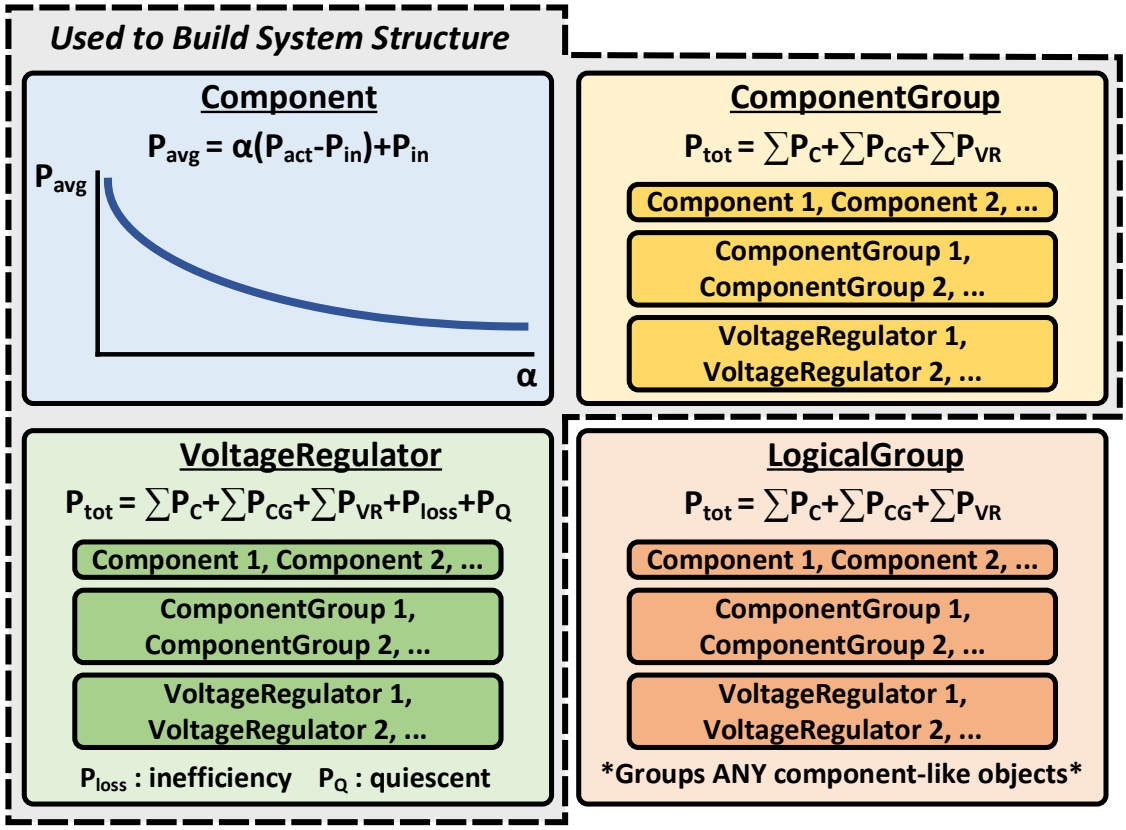


Figure 6.2: The four base types of component-like objects used to describe connection and hierarchy in terms of power consumption.

defining power consumption is best used in the earlier design phases of a system or when only a quick analysis is required.

Once Components have been defined, the user should specify how they group together. This is achieved with the "ComponentGroup" object. The ComponentGroup object is what partially helps define the hierarchy of the system by allowing an arbitrary number of component-like objects to exist "beneath it". When a system is defined in terms of voltages and currents, the ComponentGroup checks to make sure that everything is operating at the same rail level and is useful for abstracting sections of the system away. A method is available to recursively traverse through the existing structure and sum the total power or current of the particular ComponentGroup. The total power for a particular ComponentGroup that is the parent node to other component-like objects is

given by

$$P_{CG}^{N-1} = \sum_{c=1}^{C_{tot}} P_c^N + \sum_{cg=1}^{CG_{tot}} P_{cg}^N + \sum_{vr=1}^{VR_{tot}} P_{vr}^N \tag{6.2}$$

where P_c^N is the power of all the Components, P_{cg}^N is the power of all the ComponentGroups, and P_{vr}^N is the power of all the objects described next, which are VoltageRegulators.

Another class similar to the ComponentGroup is the "VoltageRegulator", which is responsible for both representing system structure and specific aspects of supply regulation. In addition to the same methods as ComponentGroup for calculating the power of components beneath it, the VoltageRegulator needs to account for the inefficiencies in supply regulation as well as static power or current consumption as a result of regulator circuitry. Both efficiency and regulator power are parameters which could be controlled by custom user models as well. Similar to the ComponentGroup, the total power for a particular VoltageRegulator is given by

$$P_{VR}^{N-1} = \sum_{c=1}^{C_{tot}} P_c^N + \sum_{cg=1}^{CG_{tot}} P_{cg}^N + \sum_{vr=1}^{VR_{tot}} P_{vr}^N + P_{VR,loss}^{N-1} + P_{VR,Q}^{N-1} \tag{6.3}$$

where the $P_{VR,loss}^{N-1}$ and $P_{VR,Q}^{N-1}$ are the power loss due to regulation inefficiency and the quiescent power consumption of the regulator itself.

The final object type isn't used to define the system hierarchy, but it is useful when considering how a user would want to analyze the results of certain modeling experiment. As mentioned before, a system has many components and subcomponents that are set on a variety of possible supply rails. For example, an AFE component may comprise of amplifiers, filters, an ADC, and digital accelerators. For improved power performance and isolation from sensitive analog components, the digital may be separated onto a separate, lower-VDD rail which could run digital from other components as well, such as the processor or memory. As a result, the complete AFE component is scattered across the hierarchy and total AFE power can't be reported by itself. The LogicalGroup is used here to specifically collect any arbitrary set of component-like objects, without respect to the real hierarchy, in order to provide customized power calculations across the system. This is the only differentiating feature it has from the ComponentGroup.

Creating a System Hierarchy. The easiest way to model the system structure is to start at the bottom and work up. Components that will be grouped together, with a ComponentGroup or VoltageRegulator, should be defined early. At instantiation for the ComponentGroup or VoltageRegulator, the relevant component-like objects should be included as parameters to the objects in one list per component-like object. As the hierarchy is built in this way, there will end up being one final ”top level” ComponentGroup or VoltageRegulator used for modeling purposes. Typically, this object will be the main point of reference when running simulations and reporting power consumption. An update needs to be applied on the top-level component-like object in order to successfully link the structure together. Optionally, the user can add or remove component-like objects depending on the experiment. However, this is better done through the use of the System class described later.

6.2.2 Customizing Component-Level Modeling

Relevant Variables. The power consumption of each component may have a large number of relevant attributes that help define it. In this modeling framework, the Variable class represents meaningful values that are relevant to the user's modeling experiments. Operating variables, such as sampling rate, determine how the system is used and can generally be redefined in software to impact power. Design variables, such as memory size or voltage rail options, are determined by the circuit designer. Even though post-fabrication this type of variable is not programmable, it can play an important part of the design process just as much as operating variables would in the modeling environment. The way Variable objects are defined and used are up to the user. They are characterized by a static operating point (the current value) and an array of values necessary for performing sweeps of the Variable. These can be changed at any time if desired and are often plotted on the x-axis of modeling results, given that they are relevant ”inputs” for the experiments.

User-Defined Models. Defining Variables would be relatively meaningless without also defining how they map to power consumption. The Model class gives the user total freedom to determine what that relationship looks like by containing an arbitrary list of Variables and a user-

defined function to relate variables to component power consumption. Inside the user-defined Python function, the relevant variables are called and used in some type of calculation to generate a result that could be set to change the associated Component's active power/current, inactive power/current, duty-factor, or average power/current. Each Model object is provided as a parameter to the associated Component object(s). A single Component can have any number of Models to describe it, but only one can be considered "active". This can be set with a method call. For example, there may be numerous Models to describe a transmitter Component where each Model has unique Variables that vary the power consumption of the transmitter. Before that Component object can be used, a specific Model would need to be defined otherwise the default values would take over.

6.2.3 Modes and Systems

Many times an entire system can be characterized by how the "sum of the parts" all work together in terms of power and performance. System modes are an important part of abstracting away the challenge of defining how each and every component operates individually from one application or context to another. The Mode class is used to achieve this by grouping two sets of objects together. Each Mode has two lists mapping what Components should be defined by which specific Models. Each mode also has another pair of lists that map Variables to specific operating points (their "static" value when not being swept). In this way, a single Mode object describes how an entire system operates, how power is calculated, and what Variables are relevant. Modes contain their associated system and numerous Modes can be made per system. To shift from one Mode to another, a basic method call is required. However the Mode class only helps the user define how the system operates to simplify the coding process for creating experiments. The System class expands on this by combining the ability to easily change Modes as well as the actual system hierarchy. The class supports an arbitrary number of system "variants" where each variant represents a literally system structure and can have an arbitrary number of Modes. In this way, the user can do the preparation ahead of time to create a large number of different system structures

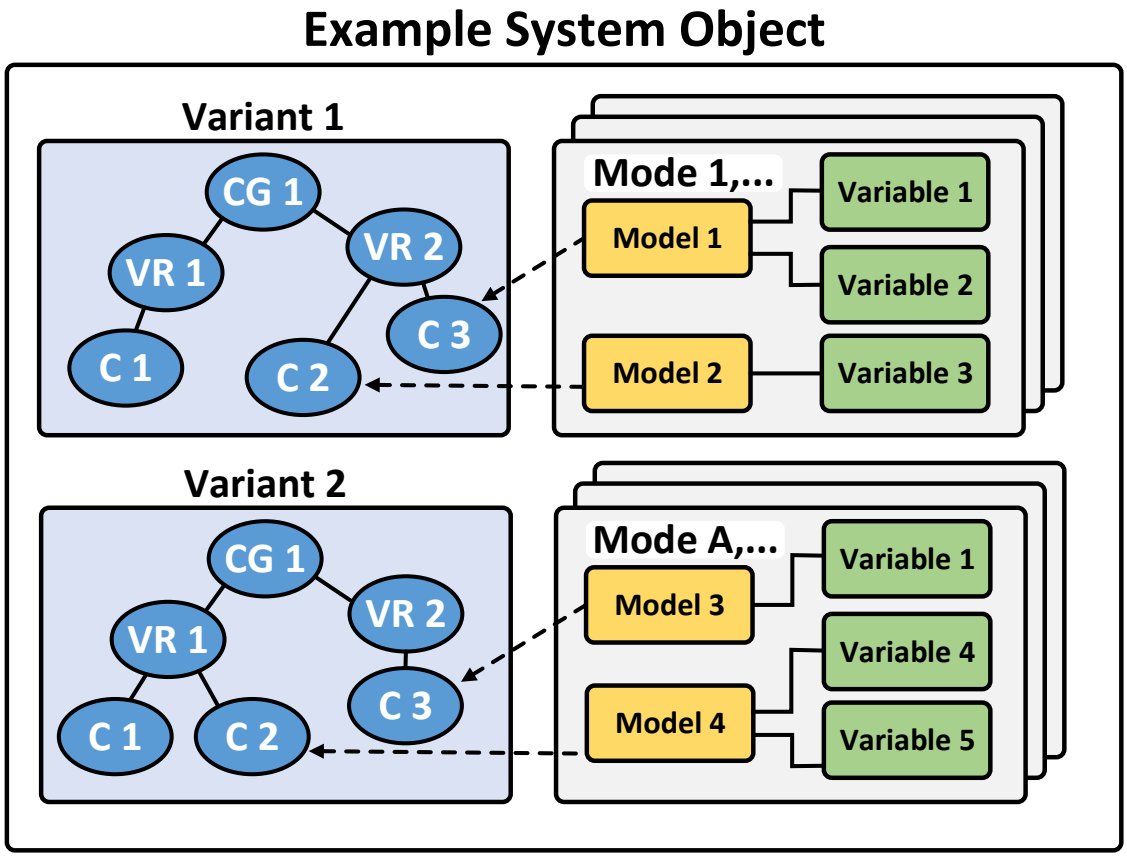


Figure 6.3: An example representation of a complete system model developed in this modeling framework.

and modes of operation and go about changing them with just one command. An example system model using all of the mentioned class structures is in Figure 6.3.

6.2.4 Built-In Experiments

All the classes presented so far only describe how the system is modeled; they aren't capable of running modeling experiments on their own. To enable this, a single common module is used to hold generic functions for calculating and plotting results. This framework utilizes the Plotly library for Python to provide a clean and more interactive interface for viewing results. There are several key built in experiment types that are reusable for a number of situations. The first function

sweeps an input variable (v_i) and reports power consumption (P) for a particular structure that the user previously defined. A simple relationship for the function can be described as $P = f(v_i)$. A similar function supports a 2D variable sweep. The second allows the user to set a target power consumption (P_t), at the system top or anywhere throughout the hierarchy, and let the tool tune a particular variable (v_t) until that hierarchy reaches the desired set point. From the user's perspective, this can be described by $v_t = f^{-1}(P_t)$. The third assumes a required power budget for the system and trades off two variables (v_1, v_2) while maintaining a constant level of power consumption ($P_{budget} = C$), within a tolerable deviation (Δ). The fourth calculates the potential lifetime of the system given a fixed amount of energy to begin with. A 2D version of this calculation is available as well. A sunburst plot is used to visualize the hierarchy itself and power consumption throughout, which is useful for checking operating points and comparing different modes of operation.

6.3 Modeling Framework Example: Compute versus Communicate

A series of realistic modeling experiments is presented in this section around the theme of "compute versus communicate" in order to present the utility of the proposed modeling framework. This concept of "compute versus communicate" is a common trade-off found in IoT applications. Often, it is advantageous to pre-process some data locally at or around the edge of the IoT rather than transmit all information to the cloud. In this example, a simple system model is created using the framework based on a real low-power IoT system for biomedical applications. The key components in the system are shown in Figure 6.4. The goal of the application is to continuously sense a user's ECG waveform through the on-chip analog front-end, then transmit the data to a nearby Bluetooth-enabled cell phone. An app was developed to display the resulting ECG information transmitted. Previous low-power systems developed in custom silicon were used to implement this application [64,65]. This current system represents the next generation of the implementation with

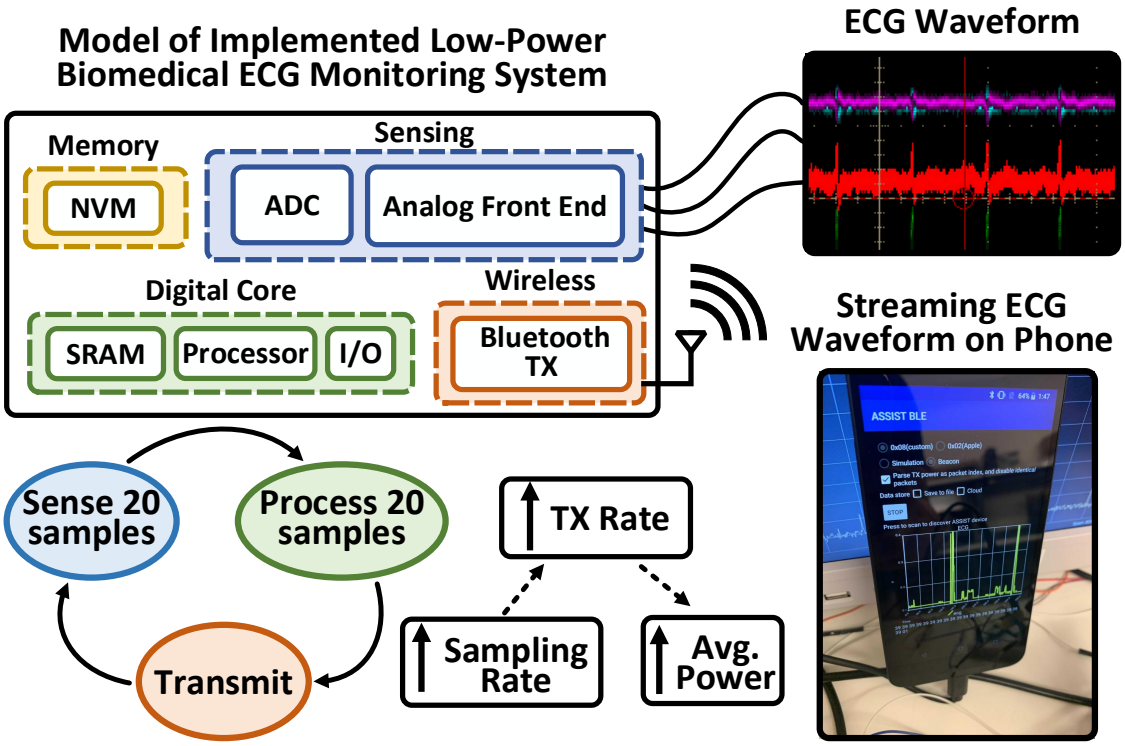


Figure 6.4: Low-power IoT biomedical system modeled as an example with the proposed modeling framework.

improved sensing and processing components. The repetitive operation intended for this application is to first sense 20 samples of ECG data, limited by the packet size of the Bluetooth transmitter. Second, the node has the opportunity to process the data if desired. Then third, the data is transmitted to the phone with asynchronous Bluetooth communication via beacon packets. Unfortunately, the asynchronous nature of this communication method suffers from missed packets because the Bluetooth receiver on the phone periodically checks all three of the advertising channels while the data only transmits over one. At best, the missed packet ratio is around $1/3$. This requires that data be re-transmitted by default, which costs extra power. In this model, the sampled data is transmitted twice with one packet right after the other. An important relationship to recognize based on the assumption of this three-state operation is that the AFE sampling rate is intrinsically tied to the transmission event rate. Thus, if the sampling rate increases then the system power can increase significantly because it modulates the TX rate. Real-time operation also matters in this context,

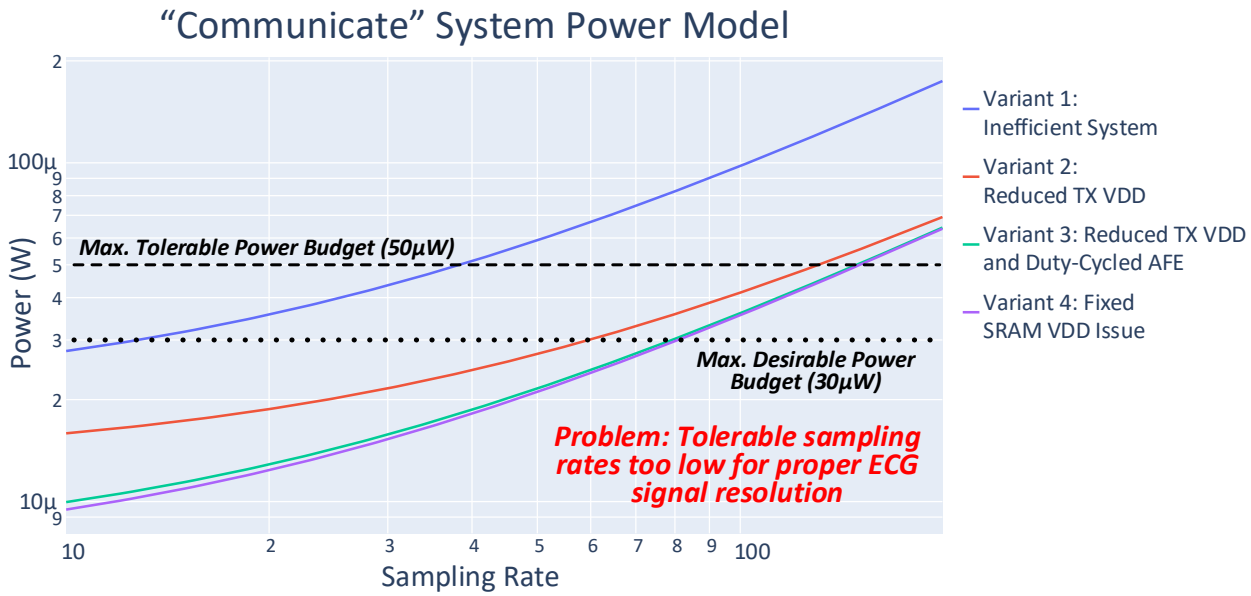


Figure 6.5: Modeled system power consumption over the course of several anticipated improvements as a function of ECG sampling rate against system power budgets.

so data should appear in a relatively fluid manner. This is the starting context for these modeling experiments as it represents the starting point for a hypothetical team developing and testing the system.

The modeling framework is used to model each of the components shown in Figure 6.4. Active and inactive current consumption values are derived from measured silicon results. Key variables, including sampling rate and transmission rate, are assigned to numerous models associated with the AFE and transmitter. A top-level System object is defined which encapsulates several different variants in the architecture of the system structure. Variant 1 represents an inefficient system architecture with several design bugs that forces it to operate at a higher power than desired. This is the current state of the system in question. The hypothetical design team has developed several solutions, in phases, to solve the challenges posed by Variant 1. Variant 2 addresses a large power bottle neck, which is the voltage rail that the transmitter operates at. In Variant 1, the custom TX only worked at 3.3V with LDOs supplying lower noise on the supply line, but in Variant 2 this is considered fixed and run at the intended 1.2V rail. In Variant 3, the AFE has a power switch

used to duty-cycle the higher power buffers on its 1.2V rail. Finally, Variant 4 represents a respin of the digital core to fix a slight issue with the SRAM so it remains compatible with the originally intended VDD at 0.6V. With all this context first built in the modeling framework, the following examples only require several lines of Python code to create.

The first modeling experiment conducted concerns the system and operation explicitly shown in Figure 6.4 based on a "Communicate" operation paradigm. The AFE sampling rate is the most relevant variable in this case, as it limits how much resolution the ECG waveform on the phone has. From tests in the lab, the designers determined that a sampling rate of several hundred samples per second was ideal for showing ECG QRS peaks on the phone. All four system variants' power consumption are plotted as a function of sampling rate in Figure 6.5 using the built-in sweepVariable function. A maximum tolerable power budget of $50\mu W$ is set with a desired maximum of $30\mu W$, so it is ideal to go below this point. Off the bat from this figure, even in Variant 4 the sampling rate is still below 100 samples per second for the target power value. As a result, something needs to be done in order to achieve the necessary resolution at the required power point.

In order to achieve the desired level of resolution on the phone screen, a change is proposed to the system such that it samples at a higher rate, 300 samples per second in this case. In addition, the digital core performs some pre-processing to search for the QRS peaks which are of primary interest to make sure they get to the phone, while leaving out most of the less useful signal information in between peaks. This revised "Compute" system is plotted as a function of transmission rate now that the sampling rate is fixed. In this case, the designers wish to see for the specified target power ($30\mu W$) what the achievable transmission rates are for every Variant. The target is to achieve greater than three transmission events per second. Figure 6.6 reports that with the tuneVariable function Variant 1 will not be sufficient to satisfy the requirements as it can only achieve 0.6Hz TX rate, so architectural changes are still required. But Variants 2, 3, and 4 all achieve TX rates of greater than or equal to three times per second which is fast enough to report in real time the ECG waveform. By improving and fixing the hardware, the relative power improvement by going from Variant 1 to 4 in the "Compute" mode is approximately 2.6x. The final experiment wishes to explore how much moving to the new computing scheme saves in terms

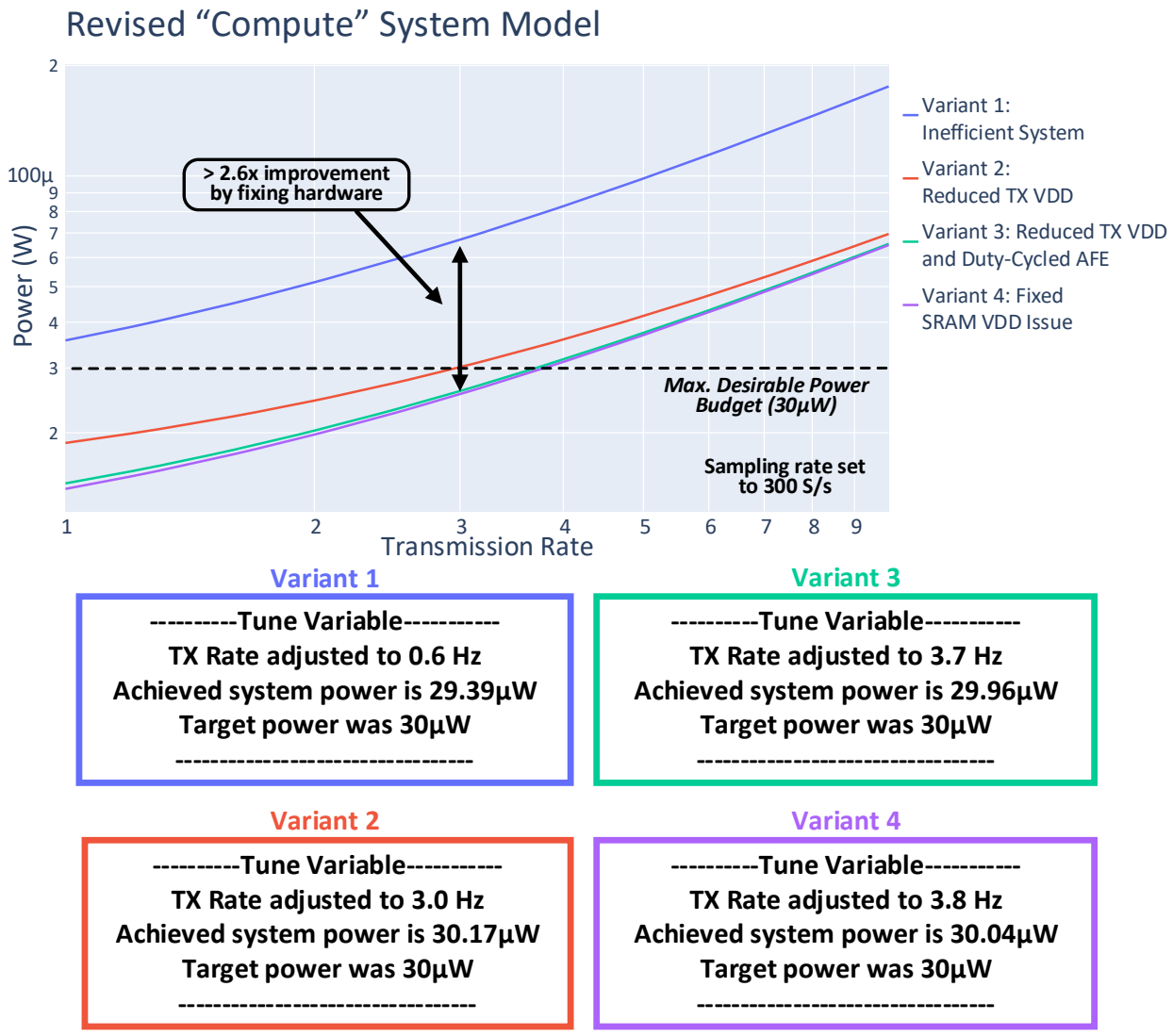


Figure 6.6: By incorporating some level of pre-processing, the system can save on average power. The tuneVariable function is used to show the achievable transmission rate at the desired system power.

of power consumption relative to the original communicate operation. This comparison assumes that both systems' AFEs are operating at the desired 300 S/s. Figure 6.7 shows the power improvement of the computing system versus the pure communication system originally described. At the desired transmission rate of three events per second, an estimated 3.2x power savings is achieved in the worst case. Interestingly, the power improvement associated with system variant

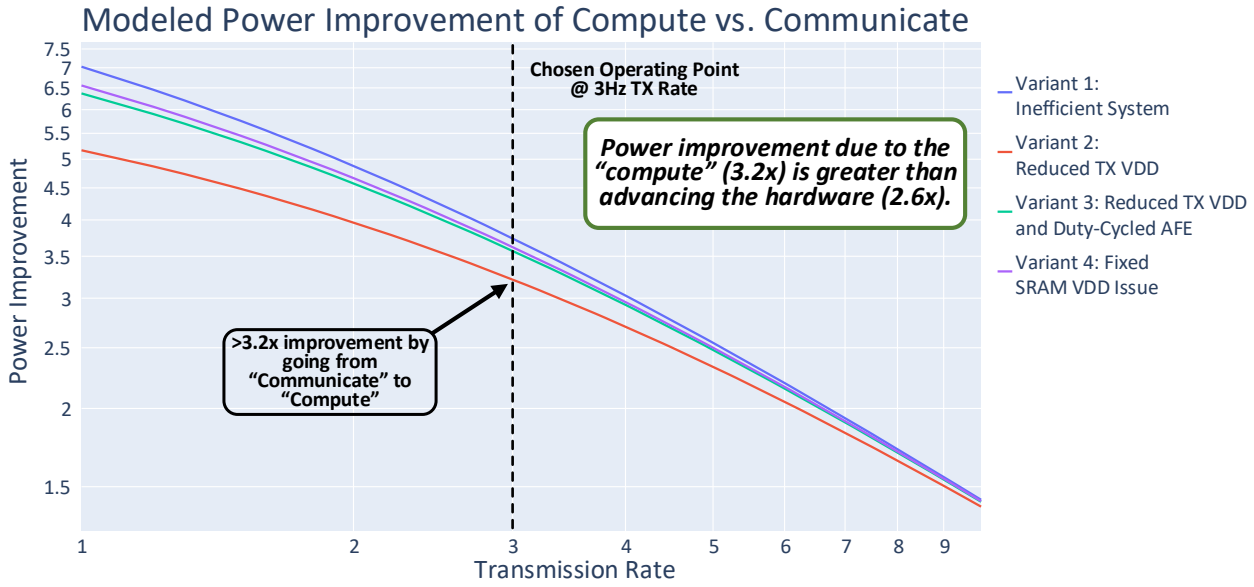


Figure 6.7: Modeled power improvement by moving from communicate to compute-based operation as a function of the transmission rate.

changes in Figure 6.6 are relatively less than that of the improvements achieved by switching to a computing-based operation. Variant 2 has the least power improvement of the group in Figure 6.7 primarily due the always-on AFE adding to the floor power. Variant 1 shows an anticipated 3.74x improvement while Variant 2 shows 3.21x improvement. With all this information in mind, the designers know that they need to move to a system which will perform more computation on the ECG data while also improving integration by at least remedying the TX voltage rail problem in order to satisfy application requirements. This result is unique to this specific instance and can vary across applications, systems, and requirements which is why this modeling framework is key to fast and effective virtual prototyping.

6.4 Conclusions

As the breadth of IoT applications and possible low-power hardware increases, quick and effective modeling methods become more necessary in order to optimize power consumption for the required operation. This chapter discussed a proposed system modeling framework to enable fast virtual prototyping of low-power IoT systems. It also covered an example use-case of the modeling framework based on a real custom silicon biomedical IoT system.

The following is a summary of the contributions of the work presented in this chapter:

1. Developed a new flexible system-level modeling framework in Python to easily analyze low-power IoT systems and inform users about how to tune their system to achieve desired levels of power consumption and operation

2. Developed widely applicable functions that allow the user to 1) sweep operating variables and report power consumption, 2) set a target system power and let the tool tune a given variable to achieve that power, 3) exchange two operating variables while maintaining a constant system power and review the trade-off, 4) calculate lifetime as a function of operating variables

3. Created a model and series of experiments based on a low-power biomedical system with a custom analog front-end, processing core, and Bluetooth transmitter to help analyze how to improve the system in a variety of aspects

CHAPTER 7

SYSTEM CONTEXT FOR WITHIN-PACKET DUTY-CYCLED WAKE-UP RECEIVERS

7.1 Modeling Experiment Context

Previous chapters in this dissertation covered the topics of WuRX analysis, discussion, and design at the component level. This chapter concerns the system-level impact of using wake-up receivers, and in particular within-packet duty-cycled wake-up receivers, by using the previously described modeling framework. Three sets of experiments are chosen that highlight important related concepts about the use of wake-up receivers and are implemented in the context of the previously described biomedical monitoring system (Fig. 7.1). The generic mode of operation in this event-driven ECG sensing node is now based on wake-ups no longer adhering to just a periodically-driven sense, process, and transmit behavior. The data supporting the WuRX model implemented in this framework is take from another model used to develop the WPDC versus PLDC advantage figure verified to be consistent with silicon results, found in Figure 5.22. The WuRX implements similar settings as well, including a window size of 5, correlator length of 3, and realistic start-up times for all the components. The wake-up radio is assumed to be integrated entirely onto the 1.2V rail available on the system. A single ”wake-up event” consists of a ”sense, process, transmit” operation. Exactly 20 samples are taken and the same data is transmitted twice, in order to account for inevitable packet loss. These are the same assumptions from the end of Chapter 6.

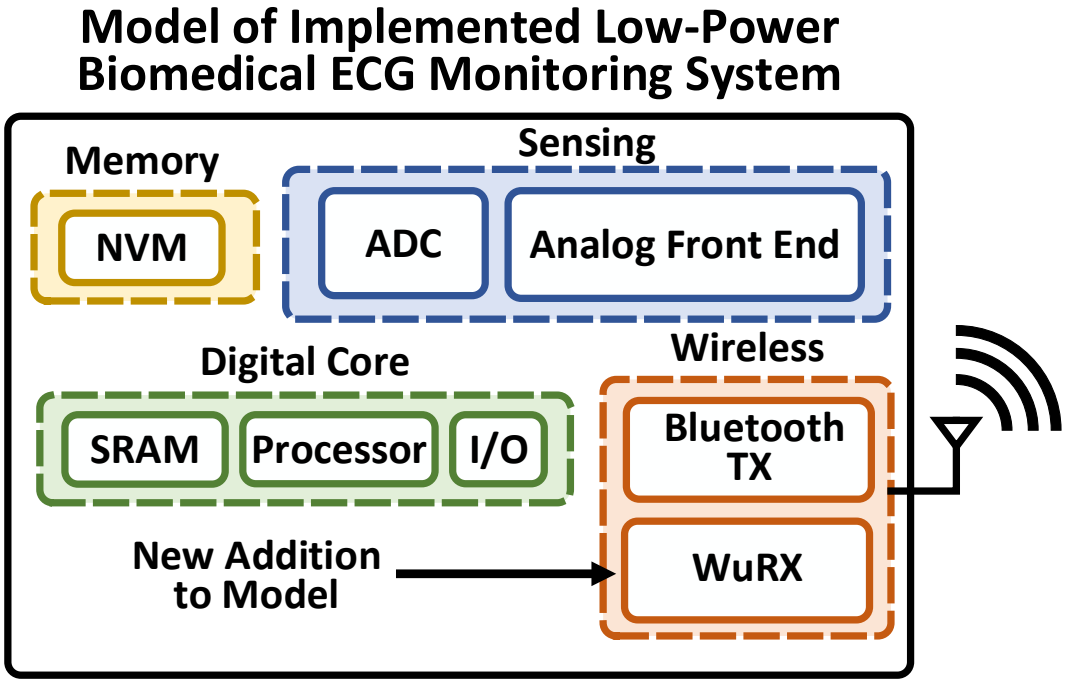


Figure 7.1: This chapter concerns the impact of adding a wake-up receiver to a system. Specifically, these modeling experiments are generated in the context of the previously described biomedical monitoring system.

7.2 Experiment 1: PLDC versus WPDC

The first experiment concerns a comparison of the packet-level and within-packet duty-cycling methods in this system context. Both wake-up receivers are programmed to operate with 100ms average latency and with a low activity factor of one event per hour. This value does not include the false wake-up rate, which differs between PLDC and WPDC operation. It was shown previously in Chapter 3 that at the component level in an active-power dominated context, the within-packet duty-cycling scheme had a fixed power advantage when compared to PLDC as a function of average latency under low activity situations. When integrated into a system, this advantage can vary depending on the power consumption of other components and activity level. To visualize this breakdown across system structure, the proposed modeling framework has a built-in ”sunburst” plotting feature to represent elements in the system as segments in a chart, where

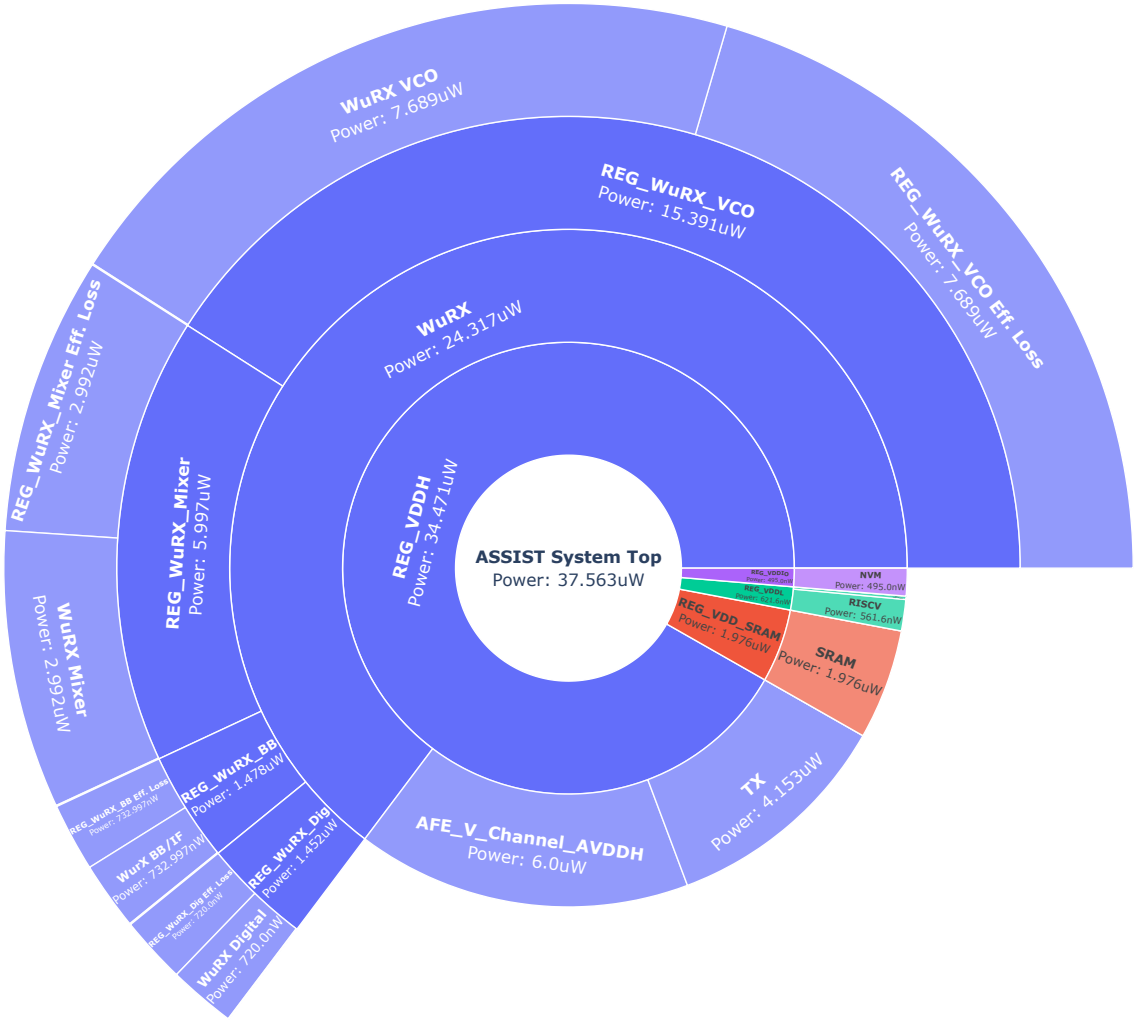


Figure 7.2: Sunburst plot showing the hierarchy of the modeled system and estimated power breakdown. The PLDC WuRX dominates the power profile in this case in part due to poor regulator efficiency (L_{avg} = 100ms and one wake-up per hour activity level).

elements ”deeper” in the hierarchy are plotted further away from center while the arc lengths of each segment represents the contributing power to the parent segment above it. Figure 7.2 shows the system with an integrated WuRX while using PLDC operation. According to the calculated power numbers below the name inside of each segment in the chart, the wake-up radio is the major power consumer in this scenario. In this example, poor voltage rail integration is one reason why

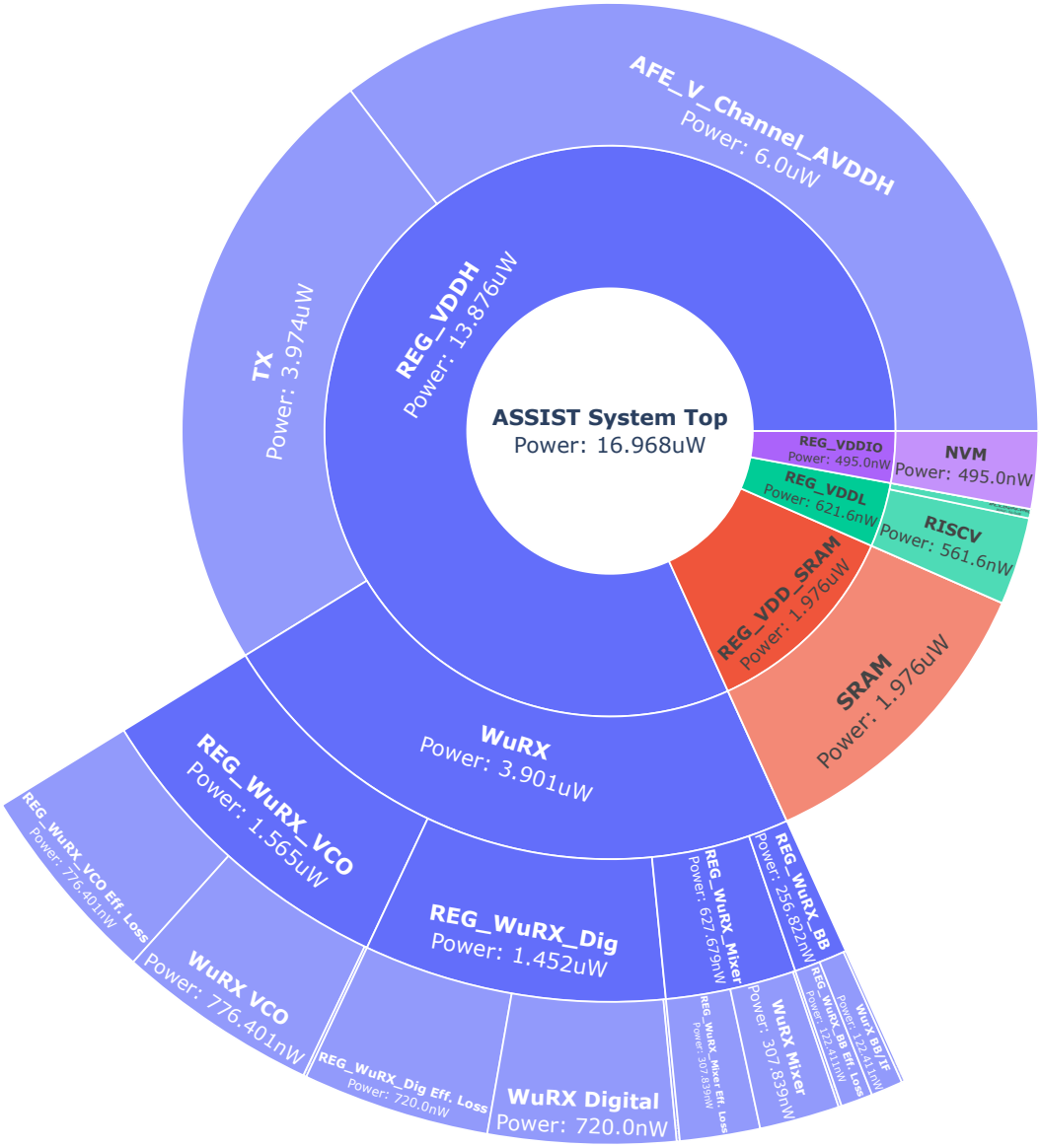


Figure 7.3: Sunburst plot of the system using a WPDC WuRX with a significantly improved component power level, but relatively the system now consumes more all together decreasing the advantage.

the WuRX consumes nearly double the power it does when standing alone. In PLDC mode, this particular WuRX should experience around 122 false wake-ups per hour based on estimations, but this actually barely increases the overall average power (less than $1\mu W$). It may be necessary to

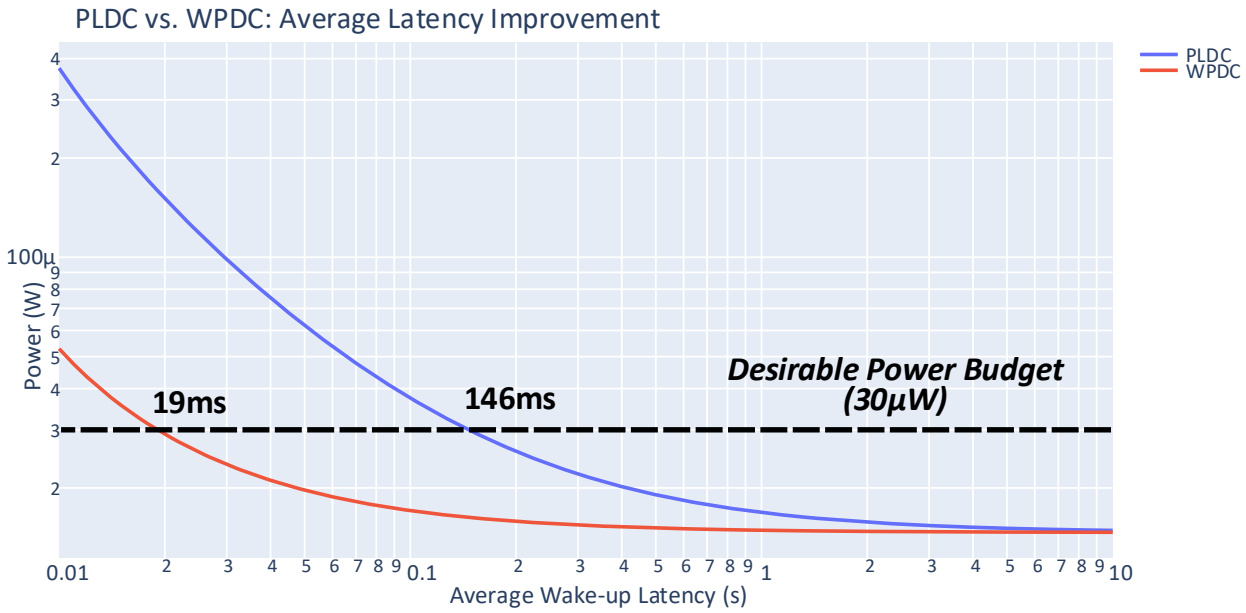


Figure 7.4: Modeled result for the effect of wake-up latency on system power consumption for PLDC and WPDC operating modes.

increase the average latency of this WuRX in order to achieve lower average power.

When the WuRX is changed to WPDC mode, the average power drops dramatically (Fig. 7.3). Based on data also in Figure 7.2, the power improvement due to WPDC at the component level is 6x. However the total system power only decreases by a factor of approximately 2.2x, assuming only one true wake-up event (meaning one sense, process, transmit event) per hour plus false wake-ups. This demonstrates how the savings from one type of wake-up receiver to another, in this case duty-cycling method, may not matter significantly because of how close they are to the floor level of the system. In WPDC mode, this system should only experience around 9 false wake-ups per hour which changes the system power by less than $1\mu W$.

While power advantage by going from PLDC to WPDC is lost at the system level, it can be partially regained in terms of average latency. For a constant power budget, in this example $30\mu W$, a horizontal line can be drawn on a plot relating latency to power consumption for PLDC and WPDC operating modes (Fig. 7.4). This modeling scenario still considered one intentional

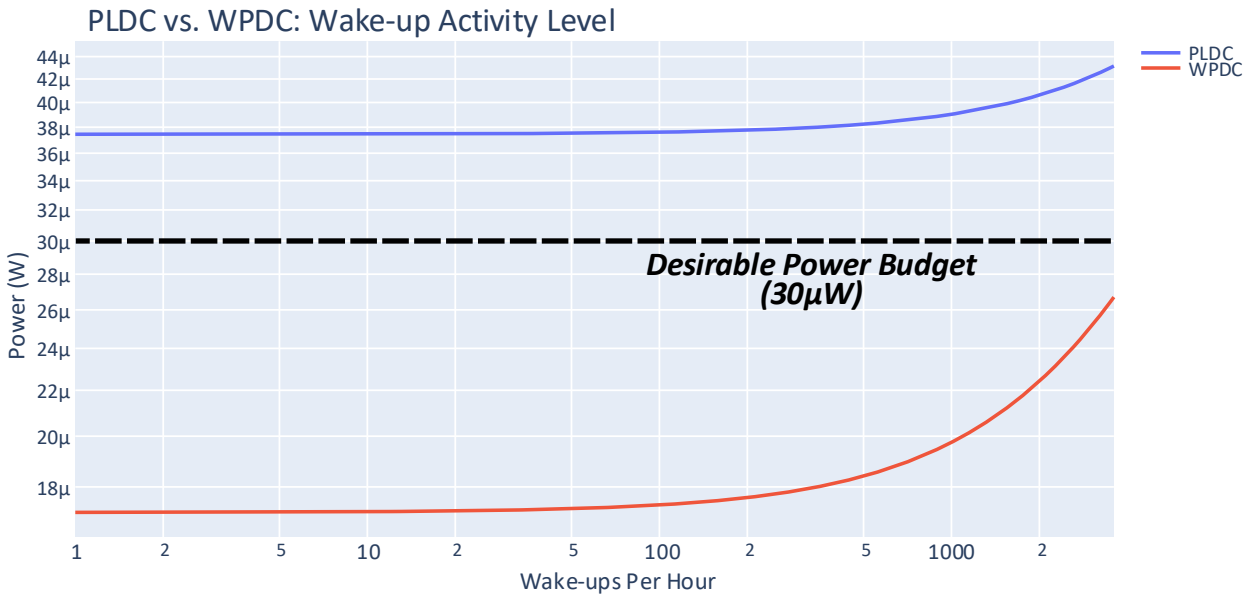


Figure 7.5: Modeled result for system activity (wake-ups per hour) increase on system power consumption.

event per hour, so the power increase is purely do to changes in WuRX operation. The WPDC setting for this system demonstrates that it can break the sub-$30\mu W$ boundary at nearly 20ms latency while the PLDC operation must be at least around 150ms average latency. This represents an improvement of around 7.7x in latency. At this system power budget with so little activity factor, the WuRX dominates system power consumption which is why in comparison to Figure 5.22 the relative ”distance” between the curves is quite similar. A new story is told in Figure 7.5 which models the effect of scaling the ”intentional” wake-ups per hour, meaning that more ”sense, process, transmit” operations are being performed. This setting is evaluated when both WuRXs are operating at 100ms average latency. As expected, at low activity levels the system power doesn't increase significantly. The WuRX operating in WPDC mode will be the first to start climbing because the average power of the system minus the WuRX will contribute a greater percentage of the power earlier on compared to PLDC mode. However even with a steeper slope, the WPDC WuRX will still not hit the $30\mu W$ line at 3000 events per hour.

7.3 Experiment 2: Event Latency and Communication Activity WuRX Trade-Off

A significant portion of the discussion in this dissertation concerning wake-up receivers has focused on the latency versus power consumption trade-off. With good reason, it is an important point to consider when considering including one in a low-power IoT system. Generally for this purpose, the wake-up receiver should be able to operate around the floor system power level so that it's own power consumption doesn't affect or strongly dominate the rest of the system which may scale in performance over time, potentially to adjust to dynamic harvesting conditions or provide longer battery life. Beyond this idea though, let's consider where the wake-up receiver is useful and when it begins to matter less.

There are three main ways that a network can be built to allow systems to initiate communication: transmitter initiated, receiver initiated, and bi-directionally initiated [66]. The most popular form is bi-directional initiated communication which enables the most flexibility because any node could request data from any other node. In the wake-up receiver context, this means that any nodes could wake up any other nodes. But all of their default states are still in listening mode, which again is part of the motivation for moving from RXs to WuRXs. However, it is foreseeable that there are also ”simple” nodes in the IoT which just transmit data periodically sending beacons to some receiver in the vicinity. Here, this type of node is referred to as ”periodically driven”. The other type of node does not encounter activity on a pre-defined schedule, hence it is ”event-driven” in nature. These two types of systems are important for the next piece of the discussion.

The new WuRX trade-off to consider is event latency versus communication activity. In other words, this represents the trade-off of ”how quickly” versus ”how often” can I send data or event-based information for low-power systems. This trade-off space consists of four extremes which are: 1) infrequent events with low latency, 2) infrequent events with high latency, 4) frequent events with low latency, and 4) frequent events with high latency (Fig. 7.6). A system operating in the first case is the primary motivation for wake-up receivers. It helps lower the event latency

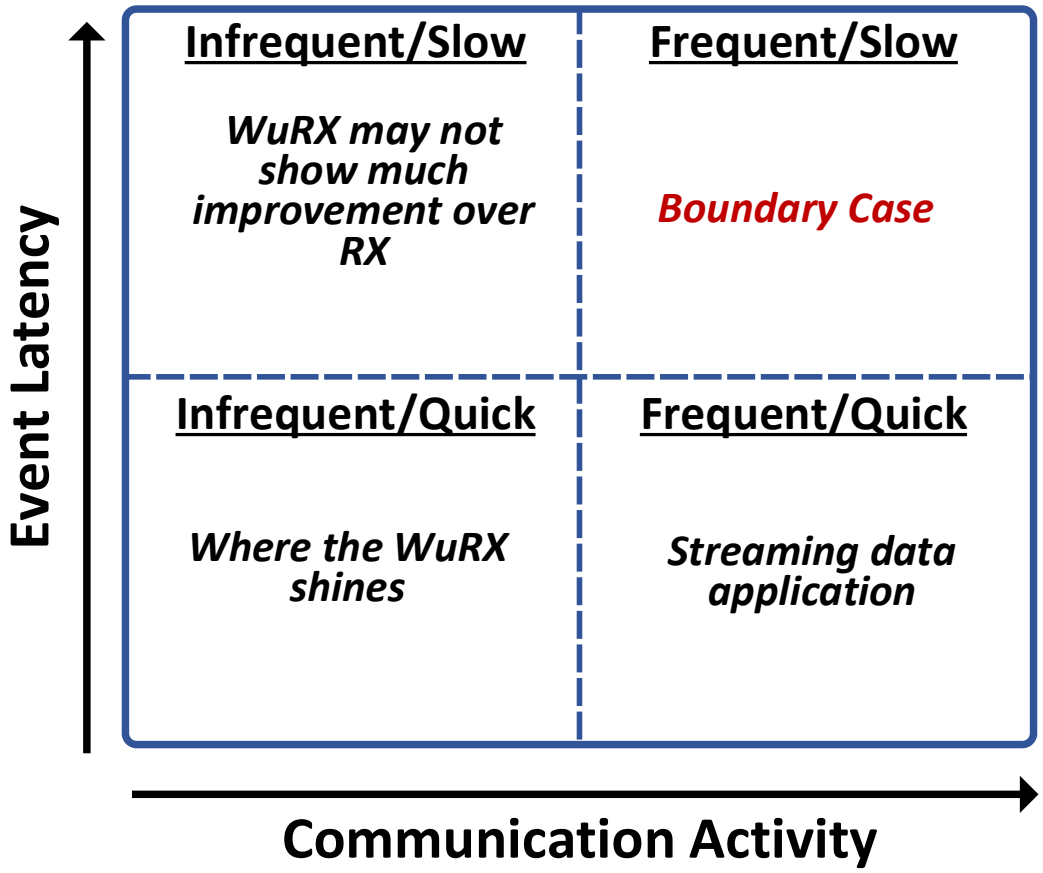


Figure 7.6: Four corner cases of how a wake-up radio fits into scenarios of high and low event latency and communication activity.

in low activity factor situations when compared to a heavily duty-cycled RX that operates at the same average power. However in applications that fall under the second category, the wake-up receiver is less impactful and a standard heavily duty-cycled RX could likely do the job nearly as well. The third category represents systems that are doing lots of tasks and reporting information often, which is classic of streaming-like operation. While these systems are still event-driven, they will have higher power relative to the original WuRX use case. The final category represents an interesting boundary case when considering its implications. If the given wake-up receiver has a particular latency, then that wake-up radio has a maximum number of events that it could handle per second. However, a system could certainly try to exceed that limit, if not bounded, which would mean that those requests would fail and therefore the application would begin to fail. In fact, this

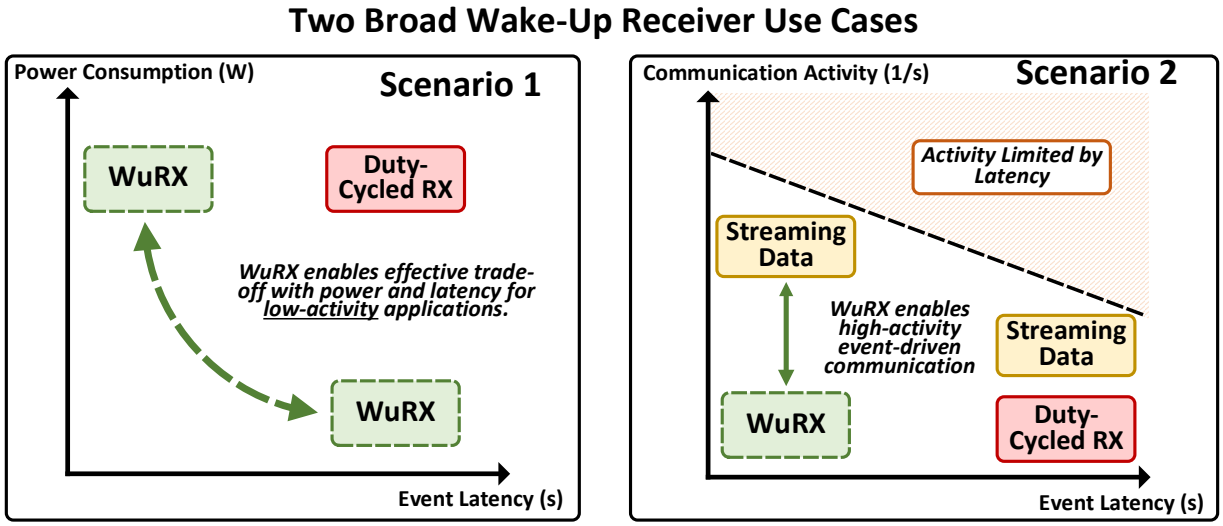


Figure 7.7: An expression of the classic latency versus power trade-off as well as a second use-case for wake-up receivers which allow for a greater event throughput or level of communication activity.

case where the event latency is inversely proportional to the communication activity describes the nature of the periodically-driven TX system. It cannot move data faster than it's latency and inherently that is also its level of activity. If one scales, then the other follows. Broadening this discussion, in comparison the event-driven system breaks this boundary because it is not forced to move back and forth on this single line but rather along the entire space underneath of it as well because it can have an arbitrary activity factor. A depiction of this concept is shown in Figure 7.7. And in this figure the next use-case for wake-up receivers appears. The traditional trade-off is shown as Scenario 1, but this new trade-off is depicted in Scenario 2. Given the four categories of operation on these two axes, the wake-up receiver is still most optimal, compared to using an RX, when acting with low latency. Therefore, when we consider how activity may be required to scale in an IoT application, an RX-based event-driven scheme will actually fundamentally limit the communication activity, which can also be thought of as event throughput, because of its higher latency. The wake-up receiver enables a greater level of possible activity, on an event-driven basis, regardless of what ends up happening to the average power consumption. This also means that a wake-up enabled node can provide power savings compared to the periodically-driven TX with a

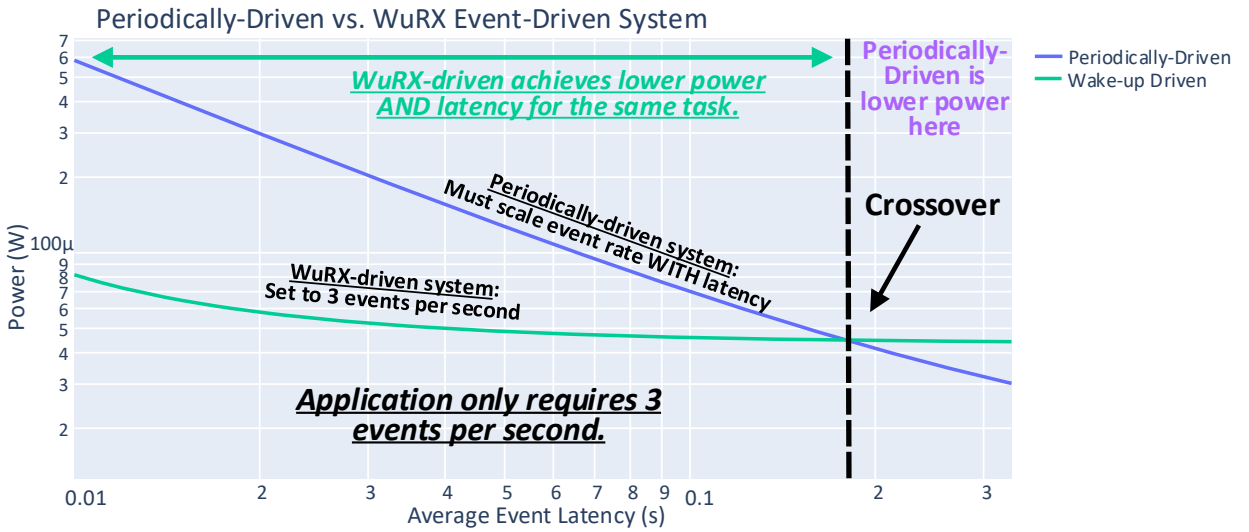


Figure 7.8: Modeling experiment showing how the WuRX-based event-driven system has a power advantage over the periodically-driven system at low latencies. While the activity level of the event-driven system is fixed (3 per second), the periodically-driven system varies in activity factor based on the inverse of the latency.

low latency at a rated event throughput.

With this in mind, several modeling experiments are presented to show part of this new trade-off. The biomedical system being modeled doesn't use an RX, but it is by default a periodically-driven TX system. This operation is compared to the WuRX, event-driven variant of the system to see when one is more preferable than the other. The first experiment considers the latency versus power trade-off for a fixed event activity level. In this scenario, the application demands three events per second. When considering the two alternatives, periodically-driven and event-driven, the periodically-driven system must scale its latency to achieve the desired activity factor. It is constantly performing the event (sense, process, transmit) in a periodic fashion. In this case, the latency would need to be approximately 0.33 seconds (based on 1/ (3 events per second)) to hit the desired activity level. However, the wake-up driven system can have a latency less than 0.33 seconds and still deliver the same activity factor. This is shown in Figure 7.8. The blue line represents the periodically-driven system and, again, its activity scales with its latency. The

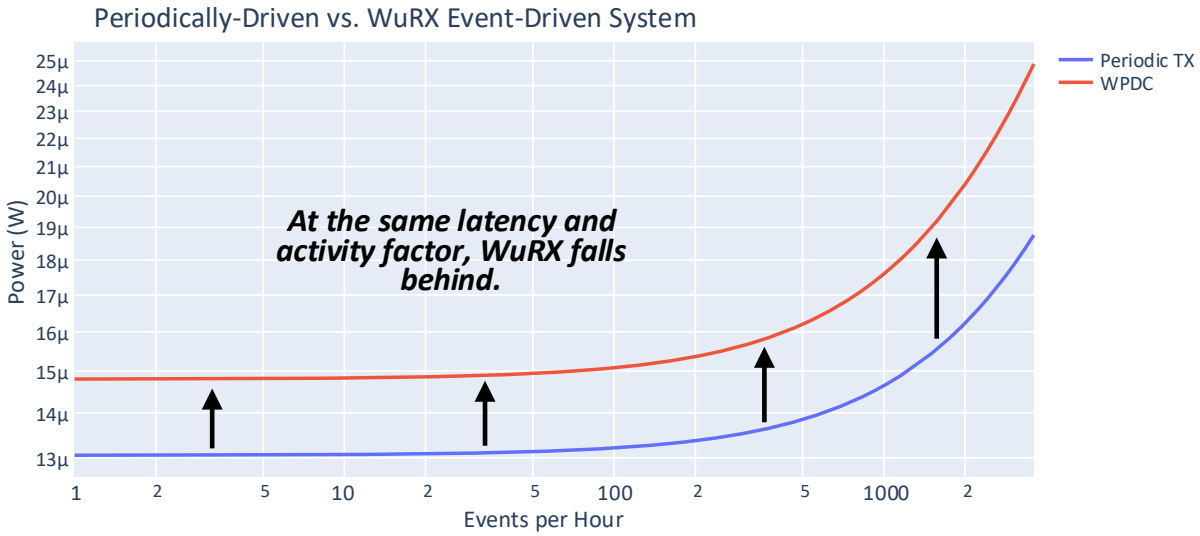


Figure 7.9: When attempting to achieve the same latency and activity factor as the periodically-driven TX, the event-driven system adds more power because of the WPDC wake-up receiver.

green line is the WuRX-driven system that performs three events per second. In this case, if the application called for an event latency below 200ms, then using an event-driven wake-up receiver would be lower power than the alternative. This improvement increases as the latency decreases. This is because fewer and fewer of the WuRX "listening" periods are triggered by a wake-up signal at low latencies when the activity level is fixed. The simple periodically-driven system however does not afford this luxury and must spend significantly more power to get lower latency. When the wake-up radio is set to have the same latency as the periodically-driven TX and the event rate is swept, the power consumption of the event-driven system will be higher. This was modeled in Figure 7.9 for a fixed latency of 100ms. The offset seen in this figure is due to the wake-up receiver's additional power consumption added on top of the rest of the system power budget.

The final experiment here looks at holding the power consumption constant for the system at the desired $30\mu W$ budget while sweeping latency and activity factor only for the WuRX, event-driven system in Figure 7.10. This is a useful plot for a system designer as it gives them an idea of how much they can get for what they spend power-wise. The two variables are swept nearly to the point where the activity level ends up being limited by the average latency. This power budget

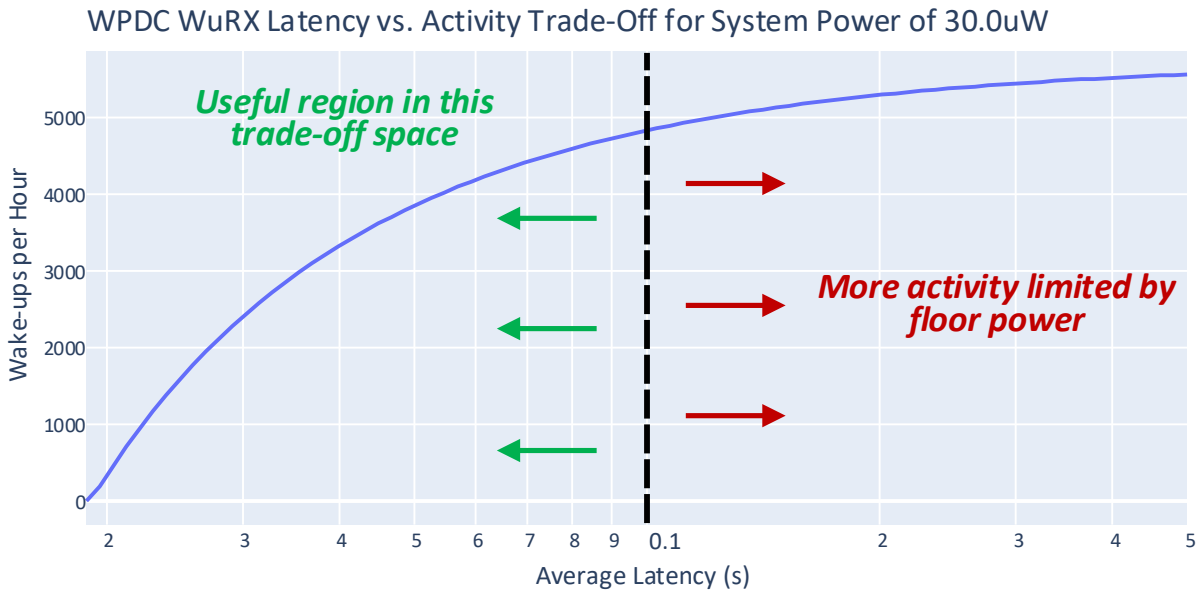


Figure 7.10: The event-driven system provides the greatest advantage at lower latencies because it scales strongly between latency and activity level.

is able to buy from nearly 0 to 5000 events per hour all with latencies below 100ms. But after this point, the trade-off becomes worse and takes hundreds of more milliseconds added latency to garner an extra 500 wake-ups per hour. This steadying out is due to the floor power becoming more prevalent in the system power budget. Basically, duty cycling out further won't help reduce the average power anymore in this example. So a designer looking at this chart would likely choose an operating point below 100ms in latency or go back to the drawing board to get better performance or even try to reduce the power if these specifications exceed the needs of the application.

7.4 Experiment 3: Dependability in WuRXs

Dependable operation for low-power systems is critical in order to ensure that they last their rated lifetimes and are able to sufficiently survive challenging energy harvesting constraints. Wake-up radios are traditionally viewed as a component that can always benefit a system, but in fact it is able to negatively impact the power profile. The primary way this occurs is through false wake-

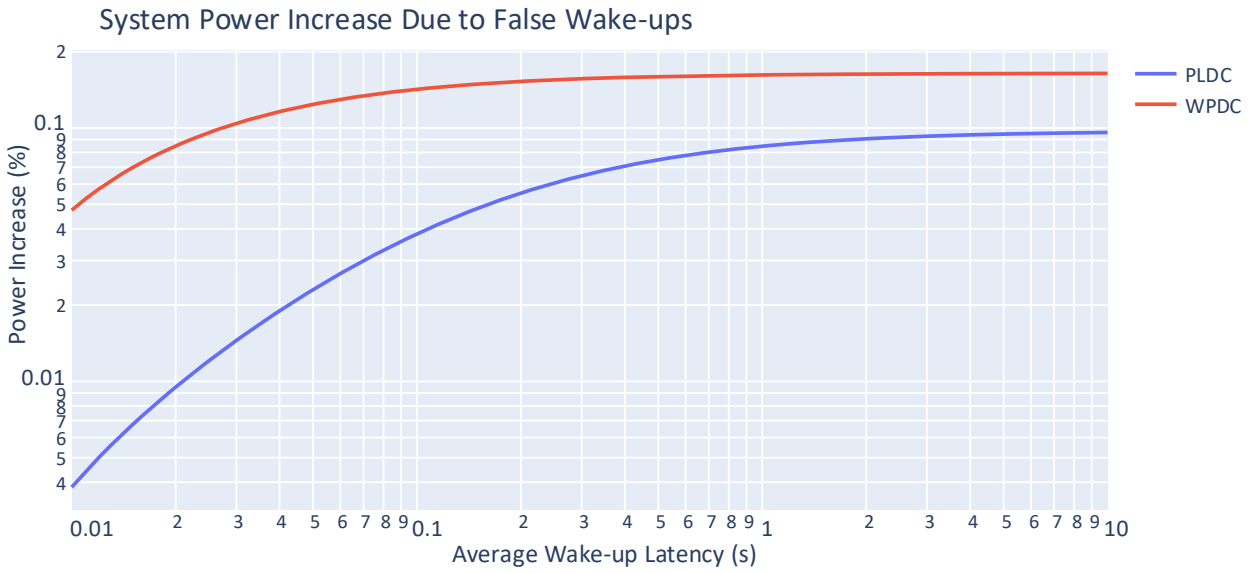


Figure 7.11: Power increase to PLDC and WPDC WuRX-enabled systems from false wake-ups in the context of the associated biomedical sensing application.

ups and carrier-senses. Bit errors are an ordinary phenomenon when thresholding data, but there are methods for correcting these errors and managing their presence to a point. However the idea of a ”wake-up” by itself implies that the entire system reacts to this one signal generated by the component. That means the WuRX wields a large amount of power, in both the figurative and literal sense, when the false wake-up rate becomes too high and negatively impacts the total system average power. False wake-ups are impossible to get rid of, but as mentioned in previous chapters there are methods for mitigating this value at the exchange of other metrics, such as sensitivity.

However, in this specific modeling experiment it turns out that false wake-ups don't negatively affect the system very much. Assuming the same WuRX settings from the measurement-verified modeling experiment in Figure 4.16, a PLDC WuRX should experience about 122 wake-ups per hour while the WPDC WuRX should only experience around 9 per hour. While the difference seems significant, it turns out that neither represent a significant impact on system power. It is also assumed that there is only one intentional event per hour. Figure 7.11 shows the percentage difference in power between an ideal PLDC and WPDC WuRX versus their counterparts that

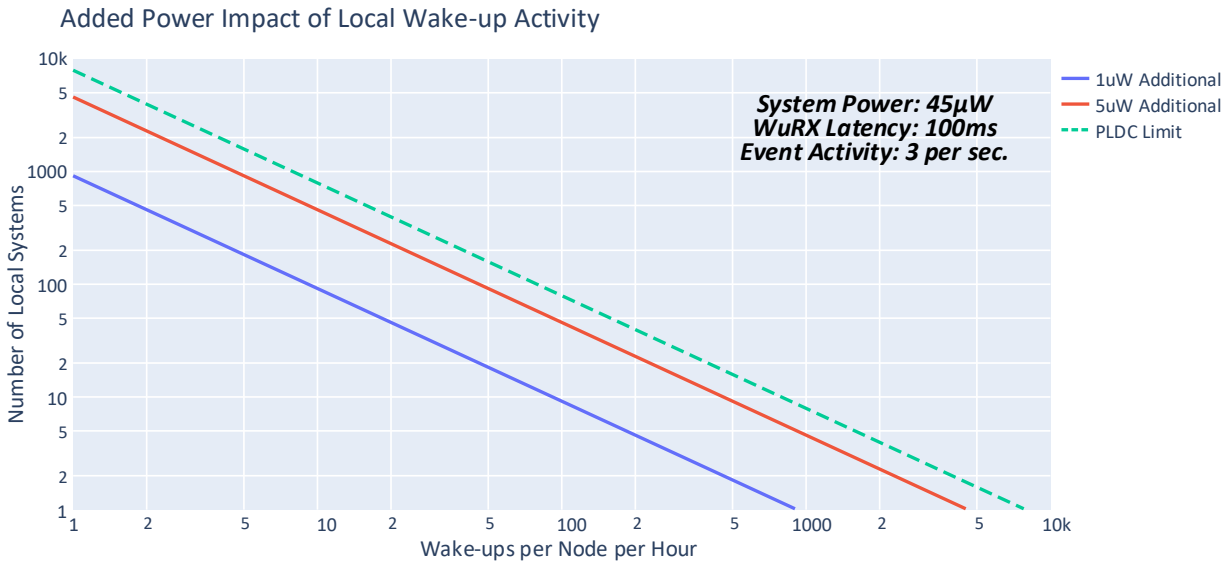


Figure 7.12: Added power consumption due to local wake-up networking activity when using WPDC-based wake-up receivers.

suffer from false wake-ups, as a function of latency. The modeled impact shows that it hovers worst case around 0.1%, which barely puts a dent in the system power budget for both WPDC and PLDC systems. The WPDC system is slightly more impacted because of the greater power savings afforded through the carrier-sense mechanism, so events end up creating a slightly larger impact.

While in this particular instance and application the false wake-up ratio doesn't impact power very much, another factor can impact it significantly. When designing low-power systems for the IoT, they are not made to be alone. Often they are part of a much larger ecosystem of similar or potentially dissimilar nodes. In the case of many systems in the local wireless vicinity, wake-ups from nodes can increase the power of other WPDC-enabled systems due to the carrier-sense mechanism. Both the level of activity per node and the actual quantity of nodes play a role in determining the added power impact for WPDC wake-up based systems. IoT system and network designers should take into account the impact on nearby listening systems when considering wake-up network scale and activity. In this particular application, the modeling experiment shown in Figure 7.12 shows the tolerable level of neighbor nodes in a local network coupled with their respective level of ac-

tivity, quantified in wake-ups per hour per node. The three plotted lines represent the boundary for the respective added power consumption to the system, $1\mu W$, $5\mu W$, and the PLDC limit where the spectrum is 100% active, due to the extra local networking activity. This trade-off with power is a linear relationship across a large region of the WuRX operating space because both activity and network scale scales the WPDC carrier-sense factor β linearly. So a system could operate at various levels of power consumption, and with reasonable certainty be operating with this same relationship. For this particular biomedical application, data is requested very often requiring many thousands of events per hour specifically 3 per second which is very high by normal low-power edge IoT standards. It was also set to have 100ms average WuRX latency, which sets the maximum activity level as well. Under these two conditions, the system consumes $45\mu W$. According to the figure, the number of local nodes as a result cannot be more than just a few in order to prevent significant changes in desired levels of system power consumption. Interestingly, if there was one more node in the local vicinity performing the same task, then both nodes would have maxed out WuRX power at the PLDC limit as the added level of activity would be over 10k events. This could present a large issue for a deployed product in the case that many similar systems all come into the local vicinity of each other. In that case, all the nodes would have a large increase in power consumption and wake-up transmissions may require something like the earlier proposed CSMA TX block to prevent wake-up message collisions unless different wake-ups were orthogonal to one another.

7.5 Conclusions

This chapter combined the proposed system modeling framework with wake-up radio techniques and measured results to demonstrate a set of modeling experiments around the impact of using wake-up receivers in a low-power IoT context. The first experiment covered the system impact of using a within-packet duty-cycled wake-up radio when compared to the traditional packet-level duty-cycling method. The second experiment focused on the question and trade-off of event latency versus communication activity for IoT applications. Lastly, this chapter presented an ex-

periment focused on how the wake-up receiver can negatively impact system power consumption through the lens of false wake-ups and carrier-sense events due to local wake-up network activity.

The following is a summary of the contributions of the work presented in this chapter:

1. Added the 2.4GHz WPDC WuRX to the previously modeled system then analyzed the system-level improvement when using WPDC compared to PLDC wake-up receivers
2. Analyzed a key trade-off of the modeled system concerning how often vs. how quickly data should be reported demonstrating another potential use-case for wake-up receivers
3. Modeled the increase in system power due to false wake-ups as well as local wake-up environments of varying size and activity factor while using the WPDC wake-up receiver

CHAPTER 8

CONCLUSION

8.1 Review of Contributions

In Chapter 3, this work discussed and analyzed the conventional bit-level and packet-level duty-cycling techniques used to reduce the power for wake-up receivers. Based on the PLDC method, the proposed within-packet duty-cycling scheme, which employs a carrier-sense mechanism, is able to further reduce power consumption compared to PLDC while retaining the same average wake-up latency. All three duty-cycling methods were modeled against average latency and duty-factor to show which techniques were best and by how much. The WPDC method was shown to have the lowest duty-factor across all latencies, assuming a low wake-up activity factor as is common in the scenarios where wake-up receivers are deployed.

In Chapter 4, a broad overview of digital baseband architectures was outlined for both bit-level and packet-level/within-packet duty-cycling wake-up receivers. Binary correlator architectures were discussed and compared and the false wake-up implications of choosing different correlator settings was mathematically provided. The carrier-sense algorithm for the within-packet duty-cycling method was outlined to describe how it uses both time and signal-strength information in order to determine if the channel is idle or active. The use of WPDC also creates more variables leading to a further complicated set of relationships when it comes to power consumption, sensitivity, and false wake-ups. The math detailing these relationships were outlined, modeled, and verified through silicon measurements in a WPDC WuRX. Further exploiting the carrier-sense mechanism, a carrier-sense multiple-access transmission block was developed and implemented to enable node-to-node wake-ups. Several techniques for automatic gain and offset control methods

were discussed primarily for OOK-based wake-up receivers. A demonstration of this control loop was also provided.

In Chapter 5, four total implemented wake-up receivers for a variety of IoT applications were proposed, discussed, and measured to demonstrate current state-of-the-art operation in power, latency, and sensitivity. The first three wake-up radios utilized bit-level duty-cycling and each emphasized a different story for kilometer-level wireless links at sub-μW powers. The first prototype focused on maximizing sensitivity while minimizing average power consumption. The second prototype was developed for a greater level of scalability in the metrics of power, latency, and sensitivity. The final prototype focused on integration and multi-channel wake-up capabilities. The last wake-up radio was a combined packet-level and within-packet duty-cycled uncertain-IF downconversion variant which was designed for near-μW and sub-second latency operation in the popular 2.4GHz ISM band compatible with the large number of existing wireless IoT devices. This wake-up radio demonstrated the power advantage from using the WPDC method in comparison to PLDC along with an event-driven PLL to recalibrate the free-running VCO useful for preventing loss of operation due to temperature drift. The ”champion” points from each design was in the end plotted against other fellow wake-up receivers demonstrating the improvements each radio made in the current literature.

In Chapter 6, a proposed system-level modeling framework was first introduced. The motivation for this framework is to help enable virtual prototyping for low-power systems to quickly relate how design and operation impacts power consumption. The Python-based framework was discussed first in terms of how it models system structure, through the use of hierarchy definitions for four different types of component-like objects. An outline was created for how the framework then models operation and design through the use of custom variable and model objects. The framework can also enable greater code simplicity through the Mode and System classes which encapsulate both a variety of relevant system structure and operational differences. Then, an example IoT system model was developed based on a measured biomedical sensing system. The discussed example took the approach to look at this system through the lens of the popular ”compute versus communicate” trade-off.

In Chapter 7, the concepts and measurements from the implemented WPDC wake-up receiver is combined with the previously implemented IoT system model in Chapter 6 in attempt to demonstrate some view of how a wake-up receiver impacts an IoT system. Three experiments were developed for this chapter. The first experiment covered the discussion of system-level power impact, compared to component-level, of the packet-level and within-packet duty-cycling methods. The second experiment sought to explore a second trade-off space for wake-up receivers in terms of event latency versus communication activity. Here, the wake-up receiver demonstrated that it is able to achieve a greater event throughput compared to a standard duty-cycled receiver operation and that at low-latencies it is most useful in this trade-off space when compared to an inefficient periodically-driven, transmitting system. The last experiment briefly covered the topic of power dependability in WPDC WuRXs. It's critical that systems can have not only dependable sources of energy availability, but also dependable power consumption profiles. The impact of false wake-ups were modeled in this section. The impact that the carrier-sense mechanism in WPDC has on power was also modeled to reveal how scaling the size of an IoT network and the level of wake-up activity can negatively impact power.

8.2 Future Work

Wake-up receivers will be an important part of future communication systems, but will require further research and development in order to be truly successful for the long term. One key area of research is improving the interference performance of the receivers. Generally, standard receiver SIR specifications are far higher than what is currently achieved in the literature with WuRXs. The off-chip RF filtering MEMS devices discussed in this dissertation help filter out some interference, but for modern communication infrastructures this will likely not be enough. It may require higher levels of power consumption combined with aggressive duty-cycling using techniques similar to the ones presented in this dissertation. In relation to this concern, wake-up radio standards should be developed further in order to allow industry to better integrate this technology into their devices. Without standards, wake-up receivers will continue to simply be an academic research

project. A wake-up standard (IEEE 802.11ba) was launched recently which will help pave the way for commercial systems using this technology. Finally, there is great opportunity to consider the networking implications of using wake-up receivers and determine the aggregate energy savings at the network-level by using WuRXs to manage handshake-related requests. It is important to know when and by how much wake-up receivers help lower system power.

There is ripe opportunity to improve system modeling capabilities, especially focused on self-powered systems. As previously mentioned, battery-powered systems will have a fixed lifetime based simply on the average power of the device. A far more complex situation arises in self-powered deployments due to energy harvesting uncertainty. Modeling efforts should be focused on developing better schemes for predicting and simulating energy harvesting environments as well as determining in a variety of ways how to calculate dependability aspects of self-powered systems. Transient time-domain modeling methods may be useful here as well. Several dependability metrics that will possibly be relevant are system availability, repairability, reliability, and performability.

8.3 Publications

8.3.1 Published

1. H. L. Bishop, P. Wang, D. Fan, J. Lach and B. H. Calhoun, ”Lighting IoT test environment (LITE) platform: evaluating light-powered, energy harvesting embedded systems,” 2018 Global Internet of Things Summit (GIoTS), Bilbao, 2018, pp. 1-6 [HB1]

2. J. Moody, A. Dissanayake, H. L. Bishop, R. Lu, N. Liu, D. Duvvuri, A. Gao, D. Truesdell, N. S. Barker, S. Gong, B. H. Calhoun, S. Bowers, ”A -106dBm 33nW Bit-Level Duty-Cycled Tuned RF Wake-up Receiver,” 2019 Symposium on VLSI Circuits, Kyoto, Japan, 2019, pp. C86-C87 [HB2]

3. O. Abdelatty, H. L. Bishop, Y. Shi, X. Chen, A. Alghaihab, B. H. Calhoun, D. Wentzloff, ”A

Low Power Bluetooth Low-Energy Transmitter with a 10.5nJ Startup-Energy Crystal Oscillator," ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC), Cracow, Poland, 2019, pp. 377-380 [HB3]

4. J. Moody, A. Dissanayake, H. L. Bishop, R. Lu, N. Liu, D. Duvvuri, A. Gao, D. Truesdell, N. S. Barker, S. Gong, B. H. Calhoun, S. Bowers, "A Highly Reconfigurable Bit-Level Duty-Cycled TRF Receiver Achieving 106-dBm Sensitivity and 33-nW Average Power Consumption," in IEEE Solid-State Circuits Letters, vol. 2, no. 12, pp. 309-312, Dec. 2019 [HB4]

5. A. Dissanayake, J. Moody, H. L. Bishop, D. Truesdell, H. Muhlbauer, R. Lu, A. Gao, S. Gong, B. H. Calhoun, S. Bowers, "A- 108dBm Sensitivity, -28dB SIR, 130nW to 41μW, Digitally Reconfigurable Bit-Level Duty-Cycled Wakeup and Data Receiver," 2020 IEEE Custom Integrated Circuits Conference (CICC), Boston, MA, USA, 2020, pp. 1-4 [HB5]

6. H. L. Bishop, P. Wang, B. H. Calhoun, "Application-Driven Model of a PPG Sensing Modality for the Informed Design of Self-Powered, Wearable Healthcare Systems," 2020 IEEE International Symposium on Circuits and Systems (ISCAS) [HB6]

7. P. Wang, R. Agarwala, H. L. Bishop, A. Dissanayake, B. H. Calhoun, " A 785nW Multimodal (V/I/R) Sensor Interface IC for Ozone Pollutant Sensing and Correlated Cardiovascular Disease Monitoring," 2020 Symposium on VLSI Circuits [HB7]

8. A. Dissanayake, H. L. Bishop, J. Moody, H. Muhlbauer, B. H. Calhoun, S. Bowers, "A Multichannel, MEMS-less -99dBm 260nA Bit-level Duty Cycled Wakeup Receiver," 2020 Symposium on VLSI Circuits [HB8]

9. H. L. Bishop*, A. Dissanayake*, S. Bowers, B. H. Calhoun, "An Integrated 2.4GHz -91.5dBm-Sensitivity Within-Packet Duty-Cycled Wake-Up Receiver Achieving 2μW at 100ms Latency," 2021 International Solid-State Circuits Conference (ISSCC) (* Equally-Credited Authors) [HB9]

10. R. Agarwala, P. Wang, H. L. Bishop, A. Dissanayake and B. H. Calhoun, "A 0.6V 785-nW Multimodal Sensor Interface IC for Ozone Pollutant Sensing and Correlated Cardiovascular

Disease Monitoring," in IEEE Journal of Solid-State Circuits, vol. 56, no. 4, pp. 1058-1070, April 2021 [HB10]

8.3.2 Accepted

1. A. Dissanayake*, H. L. Bishop*, S. Bowers, B. H. Calhoun, " 2.4GHz -91.5dBm Sensitivity Within-Packet Duty-Cycled Scalable Wakeup Receiver," 2021 Journal of Solid-State Circuits (JSSC) (* Equally-Credited Authors) [HB11]

8.3.3 Software Repositories

1. Low-Power System Modeling Framework (RLP-VLSI Github Account) [HB12]

BIBLIOGRAPHY

[1] Sravanthi Chalasani and J. M. Conrad, "A survey of energy harvesting sources for embedded systems," in *IEEE SoutheastCon 2008*, 2008, pp. 442–447.

[2] M. Prauzek, J. Konecny, M. Borova, K. Janosova, J. Hlavica, and P. Musilek, "Energy harvesting sources, storage devices and system topologies for environmental wireless sensor networks: A review," *Sensors*, vol. 18, no. 8, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/8/2446

[3] H. Milosiu, F. Oehler, M. Eppel, D. Fruehsorger, and T. Thoenes, "A 7-µw 2.4-ghz wake-up receiver with -80 dbm sensitivity and high co-channel interferer tolerance," in *2015 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)*, 2015, pp. 35–37.

[4] C. Salazar, A. Cathelin, A. Kaiser, and J. Rabaey, "A 2.4 ghz interferer-resilient wake-up receiver using a dual-if multi-stage n-path architecture," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 9, pp. 2091–2105, 2016.

[5] N. M. Pletcher, S. Gambini, and J. Rabaey, "A 52 μ w wake-up receiver with $-$ 72 dbm sensitivity using an uncertain-if architecture," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 269–280, 2009.

[6] H. Seok, O. Jung, A. Dissanayake, and S. Lee, "A 2.4ghz, 102dbm-sensitivity, 25kb/s, 0.466mw interference resistant bfsk multi-channel sliding-if ulp receiver," in *2017 Symposium on VLSI Circuits*, 2017, pp. C70–C71.

[7] M. R. Abdelhamid, A. Paidimarri, and A. P. Chandrakasan, "A 80dbm ble-compliant, fsk wake-up receiver with system and within-bit dutycycling for scalable power and latency," in *2018 IEEE Custom Integrated Circuits Conference (CICC)*, 2018, pp. 1–4.

[8] P. P. Wang and P. P. Mercier, “28.2 a 220w -85dbm sensitivity ble-compliant wake-up receiver achieving -60db sir via single-die multi- channel fbar-based filtering and a 4-dimentional wake-up signature,” in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 440–442.

[9] R. Dorrance, R. Liu, K. T. Asma Beevi, D. Dasalukunte, M. A. S. Lopez, V. Kristem, S. Azizi, and B. Carlton, “An ultra-low power, fully integrated wake-up receiver and digital baseband with all-digital impairment correction and -92.4dbm sensitivity for 802.11ba,” in *2019 Symposium on VLSI Circuits*, 2019, pp. C80–C81.

[10] X. Huang, P. Harpe, G. Dolmans, H. de Groot, and J. R. Long, “A 780–950 mhz, 64–146 μw power-scalable synchronized-switching ook receiver for wireless event-driven applications,” *IEEE Journal of Solid-State Circuits*, vol. 49, no. 5, pp. 1135–1147, 2014.

[11] N. Pletcher, S. Gambini, and J. Rabaey, “A 65 w, 1.9 ghz rf to digital baseband wakeup receiver for wireless sensor nodes,” in *2007 IEEE Custom Integrated Circuits Conference*, 2007, pp. 539–542.

[12] X. Huang, A. Ba, P. Harpe, G. Dolmans, H. De Groot, and J. Long, “A 915mhz 120w-rx/900w-tx envelope-detection transceiver with 20db in-band interference tolerance,” in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 454–456.

[13] V. Mangal and P. R. Kinget, “28.1 a 0.42nw 434mhz -79.1dbm wake-up receiver with a time-domain integrator,” in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 438–440.

[14] J. Moody, P. Bassirian, A. Roy, N. Liu, S. Pancrazio, N. S. Barker, B. H. Calhoun, and S. M. Bowers, “A 76dbm 7.4nw wakeup radio with automatic offset compensation,” in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018, pp. 452–454.

[15] N. E. Roberts, K. Craig, A. Shrivastava, S. N. Wooters, Y. Shakhsheer, B. H. Calhoun, and D. D. Wentzloff, "26.8 a 236nw 56.5dbm-sensitivity bluetooth low-energy wakeup receiver with energy harvesting in 65nm cmos," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 450–451.

[16] D. Ye, R. van der Zee, and B. Nauta, "26.2 an ultra-low-power receiver using transmitted-reference and shifted limiters for in-band interference resilience," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 438–439.

[17] P. Bassirian, D. Duvvuri, D. S. Truesdell, N. Liu, B. H. Calhoun, and S. M. Bowers, "30.1 a temperature-robust 27.6nw 65dbm wakeup receiver at 9.6ghz x-band," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 460–462.

[18] H. Jiang, P. H. P. Wang, L. Gao, C. Pochet, G. M. Rebeiz, D. A. Hall, and P. P. Mercier, "A 22.3-nw, 4.55 cm2 temperature-robust wake-up receiver achieving a sensitivity of 69.5 dbm at 9 ghz," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1530–1541, 2020.

[19] D. D. Wentzloff, "Ultra low power radio survey." [Online]. Available: http://www.eecs.umich.edu/wics/low_power_radio_survey.html

[20] "K32w061 / k32w041 ieee 802.15.4 and bluetooth le 5.0 wireless microcontroller." [Online]. Available: https://www.nxp.com/docs/en/data-sheet/K32W061.pdf

[21] "nrf52832 product specification v1.4." [Online]. Available: https://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.4.pdf

[22] Ambiq-Micro, "Apollo3 blue mcu datasheet." [Online]. Available: https://ambiq.com/wp-content/uploads/2020/10/Apollo3-Blue-MCU-Datasheet.pdf

[23] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[24] "Industrial revolution 4.0 - the future of iiot." [Online]. Available: https://www.juniperresearch.com/white-papers/industrial-revolution-4-the-future-of-iiot

[25] G. Inc, "Forecast analysis: Healthcare providers iot endpoint electronics and communications revenue, worldwide." [Online]. Available: https://www.gartner.com/doc/3990045

[26] S. Fahrni, C. Jansen, M. John, T. Kasah, B. Koerber, and N. Mohr, "Coronavirus: Industrial iot in challenging times," Apr 2020. [Online]. Available: https://www.mckinsey.com/industries/advanced-electronics/our-insights/coronavirus-industrial-iot-in-challenging-times

[27] J. Stokes, "Classic.ars: Understanding moore's law," Sep 2008. [Online]. Available: https://arstechnica.com/gadgets/2008/09/moore/

[28] E. Mollick, "Establishing moore's law," *IEEE Annals of the History of Computing*, vol. 28, no. 3, pp. 62–75, 2006.

[29] S. Lukasik, "Why the arpanet was built," *IEEE Annals of the History of Computing*, vol. 33, no. 3, pp. 4–21, 2011.

[30] DARPA, "Arpanet." [Online]. Available: https://www.darpa.mil/attachments/ARPANET_final.pdf

[31] "Internet host count history." [Online]. Available: https://web.archive.org/web/20120518101749/http://www.isc.org/solutions/survey/history

[32] K. L. Lueth, "Iot market – forecasts at a glance," May 2020. [Online]. Available: https://iot-analytics.com/iot-market-forecasts-overview/

[33] Cisco, "Cisco annual internet report (2018–2023)." [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf

[34] ——, "The internet of things how the next evolution of the internet is changing everything," Apr 2011. [Online]. Available: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

[35] Energizer, "Energizer cr2025." [Online]. Available: https://data.energizer.com/PDFs/cr2025.pdf

[36] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 6, no. 4, p. 32–es, Sep. 2007. [Online]. Available: https://doi.org/10.1145/1274858.1274870

[37] A. F. Yeknami, X. Wang, I. Jeerapan, S. Imani, A. Nikoofard, J. Wang, and P. P. Mercier, "A 0.3-v cmos biofuel-cell-powered wireless glucose/lactate biosensing system," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3126–3139, 2018.

[38] I. Lee, E. Moon, Y. Kim, J. Phillips, and D. Blaauw, "A 10mm3 light-dose sensing iot2 system with 35-to-339nw 10-to-300klx light-dose-to-digital converter," in *2019 Symposium on VLSI Circuits*, 2019, pp. C180–C181.

[39] C. J. Lukas, F. B. Yahya, J. Breiholz, A. Roy, X. Chen, H. N. Patel, N. Liu, A. Kosari, S. Li, D. Akella Kamakshi, O. Ayorinde, D. D. Wentzloff, and B. H. Calhoun, "A 1.02 w battery-less, continuous sensing and post-processing sip for wearable applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 271–281, 2019.

[40] J. K. Brown, D. Abdallah, J. Boley, N. Collins, K. Craig, G. Glennon, K. Huang, C. J. Lukas, W. Moore, R. K. Sawyer, Y. Shakhsheer, F. B. Yahya, A. Wang, N. E. Roberts, D. D. Wentzloff, and B. H. Calhoun, "27.1 a 65nm energy-harvesting ulp soc with 256kb cortex-m0 enabling an 89.1µw continuous machine health monitoring wireless self-powered system," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 420–422.

[41] K. Philips, "Ultra low power short range radios: Covering the last mile of the iot," in *ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC)*, 2014, pp. 51–58.

[42] P. P. MERCIER, *ULTRA-LOW-POWER SHORT-RANGE RADIOS*. SPRINGER INTERNATIONAL PU, 2016.

[43] T. Zheng, S. Radhakrishnan, and V. Sarangan, "Pmac: an adaptive energy-efficient mac protocol for wireless sensor networks," in *19th IEEE International Parallel and Distributed Processing Symposium*, 2005, pp. 8 pp.–.

[44] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, ser. SenSys '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 95–107. [Online]. Available: https://doi.org/10.1145/1031495.1031508

[45] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-mac: A short preamble mac protocol for duty-cycled wireless sensor networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, ser. SenSys '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 307–320. [Online]. Available: https://doi.org/10.1145/1182807.1182838

[46] B. Sklar, *Digital communications: fundamentals and applications*. Prentice Hall, 2001.

[47] B. Razavi, *RF Microelectronics*. Dorling Kindersley, 2014.

[48] J. Liu, C. Chen, Y. Ma, and Y. Xu, "Energy analysis of device discovery for bluetooth low energy," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, 2013, pp. 1–5.

[49] R. C. Carrano, D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, "Survey and taxonomy of duty cycling mechanisms in wireless sensor networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 181–194, 2014.

[50] J. Moody, P. Bassirian, A. Roy, N. Liu, N. S. Barker, B. H. Calhoun, and S. M. Bowers, "Interference robust detector-first near-zero power wake-up receiver," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 8, pp. 2149–2162, 2019.

[51] P. H. P. Wang and P. P. Mercier, "A 4.4w 92/90.3dbm sensitivity dual-mode ble/wi-fi wake-up receiver," in *2020 IEEE Symposium on VLSI Circuits*, 2020, pp. 1–2.

[52] A. S. Tanenbaum and D. Wetherall, *Computer networks*. Pearson India Education Services Pvt, Limited, 2014.

[53] V. Mangal and P. R. Kinget, "Sub-nw wake-up receivers with gate-biased self-mixers and time-encoded signal processing," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 12, pp. 3513–3524, 2019.

[54] J. Moody, A. Dissanayake, H. Bishop, R. Lu, N. Liu, D. Duvvuri, A. Gao, D. Truesdell, N. S. Barker, S. Gong, B. H. Calhoun, and S. M. Bowers, "A -106dbm 33nw bit-level duty-cycled tuned rf wake-up receiver," in *2019 Symposium on VLSI Circuits*, 2019, pp. C86–C87.

[55] ——, "A highly reconfigurable bit-level duty-cycled trf receiver achieving 106-dbm sensitivity and 33-nw average power consumption," *IEEE Solid-State Circuits Letters*, vol. 2, no. 12, pp. 309–312, 2019.

[56] A. Dissanayake, J. Moody, H. L. Bishop, D. Truesdell, H. Muhlbauer, R. Lu, A. Gao, S. Gong, B. H. Calhoun, and S. M. Bowers, "A- 108dbm sensitivity, -28db sir, 130nw to 41µw, digitally reconfigurable bit-level duty-cycled wakeup and data receiver," in *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1–4.

[57] A. Dissanayake, H. L. Bishop, J. Moody, H. Muhlbauer, B. H. Calhoun, and S. M. Bowers, "A multichannel, mems-less 99dbm 260nw bit-level duty cycled wakeup receiver," in *2020 IEEE Symposium on VLSI Circuits*, 2020, pp. 1–2.

[58] R. Liu, A. Beevi K.T., R. Dorrance, D. Dasalukunte, V. Kristem, M. A. Santana Lopez, A. W. Min, S. Azizi, M. Park, and B. R. Carlton, "An 802.11ba-based wake-up radio receiver with wi-fi transceiver integration," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 5, pp. 1151–1164, 2020.

[59] A. Boulis, "Castalia: Revealing pitfalls in designing distributed algorithms in wsn," in *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, ser. SenSys '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 407–408. [Online]. Available: https://doi.org/10.1145/1322263.1322318

[60] D. Benedetti, C. Petrioli, and D. Spenza, "Greencastalia: An energy-harvesting-enabled framework for the castalia simulator," in *Proceedings of the 1st International Workshop on Energy Neutral Sensing Systems*, ser. ENSSys '13. New York, NY, USA: Association for Computing Machinery, 2013. [Online]. Available: https://doi.org/10.1145/2534208.2534215

[61] J. Eriksson, F. Österlind, N. Finne, N. Tsiftes, A. Dunkels, T. Voigt, R. Sauter, and P. J. Marrón, "Cooja/mspsim: Interoperability testing for wireless sensor networks," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, ser. Simutools '09. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009. [Online]. Available: https://doi.org/10.4108/ICST.SIMUTOOLS2009.5637

[62] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt, "Cross-level sensor network simulation with cooja," in *Proceedings. 2006 31st IEEE Conference on Local Computer Networks*, 2006, pp. 641–648.

[63] P. Levis, N. Lee, M. Welsh, and D. Culler, "Tossim: Accurate and scalable simulation of entire tinyos applications," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*, ser. SenSys '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 126–137. [Online]. Available: https://doi.org/10.1145/958491.958506

[64] O. Abdelatty, H. Bishop, Y. Shi, X. Chen, A. Alghaihab, B. H. Calhoun, and D. D. Wentzloff, "A low power bluetooth low-energy transmitter with a 10.5nj startup-energy crystal oscillator," in *ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC)*, 2019, pp. 377–380.

[65] C. J. Lukas, F. B. Yahya, J. Breiholz, A. Roy, X. Chen, H. N. Patel, N. Liu, A. Kosari, S. Li, D. Akella Kamakshi, O. Ayorinde, D. D. Wentzloff, and B. H. Calhoun, "A 1.02 w battery-less, continuous sensing and post-processing sip for wearable applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 271–281, 2019.

[66] R. Piyare, A. L. Murphy, C. Kiraly, P. Tosato, and D. Brunelli, "Ultra low power wake-up radios: A hardware and networking survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2117–2157, 2017.