Ultra Low Power Circuits for Internet of Things and Deep Learning Accelerator Design with In-Memory Computing

by

Myungjoon Choi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2018

Doctoral Committee:

Professor Dennis M. Sylvester, Chair
Professor David Blaauw
Assistant Professor Cynthia Chestek
Assistant Professor Hun-Seok Kim

Myungjoon Choi

myungjun@umich.edu

ORCID iD: 0000-0003-0995-6642

DEDICATION

To my Lord,

my wife Dayoon, my to be born daughter,

and my parents and family

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Abstract: Collecting data from environment and converting gathered data into information is the key idea of Internet of Things (IoT). Miniaturized sensing devices enable the idea for many applications including health monitoring, industrial sensing, and so on. Sensing devices typically have small form factor and thus, low battery capacity, but at the same time, require long life time for continuous monitoring and least frequent battery replacement. This thesis introduces three analog circuit design techniques featuring ultra-low power consumption for such requirements: (1) An ultra-low power resistor-less current reference circuit, (2) A 110nW resistive frequency locked on-chip oscillator as a timing reference, (3) A resonant current-mode wireless power receiver and battery charger for implantable systems.

Raw data can be efficiently transformed into useful information using deep learning. However deep learning requires tremendous amount of computation by its nature, and thus, an energy efficient deep learning hardware is highly demanded to fully utilize this algorithm in various applications. This thesis also presents a pulse-width based computation concept which utilizes in-memory computing of SRAM.

CHAPTER 1

Introduction

Collecting data from environment and converting gathered data into information is the key idea of Internet of Things (IoT). Miniaturized sensing devices enable the idea for many applications including health monitoring, industrial sensing, surveillance, and so on. Sensing devices typically have small form factor for better bio-compatibility in body-implantable applications or for better distribution in industrial applications. Integrated circuits normally have size of less than a few tens of square millimeter, however, reducing battery sizes to a similar order limits battery capacity severely.

Figure 1.1 [1] shows power budget as a function of desired lifetime and battery size. With a $1mm^2$ Li-ion thin-film battery, average power consumption as low as 100's of nano-watt is required for 1 day continuous operation without charging or battery replacement. A down-hole monitoring sensor in Figure 1.2 is one system with such a low power operation constraint [2]. Ultra-low power circuit techniques come into play for IoT applications in this context.

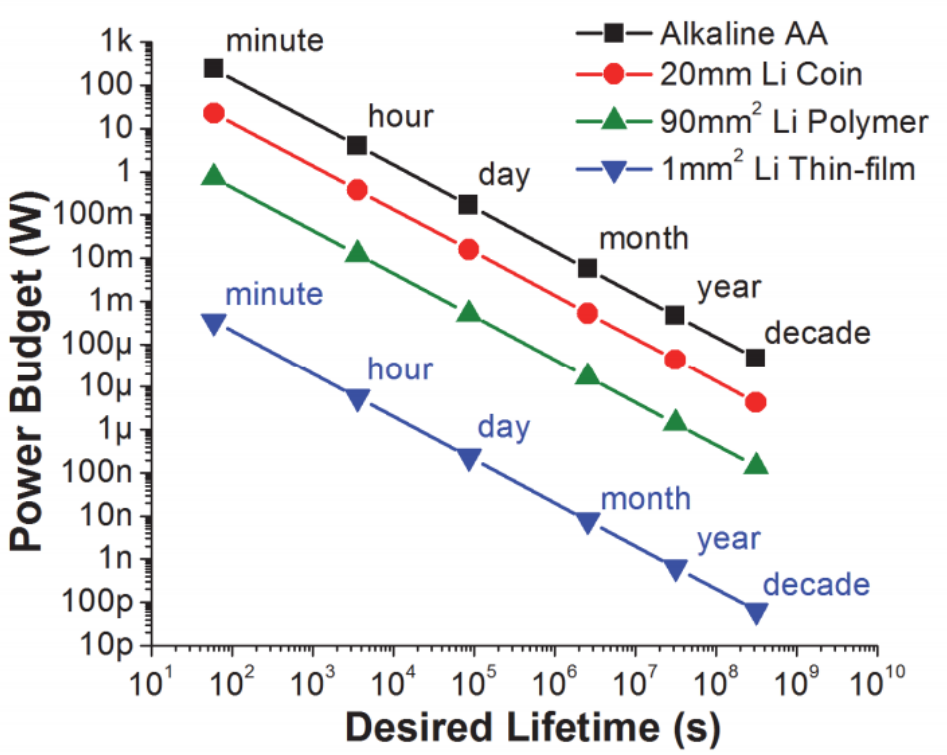
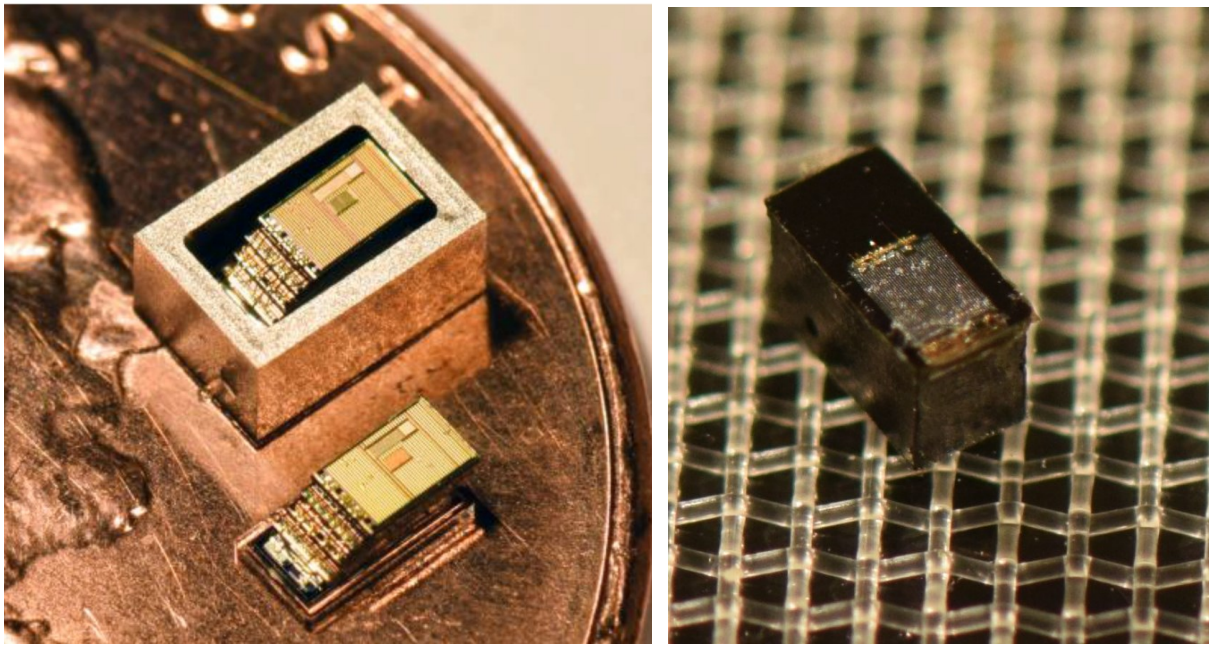


Figure 1.1 Average power budget as a function of desired lifetime and battery size [1]



(a) (b)

Figure 1.2 Down-hole monitoring sensors in (a) metal packaging and (b) epoxy packaging

This thesis introduces three analog circuit design techniques featuring ultra-low power consumption. First, an ultra-low power resistor-less current reference circuit is introduced [3]. Voltage and current reference circuits are key components of analog circuits used for biasing, frequency generation, and high precision building block such as slope-based analog to digital converters. A sub-nano-watt voltage reference circuits has been proposed [4], but this work is the first pico-watt level current reference. This also features resistor-less architecture, which is important for area-efficient implementation especially for sub-nano-ampere current generation. Secondly, a 110nW resistive frequency locked on-chip oscillator as a timing reference [5], [6] is discussed. On-chip timing reference is highly needed for small form factor systems. This work introduces methods to make a low power accurate timing reference circuit. A resonant current-mode wireless power receiver and battery charger for implantable systems [7], [8] is also presented.

This power receiver focuses on lowering the minimum harvestable input power so that an implantable system can start utilizing power at low incident power.

Raw data can be effectively transformed into useful information using deep learning. However deep learning requires tremendous amount of computation by its nature, and thus, an energy efficient deep learning hardware is highly demanded to fully utilize this algorithm in various applications. This thesis also presents a pulse-width based computation concept which uses in-memory computing of SRAM.

The dissertation is composed of 3 chapters introducing ultra-low power analog circuit techniques mentioned above and 1 chapter proposing a pulse-width based deep learning accelerator.

Chapter 2 proposes a MOSFET-only, 20pA, 780ppm/ºC current reference that consumes 23pW. The ultra-low power circuit exploits subthreshold-biased MOSFETs and a complementary-to-absolute temperature (CTAT) gate voltage to compensate for temperature dependency. The design shows low supply voltage sensitivity of 0.58%/V and a load sensitivity of 0.25%/V.

In Chapter 3, a sub-μW on-chip oscillator for fully integrated system-on-chip designs is introduced. This work introduces a resistive frequency locked loop topology for accurate clock generation. In this topology, a switched-capacitor circuit is controlled by an internal voltage-controlled oscillator (VCO), and the equivalent resistance of this switched-capacitor is matched to a temperature-compensated on-chip resistor using an ultra-low power amplifier. This design yields a temperature-compensated frequency from the internal VCO. The approach eliminates the traditional comparator from the oscillation loop; this comparator typically consumes a significant portion of the total oscillator power and limits temperature stability in conventional RC relaxation oscillators due to its temperature-dependent delay. A test chip is fabricated in 0.18 μm CMOS that

exhibits a temperature coefficient of 34.3 ppm/ºC with long-term stability of less than 7 ppm (12 second integration time) while consuming 110 nW at 70.4 kHz. A radio transmitter circuit that uses the proposed oscillator as a baseband timing source is also presented to demonstrate a system-on-chip design using this oscillator.

In Chapter 4, wireless power transfer for implantable systems is presented. Wireless power receiver circuit in implantable systems must harvest very low power levels due to low incident power on human tissues and a small receiver coil size. This chapter proposes resonant current-mode charging to reduce minimum harvestable input power and increase power efficiency at low input power levels. Avoiding rectification and voltage regulation from conventional voltage-mode methods, this work resonates an LC tank for multiple cycles to build up energy, then directly charges a battery with inductor current. A prototype is fabricated in 0.18μm CMOS technology. Minimum harvestable input power is 600nW and maximum power efficiency is 67.6% at 4.2μW input power. Power transmission through bovine tissue is measured to have negligible efficiency loss, making this technique amenable to implantable applications.

In Chapter 5, a pulse-width based deep learning accelerator with in-memory computing is introduced. This chapter proposes an all layers all weights on-chip deep learning accelerator based on pulse width modulation. Key challenges of a comprehensive deep learning system include layer to layer connections, energy and latency constraints from data movement, implementation of non-linear functions such as sigmoid and ReLU. The proposed system modulates input images into pulse width and computes dot-product as charge integration of SRAM's read buffer currents over pulse widths. With this in-memory computing approach, energy and delay for weight retrieval can be removed. Dot-product outputs are stored as analog voltages, and converted into pulse width again by comparison to non-linear reference signals. Outputs modulated in pulse width enable

layer to layer connection simple and fast, as standard digital buffers can easily drive pulse signals without information loss. Intermediate values during inferences all remain in analog domain either as voltages or pulse widths, removing analog to digital conversion overhead and allowing an energy efficient high-throughput accelerator.

Finally Chapter 6 concludes all presented works and discusses future directions.

CHAPTER 2

An Ultra-Low Power Resistor-less Current Reference Using Subthreshold MOSFETs

2.1 Introduction

Sub-nano ampere current references are of increased interest recently, as micro-scale sensor nodes and bio-implantable systems with limited power budgets gain popularity [9]. These systems use ultra-low-power mixed signal circuits such as oscillators and analog amplifiers, which require current references with low power overhead as key building blocks.

To motivate the need for an ultra-low power current reference with low temperature dependence, consider a recently reported 65nW CMOS temperature sensor [10]. This sensor uses multiple subthreshold-mode operational amplifiers, each of which consumes 100s of pA. The amplifiers make up 6% of total analog front-end power consumption at room temperature. However, due to the lack of a temperature-compensated current reference, amplifier power increases exponentially with temperature such that they consume 52% of total analog front-end power at 100ºC. Adopting the current reference circuit proposed in this paper would limit the amplifier and current reference overhead power to only 6% at 100ºC, reducing total analog front-end power from 56.2nW to 14.9nW at 100ºC.

Many conventional current reference circuits are variations of the β-multiplier current reference (Figure 2.1 (a)). However, this type of reference is unsuitable for sub-nA current generation as it requires an extremely large resistor of 1GΩ or more. Further, a start-up circuit is

needed to prevent the circuit from becoming trapped in an undesired operating point, adding area overhead. The authors of [11] replace the resistor with a MOSFET to create a subthreshold version of the β-multiplier, however the circuit remains in the nW range (88nW@1.3V).

Other proposed current references employ a reference voltage and a resistor (Figure 2.1 (b)) [12], [13], achieving a temperature coefficient (TC) as low as 24.9ppm/ºC [12]. However, those circuits consume μW's and their use of resistors complicate sub-nA current generation. Also, polysilicon resistors vary by up to ±25% [14]; this variability is independent of transistor process variation, potentially worsening process sensitivity.

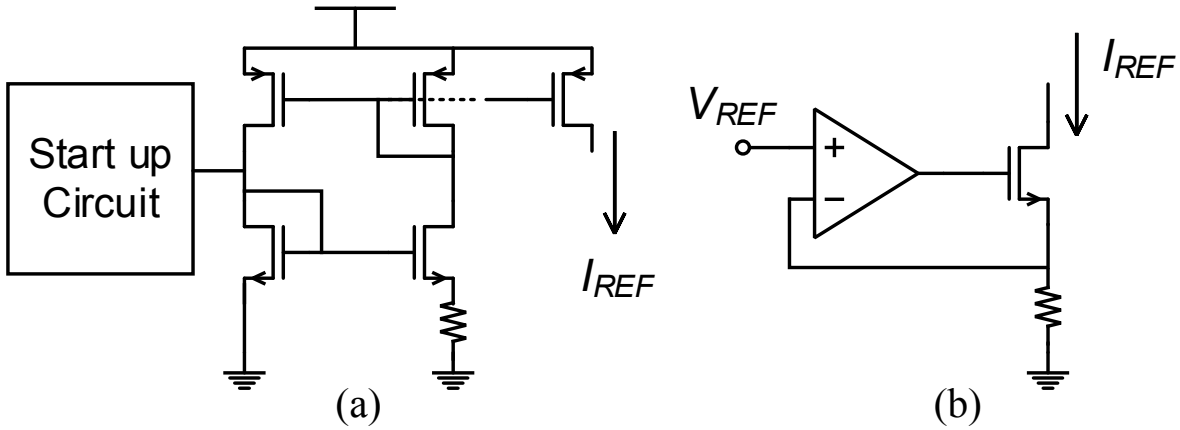


Figure 2.1 Conventional current references based on: (a) β-multiplier; (b) voltage reference divided by resistance.

This paper proposes a new topology to generate a sub-nA (20pA) level reference current with very low power overhead. It shows 780ppm/ºC TC and consumes 23pW, which is >50× smaller than the lowest power consumption reported previously [15]. This work also describes techniques to improve supply voltage and load voltage regulation.

2.2 Proposed Circuit

The basic idea of this work is to linearly reduce the gate voltage of a subthreshold-biased MOSFET as temperature increases, providing compensation (first-order) for the exponential dependence of drain current on temperature. The design challenge is to achieve this with pW-level power overhead. The proposed design has three components (Figure 2.2): an ultra-low-power line regulator, a CTAT gate voltage generator, and a cascoded subthreshold MOSFET output stage.

In addition, an optional current level selector (CLS) can be incorporated to provide a tunable range of current magnitudes.

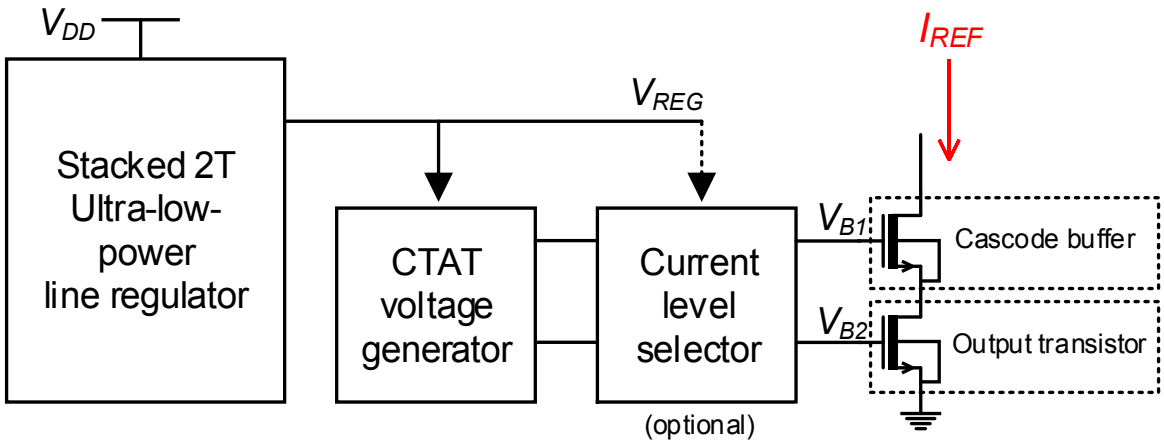


Figure 2.2 Block diagram of proposed current reference.

Figure 2.2 shows the implementation of the overall circuit, including the detailed structure of each of the three blocks in Figure 2.3.

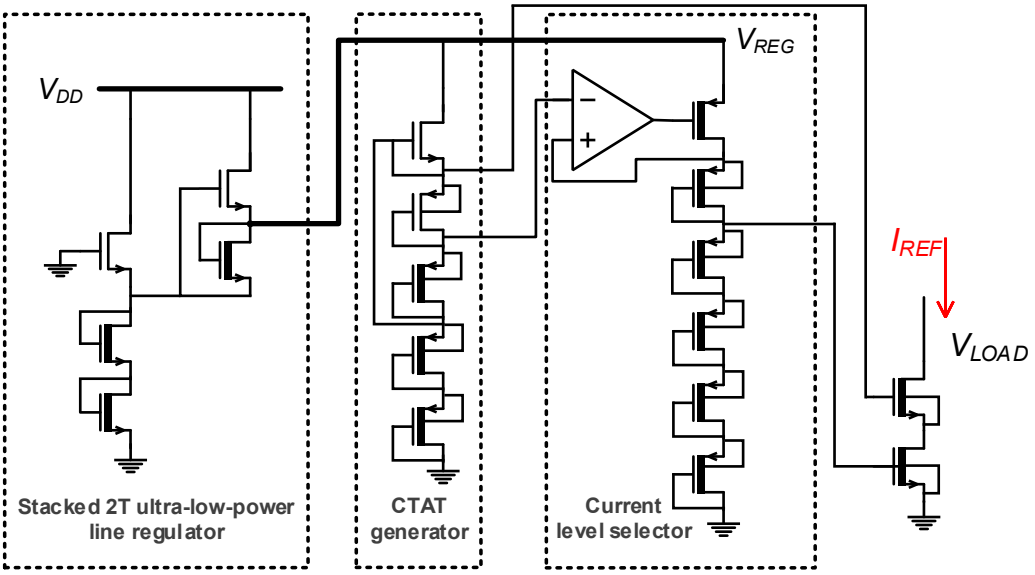


Figure 2.3 Circuit diagram of proposed current reference circuit.

A. Temperature Compensation

The output current of the proposed current reference can be derived as (1), below. Since the subthreshold current exponentially depends on both absolute temperature and gate to source voltage, by linearly decreasing the MOSFET gate voltage as temperature increases (Equation 2), transistor drain current remains nearly constant. Equation (3) shows that the remaining temperature dependent terms are *T^(1/2)* and *exp(α2/T)*, which approximately cancel out each other with respect to T. To simplify, temperature independent terms are packed into α1 and α2 (4). Differentiating (1) with respect to T gives (5). Setting it to 0 provides the temperature where the output current is temperature-independent as derived in (6). If we want to operate this circuit to be temperature-independent at room temperature (T_r), the gate voltage can be designed so that Vgs0 of (7) is met. The following B section describes how to generate this gate voltage.

$$I_{REF} = \mu(T_r)\left(\frac{T}{T_r}\right)^{-1.5} C_{ox}\frac{W}{L}\left(\frac{kT}{q}\right)^2 e^{\left(\frac{q(V_{gs}-V_{th0}+\kappa_{V_{th}}T)}{mkT}\right)} \quad (1)$$

$$V_{gs} = V_{gs0} - \kappa_{V_{gs}}T \quad (2)$$

$$I_{REF} = \alpha_1 T^{\frac{1}{2}} e^{\frac{\alpha_2}{T}} \tag{3}$$

$$\alpha_1 = \mu(T_r) C_{ox} \frac{W}{T_r^{-1.5} L} \frac{k^2}{q^2} e^{\frac{q(\kappa V_{th} - \kappa V_{gs})}{mk}}, \alpha_2 = \frac{q(V_{gs0} - V_{th0})}{mk} \tag{4}$$

$$\frac{\partial I_{REF}}{\partial T} = \alpha_1 e^{\frac{\alpha_2}{T}} T^{-\frac{1}{2}} \left(\frac{1}{2} - \alpha_2 T^{-1}\right) = 0 \tag{5}$$

$$T\left(\frac{\partial I_{REF}}{\partial T} = 0\right) = 2\alpha_2 = \frac{2q(V_{gs0} - V_{th0})}{mk} \tag{6}$$

$$V_{gs0} = mkT_r/2q + V_{th0} \tag{7}$$

where μ is mobility, Cox is oxide capacitance, and W and L are MOSFET width and length. Vgs0 is Vgs at 0K and Vth0 is threshold voltage at 0K. κVth and κVgs are temperature coefficients.

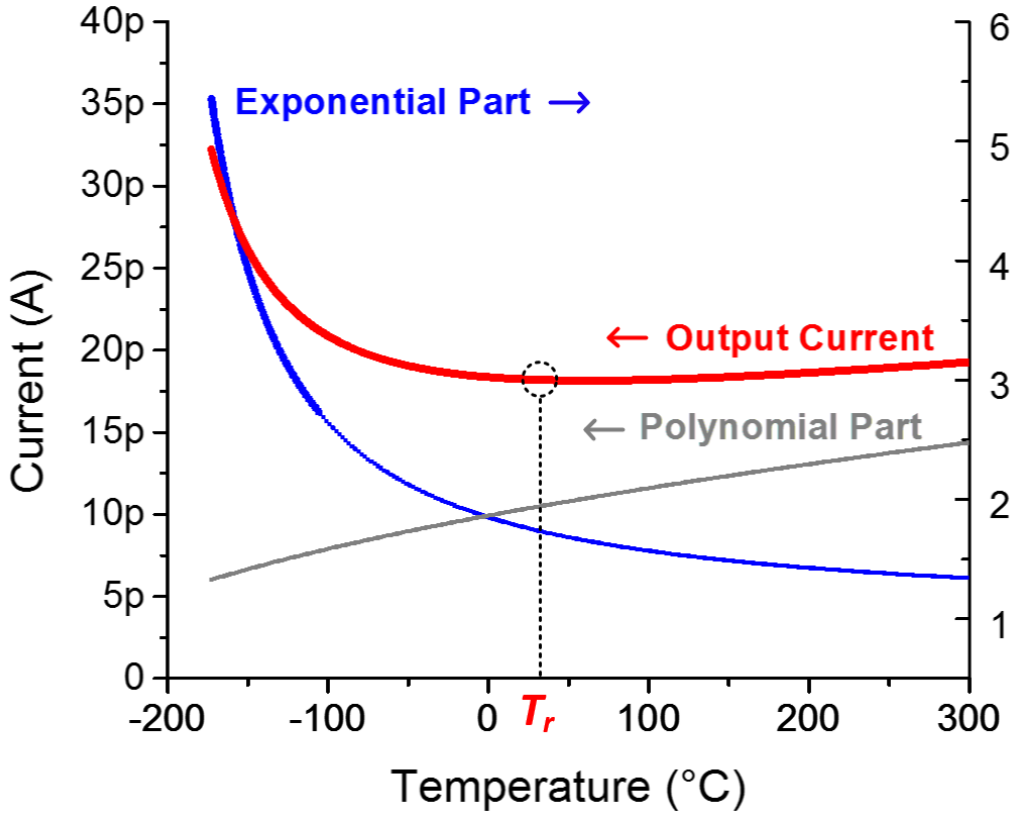


Figure 2.4 MATLAB simulation of the current model in (1)-(7)

To validate this analysis, MATLAB simulation results with the above model are plotted in Fig. 4. The *exp(α2/T)* part decreases while the *T^(1/2)* part increases across the temperature. As they cancel each other, the output current shows nearly constant behavior for the desired range centered at Tr.

B. CTAT Voltage Generator

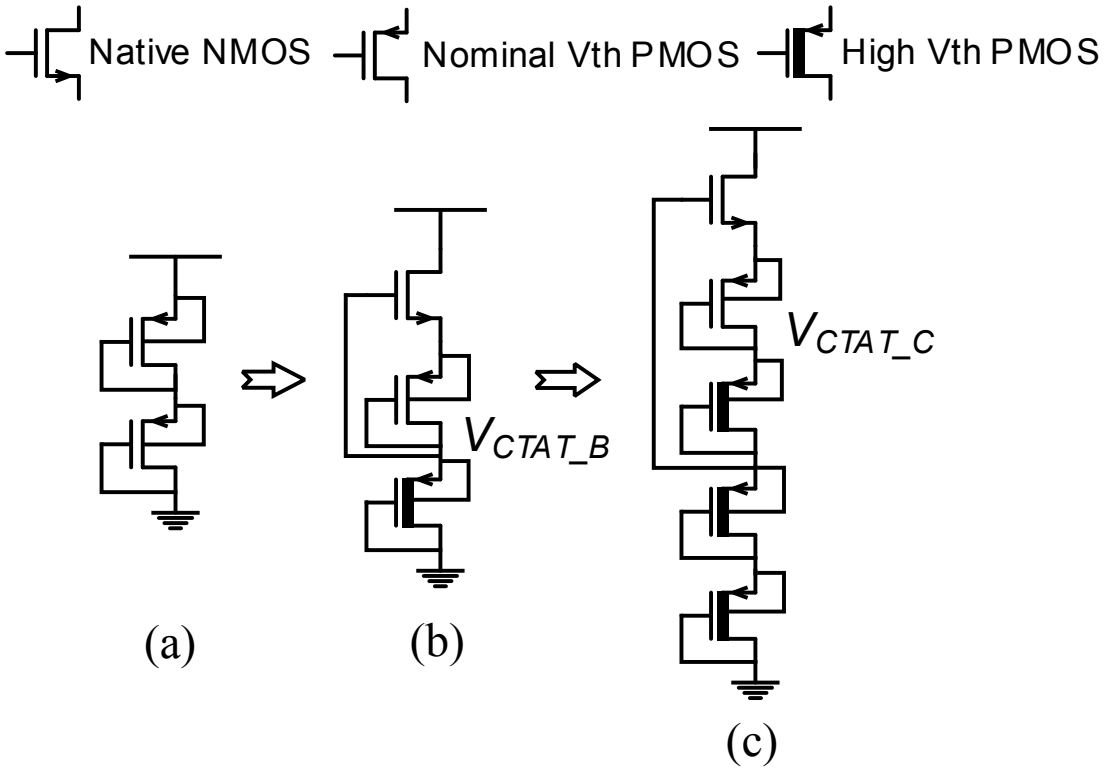


Figure 2.5 (a) Conventional CTAT voltage generator. Modified designs: (b) for better supply noise rejection; (c) additional MOSFETs for higher TC.

A linear CTAT voltage is used to compensate for the temperature dependence of Vth as shown in equation (2). We generate this voltage using a stack of diode-connected transistors with

different sizes (Figure 2.5 (a)) [4], [6]. We modify this traditional CTAT generator in several ways to achieve lower supply sensitivity, the desired temperature coefficient, and reduced power.

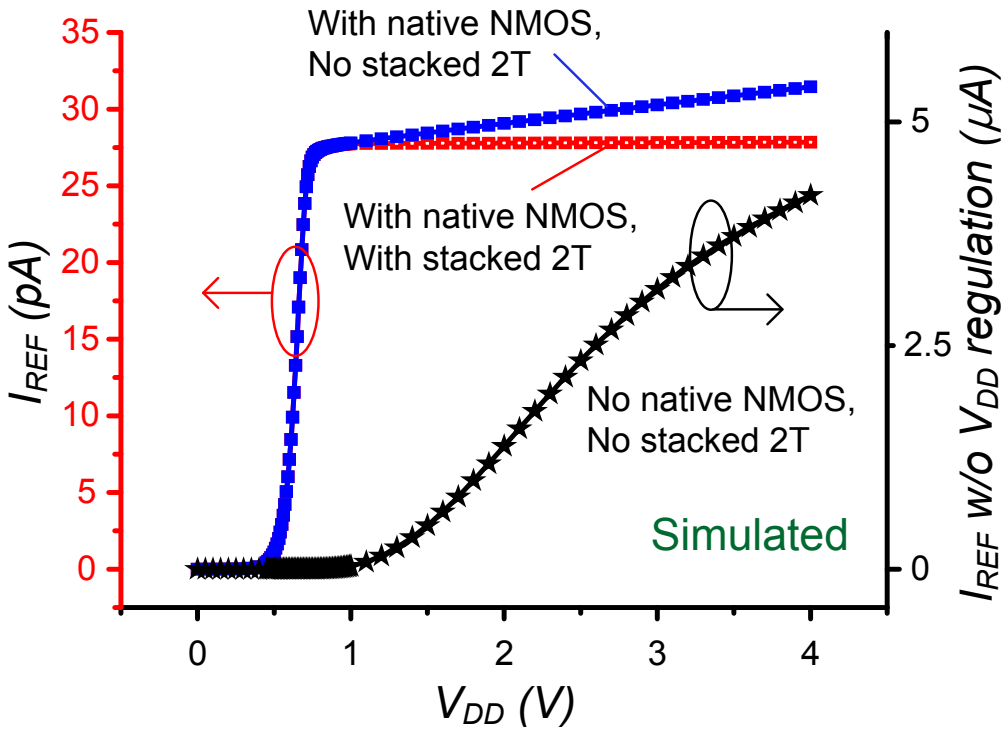


Figure 2.6 Output current vs. VDD with/without VDD regulation techniques.

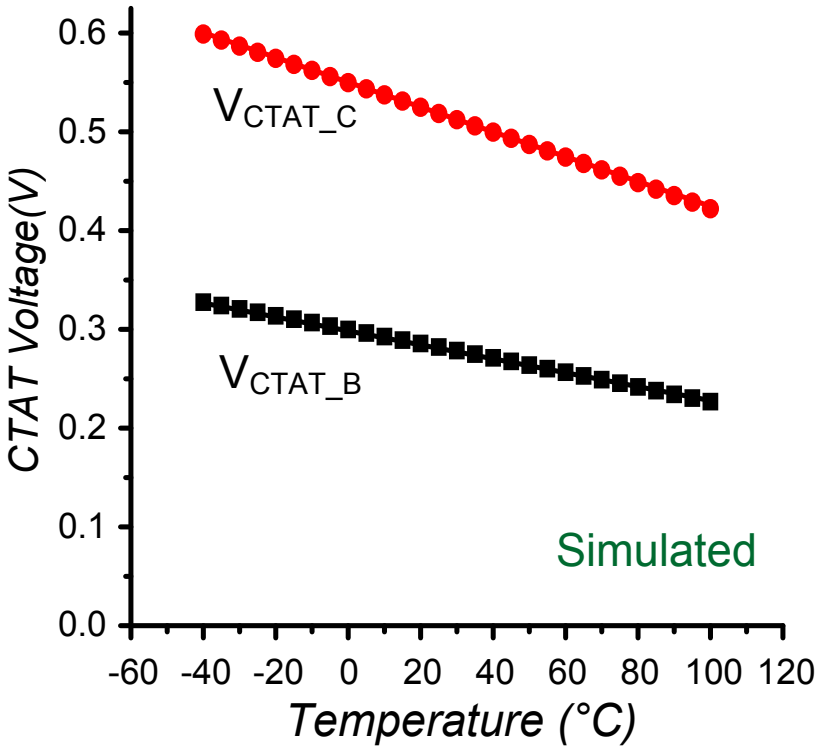


Figure 2.7 CTAT voltage generated by diode-connected transistor stacks in Figure 2.4

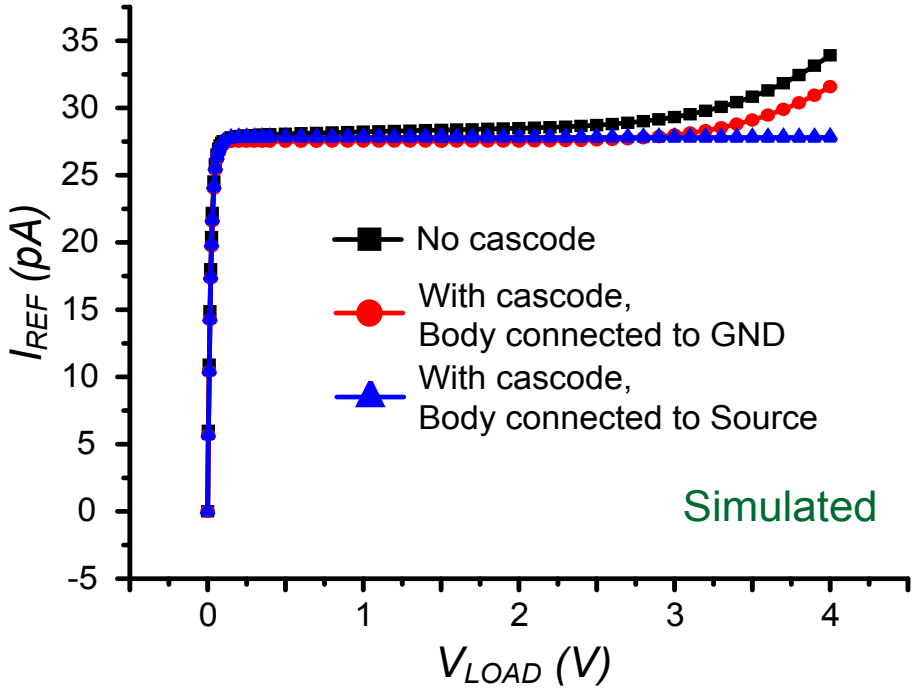
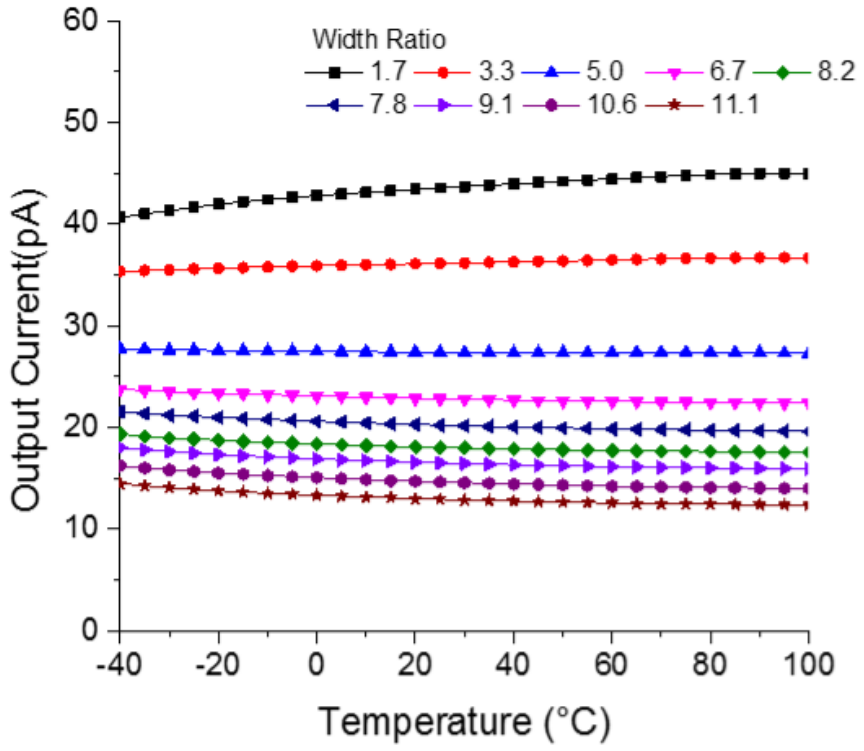


Figure 2.8 Load sensitivity of output current using techniques in Figure 2.5.



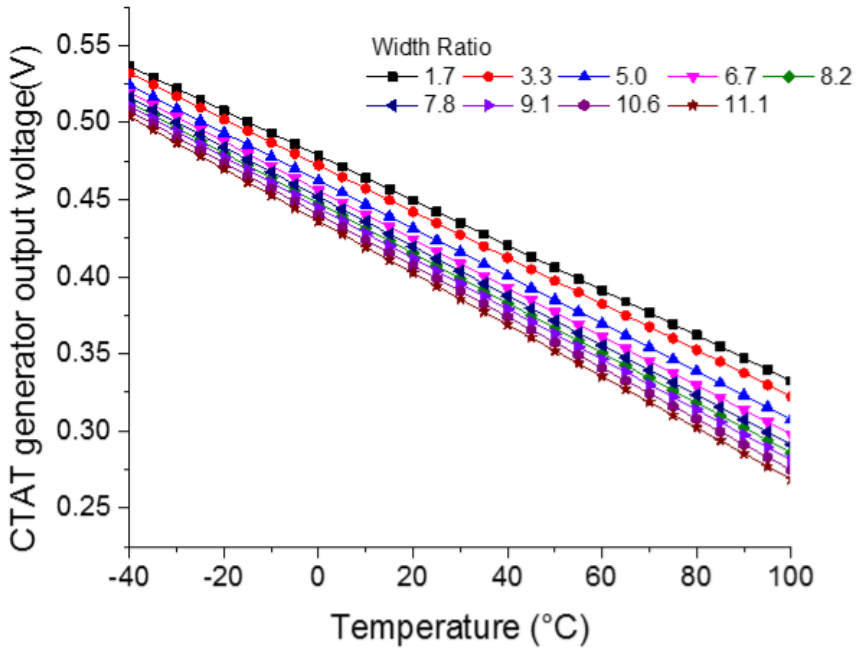


Figure 2.9 Simulation result of (a) output current, (b) CTAT generator output voltage with different ratio of PMOS width in CTAT generator stack.

As seen in Figure 2.5 (b), high-Vth devices are used to minimize power consumption while a native NMOS is added at the top of the stack to reduce supply sensitivity from 4,042%/V to 4.39%/V. An additional supply rejection stage, comprised of two-stacked 2T voltage reference [4] (Figure 2.3, left), further decreases supply voltage sensitivity by a factor of 36× (Figure 2.6). Furthermore, Fig. 5(c) shows the addition of two PMOS transistors, which increases the TC to the required value, from −0.72mV/°C to −1.26mV/°C (Figure 2.7). Figure 2.9 shows that V_{CTAT-C} slope and temperature coefficient of the output current can be controlled by changing transistor width ratio of nominal-Vth PMOS and high-Vth PMOS in CTAT generator.

The threshold voltages of the output transistors in Figure 2.2 vary across process corners, resulting in considerable change in the reference current. This is mitigated by using different device types and channel lengths in the proposed CTAT generator stack, such that the voltage levels of V_{B1} and V_{B2} track that of the threshold voltage of output stage transistors. Short-channel and high-Vth devices are used for the lower three transistors, while long-channel and nominal-Vth devices are used for the upper transistor in the CTAT generator (Figure 2.5(c)). This results in a correlation coefficient of 0.9983 between V_{B1}, V_{B2}, and the threshold voltage of output stage transistors in global corner simulation.

C. Output Stage

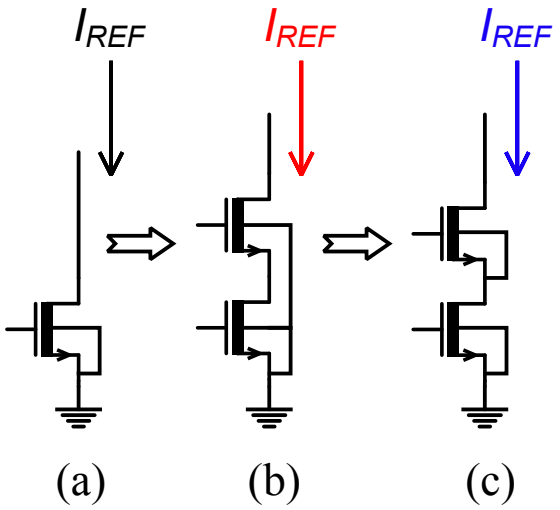


Figure 2.10 Output stage configurations: (a) one subthreshold-biased NMOS; (b) with cascode buffer; (c) body tied to its own source.

In the subthreshold regime, MOSFET drain current is nearly independent of VDS as long as it exceeds 3-4kT/q [4]. However, drain-induced barrier lowering (DIBL) increases load sensitivity to 4.83%/V (simulation). To address this, we use a cascode stack on the output transistor to buffer the drain voltage of the output transistor (Figure 2.10 (b)), reducing load sensitivity to 3.48%/V. To further reduce load sensitivity, the cascode MOSFET body is tied to its own source to prevent substrate current induced body effect (Figure 2.10 (c)). This yields a load sensitivity of 0.35%/V from 0.1V to 4V (simulation, Figure 2.8).

2.3 Measurement Results

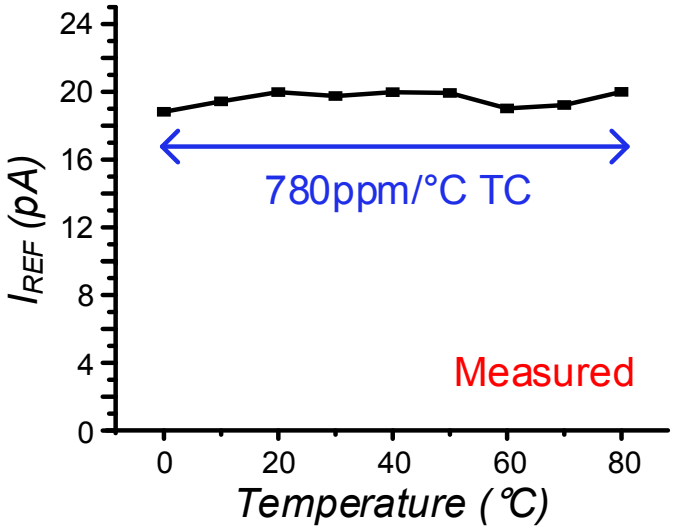


Figure 2.11 I_{REF} across temperature.

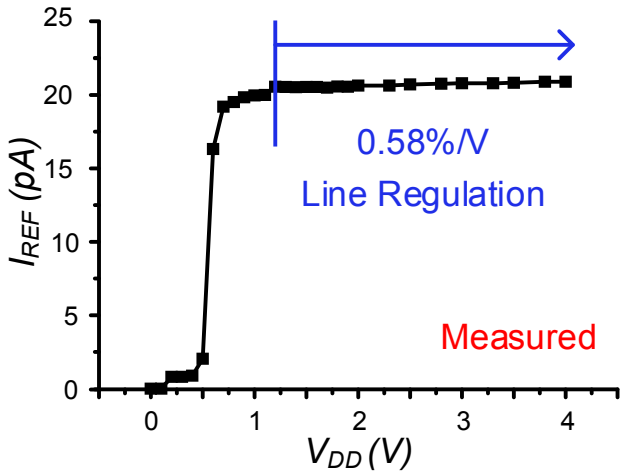


Figure 2.12 I_{REF} across supply voltages.

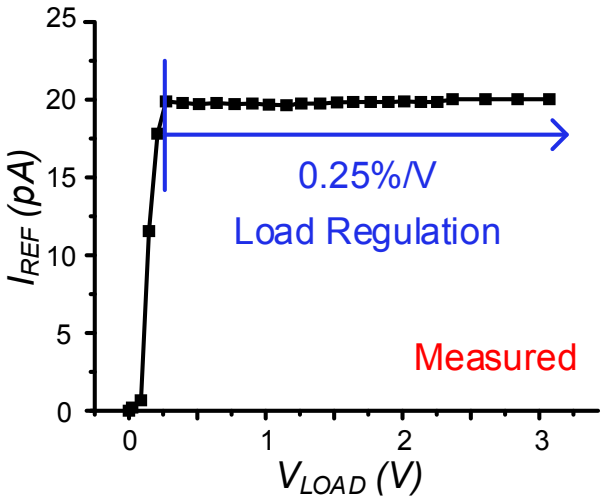


Figure 2.13. I_{REF} across load voltages.

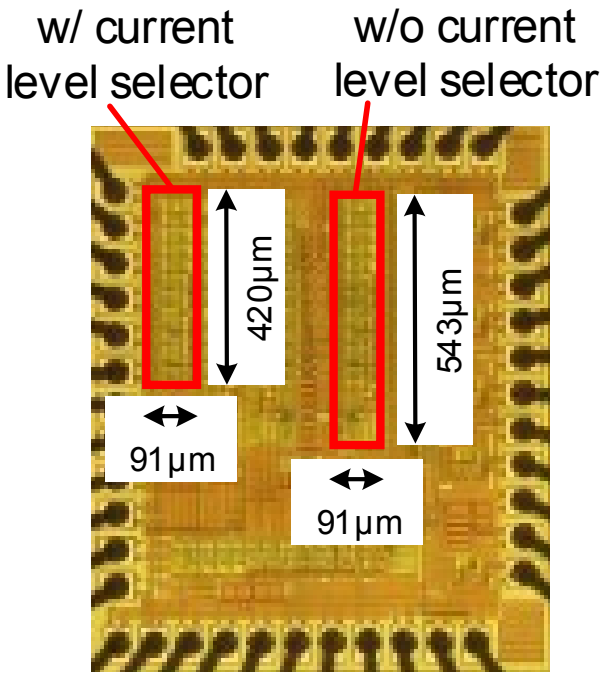


Figure 2.14 Die photo.

The proposed current reference was fabricated in 0.18µm CMOS.

Figure 2.11 shows the measured output current across temperature, which maintains its desired level within 780ppm/°C from 0°C to 80°C.

Figure 2.12 shows measured line sensitivity of 0.58%/V for VDD ranging from 1.2V to 4V. Load sensitivity measurement results are shown in Figure 2.13, showing load sensitivity of 0.25%/V for V_{LOAD} between 0.27V and 3V. Figure 2.14 is the photograph of the fabricated chip.

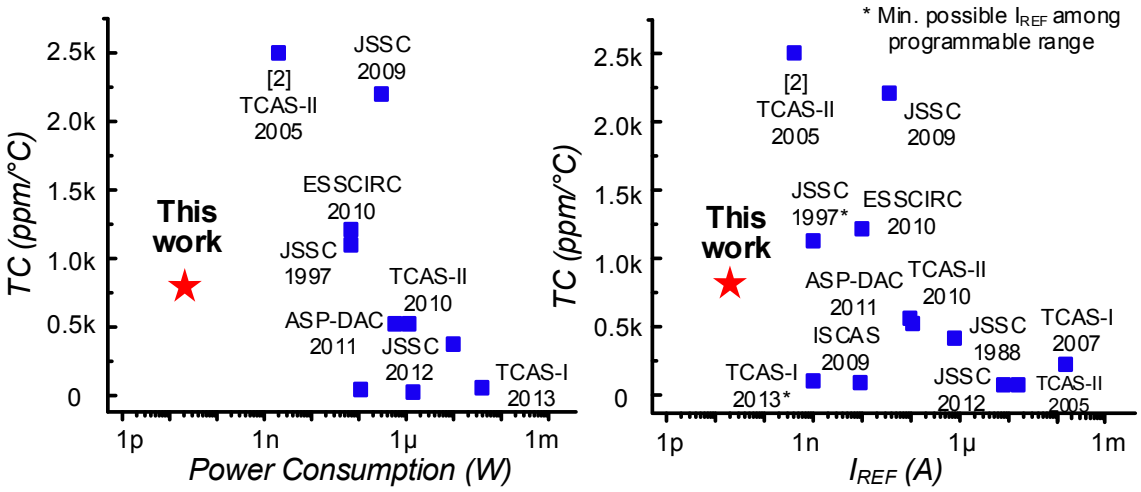


Figure 2.15 Temperature coefficient of reported current reference circuits over power consumption (left) & I_{REF} (right).

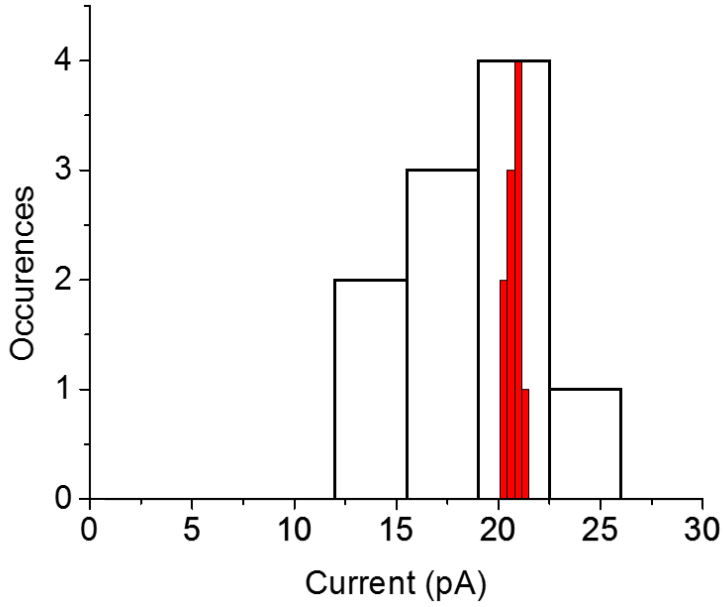


Figure 2.16 Measured current distribution before calibration (black line) and after calibration (red).

Figure 2.15 shows previously reported current references in terms of TC, power consumption and IREF. The proposed current source consumes the lowest power among the shown current references and also enables the lowest regulated output current level. The output current levels of 10 different chips were measured, with results shown in Figure 2.16, both with and without calibration. After calibration, the distribution has a mean current level of 20.79pA and standard deviation of 0.4pA.

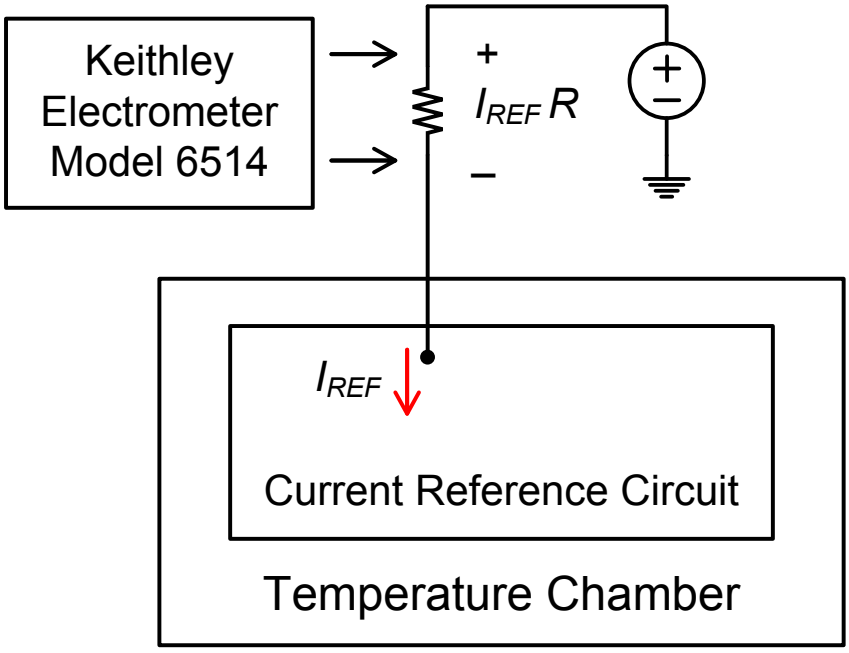


Figure 2.17 Testing scheme for sub-nA current measurement.

Measuring sub-nA currents can be challenging, even with specialized equipment such as electrometers targeting very low current measurements. Hence, we use a scheme that converts current to voltage by generating an IR drop; this eases measurement as voltage can be characterized with very high accuracy and extremely high impedance (>200TΩ) with an electrometer. As shown in Figure 2.17, current reference test chips are placed inside a temperature chamber and the output current is configured to flow through an off-chip resistor that is held at room temperature. This

off-chip resistor is chosen to be 500MΩ such that the voltage drop across it is in the proper electrometer measurement range while ΔVLOAD across temperature is negligible.

2.4 Conclusion

We propose a 23pW 780ppm/°C current reference circuit with line sensitivity of 0.58%/V and load sensitivity of 0.25%/V. This work can be adopted in micro-scale sensor applications to dramatically reduce analog power consumption across temperature.

CHAPTER 3

A 110nW Resistive Frequency Locked On-Chip Oscillator with 34.3 ppm/^{o}C Temperature Stability for System-on-Chip Designs

3.1 Introduction

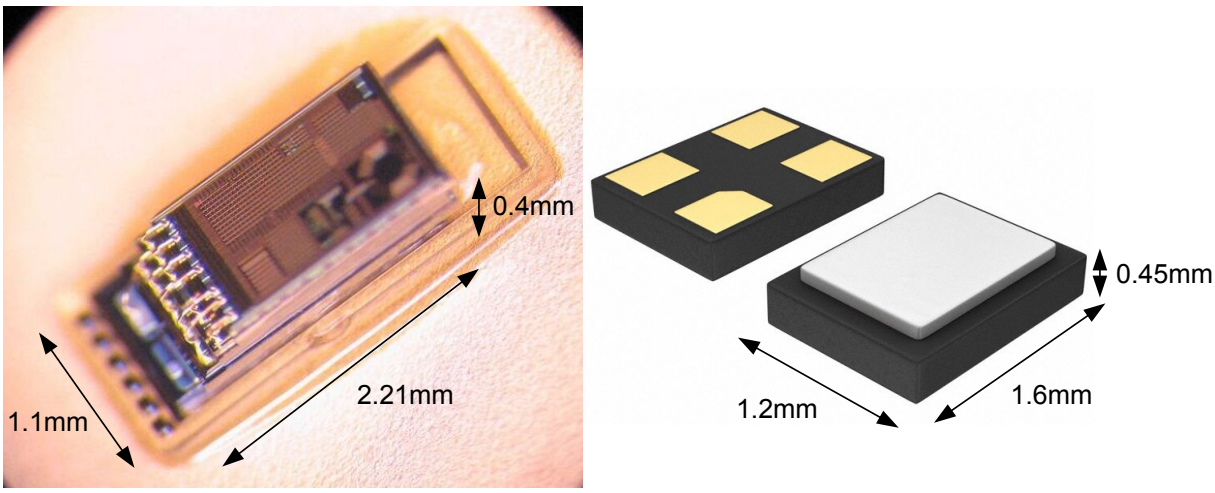


Figure 3.1(a) A picture of millimeter sized wireless sensor node with no board mounting. (b) A picture of one of the smallest off-the-shelf crystal components with its size.

A stable clock source is one of the most important requirements for integrated circuit designs. Although recently introduced techniques allow crystal oscillators to provide a very accurate clock

while consuming as little as a few nanowatts [16], [17] for applications such as a Bluetooth Low Energy sleep timer which requires ±500ppm frequency accuracy, fully integrated on-chip generation of a clock source has become more important as system-on-chip designs have proliferated. More specifically, wireless sensor nodes for Internet-of-Things (IoT) applications have a small form factor and limited board space, making it difficult to integrate crystal oscillators, especially for implantable applications. Figure 3.1 compares the physical size of a recently proposed millimeter-sized sensor system that consists of stacked dies with no board mounting [18] with that of a small-sized off-the-shelf crystal [19]. This comparison illustrates the challenge of integrating external crystal oscillators in millimeter-sized IoT devices.

An on-chip oscillator requires low power consumption and energy per cycle, frequency stability over varying ambient temperatures, long-term stability, and low supply voltage sensitivity. On-chip oscillators that consume little power and exhibit low energy per cycle are important for wireless sensor applications such as [9], [20], [21]. These systems are usually powered by millimeter-sized batteries, and thus, the total energy budget is limited. To reduce power and extend their lifetimes, these systems are highly duty-cycled. They remain in sleep mode the majority of the time and intermittently wake up to measure environmental signals, process the measured data, and wirelessly transmit the data to the outside. As an example, a millimeter-scale wireless imaging system presented in [18] consumes 304nW in its sleep mode, and an electro-magnetic energy harvesting system introduced in [21] requires 190nW in its idle mode. Low oscillator power consumption is important in a system with low activity where the standby current dominates the total power consumption, as is the case with a wake-up timer or a sleep mode timer. This type of timer is turned on even during sleep mode, when most blocks are power-gated, in order to wake

the system periodically, and its power consumption often dominates the total system power consumption in sleep mode.

Low energy consumption per cycle is an important requirement for wireless sensor nodes when the system performs frequent activities and an oscillator's energy consumption can represent a substantial portion of the total system energy. This situation can occur in clocks for a processor, a radio baseband controller, or a power management circuit with switched-capacitor networks. The proposed oscillator is adopted as a clock source to a radio baseband controller [22] which consumes 21.7nJ/bit. One of the state-of-art switched capacitor DC-DC converters [23] outputs 0.1nJ/cycle. It implies that the level of energy/cycle of oscillators should be significantly less than 0.1nJ/cycle to be integrated in such power converter systems. For such applications, the active energy consumption of an oscillator must be kept low because wireless sensor nodes have limited energy budgets.

An oscillator must also show good frequency stability. When implemented in a wake-up timer, an oscillator must maintain a constant system wake-up period across a wide range of temperatures. The frequency stability specification becomes very important when synchronizing nodes for radio transmission. For wireless communication, a transmitter and a receiver must be synchronized to ensure valid data packet transfers. If the oscillators employed in wireless nodes have a high degree of uncertainty and instability, the resulting time window for synchronization must be extended to compensate. Within a packet, the smaller the Allan deviation is, the longer a number of consecutive bits can be transmitted without a separate synchronization header. Between wakeup periods, the frequency drifts due to temperature or supply voltage variations should be smaller than what the receiver can tolerate. More details are explained in Section V. Considering

the unstable battery voltage of sensor nodes during their lifetimes and ambient temperature changes that sensors can face, low temperature and supply voltage sensitivity is important.

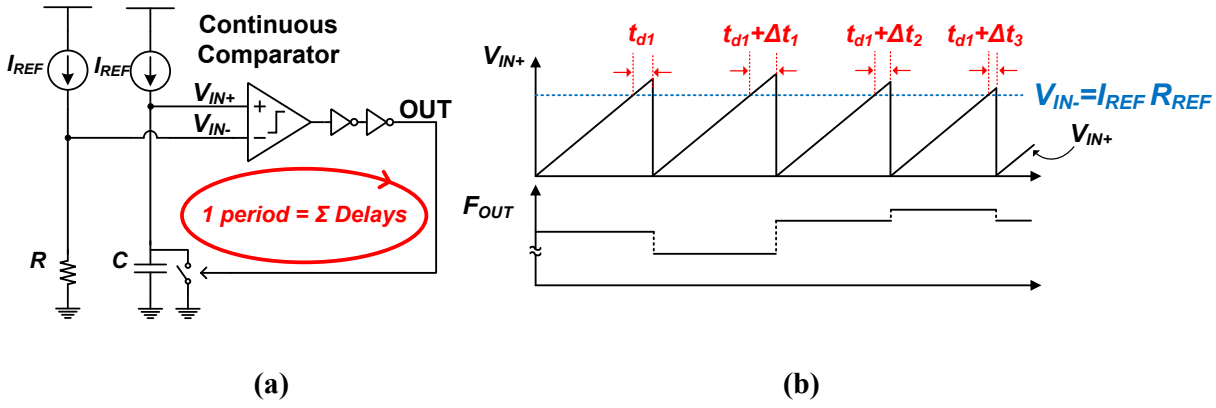


Figure 3.2 (a) Conventional RC relaxation oscillator circuit. (b) Its unstable frequency caused by comparator delay variation.

On-chip clock sources can be generated by various methods such as gate-leakage-based oscillators [24], as well as mobility-based frequency generation, LC oscillators, RC relaxation oscillators, and RC harmonic oscillators as described in [25]. Among the aforementioned approaches, one of the most common structures is an RC relaxation oscillator, illustrated in Figure 3.2 (a). This conventional RC relaxation oscillator is composed of two identical current sources, a resistor, a capacitor with a reset switch, and a comparator with buffers. The negative input voltage (V_{IN-}) of the comparator is set by the product of the source current (I_{REF}) and the reference resistance (R_{REF}). A source current on the right side charges the capacitor, and the capacitor is reset when a positive input voltage, V_{IN+}, exceeds the threshold voltage (V_{IN-}). The reset signal is generated by the comparator and then buffered by a few inverter stages. In this structure, one clock period is the sum of the RC delay, comparator delay, and buffer delays. The RC delay can be temperature-compensated to the first order fairly easily by serially combining a resistor with a positive temperature coefficient and a resistor with a negative temperature coefficient and

trimming the breakdown between the two resistors. On-chip capacitors made of MIM capacitors or traditional metallization layers have negligible temperature coefficients.

However, reducing comparator and buffer delay variation across temperature requires complicated design techniques and remains the main source of temperature instability, as shown in Figure 3.2 (b). In this figure, the varying comparator delay at each cycle is expressed as Δt_x, and is added to each cycle period (t_{d1}). To address this issue, a feed-forward period control was introduced in [26] to cancel comparator delay variation by measuring it and removing the effect with boost charging. However, the replica circuits to measure comparator delay nearly double the required area and power. A comparator offset cancellation technique was proposed that switches the comparator input polarity every half period to tackle temperature-dependent comparator offset voltage [27]. However, comparator delay itself remains in the oscillator period, and thus a significant amount of power is consumed to render the delay of comparator and buffers to be less than 0.4% of oscillator period. A current-mode RC relaxation oscillator [28] eliminates capacitor resetting delay by dual-phase operation, but still has comparator delay in the oscillator period. It is important to note that this comparator delay issue creates a power and temperature stability tradeoff relationship in traditional RC relaxation oscillators, meaning that power consumption increases as temperature stability is improved. Another design [29] achieves 38.2 ppm/ºC with a circuit technique called local supply tracking threshold voltage, but it relies on a dedicated implant process for a zero temperature coefficient poly resistor, which is not always available in other technologies. A constant charge subtraction method was suggested to address comparator delay variation in [30], but the output frequency is limited by the low power amplifier, producing an 11 Hz clock that can only be used in low frequency applications.

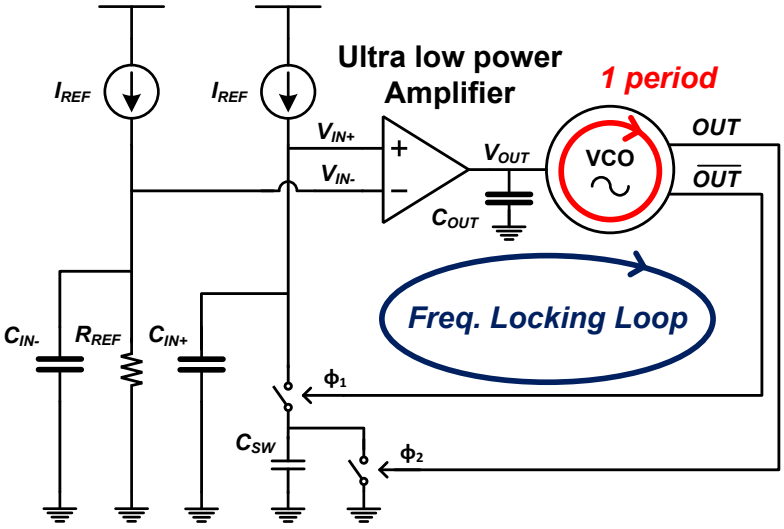
To stabilize comparator delay, a supply-regulated ring oscillator in a temperature compensated loop was introduced in [31]. However, [31] targets a much faster frequency of 10 MHz with a correspondingly much higher power consumption (80 μW) than the work presented here, and also exhibits a relatively high temperature coefficient. An approach that relies on the use of the RC zero voltage crossing time as a timing reference and locks a VCO frequency to this reference time was developed in [32], however the system consumed substantial energy per cycle (11.3 pJ/cycle).

In this paper, a Resistive Frequency Locked on-chip Oscillator (RFLO) is proposed to solve the problems caused by the use of a comparator in existing RC relaxation oscillator structures. This paper is an extension of [5]. This RFLO is based on the principle that a switched-capacitor circuit controlled at a certain frequency can function as a resistor [33], [34]. The RFLO structure replaces a comparator with an ultra-low power amplifier and uses this amplifier to match the resistance of a switched-capacitor circuit to that of a temperature-compensated on-chip resistor. The frequency of the control signal for the switched-capacitor circuit is the output frequency of this oscillator, and the frequency is stabilized by the resistive frequency locked loop. A recent work [35] embedded a VCO in a feedback loop with a current comparator and a frequency-to-current converter to generate a temperature-compensated clock source. However, that work relies on transistor matchings in a current comparator instead of the active control of an amplifier, resulting in a relatively high temperature sensitivity of 90 ppm/ºC and supply voltage sensitivity of 4%/V.

The proposed 70.4 kHz oscillator achieves a temperature sensitivity of 34.3 ppm/ºC, supply voltage sensitivity of 0.75%/V, and long term stability of 7 ppm after an integration time of 12 seconds while consuming 110 nW at room temperature. This paper is organized as follows. Section II describes the operating principles of the RFLO and its design. Section III describes the sources

of temperature-dependent frequency instability and introduces techniques to address each source. Section IV describes the measurement results and chip implementation. Section V shows a radio transmitter circuit integrated with the proposed RFLO as an example of a fully integrated system-on-chip design. Finally, Section VI concludes the paper.

3.2 Resistive Frequency Locked Oscillator



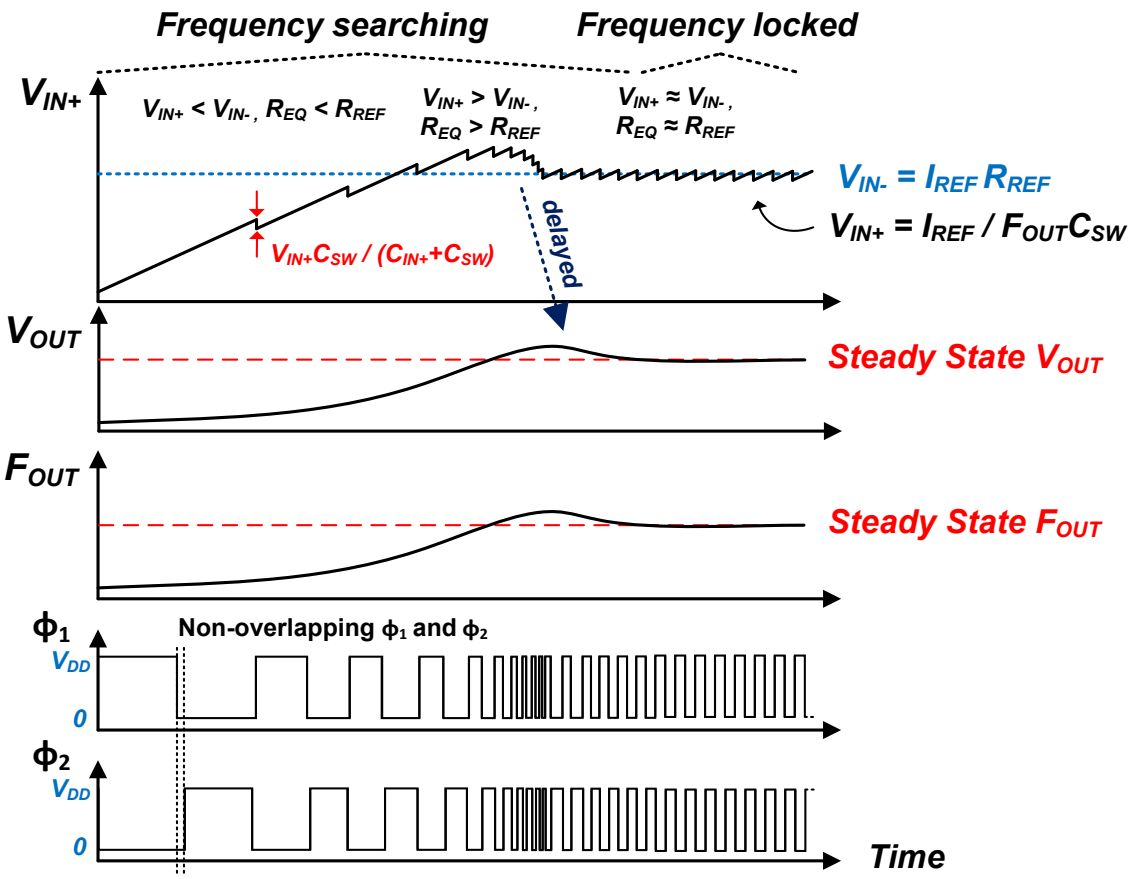


Figure 3.3 (a) Circuit diagram of proposed Resistive Frequency Locked on-chip Oscillator. (b) Its conceptual operating waveforms.

An RFLO removes the comparator from the oscillation loop and adopts a frequency locked loop with an ultra-low power amplifier. The simplified circuit diagram and its operating waveforms are illustrated in Figure 3.3. The basic principle is to generate a stable frequency by matching the equivalent resistance of a switched-capacitor circuit (C_{SW}) to a temperature-compensated on-chip resistor (R_{REF}). A first-order analysis is introduced in this section, followed by an analysis with second-order effects in Section III.

A reference current I_{REF} is injected into R_{REF} to develop a reference voltage $V_{IN-} = I_{REF} \times R_{REF}$, and this voltage is connected to a negative input of an amplifier. The amplifier forces this voltage to match the voltage of positive input node V_{IN+}, where the same I_{REF} flows through C_{SW}.

This V_{IN+}, which is the product of the current and resistance at the node, can be expressed as Equation (1) because an equivalent resistance of a switched-capacitor circuit operating at a frequency of F_{OUT} is $1/(C_{SW}F_{OUT})$. Here, F_{OUT} is the VCO frequency controlled by the amplifier output, V_{OUT}. By equating V_{IN+} and V_{IN-} as shown in Equation (2), the VCO frequency F_{OUT} is defined by R_{REF} and C_{SW} as derived in Equation (3). As the two reference current terms (I_{REF}) in V_{IN+} and V_{IN-} cancel out in the equation, F_{OUT} is insensitive to I_{REF}. Furthermore, assuming the reference currents are independent of the supply voltage, F_{OUT} is also insensitive to supply voltage fluctuation as the supply voltage does not appear in Equation (3).

$$V_{IN+} = \frac{I_{REF}}{C_{SW}\, F_{OUT}} \tag{1}$$

$$V_{IN+} = V_{IN-} (= I_{REF} R_{REF}) \tag{2}$$

$$F_{OUT} = \frac{1}{R_{REF}\, C_{SW}} \tag{3}$$

R_{REF} is temperature compensated in this implementation, and C_{SW} is a MIM capacitor with very low temperature dependency, and thus, a highly temperature-stable frequency is generated.

Figure 3.3 (b) shows the RFLO locking process in time domain, starting from a point where the VCO frequency, F_{OUT}, is lower than the target frequency. In this condition, the charge pumped out of V_{IN+} by C_{SW} is less than the charge flowing in from I_{REF}, and thus, V_{IN+} rises. When V_{IN+} matches V_{IN-}, the VCO frequency can be locked depending on the damping ratio of the frequency locked loop. As an example, an overshooting case is shown in this figure. After V_{IN+} equals V_{IN-}, it slightly exceeds V_{IN-}, and the VCO frequency increases because the VCO is biased at a higher voltage than before. The VCO resets C_{SW} more frequently, and thus, the charge pumped out of V_{IN+} is now greater than the charge flowing in from I_{REF}. Thus, V_{IN+} decreases and again approaches V_{IN-}, and the VCO frequency locks. For more quantitative analysis, impedance at the

node V_{IN+} (Z_{VIN+}) and its partial derivative with respect to F_{OUT} can be expressed as Equation (4). At steady-state, output frequency settles at F_{OUT0}, which is $1/R_{REF}C_{SW}$.

$$Z_{VIN+} = \frac{1}{sC_{IN+}} \parallel \frac{1}{C_{SW}F_{OUT}}, \qquad \frac{dZ_{VIN+}}{dF_{OUT}} = \frac{-C_{SW}}{(sC_{IN+} + C_{SW}F_{OUT0})^2} = \frac{-C_{SW}}{\left(sC_{IN+} + \frac{1}{R_{REF}}\right)^2} \quad (4)$$

Frequency response of the frequency locked loop is derived in Equation (5). A_V is an amplifier gain, R_{OUT} (0.3GΩ from simulation) is an amplifier output resistance, C_{OUT} (10pF) is a capacitor at the node V_{OUT}, and K_{VCO} (1.41 MHz/V from simulation) is the gain of VCO. This loop has one pole at $-\frac{1}{C_{OUT}R_{OUT}}$, and 2 poles at $-\frac{1}{C_{IN+}R_{REF}}$. The dominant pole is located at ω=330 rad/s, and two non-dominant poles are located at ω=7.5 k rad/s.

$$\begin{aligned} F_{OUT}(s) &= A_V \frac{\frac{1}{sC_{OUT}}}{R_{OUT} + \frac{1}{sC_{OUT}}} K_{VCO} I_{REF} \left(- \frac{C_{SW}}{(sC_{IN+} + \frac{1}{R_{REF}})^2} \right) \\ &= A_V \frac{1}{1 + sC_{OUT}R_{OUT}} K_{VCO} I_{REF} \left(- \frac{C_{SW}}{(sC_{IN+} + \frac{1}{R_{REF}})^2} \right) \end{aligned} \quad (5)$$

The ripples caused by capacitor switching exist on the V_{IN+} node, but their amplitude is small due to the C_{SW}/C_{IN+} ratio of 0.09 (=0.9pF/10pF). The low bandwidth of the ultra-low power amplifier works as a low pass filter and helps to further reduce ripples and to stabilize F_{OUT}. The amplifier of this design has gain of −9dB at 70.4 kHz from simulation. C_{OUT} (=10pF) and an output resistance of the amplifier (=0.3GΩ) make a first-order low pass filter with cutoff frequency of 53Hz, yielding gain of −62dB at 70.4 kHz. This combination results in a gain of −71dB at the ripple frequency, suppressing voltage ripple at the node V_{OUT} as low as 4μV. Two clock signals, ϕ1 and ϕ2, are non-overlapping clocks. Changes in V_{IN+} and V_{IN-} appear at V_{OUT} with some delays

due to the limited bandwidth of the amplifier as shown in Figure 3.3 (b). However, in a steady-state where the frequency is stabilized, this low bandwidth of the amplifier does not disturb accurate clock generation.

The proposed topology has the following key advantages over a traditional RC relaxation oscillator topology. First, it removes the traditional comparator from the oscillation loop, thereby removing the power and temperature stability trade-off introduced by the comparator. Second, the amplifier, which replaces the power consuming comparator and provides frequency locking, consumes very little power. This is possible because the amplifier must only track the impact of ambient temperature changes on the VCO. These temperature changes are slow, and hence the amplifier can be low-bandwidth and ultra-low power. Third, this structure shows good long-term stability. Any slight deviation in frequency in a particular cycle i results in a slight difference in the charge flowing into and out of node V_{IN+} (the charge is noted as ΔQ_i). Unlike a traditional relaxation oscillator in which the circuit is reset every cycle, and hence the charge discrepancy is lost, this topology carries ΔQ_i over from one cycle to the next and accumulates it on capacitor C_{IN+}, as shown in Equation (6). If most of the frequency error comes from random noise sources, the sum of ΔQ_i over many cycles approaches zero, resulting in a V_Δ of nearly 0V.

$$V_\Delta = \sum_{i=0}^{N} \frac{\Delta Q_i}{C_{IN+}}, \quad \text{for N cycles} \tag{6}$$

Even with a non-zero V_Δ after N cycles, the amplifier compensates for the error by adjusting the frequency in subsequent cycles, providing excellent long-term frequency stability. The frequency stability after long integration time is mainly limited by flicker noise. Also, an amplifier offset varying over time, long-term drifts of resistors and capacitors can be the sources of the long-term frequency inaccuracy. Finally, this structure shows low supply voltage sensitivity because the

frequency is only defined by R_{REF} and C_{SW}, as shown in Equation (3) in the first order analysis. In the second order, gain and an offset voltage of the amplifier slightly vary with the supply voltage. I_{REF}s generated by an internal current reference circuit change with the supply voltage affecting the amplifier DC input operating points. These non-idealities affect how accurately V_{IN+} and V_{IN-} match which defines the frequency accuracy. The VCO frequency is controlled by the amplifier output voltage, and this bias voltage is automatically adjusted by a frequency locked loop when the supply voltage changes.

3.3 Sources of Temperature Instability and Solutions

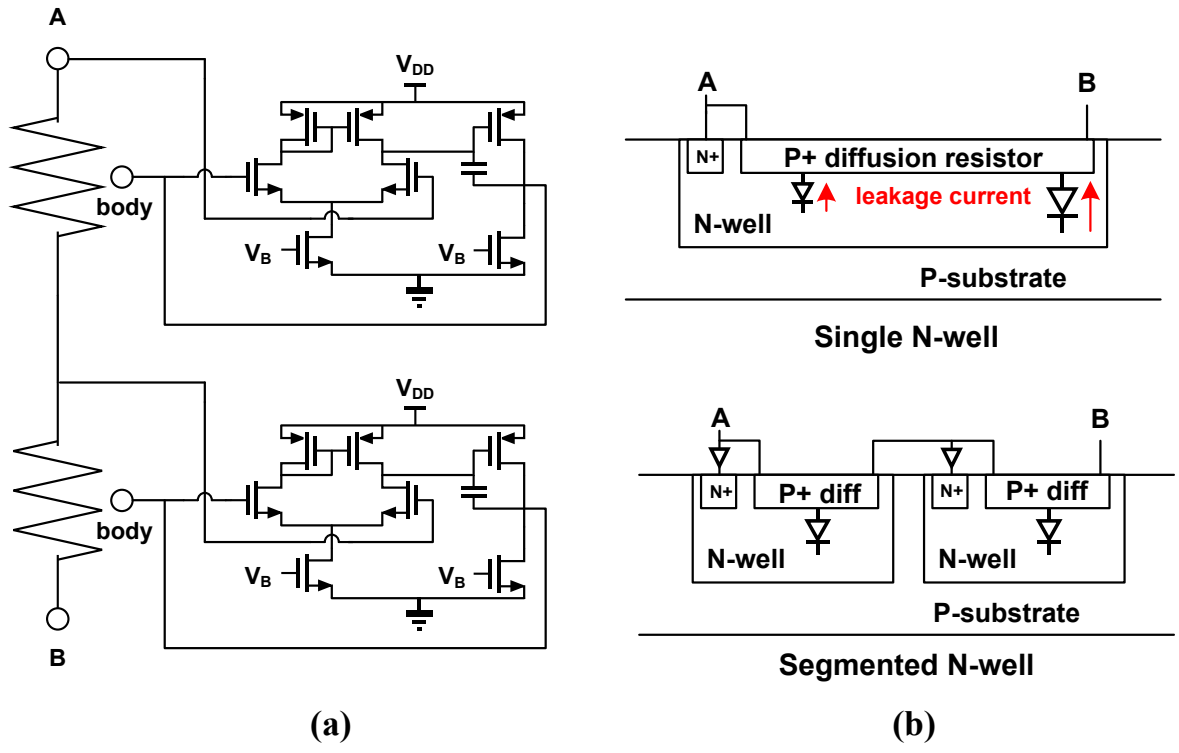


Figure 3.4 (a) A schematic of segmented N-well technique to address well leakage current. (b) Cross section of segmented N-well technique shown as a physical layout.

This section discusses possible sources of temperature instability in the proposed RFLO design and describes solutions for each source. To begin with, F_{OUT} is only defined by $1/R_{REF}C_{SW}$

under ideal conditions. While MIM capacitors have a very low temperature coefficient, on-chip resistors show a non-zero temperature coefficient. In this work, a negative temperature coefficient (TC) poly resistor without silicide is serially combined with a positive TC diffusion resistor without silicide in order to cancel their temperature dependencies [26]. The ratio between the two resistors is 2-point on-chip trimmed after fabrication to compensate for chip-to-chip process variation. Typical values for a poly resistor and a diffusion resistor are 11.8MΩ and 1.5MΩ, and a typical current value for each I_{REF} is 12nA. For the diffusion resistor, leakage current through a reverse-biased well diode introduces a non-linear temperature dependency as much as 0.18% of voltage error across the resistor at 80ºC in simulation. This error translates to 14.6ppm/ºC of frequency inaccuracy. Furthermore, as this error is non-linear, it cannot be effectively corrected by two point on-chip trimming. A segmented N-well technique shown in Figure 3.4 [30] is adopted to address this well leakage current, increasing the maximum operating temperature from 50ºC to 80ºC in simulation. Figure 3.4 (a) and (b) illustrate a schematic and a cross-section of a physical layout of the segmented N-well technique, respectively. For a diffusion resistor, the leakage current of a reverse-biased diode increases as the voltage difference between P+ diffusion and the N-well increases. With this technique, the diffusion resistor is divided into two segments so that the maximum voltage difference is reduced by half. In addition, the inserted buffers prevent leakage currents into the N-well from altering the total current flowing through Terminal A to Terminal B, as shown in Figure 3.4. Power and area overhead for this technique are 0.1% and 3.4% of the total circuit.

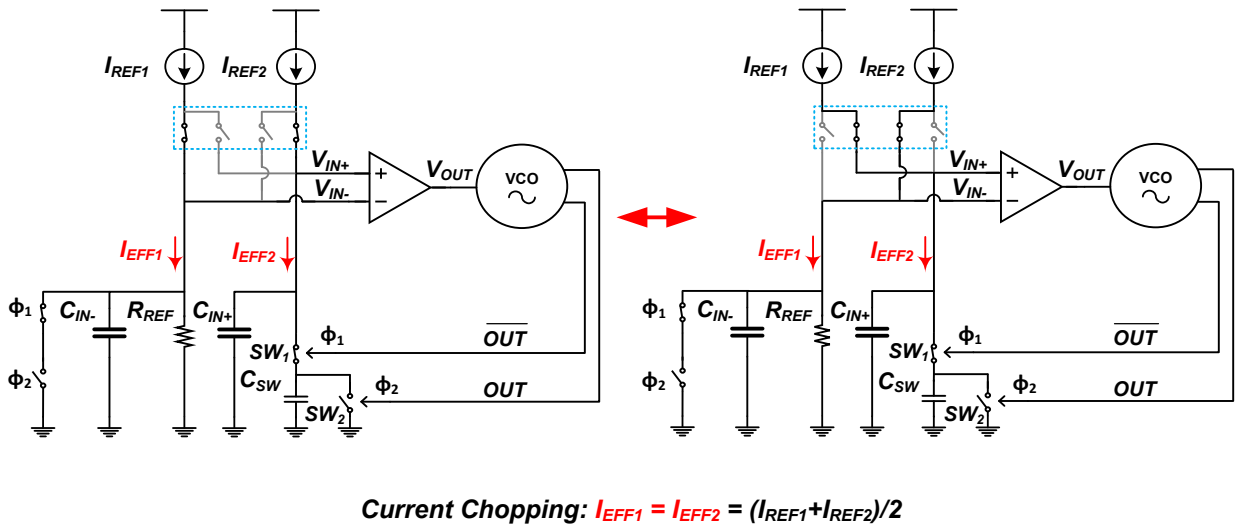


Figure 3.5 A schematic of current chopping technique.

Mismatch between I_{REF1} and I_{REF2} does not affect temperature stability if the mismatch is constant across temperature; a fixed current mismatch only introduces a fixed frequency offset. However, if the current mismatch varies over temperature, it impacts the temperature stability of F_{OUT}. To address this problem, two current sources, I_{REF1} and I_{REF2}, alternate their connections to each input node of the amplifier as illustrated in Figure 3.5. VCO outputs control this alternation. Each amplifier input node is connected to I_{REF1} for one half of its operating time and to I_{REF2} for the other half of its operation. As a result of this chopping scheme, the effective current at each input is the average of I_{REF1} and I_{REF2}, removing frequency errors caused by current mismatch.

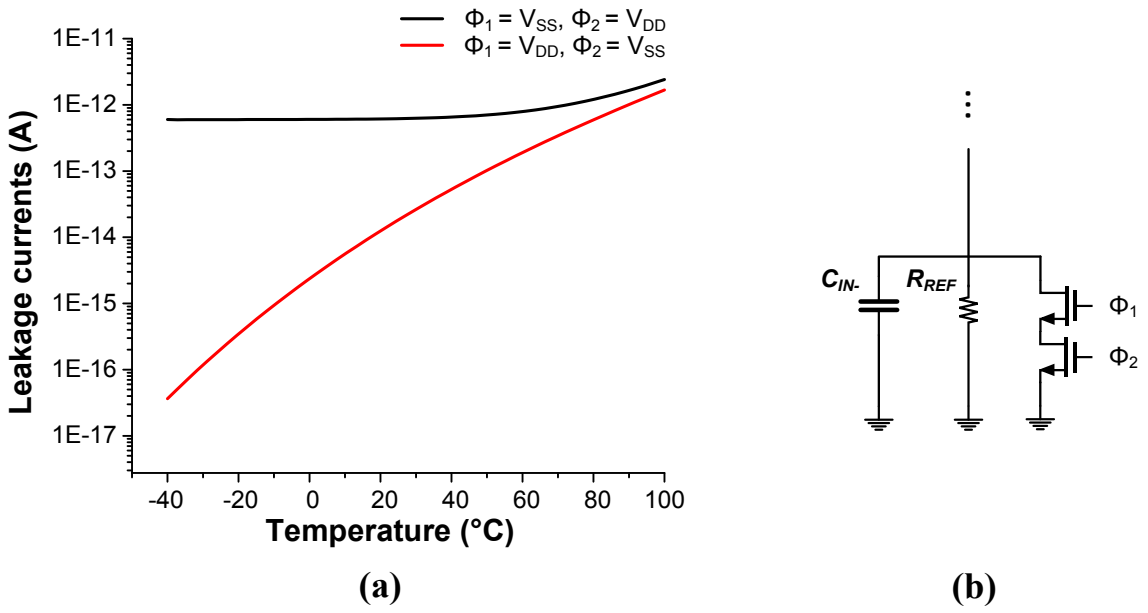


Figure 3.6 (a) Simulation results of leakage current of Switches 1 and 2 at different temperatures. (b) A schematic of dummy switches.

Switches, shown as $SW_{1,2}$ in , also have minor impact on temperature stability. If the V_{IN+} or V_{IN-} levels change, the switch parasitic capacitance (C_{PAR}), which consists of transistor gate-to-drain and body-to-drain capacitances, varies non-linearly and alters the total capacitance at V_{IN+}. To reduce this effect, C_{SW} is sized so that C_{PAR} is less than 0.02% of C_{SW}. In addition to this parasitic capacitance issue, the leakage current (I_{leak}) of $SW_{1,2}$ should be properly dealt with. I_{leak} increases from sub-pA levels at −20ºC to ~2pA at 80ºC in simulation as shown in Figure 3.6 (a), which can create 167ppm of frequency inaccuracy. The magnitude of I_{leak} differs by switching phases. I_{leak} changes the effective current at V_{IN+} node and worsens temperature stability. To address this effect, identical dummy switches are added at the V_{IN-} node as shown in Figure 3.6 (b). Equations (7)-(9) summarize how the described current-chopping technique and dummy switches eliminate error sources for frequency accuracy. Equation (7) describes the effective current flowing at V_{IN-} node considering current chopping and switch leakage current. The same amount of current flows

through V_{IN+} as described in Equation (8). As a result, two identical effective currents are cancelled in Equation (9), and the first-order errors affecting F_{OUT} are mitigated.

$$V_{IN-} = \left(\frac{I_{REF1} + I_{REF2}}{2} - I_{leak}\right) R_{REF} \tag{7}$$

$$V_{IN+} = \frac{\frac{I_{REF1} + I_{REF2}}{2} - I_{leak}}{F_{OUT} \times C_{SW}} \tag{8}$$

$$F_{OUT} = \frac{\frac{I_{REF1} + I_{REF2}}{2} - I_{leak}}{\left(\frac{I_{REF1} + I_{REF2}}{2} - I_{leak}\right) \times R_{REF} \times C_{SW}} = \frac{1}{R_{REF} \times C_{SW}} \tag{9}$$

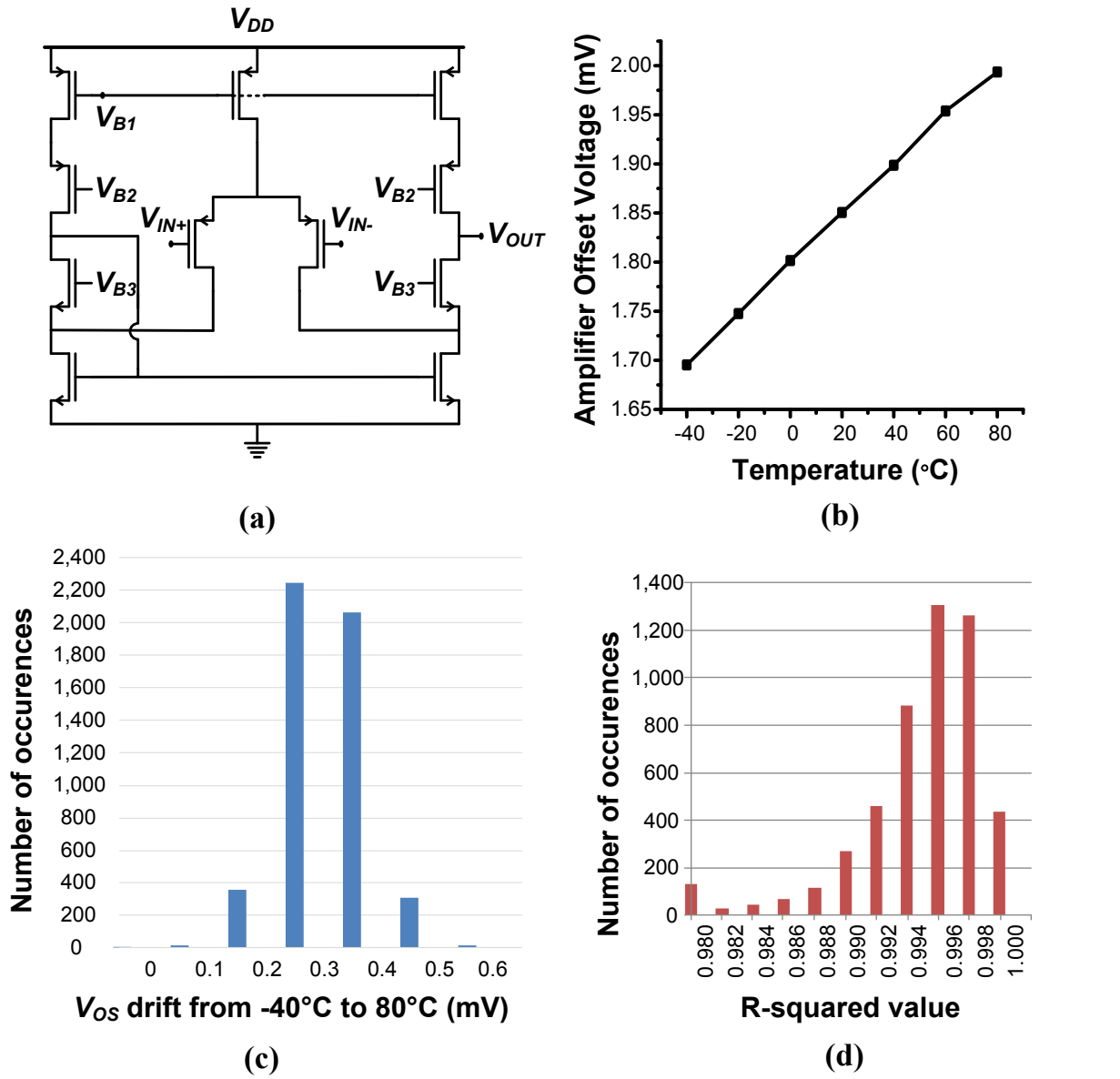


Figure 3.7 (a) A schematic of a subthreshold mode ultra-low power amplifier. (b) An average result of Monte Carlo simulation of the amplifier offset voltage at different temperatures. (c) Histogram of total V_{OS} drift from −40°C to 80°C of each Monto Carlo runs (total 5,000 runs). (d) Histogram of R-squared value (offset voltage vs. temperature) of each Monto Carlo runs (total 5,000 runs).

The amplifier used in the proposed design is a 1-stage folded cascode structure operating in the subthreshold region (Figure 3.7 (a)). This amplifier provides 85 dB DC gain, 1.8 kHz bandwidth, and wide output range of 0.4—0.8 V while consuming only 3.6 nW at room

temperature in simulation. An offset voltage of this amplifier (V_{OS}) does not affect temperature stability if V_{OS} is constant over temperature. However, V_{OS} drift over temperature affects temperature stability. Average values of Monte Carlo simulation results of V_{OS} at different temperatures are shown in Figure 3.7 (b) and a histogram of the total V_{OS} drift from −40°C to 80°C of 5,000 run Monte Carlo simulation is shown in Figure 3.7 (c). The average V_{OS} drift is 0.298mV, which corresponds to 16.7ppm/°C. An auto-zeroing technique introduced in [5] can reduce V_{OS} itself and V_{OS} drift, but if the V_{OS} changes linearly with temperature, it can be cancelled out by 2-point on-chip trimming in the first order without overhead of auto-zeroing. Simulation results in Figure 3.7 (d) show that 87% of 5,000 Monte Carlo runs have R-squared value greater than 0.99 (the closer R-squared value is to 1, V_{OS} with respect to temperature is more linear), and the average value of R^2 is 0.993, which de-emphasizes necessity of auto-zeroing. A finite gain of the amplifier (A_V) also generates a frequency offset, and the resulting frequency at steady state is derived in Equation (10-11). An error from the finite gain decreases as A_V and K_{VCO} increases, and an error from V_{OS} decreases as I_{REF} increases. Each current source is simplified to I_{REF} as current non-idealities are already analyzed in Equation (9).

$$V_{IN+}\left(=\frac{I_{REF}}{C_{SW}F_{OUT}}\right)=\left(1+\frac{1}{A_V K_{VCO}}\right)V_{IN-}+V_{OS}=\left(1+\frac{1}{A_V K_{VCO}}\right)I_{REF}R_{REF}+V_{OS} \quad (10)$$

$$F_{OUT}=\frac{1}{\left(1+\frac{1}{A_V K_{VCO}}\right)C_{SW}R_{REF}+\frac{V_{OS}}{I_{REF}}C_{SW}} \quad (11)$$

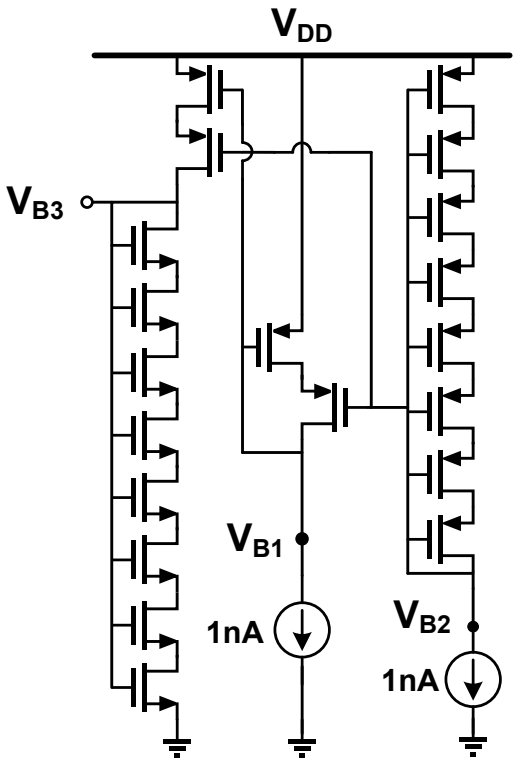


Figure 3.8 A bias voltage generation circuit for the ultra-low power amplifier in Fig. 7.

Figure 3.8 describes a bias voltage generation circuit for this amplifier. The 1nA current references described in Figure 3.8 are implemented on-chip with the resistor-less techniques introduced in [3].

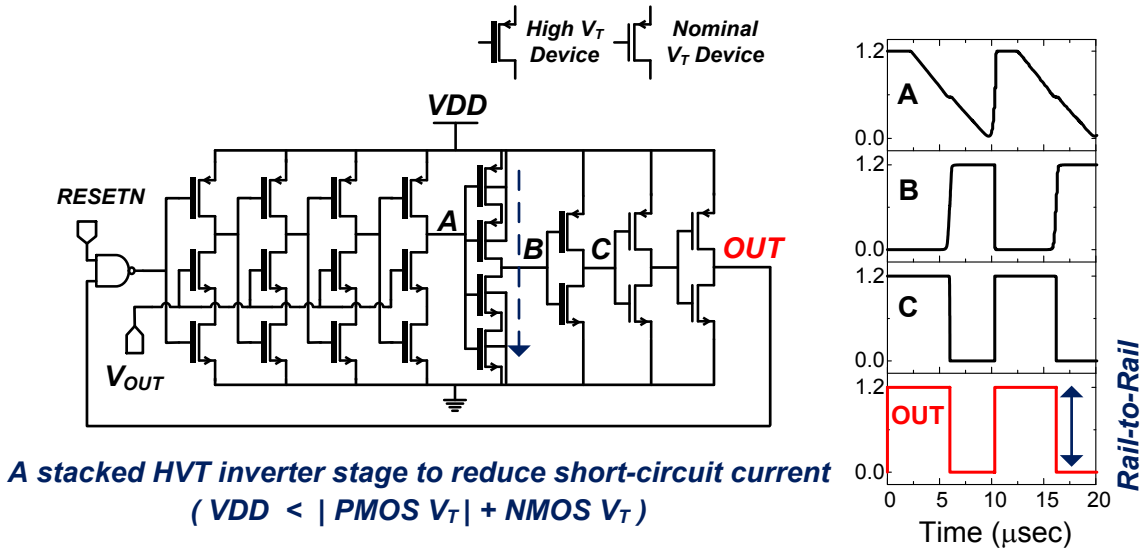


Figure 3.9 A schematic of rail-to-rail voltage controlled oscillator and its simulated waveforms.

Figure 3.9 shows a circuit diagram of the VCO used in this work. This VCO operates rail-to-rail with a wide frequency range and low power consumption. The VCO frequency is designed to be highly sensitive to bias voltage V_{OUT}. This high sensitivity relaxes the required output operating range of the amplifier. To achieve this high frequency sensitivity to bias voltage, the delay of the first four stages is designed to be exponential with V_{OUT} using high V_T NMOS transistors operating in their subthreshold region where drain current is exponential with gate-to-source voltage. The next four stages are buffers to restore the slew rate with low short-circuit current. The first stage is a stacked inverter with high V_T devices, the second stage is an inverter with high V_T devices, and the last 2 stages are inverters with normal V_T devices. Using this manner of staged output buffers reduces VCO power (10.3 nW in simulation) by 67× at a supply voltage of 1.2V through minimizing the short-circuit current while reducing the signal transition time. As F_{OUT} is set by $1/R_{REF}C_{SW}$ and controlled in a closed loop, the RFLO does not require the VCO to have a linear voltage-frequency relation, thereby relaxing the VCO specification.

3.4 Measurement Results

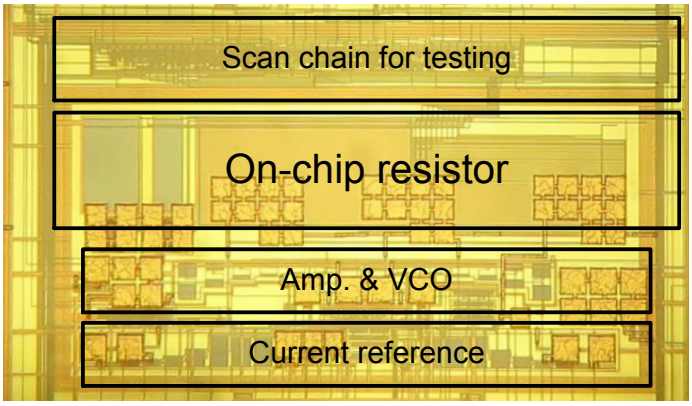


Figure 3.10 Die photo of the proposed RFLO in 0.18 μm CMOS.

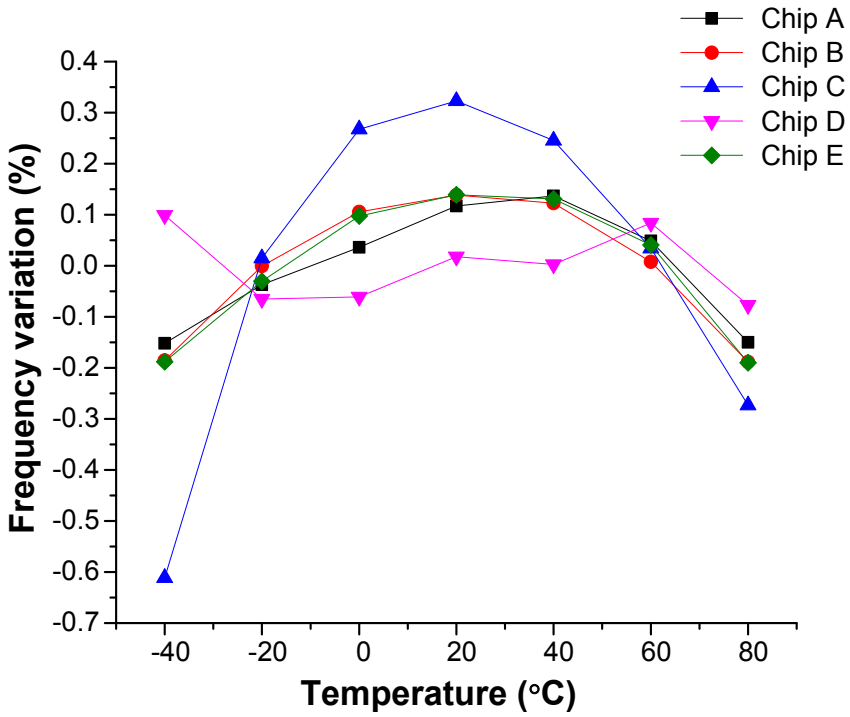


Figure 3.11 Measured frequency variation with respect to temperature.

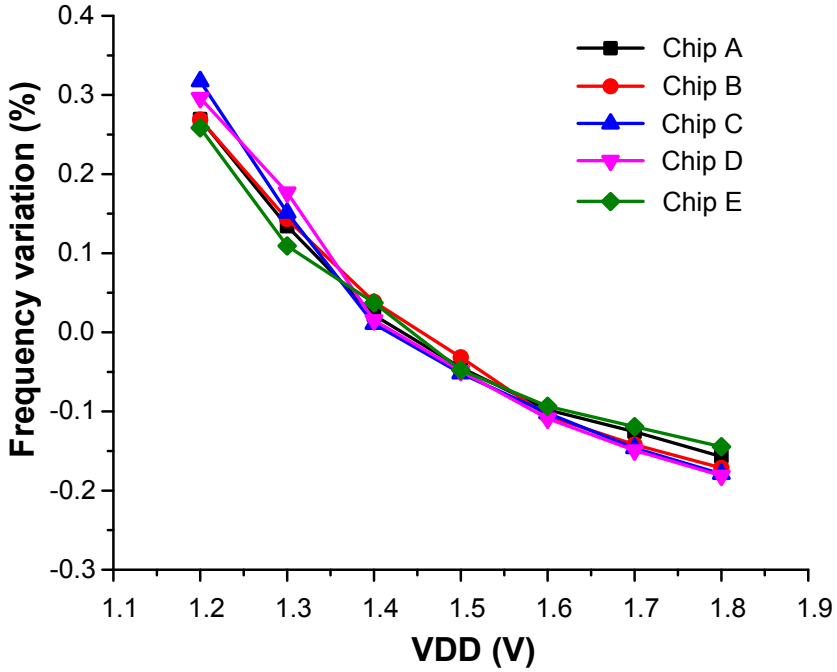


Figure 3.12 Measured frequency variation with respect to supply voltage.

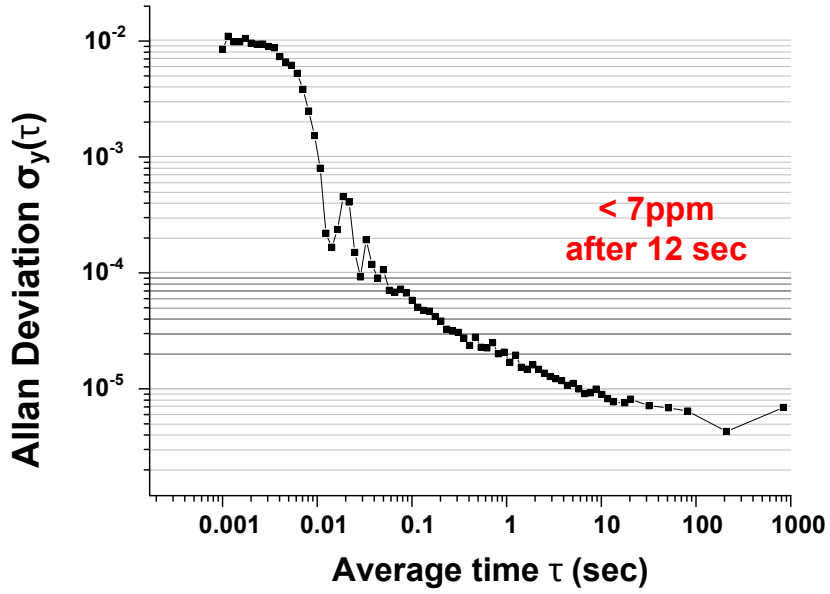


Figure 3.13 Measured Allan Deviation.

The proposed design was fabricated in 0.18 μm CMOS with total area of 0.26 mm^2. Figure 3.10 shows the die photo. The area occupied by transistors can be reduced by porting this design to advanced technologies. The total area can be further reduced by adopting a duty-cycled resistor

technique [36] as the temperature compensated on-chip resistor occupies 0.11mm^2 (42.3% of the total area) in this design. The duty-cycled resistor technique increases an equivalent resistance of a resistor by 1/duty cycle. The clock frequency of this design is 70.4 kHz and has an average temperature stability of 34.3 ppm/ºC between −40ºC and 80ºC for five measured chips as shown in Figure 3.11. The ratio of the positive TC on-chip resistor to negative TC resistor is trimmed on-chip at 2 temperatures to have the lowest TC, and the same single setting is maintained for the entire temperature range. The measured frequency is not calibrated off-chip after measurements. This temperature coefficient is the lowest among the reported sub-μW on-chip oscillators shown in Table 1. The lowest temperature coefficient measured among the five samples is 14.7 ppm/ºC from Chip D, where the first and second order temperature dependencies are cancelled, and the remaining higher order temperature dependencies are exhibited. The clock frequency shows an average supply voltage sensitivity of 0.75%/V in the range of 1.2-1.8 V for the five chips measured, as shown in Figure 3.12. A typical supply voltage is 1.3V. In this fabrication process, nominal V_T devices and high V_T devices in Figure 3.9 are 1.8V devices and 3.3V IO devices, respectively. All transistors in the amplifier, the amplifier bias generator in Figure 3.8, current references, switches for the switched capacitor resistor are 3.3V IO high V_T devices. A digital controller that generates current chopping signals and switch control signals from VCO outputs is composed of only 1.8V nominal V_T devices. 1.8V devices are designed to work up to 1.8V, but for testing, in the short-term, the circuit operated up to 3V without reliability issues. We used up to 3V to more extensively verify the circuit techniques for supply sensitivity. The long-term stability (Allan deviation) is less than 7 ppm for an integration time of longer than 12 seconds, as shown in Figure 3.13. This long-term stability is the second best among the prior state-of-the-art sub-μW on-chip oscillators.

It is important to analyze how each error sources contribute to absolute frequency inaccuracy. From the measurement results, a typical commercial temperature range of 70°C results in an inaccuracy equivalent to 479.5mV supply voltage change for the proposed oscillator. Hence, with a typical supply voltage of 1.3V of this oscillator, the line sensitivity is less critical than the temperature coefficient, as 40% of supply voltage fluctuation is not usually expected. The temperature coefficient and line sensitivity can be directly compared with each other because both measure instant frequency change due to short-term changes. Allan deviation measures stability due to noise processes rather than environmental effects. Given that, to reduce the line sensitivity further to the level of Allan deviation, embedding a linear regulator can be one option at the expense of some power overhead and actually, we adopted a linear regulator for system integration as described in Section V. Alternatively, temperature and voltage sensitivity can be further reduced using more extensive on-chip trimming as previous works such as [24] executed.

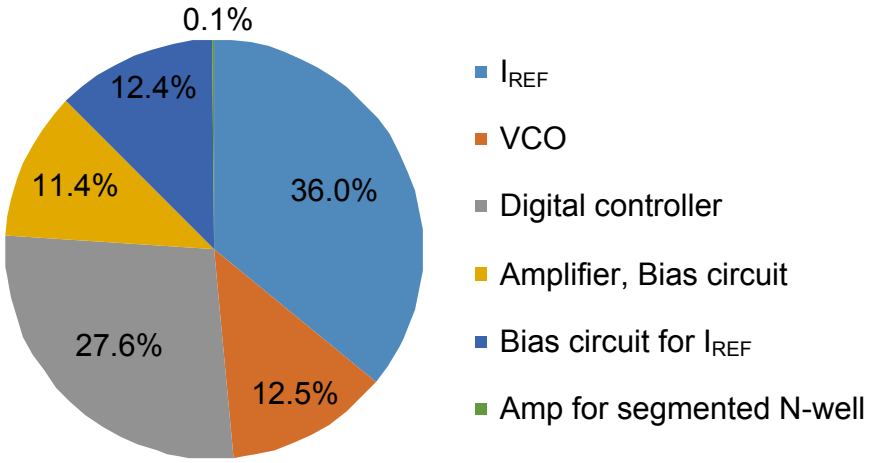


Figure 3.14 Breakdown of power consumption.

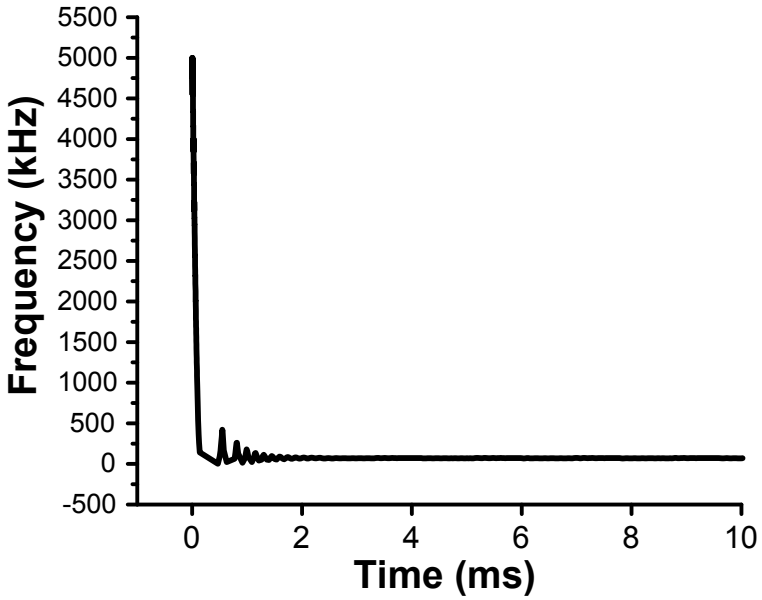


Figure 3.15 Measured start-up response of the proposed oscillator.

The design consumes 110 nW at room temperature, yielding the second lowest energy consumed per cycle, 1.56 pJ/cycle, among the previous works listed in Table 1. Power consumption for each part of the oscillator is described in Figure 3.14. A digital controller in the pie chart generates switch control signals and current chopping control signals from VCO outputs. A start-up response is measured as shown in Figure 3.15. From this figure, frequency overshoots

multiple times before it settles. The frequency locked loop is under-damped from the measurement and the start-up latency is less than 2.5msec. This latency can be shortened by increasing the frequency locked loop bandwidth and a damping ratio of the loop, but it is only allowed as long as the loop stays stable.

3.5 System Integration

The proposed RFLO is integrated in a single-chip radio system for wireless sensor nodes [22] demonstrating its capability to serve as an on-chip clock for a radio baseband controller. A substantial level of frequency accuracy is required for the radio baseband timer since the baseband controller determines the bit rate which needs to match the bit rate of the paired transceiver so that the data modulation/demodulation does not fall out of synchronization. The accuracy of the baseband timer therefore directly impacts the length of the data packet that can be transmitted. This radio implements pulse position modulation (PPM) with pulse position resolution (T_{PR}) of 4μs and separation between bits (T_{SP}) of 128μs. For M-ary PPM, a symbol length (T_{SYM}) is $M \times T_{PR} + T_{SP}$. A jitter for the Nth symbol position is $N \times T_{SYM}$ multiplied by Allan deviation ($\sigma_{y(N \times TSYM)}$) at integration time of $N \times T_{SYM}$. This jitter should be less than T_{PR} as derived in Equation (12).

$$\frac{T_{PR}}{NT_{SYM}} < \sigma_{y(NT_{SYM})} \tag{12}$$

With higher N or longer T_{SYM}, this condition becomes harder to satisfy as the Allan deviation improves only sub-linearly with increasing integration time. The radio system requires a packet of 192bits, and it is able to communicate with T_{SYM} of 136μs, but bit errors occur with longer T_{SYM}. This corresponds to the proposed calculation as T_{PR}/NT_{SYM} is 153.2ppm and Allan deviation at

integration time of 26.1ms from Figure 3.13 is in the range of 150 – 200ppm. For more conventional M-PPM modulation, T_{PR} is $T_{SYM}/2$ and T_{SP} is 0, simplifying Equation (12) to Equation (13).

$$\frac{1}{2N} < \sigma_{y(NT_{SYM})} \tag{13}$$

In this prototype, a FPGA-based demodulation code can tolerate a center frequency drift of ±2,000ppm (±500 Hz). Between wakeup periods, if temperature changes by ±58.3ºC or the supply voltage fluctuates by ±267mV, communication fails. To compare this work with previous generations, the oscillator with similar power consumption [27] can operate in the same system within ±51.9ºC temperature variation and ±200mV supply voltage change. Allan deviation near integration time of 26.1ms (estimated from their figure) is similar to this work. However the long term Allan deviation is 2.9× higher than this work, lowering N or T_{SYM} by the same ratio.

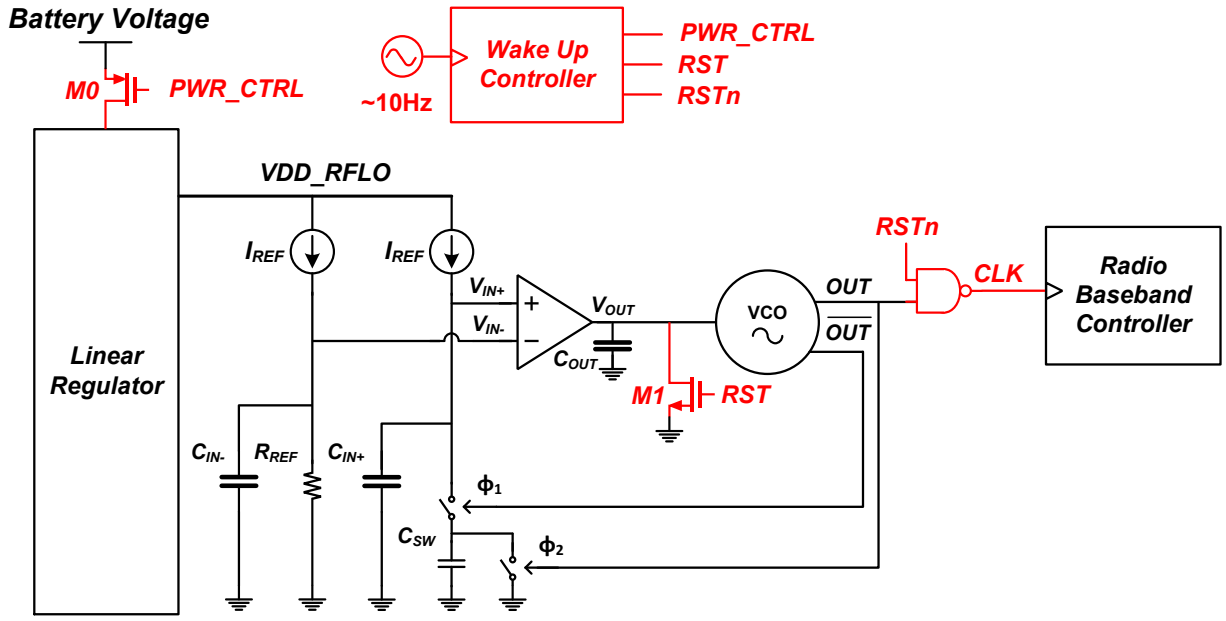


Figure 3.16 An RFLO combined with a wake-up controller to function as a clock source for radio baseband controller.

Unlike conventional radio systems that adopt a crystal oscillator as a clock source, this work is fully integrated, reducing the volume of a millimeter-scale system. An important component required to integrate an RFLO within a wireless sensor node is a wake-up controller, as the sensor node periodically sleeps and the wake-up timer will enable the RFLO during active periods only. The key functions of the wake-up controller are minimizing RFLO power consumption during the system sleep mode and ensuring that upon wake-up the RFLO clock feeds into the system only after it is stabilized. An RFLO combined with a wake-up controller is shown in Figure 3.16. During the system's sleep mode, RFLO leakage current is constrained to 20 pA (simulated value) by M0, a high threshold voltage PMOS header. The VCO control voltage (V_{OUT}) is clamped to ground by M1, a high threshold voltage NMOS.

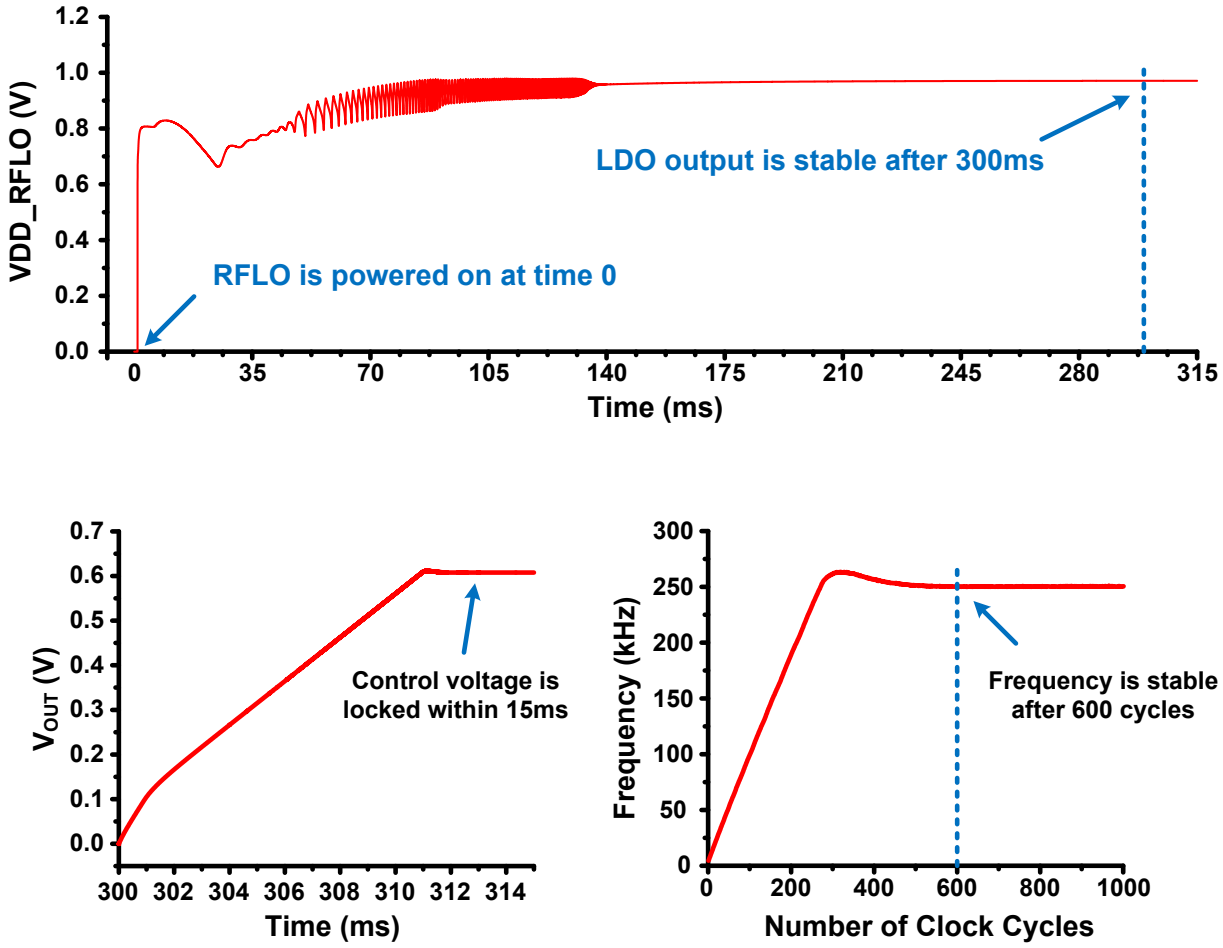


Figure 3.17 Simulation results of RFLO supply voltage (VDD_RFLO), VCO control voltage (VOUT), and RFLO frequency during a wake-up period.

At the beginning of wake-up mode, a linear regulator is powered and generates a supply voltage (*VDD_RFLO*) for the RFLO from the battery voltage. This takes approximately 300 ms, mostly due to the slow stabilization of the voltage reference inside the linear regulator; the reference is not explicitly shown in the figure as it is not the main focus of interest. Using a linear regulator improves frequency stability with respect to supply voltage and jitter performance at the cost of power overhead. A sub-100-pA watchdog timer with a 10-Hz frequency is implemented to measure the *VDD_RFLO* stabilization time. After *VDD_RFLO* is stabilized, the wake-up

controller toggles *RST* to low, causing the VCO to start oscillation after which the VCO frequency converges to the target value by the frequency locked loop. Before the frequency stabilizes at its target frequency, the clock is isolated from the radio system with a NAND gate. Simulation verified that a maximum of 600 cycles is required to stabilize the RFLO. An internal counter connected to the RFLO counts to 600, after which the RFLO clock is fed into the baseband controller of the radio system. Figure 3.17 shows simulated results of *VDD_RFLO*, V_{OUT}, and RFLO frequency during a wake up period. In order to meet the required frequency of the radio system, C_{SW} is reduced by 3.5 times, yielding 250 kHz clock frequency. A lower *VDD_RFLO* of 1V is used as this is the voltage level the radio system could provide. Simulated power consumption of the oscillator is 110nW.

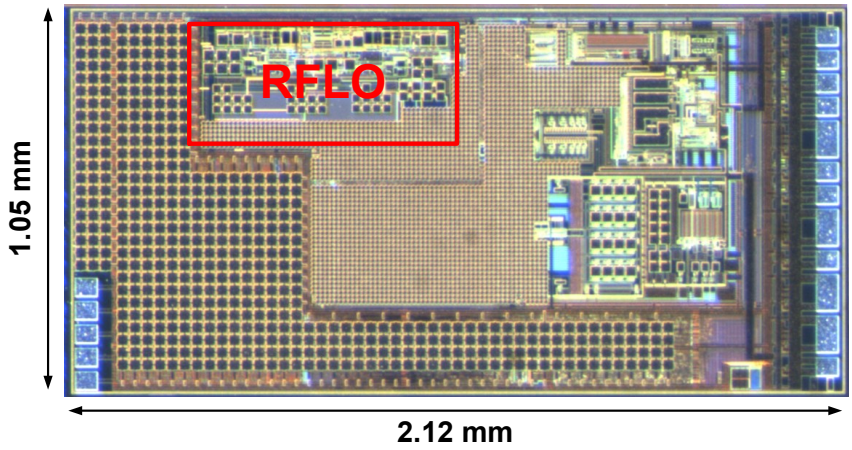


Figure 3.18 Die photo of a radio system integrated with a RFLO.

For the 300 ms required for *VDD_RFLO* stabilization, the entire system shown in Figure 3.16 consumes 35 nJ, while during the additional 600 cycles needed for RFLO stabilization the entire system consumes an additional 1.6 nJ, which is only 7.4% of 1 bit transmission. This 36.6 nJ of total energy consumption during the stabilization period corresponds to less than 2 equivalent

transmitted bits of wake-up overhead as a transmission energy consumption is 21.7 nJ/bit. However, reducing the stabilization time for a resistive frequency locked oscillator is still an important issue for systems that periodically sleep and wakeup, and it can be an interesting future work. This radio system was fabricated in 0.18 μm CMOS technology, and a die photo is shown in Figure 3.18.

3.6 Conclusion

An RFLO is introduced in this work. The proposed topology removes the comparator present in traditional RC relaxation oscillators, which is one of the main sources of temperature instability. Instead, an ultra-low power amplifier forms a frequency locking loop with a switched-capacitor circuit to generate a temperature-compensated clock signal. This oscillator produces a 70.4kHz clock with an average temperature coefficient of 34.3 ppm/ºC in the −40ºC to 80ºC range, an average supply voltage sensitivity of 0.75%/V in the 1.2 V to 1.8 V range for five samples, and long-term stability of less than 7 ppm after an integration time of 12 seconds while consuming 110 nW at room temperature. By avoiding external components, this oscillator targets fully integrated system-on-chip designs, and a radio transmitter system integrated with the oscillator is implemented and characterized.

CHAPTER 4

A Resonant Current-mode Wireless Power Receiver and Battery Charger with -32dBm Sensitivity for Implantable Systems

4.1 Introduction

Continuous health monitoring has become feasible, in part due to miniature implantable sensor systems such as [37]–[40]. Battery recharging capability is essential for such implantable systems because changing a system battery may incur a surgery, making implantable systems less attractive. For this purpose, wireless power transfer is a popular option since it is non-invasive. However, there are two main challenges. First, strict safety regulations of power exposure on human tissue limit the available incident power at the receiver coil. The specific absorption rate (SAR) limit set by the Federal Communication Commission (FCC) is 4W/kg, and standards setting organizations typically use 1/10 of this value. In addition to tissue heating issues, non-thermal effects such as altered cell membrane permeability or central nervous system effects can be caused by exposures less than $10mW/cm^2$ [41]. Secondly, implanted systems favor small coils for better biocompatibility and reduced invasiveness. For example, a glucose sensor [38] employs a contact lens form factor with a diameter of 1cm and a neural recording circuit [39] adopts a receiving

power coil with a diameter of 2cm. The small size of the receiver coil, combined with low incident power, reduces the received power at the coil, making it difficult to obtain sufficient power for implanted devices. This points to the need for high power efficiency transfer techniques, especially at very low received power levels.

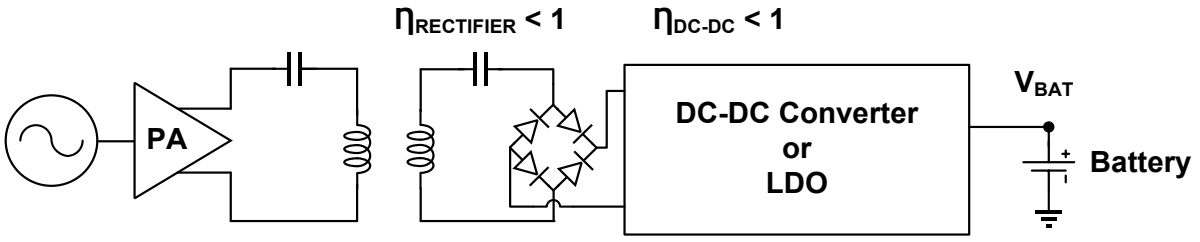


Figure 4.1 A block diagram of conventional voltage-mode wireless power transfer system.

As illustrated in Figure 4.1, most conventional wireless power receivers are composed of a rectifier for AC-DC conversion, followed by a DC-DC converter or linear regulator to generate an accurate voltage to safely charge a battery. In this voltage-mode approach, the input power (P_{IN}) at the receiver coil must be high enough to overcome the rectifier threshold voltage ($V_{TH,RECT}$), which is set by twice the diode built-in voltage in addition to an input voltage of a DC-DC converter or a linear regulator ($V_{IN,DC\text{-}DC}$). Any input power resulting in a voltage less than this cannot be harvested, limiting the minimum harvestable input power ($P_{IN,MIN}$). To address the rectifier threshold voltage issue, transistors with very low threshold voltage can be used as a diode. However, this generally increases the reverse diode current, and often requires additional fabrication steps. Active rectifiers composed of transistors and control circuitry can reduce the diode drop, as used in a previous work [42]. However, the operating frequency of [42] is 13.56MHz, which is 271.2× faster than this work, requiring a very high bandwidth of the control circuitry to generate accurate switching timing, and thus, consuming substantial power. Even with ideal diodes, the receiver LC tank peak voltage must exceed $V_{IN,DC\text{-}DC}$ to harvest. Also, the charging voltage

needs to be regulated to ensure battery safety, using a DC-DC converter or a linear regulator, which further reduces power efficiency.

Wireless power receivers can be categorized into two types: coil-based near-field receivers and antenna-based far-field receivers. Coil-based power receivers have relatively high power efficiency, but $P_{IN,\ MIN}$ ranges from 100's of μW to W [43]–[45]. These systems transmit and receive high power, and thus target high power applications including wireless cellular phone charging rather than ultra-low power implantable device charging. Far-field RF power receivers report lower $P_{IN,MIN}$ of several μW [46], [47], but power efficiency is comparatively low with power efficiency of 15% at 10μW [48]. Most recently, rectifier-antennas co-design methodology [44], [45] achieved sensitivity of −30.7dBm and −34.5dBm with a rectifier output voltage of 1V and 1.6V, respectively.

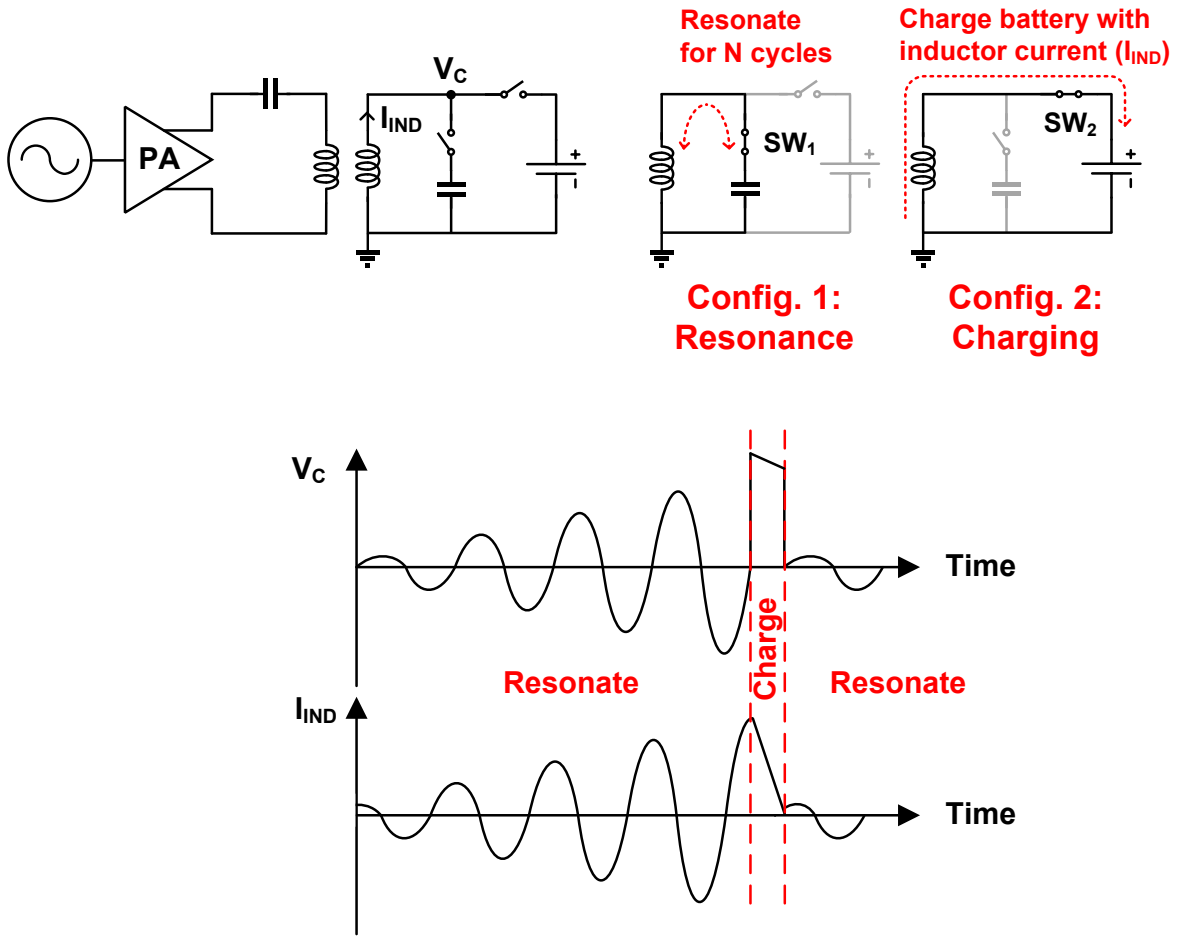


Figure 4.2 A block diagram of the proposed resonant current mode wireless power transfer system (top) and its conceptual waveforms at a resonance mode and a charging mode (bottom).

This paper is an extension of [7] and introduces a resonant current-mode approach that avoids rectification and voltage regulation. Instead, this method places a capacitor in parallel with a receiver coil to form an LC tank, and then resonates the LC tank for *multiple cycles* to accumulate energy (config. 1). It then transfers this energy to the battery in a boost-converter fashion (config. 2) as shown in Fig. 2. This method has three advantages. First, it improves $P_{IN,MIN}$, as it is no longer limited by $V_{TH,RECT} + V_{IN,DC\text{-}DC}$. Secondly, resonating an LC tank for multiple cycles can optimally balance different types of losses, reducing $P_{IN,MIN}$. In contrast, a non-resonant power receiver [51] employing current-mode charging could not collect power across multiple cycles, which limited

its power efficiency at low power levels and resulted in a relatively large $P_{IN,MIN}$ of 7.8μW. It should be noted that a tradeoff exists to implement resonance. The non-resonant receiver [51] does not require an off-chip capacitor for tuning a resonant frequency, and thus, can have a smaller system form factor. Also power efficiency of [51] is less sensitive to operating frequency variation. Thirdly, because the proposed method directly charges a battery with inductor current, it adopts the advantage of typical current mode charging, which does not require voltage regulation during most of the battery charging phase. Voltage-mode charging demands accurate output voltage to safely charge the battery. Given a process-dependent $V_{TH,RECT}$, a DC-DC converter requires wide input range, wide conversion ratio, and input voltage detection. Removing a voltage regulation eliminates power efficiency loss derived at this step. However, at the final phase of battery charging, typically constant voltage method is preferred as it guarantees safe and accurate full charging. To fully exploit these advantages, a maximum efficiency tracker is designed to optimize key parameters including the number of resonant cycles (N_{RESO}), bias current of a zero crossing detector (I_{BIAS}), and frequency of a V_{BAT} detector (F_{DET}) across a range of input power.

This paper is organized as follows. Section II describes the operating principles of the resonant current-mode wireless power receiver and battery charger, and analytically compares $P_{IN,MIN}$ of the proposed method with that of conventional voltage-mode charging approaches. Section III describes circuit implementations of each block. Section IV analyzes different types of energy losses and power efficiency, and Section V describes the measurement results. Finally, Section VI concludes the paper.

4.2 Resonant Current-Mode Charging

A. Operating Principles

A simplified diagram and conceptual waveforms of the proposed wireless power transfer system are shown in Figure 4.2. A wireless power transmitter, described in the left side, is composed of a sinusoidal signal generator, a power amplifier, an inductor, and a capacitor. A power amplifier amplifies a sine wave generated by the signal generator, and the amplified signal drives the LC tank. On the right side, the proposed wireless power receiver and battery charger are shown. The receiver part has a receiver coil, a parallel capacitor, two switches, a battery, and control circuitry. Resonant frequencies of LC tanks in both sides are tuned to the sine wave frequency of 50kHz.

This method has two modes: resonance and charging. In a resonance mode, switch 1 is closed and switch 2 is open, and thus, the receiver coil is connected to a parallel capacitor (C_{RX}) and forms an LC tank. As the receiver collects power, V_C amplitude continuously increases across resonant cycles and asymptotically approaches its final value as shown in the bottom of Fig. 2. When V_C is 0V and rising, all energy in the LC tank is stored in a receiver coil as $E_L=LI_{IND}^2/2$ where I_{IND} is inductor current. A zero crossing detector detects this condition and a digital counter counts the number of resonant cycles. When the count reaches a predetermined value, control circuitry switches the circuit to a charging mode. In this mode, switch 1 is open and switch 2 is closed, which disconnects the receiver coil from C_{RX} and connects it directly to the battery. At this point the energy stored in the inductor charges the battery like a boost converter. As a result, V_C instantly rises to the battery voltage (V_{BAT}) plus $I_{IND}\times R_{SW2}$, and then decreases as E_L is transferred to the battery. R_{SW2} is the on-resistance of switch 2. Energy transfer is complete when current flowing through switch 2 becomes zero. This condition is sensed by detecting when V_C equals V_{BAT}. When this condition is met, the circuit switches back to resonance mode. The proposed receiver charges a battery by continually repeating this routine.

B. ***Analysis of Minimum Harvestable Input Power***

This sub-section compares the minimum harvestable input power ($P_{IN,MIN}$) of the proposed resonant current-mode method and conventional voltage-mode method. The analysis starts by calculating the amplitude of V_C. When V_C is saturated, all energy received per cycle is dissipated in the LC tank at each cycle. The saturated voltage amplitude of V_C ($V_{C,SAT}$) and the saturated current amplitude of I_{IND} ($I_{IND,SAT}$) are given in (1)-(2). Here Q is the quality factor of the LC tank. T_{Cycle} is one period of the received sine wave.

$$E_{Stored\ in\ LC} = \frac{Q}{2\pi}E_{Loss/cycle} = \frac{Q}{2\pi}E_{Received/cycle} = \frac{Q}{2\pi}P_{IN}T_{Cycle} = \frac{LP_{IN}}{R_{IND}} = \frac{1}{2}C_{RX}{V_{C,SAT}}^2 = \frac{1}{2}L{I_{IND,SAT}}^2 \quad (1)$$

$$V_{C,SAT} = \sqrt{\frac{2LP_{IN}}{R_{IND}C_{RX}}},\ I_{IND,SAT} = \sqrt{\frac{2P_{IN}}{R_{IND}}} \quad (2)$$

For the conventional rectifier and DC-DC converter structure, as derived in (3)-(4), $P_{IN,MIN}$ is the power such that the resulting V_C equals $V_{TH,RECT} + V_{IN,DC\text{-}DC}$ [51]. The lowest $V_{IN,DC\text{-}DC}$ found from the literature ranges from 0.12V to 0.15V [52], [53].

$$V_{C,SAT,MIN} = \sqrt{\frac{2LP_{IN,MIN}}{R_{IND}C_{RX}}} = V_{TH,RECT} + V_{IN,DC-DC} \quad (3)$$

$$P_{IN,MIN} = \frac{R_{IND}C_{RX}(V_{TH,RECT} + V_{IN,DC-DC})^2}{2L} \quad (4)$$

However, in this proposed current-mode charging, $P_{IN,MIN}$ can be lower than given by (4). If the energy stored in the receiver coil at the end of a resonance mode ($E_{LC,RES}$) can overcome the conduction losses from coil ESR (R_{IND}) and R_{SW2}, the switching loss for mode transitions between a resonance and a charging mode, and the energy overhead of control circuitry, the receiver can harvest power, as described in (5):

$$E_{LC,RES,MIN} = E_{Conduction} + E_{Switching} + E_{control} \quad (5)$$

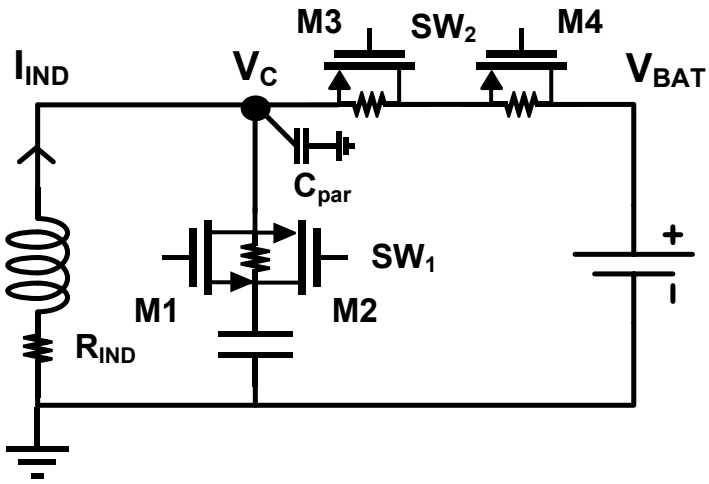


Figure 4.3 A schematic of the proposed wireless power receiver with parasitic resistors and capacitors.

To concisely compare two charging methods, we assume that the number of resonant cycle is large enough so that V_C and I_{IND} are saturated in the analysis of (6)-(11). More detailed analysis on the number of resonant cycle is given in Section IV. The circuit in Figure 4.3 is used for this analysis. Saturated V_C and I_{IND} in current mode ($V_{C,SAT,CM}$ and $I_{IND,SAT,CM}$, respectively) are smaller than those of a voltage-mode case for the same input power level, since on-resistance of the switch 1 (R_{SW1}) adds conduction loss in the LC tank, as derived in (6). As the parasitic resistance of a capacitor (R_{CAP}) is insignificant compared to R_{IND}, R_{CAP} is not included in the analysis.

$$V_{C,SAT,CM} = \sqrt{\frac{2LP_{IN}}{(R_{IND}+R_{SW1})C_{RX}}},\ I_{IND,SAT,CM} = \sqrt{\frac{2P_{IN}}{R_{IND}+R_{SW1}}} \quad (6)$$

$E_{Conduction}$ in (5) is the energy I_{IND} dissipates through R_{IND} and R_{SW2} in the charging mode. I_{IND} starts at $I_{IND,SAT,CM}$ and reduces to zero as inductor energy transfers to the battery. Because this charging time (T_{ch}) is very short compared to the resonant period formed by receiver coil inductance and

battery capacitance, the inductor current curve can be approximated as linear. As a result, $E_{Conduction}$ can be expressed as (7) below.

$$\begin{aligned} E_{Conduction} &= \int_{t=0}^{Tch} I_{IND}(t)^2 \, (R_{IND} + R_{SW2}) dt \\ &= \int_{t=0}^{Tch} \{ \sqrt{\frac{2P_{IN}}{R_{IND} + R_{SW1}}} (1 - \frac{t}{T_{ch}}) \}^2 \, (R_{IND} + R_{SW2}) dt \\ &= \frac{2P_{IN} T_{ch} (R_{IND} + R_{SW2})}{3(R_{IND} + R_{SW1})} \end{aligned} \quad (7)$$

Equation (5) can be expanded as (8), and solving it gives the minimum harvestable input power in the current-mode charging ($P_{IN,MIN,CM}$) shown in (9). Equation (10) describes switching energy loss for one charging event.

$$E_{LC,RES,MIN} = \frac{LP_{IN,MIN,CM}}{R_{IND} + R_{SW1}} = \frac{2P_{IN,MIN,CM} T_{ch} (R_{IND} + R_{SW2})}{3(R_{IND} + R_{SW1})} + \sum C_i V_i^2 + E_{Control} \quad (8)$$

$$P_{IN,MIN,CM} = \frac{3(\sum C_i V_i^2 + E_{Control})(R_{IND} + R_{SW1})}{3L - 2T_{ch}(R_{IND} + R_{SW2})} \quad (9)$$

$$\sum C_i V_i^2 = \left(C_{G,M1} + C_{G,M3}\right) {V_{1.2V}}^2 + \left(C_{G,M2} + C_{G,M4}\right) {V_{BAT}}^2 + C_{par}(V_{BAT} + I_{IND} R_{SW2})^2 \quad (10)$$

From the above equations, the minimum harvestable input power in the proposed approach is clearly no longer related to rectifier threshold voltage and DC-DC converter input voltage. By careful choices of switches and inductor along with low power control circuit design, this method can overcome the power sensitivity limits of conventional voltage-mode charging.

4.3 Circuit Implementation

A. Power switches

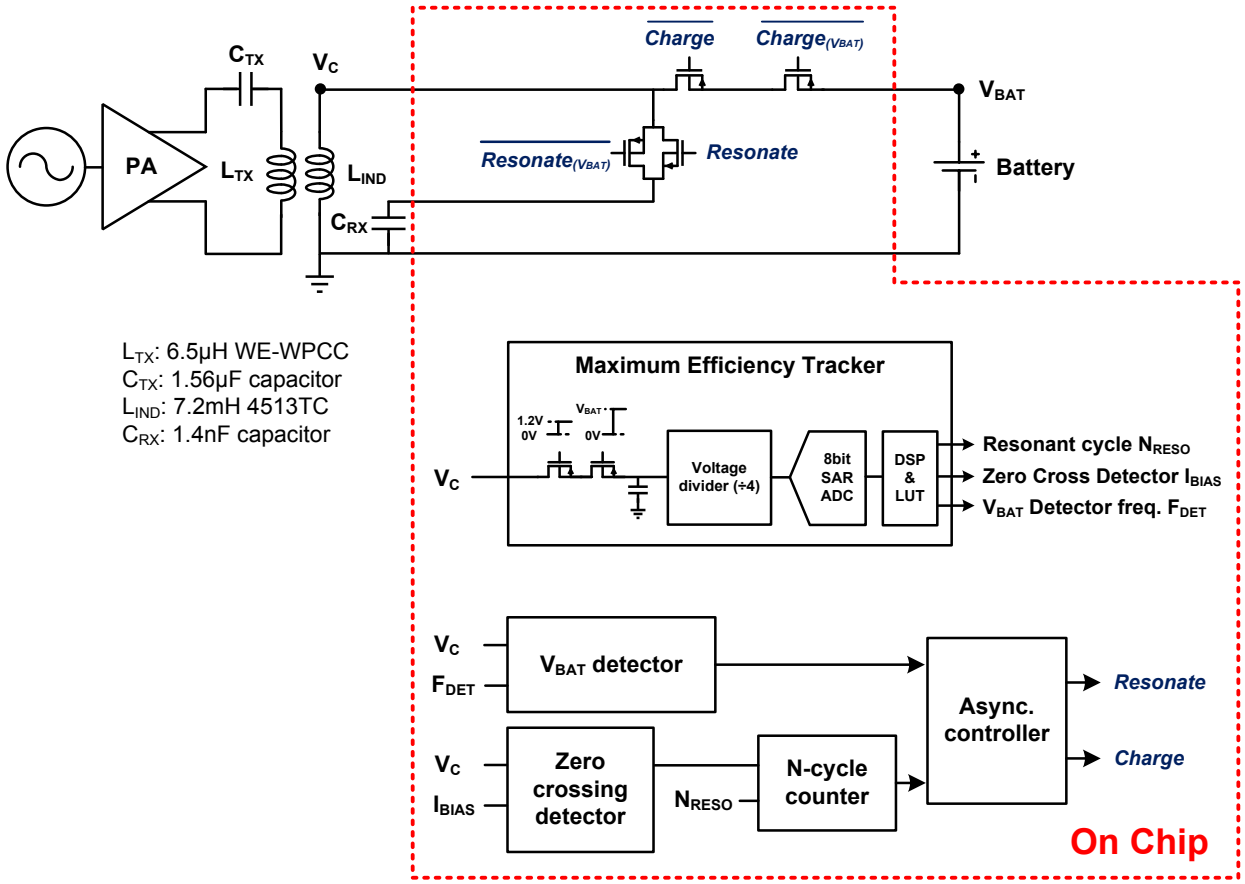


Figure 4.4 A system diagram of the proposed wireless power transfer system including block diagrams of control circuitry.

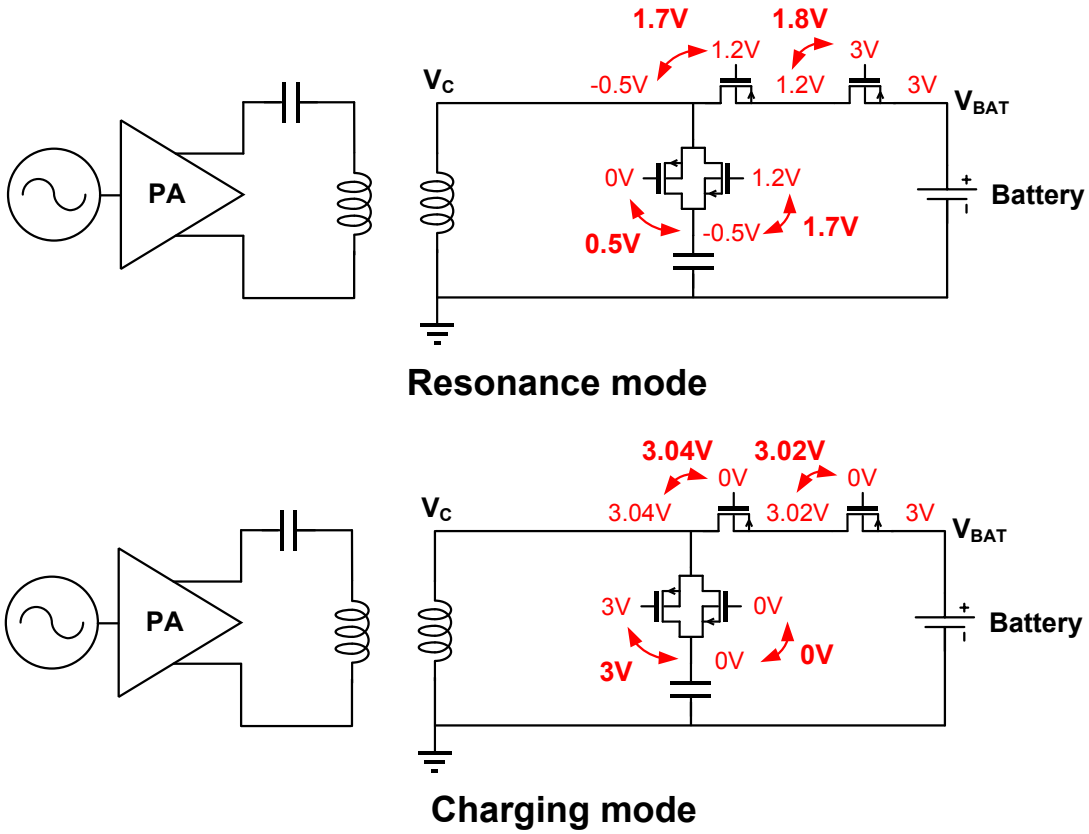


Figure 4.5 A proposed receiver at two modes with notation of voltage drops across oxides.

Figure 4.4 shows the proposed system diagram. The power transmitter is drawn on the left side and the proposed power receiver and battery charger is shown at right. All parts inside the red dotted line are integrated on-chip. All four power transistors are 3.3V I/O devices. Switch 1 in Figure 4.2 is implemented with one PMOS transistor and one NMOS transistor connected in parallel. The PMOS transistor and the NMOS transistor are controlled by V_{BAT}-level and 1.2V-level signals, respectively. Using a 1.2V signal prevents large source/drain to gate voltages that can cause oxide breakdown when V_C swings to a large negative voltage level during resonance mode. Possible voltages across oxides at resonance mode are shown in Figure 4.5. Switch 2 in Figure 4.2 consists of two PMOS transistors in series. The left and right PMOS transistors are controlled by 1.2V and V_{BAT}-level signals, respectively, for the same reason. Oxide voltages in

charging mode are also shown in Figure 4.5. In this implementation, the 1.2V supply is externally provided and power consumption from this source is included in efficiency calculations.

Power transistor sizing should consider the tradeoff between switching and conduction losses. As transistor width increases, switching losses increase with higher capacitance while conduction losses decrease with lower on-resistance. Transistor lengths are set to minimum values. Two prototypes of this work are fabricated with different switch sizes. The first version has M1 of 70μm/350nm, M2 of 140μm/300nm, and M3/M4 are both 35μm/300nm. The second version increases M1-M4 widths by 2× compared to the first version. Switching losses are constant across different input power while conduction losses increase with input power. As a result reducing switch sizes lowers $P_{IN,MIN}$ and increasing switch sizes enhances efficiency at high P_{IN}. Measured results are introduced in Section V to support these expected trends.

B. Zero Crossing Detector

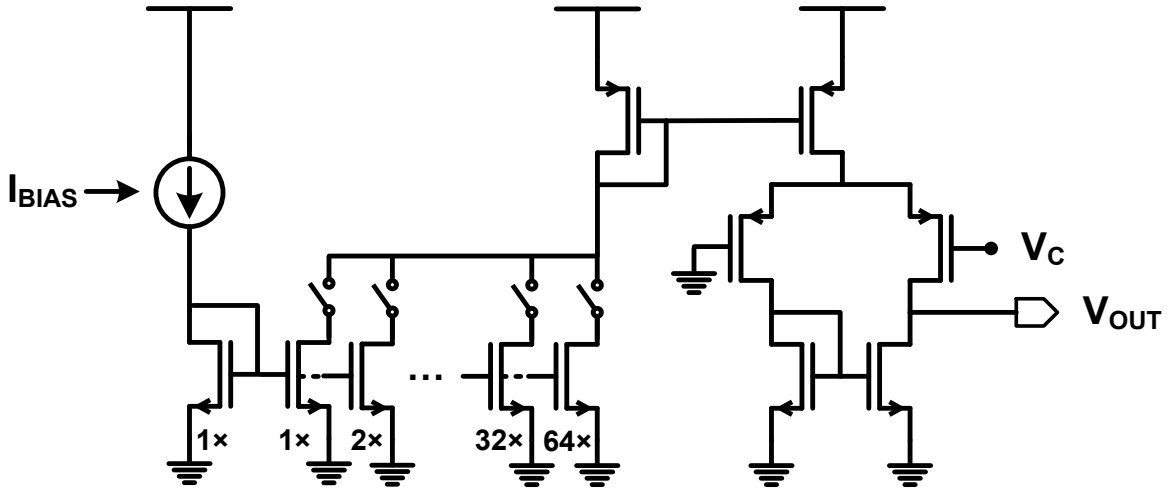


Figure 4.6 A schematic of a zero crossing detector.

After the LC tank builds up enough energy to harvest, the circuit should switch from resonance mode to charging mode. This transition should take place when the inductor stores all the LC tank

energy and the parallel capacitor has no energy, and thus V_C is zero. A zero crossing detector detects this condition. It is implemented with a standard one-stage amplifier with differential inputs and a single-ended output as shown in Figure 4.6. The two inputs are connected to ground and V_C. PMOS transistors are used as input pairs as the amplifier needs to operate near 0V. Bias current (I_{BIAS}) is programmable from 3nA to 200nA by a maximum efficiency tracker, and the current mirror multiplies the current by 10×.

Limited bandwidth of the zero crossing detector results in a switching voltage error, V_{err}. As a result, C_{RX} has a remaining energy of $C_{RX}V_{err}^2/2$ rather than the ideal 0J at the end of resonance mode. This energy stored in C_{RX} is wasted by charge redistribution in charging mode and conduction loss in the next resonance mode. Increasing I_{BIAS} reduces this loss by improving zero crossing detector bandwidth, but increases its power consumption. V_{err} is derived in (11), assuming that V_{err} and t_{err} are small. Here α_1 is the ratio of V_C amplitude at ith resonant cycle ($V_{C,cycle\ i}$) to V_C amplitude at saturation ($V_{C,peak}$), as shown in (11). The input wave frequency is f_{IN}. Because increasing I_{BIAS} also directly increases amplifier bandwidth, t_{err} can be expressed as in (12). To maintain constant V_{err} and energy loss caused by V_{err} with increasing $V_{C,peak}$, I_{BIAS} should increase linearly as well. A maximum efficiency tracker therefore measures $V_{C,peak}$ and sets I_{BIAS} accordingly.

$$V_{err} = \alpha_1 V_{C,peak} \sin(2\pi f_{IN} t_{err}) \approx \alpha_1 V_{C,peak} \times 2\pi f_{IN} t_{err} = \frac{2\pi\alpha_1 V_{C,peak} f_{IN}}{\alpha_2 I_{BIAS}} \tag{11}$$

$$\alpha_1 = \frac{V_{C,cycle\ i}}{V_{C,peak}}, \qquad \frac{1}{t_{err}} = \alpha_2 I_{BIAS} \tag{12}$$

C. VBAT Detector

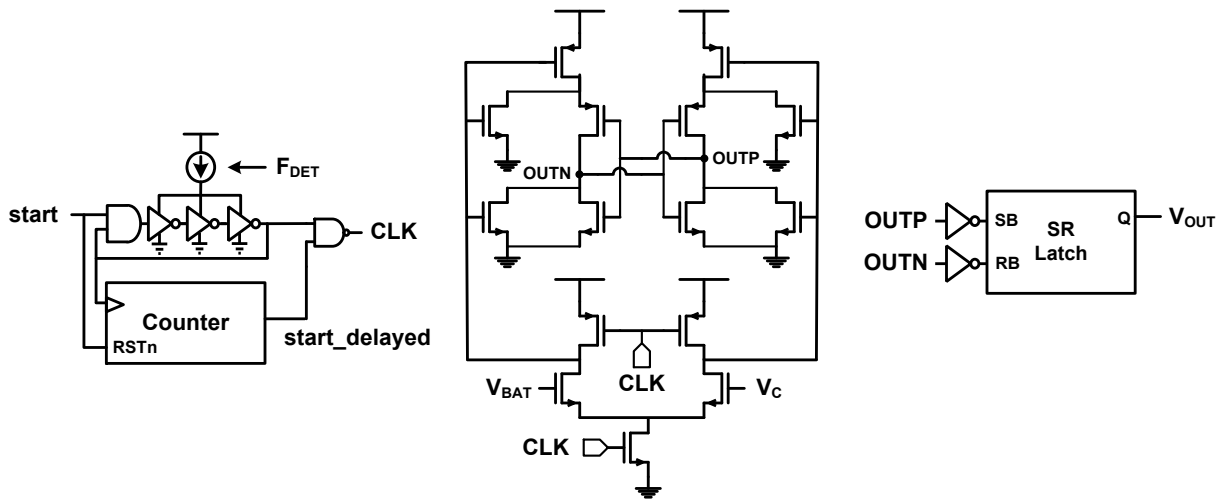


Figure 4.7 A schematic of a V_{BAT} detector with its clock generator at left.

In charging mode, I_{IND} flows through Switch 2 creating a voltage drop of $I_{IND} \times R_{SW2}$ across the switch. A V_{BAT} detector detects when the voltage drop decreases to zero, as the energy transfer is complete at that point. The detector is a dynamic comparator based on [54], as shown in Figure 4.7. Its inputs are connected to V_{BAT} and V_C, and the comparator outputs are captured by an SR latch. A clock signal for the comparator is provided by an internal current-starved ring oscillator, of which the frequency is controlled by the maximum efficiency tracker. Charging time is defined as (13) and this time is approximated by the maximum efficiency tracker as follows. L and C_{RX} are fixed, and the peak voltage of V_C at resonance mode ($V_{C,peak}$) is detected with a maximum efficiency tracker. V_{BAT} does not change widely as battery operating voltages are fixed in certain ranges such as 2.25V to 3V [55] or 0.9V to 1.6V [56] for commercial lithium batteries. With estimation of charging time, V_{BAT} detector is turned on with some delay after the energy transfer starts, in order to reduce power consumption. The delay can be set by a counter and a clock gating NAND gate as in Figure 4.7.

$$T_{ch} = \frac{LI_{IND,peak}}{V_{BAT}} = \frac{L\sqrt{C_{RX}V_{C,peak}^2/L}}{V_{BAT}} = \frac{V_{C,peak}\sqrt{C_{RX}L}}{V_{BAT}} = \frac{\alpha_3}{F_{DET}} \quad (13)$$

A mistimed transition from charging mode to resonance mode leads to energy loss. When the mode is switched too early, inductor energy is not completely transferred to the battery. When the mode is switched too late, all inductor energy is transferred to the battery after which the battery begins discharge to the inductor. From (13), in order to maintain a relative timing resolution (α_3) of detector clock period to T_{ch}, a higher detector frequency (F_{DET}) is required at lower $V_{C,peak}$. A maximum efficiency tracker sets F_{DET} with $V_{C,peak}$ information to keep α_3 constant.

D. *Asynchronous Controller*

Transitions between the two modes are controlled by event-driven asynchronous logic to eliminate dynamic power during a given configuration. If implemented with synchronous logic, the clock speed is set by the fastest detection speed among trigger events, which is T_{ch}. This timeframe can be shorter than 1μs at low input power; to detect this transition with precise timing resolution (i.e., 1%), the controller clock frequency must be 100MHz, consuming several μW and making sub-μW harvesting impossible.

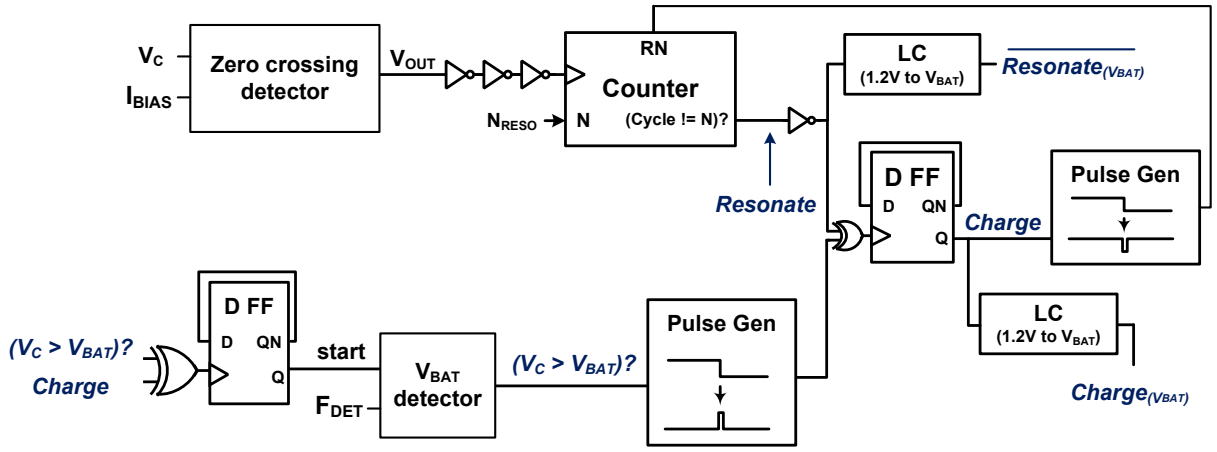


Figure 4.8 A block diagram of an asynchronous controller.

Figure 4.8 describes the asynchronous controller. The zero crossing detector converts the sinusoidal V_C into a rectangular signal, which serves as the clock for the following counter. The counter outputs a *Resonate* signal when the number of received rising edges reaches a predetermined value, N_{RESO}. N_{RESO} is provided by the maximum efficiency tracker. Level converters generate V_{BAT}-level signals from 1.2V-level signals. V_{BAT} detector outputs a logic 1 when V_C exceeds V_{BAT}. Pulse generators provide clock inputs to flip-flops as no external clock is available.

E. *Maximum Efficiency Tracker*

The proposed wireless power receiver and battery charger has three programmable system parameters that can maximize power efficiency across varying input power levels. Input power can vary when the transmitter power changes, or when TX/RX coil separation varies. If the resonant frequency deviates from the operating frequency, input power also changes. In this work, resonant frequency deviation of 100,000 ppm from the operating frequency of 50 kHz decreases the energy stored in C_{RX} at $V_{C,peak}$ by more than 4×. Although dynamically re-tuning the resonant

frequency can recover input power most directly, it requires additional capacitor array. Maximizing efficiency at a given input power can be an alternative solution, as adopted in this work.

The maximum efficiency tracker measures an input power level and set values for these parameters. During initial operation, the system stays in resonance mode and the amplitude of V_C increases. When V_C is saturated, its peak voltage is captured by a sample and hold circuit inside the maximum efficiency tracker. To find the phase where V_C is at its peak, an internal ring oscillator with a high frequency runs for one period of V_C, and a counter counts the number of oscillator cycles. When the count is at half of the number of oscillator cycles, V_C peaks. The sampled V_C peak voltage is then digitized with a standard 8-bit SAR ADC, and a simple on-chip signal processing block and look-up table sets the three parameters to maximize power efficiency: N_{RESO}, I_{BIAS}, and F_{DET}. From (11) and (13), I_{BIAS} is proportional to $V_{C,peak}$, and F_{DET} is inversely proportional to $V_{C,peak}$, so they can be easily calculated. N_{RESO} decreases with increasing $V_{C,peak}$, but the relationship is more complex and is analyzed in Section IV. The SAR ADC operates only once before the charging operation to detect input power level, and is then power gated. Power consumption of the SAR ADC, a power gating controller, and a clock generator is 13.7μW from simulation. It takes 0.85μs for 1 analog to digital conversion, and consumes 11.67pJ. When power gated, the block consumes 46.8pW.

4.4 Efficiency Analysis

This section analyzes the four kinds of energy losses present in this system: conduction losses and switching losses in each of the resonance and charging modes. The impact of the number of resonant cycles on power efficiency is also analyzed. First, conduction loss in resonance mode

($E_{L,CON,RES}$) is defined as the energy dissipated in R_{IND} and R_{SW1} as derived in (14). $I_{IND,rms,n}$ is a root mean square value of I_{IND} at the n^{th} cycle. N is the number of cycles in resonance mode.

$$\begin{aligned} E_{L,CON,RES} &= \sum_{i=1}^{N} {I_{IND,rms,n}}^2 (R_{IND} + R_{SW1}) T_{Cycle} \\ &= \sum_{i=1}^{N} 2\pi {I_{IND,rms,n}}^2 (R_{IND} + R_{SW1}) \sqrt{LC_{RX}} \end{aligned} \tag{14}$$

Total energy received during resonance mode is the sum of energy stored in the inductor and the conduction energy loss in the LC tank as derived in (15). Solving this equation for $I_{IND,rms,n}$ gives (16), and substituting this into (14) yields $E_{L,CON,RES}$ in (17). When N is small, this loss term increases rapidly with N with slope approaching $P_{IN}T_{Cycle}$. Also, $E_{L,CON,RES}$ is proportional to P_{IN}.

$$nP_{IN}T_{Cycle} = \frac{1}{2}L(\sqrt{2}I_{IND,rms,n})^2 + \sum_{i=1}^{n} 2\pi {I_{IND,rms,i}}^2 (R_{IND} + R_{SW1}) \sqrt{LC_{RX}} \tag{15}$$

$$I_{IND,rms,n} = \sqrt{\frac{2\pi\sqrt{LC_{RX}}P_{IN}}{L + 2\pi\sqrt{LC_{RX}}(R_{IND} + R_{SW1})} \sum_{i=1}^{n} \left\{ \frac{L}{L + 2\pi\sqrt{LC_{RX}}(R_{IND} + R_{SW1})} \right\}^{i-1}} \tag{16}$$

$$E_{L,CON,RES} = \frac{4\pi^2 (R_{IND} + R_{SW1}) LC_{RX} P_{IN}}{L + 2\pi\sqrt{LC_{RX}}(R_{IND} + R_{SW1})} \sum_{n=1}^{N} \sum_{i=1}^{n} \left\{ \frac{L}{L + 2\pi\sqrt{LC_{RX}}(R_{IND} + R_{SW1})} \right\}^{i-1} \tag{17}$$

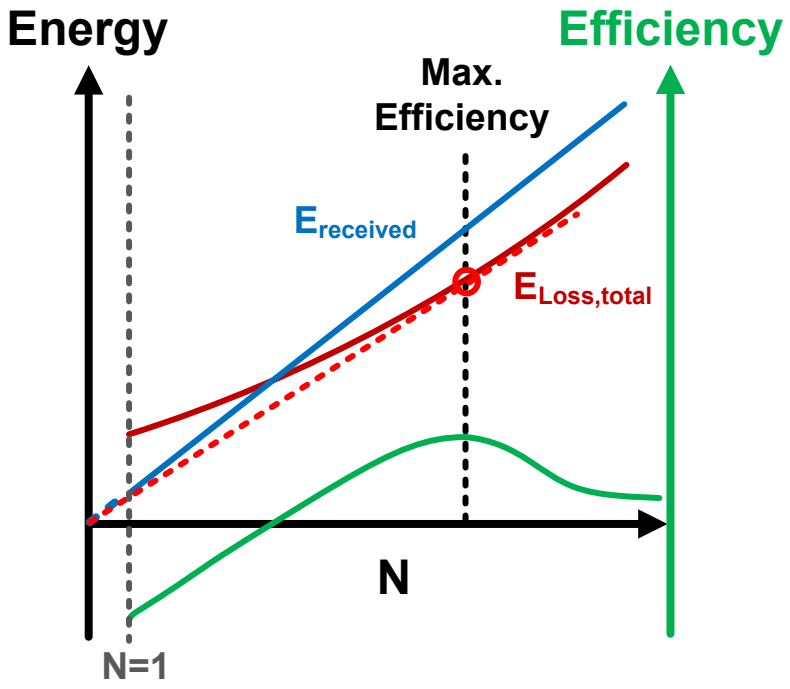


Figure 4.9 Conceptual graphs of a received energy, a total energy loss, and power efficiency with respect to number of resonant cycles (N).

Secondly, switching loss when moving from resonance mode to charging mode is $C_{gate,M2}V_{BAT}^2$. Here, only the transistors that draw energy from supply voltages are included. This switching loss is independent of N and P_{IN}, as mode switching happens only once per single charging event. Thirdly, conduction loss in charging mode ($E_{L,CON,CH}$) is the energy that I_{IND} dissipates through R_{SW2} and R_{IND}. $I_{IND,N}$ is the peak inductor current at the N^{th} cycle and is $\sqrt{2}$ × higher than $I_{IND,rms,N}$. $E_{L,CON,CH}$ is derived in (18), which is generalized version of (7); note that (7) only considers the case when N is large enough such that I_{IND} is saturated. $E_{L,CON,CH}$ increases with N, but also saturates when I_{IND} saturates. It increases with N faster than P_{IN} does, because it is cubically proportional to $I_{IND,N}$ while P_{IN} is proportional to the square of $I_{IND,N}$.

$$E_{L,CON,CH} = \int_{t=0}^{Tch} I_{IND}(t)^2 (R_{SW2} + R_{IND})dt = \int_{t=0}^{Tch} \{I_{IND,N}(1 - \frac{t}{T_{ch}})\}^2 (R_{SW2} + R_{IND})dt$$

$$= \frac{{I_{IND,N}}^2(R_{SW2} + R_{IND})}{3} T_{ch} = \frac{{I_{IND,N}}^2(R_{SW2} + R_{IND})LI_{IND,N}}{3V_{BAT}} \quad (18)$$

$$= \frac{{I_{IND,N}}^3(R_{SW2} + R_{IND})}{3}$$

Finally, switching loss when transitioning from charging mode to resonance mode is $(C_{gate,M1} + C_{gate,M3})V_{DD}^2 + (C_{gate,M4} + C_{par})V_{BAT}^2$. C_{par} is the parasitic capacitance at node V_C in Figure 4.3. This loss is independent of N and P_{IN}.

Resonating the LC tank more than 1 cycle during resonance mode improves power efficiency at low input power levels. This is highlighted by the fact that if the energy stored in an LC tank for one resonant cycle is less than the switching losses of Switch 1 and Switch 2, conduction loss of R_{SW1}, R_{SW2} and R_{IND}, and other control overhead, the system cannot charge the battery. However, if the LC tank resonates for additional cycles, the LC tank builds up sufficient energy to overcome these losses, enabling harvesting at the same (small) input power level. However, resonating for too many cycles can decrease power efficiency as $E_{L,CON,RES}$ grows with N at the same rate the LC tank energy does while $E_{L,CON,CH}$ increases with N more rapidly than LC tank energy. At the same time, loss due to charging events per unit time decreases as N increases. In this way a given input power exhibits a corresponding optimal N that balances the aforementioned losses. Increasing N is more beneficial for low P_{IN}, since at high P_{IN} the large I_{IND} results in high conduction loss, which limits gains from large N. Conceptual waveforms of the total energy loss, energy received, and power efficiency with respect to N are plotted in Figure 4.9. The N that maximizes power efficiency occurs where a straight line from N=0 touches the loss curve.

4.5 Measurement Results

Two versions of the proposed work are fabricated in 0.18μm standard CMOS technology with different sizes of power transistors, as mentioned in Section III. Measurement results of the first version are reported in [7]. The system includes a 7.2mH Coilcraft 4513TC receiver coil with Q-factor of 51 and 1.4nF off-chip capacitor. Average on-resistances of the parallel connection of M1 and M2 are 56 Ω for version 1 and 28 Ω for version 2. Parasitic resistance of C_{RX} is negligible. From Figure 4.2, config. 1, I_{IND} flows the loop formed by L_{IND} and C_{RX} in series. LC tank's Q-factors of version 1 and 2 are 15.4 and 19.1 from (19), respectively.

$$Q = \frac{1}{R_{IND} + R_{SW1}} \sqrt{\frac{L_{IND}}{C_{RX}}} \tag{19}$$

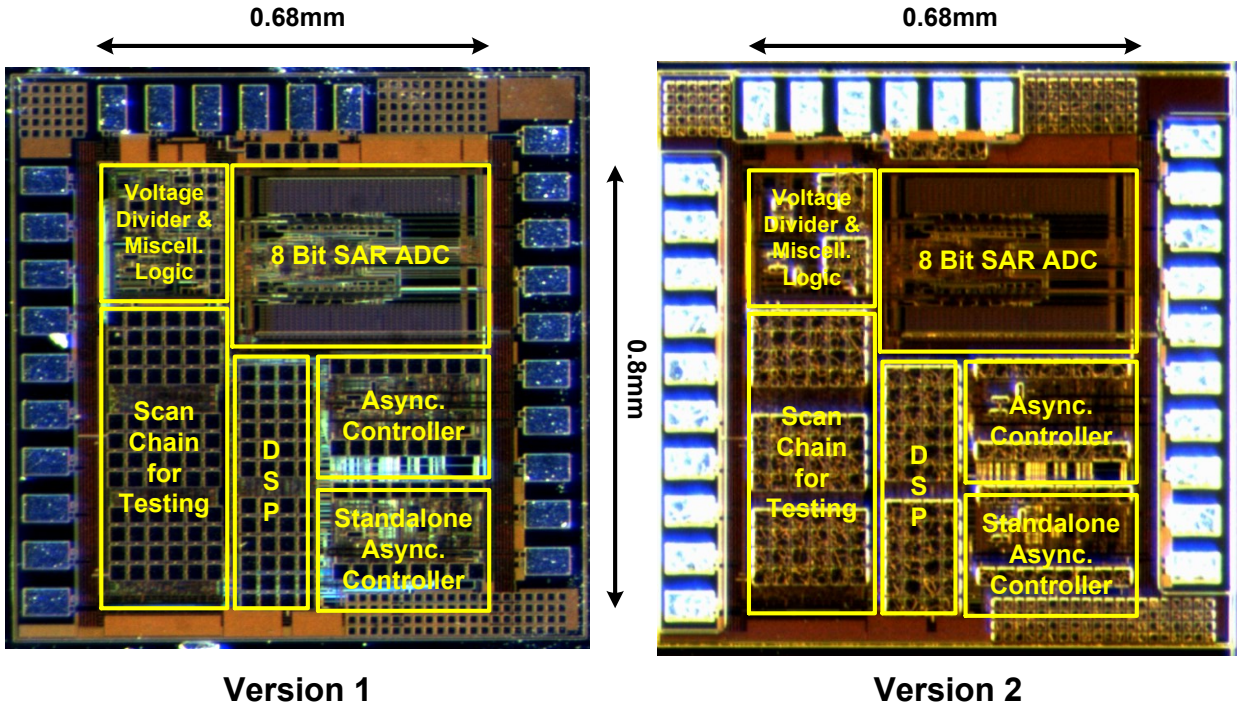


Figure 4.10 Microphotograph of two 0.18µm test chips (0.68 × 0.8mm^2 each).

The chip area is 0.544mm^2 for each version, as seen in Figure 4.10. The design is composed of an asynchronous controller, a maximum efficiency tracker (consisting of a 8-bit SAR ADC, a digital signal processor, and a voltage divider with miscellaneous logic gates), and a scan chain for testing. A standalone asynchronous controller is added for testing purposes. The receiver coil is 11.7mm × 3.5mm × 2.6mm, which is sufficiently small to be implanted in applications such as neural recorders and cochlear implants [39], [40].

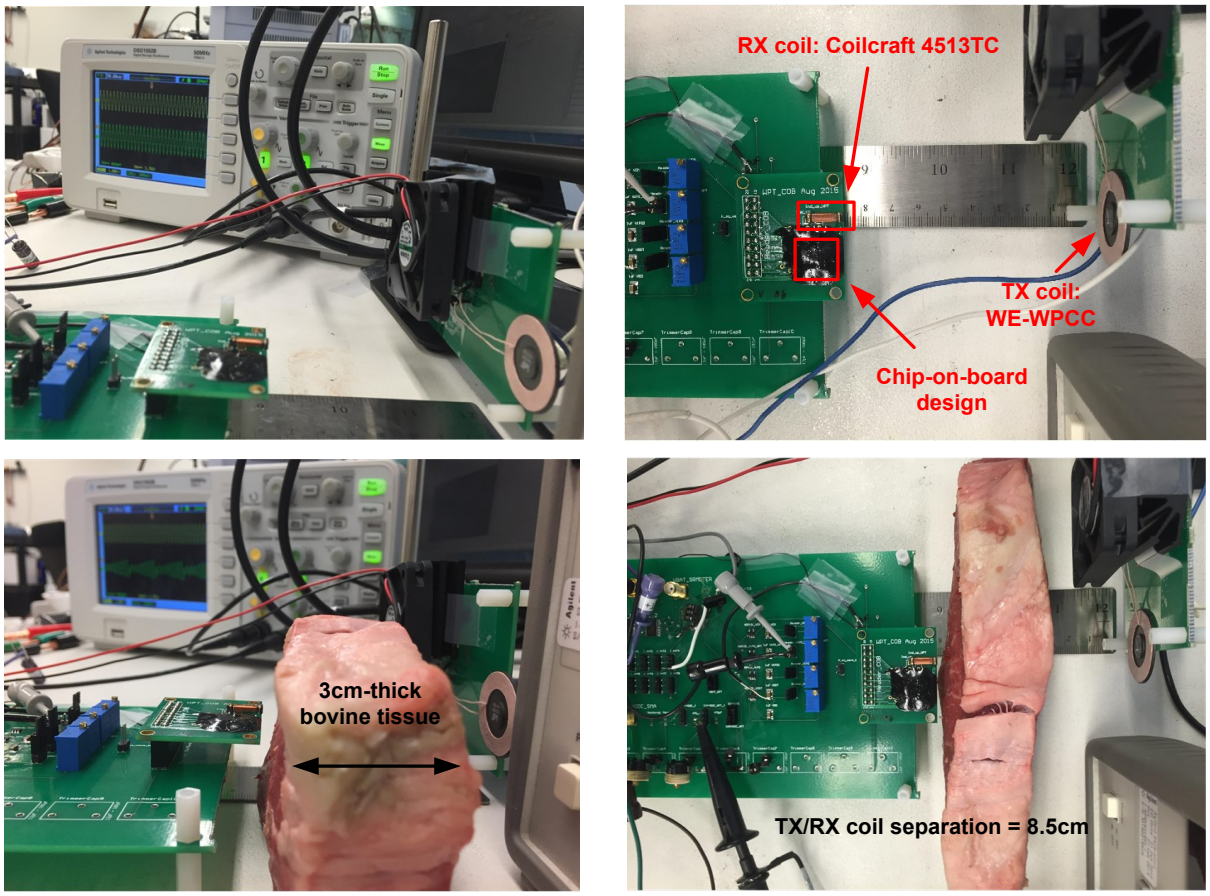


Figure 4.11 Measurement setups of the wireless power transfer system.

Figure 4.11 shows the testing setup. To minimize parasitic capacitance at the inductor node (C_{par}), chip-on-board packaging is used. A fabricated chip is wire-bonded and encapsulated in black epoxy. A WE-WPCC wireless power charging transmitter coil is chosen as a TX coil, and a ceramic off-chip capacitor forms a TX-side LC tank. The TX coil has inductance of 6.5µH. A board spacer and holder is used to accurately control TX/RX separation with 1mm resolution. Because this charging method injects current to a battery for a short time, this current cannot be captured accurately by equipment such as a sourcemeter. Instead, an off-chip capacitor (C_{OUT}) with known capacitance is connected to the output node, and the voltage change (ΔV_{OUT}) over a known time ($t_{measure}$) is measured. An Aluminum electrolytic capacitor is used. The capacitor has leakage current I_{Leak}, so this self-discharge rate is measured separately and calibrated out. A battery is also

measured to prove charging capability and is shown to be functional. However, because the voltage to charge capacity curve is not linear and varies over recharging cycles, the output power cannot be accurately measured using a battery load. The output power (P_{OUT}) is calculated from (20).

$$P_{OUT} = \int \left(C_{OUT} \frac{\Delta V_{OUT}}{t_{measure}} - I_{Leak} \right) V_{OUT}(t)\, dt \tag{20}$$

Measured minimum harvestable input power from version 1 (600nW) is 3.9× lower than [44], which exhibited the lowest harvestable input power at the time of publication, and 13× lower than [51], which uses the same size coil. A recently published receiver [49] achieves 1.69× lower sensitivity than this work using a self-oscillating technique and a receiver coil with higher Q-factor of 120, but a lower peak efficiency of 27.7%. Version 2 harvests at input power levels above 890nW. This version's larger switches reduce conduction losses but increase switching losses for mode transitions, and the fixed amount of switching losses has the strongest impact on minimum harvestable input power.

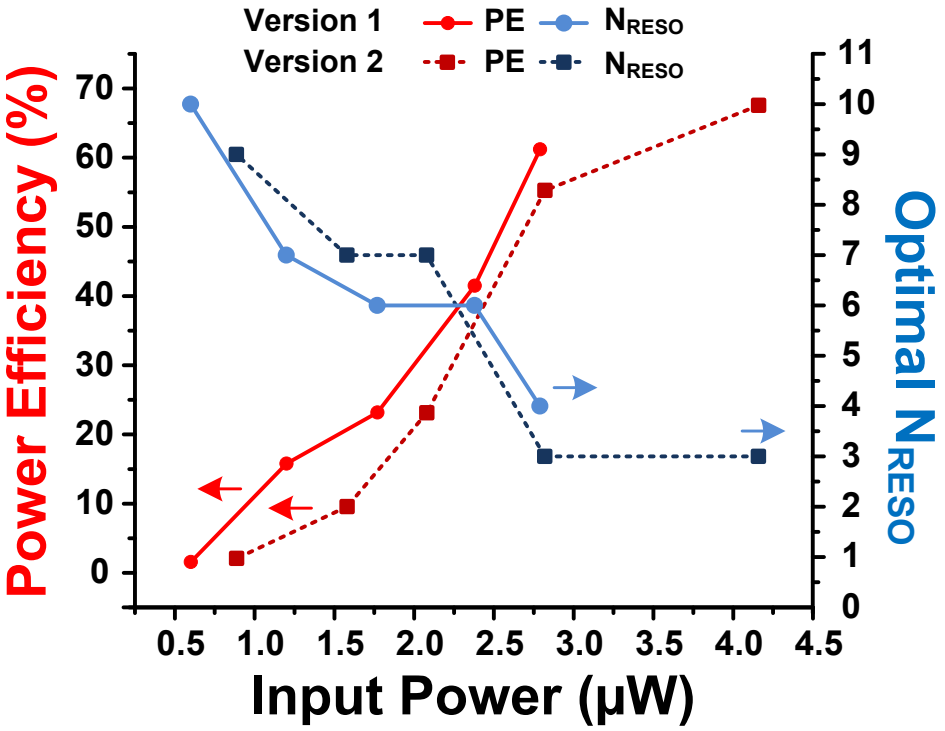


Figure 4.12 Measured power efficiencies and optimal N_{RESO}s of version 1 and 2 with respect to input power.

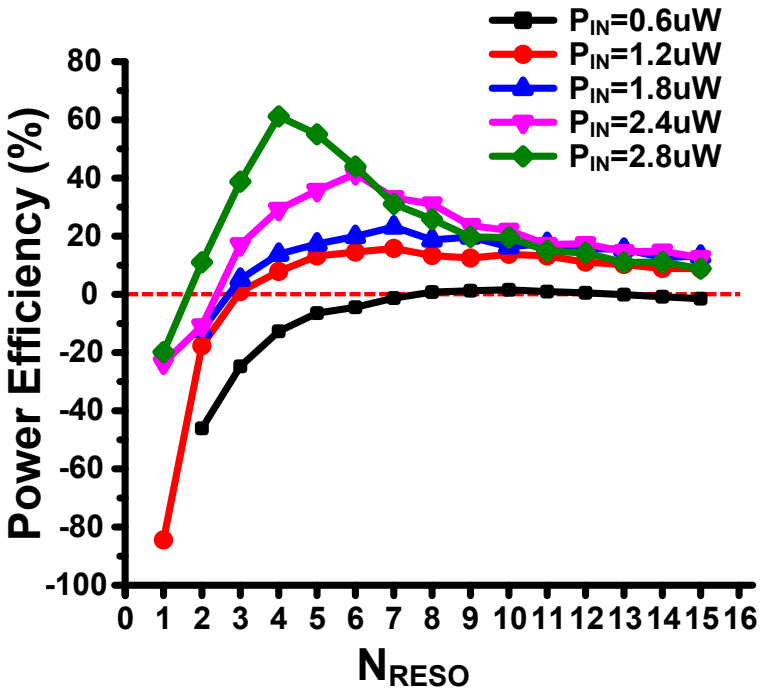


Figure 4.13 Measured power efficiency at different input power with respect to N_{RESO} of version 1.

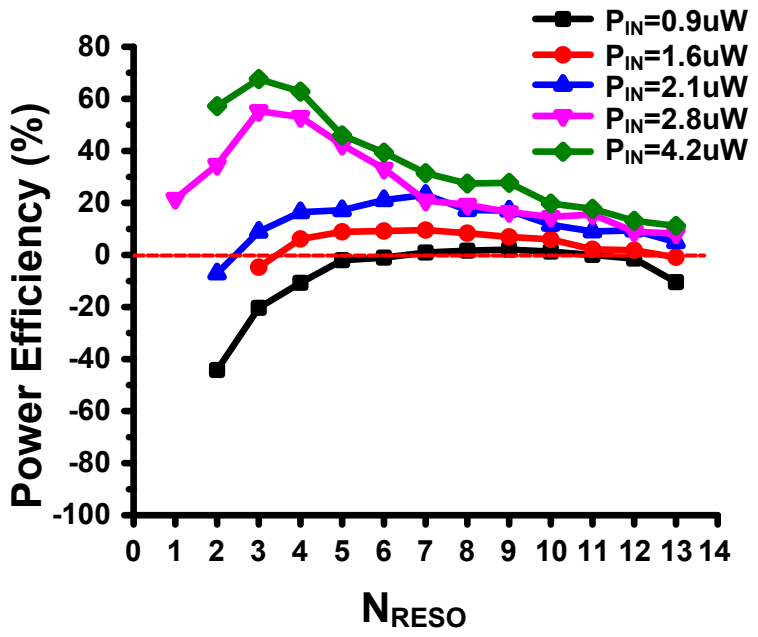


Figure 4.14 Measured power efficiency at different input power with respect to N_{RESO} of version 2.

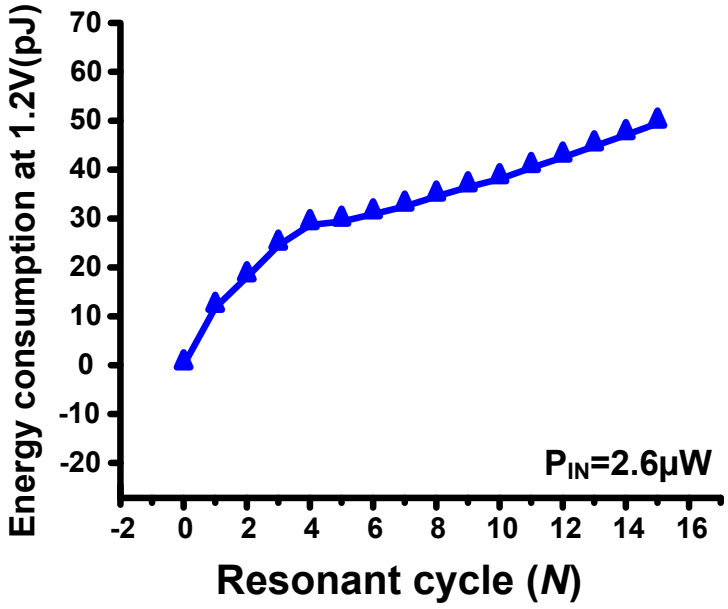


Figure 4.15 Measured energy consumption at external 1.2V supply with respect to N.

Power efficiency increases as input power increases, reaching 61.2% at P_{IN} = 2.8μW and 67.6% at P_{IN} = 4.2μW for versions 1 and 2, respectively (Figure 4.12). In version 1, 600nW P_{IN} is harvestable when N reaches 7 with optimal N of 10. For version 2, 890nW P_{IN} becomes harvestable when N exceeds 7 with an optimal N of 9. Optimal N decreases as P_{IN} increases. For version 1, at

P_{IN}=2.8μW, optimal N is 4, and for version 2, at P_{IN}=4.2μW, optimal N is 3. Figure 4.13 and Figure 4.14 show measured efficiencies at different input power levels with respect to N for version 1 and 2, respectively. The maximum allowable V_C amplitude without device breakdown issues can be found from a resonance mode in Figure 4.5. When 1.2V is used for M1 and M2 gates (3.3V IO transistor), maximum V_C amplitude is 2.1V, resulting in the maximum harvestable input power of 46μW. With a 20mW transmitter the maximum separation of TX/RX coils is 8.5cm in air. Identical performance is measured through 3cm of bovine tissue and 5.5cm air. This is expected since theoretically tissue absorbs negligible power at 50kHz. According to [57], theoretical power loss, $P=P_0e^{-2\alpha\sqrt{FD}}$, ($P_0$=incident power, $\alpha=2\times10^{-3}\text{sec}^{1/2}\text{m}^{-1}$ for muscle, F=50kHz, D=3cm) is less than 2.7%. This result supports our target application where an implantable system is charged by an external transmitter under the energy exposure limits of human tissue. Energy consumption from the external 1.2V supply voltage is measured with respect to N at input power of 2.6μW as shown in Figure 4.15. This energy is the energy consumption sum of a zero crossing detector, an asynchronous controller and power transistors M1 and M3. This work assumes an external 1.2V source and V_{BAT} to be greater than 1.2V. These assumptions lower $P_{IN,MIN}$. Some of the previous works [46]–[48] can start harvesting with no external sources, which can charge overly depleted batteries. Such an assumption runs counter to applications where a cold start is necessary, but for applications where transmitted power is limited, this work can start harvesting from a lower input power.

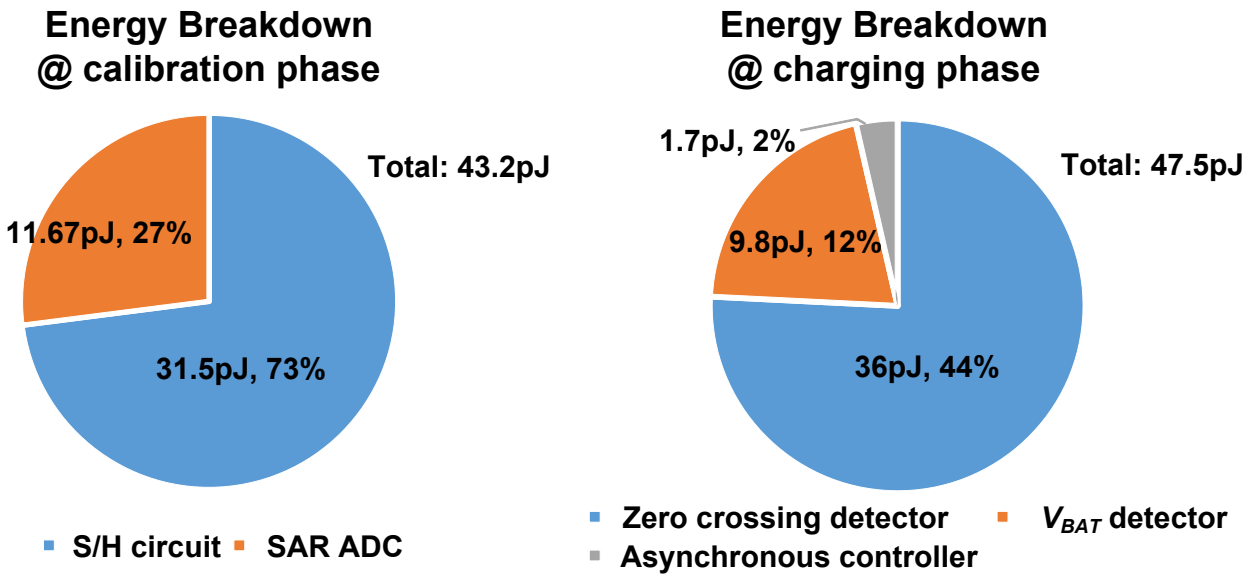


Figure 4.16 Energy breakdown at calibration and charging phase for version 2 with P_{IN}=4.2µW, and N=3 (simulated).

Energy breakdown of each block is discussed here. The following analysis is for the case of N=3, P_{IN}=4.2µW for version 2. System operation is divided into two phases: calibration and charging operation. In calibration mode, a maximum efficiency tracker is on and all other blocks do not operate. A sample and hold circuit including an amplifier consumes 31.5pJ for 0.1µs and the 8bit SAR ADC consumes 11.67pJ for 0.85µs. After calibration, the maximum efficiency tracker is power gated and the system switches to normal charging operation, where system energy consumption is divided into zero crossing detector energy, V_{BAT} detector energy, and asynchronous controller energy. The sum of all block's energy consumption for one charging event is 47.5pJ. The zero crossing detector, the V_{BAT} detector, and the asynchronous controller consume 36pJ, 9.8pJ, and 1.7pJ, respectively. Energy breakdowns in these two phases based on simulations are shown as pie charts in Figure 4.16.

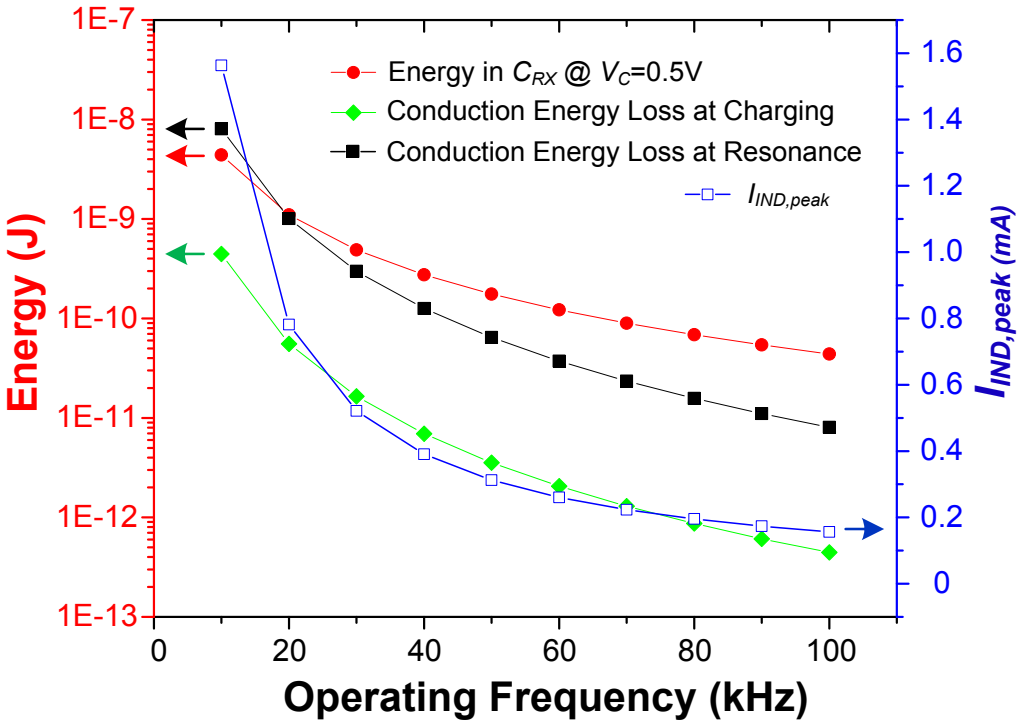


Figure 4.17 Stored energy in C_{RX}, energy losses, and $I_{IND,peak}$ with respect to operating frequency (simulated).

Increasing operating frequency with a given receiver coil requires reducing C_{RX}. This decreases the energy stored in C_{RX} at the same V_C, I_{IND} amplitude, and thus amount of conduction energy losses. Meanwhile switching energy loss per 1 charging event is fixed for given power switch sizes regardless of the operating frequency. These trends are analyzed in Figure 4.17. To concentrate on the effect of operating frequency, a few assumptions are applied for the analysis of Figure 4.17: switch sizes are the same as those of version 1 and N is large enough so that I_{IND} is saturated.

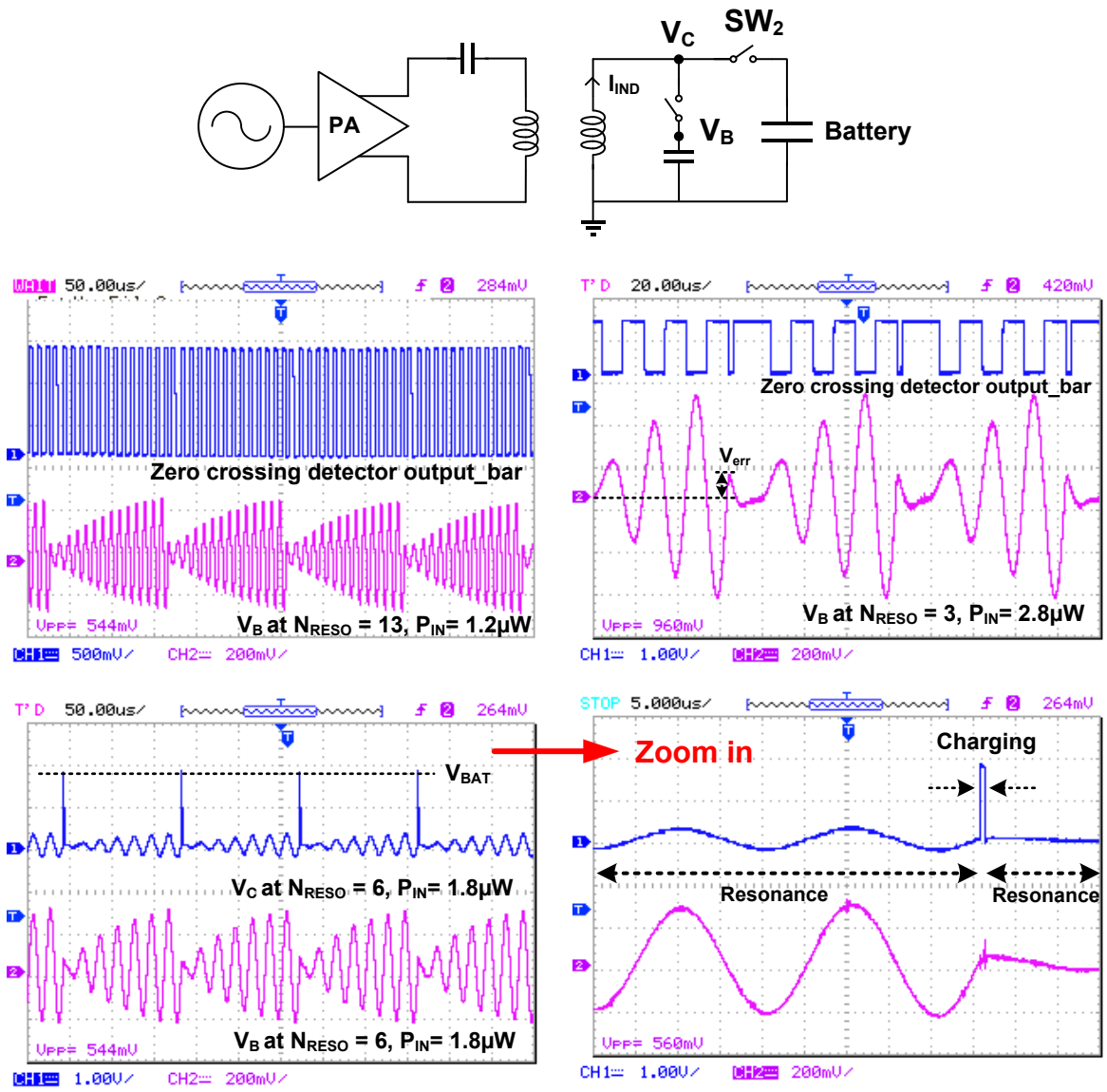


Figure 4.18 Measured waveforms of voltages at V_B and V_C, and inverted zero crossing detector output with oscilloscope.

Oscilloscope waveforms show the zero crossing detector output as blue lines and V_B as red lines. The top left figure shows V_B building up during resonance mode. At top right, V_{err} caused by finite bandwidth of the zero crossing detector is captured. In the bottom left, V_C is measured. In charging mode, V_C rises past V_{BAT} to allow charging and in resonance mode, it tracks V_B. A

zoomed-in waveform is captured at bottom right, clearly showing the behavior of V_C in charging and resonance modes.

Table 4.1 Performance Summary and Comparison Table

	This Work	[15] RFIC 2015	[11] VLSIC 2013	[12] JSSC 2008	[17] ESTPE 2015	[9] BCAS 2012	[13] MTT 2015	[8] ISSCC 2015	[10] ISSCC 2015
Technology (μm)	0.18	0.065	0.09	0.25	0.18	0.065	Off-chip	0.35	0.13
Chip area(mm2)	0.544	0.8	0.029	0.4	0.26	0.6	N/A	5.415	14.44
Frequency (MHz)	0.05	2,400	868	906	0.125	1,860	900 / 1,800 / 2,100 / 2,450	13.56	6.78
Min. Harvestable PIN (μW)	0.6	0.85	2.34	5.5	7.8*	200	N/A	N/A	N/A
Max. Receiver Efficiency @ PIN	67.7% @ 4.2μW	38% @ 5μW**	31.5% @ 31.6μW	60% @ 158μW	84% @ 660μW	31.9% @ 500μW	84% @ 3.8mW	92.5% @ 59.45mW	84.6% @ 7.09W
Pickup Coil Size	2.6×3.5×11.7mm3	1.33 cm2	20.9cm2	30cm2	2.6×3.5×11.7mm3	2mm× 2mm	10cm× 10cm	9.5mm diameter	N/A
Coil/ Antenna	Coil	Antenna	Antenna	Antenna	Coil	Antenna	Antenna	Coil	Coil
Measured distance @ TX Power	8.5cm @ 20mW	20m @ 4W	25 m @ 1.78W	15m @ 4W	7cm @ N/A	5cm @ 2W	50m @ 1.2mW/m2	1.8cm @ 50mW	6mm @ N/A
Charging method	Resonant current-mode	Voltage-mode	Voltage-mode	Voltage-mode	Current-mode	Voltage-mode	Voltage-mode	Voltage-Mode	Voltage-mode
Off-chip components in receivers	L (inductor), C (capacitor)	L	L	L	L	L	L, C	L	N/A

Table 1 summarizes performance of this work and compares to prior art. This work shows a sub-μW minimum harvestable input power and maximum power efficiency of 67.6% at >7.5× lower input power than state-of-the-art works. Measured distance between TX/RX coils is 8.5cm with the lowest TX power of 20mW among specified TX powers.

4.6 Conclusions

This paper proposes a resonant current-mode wireless power receiver and battery charger. The proposed prototype is fabricated in 0.18μm CMOS technology with area of 0.544mm^2. Unlike a conventional voltage-mode receiver that rectifies input wave and converts the rectifier output with a DC-DC converter or a linear regulator, this method directly charges a battery with inductor current. Furthermore, the LC tank resonates for multiple cycles to maximize its power efficiency by balancing switching and conduction losses. This work achieves a very low minimum harvestable input power of 600nW, and maximum efficiency greater than 60% at >7.5× lower input power than related work. Power transmission through bovine tissue is demonstrated to validate operation in implantable applications.

CHAPTER 5

A Pulse-Width Based Deep Learning Accelerator with In-memory Computing

5.1 Introduction

Several ASIC systems have been recently proposed in order to enhance energy efficiency, throughput, and latency of deep learning, including not only conventional digital circuit systems [58]–[59], but also analog circuit systems based on in-memory computing [60]–[62]. However, most of the existing systems can only process one layer, while layer to layer connections and interface are handled off-chip. Considering that significant portion of energy and latency are consumed in data movement, a comprehensive system implementing entire deep learning layers is greatly demanded.

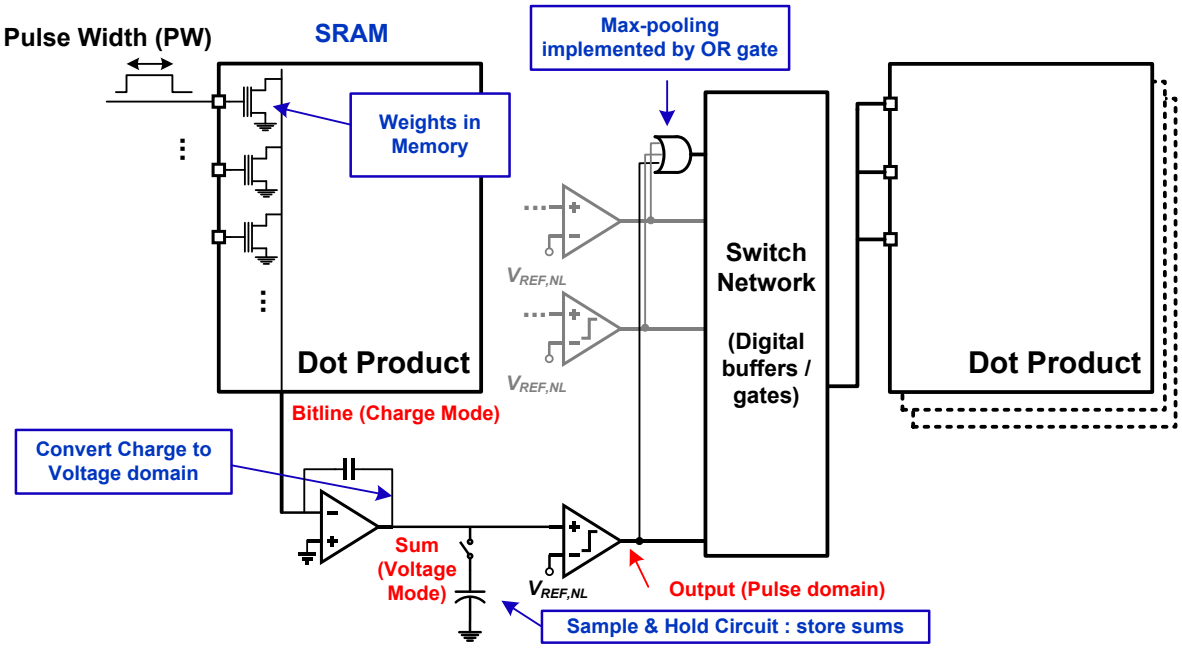


Figure 5.1 A layer of pulse-width based deep learning accelerator.

In this work, we propose an all layers all weights on-chip deep learning accelerator based on pulse width modulation as shown in Figure 5.1. Key challenges of a comprehensive deep learning system include layer to layer connections, energy and latency constraints from data movement, implementation of non-linear functions such as sigmoid and ReLU. Our proposed system modulates input images into pulse width and computes dot-product as charge integration of SRAM's read buffer currents over pulse widths. With this in-memory computing approach, energy and delay for weight retrieval can be removed. Dot-product outputs are stored as analog voltages, and converted into pulse width again by comparison to non-linear signals. Outputs modulated in pulse width enable layer to layer connection simple and fast, as standard digital buffers can easily drive pulse signals without information loss. Intermediate values during

inferences all remain in analog domain either as voltages or pulse widths, removing analog to digital conversion overhead and allowing an energy efficient high-throughput accelerator.

5.2 Operating Principles

A. Charge mode Dot-product

The most prevailing computation element in convolutional neural network is dot product. As described in (1), an activation function of neural network is composed of a number of multiplication and addition depending on filter sizes of applied neural network structures.

$$0 = \sum w_i x_i + b \tag{1}$$

This work utilizes the relationship of charge, current and time. Total accumulated charges (Q) from a current source (I) over a given time (t) is given by Q = It. If multiple current sources (I_i) are independently controlled by pulse-width modulated switches, so that the i^{th} current source turns on for t_i time, the accumulated charges from all current sources combined are described as (2) which is the same with (1) except for a bias value *b*.

$$0 = \sum I_i t_i \tag{2}$$

Figure 5.2 shows a structure of a conventional 8T SRAM cell with 6T memory cell and 2T read buffer. In-memory computation in analog domain utilizing SRAM cell has been introduced in prior works including [58]–[63]. This 8T structure decouples 6T data storage part from 2T read buffer so that in-memory computation does not affect read, write margin and data retention of the cell at the expense of area and cell complexity. However, it suffers from linearity degradation for dot product calculation. The bit-line for reading (RBL) is pre-charged to VDD at the beginning of in-memory calculation. As a pulse input comes in and turns on the NMOS transistor, RBL will

start to discharge and its voltage is reduced. The drain source voltage of read transistor changes over time resulting in unequal I_i over time. This issues has been one of the most critical issue limiting resolution of analog domain dot product.

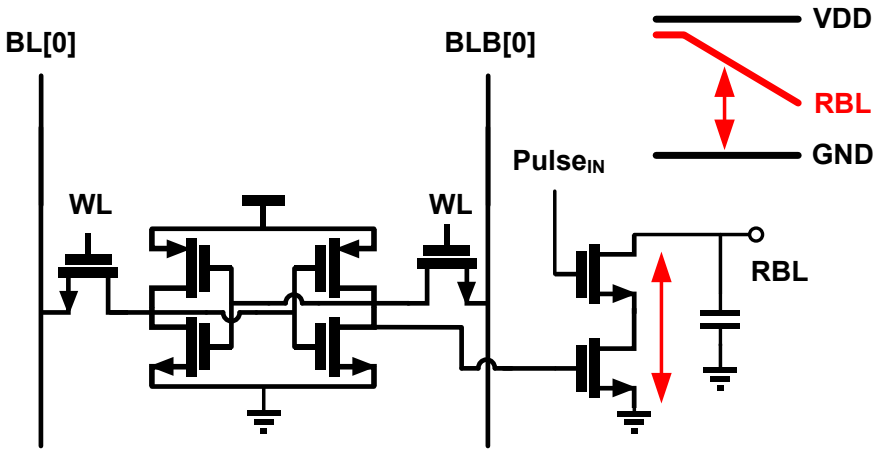


Figure 5.2 Linearity issue of charge mode dot product with conventional 8T SRAM cell

To improve linearity, this work proposes a new structure shown in Figure 5.3. It adds one cascode buffer transistor to protect drain voltage fluctuation of read transistors and an active integrator with a feedback capacitor. The integrator holds its negative input voltage to be the same with its positive input voltage. When pulse input comes in, charges are accumulated at the output of the integrator and the drain voltage of the read buffers remain constant, eliminating the problematic reduction of drain voltage over time. The linearity improvement becomes more pronounced in advanced technologies because transistor channel length modulation is more severe in such technologies. Simulation results in Figure 5.4 show INL errors in 7bit resolution with and without analog techniques for dot product output voltages (voltages at integrator output nodes) with respect to input pulse width. With analog techniques including a cascode buffer and an integrator, dot product linearity versus input pulse width improves by 21×.

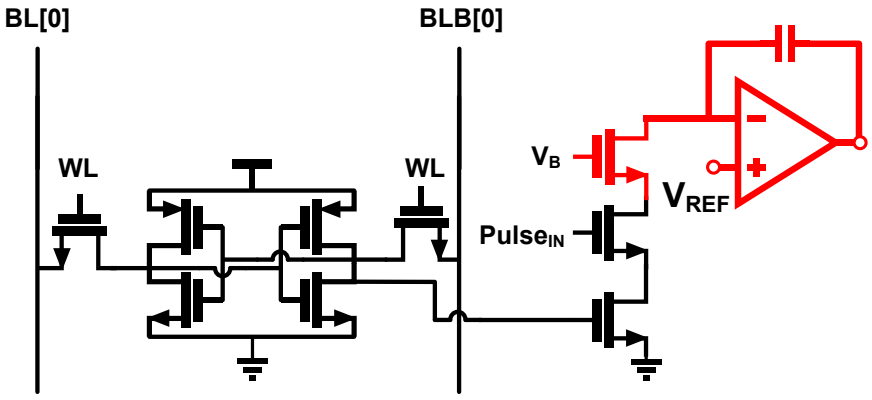
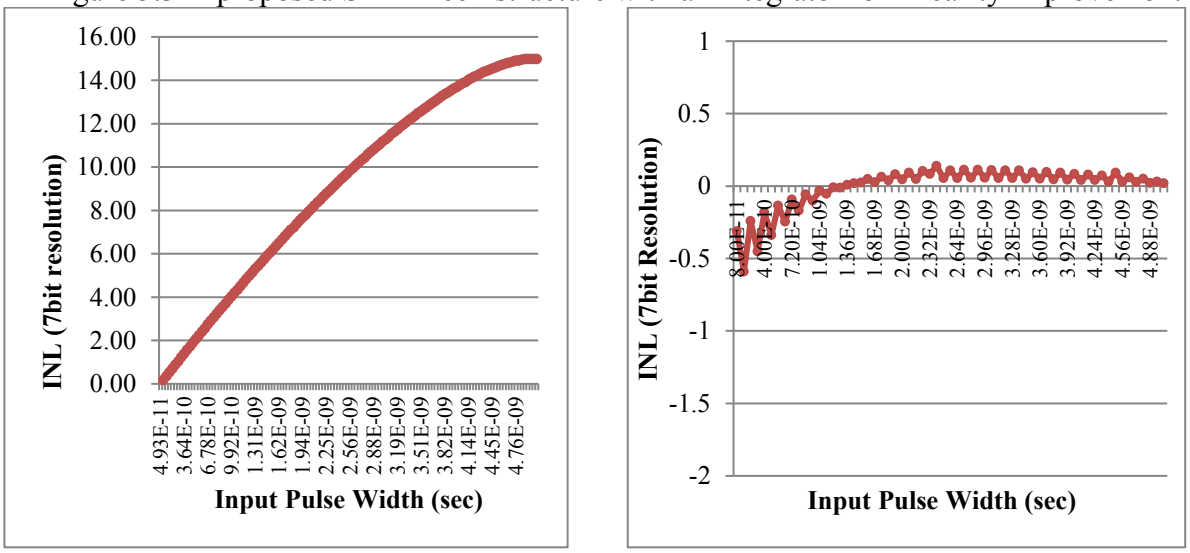


Figure 5.3 A proposed SRAM cell structure with an integrator for linearity improvement



(a) (b)

Figure 5.4 Linearity simulation results (a) without (b) with a cascode buffer and an integrator

B. Nonlinear Transfer function in Pulse Width domain

After a dot product is calculated, nonlinear functions are applied to the outcome to activate a neuron. For example, one of the most common nonlinear function is rectified linear unit (ReLU) as stated in (3). Previous analog machine learning accelerators suffered from lack of nonlinear

functions in analog domain, resulting in analog to digital conversion to realize nonlinear functions. However, such analog to digital conversion and succeeding digital to analog conversion add energy, area and latency overhead, making analog computation less effective and less attractive.

$$f(x) = \max(x, 0) \tag{3}$$

This work introduces a concept of analog domain non-linear function using a nonlinear reference signal and a continuous comparator as illustrated in Figure 5.5. Three transistors on the left side show a read buffer connected to SRAM cell. Dot product calculation output builds up at the integrator output node. A continuous comparator detects a point where this output voltage becomes less than a nonlinear reference signal ($V_{REF,NL}$), and generates a pulse signal.

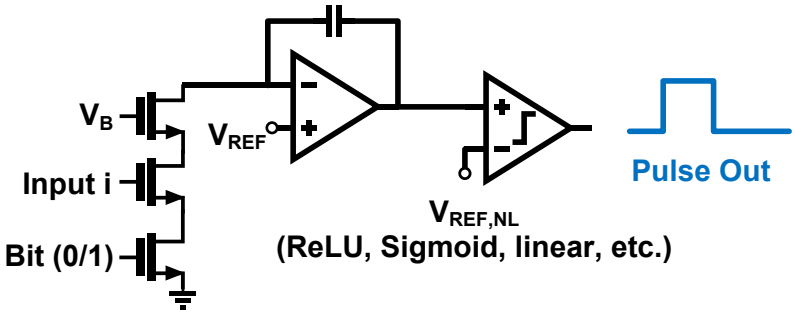


Figure 5.5 Concept of nonlinear transfer function from voltage to pulse width

To understand this concept, linear transfer case is first explained as follows. Figure 5.6 conceptually illustrates charge mode dot product calculation with 3 pulse inputs (i, i+1, i+2) and linear transfer of the output voltage. From the left figure, a red line is $V_{REF,NL}$ for this case. The comparator output starts with 0 by a resetting logic which is not shown in Figure 5.5 and changes to 1 as soon as the comparator fires. At the point where the red line crosses the output voltage, the comparator output becomes 0 and a pulse width is defined accordingly. Another case with lower dot product output voltage is described in the right figure. With lower voltage, corresponding pulse

width decreases, resulting in the linear transfer curve from input voltage to output pulse width as shown in Figure 5.7. Simulation results for voltage to pulse width conversion is shown in Figure 5.8.

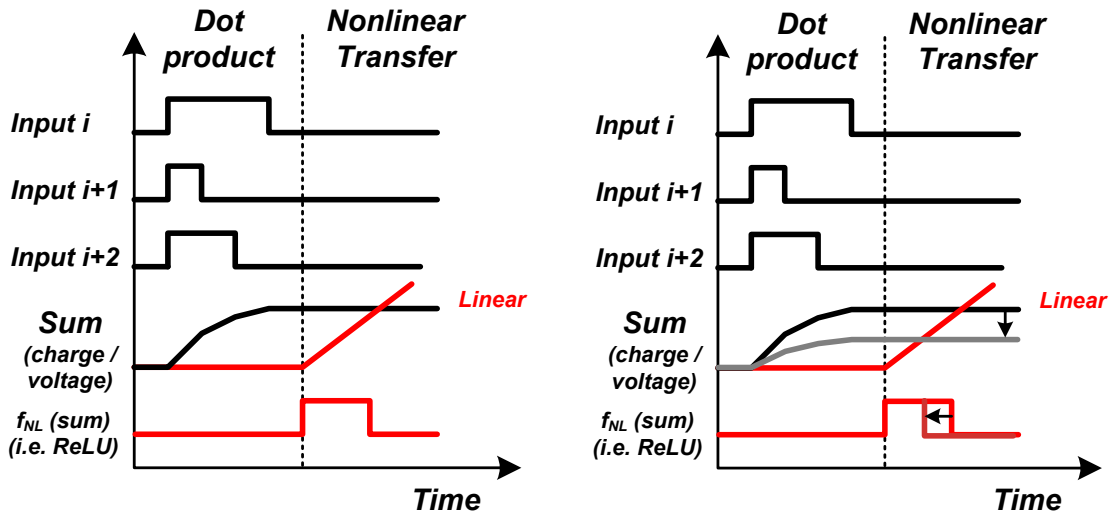


Figure 5.6 Conceptual diagram showing charge mode dot product and voltage to pulse width transfer

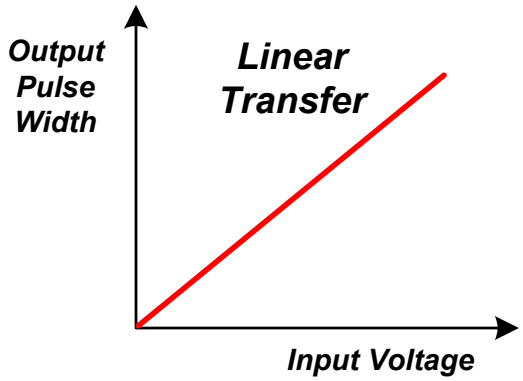


Figure 5.7 Input voltage vs. output pulse width of the introduced transfer circuit

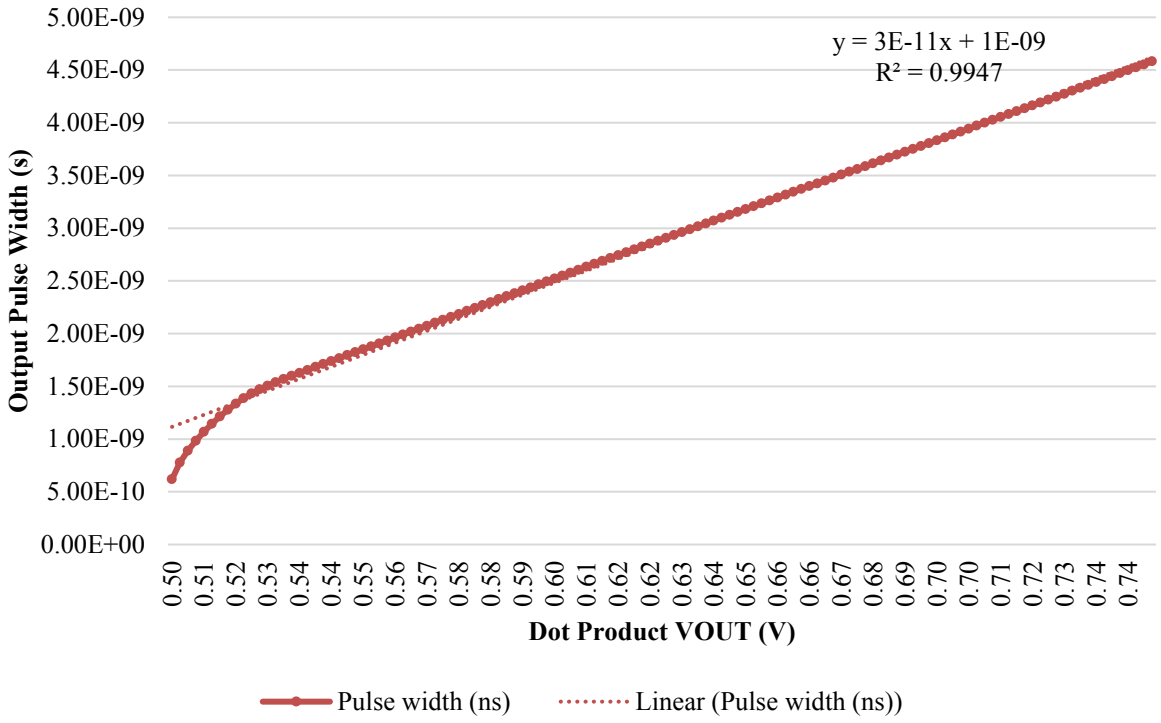


Figure 5.8 Linearity simulation result of voltage to pulse width conversion

Now, for ReLU transfer case, the same principle is applied with the previous case, but $V_{REF,NL}$ becomes a green curve in Figure 5.9. $V_{REF,NL}$ starts from an intermediate voltage which corresponds to number 0 in voltage domain. Thus any dot product output voltage less than this intermediate voltage will generate zero pulse width, and above that voltage, transfer curve will be linear as shown in the right figure. Generating the nonlinear reference signal is achieved by using an identical integrator with the one used for charge integration in Figure 5.5. Instead of the read buffer from SRAM cell, a current source draws current from the negative input node of the integrator. With this approach, systematic nonlinearity of dot product introduced by non-idealities of integrator circuit can be cancelled out by the same nonlinearity introduced by $V_{REF,NL}$ to a certain extent. Ideally, $V_{REF,NL}$ can be shared by many comparators to provide uniform reference signal and also to save power and area overhead. In real, the number of shared comparators is limited by

signal integrity of the reference signal because comparator outputs feed back to the reference signal through capacitive coupling and affect its shape.

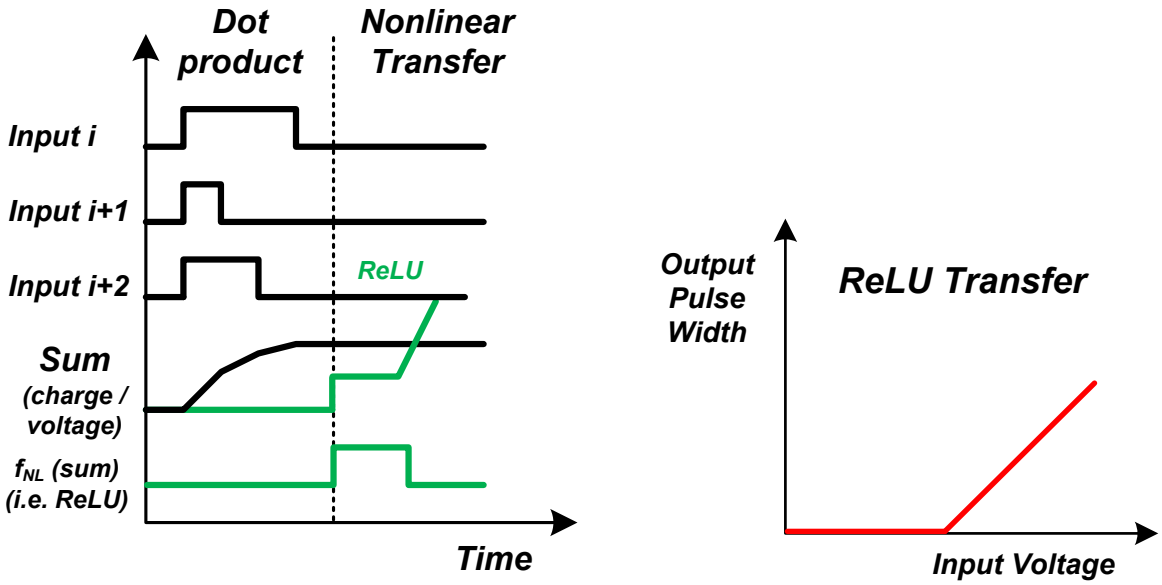


Figure 5.9 ReLU transfer of voltage into pulse width, and its transfer curve

5.3 Architecture

From this section, architecture of this proposed work will be explained in detail. Deep learning algorithms typically have more than 4 layers and even from a few tens to more than thousand layers [64]. Thus, it is very important to design layer to layer interconnect not just intra-layer structure to optimize energy efficiency, throughput and latency. First, input images are sent to an input layer in digital domain. The input layer converts digital values into voltages using a digital to analog converters and then to pulse signals using the same continuous comparator structure in Figure 5.5. The only signals that feed into the next layer are pulse signals. Pulses are advantageous than analog voltage or current because they can be transmitted using standard digital gates such as inverters and buffers. Pulses will be skewed depending on capacitance and resistance

of each path. However only the pulse width, not the absolute pulse arrival time, is an input for calculation, so this architecture has immunity on skew issues. In addition, pooling functions like Max-pooling can be easily achieved by an OR-gate (Figure 5.10) instead of digital adders and comparison logics needed in digital calculation.

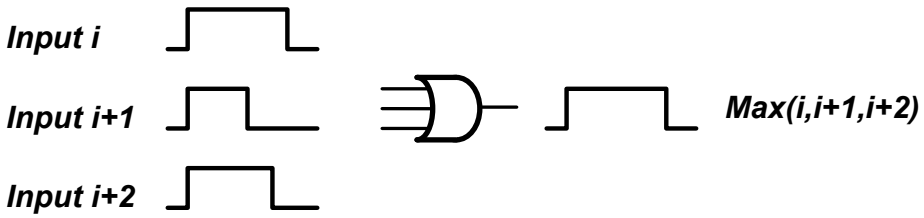


Figure 5.10 Max-pooling in pulse-width domain by OR-gate

After the input layer which converts digital data into pulse widths, the following layer is an integrated layer of a convolutional layer, nonlinear transfer layer (i.e. ReLU), and a pooling layer. The structure shown in Figure 5.5 performs convolution and ReLU, and digital gates are used for pooling. It is worth noting that inputs to this integrated layer are pulse width modulated (PWM) digital signals, and outputs of this layer are also PWM digital signals. There are no A to D or D to A conversion inside or between layers, reducing latency and energy/area overheads.

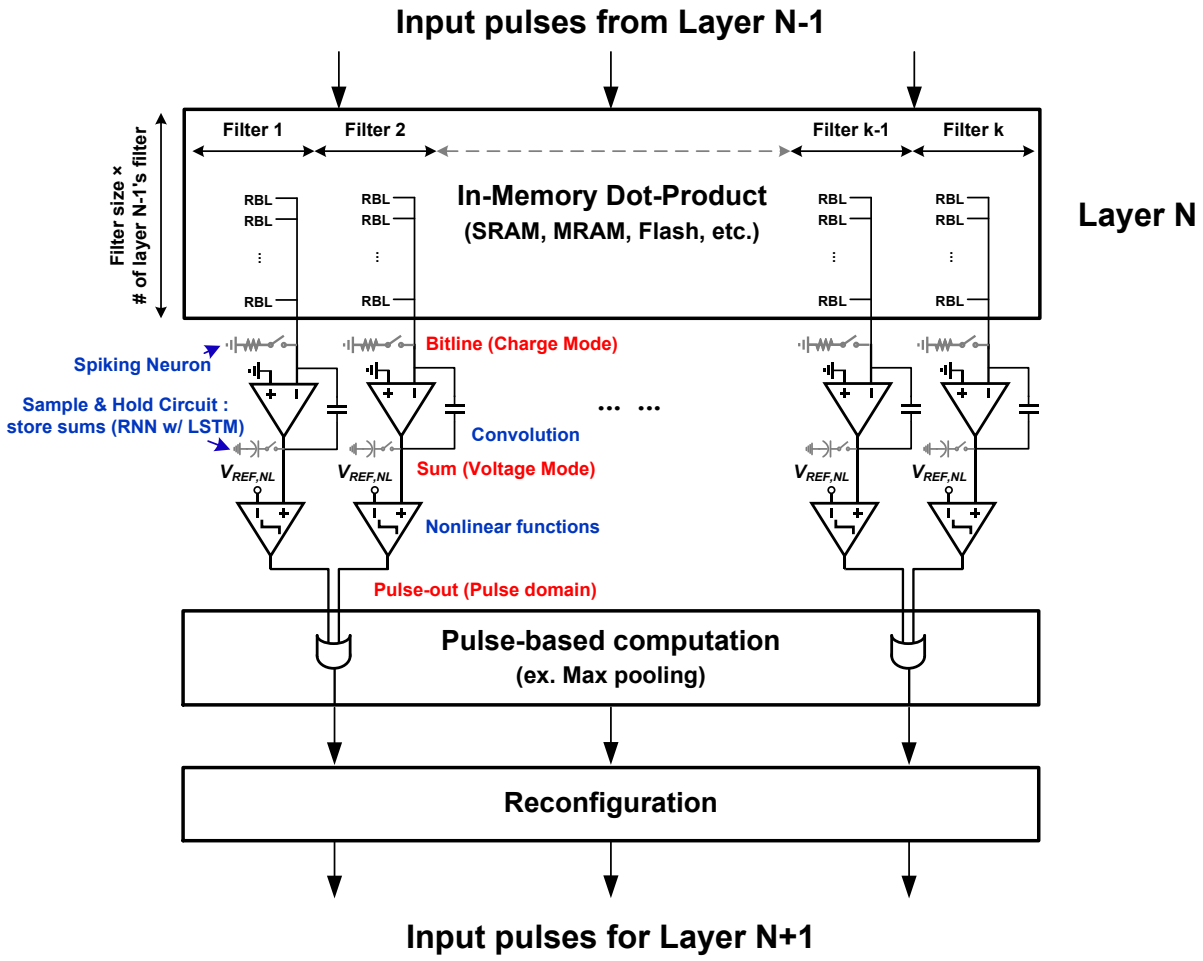


Figure 5.11 An integrated layer of convolutional, nonlinear transfer, and pooling layers

Figure 5.11 shows a block diagram of the integrated layer. A filter inside the in-memory dot-product is duplicated for k times, which is the number of filters per layer. Reconfiguration block assigns which output signals should connect to which location in the next layer. This work implemented 17-layer residual learning algorithm [64] for 10-way image classification scenario. A detailed algorithm architecture is described in Figure 5.12.

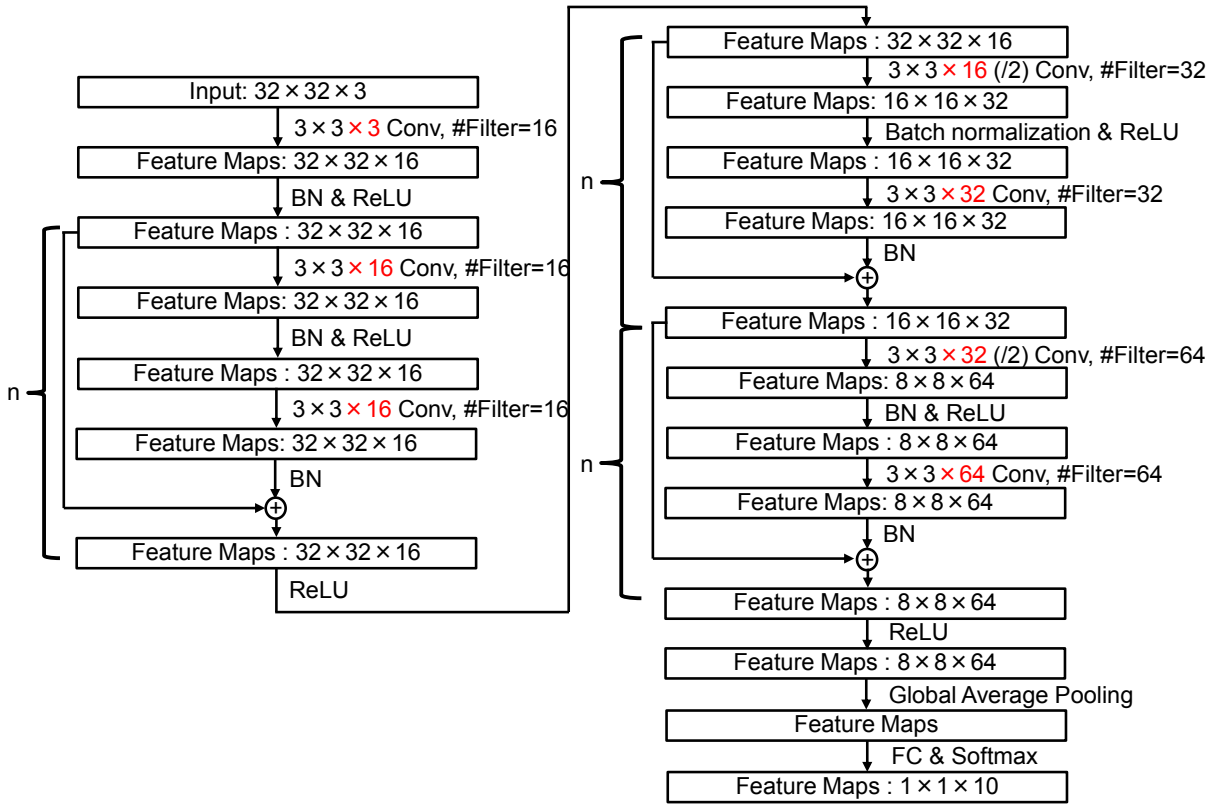


Figure 5.12 Architecture of residual learning network for 10-way image classification

5.4 Future Direction

A general purpose in-memory computing deep learning accelerator which can dynamically update fabrics will be a powerful substitute for a general purpose GPU, improving energy efficiency and throughput by orders of magnitude. A major challenge in achieving plasticity is rapidly evolving deep learning algorithms. Even though core dot-product module remains the same, many system parameters differ by algorithms such as filter sizes, number of filters per layer, number of layers, pooling methods (average pooling, max pooling, stride of greater than 1, etc.) and non-linear functions (sigmoid, ReLU, etc.). Moreover, state-of-the-art algorithms feature custom-shaped modules such as an inception module from GoogLeNet and a residual module and a bottleneck module from ResNet. The proposed system is highly amenable to structural plasticity.

Pulse width modulation is a promising method for plasticity because it simplifies interconnect and multiplexing structures. Charge integration method can convert non-spiking system into a spiking neural network by simply adding a constant current leakage path (or a resistor). A digital OR gate implements max pooling in pulse width domain, and convolutional layer can be used as an average pooling layer with weight of all 1. Non-linear functions can be achieved by a DAC and a look-up table, with high level of programming freedom.

CHAPTER 6

Conclusions

Internet of Things has gained lots of interest and popularity in recent years, and it started to find tangible applications and markets. A millimeter sized sensor for downhole oil well exploration can be one example [2]. Along with low power digital circuit techniques such as DVFS, power and clock gating, and so on, ultra-low power analog circuit techniques have facilitated realization of such miniature sensors. Design of low power analog circuits is challenging in that it usually requires new structures and novel views of circuit, not just lowering supply voltages or operating frequencies.

This dissertation introduced various low power analog circuit techniques. Chapter 2 presented a new topology to generate a sub-nA (20pA) level reference current with very low power overhead. It shows temperature coefficient of 780ppm/ºC and consumes 23pW, which is more 50 times smaller than the lowest power consumption reported previously [15]. This work also described techniques to improve supply voltage and load voltage regulation.

In Chapter 3, we discussed an on-chip oscillator topology that replace the comparator present in traditional RC relaxation oscillators with an ultra-low power amplifier. The comparator and buffer delays are one of the main sources of temperature instability. In this work, a period is formed by a frequency locking loop with a switched-capacitor circuit. This oscillator produces a 70.4kHz clock with an average temperature coefficient of 34.3 ppm/ºC in the −40ºC to 80ºC range,

an average supply voltage sensitivity of 0.75%/V in the 1.2 V to 1.8 V range for five samples, and long-term stability of less than 7 ppm after an integration time of 12 seconds while consuming 110 nano-watt at room temperature.

Chapter 4 proposed a resonant current-mode wireless power receiver and battery charger. This method directly charges a battery with inductor current, and furthermore, resonates the LC tank for multiple cycles to maximize its power efficiency by balancing switching and conduction losses. With this new approach, this work achieves minimum harvestable input power of 600nW, and maximum efficiency greater than 60% at >7.5× lower input power than related work.

Finally in Chapter 5, we introduced a deep learning accelerator design concept in analog domain using in-memory computation of SRAM. Extracting useful information from gathered data is a key function of Internet of Things. This work proposed an alternative method to conventional digital computation which can enhance energy efficiency and throughput by reducing data movement power and latency overhead.

Some of the proposed analog techniques are already embedded in a millimeter-sized sensor system [22] and they will possibly find further applications. In the future, analog reference circuits should find a way to achieve great accuracy and robustness to the extent that can be applicable to industry standards with ultra-low power consumption. Proposed reference circuits require two-point calibrations for temperature stability, and it will reduce cost for manufacturability and testing if one point calibration or no calibration is needed. The suggested wireless power transfer circuit introduced a novel technique of wireless charging. To be implanted into body, miniaturization of the entire system including an off-chip inductor and capacitor, and correspondingly, optimization of operating frequency will be advantageous. The proposed pulse width based deep learning accelerator shows great potential to improve energy efficiency of deep learning hardware. Further

research opportunities are wide open to achieve an analog hardware for flexible algorithms and robustness to PVT variations.

BIBLIOGRAPHY

[1] D. Blaauw *et al.*, "IoT design space challenges: Circuits and systems," in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1–2.

[2] M. Choi *et al.*, "Autonomous Microsystems for Downhole Applications: Design Challenges, Current State, and Initial Test Results," *Sensors*, vol. 17, no. 10, p. 2190, Sep. 2017.

[3] M. Choi, I. Lee, T. K. Jang, D. Blaauw, and D. Sylvester, "A 23pW, 780ppm/C resistor-less current reference using subthreshold MOSFETs," in *ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC)*, 2014, pp. 119–122.

[4] M. Seok, G. Kim, D. Blaauw, and D. Sylvester, "A Portable 2-Transistor Picowatt Temperature-Compensated Voltage Reference Operating at 0.5 V," *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2534–2545, Oct. 2012.

[5] M. Choi, S. Bang, T. K. Jang, D. Blaauw, and D. Sylvester, "A 99nW 70.4kHz resistive frequency locking on-chip oscillator with 27.4ppm/C temperature stability," in *2015 Symposium on VLSI Circuits (VLSI Circuits)*, 2015, pp. C238–C239.

[6] M. Choi, T. Jang, S. Bang, Y. Shi, D. Blaauw, and D. Sylvester, "A 110 nW Resistive Frequency Locked On-Chip Oscillator with 34.3 ppm/C Temperature Stability for System-on-Chip Designs," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2106–2118, Sep. 2016.

[7] M. Choi, T. Jang, J. Jeong, S. Jeong, D. Blaauw, and D. Sylvester, "21.5 A current-mode wireless power receiver with optimal resonant cycle tracking for implantable systems," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 372–373.

[8] M. Choi, T. Jang, J. Jeong, S. Jeong, D. Blaauw, and D. Sylvester, "A Resonant Current-Mode Wireless Power Receiver and Battery Charger With -32 dBm Sensitivity for Implantable Systems," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 2880–2892, Dec. 2016.

[9] Y. Lee *et al.*, "A Modular 1 mm Die-Stacked Sensing Platform With Low Power I C Inter-Die Communication and Multi-Modal Energy Harvesting," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 229–243, Jan. 2013.

[10] S. Jeong, J. y Sim, D. Blaauw, and D. Sylvester, "65nW CMOS temperature sensor for ultra-low power microsystems," in *Proceedings of the IEEE 2013 Custom Integrated Circuits Conference*, 2013, pp. 1–4.

[11] T. Hirose, Y. Osaki, N. Kuroki, and M. Numa, "A nano-ampere current reference circuit and its temperature dependence control by using temperature characteristics of carrier mobilities," in *2010 Proceedings of ESSCIRC*, 2010, pp. 114–117.

[12] J. Lee and S. Cho, "A 1.4μW 24.9-ppm/C Current Reference With Process-Insensitive Temperature Compensation in 0.18μm CMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2527–2533, Oct. 2012.

[13] G. Serrano and P. Hasler, "A Precision Low-TC Wide-Range CMOS Current Reference," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 558–565, Feb. 2008.

[14] B. Fotouhi, "All-MOS voltage-to-current converter," *IEEE J. Solid-State Circuits*, vol. 36, no. 1, pp. 147–151, Jan. 2001.

[15] E. M. Camacho-Galeano, C. Galup-Montoro, and M. C. Schneider, "A 2-nW 1.1-V self-biased current reference in CMOS technology," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 52, no. 2, pp. 61–65, Feb. 2005.

[16] D. Yoon, D. Sylvester, and D. Blaauw, "A 5.58nW 32.768kHz DLL-assisted XO for real-time clocks in wireless sensing applications," in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 366–368.

[17] K. J. Hsiao, "17.7 A 1.89nW/0.15V self-charged XO for real-time clock generation," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 298–299.

[18] G. Kim *et al.*, "A millimeter-scale wireless imaging system with continuous motion detection and energy harvesting," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, 2014, pp. 1–2.

[19] “ABM13.pdf.” .

[20] Y. P. Chen *et al.*, “An Injectable 64 nW ECG Mixed-Signal SoC in 65 nm for Arrhythmia Monitoring,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 375–390, Jan. 2015.

[21] H. Reinisch *et al.*, “An Electro-Magnetic Energy Harvesting System With 190 nW Idle Mode Power Consumption for a BAW Based Wireless Sensor Node,” *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1728–1741, Jul. 2011.

[22] Y. Shi *et al.*, “26.7 A 10mm^3 syringe-implantable near-field radio system on glass substrate,” in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 448–449.

[23] W. Jung *et al.*, “8.5 A 60%-efficiency 20nW-500µW tri-output fully integrated power management unit with environmental adaptation and load-proportional biasing for IoT systems,” in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 154–155.

[24] Y. Lee, B. Giridhar, Z. Foo, D. Sylvester, and D. Blaauw, “A 660pW multi-stage temperature-compensated timer for ultra-low-power wireless sensor node synchronization,” in *2011 IEEE International Solid-State Circuits Conference*, 2011, pp. 46–48.

[25] *Electrothermal Frequency References in Standard CMOS | S. Mahdi Kashmiri | Springer*.

[26] T. Tokairin *et al.*, “A 280nW, 100kHz, 1-cycle start-up time, on-chip CMOS relaxation oscillator employing a feedforward period control scheme,” in *2012 Symposium on VLSI Circuits (VLSIC)*, 2012, pp. 16–17.

[27] A. Paidimarri, D. Griffith, A. Wang, A. P. Chandrakasan, and G. Burra, “A 120nW 18.5kHz RC oscillator with comparator offset cancellation for ±0.25% temperature stability,” in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 184–185.

[28] S. Dai and J. K. Rosenstein, “A 14.4nW 122KHz dual-phase current-mode relaxation oscillator for near-zero-power sensors,” in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, 2015, pp. 1–4.

[29] D. Griffith, P. T. Røine, J. Murdock, and R. Smith, "17.8 A 190nW 33kHz RC oscillator with ±0.21% temperature stability and 4ppm long-term stability," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 300–301.

[30] S. Jeong, I. Lee, D. Blaauw, and D. Sylvester, "A 5.8 nW CMOS Wake-Up Timer for Ultra-Low-Power Wireless Applications," *IEEE J. Solid-State Circuits*, vol. 50, no. 8, pp. 1754–1763, Aug. 2015.

[31] J. Lee and S. Cho, "A 10MHz 80 μW 67 ppm/C CMOS reference clock oscillator with a temperature compensated feedback loop in 0.18 μm CMOS," in *2009 Symposium on VLSI Circuits*, 2009, pp. 226–227.

[32] J. Lee, P. Park, S. Cho, and M. Je, "5.10 A 4.7MHz 53 μW fully differential CMOS reference clock oscillator with -22dB worst-case PSNR for miniaturized SoCs," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.

[33] T. O'Shaughnessy, "A CMOS, self calibrating, 100 MHz RC-oscillator for ASIC applications," in *Proceedings of Eighth International Application Specific Integrated Circuits Conference*, 1995, pp. 279–282.

[34] B. R. Gregoire and U. K. Moon, "A Sub 1-V Constant Gm ndash; C Switched-Capacitor Current Source," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 54, no. 3, pp. 222–226, Mar. 2007.

[35] K. Ueno, T. Asai, and Y. Amemiya, "A 30-MHz, 90-ppm/C fully-integrated clock reference generator with frequency-locked loop," in *2009 Proceedings of ESSCIRC*, 2009, pp. 392–395.

[36] T. Jang, M. Choi, S. Jeong, S. Bang, D. Sylvester, and D. Blaauw, "5.8 A 4.7nW 13.8ppm/C self-biased wakeup timer using a switched-resistor scheme," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 102–103.

[37] M. H. Ghaed *et al.*, "Circuits for a Cubic-Millimeter Energy-Autonomous Wireless Intraocular Pressure Monitor," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 60, no. 12, pp. 3152–3162, Dec. 2013.

[38] Y. T. Liao, H. Yao, A. Lingley, B. Parviz, and B. P. Otis, “A 3- CMOS Glucose Sensor for Wireless Contact-Lens Tear Glucose Monitoring,” *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 335–344, Jan. 2012.

[39] D. A. Borton, M. Yin, J. Aceros, and A. Nurmikko, “An Implantable Wireless Neural Interface for Recording Cortical Circuit Dynamics in Moving Primates,” *J. Neural Eng.*, vol. 10, no. 2, p. 026010, Apr. 2013.

[40] M. Yip, R. Jin, H. H. Nakajima, K. M. Stankovic, and A. P. Chandrakasan, “A Fully-Implantable Cochlear Implant SoC With Piezoelectric Middle-Ear Sensor and Arbitrary Waveform Neural Stimulation,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 214–229, Jan. 2015.

[41] “RF_Microwave_Safety_Program_Guide.pdf.” .

[42] Y. Lu, X. Li, W. H. Ki, C. Y. Tsui, and C. P. Yue, “A 13.56MHz fully integrated 1X/2X active rectifier with compensated bias current for inductively powered devices,” in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 66–67.

[43] X. Li, C. Y. Tsui, and W. H. Ki, “12.8 Wireless power transfer system using primary equalizer for coupling- and load-range extension in bio-implant applications,” in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.

[44] D. Pivonka, A. Yakovlev, A. S. Y. Poon, and T. Meng, “A mm-Sized Wirelessly Powered and Remotely Controlled Locomotive Implant,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 6, pp. 523–532, Dec. 2012.

[45] K. G. Moh *et al.*, “12.9 A fully integrated 6W wireless power receiver operating at 6.78MHz with magnetic resonance coupling,” in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.

[46] M. Stoopman, S. Keyrouz, H. J. Visser, K. Philips, and W. A. Serdijn, “A self-calibrating RF energy harvester generating 1V at -26.3 dBm,” in *2013 Symposium on VLSI Circuits*, 2013, pp. C226–C227.

[47] T. Le, K. Mayaram, and T. Fiez, “Efficient Far-Field Radio Frequency Energy Harvesting for Passively Powered Sensor Networks,” *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1287–1302, May 2008.

[48] V. Kuhn, C. Lahuec, F. Seguin, and C. Person, “A Multi-Band Stacked RF Energy Harvester With RF-to-DC Efficiency Up to 84%,” *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 5, pp. 1768–1778, May 2015.

[49] J. Kang, P. Y. Chiang, and A. Natarajan, “21.6 A 1.2cm2 2.4GHz self-oscillating rectifier-antenna achieving -34.5dBm sensitivity for wirelessly powered sensors,” in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 374–375.

[50] J. Kang, P. Chiang, and A. Natarajan, “A 3.6cm2 wirelessly-powered UWB SoC with -30.7dBm rectifier sensitivity and sub-10cm range resolution,” in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2015, pp. 255–258.

[51] O. Lazaro and G. A. Rincón-Mora, “A Nonresonant Self-Synchronizing Inductively Coupled 0.18- m CMOS Power Receiver and Charger,” *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 3, no. 1, pp. 261–271, Mar. 2015.

[52] P. H. Chen *et al.*, “A 120-mV input, fully integrated dual-mode charge pump in 65-nm CMOS for thermoelectric energy harvester,” in *17th Asia and South Pacific Design Automation Conference*, 2012, pp. 469–470.

[53] J. Kim, P. K. T. Mok, and C. Kim, “23.1 A 0.15V-input energy-harvesting charge pump with switching body biasing and adaptive dead-time for efficiency improvement,” in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 394–395.

[54] M. Miyahara, Y. Asada, D. Paik, and A. Matsuzawa, “A low-noise self-calibrating dynamic comparator for high-speed ADCs,” in *2008 IEEE Asian Solid-State Circuits Conference*, 2008, pp. 269–272.

[55] “39mlseries_datasheets_merged.pdf.” .

[56] “Seiko Instruments Inc. Micro Energy Division,” *Seiko Instruments Inc. Micro Energy Division*. [Online]. Available: http://www.sii.co.jp/en/. [Accessed: 10-Apr-2017].

[57] “NANOMEDI.pdf.” .

[58] B. Moons and M. Verhelst, “An Energy-Efficient Precision-Scalable ConvNet Processor in 40-nm CMOS,” *IEEE J. Solid-State Circuits*, vol. 52, no. 4, pp. 903–914, Apr. 2017.

[59] Y. H. Chen, T. Krishna, J. S. Emer, and V. Sze, “Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks,” *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 127–138, Jan. 2017.

[60] B. E. Boser, E. Sackinger, J. Bromley, Y. L. Cun, and L. D. Jackel, “An analog neural network processor with programmable topology,” *IEEE J. Solid-State Circuits*, vol. 26, no. 12, pp. 2017–2025, Dec. 1991.

[61] J. Lu, S. Young, I. Arel, and J. Holleman, “A 1 TOPS/W Analog Deep Machine-Learning Engine With Floating-Gate Storage in 0.13µm CMOS,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 270–281, Jan. 2015.

[62] J. Zhang, Z. Wang, and N. Verma, “In-Memory Computation of a Machine-Learning Classifier in a Standard 6T SRAM Array,” *IEEE J. Solid-State Circuits*, vol. 52, no. 4, pp. 915–924, Apr. 2017.

[63] M. Kang, S. Gonugondla, A. Patil, and N. Shanbhag, “A 481pJ/decision 3.4M decision/s Multifunctional Deep In-memory Inference Processor using Standard 6T SRAM Array,” *ArXiv161007501 Cs*, Oct. 2016.

[64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *ArXiv151203385 Cs*, Dec. 2015.