Washington University in St. Louis
Washington University Open Scholarship

Engineering and Applied Science Theses & Dissertations

McKelvey School of Engineering

Summer 8-15-2018

Self-powered Time-Keeping and Time-of-Occurrence Sensing

Liang Zhou
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Part of the Computer Engineering Commons, Computer Sciences Commons, and the Electrical and Electronics Commons

Recommended Citation
Zhou, Liang, "Self-powered Time-Keeping and Time-of-Occurrence Sensing" (2018). *Engineering and Applied Science Theses & Dissertations*. 385.
https://openscholarship.wustl.edu/eng_etds/385

This Dissertation is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in Engineering and Applied Science Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST.LOUIS

School of Engineering & Applied Science
Department of Computer Science and Engineering

Dissertation Examination Committee:
Shantanu Chakrabartty, Chair
Roger Chamberlain
Raj Jain
Chuan Wang
Xuan Silvia Zhang

Self-powered Time-Keeping and Time-of-Occurrence Sensing

by

Liang Zhou

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
St. Louis, Missouri

© 2018, Liang Zhou

Table of Contents

List of Figures

List of Tables

Acknowledgments

It would be impossible to finish this dissertation without others' help and support. I would like to take this opportunity to thank them particularly.

First I would like to thank my advisor Professor Shantanu Chakrabartty who sponsored me the opportunity for my research. He has been always supportive and provided numerous and inspiring help that guided me through the research process. His advising in the past several years stands as a beacon and enlightened my path towards the goal of my study. I really appreciate him for sharing those brilliant ideas and thoughts with me. I also want to thank him for all the effort in training me as an eligible researcher. Without his guidance, it would not be possible for me to achieve this and finish this dissertation.

I would like to thank my collaborators, Professor Jian Ren and Mr. Afifi. Thanks for their supportive and excellent collaboration work that improves and expedites my research.

I thank my advisory committee members, Professor Roger Chamberlain, Professor Raj Jain, Professor Xuan Silvia Zhang and Professor Chuan Wang for their great support and advising in my study and research process. I am grateful for their time and effort in reviewing my progress and my dissertation.

I am also grateful to my labmates, Tao Feng, Mingquan Yuan, Kenji Aono, Hassan Khan, SriHarsha Kondapalli, Darshit Mehta, Ahana Gangopadhyay, Oindrila Chatterjee, Yarub Alazzawi and Brittany Scheid. I really enjoy the time we spent together, and thank them for making the lab a pleasant place to stay. I am really grateful for all the useful discussions we have that boosted my research. I would like to thank Kenji particularly for his help and support to accelerate and improve my research.

I would like to thank all my friends, whom I can not list here, for all the time we spent together. It is you guys who make my study a really pleasant and enjoyable experience. Thank you for making my life so colorful and so many sweet memories we have together.

Finally, I want to thank my family for their great support during my study. Without their persistent and selfless support, I cannot focus on my study without any concern. Particularly, I would like to thank my wife, Juqian Chen, who has continuously stood behind me and supported all the decisions I have made. Thanks for your patience and love.

Liang Zhou

Washington University in Saint Louis

August 2018

Dedicated to my family.

ABSTRACT OF THE DISSERTATION

Self-powered Time-Keeping and Time-of-Occurrence Sensing

by

Liang Zhou

Doctor of Philosophy in Computer Engineering

Washington University in St. Louis, 2018

Professor Shantanu Chakrabartty

Self-powered and passive Internet-of-Things (IoT) devices (e.g. RFID tags, financial assets, wireless sensors and surface-mount devices) have been widely deployed in our everyday and industrial applications. While diverse functionalities have been implemented in passive systems, the lack of a reference clock limits the design space of such devices used for applications such as time-stamping sensing, recording and dynamic authentication. Self-powered time-keeping in passive systems has been challenging because they do not have access to continuous power sources. While energy transducers can harvest power from ambient environment, the intermittent power cannot support continuous operation for reference clocks. The thesis of this dissertation is to implement self-powered time-keeping devices on standard CMOS processes.

In this dissertation, a novel device that combines the physics of quantum tunneling and floating-gate (FG) structures is proposed for self-powered time-keeping in CMOS process. The proposed device is based on thermally assisted Fowler-Nordheim (FN) tunneling process across high-quality oxide layer to discharge the floating-gate node, therefore resulting in a time-dependent FG potential. The device was fully characterized in this dissertation, and it does not require external powering during runtime, making it feasible for passive devices and systems.

Dynamic signature based on the synchronization and desynchronization behavior of the FN timer is proposed for authentication of IoT devices. The self-compensating physics ensure that when distributed timers are subjected to identical environment variances that are common-mode noise, they can maintain synchronization with respect to each other. On the contrary, different environment conditions will desynchronize the timers creating unique signatures. The signatures could be used to differentiate between products that belong to different supply-chains or products that were subjected to malicious tampering. SecureID type dynamic authentication protocols based on the signature generated by the FN timers are proposed and they are proven to be robust to most attacks. The protocols are further analyzed to be lightweight enough for passive devices whose computational sources are limited.

The device could also be applied for self-powered sensing of time-of-occurrence. The prototype was verified by integrating the device with a self-powered mechanical sensor to sense and record time-of-occurrence of mechanical events. The system-on-chip design uses the timer output to modulate a linear injector to stamp the time information into the sensing results. Time-of-occurrence can be reconstructed by training the mathematical model and then applying that to the test data. The design was verified to have a high reconstruction accuracy.

Chapter 1

Introduction to Self-powered Timers

Time has penetrated into almost every aspect of human beings' activities. At the micro-scale, each person has their own time-keeping devices such as a wrist-band watch or a smartphone to schedule their daily activities to maintain a regulated and efficient lifestyle. At the macro-scale, delicate systems require accurate time-keeping or synchronization to perform accurately and correctly. For instance, modern computing systems rely on a global clocking system to ensure correct order of operation. Wireless communications also require accurate synchronization across different networks for successful information exchange. In this chapter, the history of designing and using time-keeping devices will be presented first, followed by the introduction of self-powered and passive devices, which motivate the design of self-powered timers. Then the challenges in designing time-keeping devices for passive systems are discussed. Finally the contributions of this dissertation are summarized.

1.1 Time-keeping Devices

The Chinese saying ”work at sunrise, rest at sunset” describes in the ancient time, people used sun as the indication of time for everyday activities. Time has been playing a significant role since the start of human age. It provides the reference to regulate the schedule of events, such as farming, eating and sleeping. Navigation for sailing on the sea requires more precise time-keeping for accurate estimation of location. When entering the modern life, time-keeping becomes even more important, spanning a variety of applications ranging from global positioning systems (GPS), network synchronization to high-performance computation. Therefore, designing high-performance time-keeping systems has been always a motivating task for scientists and engineers.

Clocks are designed to measure, track and indicate time. Early clocks in human history take advantage of particular natural moves to keep track of time. For instance, sundial clocks use the movement of sun hence the location of its shadow on a dial to indicate time, while sand clocks and water clocks use the movement driven by gravity to record time. Promoted by the advance of science and technology, making more accurate clocking devices became feasible. The first pendulum clock was built in 1657 based on the scientific discovery of fixed periods in dangling pendulums and it significantly improved the time-keeping accuracy. Motivated by the navigation on the sea, marine chronometers based on mechanical oscillation were invented in the 18th century to avoid the drawbacks of pendulum clocks which are prone to rocking and shaking. It could achieve a time-keeping accuracy of 5 seconds for a 10-week period [37]. The discovery of quartz crystal together with the development of electronics enabled another jump in the watch industry. The oscillation of quartz crystal is relatively stable and resistant to environment changes, making it feasible for compact time-keeping. After 1970s, electronic watch based on quartz crystal started to dominate the watch market. The development of

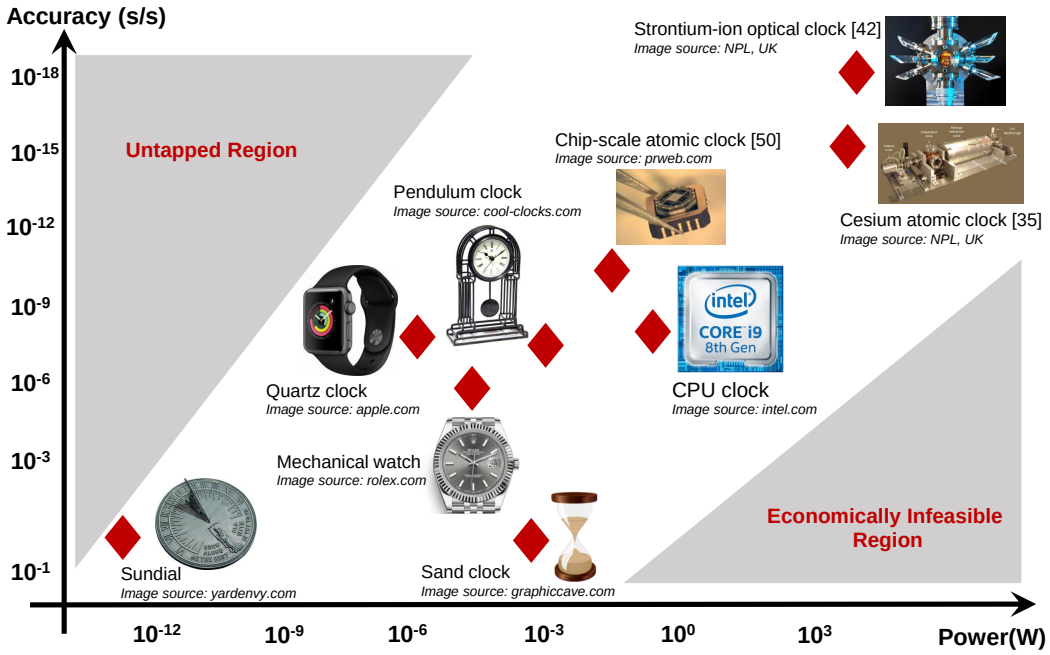


Figure 1.1: Comparison of clocks with different operation principles.

quantum physics and molecular-atomic physics, together with engineering techniques makes it possible for super accurate time-keeping. Atomic clock based on the quantum transition of electrons across different energy levels can achieve accuracy as high as 10^{-18}. This super accuracy enables the proliferation of modern applications such as the GPS all over the world.

Design of clocks is a trade-off over multiple factors such as power, accuracy and footprint. Fig. 1.1 illustrates the power-accuracy performance of typical clocks. In the ancient time, it was difficult to store energy in a compact style, therefore the size of clocks are usually large. Although sundial clock itself does not consume energy, it significantly depends on the direct light from the sun and calibration of the absolute direction, therefore it has to be anchored at a place and track time at sunny days. Pendulum clock as its name indicates requires a pendulum to maintain time. Since the dangling period depends on the length of the pendulum, the size can not be scaled. Also as previously mentioned, pendulum clock is prone

to shaking, therefore, it also needs to be anchored for accurate time-keeping. Mechanical clock and electronic clock built upon quartz crystal shrink the size making it wearable at the wrist, while maintaining a low power consumption down to 10^{-6}W, therefore pushing the time-keeping accuracy limit for daily applications.

In electronic systems, such as a smartphone or a laptop, the von Neumann architecture requires the system to work in a synchronized way, which needs a synchronous clock. When the clock speed reaches GHz range, the maintenance of synchronization becomes much more challenging. As a result, the design of CPU clock put more effort on the optimization of form factors, synchronization and power consumption, instead of the absolute timing accuracy. However, in applications such as network synchronization and GPS, the absolute timing accuracy is key to the system performance because such network spans a large spatial scale. For instance, the global positioning system using the triangular principle to calculate the position. The distances between the position of the target and the satellites, which are calculated from the speed of electro-magnetic wave (speed of light) multiplying by transmission time, determine the positioning accuracy. Since the speed of light is large, super accurate time is required to achieve small positioning error. A timing error of 1µs will result in a 300 meters location error, which is unacceptable for GPS applications. Atomic clock, which uses the stability of quantum transitioning between different energy levels of an atom for time counting, is employed for GPS systems to provide an accurate timing. To stabilize the emitted wave frequency which determines the timing accuracy, the system needs to be cooled down to extremely low temperature to mitigate thermal interference, leading to increasing design complexity and power consumption. A typical atomic timer can achieve a timing accuracy ranging from 10^{-15} to 10^{-18}[35, 42, 50].

In the era of Internet-of-Things (IoTs), more dimensions are introduced into the design space of time-keeping devices. Low-power chip-scale clocks are attracting a lot of attention because

they provide good leverage in performance metrics such as volume, power and accuracy. For instance, wireless network applications usually put more effort in designing timers with small power consumption and form factors. In [55], Lee et.al. proposed a watchdog clock for wireless sensor applications which operates at sub-kHz frequency range and consumes sub-nW power. The clock is mainly used for digital computation synchronization, and cannot provide time counting function. On the contrary, chip-scale atomic clocks were proposed in literature to maintain accurate timing locally [50]. The power consumption of such timers can be optimize to less than 100 mW with timing accuracy up to 10^{-10}.

However, it is challenging to design continuous timers for passive devices, because passive devices do not have access to continuous power sources and also the system volume is too limited to implement specific types of timers such as quartz clocks. With the expansion of applications based on passive devices, the necessity of designing self-powered timers for passive systems is increasing. In the following section, we will briefly introduce self-powered and passive devices.

1.2 Self-powered and Passive Devices

Self-powered and passive devices eliminate the need of continuous power sources such as batteries or power grids, therefore enable rich forms of implementation. The light-weight and small form-factor features expedite the broad adoption of them in numerous applications. Based on the functionality of passive devices, they can be classified into three major categories: (a) self-powered sensors, (b) passive tags and cards, and (c) passive electronic components.

Self-powered sensors target at sensing one or a group of physical or chemical parameters and operate by harvesting energy from ambient environment using energy transducers. As illustrated in Fig. 1.2 [10], self-powered sensors are attractive across multiple applications

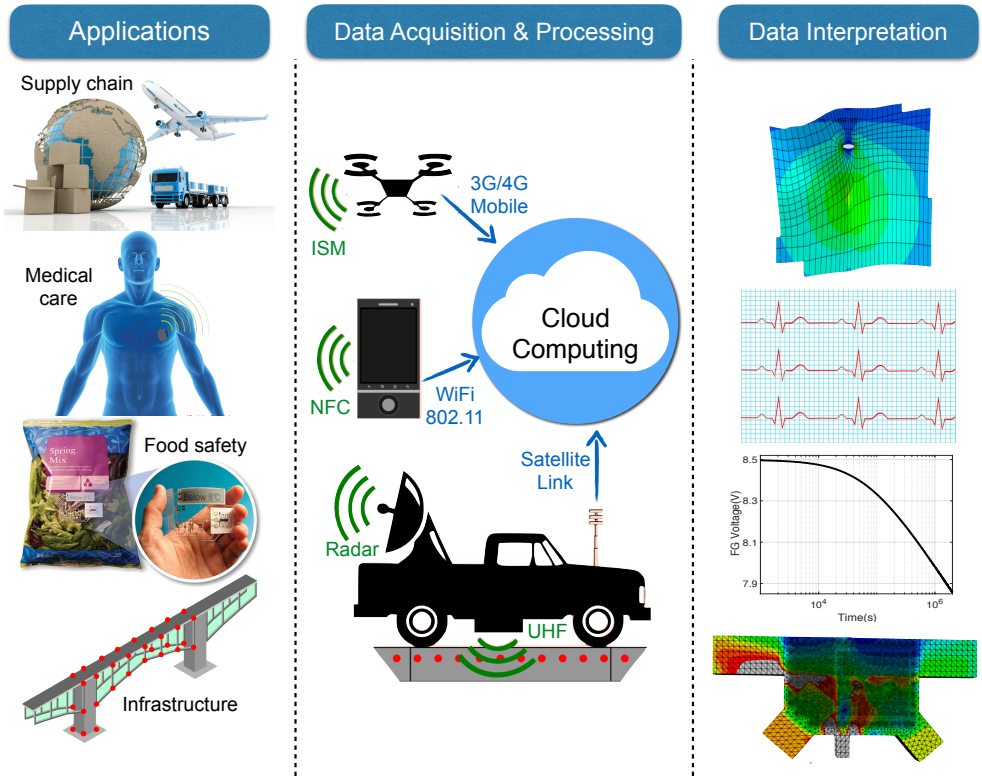


Figure 1.2: An overview of self-powered sensors with implementation, data collection and analysis.

such as supply chain monitoring [54, 57, 96], biomedical monitoring [24, 79], food safety [88, 92, 93] and infrastructure monitoring [10, 44]. A self-powered sensor composes three parts (as illustrated in Fig. 1.2): a front-end sensor that collects data, a middle-ware data storage and acquisition interface and a back-end data analysis and interpretation server. The sensor itself harvests energy from diverse ambient energy sources such as solar, mechanical, chemical and thermal fluctuations. Table. 1.1 summarizes the power density and I-V features of typical energy sources. The sensed data can be stored as an electrical signal in nonvolatile memories or just used some optical or physics parameters such as color, impedance and diffusion length. The data acquisition methods depend on the architecture and operation principle of the designed sensor. For instance, the information harvested by a QR sensor [94] can be easily accessed by a smartphone scanner. For integrated sensors, the data can be accessed in a

wireless way using RF [44] or ultrasonic [31] interface, or just using a plug-and-play interface [29]. Data interpretation and analysis can be done either locally or remotely on the server. Since the server is not limited in the computational resources, it is the most common way for data analysis.

Table 1.1: Characteristics of Typical Energy Sources

Type	Thermal	Electromagnetic	Mechanical	Biological
Transducer	Thermal couple	L-C coupling, antenna	Electromagnetic, electrostatic, piezoelectric	Biofuel cell
Power Density	2000 µW/cm^2 with 12°C gradient [25]	40 µW/cm^2 at 10m [108]	3.89 – 830 µW/cm^3 [48, 72]	2 – 4 mW/cm^2 [20]
Open-circuit Voltage V_{OC}	10 mV - 10 V	100 mV - 5 V	10 - 50 V	100 mV-1 V
Frequency f_S	DC	100 kHz – 5 GHz	0.1 Hz – 1 kHz	DC
Source Impedance Z_S	Resistive impedance 1 – 100s of Ω	Inductive impedance Low kΩ	Capacitive impedance 10s of kΩ – 100 kΩ	10s of kΩ – 100 kΩ

The second type of passive devices, namely passive tags and cards, are commonly seen and used in everyday life as shown in Fig. 1.3. Such examples include the use of RFID tags for supply-chain monitoring [8], retail-pass and access control, personal identification such as state ID and passports, and financial cards (e.g. credit and debit cards). Passive tags and cards only operate when external power sources are accessible. For instance, RFID tags can only be activated when an active reader is within the operational range, and credits cards can only be accessible when plugged into or swiped by a card reader. All the elements in the devices are passive and static.

The third type of passive devices namely the electronic components are extensively deployed in electronic systems such as smart phones, tablets and laptops. The surface-mount feature makes them lightweight and small, making them popular in all electronic platforms.

Image Source: michigan.gov, offgridweb.com, retail-innovation.com, freerepublic.com, uruktech.com, tpvcorentia.com

Figure 1.3: Use of passive tags and cards in daily life.

While self-powered and passive devices are increasingly popular because of the excellent trade-off in power, performance and form factors, their restricted resources limit the functionality of the system. For instance, the accessible power for RFID devices are usually as low as several μW during operational modes, therefore, the computing capability is not strong enough to support complex functionalities. One drawback is the security function in such systems is weak, making them vulnerable to attacks such as eavesdropping. Another limitation is that the static and simple architecture lead to low counterfeiting cost through reverse engineering. Moreover, the lack of continuous power makes it impossible to integrate reference clocks in the system. In this dissertation, we address the challenge by integrating a self-powered system timer in passive devices and systems. The added dimension of dynamic time-keeping will enlarge the design space and enable diverse functions including time-stamp sensing and security enhancement.

1.3 Design Challenges

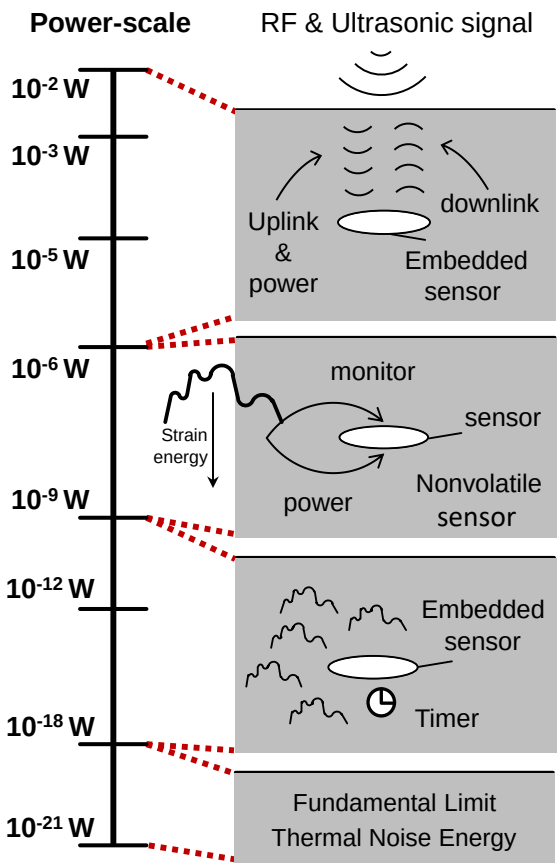


Figure 1.4: Power scale of self-powered sensing systems.

The main task of this thesis is to integrate self-powered timers in passive devices and systems for diverse function expansion. The challenge is how to keep time continuously in an accurate and robust manner without any external powering source. At the fundamental level, the only perennial and omnipresent source of energy that can be exploited in an unconnected, passive device is thermal (ambient temperature), which manifests itself as thermal-noise and fluctuations in electronic devices. Unfortunately, as shown in Fig. 1.4, while ambient energy sources such as RF signals and mechanical variations can provide intermittent power ranging

from microwatts up to milliwatts, the power levels in thermal-fluctuations lie well below 10^{-18}-10^{-16}W which is too scarce for any conventional electronic devices to function. In [14], C. Bennett proposed a way to approach the fundamental limits of computation based on thermally-driven process, where a skew of probability distribution in random walk direction can result in reliable computation, an example of which is the DNA copy operation. In this thesis, the challenge is to investigate thermodynamically driven electron transport phenomena as a way to implement self-powered timers that can continuously operate at the range of several years.

The successful implementation of self-powered timers in passive systems also requires an effective yet efficient interface between the timer circuit and the application circuitry. For instance, how to combine the timer value with the sensing process to achieve time-stamped sensing is challenging, because the interfacing circuits also need to operate in a low-power manner. Moreover, for authentication purpose, building the protocol upon the dynamic response of the timer requires careful consideration from a system-level integration point of view. In general, the challenges are two-fold:

- Robust time-keeping using thermal noise fluctuations for operational periods in the range of years.
- Application-dependent interfacing circuit which is highly accurate and power efficient.

1.4 Contributions

There are four major contributions in this dissertation and they are summarized as follows.

1. Self-powered timing devices based on thermally assisted quantum tunneling are proposed and characterized. The devices employ the physics of quantum-mechanical tunneling of

electrons through synthetically introduced oxide-traps and through thin-oxide layers. The tunneling process is thermodynamically driven and is able to operate at power levels well below 10^{-18}-10^{-16} W (thermal-noise power level). The device based on Fowler-Nordheim tunneling is characterized using prototypes fabricated on multiple CMOS processes. An accurate mathematical model is proposed to capture the dynamic response of the device. Measurement results show that the device is robust, reliable, scalable and supports long-term operation.

2. The mechanisms that lead to excellent synchronization across different timers are investigated and studied. Built upon the synchronization attribute, a dynamic footprint detection sensor using a couple of timers is proposed and characterized. Desynchronization mechanisms are proposed and verified using fabricated prototypes, and a mathematical model originated from the timer model can describe the desynchronization performance accurately. The temperature impact on the synchronization performance is further explored and used to design time-temperature indicators.

3. Dynamic signature based security architecture and protocol is proposed based on the timing device for passive IoT devices. The protocol combines the timer's feature of synchronization with the advantage of random number generation and hash function, and achieves security performance that is resistant to most attacks. The hardware and computation cost is analyzed and lightweight enough for passive devices.

4. System-on-chip (SoC) designs are proposed to sense the time-of-occurrence of mechanical events. The system consists of two SoC chips, with one as the timer SoC and the other one as injector SoC. The design employs an interfacing module that can connect the output of the timer with the injector, and achieves time-stamped sensing. The design is verified using

fabricated prototypes on standard CMOS process, and demonstrates good time-stamping accuracy.

The dissertation is organized as follows. In Chapter 2, the operation principle of the proposed timing device is discussed in detail and its dynamic behavior is fully characterized using fabricated prototypes on standard CMOS processes. Chapter 3 explores the synchronization attribute of the timer device, and this attribute is used for generating dynamic signatures and time-temperature indicators. Chapter 4 introduces the details of the lightweight authentication protocols based on the self-powered timer device for passive IoT devices. In Chapter 5, the time-stamped SoC sensor is demonstrated and the performance is characterized. Chapter 6 summarizes this dissertation and proposes potential future research directions.

Chapter 2

Self-powered Time-keeping in CMOS Process

In this chapter, we present the design and characterization of the proposed self-powered time-keeping devices in standard CMOS processes. The target is to realize robust time-keeping for at least 3 years which is approximately the lifetime of a typical passive device. The concept and essential elements that compose a self-powered timer is first introduced, followed by exploration and analysis of possible physics mechanisms and corresponding devices that can be used for self-powered time-keeping. Two potential timers based on thermal-assisted quantum tunneling, namely trap-assisted tunneling (TAT) and Fowler-Nordheim tunneling (FNT), are proposed to overcome the challenge in designing long-term self-powered timers. The operational principle, mathematical model, design implementation, and experimental characterization of the tunneling timers based on TAT and FNT are presented in detail to validate the designs.

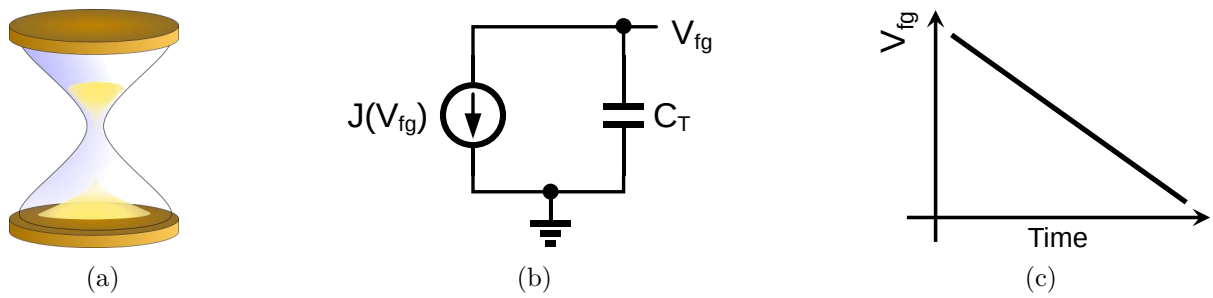


Figure 2.1: Concept of self-powered timers: (a) analogy of sand clock for time-keeping, (b) circuit model and (c) ideal response of a linear self-powered timer.

2.1 Concept of Self-powered Timers

The general idea of implementing a self-powered timer in CMOS process comes from the sand clock used in ancient times as show in Fig. 2.1(a). The sand stored on the top level continues to leak down to the lower level driven by gravity and the residual sand is an indication of time. In CMOS process, we can implement a similar design consisting of a capacitor and a discharging element, the model of which is illustrated in Fig. 2.1(b), where the discharging element is modeled as a leakage current $J(V_{fg})$. Initially, the capacitor C_T stores a predetermined amount of charge like the sand in the sand clock, which will continuously leak through $J(V_{fg})$. V_{fg} will evolve with respect to time hence can be used to track time. Fig. 2.1(c) shows an ideal case where V_{fg} depends on time linearly. To implement a well-defined timer, the capacitor need to maintain the charge well and the charge should be only altered by $J(V_{fg})$. $J(V_{fg})$ should show weak dependence on V_{fg} to approach a linear response. In the following sections, we will discuss the structures in CMOS to implement the timers.

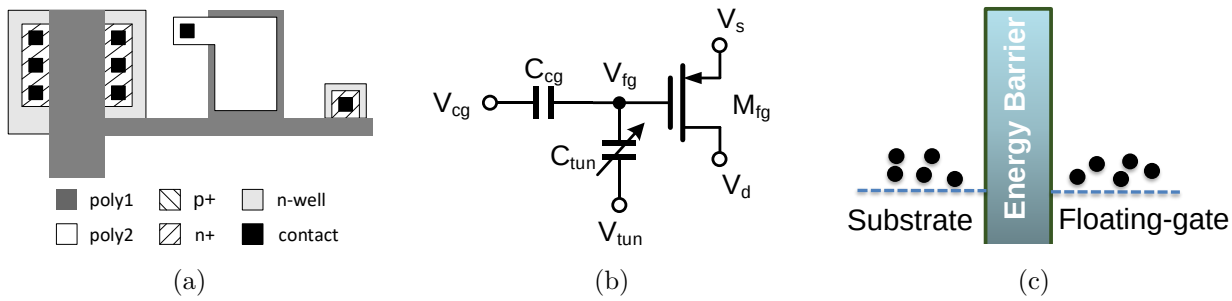


Figure 2.2: Structure of floating-gate transistors: (a) layout, (b) schematic and (c) energy band diagram.

2.1.1 Charge Storage Capacitors

Capacitors are used to store charge in electronic systems. In CMOS process, multiple structures can be employed to implement a capacitor, such as metal-oxide-metal (MOM) capacitors, metal-insulator-metal (MIM) capacitors and MOS capacitors. An ideal candidate of the charge capacitor in CMOS process would be the floating-gate structure which comprises of a strip of poly-crystalline silicon (polysilicon) that is completely insulated by high-quality, thermally-grown silicon-dioxide, as shown in Fig. 2.2(a) and (b). This polysilicon strip, also referred to as a floating-gate, serves as a reservoir of electrons and the surrounding silicon-dioxide will serve as an energy barrier (shown in Fig. 2.2(c)) that prevents the electrons to leak out (by thermal excitation or quantum tunneling). The contribution due to the trap states in the surrounding insulation layers can be minimized by the high-quality oxide. For a standard 500 nm CMOS process, the charge can be retained at an accuracy of 8 bit for 8 years with dynamic range of 0-4 V [41]. Therefore, we employ floating-gate structures to implement the charge storage capacitor. The floating-gate structures have been broadly used as nonvolatile memories in integrated circuits.

The common method for programming the initial value of FG nodes is either by using FN tunneling or by using hot-electron injection [46]. FN tunneling removes the electrons from FG

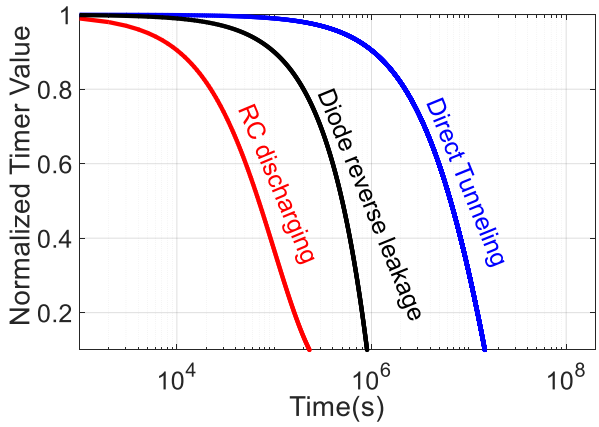


Figure 2.3: Characteristics of timers using different leakage physics.

node by applying a high-voltage (> 15 V for 13nm oxide thickness) across a parasitic nMOS capacitor acting as a programming junction, as illustrated in Fig. 2.2(a). Although single channel programming can be achieved [100], FN tunneling is usually employed to erase the FG value globally. Hot-electron injection, on the other hand, requires lower voltage (> 4.2 V in 0.5-μm CMOS process) than tunneling and hence is the primary mechanism for accurate programming of floating-gates. The hot-electron programming procedure involves applying a voltage larger than 4.2 V across the source and drain terminals. The large electric field near the drain of the pMOS transistor creates impact-ionized hot-electrons whose energy when exceeds the gate-oxide potential barrier (≈3.2 eV) can get injected onto the floating-gate. A combination of FN tunneling and hot-electron injection can program the FG voltage to target value.

2.1.2 Leakage Mechanisms in CMOS Process

Implementing a reliable time-reference by exploiting the discharge characteristics on a capacitor is challenging because it requires a precise on-chip leakage current sink $J(V_{\mathrm{fg}})$ that is continuously active for durations greater than several years. For instance, ensuring a 1 V change across an 1 pF on-chip capacitor over a duration of one year would require an average

discharge current of 3.2×10^{-20} A (equivalent to a rate of ten electrons per minute). The leakage current should be robust, reliable, and show weak dependence on V_{fg} to approach a linear response. In Fig. 2.3 we show the discharge characteristics of three common processes in CMOS process. Implementation using resistors would require a prohibitively large leakage resistance of $10^{18}\,\Omega$ which is difficult to fabricate in CMOS process. Reverse-leakage currents across standard p-n junctions are too large (on the order of femtoamperes) and any active cancellation mechanisms to achieve atto-watt leakage [62] would require external powering.

Electrons can leak through the oxide using three mechanisms of quantum tunneling: (a) direct tunneling (DT) where carriers directly tunnel through the rectangular barrier formed by the dielectric layer, (b) trap-assisted tunneling (TAT) where carriers tunnel through the dielectric layer with the assistance of trap states, and (c) Fowler-Nordheim tunneling (FNT) where carriers tunnel through a triangular-shape barrier resulted from a strong electric field. Direct tunneling (DT) of electrons through the gate-oxide could potentially lower the desired discharge rate; however, the underlying translinear response is still too steep (as shown in Fig. 2.3) and is determined by process parameters (for example oxide-thickness). For instance, the tunneling rate in oxide layer with thickness of 2 nm can easily exceed 10^{-14} A/um^2 [91], which is too large to implement long-term self-powered timers. It is a process specific parameter and cannot be modulated or controlled. In standard 500 nm CMOS processes, it is negligible because the minimum oxide thickness is larger than 10 nm.

Trap-assisted tunneling depends on the number and distribution of traps and is negligible in CMOS floating-gate structures because the oxide is thermally-grown, thus the interface between the polycrystalline silicon and silicon-di-oxide has very few defects or electron traps. However, this process can be aggravated by introducing external trap states by creating defection states in the oxide. FN tunneling, on the other hand, depends on the shape of the energy barrier across the dielectric layer which can be controlled and modulated by changing

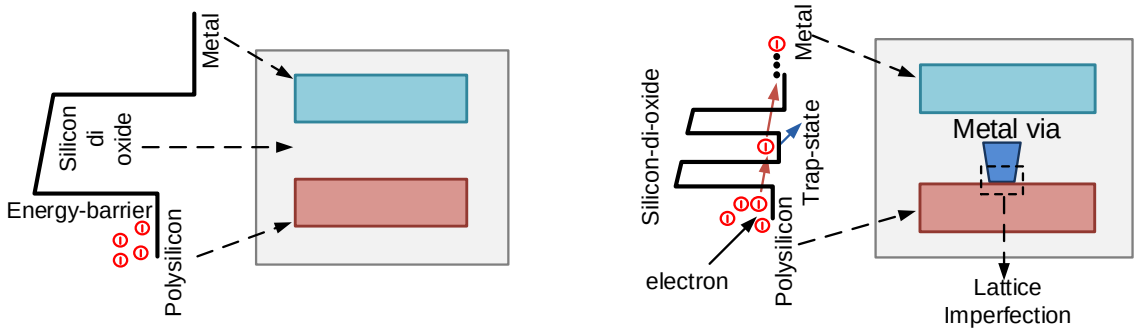
the electric field across the dielectric layer. Therefore, in the following parts, we will explore the use of trap-assisted tunneling and FN tunneling to implement long-term self-powered timers.

2.2 Trap-Assisted Tunneling Timers

In trap-assisted tunneling (TAT) timers [106], the device harvests thermal fluctuation energy from ambient environment to activate the tunneling process for time-keeping. With the assistance of trap-states in the oxide, an intermediate conduction band can be formed and accelerate the tunneling process. As a result, the tunneling rate can be tuned to specific ranges for timer integration.

2.2.1 Trap-assisted Electron Transportation

As shown in Fig. 2.4(a), in ideal case, floating-gate node is surrounded by high-quality oxide which forms an energy barrier that prevents the electrons to either surmount or tunnel through the barrier. The high-quality barrier and hence the retention of charge is determined by the quality of polysilicon silicon-di-oxide interface, which for thermally grown oxide exhibits ultra-low density of imperfections. However, when a metallic interconnect (or a via) is formed on the polysilicon surface (as shown in Fig. 2.4(b)), the surface is strained due to metal-polysilicon lattice mismatch which results in spurious traps at the interface. Therefore, even if the metallic junction is left floating, the charge on the polysilicon floating-gate leaks out over long-duration of time (as shown in Fig. 2.4(b)). In literature, this leakage is typically considered to be a nuisance and several methods have been proposed [15] to reduce this artifact. In this work, the leakage characteristics is exploited to implement a timer that discharges the electrons on the floating-gate over a long-duration of time.



(a) (b)

Figure 2.4: Leakage control: (a) the ideal FG structure without traps and its energy band; (b) the floating gate structure with metal via and its energy band.

When a potential difference exists between the floating-gate and the surrounding metal the transport of electrons could occur due to the following three physical phenomena [69]: (a) trap-assisted tunneling where the electrons move to an unoccupied trap-state by defects close to the metal-polysilicon-oxide interface; (b) Modified Poole-Frenkel (MPF) or internal Schottky emission of the trapped electrons into the conduction band of the silicon-di-oxide; and (c) Thermal-field emission of Fermi-level electrons from the metal directly into the conduction band of the silicon-di-oxide. While trap-assisted tunneling mainly affects the leakage current on short time-scales, MPF and thermal-field emission of electrons is dominant for long time-scales. However, for this preliminary application our focus will be to exploit the combined effect of all the different leakage mechanism. However, one common attribute between different leakage mechanism is that the leakage current increases with the increase in defects at the polysilicon interface. Therefore, we can control the leakage by effectively controlling the number of metallic vias connected the floating-gate.

Exploiting the oxide-leakage current to implement integrators and timers would, however, require precision measurement of the drift in floating-gate voltage. For instance, a timer that is operational over a period of 20 years and with a voltage range of 1 V would require

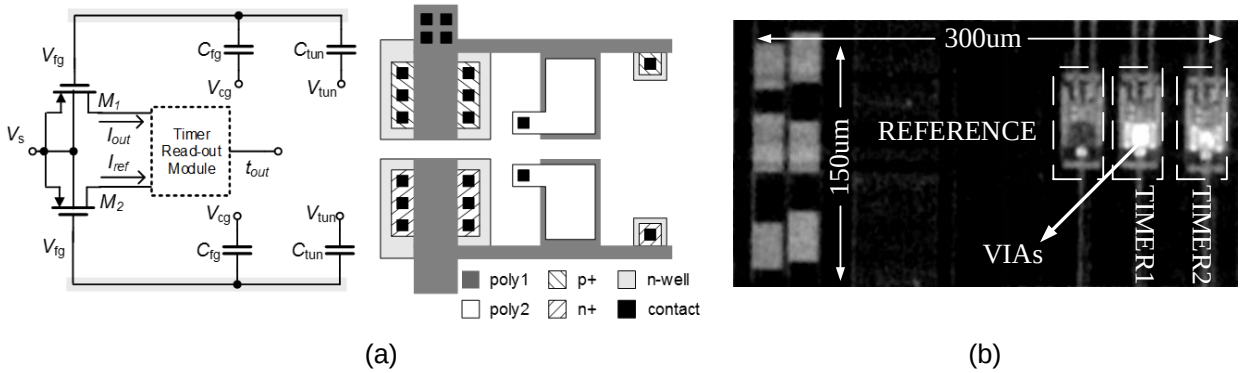


Figure 2.5: Implementation of trap-assisted timers: (a) schematic of differential architecture for accurate readout; (b) micro-photograph of fabricated die.

measuring voltage drifts of less than 10 μV/hour. To accurately extract the change, we couple the floating-gate to the gate of a pMOS transistor and measure the drain-current to infer the charge retained on the floating-gate. The layout of structure is shown in Fig. 2.5(a) which shows a differential architecture with a floating-gate transistor M_2 (with no metallic vias) acting as a reference structure. Note that the floating-gate of transistor M_1 has multiple floating metallic contacts. Ideally, the current through M_2 should remain unchanged (no leakage) once the charge on its floating-gate has been programmed. Under sub-threshold biasing, the drain current I_{ref} through M_2 can be expressed in terms of its floating-gate charge Q_{ref} as

$$I_{ref} = I_0 e^{-\frac{\kappa Q_{ref}}{U_T C_T}} e^{\frac{V_s}{U_T}} \tag{2.1}$$

This reference current I_{ref} could now be used to compensate for effects of temperature variations in the read-out current I_{out}. As shown in Fig. 2.5(a), a timer read-out module measures the reference current I_{ref} and the timer current I_{out} at two different values of the source voltages V_{s1}, V_{s2}. Assuming ideal matching of the transistors M_1 and M_2, the voltage

variance of the timer floating-gate in can be estimated to be

$$\begin{aligned}\Delta Q_{FG,i} &= Q_{FG,i+1} - Q_{FG,i} \\ &= -\frac{C_T}{\kappa}[U_{T,i+1}\ln\frac{I_{out,i+1}}{I_{ref,i+1}} - U_{T,i}\ln\frac{I_{out,i}}{I_{ref,i}}]\end{aligned} \tag{2.2}$$

Where i represents the ith measurement at time t_i. As the technology current characteristic value of I_0 in equation 2.1 could not be extracted just from measurements of I-V characteristics, the charge variance which gives enough information for timer will be used instead as stated in equation 2.2. U_T is the thermal voltage which can be obtained by the reference current at different V_S as

$$U_T = \frac{V_{s1} - V_{s2}}{\ln I_{ref,s1} - \ln I_{ref,s2}} \tag{2.3}$$

Note that even though the effect of temperature variations during read-out has effectively been compensated, the leakage current is still a function of temperature. One future direction in this area will focus on compensating for effects of temperature variation of the oxide-leakage current.

As the temperature variance comes from the ambient, it is reasonable to assume the environment temperature of two tests during a short time period changes slightly which could be neglected. The accurate U_T makes it possible to get the pure leakage effect based on equation 2.2. Since κ and C_T almost keep constant with time and temperature, we can view $\frac{\kappa\Delta Q_{FG}}{C_T}$ instead of ΔQ_{FG} as the process unit. Using the temperature information obtained by reference structure, according to equation 2.1, we could deduce the relationship between $\frac{\kappa\Delta Q_{FG}}{C_T}$ and time which excludes the temperature effect.

Note that the leakage at the beginning is decided by the TAT transient process while the other two factors could be neglected, and then MPF and Schottky emissions will conduct the

leakage process, the measurement circuits will start to work until the current becomes stable after the transient process. Considering the small leakage current, the FG voltage could be viewed as a quasi-static condition, thus the leakage current can be assumed as [83]

$$I_{lkg} = k\Delta V \tag{2.4}$$

where k is a coefficient decided by defect state distribution and temperature. Deduced from equation 2.4 we have the variance of Q_{FG} respect to time as

$$\frac{\kappa \Delta Q_{FG}}{C_T}(t) = \frac{\kappa Q_{FG,0}}{C_T}\left(1 - e^{-\frac{k}{C_T}t}\right) \tag{2.5}$$

The equation implies the charge on the FG will change exponentially to time, which has the same characteristics as a RC time delay circuit. After a time period comparable to the time constant, the voltage will change slowly due to the exponential characteristics. The timer requires the voltage variance to be measurable. Therefore, in order to achieve a timer which can measure time duration as long as T, the time constant must have the same level as T. The time constant has a form as $\tau = \frac{C_T}{k}$ decided by two factors, k and C_T in our structure.

Now we can build a timer with timing ability we want. The basic two methods are to adjust the k or C_T. The C_T is decided mainly by the area of the gate of the transistor, thus method can be easily achieved. However, the value of k is more complicated to be decided accurately. According to the analysis and test results of [69], the leakage rate has a strong relationship with the density of traps and the distance between two electrodes of the capacitor. The former could be increased by adding metal vias, while the latter one can be changed through layout methods. Therefore, through adjusting the timer's structure, different time durations could be achieved.

2.2.2 Implementation and Measurement Results

The proposed timer has been fabricated on a 0.5 μm standard CMOS process. P-type FG MOS timers were used to validate the characteristic of leakage in FG, whose die-photograph is shown in Fig. 2.5(b). Two distinct timers were built to verify the factors that influence the timers' leakage current. As shown in Fig. 2.5(b), the timers have exactly the same layout except the number of metal vias. The reference timer has no vias, while timer1 and timer2 have 35 and 1 via respectively.

In the first experiment, the current of all the timers are obtained, and Fig. 2.6(a) demonstrates the current dependence on the temperature indicated by equation 2.1. The measured characteristics of timer1 is illustrated in Fig. 2.6(b), which contain the temperature impact on the channel current. Using the differential structures, a pure leakage characteristics which excluded the temperature can be acquired, and the leakage characteristics of timer1 and timer2 is shown Fig. 2.6(c) .

By fitting the voltage using the proposed model, two curves could be obtained that match with the measurement results well. It can be observed that the curve of timer1 matches better than that of timer2, which is reasonable since the initial voltage difference between the FG and surrounding metal of timer2 is relatively smaller than that of timer1, thus it is more vulnerable to measurement error. Neglecting temperature effect on leakage current also induces errors to the timers.

The time-constants can be derived from the fitted results, and timer1 and timer2's time constant are estimated as 3.6 hours and 24.5 hours respectively. This result strengthens the conclusion that the number of metal vias could affect the leakage current: more vias lead to a smaller time-constant. This was further verified by measured results from multiple timers

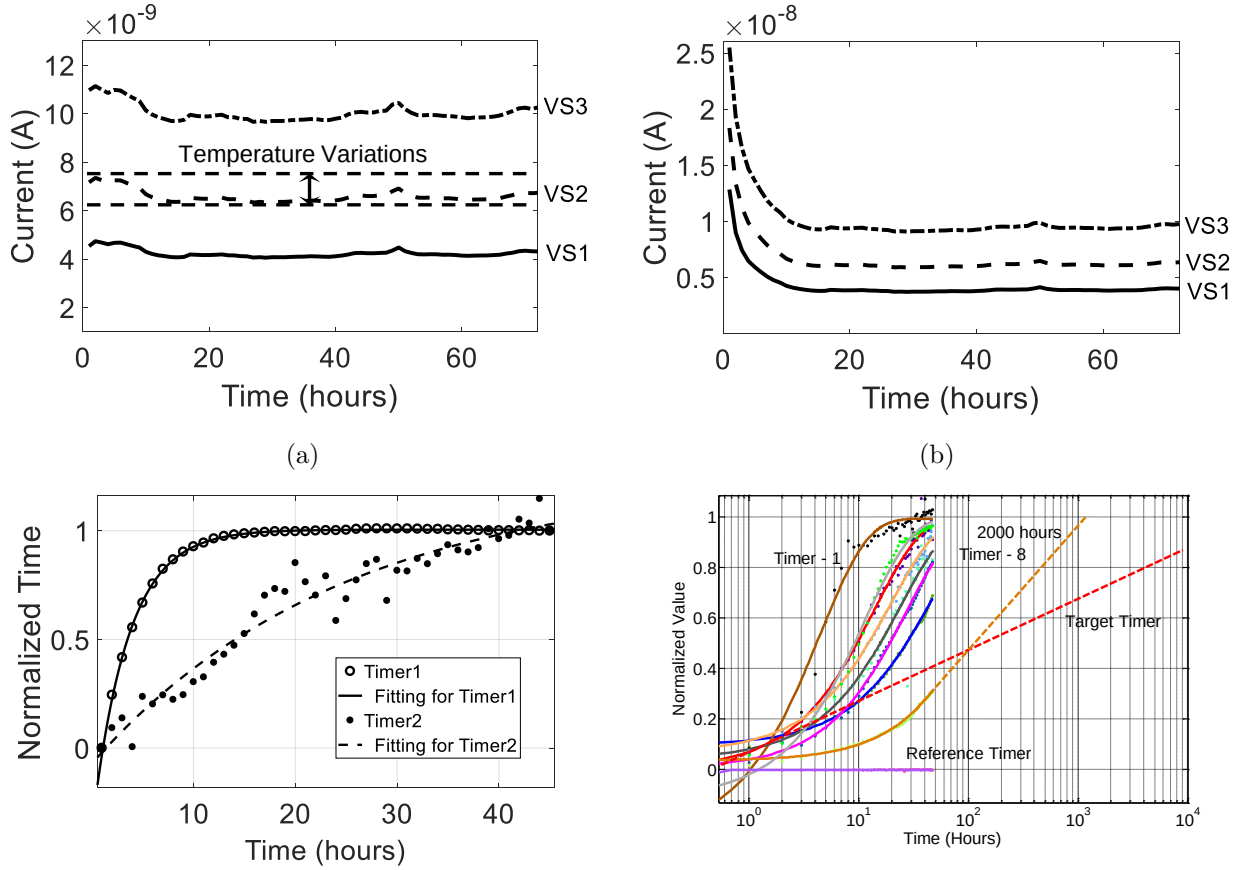


Figure 2.6: Measured response from fabricated TAT timers: (a) the current of ideal FG structure respect to time; (b) the timer's current without eliminating the temperature effect; (c) temperature compensated leakage characteristics of two timers and (d) response of multiple timers with different via numbers.

with different configurations, as shown in Fig. 2.6(d). The TAT timers have the potential to support operational time as long as several months.

2.2.3 Discussions

Trap-assisted tunneling timers were prototyped and verified in this section. The leakage characteristics of the FG transistors and the possible methods to control the leakage rate are measured and verified. It has been proven that with proper regulation, it is possible to

use thermal energy to realize a long-term timer. Temperature influence on the measurement results can be eliminated by employing a differential architecture. TAT timers could be integrated into systems without adding complexity to the technology. However, the difficulty to accurately control the time constant limits the application of the TAT timers. The time constant depends heavily on the number and distribution of the trap states created by the floating vias, which demonstrates random variations in manufacturing process. This implies the time constants of timers with identical structure could show significant difference, therefore it is difficult to synchronize them. It is necessary to characterize each timer individually before implementation. This obviates the design goal of a unified and reliable time reference.

2.3 Fowler-Nordheim Tunneling Timers

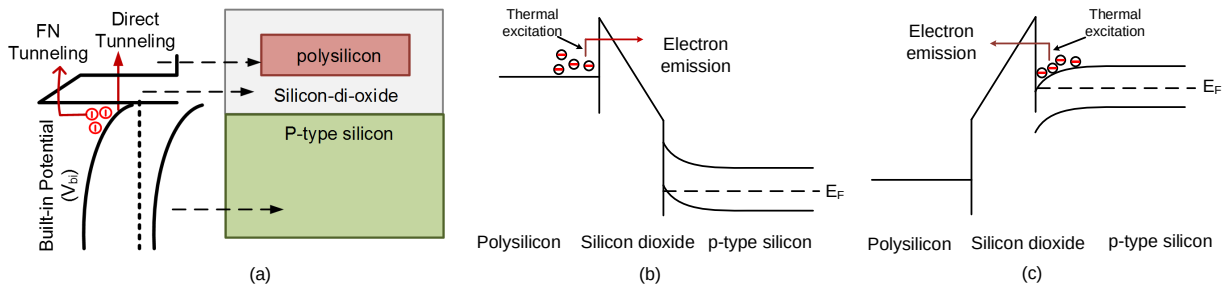


Figure 2.7: Energy-band for FN tunneling in Floating-gate structure: (a) cross-sectional layout and its energy band diagram; (b) negative bias for FN tunneling and (c) positive bias for FN tunneling.

Compared to TAT timers, tunneling mechanism based on FN tunneling demonstrates more favorable attributes [105]. Electron leakage due to FN tunneling dominates when the field across the oxide layer reaches a critical threshold. This is illustrated using the energy-band diagrams of MOS structures with positive and negative bias applied to the gate terminal shown in Fig. 2.7. Since the energy barrier at the interface of holes (3.8 eV) is larger than

that of electrons (3.2 eV), the probability of hole emissions is negligible compared to that of electrons. Positive bias of the gate electrode results in degenerate silicon surface in the n-type substrate, therefore inducing metal-like behavior because the conduction band of the silicon is lower than the Fermi potential at the interface, as illustrated in Fig. 2.7(c). As a result, most of the voltage drop occurs across the silicon-di-oxide barrier. Therefore, in this work, positive bias of FG node for electron tunneling is used to implement self-powered timers.

2.3.1 Mathematical Model

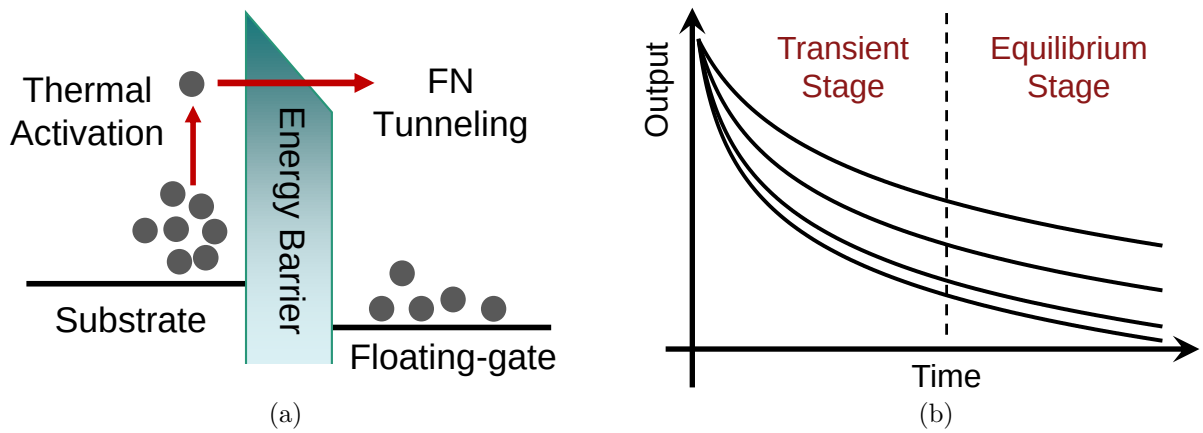


Figure 2.8: Operation of the self-powered FN tunneling timer device. (a) Energy band illustrating two-step tunneling process. (b) Temporal response illustrating the transient and equilibrium regions.

FN tunneling depends on the shape of the energy barrier across the dielectric (oxide) layer which can be controlled and modulated by changing the electric field applied across the dielectric layer. The physics of FN tunneling, as illustrated in Fig. 2.8(a) is a two-step process. Electrons are first thermally excited to an energy level E which then tunnel through the triangular barrier into the floating-gate. Note that at the bottom of the energy barrier, the oxide thickness is large enough (greater than 10 nm in a 0.5 μm CMOS process) that the

probability of electrons directly tunneling through is negligible. Thus, FN tunneling can be modeled by an equivalent energy scavenging circuit where the input energy source is the ambient thermal-activation or thermal-noise and rectification diode is formed by the tunneling barrier whose output is the floating-gate capacitance. Mathematically the combination of thermal activation and electron tunneling can be expressed by the FN tunneling current density J (A/m^2) as [28]:

$$J = \frac{q}{h}\gamma \int_{-\infty}^{\infty} P_T(\zeta)T(\zeta)d\zeta \tag{2.6}$$

where $P_T(\zeta)$ is the probability density function corresponding to an electron occupying an energy level ζ and $T(.)$ represents the tunneling probability of the electron and is a function of the barrier thickness. The parameters h and q correspond to the Plank's constant and charge of free electrons respectively. γ represents a transmission parameter that is a function of the interface properties. In its general form shown in equation 2.6, it is practically impossible to obtain a closed form expression for FN tunneling current density J, let alone solve a coupled differential equation involving J. Therefore, for the sake of simplicity we will ignore several second order effects (e.g. effect of image force and temperature variations) and consider the following mathematically tractable form of J[56]

$$J = \alpha E^2 \exp\left(-\frac{\beta}{E}\right) \tag{2.7}$$

where E represents the electric field across the oxide-barrier and where the parameters α and β are a function of the material properties and are given by

$$\alpha = \frac{mq^3}{8\pi m^* h\phi} \tag{2.8}$$

$$\beta = \frac{4(2m^*)^{\frac{1}{2}}\phi^{\frac{3}{2}}}{3hq} \tag{2.9}$$

with m^* being the effective mass of electrons in the forbidden gap of the silicon-di-oxide, m being the mass of a free electron, ϕ the barrier height at the interface, and h the Planck's constant.

The core structure of the timer can be modeled as a parallel connection of the tunneling current source and the gate capacitance, as illustrated in Fig. 2.1(b). If the floating-gate capacitance is assumed to be C_T and the cross-sectional area of the tunneling junction is assumed to be A, then the incremental change in floating-gate voltage $V_{fg}(t)$ (equivalently floating-gate charge) can be expressed by the first order differential equation

$$dV_{fg}(t) = \frac{dQ}{C_T} = \frac{AJ(E)dt}{C_T}. \tag{2.10}$$

Under the assumption of a triangular barrier, V_{fg} and E are related by the oxide-barrier thickness t_{ox} as

$$V_{fg}(t) = t_{ox}E(t) + V_{sub} \tag{2.11}$$

where V_{sub} is the effective voltage drop across the n-type substrate. By integrating equations 2.7, 2.10 and 2.11, the dependence of electric field E on time t can be expressed in the form of

$$E(t) = \frac{\beta}{\ln(k_1 t + k_0)} \tag{2.12}$$

where k_0 and k_1 are constants that are given by

$$k_0 = \exp\left(\frac{\beta}{E_0}\right), \; k_1 = \frac{A\alpha\beta}{C_T t_{ox}} \tag{2.13}$$

here E_0 is the initial electric field across the gate oxide. Substituting E into equation 2.11, the floating-gate voltage change over time can be expressed as

$$V_{fg}(t) = \frac{k_2}{\ln(k_1 t + k_0)} + V_{sub} \tag{2.14}$$

where

$$k_2 = \beta t_{ox} \tag{2.15}$$

Fig 2.8(b) illustrates typical responses of the timer according to equation 2.14.

2.3.2 Robustness Analysis

Robustness is one of the key features to successfully implement the timer in passive devices. As illustrated in Fig. 2.9(a), deployed timers on each passive tag should maintain synchronization with respect to the one on the server. This requires the timer device to be highly robust to mismatch. In the derivation of the model expressed by (2.14) we assume a uniform tunneling junction which is not true due to manufacturing variations. The total tunneling current is the sum of the tunneling current over all tunneling junctions, as demonstrated in Fig. 2.9(b). As shown in Fig. 2.9(c), the tunneling junction is formed by the gate oxide of a pMOS transistor and the junction oxide thickness is not uniform. As a result, the oxide thickness is a function of location. The fact that potential drop V_{FN} is identical across the whole junction leads to varying electric fields at different locations. The current through the junction at location (x, y) with thickness $t_{ox}(x, y)$ can be expressed as:

$$J(x, y) = \alpha \frac{V_{FN}^2(t)}{t_{ox}(x, y)^2} \exp\left[\frac{-\beta t_{ox}(x, y)}{V_{FN}(t)}\right] \tag{2.16}$$

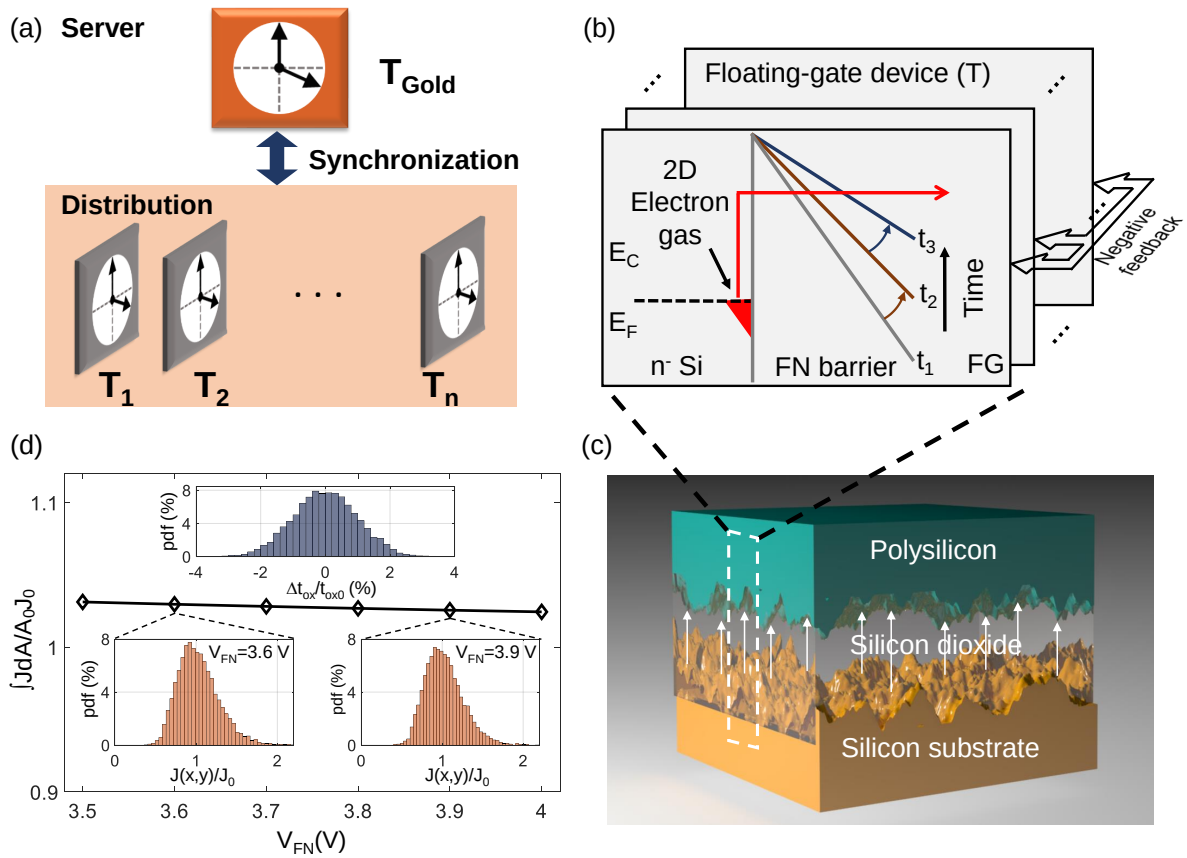


Figure 2.9: Robust analysis of the FN tunneling timers. (a) Requirement of synchronization for passive tags. (b) Interaction and combination of different tunneling junctions. (c) Micro-photograph of a fabricated timer device. (d) Illustration of non-uniform tunneling junction oxide thickness distribution.

Correspondingly, the tunneling current is the integrated value of the distribution as

$$J_{tun} = \int J(x,y)dxdy \tag{2.17}$$

Let t_{ox0} denote the average oxide thickness and let J_0 denote the FN tunneling current density when the potential $V_{\text{FN}}(t)$ is applied to a barrier of thickness t_{ox0}. Then,

$$J(x,y) = J_0 \frac{J(x,y)}{J_0} = J_0 \left(1 + \frac{\Delta t_{\text{ox}}}{t_{\text{ox0}}}\right)^{-2} \exp\left[-\frac{\beta t_{\text{ox0}}}{V_{\text{FN}}(t)} \frac{\Delta t_{\text{ox}}}{t_{\text{ox0}}}\right] \tag{2.18}$$

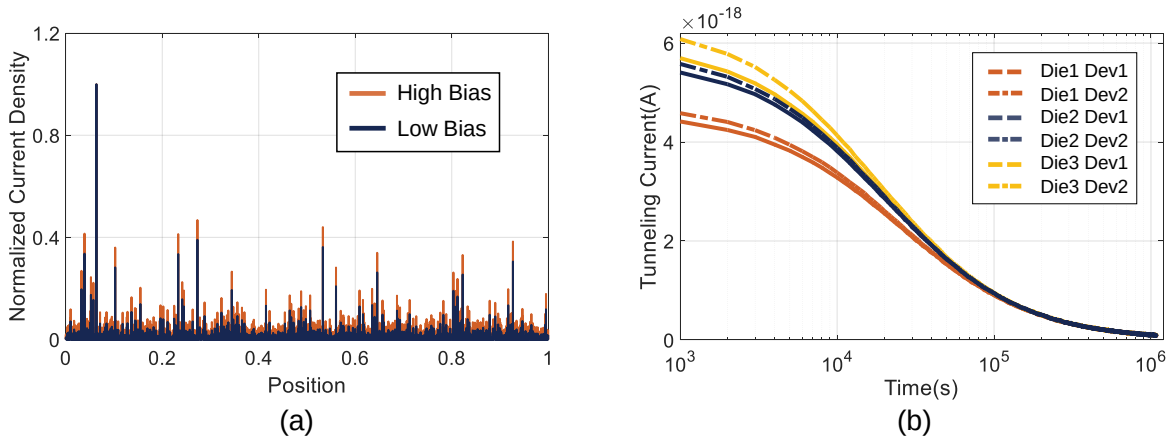


Figure 2.10: Current characterization in FN timers. (a) Relative current distribution under different bias voltages. (b) Measured tunneling currents converge to the same value even with mismatch.

where Δt_{ox} is the deviation of the oxide thickness from the average t_{ox0}. Under the realistic assumption of less than 1% relative standard deviation in the gate-oxide thickness, equation (2.18) can be approximated as

$$J(x,y) \approx J_0 \left[1 - \left(2 + \frac{\beta t_{\text{ox0}}}{V_{FN}} \right) \frac{\Delta t_{\text{ox}}}{t_{\text{ox0}}} \right] \tag{2.19}$$

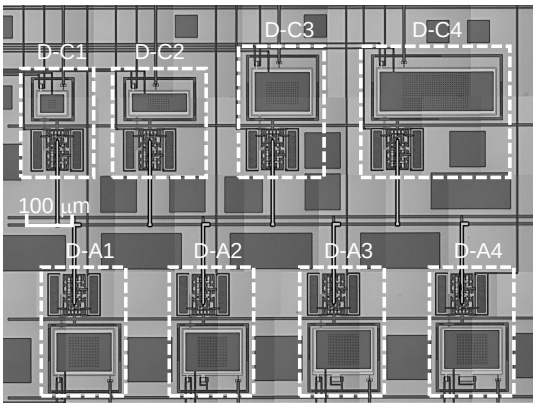
This approximation is valid because in practice the change in V_{FN} will be less than 10% in the region of interest. By integrating $J(x,y)$ over the cross-sectional area A_0, the second item in (2.19) cancels due to property of the error term Δt_{ox}. This implies that the FN tunneling current density can be assumed to be a function of the average oxide thickness t_{ox0}. Fig. 2.9(d) shows the simulated relationship between the current calculated from distributed oxide thickness and average oxide thickness. It is relatively constant across a large dynamic voltage range, indicating that the error can be calibrated with a constant coefficient.

For time intervals $t > t_0 \gg k_0/k_1$, the change in V_{FN} can be estimated by

$$\Delta V_{\mathrm{FN}}(t) \approx k_2 \left(\frac{1}{\ln \frac{t}{\tau}} - \frac{1}{\ln \frac{t_0}{\tau}} \right) \tag{2.20}$$

where τ could be chosen approximately as $1/k_1$ to improve the accuracy of the approximation. Equation (2.20) shows that after a certain time of operation, the response of the FN device shows relatively weak dependence on the initial conditions and fabrication artifacts except for the single parameter t_{ox0}. The temporal response of the timer can be divided into two stages: (a) transient phase where the timers deviate from each other due to mismatch and (b) equilibrium stage where the timers' response becomes parallel (or synchronized) with respect to each other with the relative synchronization accuracy determined only by the mismatch in the average oxide thickness. Given that the gate-oxide thickness is a very well controlled parameter in most modern semiconductor process, it is expected that t_{ox0} will not significantly deviate across different devices.

The robust synchronization behavior can be explained intuitively as a self-compensation mechanism. Assume two identical timers show different tunneling rates initially at the same bias. The one with larger initial tunneling rate will have a faster decrease rate in the floating-gate voltage V_{fg} hence a faster decrease of V_{FN}. As implied by (2.16), the tunneling current will decrease faster accordingly. The device demonstrating faster initial tunneling current will induce an larger acceleration speed in the decrease of the current itself, leading to converging tunneling rate. This self-compensation mechanism functions as a negative feedback in the device to maintain an identical tunneling rate across timers. Fig. 2.10(b) shows the extracted tunneling current from measured response of 6 timers with identical structure on three different dies. The 6 timers were programmed to the same initial value. The tunneling current of each timer converges to the same value and get synchronized as expected.



Parameters	Capacitance (pF)	Tunneling Area (μm²)
D-A1	2	72
D-A2	2	108
D-A3	2	144
D-A4	2	180
D-C1	2	54
D-C2	4	54
D-C3	8	54
D-C4	16	54

Figure 2.11: Die photograph of the fabricated FN timers and their form factors.

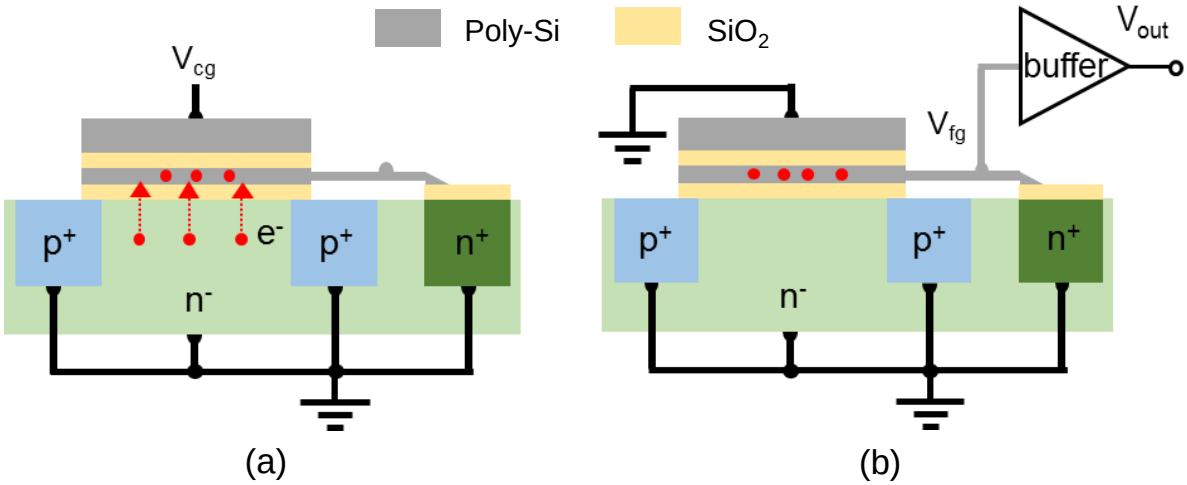
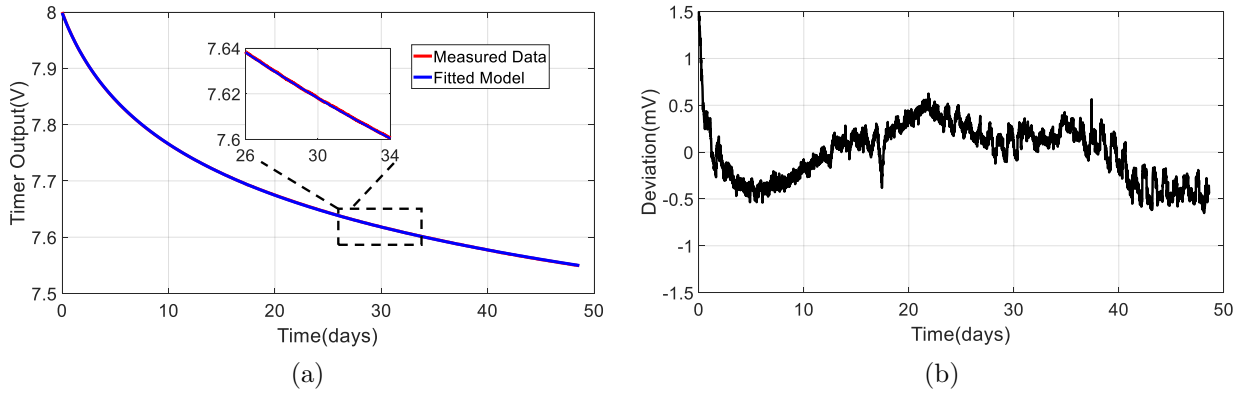


Figure 2.12: Operational modes of FN tunneling timers: (a) self-powered timing mode and (b) readout mode.

2.3.3 Measurement Results

To characterize the behavior and performance, totally eight variants of FN timers were fabricated on a standard 0.5-μm double-poly CMOS process and the micro-photograph of the die is shown in Fig. 2.11, where the form factors of the timers are also summarized. The device can be initialized to operate in different modes by programming the charge on the floating-gate. While it is easy to program the timer into FN tunneling region where



(a) (b)

Figure 2.13: Measured timer response and comparison with the behavioral model: (a) measured response and the fitted model, (b) deviation between the data and model.

electrons can continuously tunnel through the gate oxide, the readout circuits need to be carefully designed to avoid interference with the FG potential. For this purpose, we used a modulation approach which is illustrated in Fig. 2.12(a) and (b). Initially, we programmed the FG node voltage to a value around 3 V. If we want to activate the FN tunneling timing process, we connect V_{cg} to a fixed voltage such as 6 V (as shown in Fig. 2.12(a)). Since the pMOS transistor is biased at accumulation mode, all the capacitors can be assumed to be constant and FG voltage depends linearly on V_{cg}. To measure the FG potential, we connect V_{cg} to ground, which pulls the FG voltage below 3 V and hence can be interrogated using a standard unity-gain buffer, as illustrated in Fig. 2.12(b). The buffer input is directly connected to the FG node using the same polysilicon layer which avoids the use of metallic vias that might introduce trap states in the surrounding oxide.

Model Verification

The first set of experiment were conducted to verify the proposed model of the timer device. The timer was put in an environmental chamber maintaining a constant temperature of 20 °C. We sampled the output of the timer every 1000 seconds for a total duration of 4.2×10^6

seconds (approximately 48 days) and the measured response is plotted as the red line in Fig. 2.13(a). After obtaining the data, we used the model expressed by (2.14) to fit the data and the result is marked as blue in Fig. 2.13(a). The parameters fitted from the measured response are summarized in Table. 2.1. To analyze the model accuracy, the deviation between the measured results and the fitted model was calculated and plotted in Fig. 2.13(b). The maximum deviation is calculated to be less than 1.5 mV across a dynamic range larger than 400mV, implying model accuracy better than 46 dB. The results validate the efficacy of the proposed model for timer characterization. Unlike TAT timers, FN tunneling timers show more predictable and clean response. Moreover, the accurate model that can capture the dynamic response makes it feasible to recover the time using measured output.

Long-term Operation Validation

As indicated by model (2.14), the proposed timing devices demonstrate asymptotic $\frac{1}{\ln t}$ behavior which is neither like a R-C timer which saturates very quickly, nor like direct tunneling timers that only depends on the oxide thickness which is also very fast. Therefore, it demonstrates the potential to support long-term operation. In this experiment, we investigate the capability of using FN tunneling timers for long-term robust time-keeping.

While it is not realistic to conduct experiments running for years to verify the timer's performance, we use an alternative way to do this. The timer was biased at three different stages corresponding to different time ranges, with each stage running up to two weeks. The

Table 2.1: Model timer parameters estimated using measured data

Parameter	Value
k_2	83.97
k_0	1.103×10^9
k_1	3.294×10^3
V_{sub}	3.966

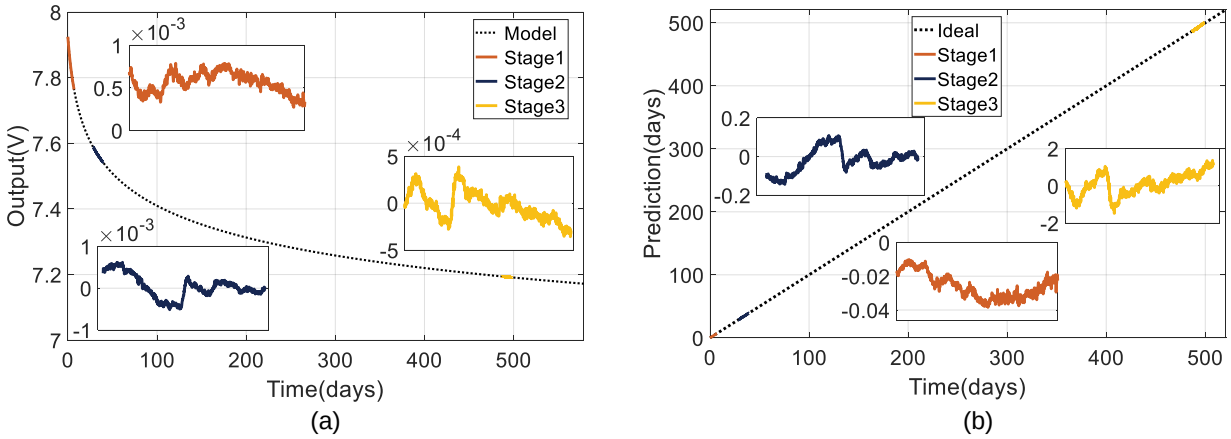


Figure 2.14: Interpolation of three stages to extrapolate the long-term response of the FN timer: (a) measured response, (b) time prediction.

response was recorded for each stage and the first stage was used to fit the model parameters. By checking the model accuracy over all the three stages, we are able to verify whether the timer can last for years.

Fig. 2.14(a) shows the measured results and the fitted model, where the deviation between the two is also plotted in the insets. Two observations can be obtained from the results: (a) the magnitude of the residual error shown in Fig. 2.14 for each of the regions is similar, highlighting that noise-floor is determined by measurement or buffer noise. This fact indicates that the model captures the dynamic behavior of all the three stages. (b) The third stage is biased at the operational range up to 1.5 years, which still matches the model well. The prediction using the fitted model from the measured results is illustrated in Fig. 2.14(b). Since the timer shows saturation behavior in time domain, the absolute prediction error increases when the monitoring period is longer, but the relative error still maintains the same level. Therefore, FN tunneling timers are feasible for robust time-tracking as long as several years and the model can dedicatedly capture the dynamic behavior within this range.

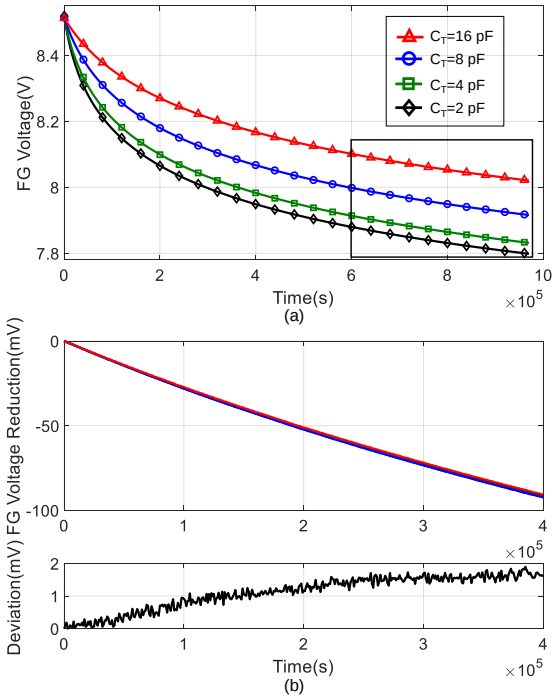


Figure 2.15: Comparison of timers for different gate capacitances C_T with tunneling area A=54 μm^2. (Marked points represent measured data and solid lines correspond to the behavior model.)

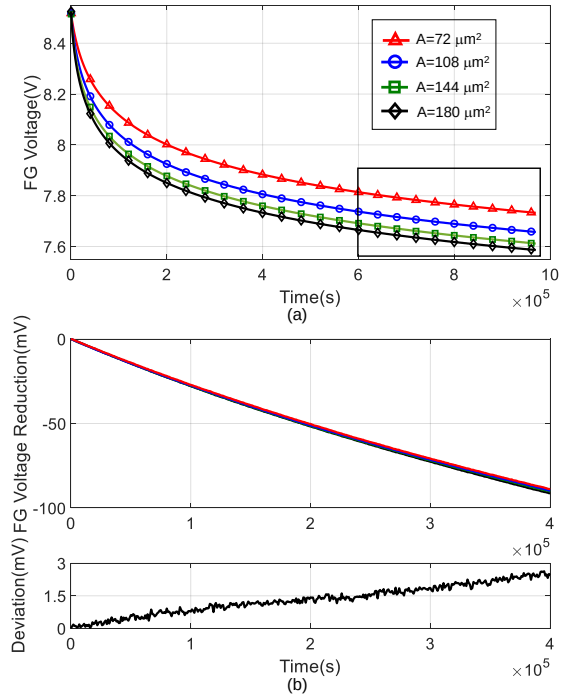


Figure 2.16: Comparison of timers for different tunneling junction areas with C_T= 4 pF. (Marked points represent measured data and solid lines correspond to the behavior model.)

Robustness and Mismatch Characterization

In this set of experiments, we verified the robustness of the timer's response to nonideal factors such as mismatch and temperature variations. The first group of experiments were designed to verify the timer responses for different values of: (a) floating-gate capacitances; and (b) tunneling junction areas. Fig. 2.15 shows the measured values (highlighted by marked points) for floating-gate capacitances of 2 pF, 4 pF, 8 pF and 16 pF respectively. For this experiment, the tunneling junction areas for all the four timers were chosen to be 54 μm^2. The corresponding timer responses estimated using the behavioral model for each of the parameters (FG gate capacitance and tunneling junction area) is also plotted as a solid-line overlaying the measured data in Fig. 2.15(a). Two observations can be inferred from

the measured results: (a) the behavioral model can accurately predict the response of the fabricated timers for different device parameters; and (b) a smaller capacitance produces a faster change in the timer response. As indicated by equation 2.20, after the initial differences in respective timer responses, all the timers change in a nearly identical fashion. This is shown in Fig. 2.15(b) which plots the change in the timer outputs measured with respect to the output measured at a reference time ($t_o = 6 \times 10^5$ s or 167 hours). The result show that the response changes by less than 2% even if the capacitances changes by more than 800%. A similar experiment was conducted for timers with different tunneling junction areas of 72 μm^2, 108 μm^2, 144 μm^2 and 180 μm^2 and for a fixed gate capacitance of C_T=4 pF. The measured result is shown in Fig. 2.16 demonstrating a similar trend as before where a larger junction area produces a larger initial change in the timer responses; but after a reference time, all the timers exhibit a nearly identical response. This is verified in Fig. 2.16 which shows the change in timer values (from the value measured at a reference time instant) for different tunneling junction areas. The measured results again validate the robustness of the timing device exhibiting less than 3% variation for junction areas that can vary by more than 100%.

In the next set of experiments we verified the mismatch in the responses of identical timers fabricated on different silicon dies. For this experiment all the timers were simultaneously programmed to "approximately" the same initial voltage and the set up was housed in an environment with similar conditions (temperature and humidity). Fig. 2.17(a) shows the measured timer responses which as expected shows a similar trend, where the initial timer responses vary due to mismatch, but then the responses reach an identical steady state response. Fig. 2.17(b) show the relative deviation in the timer value with respect to each other and the measured result exhibits less than 500 μV variation over a 100 mV operating span. This amounts to a synchronization accuracy greater than 46dB. Note that for this

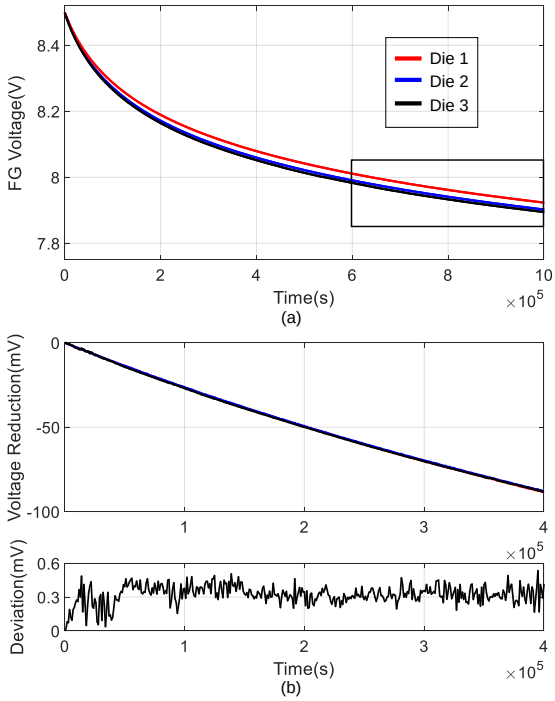


Figure 2.17: Comparison of the measurement results from timers with the same structure on different dies with gate capacitance of C_T= 8 pF and tunneling junction area of A= 54 μm^2.

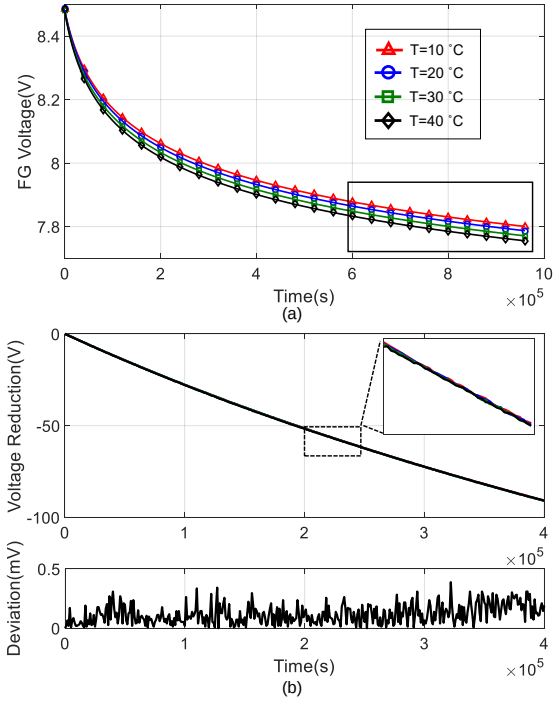


Figure 2.18: Measured dependence of timer's characteristics on temperature with C_T=2 pF and A=54 μm^2.

experiment, the noise and drift in the read-out circuits could also affect the measured result; therefore it is possible that the synchronization accuracy could be higher than what has been measured.

The final group of experiments were conducted to measure the temperature dependence of the timer. A fabricated timer with C_T=2 pF and A=54 μm^2 was housed in a temperature controlled environment chamber and timer responses were obtained for four different temperature settings: 10 °C, 20 °C, 30 °C and 40 °C. The measured responses are shown in Fig. 2.18 and as expected the initial timer responses show a faster rate with increase in temperature. This is because for FN tunneling, electrons have to be thermally excited to cross the triangular FN barrier. However, the long-term response of the timer exhibits

a self-compensating effect similar to that of the timers with different device parameters. Fig. 2.18 shows the measured change in timer responses after a reference time of 6×10^5s, and the FG voltage reduction shows a mismatch less than 400μV across a range of 100 mV. The measured results demonstrate that two timers maintained at two different temperature levels (measured range of 30 ^oC) can still be synchronized with respect to each other up to an accuracy of 0.5%.

Technology Scalability

While integrated passive devices can be implemented using diverse technologies to leverage the performance and cost, we want to verify whether the proposed device is consistent with the scaling of CMOS technologies. To do so, we fabricated the designs on a standard 180 nm CMOS process with the same circuit topologies yet different oxide thicknesses. Fig. 2.19 shows the measured response from the fabricated timers with two oxide thicknesses. As illustrated by Fig. 2.19(a), the timer with an average oxide thickness of 12 nm shows temporal response that matches with model as well as our previously presented timers on 0.5 μm CMOS process with 13 nm oxide thickness. For the other timer with 3.5 nm oxide thickness, while the measured response still matches well with the model, deviation is relatively larger because the oxide thickness is so small that the direct tunneling through the rectangular energy barrier starts to affect the tunneling process. Therefore, the general rule of applying the proposed timing device in different technologies is that thick oxide options are available, which is true because most of the advanced technologies provide I/O transistors with thick oxide for ESD protection and I/O modules design. In conclusion, the FN tunneling timer device is scalable to more advanced CMOS processes.

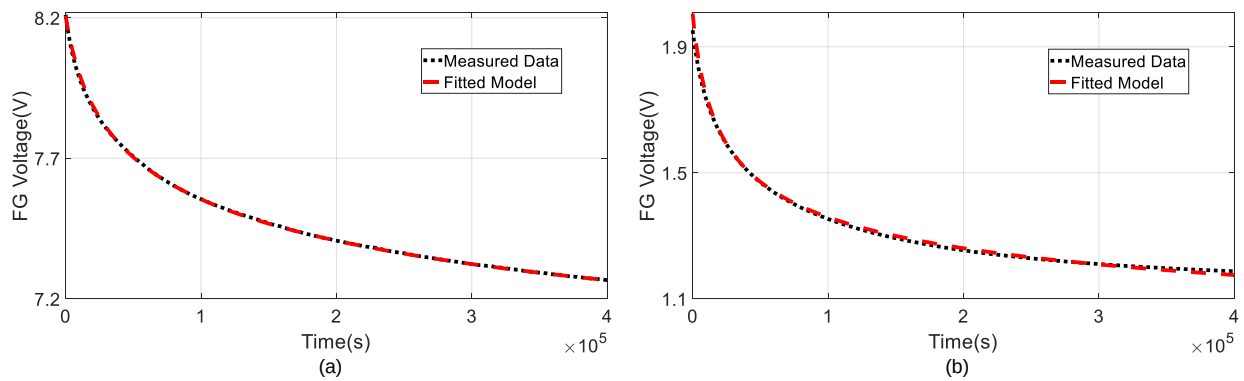


Figure 2.19: Measured results of timers fabricated on 180nm process with oxide thickness: (a) 12 nm, (b) 3.5 nm.

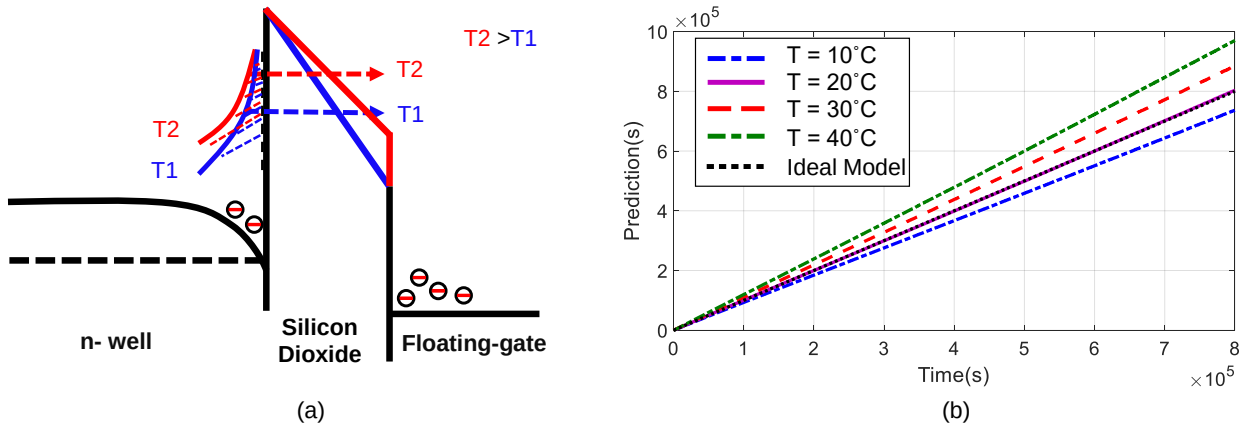


Figure 2.20: Temperature impact on timer's behavior: (a) temperature affects the electron's distribution, (b) temperature impact on time prediction accuracy.

2.3.4 Timing Accuracy Analysis

Timing accuracy is the key performance in implementing the proposed device. Due to nonideal factors during operation, the accuracy can be degraded. Error sources such as measurement noise could affect the output of the timer and affect the time recovery precision. In this section, major sources of error is discussed and analyzed.

Model Inaccuracy

In the derivation of the FN timer model, high-order nonideal factors were neglected, of which the barrier lowering and temperature effect impact the timer behavior most. They will induce model inaccuracy, therefore degrade the time-prediction accuracy. A more complete expression for the FN tunneling current density J that captures these effects is given by [56]

$$J = \alpha \frac{1}{w^2(y)} \gamma(T) E^2 \exp(-\frac{\beta v(y)}{E}) \tag{2.21}$$

which includes an explicit dependence on temperature through $\gamma(T)$ and two correction terms $w(y)$ and $v(y)$ which captures the lowering of the triangular tunneling barrier through an image force effect [70]. Here T represents the ambient temperature. These correction terms are tabulated elliptic integrals, and y is a function of the barrier height and electric field as

$$y = \frac{1}{\phi}(\frac{q^3 E}{4\pi\epsilon_r\epsilon_0})^{\frac{1}{2}}. \tag{2.22}$$

Although the tunneling process itself is temperature independent, the number of electrons of a given incident energy on the barrier is a function of temperature and the barrier height ϕ also depends on temperature. The dependence of electron momentum distribution on temperature as shown in Fig. 2.20(a) can be corrected using $\gamma(T)$ which is given by

$$\gamma(T) = \frac{\pi c k T}{sin(\pi c k T)} \tag{2.23}$$

where

$$c = \frac{4\pi(2m^*\phi)^{\frac{1}{2}} t(y)}{hqE} \tag{2.24}$$

Ravindra [70] proposed a linear dependence model to accommodate the temperature effect on the barrier height, and can be expressed as

$$\phi(T) = \phi_0 + \lambda \Delta T \tag{2.25}$$

where λ is a constant parameter. Notice that the barrier height ϕ should be replaced with the temperature dependent form $\phi(T)$.

In general, high temperature will cause accelerated tunneling in the timer. When using a model trained at predetermined temperature condition, the increase of temperature leads to over counting of the time value. Fig. 2.20(b) characterizes the impact of temperature on timing accuracy. The time are predicted using the model derived from the measured results at a constant temperature of 20 °C. The results show that the variations in the predicted time due to small temperature fluctuations is linear, thus small statistical variations in temperature can be averaged out. However, long-term temperature drifts cannot be averaged out and temperature drift of one degree will introduce approximately one percent error in the timing accuracy.

Measurement Inaccuracy

Measurement can introduce noise in two ways. On one hand, the measurement interface will change the timer output. For instance, the unity-gain buffer used in this work directly connects the output node of the timer to the input of the buffer, which is the gate of MOS transistor. The source-to-gate voltage V_{sg} of the input MOS transistor will alter the potential on the timing node, affecting the timing accuracy. On the other hand, the measurement circuit itself is an error source which results from nonideal factors such as thermal noise, offset, finite gain, finite CMRR and PSRR.

For the first type of measurement noise, it can be minimized if we can ensure the coupling from the measurement circuit is a constant value. In such scenario, it functions as a constant DC offset to the timer output. This offset will be taken into the model parameter V_{sub} in equation 2.11 and does not affect time prediction accuracy. However, in practical case, since the output of the timer changes dynamically with respect to time, this coupling value shows variations, and its amplitude depends on the measurement circuit topology and the bias.

The second type of measurement noise, is usually more significant compared to that of the first type, because this noise source is random and show time dependent features. For instance, thermal noise in the channel of MOSFET transistors will directly be coupled to the output of the buffer. The input offset of the buffer amplifier depends on temperatures and demonstrates time-varying features during the operational period. With a constant temperature, this type of noise is relatively constant across the whole lifetime and is the major limiting factor that determines the time recovery accuracy. As indicated in Fig. 2.14, the measurement noise is relatively constant, however, due to the saturating behavior of the timer output, the absolute time prediction accuracy degrades when the timer has a lower tunneling rate. It should be minimized when designing the measurement interface.

Fundamental Limits

As the model indicates, the timer shows asymptotically saturating behavior captured by $\frac{1}{\log t}$. Assume the model and measurement interface does not introduce any inaccuracy into the time prediction, then the accuracy will be determined by the statistical tunneling process. Assume the tunneling rate follows a time-dependent statistical distribution using the number of electrons as

$$R_e(t) \sim f(\mu(t), \sigma(t)) \tag{2.26}$$

where $\mu(t)$ and $\sigma(t)$ are the average tunneling rate and standard deviation at time instance t, here the tunneling rate is expressed as the number of electrons per second. While $\mu(t)$ represents the average value and is stable across the operational life, due to the discrete nature of electrons, $\sigma(t)$ will introduce fluctuations into the timing process. The fundamental limit is determined by the discrete and random nature of tunneling together with the average tunneling rate.

To characterize the impact of this statistical process on the timing accuracy, we formulate the relationship between tunneling rate and time in a different way. The time scale is normalized to the mean duration duration of single electron tunneling, which can be indicated as

$$t_n = \frac{1}{\mu(t)} \tag{2.27}$$

Now we discuss the random tunneling process within time frame t_n. When the tunneling rate is small enough, we can safely assume the tunneling rate does not change within the time frame of t_n, therefore tunneling of each electron is independent with each other. The number of electrons tunneling within time frame t_n then can be approximately using a Poisson distribution with expectation of 1 as [49]:

$$P(n = k) = \frac{1}{ek!} \tag{2.28}$$

According to the properties of Poisson distribution, its standard deviation $\sigma'(t)$ equals to its mean value. If we use the standard deviation as the expected error in accuracy, then we can calculate the inaccuracy at time t as

$$t_{err}(t) = \sigma'(t) * t_n = \frac{\sigma'(t)}{\mu(t)} = \frac{1}{\mu(t)} \tag{2.29}$$

Equation 2.29 indicates that in ideal scenario, the timing accuracy is limited by the tunneling rate. When the FN timer enters into the single electron tunneling region, its timing accuracy basically is determines by the tunneling rate itself.

2.4 Summary

In this chapter, the topologies of possible timing devices were studied and explored. Floating-gate structures are used to store charge due to their good charge retention features. Multiple leakage mechanisms in standard CMOS process were examined and quantum tunneling happened across oxide layers are found to be feasible to discharge the timer. The properties of self-powered timers based on trap-assisted tunneling and FN tunneling were extensively studied. TAT timers that take advantage of the leakage caused by trap states were first verified and it was shown that although they support long-term operation, the dynamic behavior is intractable and difficult to control, because the distribution of the trap-states depends on fabrication process and demonstrates randomness. On the contrary, FN tunneling is found to be more steady across the operational period. A mathematical model was derived by combining FN tunneling physics and the circuit model, and the model was validated to have accuracy greater than 46 dB. The origin of the robust operation was analyzed in detail and the synchronization performance was shown to be better than 0.5%. The technology scalability was further verified by fabricating the prototypes on more advanced technology nodes. Long-term study was conducted to verify that the device is feasible for operational life in the range of several years, validating its usage in passive IoT devices.

Chapter 3

Dynamic Signature Based on FN Timers

The ability to precisely synchronize and desynchronize two spatially distributed dynamical systems according to the changes in their respective operating environment provides a powerful mechanism for authentication and verifying trust. When such synchronization can be achieved at a micro/nano-scale and without any external powering, the approach could be used to identify every product in a supply-chain and to determine if the product has been subjected to any abnormal environmental conditions. In this chapter, we use the self-powered floating-gate timers to implement such a spatially distributed and synchronized dynamical system. We will demonstrate that the self-powered FG device can be used to differentiate between supply-chain trajectories, each of which has been subjected to different temperature variations or variations in its ambient radio-frequency environment. The proposed solution could provide a cost-effective and scalable solution for authenticating many different types of passive assets ranging from credit cards, packaged chipsets, to pharmaceuticals.

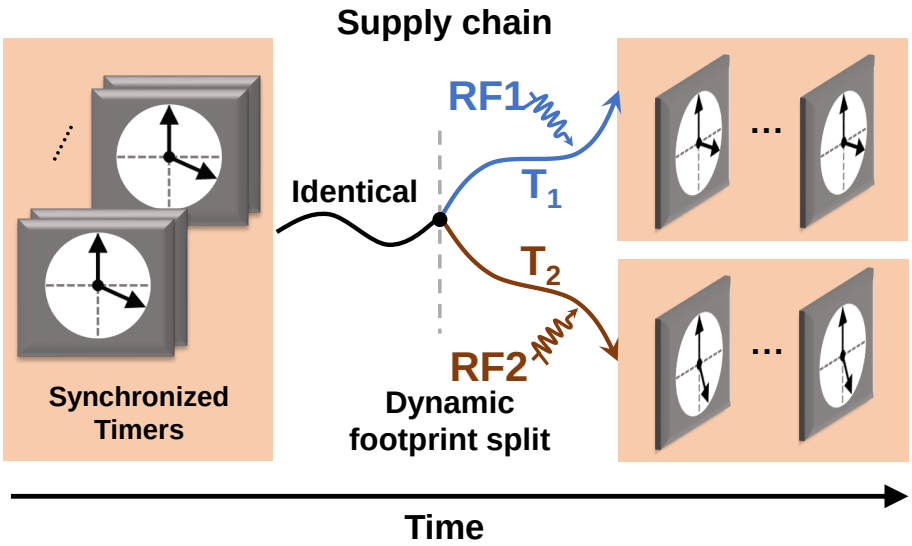


Figure 3.1: Synchronized timers can be desynchronized when they go through different dynamic footprint.

3.1 Principle of Dynamic Signature

3.1.1 Timer Desynchronization Model

Synchronized timers will follow the same dynamic path if they go through identical trajectories. However, if there is any deviation in the environment, it is possible for them to desynchronize from each other. As illustrated in Fig. 3.1, the radiation and temperature deviation in the environment of the two coupled timers will lead to different behavior in the dynamic response. Generally, the desynchronization source can be modeled as a modulation voltage source V_m connected to the modulation terminal V_T as shown in Fig. 2.12. Since the tunneling current is a function of the potential drop across the oxide barrier V_{FN} indicated by (2.16), with modulation V_m, V_{FN} will be boosted as

$$V'_{FN} = V_{FN} + \gamma V_m \tag{3.1}$$

where γ is capacitive coupling coefficient and we assume it to be 1 for simplicity, V_{FN} is the potential drop across the oxide layer of the tunneling junction. By replacing V_{FN} with V'_{FN} in (2.16) and assuming a uniform oxide thickness, we have

$$\frac{A\alpha}{Ct_{ox}^2}(V_{FNm}+V_m)^2\exp(-\frac{\beta t_{ox}}{V_{FNm}+V_m})dt = -dV_{FNm} \tag{3.2}$$

where V_{FNm} is the new potential drop across the oxide induced by the residual charge on the FG node. Considering the case where $V_m \ll V_{FNm}$, the equation can be simplified as:

$$\frac{A\alpha}{Ct_{ox}^2}\exp(\frac{k_2V_m}{V_{FNm}^2})dt \approx -V_{FNm}^{-2}\exp(\frac{k_2}{V_{FNm}})dV_{FNm} \tag{3.3}$$

It is reasonable to use V_{FNm0} to replace V_{FNm} in the left-hand side to further simplify the equation. Then we have

$$k_1\int\exp(\frac{k_2V_m}{V_{FNm0}^2})dt = \exp(\frac{k_2}{V_{FNm}}) - \exp(\frac{k_2}{V_{FNm0}}) \tag{3.4}$$

If $V_m = 0$, (3.4) will reduce to (2.14). If V_m is a AC signal with small amplitude and has the form of $V_m = v_m\sin(\omega t)$, substitute it into (3.4) we have

$$k_1\int(1+\frac{k_2v_m\sin(\omega t)}{V_{FNm0}^2})dt = \exp(\frac{k_2}{V_{FNm}}) - \exp(\frac{k_2}{V_{FNm0}}) \tag{3.5}$$

If the period of the modulation signal is comparatively small compared with the operational period of the timer, (3.5) will again reduce to (2.14). This implies when neglecting the second-order nonlinearity, symmetric AC modulation signals can be canceled out and does not impact the timer's behavior.

When V_m is constant, (3.4) can be simplified as:

$$\Delta V = V_{FN} - V_{FNm} \approx \frac{k_1 V_m t}{k_1 t + k_0} \tag{3.6}$$

where V_{FN} is the timer's response without modulation but with exactly same structures and initial conditions. This is consistent with our previous self-compensation analysis. ΔV will converge to V_m with t large enough and compensate the modulation signal.

From the above analysis, the timer device functions as an integrator and therefore a low-pass filter for the modulation signal. High frequency signals will be filtered out and have no effect on the timer's output, while DC signals can be transferred to the output without loss assuming the integration time is large enough.

3.1.2 Desynchronization Sources

If two timers are identical, they will show exactly the same temporal response provided the same initial conditions and environment conditions. However, the mismatch due to fabrication variations cannot be avoided and will affect the response. It can be approximated using a modulation signal V_{mis}, and the amplitude depends on the variance in the mismatch. Generally, a larger mismatch will result in a large equivalent signal V_{mis}.

While timers going through the same environmental conditions can maintain synchronized, the synchronization can be broken if two timers go through different trajectories where the ambient factors including thermal variance, dispersive RF signals and light radiations are different. The temperature impact on the timers' behavior was analyzed in [96]. It mainly altered the tunneling current in two ways. First of all, the energy distribution of the electrons in the substrate is a function of temperature which can be characterized by Boltzmann distribution, leading to temperature dependent tunneling current. The second factor that

affects the tunneling probability is the barrier height which is also a function of temperature. While the dependence of the tunneling current on temperature can be characterized by a mathematical model [96], in this work, we model the temperature impact as a modulation signal V_{mT} expressed as a temperature-controlled voltage source:

$$V_{mT} = f(T - T_0) + V_{m0} \tag{3.7}$$

V_{m0} is the value of modulation signal at T_0, and $f(\bullet)$ is the function describing the relationship. In the simplest scenario, we can use a linear model to approximate the relationship as

$$V_{mT} = \zeta(T - T_0) + V_{m0} \tag{3.8}$$

where ζ is the temperature coefficient[96] assuming $\gamma = 1$. Since V_{mT} is a function of ambient temperature, two synchronized timers going through ambient environment with different temperature variations could be desynchronized.

Another technique to desynchronize two timers is to employ energy transducers which can harvest environmental energy and convert it into an electrical signal. The electric signal is then used to modulate the timer's behavior. Since it is impossible for two trajectories to share exactly the same environment, it is highly unlikely the electrical signals are identical for the two timers. One pervasive energy source in the environment is the radio-frequency radiation. The energy required for modulation has been shown to be less than 1 pJ[104], which is so low that it basically functions as an open circuit to the RF antenna. Therefore, even dispersive RF sources should be capable of altering the timer behavior. RF signals need to be rectified before driving the timer. The voltage has a relationship with the power as

$$V_{mrf} = \sqrt{2R_e\lambda P} \tag{3.9}$$

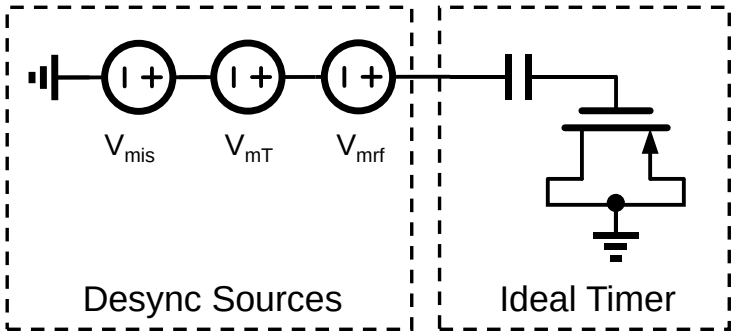


Figure 3.2: System model of the timer with multiple desynchronization sources.

where P is the received power at the antenna, λ is the power loss coefficient and R_e is the effective load resistance.

Combining all the desynchronized sources, the timer can be modeled as an ideal timer without any mismatch maintained at an ideal environment, modulated by the nonideal sources, as illustrated in Fig. 3.2. The amplitude of the modulation signal that takes into all the three aforementioned desynchronization sources can be expressed as:

$$V_m = V_{mis} + V_{mT} + V_{mrf} \tag{3.10}$$

Considering the fact that modulation of the timer's behavior requires extremely low power and low input amplitude, most of the transducers that can harvest ambient energy into electrical form can be used to desynchronize the timers. For instance, piezoelectric transducers can convert mechanical variations into electrical signals and be used to modulate the tunneling behavior and desynchronize the timers. Other types of signals from transducers such as photodiode and thermocouple are also feasible for desynchronizing the timers. The selection of transducers depends on the type of sensing target.

3.2 Experimental Validation

3.2.1 Synchronization Performance

To characterize the inter and intra-die mismatch, totally three chips with two identical timers on each were used for validation. The 6 timers were programmed to the same initial voltage of 8.4 V and placed at the room environment. Outputs of the timers were monitored periodically every 1000 seconds for more than 11 days. In ideal case, all the timers should show identical response, however, due to mismatch, they will deviate from each other. The difference between the timers were extracted and used to fit model 3.6 and the recorded response of each timer were presented in Fig. 3.3(a). During transient stage each timer follow different trajectories, as seen in Fig. 3.3(a), and get synchronized after the transient stage. The self-compensation mechanism ensures the timers enter into this equilibrium stage (synchronization) gradually where the response across timers is almost parallel to each other as highlighted in Fig. 3.3(a). To characterize the synchronization performance, we extracted the voltage reduction and the results were illustrated in Fig. 3.3(b). The responses are close to each other and the maximum absolute deviation across the 6 timers is shown in Fig. 3.3(c), which is smaller than 0.8 mV with a dynamic range larger than 80 mV, indicating a synchronization accuracy larger than 40 dB, which is consistent with the conclusion in Chapter 2.

As discussed in last section, the mismatch can be modeled as a constant modulation signal coupled to the floating-gate node. To verify this, we took the output deviation between two timers, device 1 on die 1 and device 1 on die 3 respectively, and the sampled output is plotted in Fig. 3.3(d) marked as circles. Model 3.6 is then used to fit the result which is also plotted as solid line in Fig. 3.3(d). The model fits with the measured data well and the mismatch between the two devices can be modeled as an equivalent voltage source of 75.8 mV coupled

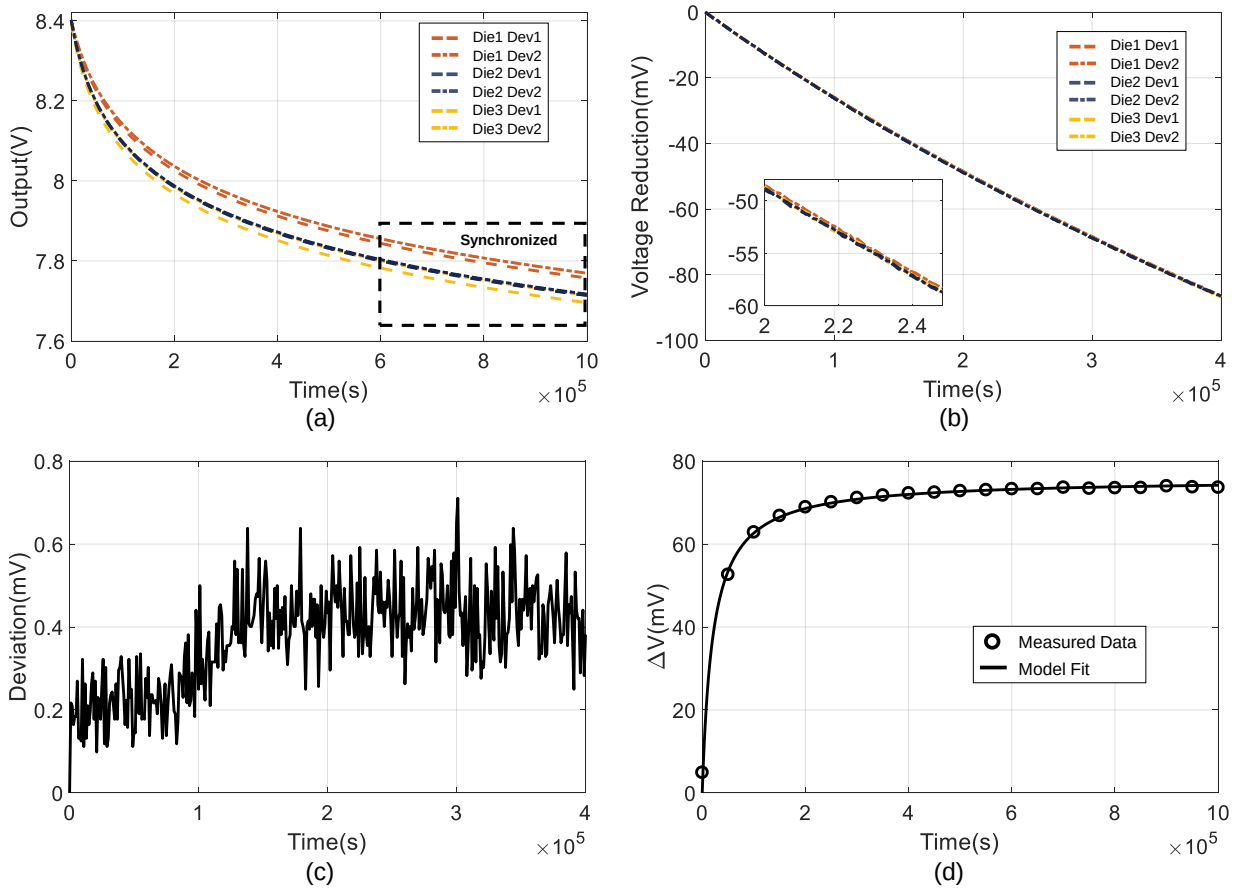


Figure 3.3: Characterization of desynchronization due to mismatch.(a) Measured response of 6 identical timers on 3 dies. (b) Synchronized response at the equilibrium region marked in the rectangular region in (a). (c) Maximum deviation across all the timers in equilibrium region. (d) Measured and fitted temporal evolution of the desynchronization due to mismatch.

to the floating-gate node. This relationship is also validated with the mismatch between all other pair of timers.

3.2.2 Temperature Signature

Besides mismatch, synchronized timers could be desynchronized if they go through different temperature trajectories, as illustrated in Fig. 3.4(a). For verification, we separate 4 identical timers into two pairs, with the pair of Timer 1 and Timer 2 in the room environment and the pair of Timer 3 and Timer 4 in an environment chamber, whose temperature can be

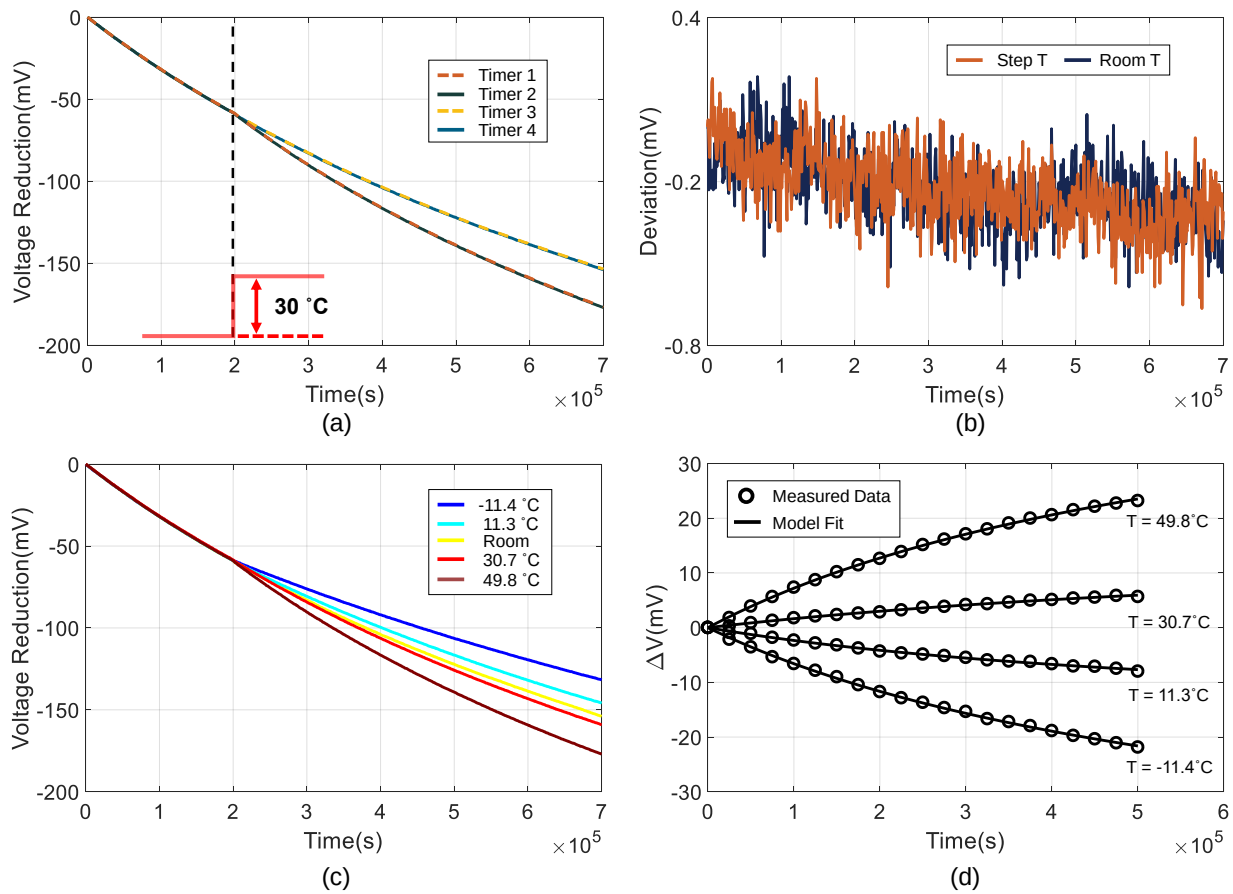


Figure 3.4: Characterization of desynchronization due to temperature. (a) Measured response illustrating that two pairs of synchronized timers are desynchronized by a step temperature change of 30 °C. (b) The deviation within each pairs is less than 0.7 mV indicating good synchronization. (c) Desynchronization response with different step changes in temperature. (d) Measured and fitted response of the desynchronization for different temperatures with respect to room temperature.

controlled accurately. Four timers were programmed to the same initial voltage and all of them were put at the room temperature (average temperature is around 22 °C) for about 8 days. The mismatch will be compensated and they get synchronized with each other due to the self-compensation physics. Then the temperature of the environment chamber was programmed to 50 °C at the time instance of 7×10^5 second. As shown in Fig. 3.4(a), the higher temperature induces a larger tunneling rate, therefore Timer 3 and Timer 4 show faster response and desynchronize with Timer 1 and Timer 2 as expected. However, since Timer

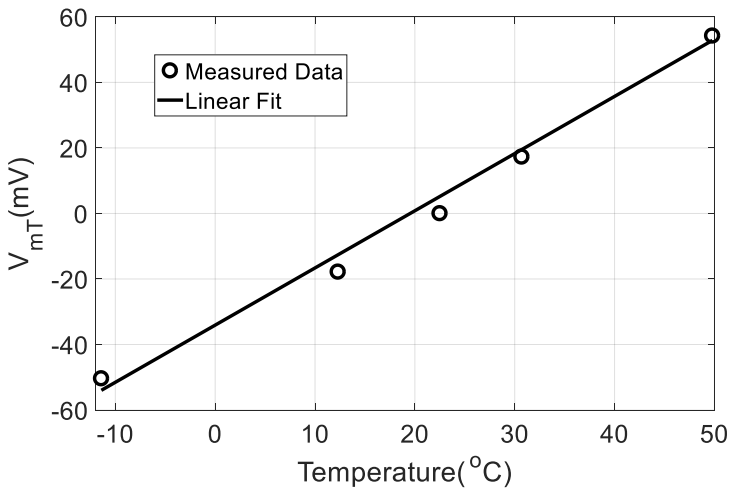


Figure 3.5: Dependence of equivalent modulation voltage on temperature.

3 and Timer 4 still experience the same environmental variations, they keep synchronized across the whole operational period. The deviations within each pair of the two are plotted in Fig. 3.4(b), where the deviation amplitude of the pair with 50 °C maintains at the same level. The temperature variation does not deteriorate the synchronization performance. This experiment validates the conclusion that if two timers go through the identical environmental trajectory, even with mismatch and ambient variations such as temperature, they can maintain synchronization with each other with high accuracy.

The quantitative characterization of temperature impact was conducted by programming the temperature of the environment chamber to different values and measuring the temporal response of the timers. The settings are the same as the previous experiment except that the temperature was programmed to -10 °C, 10 °C, 30 °C and 50 °C respectively. The actual temperature of the environment chamber was measured by an external temperature sensor for accurate calibration. As shown in Fig. 3.4(c), the timers were initially put in the room environment and got synchronized, and then went through a step temperature change. The tunneling rate shows positive dependence on the temperature.

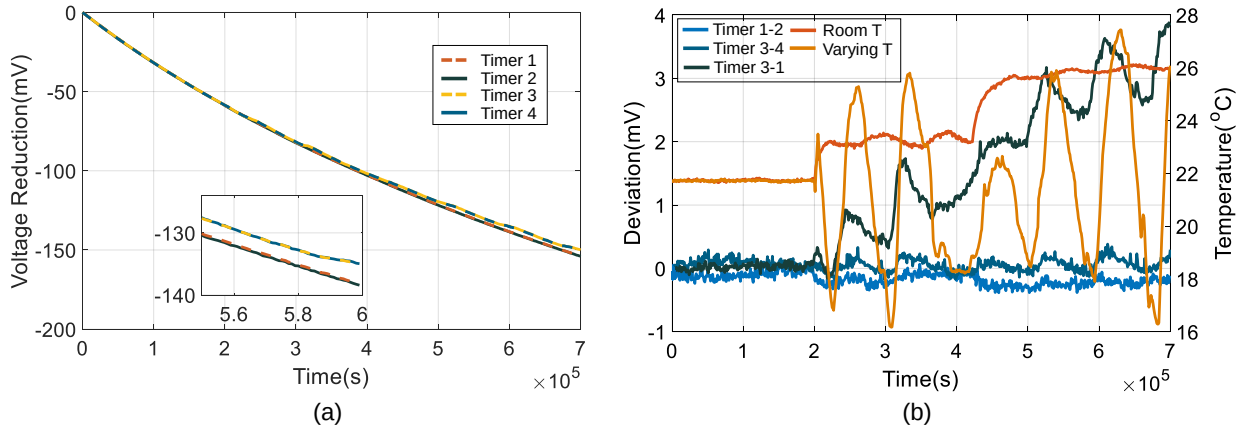


Figure 3.6: (a) Measured synchronization response of the timers in a uncontrolled room environment and a environment chamber with temperature varying in a natural way. (b) The temperature curve in the room and in the environment chamber together with the deviation between different pairs of timers.

To verify the model, we extract the ΔV between the timers in the chamber and in the room environment, and then use model (3.6) to fit ΔV. As shown in Fig. 3.4(d) where the measured data is marked with circle and fitted model is marked with solid line, the model can fit with the data well. The equivalent V_{mT} was extracted from the fitted model and plotted in Fig. 3.5. A first-order linear model was used to fit the dependence of V_{mT} on temperature, validating the models represented by (3.6) and (3.8) with temperature ranging from -10 °C to 50 °C. The temperature coefficient was calculated to be 1.73 mV/°C, which is significantly larger than the synchronization error indicated by Fig. 3.4(b) which is less than 1 mV. This further validates that we can use the desynchronization behavior to detect whether two timers go through identical environmental trajectory.

To evaluate the response in a natural environment, we programmed the temperature in the environment chamber to emulate natural temperature variations. The temperature data was extracted from public online database record of the city St. Louis, MO, USA from July 22nd to July 27th, 2017. Data consists of hourly data and it is fed accordingly to the environment chamber by ramping the temperature, smoothly interpolated between

data points. Experiments were then run in a controlled environment emulating the natural temperature variations and the responses of each timer were recorded by the same procedure mentioned above. Fig. 3.6(a) shows the measured response of the two pairs of timers at the room temperature and emulated environment chamber respectively. The timers were initially synchronized when they went through the same temperature conditions before the time instance of 7×10^5 second, and then got desynchronized because of the environment change in the chamber. The temperature conditions and the deviations between each pair of timers were plotted in Fig. 3.6(b). The pattern of the deviation follows the pattern of the temperature conditions, therefore validating the performance of the timers.

3.2.3 RF Signature

Similar to that of the temperature difference, RF signals can be employed as a source of desynchronization. Similar to the settings of temperature desynchronization, we first eliminated the effect of mismatch. Two chips were put in room environment. Initially, all the 4 timers were programmed to the same initial value and were running for a duration of 7×10^5 seconds when all the timers got synchronized. Then the rectified RF signal was coupled to the two identical timers on one chip. The output was sampled every 1000 second. A stable 915MHz RF source form software defined radio (NI-USRP) is used in our RF experiments. Synchronized timers going through different paths will experience different RF trajectories with high probability. The difference in the RF signal strength can be used to modulate the behavior and desynchronize the timers. To verify the concept, an n-stage charge pump based on Dickson's architecture was employed to harvest RF energy to generate amplified DC signal as plotted in Fig. 3.7(a). Zero-threshold Schottky diodes are used to rectify the low-power ambient RF signals. The transmitter emits a total power of 20 dBm at a distance of 2 meters from the receiver antenna.

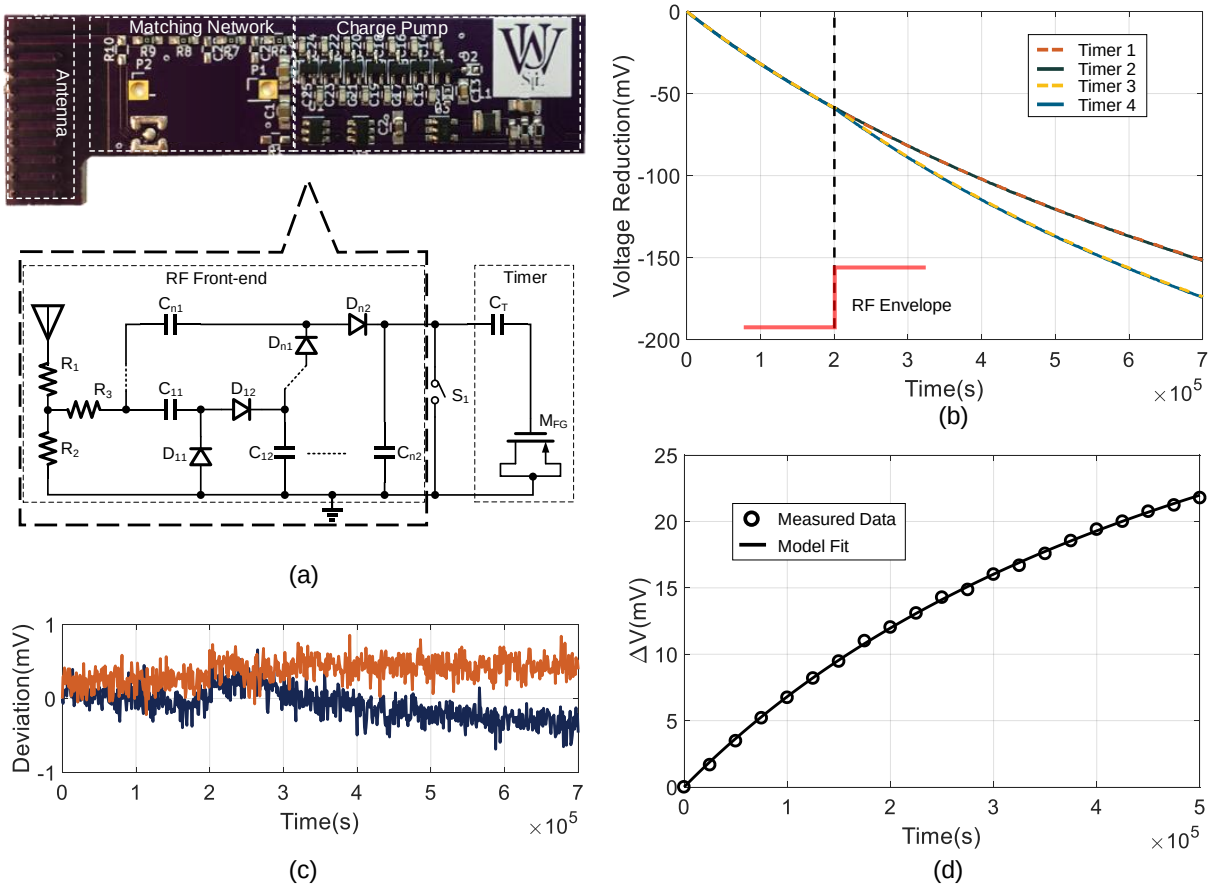


Figure 3.7: Characterization of desynchronization performance with RF modulation. (a) Photo of RF front-end PCB board illustrating the front-end signal rectification and modulation of the timer shown in the inset. (b) Measured response of the timers with and without RF modulation. (c) Synchronization performance within each pair of timers shown in (b). (d) Measured and fitted response of the desynchronization between timers with and without RF modulation.

Four identical timers on two different dies were first synchronized without any RF modulation. Then two of them on the same die was modulated by the output from the first stage of the charge pump for desynchronization. The measured results are illustrated in Fig. 3.7(b), where the four timers are divided into two pairs. Within each pair, the two timers maintain synchronous across the whole monitoring period. However, due to RF modulation, the two pairs start to deviate from each other at the point when the pair of Timer 3 and Timer 4 were modulated by the output from the rectified antenna input. Fig. 3.7(c) illustrates the

deviation within each pair. The deviation does not significantly deteriorate in the case of RF modulation. This further validates synchronous timers can keep synchronized with each other when going through the same environment. To verify the correctness of desynchronization model (3.6) on RF modulation, we use the model to fit ΔV between the timer with and without RF modulation, as illustrated in Fig. 3.7(d). Because the RF environment is well regulated leading to a stable output at the RF rectifier, the model matches with the measured data indicating the accuracy of the model, and the equivalent modulation voltage V_{mrf} is measured to be 49.6 mV.

3.2.4 Discussions and Conclusions

As indicated in [104], the energy required to modulate the FN device is as low as 1 fJ so that it can be harvested from ambient environment easily. In this section, the thermal-noise energy manifested as temperature has been directly harvested and used to drive the sensor without the assistance of any external power source. The property can be employed to implement environment temperature sensor which could possibly record the time-temperature history of the sensor. In the next section, we will introduce the detail of using it for time-temperature history sensing.

The derivation of model 3.6 assumes the amplitude of modulation signal is relatively small compared to the value of the floating-gate voltage to maintain model accuracy. A larger modulation signal will degrade the model accuracy because the high order items in the Taylor expansion can not be neglected. Therefore, there is a trade-off when designing the footprint sensor. A larger modulation signal usually can increase the detection sensitivity, however, at the cost of a degraded model accuracy and a larger system volume. With a non-constant modulation input, it will desynchronize the couple of timers, however, the deviation does not follow the model anymore. This is not an issue in applications where whether the timers

go through identical environment or not matters. A linear model was used to model the temperature impact on the timer's behavior. As illustrated in Fig. 3.5, the linear model cannot fully capture the temperature impact. In the next section, we will train the data using more accurate models to extract the time-temperature history.

3.3 Time-Temperature Indicators

Continuous monitoring of time-temperature history is one of the keys towards ensuring product quality in cold supply-chain management. For perishable products like vaccines and antibiotics, exposure to out-of-range temperatures could lead to a significant reduction in drug efficacy and also increase in the likelihood of spoilage [17, 52], an example being the 2015 incident [1] involving improperly stored medicines. Traditional cold supply-chain regulatory systems can monitor these assets in large batches (using electronic temperature sensors inside containers) or using passive labels or tags that indicate the shelf-life of each product. This procedure could potentially lead to unnecessary wastage of food because the actual quality of the product could be still good beyond the indicated expiry date. Also, the use of static expiry date labels could lead to the insertion of spoiled or bad quality products in the supply-chain due to varying storage conditions. This is especially true for perishable products like antibiotics and perishable foods which are sensitive to temperature and their respective quality shows time-temperature dependent deterioration [85] as shown in Fig. 3.8(a).

While time-temperature indicators (TTI) are popular in literature such as [84, 95], they are based on chemical or diffusion processes that require careful calibration and compensation. Also, other passive solutions can potentially monitor instantaneous changes in storage temperature, these devices cannot detect and store historical data of temperature. This is

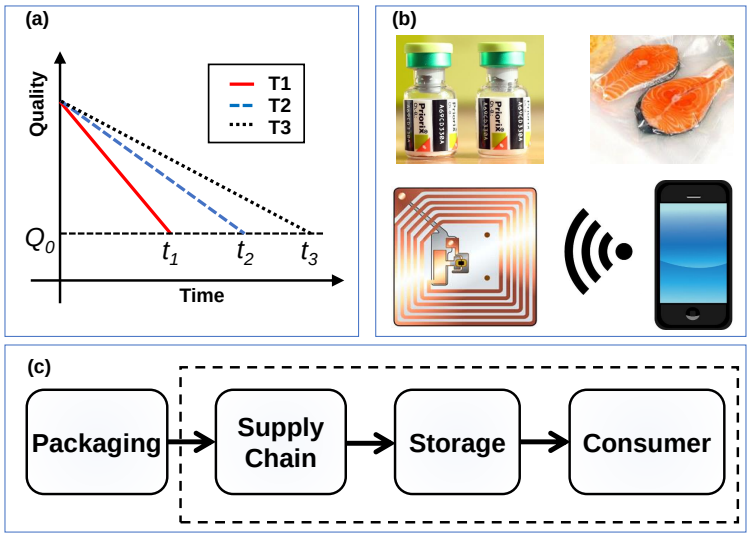


Figure 3.8: Time-temperature monitoring in cold supply-chain management: (a) time-temperature dependent behavior of product quality, (b) use of the proposed sensor for self-powered monitoring of time and temperature and (c) monitoring stages across a supply chain.

because the passive devices do not have access to continuous source of power required for historical monitoring of temperature over time. As illustrated in last section, FN timers show temperature-dependent integration behavior, hence can continuously monitor the ambient temperature over the lifetime of the product supply-chain. An example use case is illustrated in Fig. 3.8(b) where a passive RFID tag integrated with the proposed sensor is attached to every product such as vaccine in the supply-chain. During different stages of the supply-chain including packaging, transportation, storage and the consumer shelves (as illustrated in Fig. 3.8(c)), the sensor would continuously monitor the ambient temperature with respect to time, and functions as a time-temperature indicator.

3.3.1 Operation Principle

As discussed in Chapter 2, the timer's dynamic response is a function of both time and temperature. The timer's output characteristics can be described by equation 3.6 provided

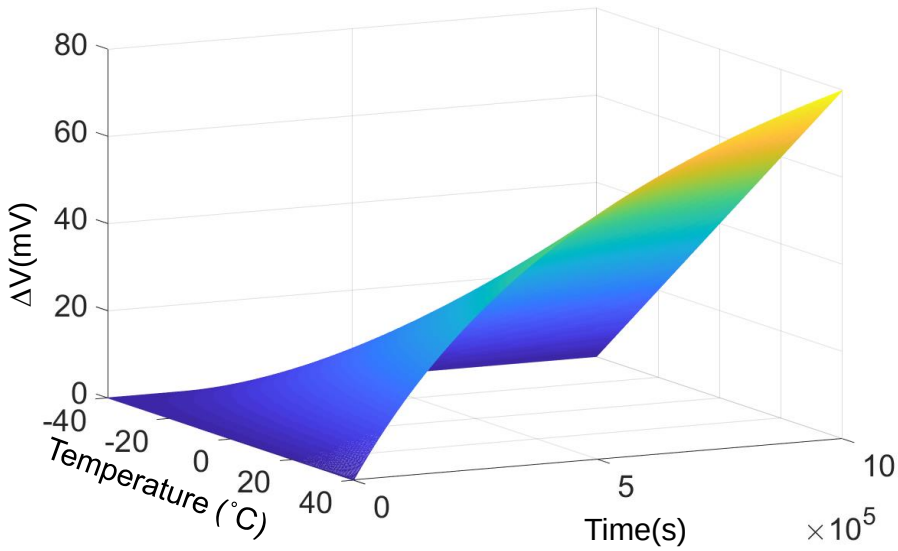


Figure 3.9: Dependence of differential output on time and temperature indicated by equation 3.11 with reference temperature at -40 °C.

the environment temperature is constant, and the magnitude of the temperature impact is included in V_m implicitly. In previous section, it was shown that V_m has an approximately linear relationship with temperature, which can be expressed by equation 3.8, and the coefficient ζ is measured to be around 1.73 mV/°C. By substituting equation 3.8 into equation 3.6 and taking the timer's response at T_0 as reference, we have

$$\Delta V = \frac{k_1 \zeta (T - T_0) t}{k_1 t + k_0} \tag{3.11}$$

The accuracy of this model relies on a constant environment temperature T. If temperature is varying, then an equivalent temperature can be used to evaluate the model. By measuring the timer output ΔV, the relationship between T and t can be obtained.

Fig. 3.9 illustrates the dependence of ΔV on temperature T and time t. For a fixed ΔV, all possible combinations of (T, t) form a contour line in the temperature-time space. If t is

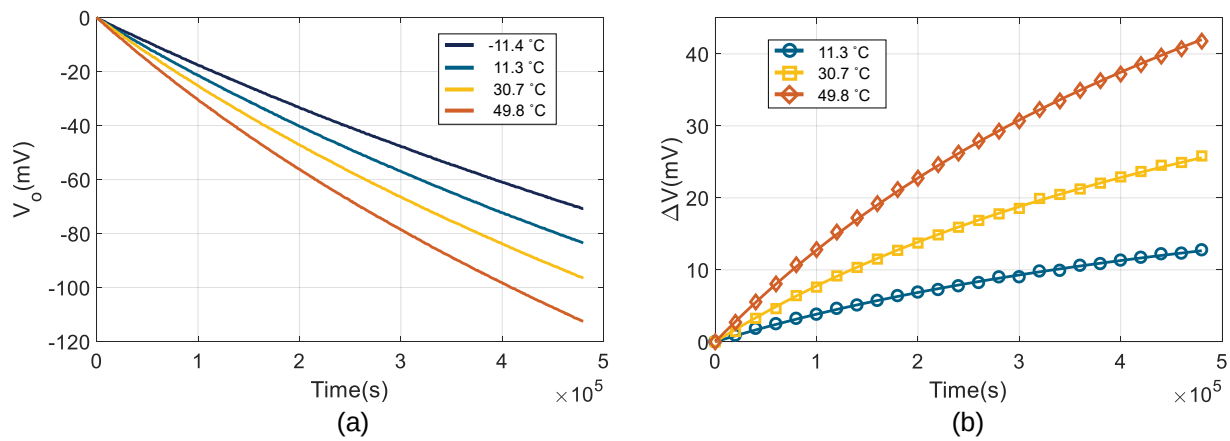


Figure 3.10: Characterization of the time-temperature indicator: (a) measured response of the sensor at different temperature, and (b) fitted result using model 3.11 matches with the measured data well (fitted model is marked as solid line and measured data is marked with discrete points).

known, T can be derived using model 3.11 as

$$T = \frac{\Delta V}{\zeta}(1 + \frac{k_0}{k_1 * t}) + T_0 \tag{3.12}$$

3.3.2 Implementation and Measurement Results

To validate the performance of the sensor as time-temperature indicators, it was programmed to a fixed initial voltage, and put in an environment chamber whose temperature can be accurately controlled. The response of the sensor at different chamber temperature points was measured and recorded. Fig. 3.10(a) shows the measured response of the sensor. High temperature leads to faster voltage reduction at the floating-gate output. If the response at -11.4 °C is used as reference, then the deviations of response at other temperatures can be extracted and are plotted in Fig. 3.10(b). Model 3.11 was used to fit the measured data and the fitted results are plotted in Fig. 3.10(b) as solid lines. The model fits with the data pretty well, hence can be used to derive the time-temperature history of the sensor.

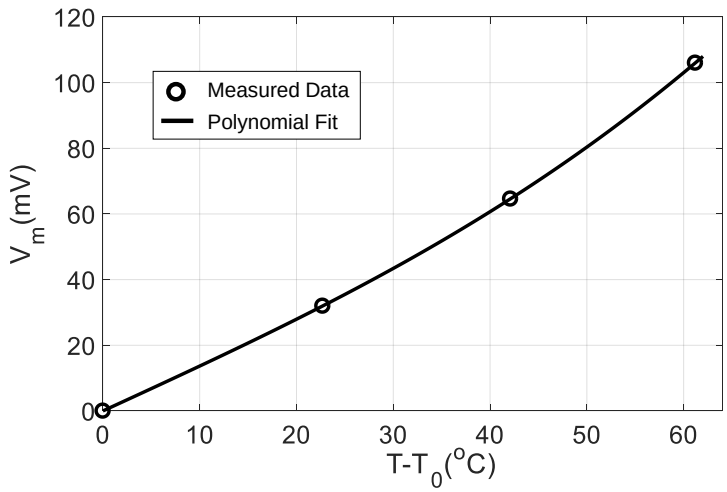


Figure 3.11: Polynomial fit of equivalent modulation voltage on temperature difference.

To accurately extract the temperature from the sensor's output, an accurate model between V_m and temperature is required. Since the physics behind this model is related to the Boltzmann distribution and tunneling probability distribution across a triangular energy barrier, which is too complex to derive an analytical relationship, in this work, we use polynomial functions to model the relationship. As indicated in last section, although linear model can estimate the dependence of V_m on T, the accuracy can be further improved using a high-order polynomial function. To avoid overfitting, a third-order polynomial model is used to extract the dependence. As shown in Fig. 3.11, the third-order model can achieve an error less than 0.1 mV, or equivalently, the temperature accuracy is better than 0.06 °C.

Temperature recovery accuracy was validated using the measured data and the proposed model. From the measured data points of time and output, we can estimate the equivalent modulation voltage V_m. To extract the average temperature, the relationship between temperature and V_m indicated by Fig. 3.11 is used to extract the temperature value. As shown in Fig. 3.12(a), three groups of data were measured and the temperature was extracted respectively. The temperature measured by an external calibration temperature sensor is also plotted in the figure as black lines. Fig. 3.12(b) plots the deviation between the recovered

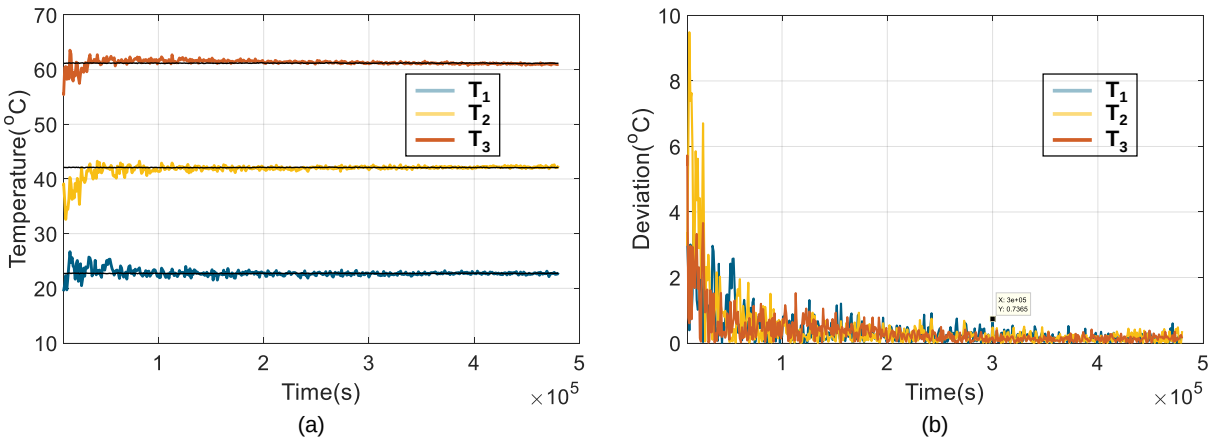


Figure 3.12: Temperature reconstruction from the measured results of the proposed TTI: (a) recovered temperatures and (b) reconstruction error. The measured temperature using a calibration temperature sensor is marked as solid black line.

temperature and the actual temperature. Initially, the deviation is relatively large which can be explained by the model 3.12. Since initially t is small, the same measurement noise at ΔV will lead to larger deviation in the temperature. The impact of measurement noise degrades with the increase of t. As shown by Fig. 3.12(b), the error is less than 1 °C after 2×10^5 seconds.

3.3.3 Discussions

In this section, a self-powered time-temperature sensor based on the FN tunneling device is introduced and characterized. While the sensor cannot detect the instantaneous temperature values, it can be used to estimate the long-term average temperature. This is useful for cold chain monitoring applications where the time information can be easily tracked and retrieved on the server. The sensor can achieved a temperature accuracy better than ±1 °C when the monitoring period is larger than 2×10^5 seconds.

3.4 Remarks and Conclusions

In this chapter, the synchronization performance was further characterized and the desynchronizing mechanisms were analyzed. A pair of synchronized timers were coupled together to detect environmental deviations. A mathematical model with concise form was proposed to account for the temporal behavior of the desynchronization process with high accuracy. Multiple sources of desynchronization were added to the sensor and the dynamic behavior was characterized respectively.

Considering the fact that the sensor is sensitive to signals with extremely low power, it can be optimized to record the footprint of other signals. For instance, although dispersive RF signals are not strong enough for passive data transmission, they are strong enough to modulate the sensor's behavior. Therefore, the proposed sensor should be able to detect ambient RF footprint. Other types of energy sources, such as mechanical and solar power, are also strong enough to change the behavior of the sensor for sensing purpose. Since the power requirement is very low, the volume of the sensor could be minimized without losing the sensitivity. In the future, we will integrate other types of transducers to detect the physical or chemical history of the sensor's footprint and characterize their performance.

Chapter 4

Dynamic Authentication Using Self-powered Timers

4.1 Introduction and Background

An infrastructure of Internet-of-Things (IoT) consisting of servers, readers and tags provides connectivity between systems and devices thus enabling a vast range of applications such as smart homes, wearables, retails, health-care, automotive and agriculture [4, 7, 27, 73, 87]. At the core of this infrastructure are tags (for example RFID tags), which are generally responsible for data collection or exchange with readers that are connected to a server. As these tags operate in an insecure and shared environment, the unprotected communications between tags and readers over a wireless channel can disclose the data collected by the tags and their locations. This raises serious concerns about security of participating tags and makes them susceptible to different security attacks [5, 47, 60].

Denial-of-Service (DoS) attacks are attacks in which an attacker forces tags to disfunction by disturbing or blocking the communication sessions between tags and readers. In ***tag impersonation***, an attacker can intercept sessions between a target tag and the reader by eavesdropping open wireless channel. Based on the intercepted sessions, the attacker can impersonate the tag without knowing its secret. It could communicate with readers instead of the tag to get the authentication from the back-end server. In ***replay attacks***, an attacker reuses communications from previous sessions to perform a successful authentication between a tag and the back-end server. ***De-synchronization attacks*** are used by an attacker to update the values in only one part of the network, either the tag or the reader. In such attacks, the tag and the reader can no more synchronously update their secrets. This makes future authentication impossible and in turn prevents proper functioning of the tag. While the described attacks do not require the attacker to compromise a target tag, there are stronger attacks that result from the physical possession of an attacker to a target tag. In ***backward traceability***, given the internal state of a target tag at time t, the attacker is able to identify tag's sessions that occurred at a time $t_i < t$ [58]. That is, knowledge of a tag's current state could help identify the tag's past sessions, which may allow tracking of the tag's past behavior. On the other hand, in ***forward traceability*** a tag's state at time t can help to identify tag sessions that occur at a time $t_i > t$. That is, knowledge of a tag's current state could help identify the tag's future sessions.

In order to tackle these concerns, it is essential to use secure cryptographic protocols to guarantee the security of tags and their data. However, tags used in such systems are generally passive, i.e., they typically do not possess an on-board source of power. Instead, they gain power by harvesting energy from the reader. This limited power availability severely constrains the computing resources of the tag as well as its storage resources. As a result of these limitations, it is therefore extremely challenging to design a secure cryptographic

protocol that provides security while efficiently utilizing the available resources. Therefore, to solve the security problems of the system, many lightweight authentication protocols have been proposed in recent years. Based on the difficulty of inverting the one-way hash function, it turns out to be the best candidate for most of these authentication protocols. Although some of these protocols are implementable by the resource constrained system, most of them have serious security problems.

An authentication protocol basically defines a set of communications and computations performed between tag, reader and back-end server. While the basic requirement of an authentication protocol is to generally authorize a tag if its ID is recognizable by the back-end database and otherwise unauthorizes it, the designed authentication protocol is also required to follow some guidelines to prevent different types of security attacks such as those described above. These guidelines are to: (1) provide dynamic responses to reader queries to avoid traceability attacks where a current session intercepted by the attacker does not enable him to identify neither tag's past nor future sessions, (2) guarantee that the sessions intercepted by the attacker do not qualify him to further be authenticated as a legitimate tag to avoid tag impersonation and replay attacks and (3) maintain the same shared secret key between the reader and the tag throughout the life-time of the tag to avoid de-synchronization attacks. An IoT system is assumed to be secure if it can consistently follow these guidelines to overcome different attacks on security.

In this work we propose an authentication protocol that guarantees a customizable level of security of tags and their data. More specifically, the proposed protocol utilizes a set of self-powered timers, reported in [105], to perform authentication. The timers provide a mechanism to achieve temporal synchronization between two passive devices without the need for any external powering or clocks. As a result the timers could be used to implement dynamic SecureID type authentication involving random keys and tokens that need to be

periodically generated and synchronized [2]. To authenticate any given tag, values of these timers are compared to a gold standard tag at the reader's side. These values are dynamic where they are essentially periodically updated. Synchronization between the tag and the legitimate reader is efficiently maintained by the timers design and the underlying reliable timer model. While the values from timers at the tag side would not perfectly match values of the gold standard due to measurement and fabrication artifacts, we tolerate an error margin in a more robust and customizable version of the proposed authentication protocol. Threshold of this margin is customized and predetermined based on the deterioration rate of the fabricated models. We also provide a comparison of our protocol with other existing protocols in terms of security, cost and performance.

4.2 Authentication Protocol

4.2.1 Notation

For easy read, the following notations in Table. 4.1 will be used throughout the rest of this section.

4.2.2 Related Work

In order to protect IoT systems from different attacks, many authentication protocols and strategies have been proposed to meet different security requirements. All authentication protocols typically aim to protect tag's security, with minimizing impact on the available limited resources. In this section, to get an idea of how they overcome different attacks, we provide an overview of these authentication protocols. We briefly discuss the design model for each protocol, and analyze their limitations.

Table 4.1: Notations

T	The tag in request.
R	The reader requesting the tag.
S	The server possessing the database.
K	The shared private key.
N_T	The number of tags in the system.
$h(x, y)$	The cryptographic hash function.
ID_T	The ID of the tag T.
n	The current number of tag readings.
n'	The expected number of tag readings.
A	The authentication value (a-bits).
V	The timer value (v-bits).
r	The random bits periodically generated by the timer.

In a first attempt to achieve authentication between tag and reader, Hash-Lock protocol was proposed in [76]. To achieve privacy, instead of using the tag's ID, this protocol uses the pseudonym of the tag, metaID. However, since eventually the secret key and the ID are sent in plain-text, an attacker can eavesdrop the key and the tag can later be impersonated. Therefore Hash-Lock is vulnerable to attacks such as impersonation, replay and tracking attacks. In an attempt to avoid the drawbacks of Hash-Lock protocol, a randomized version of the Hash-Lock protocol was proposed in [89]. In this protocol tags respond to reader's queries by generating a random value. This random value is then concatenated with the hash of the ID and sent to the reader. The reader identifies a tag by searching its database for the ID that corresponds to the hash value. The ID is then sent to the tag in plain-text. While the tag's response varies in each session, it is easy for an adversary to eavesdrop and obtain the identity of the tag. Moreover, the tag's holder is easily traced if the tag's ID is leaked. A hash-chain protocol was proposed in [63]. In this protocol, the tag always replies to the reader queries with different responses. To achieve this, it mainly depends on incorporating two different hash functions. Although this protocol introduces the dynamic property in tag

responses, an attacker can disguise a legitimate tag by resending an intercepted authentication message to the reader. Therefore, the protocol is vulnerable to replay attacks.

In [81], a hash function, a pseudo-random number generator, and an XOR operator are used in an authentication protocol for low cost tags. However, as shown in [107], this protocol is vulnerable to replay and denial of service attacks. In [107], a lightweight anti-desynchronization RFID authentication protocol was proposed. In this protocol, the server keeps track of the updated random key to prevent the active attackers from desynchronizing the shared secret between the tag and the server. Although this technique prevents the replay attack, it is prone to denial of service attacks. Finally, in [33], a scalable pseudo random based scheme was proposed. This scheme utilizes symmetric key cryptography, random number generators, and hash functions for authentication. In this scheme, although the random number generation makes it difficult to predict the next random value, it is susceptible to reverse engineering due to the static structure of the seed.

4.2.3 Preliminaries

A cryptographic hash function h is a mathematical algorithm that maps data of arbitrary size to a bit string of fixed size. It is cryptographically secure if it satisfies the following:

- *Preimage-Resistance*: It should be computationally infeasible to find any input for any pre-specified output which hashes to that output, i.e. for any given y, it should be computationally infeasible to find an x such that $h(x) = y$.
- *Weak Collision Resistance*: For any given x, it should be computationally infeasible to find $x' \neq x$ such that $h(x') = h(x)$ [61].

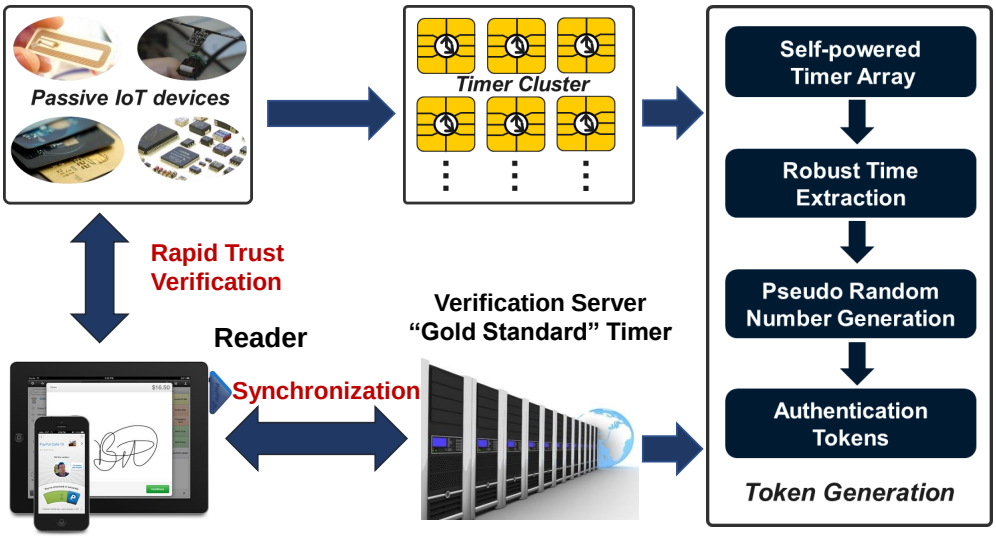


Figure 4.1: Rapid authentication of passive IoT devices using self-powered timers.

- *Strong Collision-Resistance*: It should be computationally infeasible to find any two distinct inputs x and x', such that $h(x) = h(x')$ [13, 26].

4.2.4 The System and Adversarial Model

The System Model

The system usually consists of three components: a tag T, a reader R and a back-end server S. A tag T is basically a chip that has small storage, limited computation resources and constrained communication capabilities. It requires power to perform different operations such as hash computations. Passive tags are battery-less devices operated by energy harvested from the reader. Since they have very limited power resources, these tags are assumed to receive and transmit data within a very short range. A reader R is a powerful device which is authorized by the back-end server to authenticate a group of tags through a set of communication sessions. A back-end server S provides the database for tags and participates with the reader in the tag authentication. The server is also in charge of deciding the authorization of the

set of operating readers. We particularly consider the case of a centralized system, where any reader R from the set of operating readers is continuously online and connected to a centralized server S. We denote the number of tags in a system by N_T, and let T_i for $1 \leq i \leq N_T$ denote the identifier for the ith tag in the system. The back-end server and reader are usually considered to be resource-abundant. They are generally capable of performing intensive cryptographic operations. Therefore the link between the back-end server and the reader is assumed to be secure. Moreover, the server and reader are considered to be a single entity in most of the scenarios.

The Adversarial Model

The adversary could be either passive or active. An active adversary can control a certain number of tags and readers, and modify the conversations between them enabling himself to initiate and terminate a session. A passive adversary eavesdrops the channel between a tag and a reader to learn the output of the communication sessions. The adversary may then deduce information and combine messages to later impersonate or trace a tag.

4.2.5 Dynamic Authentication Protocol

As shown in Fig. 4.1 from [103], the system model comprises one or a set of timers on-board of the tags, namely self-powered timers. These timers periodically generate random numbers that are exploited to generate authentication tokens in the proposed authentication protocol. Each of the passive asset, shown in Fig. 4.1, is integrated with a self-powered timer, which is synchronized to a "Gold-standard timer" located on an authentication server. The tokens are generated by seeding the input of a pseudo-random number generator using the output of the zero-power timers. Rapid dynamic authentication is then achieved by comparing the synchronized random tokens that are generated locally on the asset with the ones that are

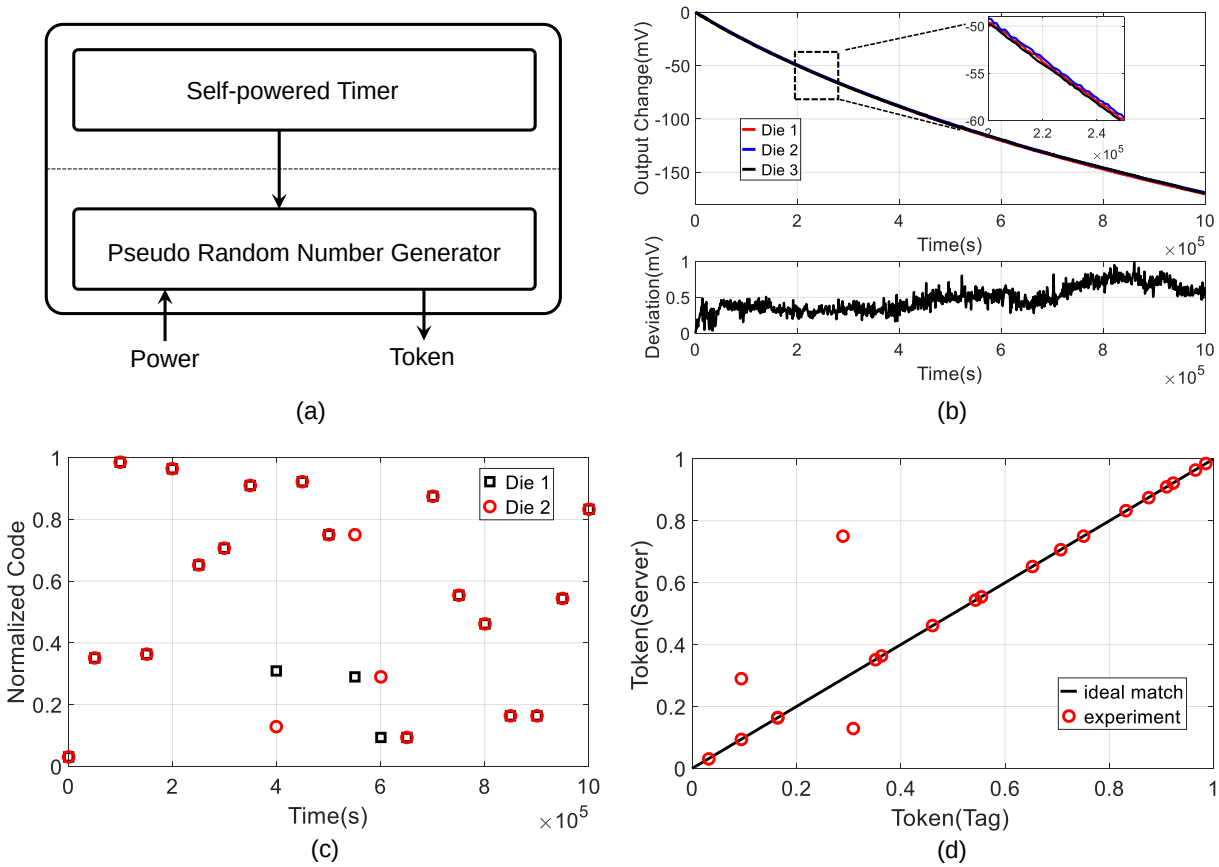


Figure 4.2: Using self-powered timer for token generation: (a) Token generation system proposed in [103], (b) measured data across different dies showing that the timer is robust and showing synchronization accuracy greater than 0.5%, (c) normalized random tokens generated using the timer output from measured response in (b) and (d) the matching result of the random tokens generated from two synchronized timers.

received from the authentication server. Because the output of the timer varies with time, the proposed approach produces a sequence of random number tokens that will be very difficult to predict and hence model using reverse-engineering techniques.

In [103] we combined the output of the timer with a pseudo random number generator (PRNG) to produce authentication tokens. The system comprises of two modules (as shown in Fig. 4.2(a)): (a) the timer which is self-powered and continuously keeps track of time; and (b) a PRNG which is externally powered when an authentication value is requested from

the tag. When a request signal is sent to the tag, the timer value is readout and digitized. The digitized value is then used to feed the PRNG such as a Linear Feedback Shift Register (LFSR) as a seed [40]. After a certain number of cycles of shift operations, the generated random code V_i can be further used in the proposed authentication protocol at any time instance t_i. The time-variant seeds break the pattern of the PRNG and makes it function like a true RNG. A synchronized timer stored on the server goes through the same process and should generate identical random number in ideal cases. By comparing the synchronicity between the two generated random number tokens, authentication can be achieved. On one hand, due to the existence of the PRNG, the timer value can be masked and protected from machine learning attacks. On the other hand, the timer breaks the pattern of the PRNG and therefore makes it difficult to predict the random output. Fig. 4.2(b) shows the measured results from timers with identical structures on three different dies. As discussed in Chapter 2, the timers demonstrates good synchronization performance with larger than 40 dB accuracy. Fig. 4.2(c) shows the normalized random tokens generated using the output from two synchronized timers (as shown in Fig. 4.2(b)) to feed a software version of PRNG. As can be observed, at some time instants, the codes deviate from each other due to the mismatch and quantization error of the digitization process. This issue can be easily tackled by searching a predetermined range of the reference timer values, therefore providing a level of tolerance. If two synchronized timers are integrated on a tag and server respectively, the tokens generated using the described strategy can be used for authentication. As illustrated in Fig. 4.2(d), in ideal cases, the token on the tag should always be equal to that on the server (plotted as the black solid line), while the real tokens can be different at a small portion of random scattered points due to non-ideal artifacts.

The robustness of the self-powered timer is key to successful implementation of the proposed protocol. In Chapter 2, timers with different combinations of form factors were fabricated

and tested at different temperatures. While the device shows various temporal behavior at the initial transient stage, the measurement results verify that at the equilibrium stage, the fabricated designs show high robustness to device mismatch and temperature variations, and the overall synchronization performance is better than 40 dB. The operational life of the device was also studied and verified to be as long as 3 years for passive operation.

The Single Timer Model

In the case of single timer model, only one timer is on-board of the tag. At each reading attempt, after being involved in a simple cryptographic operation, the timer value is compared to that of the corresponding gold standard tag at the server side. The details of the proposed authentication protocol are summarized in Algorithm 1 and described as follows.

The tag T and the reader R are assumed to share the private key K. At any authentication instance t_i, the authentication session is initiated when the tag is in the reader's range. R sends a request to T as an interrogating signal for identification information. T responds or broadcasts to R its identification ID_T and the authentication value

$$A_i = h(K, V_i), \tag{4.1}$$

where V_i is the v-bits timer value and A_i is a bits. T then sends A_i to R for authentication. Similarly, R computes $\tilde{A}_i$ and checks

$$A_i \stackrel{?}{=} \tilde{A}_i.$$

If this holds true, R authenticates T. Otherwise, T is unauthenticated.

As a matter of fact, the objective of any authentication protocol is to minimize the probability of ***false positive*** and ***false negative*** decisions. In false positive, the tag is erroneously

Algorithm 1 The proposed dynamic authentication protocol

Initialization secret (K)

shared between the Tag T and the Reader R.

At authentication time t_i

- R sends request to access T.
- T computes $A_i = h(K, V_i)$, where V_i is the v-bits timer value, and replies with the pair (ID_T, A_i).
- R retrieves T's information from the Server S, computes $\tilde{A}_i$ and checks $A_i \stackrel{?}{=} \tilde{A}_i$. If true, R authenticates T. Else, T is unauthenticated.

indicated to be authentic while it is not. On the other hand, in the false negative, the tag is erroneously indicated to be un-authentic while it is authentic. While this protocol obviously achieves dynamic authentication by sending different and unpredictable authentication values at each session, we have not yet elaborated how it is able to continuously re-synchronize the tag with the server and minimize the probability of false negative decisions. As tags are naturally assumed to operate in a non-secure environment, they generally receive frequent attempts to be read by authentic and non-authentic readers.

We consider a scenario where a tag T is attempted to be accessed by a non-legitimate reader. Since the tag updates its authentication value A_i according to equation (4.1), the authentication value is therefore neither dependent on past nor future tag accesses. Moreover it is also independent of the number of reading the tag has been read. Therefore, the proposed protocol guarantees the synchronization between the tag and the reader at any time instance t_i. As we will show later, this feature also enables our protocol to tackle numerous kinds of security attacks.

We also consider a typical security attacking scenario where at the time period between two authentication values update denoted as T_o, an adversary might attempt to reuse the

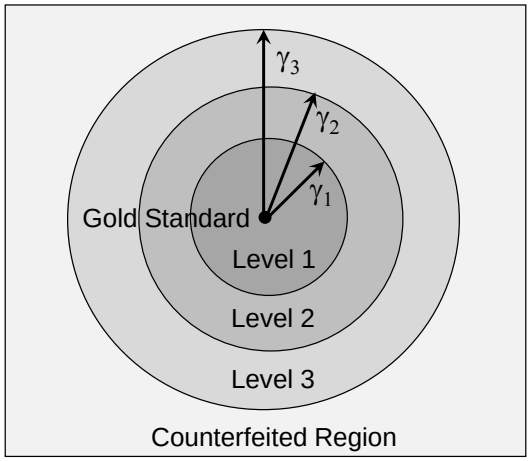


Figure 4.3: Classification based on statistical distance to the gold standard.

intercepted authentication value A_i to get authenticated. We therefore have the following remark.

Remark 1. *At any time instance t_i, when a certain tag ID_T is accessed by a legitimate reader based on a valid authentication value A_i, the server no more accepts re-accessing this tag for a predetermined time period T_o until the authentication value is updated. In other words, during a time period T_o, any tag T can only be accessed once. Any further authentication attempts from the tag ID_T during T_o are considered to be illegitimate.*

The T_o can be set dynamically by the server in a way that the server does not accept consecutive requests with identical authentication value. In other word, if the tag is successfully authenticated with value A_i, the server no longer accepts authentication with value equal to A_i. To initialize another successful authentication process, the timer value needs to be updated leading to an updated A_i. The minimum time duration between two successful authentications can be defined as the lower bound of T_o.

The Multiple Timers Model

To add more robustness to the proposed design, we consider the incorporation of a set of M timers on-board of the tag. The main motivation behind this model is to account for any possible error in the timer values as result of aging or possible security manipulation. In the case of multiple timers, each of these timers generates its own value to be involved in the same protocol as in Algorithm 1. Specifically, at each authentication time t_i between an authentic reader R and any given tag T, R is expecting M authentication values from the M self-powered timers on-board of T, computed as,

$$A_i^j = h(K^j, V_i^j) \quad \text{for } j = 1, 2, \ldots, M, \tag{4.2}$$

where V_i^j is the v-bits timer value of the jth timer and A_i^j is a bits. Based on these values the reader decides the authentication confidence level of any given tag. The resulting M authentication values $\{A_i^1, A_i^2, \ldots, A_i^M\}$ from equation ((4.2)) are compared to the set of expected authentication values at the reader's side $\{\tilde{A}_i^1, \tilde{A}_i^2, \ldots, \tilde{A}_i^M\}$,

$$A_i^j \stackrel{?}{=} \tilde{A}_i^j \quad \text{for } j = 1, 2, \ldots, M.$$

The matches between the two sets are used to compute the authentication confidence level as follows,

$$\text{Confidence Level} = \frac{\text{Number of matches}}{M}.$$

To tolerate possible errors in readings of timers-values between the operating tags and their corresponding gold standard at the reader side, we design the authentication model such that it tolerates a predetermined error threshold γ. This setting enables us to present a customizable version of our protocol summarized in Algorithm 2. The modified protocol

Algorithm 2 The multiple timers version of the proposed dynamic authentication protocol

Initialization secret (K^j)**,** $j = 1, 2, \ldots, M$

shared between the Tag T and the Reader R.

At authentication time t_i

- R sends request to access T.
- T computes $A_i^j = h(K^j, V_i^j)$ for $j = 1, 2, \ldots, M$, where V_i^j is the v-bits timer value of the jth timer, and replies with the pair (ID_T, A_i^j).
- R checks $A_i^j \stackrel{?}{=} \tilde{A}_i^j$, computes the Confidence Level and checks $1 - \text{Confidence Level} \stackrel{?}{\leq} \gamma$. If true, R authenticates T. Else, T is unauthenticated.

provides the flexibility to tolerate different levels of errors corresponding to different thresholds. These thresholds will create different safe regions with different confidence levels as shown in Fig. 4.3. We define the safe region as follows,

Definition 1 (**Safe Region**)**.** *The safe region is defined as the zone where a tested tag is legitimately following the behavior of the gold standard. This region is uniquely determined by a threshold radius* γ*.*

Selection of the threshold radius γ depends on the types of applications, prior estimation of the implementation environment and the expected security level. As illustrated in Fig. 4.3, a larger γ implies a looser restriction on the authentication process, leading to a higher authentication success rate. However, this could possibly cause a higher false positive rate and increase the risk of malicious access. As a result, the trade-off between the security level and successful authentication rate determines the selection of γ. Optimization of the threshold radius leverages the consideration of the ambient environment and security requirements. Generally, a more secure system prefers smaller thresholds such as γ_1 in Fig. 4.3.

The multiple timer version of the proposed protocol is practically an M times application of the Algorithm 1. However, in this case, R receives $\{A_i^1, A_i^2, \ldots, A_i^M\}$ and checks if

$$1 - \text{Confidence Level} \overset{?}{\leq} \gamma.$$

If true, R verifies that T falls in the safe region defined by the threshold γ. R, therefore, authenticates T and updates the state for the next session. Otherwise, T is unauthenticated.

As we previously mentioned, the tag usually operates in an insecure environment. Illegitimate readers may continuously attempt to maliciously access the tags. Thus, between every two legitimate readings, the tag probably had a number of attempts to be accessed of $e = n - n'$ times, where n and n' are the current and the expected number of tag readings respectively. We point out that between two consecutive legitimate tag accesses, no matter how many malicious access attempts have been done, correctness of the following legitimate authentication session still holds. This is a result of the independence of the authentication value of the number of tag readings. However, it might be useful for the reader to keep track of the number of illegitimate attempts e to access the tag. In particular this gives valuable information about the environment and moreover the reader would adaptively adjust the threshold γ based on this information.

Based on the statistical real-life modeling of the incorporated timers, the reader is able to decide whether the deviation in the tag's behavior is natural or it is a result of some malicious act. We therefore give the following definitions.

Definition 2 (**Natural Deviation**). *A natural deviation describes the tag's behavior as a result of natural practical circumstances.*

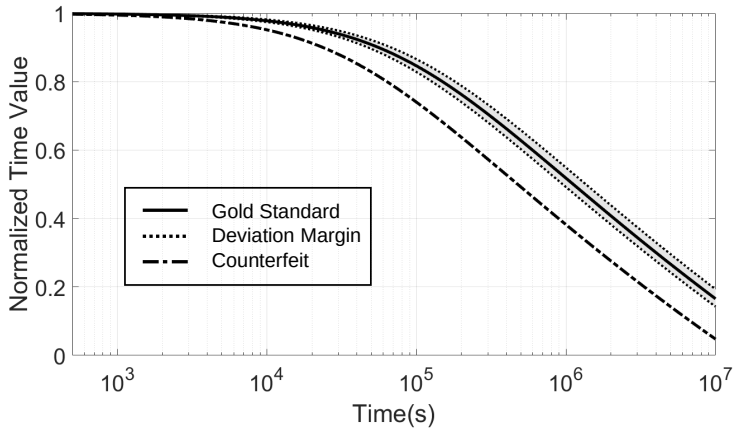


Figure 4.4: Deviation of the timer response from the reference gold-standard timer.

Definition 3 (**Malicious Deviation**). *A malicious deviation describes the tag's behavior as a result of any malicious act, where a tested tag fails to continue following the gold standard deviation pattern or follows it with an unacceptable error.*

Intuition of these definitions is clearly illustrated in Fig. 4.4. Due to nonideal artifacts such as temperature variations and mismatch, the timer device will show natural deviation from the ideal case, however, this deviation is usually within a small range of the gold standard response, as illustrated in Fig. 4.4 marked as "Deviation Margin". Therefore, by searching a predefined small range of the gold standard timer at the server end and selecting a proper threshold radius γ, the natural deviation can be easily eliminated and will not affect the authentication process. However, a malicious deviation is either because of malicious tampering or counterfeited tags that are not synchronized with that on the server. In either case, it is desynchronized and the value of the timer will be far from that stored on the server as shown in Fig. 4.4. It is obvious that a malicious deviation will definitely lead the tag to be un-authenticated. Therefore the proposed protocol enables us to detect counterfeited or malicious tags not only through instant authentication at the beginning of its operation but also through statistical means at any time during its operation lifetime.

4.2.6 Security and Performance Analysis

In this section we analyze the security and performance of the proposed authentication protocol. We begin by investigating the security of the protocol against different kinds of attacks. To be able to do this analysis, we first need to set on two key characteristics of the protocol. One is the secret shared between the tag T and the reader R. The other one is the transmitted messages at each communication session between the tag and the reader. In the proposed protocol, the secret is the private key K. The transmitted messages are basically the tag identification ID_T and the authentication value A_i.

In any authentication attempt at time instance t_i, while the tag sends the value of the same hash function in equation (4.1), both of the hash function arguments K and V_i are secure. More specifically, K is a private key that is never exposed to the adversary in clear-text and is computationally infeasible to derive. V_i is dynamically and continuously updated with the fresh r-bits output from the self-powered timers leading to an unpredictable authentication value.

Most importantly, it is worth to point out that under the assumption that the underlying hash functions have the previously explained characteristics, namely the *preimage-resistance*, *second-preimage-resistance* and *collision-resistance*, the proposed protocol is as secure as the hash functions. Moreover, to achieve the maximum possible security of the hash functions, the proposed protocol is designed to make it infeasible for an adversary, by any means other than exhaustive search, to guess the authentication value, even by overhearing the transmission channel between the tag and the reader. In particular, the adversary can guess a correct a-bits authentication value $A'_i = A_i$ with probability,

$$Pr[(A'_i = A_i)] = 2^{-a}.$$

Table 4.2: Security comparison against various attacks

	Weis et al.[89]	Ohkubo et al.[63]	Song et al. [81]	Fu et al.[33]	Zhou et al. [107]	Our Protocol
DoS Attack			✓	✓		✓
MITM Attack			✓	✓	✓	✓
Traceability Attack		✓	✓		✓	✓
Replay Attack			✓	✓	✓	✓
De-synchronization Attack					✓	✓

We now show how the proposed protocol is secure against most kinds of popular attacks.

Theorem 1. *Our protocol is secure against* de-synchronization attacks.

Proof. Equation 4.1 implies that the authentication value is determined by the current timer value. The robustness of the timer behavior ensures that the timer on the tag will keep synchronized with the timer on the server. The timer's dynamic response cannot be programmed or altered by the reader in the authentication process. As a result, in the case of malicious tag access from an illegitimate reader, the authentication values at any future time instance are independent of the previous readings, hence cannot be altered. The synchronization between the tag and the reader is continuously maintained by the self-powered timers and is resistant to de-synchronization attacks. □

Theorem 2. *The proposed protocol is secure against* tag impersonation attacks *based on the security provided by the combination of the PRNG and the hash function.*

Proof. The protocol features three levels of security that make the impersonation of a legitimate tag infeasible:

- Conventional technique based on the private key K only shared by the tag and the legitimate readers provides the initial level of security.
- The dynamic timer significantly enhances the performance of the RNG, enabling unpredictable output V_i.

- The choice of hash functions make it computationally infeasible for an adversary to find K and V_i.

Therefore, even if the adversary intercepts arbitrary number of messages at time $t < t_i$, it is practically difficult to guess the output A_i at t_i for impersonation. □

Theorem 3. *Our protocol is secure against* replay attacks.

Proof. When a tag is authenticated at time t_i, it goes into an idle mode for a predetermined time period T_o. As explained in remark 1, during this time period, the reader denies any attempts from the authenticated tag to be reaccessed. Therefore, for $t_i < t < t_i + T_o$, a tag T_i is only authenticated once. This prevents any attempts of replay attacks, where an intercepted authentication value A_i is useless during this time period. □

Theorem 4. *The proposed protocol is secure against* backward and forward traceability *attacks based on the security of the hash function.*

Proof. The key to avoid traceability attacks is to avoid using any static or predetermined messages throughout all of the authentication attempts. Our protocol employs the combination of a dynamic timer and a PRNG to generate "true" random numbers that are not predictable. This random feature makes it hard to trace the pattern. The hash function further enhances this attribute. The communicated messages during authentication at time instance t_i can not be inferred from other communicated messages at any other time t_j, where $i \neq j$. Therefore, the authentication protocol is immune to forward or backward traceability attacks. □

Table 4.2 compares the security ability of the proposed protocol to some state-of-the-art protocols proposed in literature.

Table 4.3: Cost comparison

	# of communications	Computation cost for the Tag
Weis et al.[89]	1	$C_C + C_H + C_R$
Ohkubo et al.[63]	1	$2C_H$
Song et al. [81]	3	$6C_X + 3C_H + C_R$
Fu et al.[33]	4	$5C_C + 2C_X + 4C_H + 2C_R$
Zhou et al. [107]	3	$6C_X + 5C_H$
Our Protocol	1	$C_R + C_H$

C_C: Concatenation cost, C_R: Random number generation cost,
C_X: XOR cost, C_F: Flip operation cost,
C_H: Hash function cost, C_S: Circular shift cost.

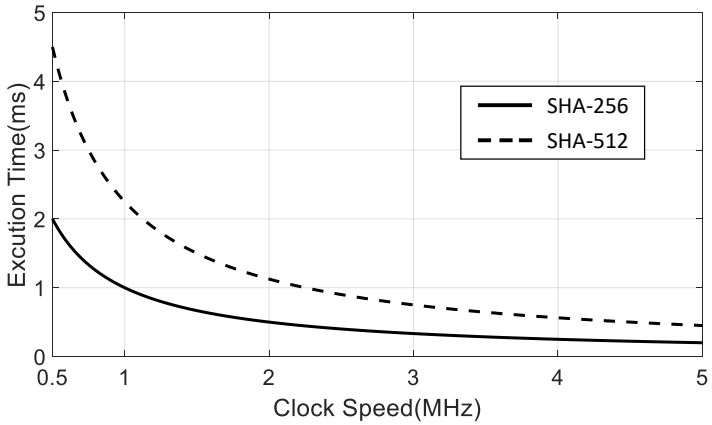


Figure 4.5: Dependence of execution time of SHA-256 and SHA-512 on clock speed.

To evaluate the performance of the proposed protocol we analyze the design from two main aspects: storage and efficiency. Since tags are typically very resource constrained, this analysis is extremely important to evaluate and compare different designs. Generally, the tag is the part of the system with the least storage and power resources. Therefore, in our analysis, while we study the resources required by both the tag and the reader, the resources required by the tag are rather more important. This is a result of the reader being assumed to be powerful and has sufficient storage as compared to the tag. We begin by investigating the amount of storage that our protocol requires. The tag basically needs to permanently store its private key K and ID_T. This amount of storage is, to the best of our knowledge,

equivalent to the least we have seen in literature. In terms of communication cost, with only one transmission from the tag to the reader, the proposed protocol is by far the most efficient we have seen in literature. Moreover, for the performance of tags in terms of hash function computation, we compute the execution time per output of the most well-known hashing algorithm, Secure Hash Algorithm (SHA). While it is benchmarked in [3] that Cycles Per Instruction (CPI) for SHA 256 and SHA 512 are 31.6 and 35.4 cycles/byte respectively, based on these values, we compute the execution time as follows,

$$\text{Execution Time} = \text{CPI} * \text{Bytes} * \text{Cycle Time}.$$

The results of performance comparison are depicted in Figure 4.5 where the execution time is measured at a tag's clock-rate ranging from 0.5 to 5 MHz. In Table 4.3, we present a cost analysis comparison between the proposed protocol and some of the state-of-the-art protocols.

4.3 Mutual Authentication Protocol for Passive Devices

4.3.1 Hash-Based Mutual Authentication Protocol (H-MAP)

IoT systems mostly operate in vulnerable environments. Not only malicious tags attempt to deceive legitimate readers but also the tags are usually susceptible to be accessed by malicious readers. To avoid such acts, mutual authentication is preferred to improve the system security performance. In this enhanced protocol, the tag comprises two independent self-powered timers: $Timer_1$ and $Timer_2$. As previously shown, the utilized timers dynamically modify their values. One of them is used to authenticate the reader while the other is used to prove the tag's authenticity. Before tag access, not only the tag has to prove its authenticity but

also the reader is required to provide a proof of legitimacy. As shown in Fig. 4.6, the tag first gets authenticated by the legitimate reader by securely proving the knowledge of the timer's value V_1. After authenticating the tag, the reader securely proves the knowledge of the self-powered timer's value V_2. If the tag is convinced about the legitimacy of the reader, full access to its data is granted. The protocol comprises the three following phases.

Initialization: The tag and the server initially share a secret key K at the registration phase.

Tag Authentication: At any authentication time instance t_i the reader sends a request to access the tag. To get authenticated by the reader, the tag computes

$$A_{1,i} = h(K, V_{1,i}),$$

where $V_{1,i}$ is the value of $Timer_1$ at time t_i. The tag replies with the pair $(\mathsf{IDT}, A_{1,i})$. The legitimate reader forwards this pair to the server which is able to retrieve the tag information $(K, \tilde{V}_{1,i}, V_{2,i})$ corresponding to IDT from the server, where $\tilde{V}_{1,i}$ is the value of the corresponding timer at the server side. The server computes $\tilde{A}_{1,i} = h(K, \tilde{V}_{1,i})$, and verifies the authenticity of the tag by checking

$$A_{1,i} \stackrel{?}{=} \tilde{A}_{1,i}.$$

If this does not hold true, the tag is unauthenticated. Otherwise, tag is authenticated and proceeds to authenticate the reader.

Reader Authentication: To achieve mutual authentication, the server computes

$$A_{2,i} = h(K, V_{2,i}),$$

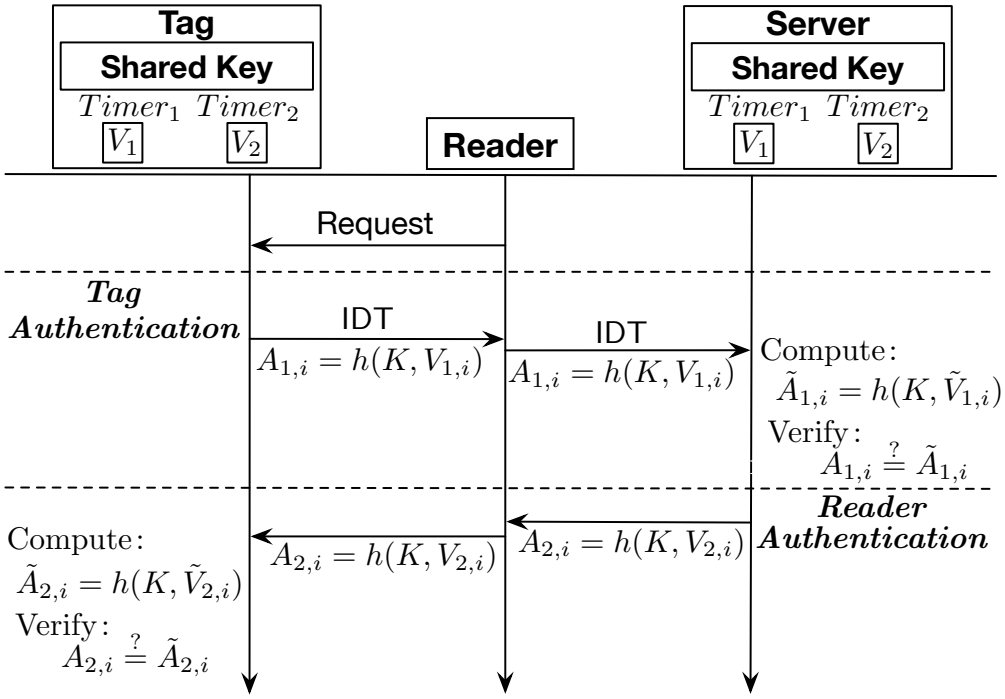


Figure 4.6: The proposed mutual authentication protocol H-MAP.

where $V_{2,i}$ is the value of $Timer_2$ at time instance t_i. $A_{2,i}$ is sent to the reader which forwards it to the tag. To authenticate the reader, the tag computes $\tilde{A}_{2,i} = h(K, \tilde{V}_{2,i})$, where $\tilde{V}_{2,i}$ is the value of the corresponding timer at the tag side. The tag verifies the authenticity of the reader by checking

$$A_{2,i} \stackrel{?}{=} \tilde{A}_{2,i}.$$

If the inequality does not hold true, the reader is unauthenticated. Otherwise, the reader is authenticated. Details of the H-MAP are described in Fig. 4.6. We note that the proposed protocol gives the flexibility to the IoT system to require the mutual authentication or not depending on the application.

Easy Revocation of Tags and Readers: Tags and readers revocation is a trivial setting in any IoT system. We believe that an authentication protocol should enable the server to easily revoke any tag or reader without having any security concern. The server should

guarantee that the communicated messages throughout previous authentication sessions shall not give any advantage to a revoked tag or reader over any regular attacker.

Assuming that a server decides to revoke a tag at time t_{rev}. In H-MAP, a revoked tag possesses the values $(K, V_{1,i}, V_{2,i})$ for $i = 0, \ldots, T_{Life}$, where T_{Life} is the tag's lifetime. At any time instance $t_i \geq t_{rev}$, a revoked malicious tag attempts to get authenticated by sending the pair $(\mathsf{IDT}, A_{1,i})$. The server will simply recognize that the tag IDT has been revoked from the system. Therefore the values possessed by the revoked tag do not give her any advantage over a regular malicious tag in attempting to get illegitimately authenticated. Similarly, assuming that a server decides to revoke a reader at time t_{rev}. For any targeted tag, an extremely lucky revoked reader possesses all the values $A_{1,i}$ and $A_{2,i}$ for all $t < t_i$. At any $t_i \geq t_{rev}$, a revoked malicious reader will fail to maliciously convince a tag of her authenticity where all the possessed values at $t < t_i$ do not give the revoked reader any advantage to guess the authentication value $A_{2,i}$ over the guess of any regular attacker. Therefore a revoked reader fails to maliciously access a tag at any $t_i \geq t_{rev}$.

4.3.2 Privacy-Preserving Mutual Authentication

In this section we introduce the Hash-based Privacy-preserving Mutual Authentication Protocol (HPMAP). The protocol comprises the three following phases.

Initialization: The tag and the server initially share a secret key K at the registration phase.

Tag Authentication: At any authentication time instance t_i the reader sends a request to access the tag. For identification, instead of sending IDT, the tag computes the XOR operation of IDT and the $Timer_1$'s value $V_{1,i}$: $\mathsf{IDT}'_i = \mathsf{IDT} \oplus V_{1,i}$. Also, to get authenticated

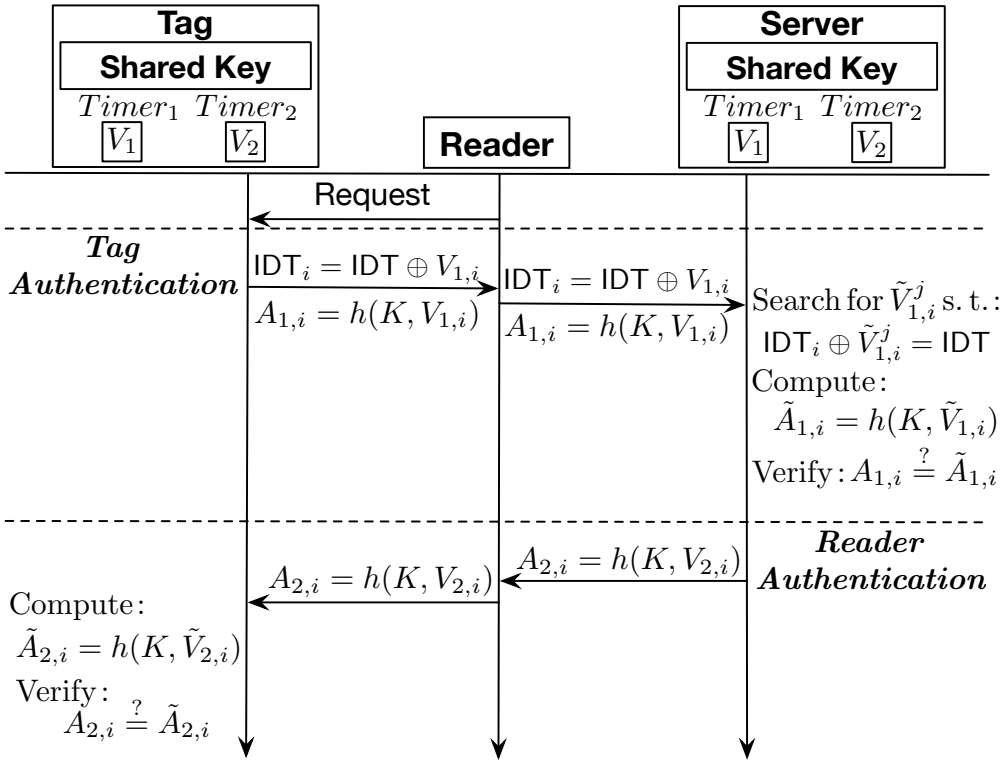


Figure 4.7: The proposed privacy-preserving mutual authentication protocol (HPMAP).

by the reader, the tag computes

$$A_{1,i} = h(K, V_{1,i}),$$

where $V_{1,i}$ is the value of $Timer_1$ at time t_i. The tag replies to the reader with (IDT'_i, $A_{1,i}$). The reader forwards this pair to the server which searches its database for $\tilde{V}^j_{1,i}$ such that:

$$\mathsf{IDT}'_i \oplus \tilde{V}^j_{1,i} = \mathsf{IDT} \oplus V_{1,i} \oplus \tilde{V}^j_{1,i} = \mathsf{IDT},$$

where $j = 1, \cdots, N$ and N is the number of tags registered on the server's database. The server is then able to retrieve the tag information $(K, \tilde{V}_{1,i}, V_{2,i})$ corresponding to IDT'_i. The server computes $\tilde{A}_{1,i} = h(K, \tilde{V}_{1,i})$, where $\tilde{V}_{1,i}$ is the value of the corresponding timer at the server side. The server then verifies the authenticity of the tag by checking

$$A_{1,i} \stackrel{?}{=} \tilde{A}_{1,i}.$$

If this does not hold true, the tag is unauthenticated. Otherwise, tag is authenticated and proceeds to authenticate the reader.

Reader Authentication: To achieve mutual authentication, the server computes

$$A_{2,i} = h(K, V_{2,i}),$$

where $V_{2,i}$ is the value of $Timer_2$ at time instance t_i. $A_{2,i}$ is sent to the reader which forwards it to the tag. To authenticate the reader, the tag computes $\tilde{A}_{2,i} = h(K, \tilde{V}_{2,i})$, where $\tilde{V}_{2,i}$ is the value of the corresponding timer at the tag side. The tag verifies the authenticity of the reader by checking

$$A_{2,i} \stackrel{?}{=} \tilde{A}_{2,i}.$$

If the inequality does not hold true, the reader is unauthenticated. Otherwise, the reader is authenticated. Details of the HPMAP are described in Fig. 4.7. We note that HPMAP also provides easy revocation similar to the analysis in Section 4.3.1.

In the worst case scenario the server will perform N XOR operations to find the IDT corresponding to the pair (IDT'_i, $A_{1,i}$). The N computations hold true if the system is ideal. However, due to the possible error, latency, or miss-synchronization, we recommend that, for each tag, the server saves m authentication values centered at t_i. Although this requires more work at the server's side, it guarantees a better performance by substantially minimizing the probability of false negatives. Thus, practically, the server is required to perform $m * N$ XOR operations at the worst case scenario to find the IDT corresponding to the pair (IDT'_i, $A_{1,i}$).

In case of large scale systems where a server finds these operations computationally time consuming, a further recommended approach is to divide the database into L groups. During a registration phase, each tag is randomly assigned to a group GID_l where $l = 1, \ldots, L$. During the identification, the tag concatenates the group GID_l to the transmitted pair (IDT'_i, $A_{1,i}$).

The server is now required to only perform $m * N/L$ operations rather than $m * N$. This minimizes the work required by a factor of L. The server can then accordingly design the system parameters (m, L) based on the available storage and computational resources.

4.3.3 Security and Performance Analysis

Security Analysis: The proposed mutual authentication protocols rely on the security of the cryptographic hash function. Specifically, a hash function is said to be cryptographically secure if it satisfies the *Preimage*, *Weak Collision* and *Strong Collision-Resistances*. The first implies that it should be computationally infeasible to find any input for any pre-specified output which hashes to that output, i.e. for any given y, it should be computationally infeasible to find an x such that $h(x) = y$. The second implies that for any given x, it should be computationally infeasible to find $x' \neq x$ such that $h(x') = h(x)$ [61]. Finally, the third implies that it should be computationally infeasible to find any two distinct inputs x and x', such that $h(x) = h(x')$ [26].

In H-MAP, for any authentication attempt at time instance t_i, the exchanged messages between the tag and the reader are $A_{1,i} = h(K, V_{1,i})$ and $A_{2,i} = h(K, V_{2,i})$. Both of the hash functions arguments $(K, V_{1,i})$ and $(K, V_{2,i})$ are secure. More specifically, K is a private key that is never exposed to the adversary in clear-text and is computationally infeasible to derive. $V_{1,i}$ and $V_{2,i}$ are dynamically and continuously updated with the fresh bits output from the self-powered timers leading to an unpredictable authentication value. Thus, it is infeasible for an adversary, by any means other than exhaustive search, to guess an authentication value, even by overhearing the transmission channel between the tag and the reader. In particular, the adversary can guess a correct a-bits authentication value $A'_i = A_i$ with probability,

$$Pr[(A'_i = A_i)] = 2^{-a}.$$

Table 4.4: Security comparison against various attacks

	Weis et al.[89]	Ohkubo et al.[63]	Song et al. [81]	Fu et al.[33]	Zhou et al. [107]	H-MAP	HPMAP
DoS Attack			✓	✓		✓	✓
MITM Attack			✓	✓	✓	✓	✓
Traceability Attack		✓	✓		✓	✓	✓
Replay Attack			✓	✓	✓	✓	✓
De-synchronization Attack					✓	✓	✓
Mutual Authentication	✓		✓	✓	✓	✓	✓
Privacy Preservation		✓	✓	✓	✓		✓

Therefore, under the assumption that the underlying hash functions have the previously explained characteristics, the proposed protocol is as secure as the cryptographic hash.

In HPMAP, for any authentication attempt at time instance t_i, the exchanged messages between the tag and the reader are $\mathsf{IDT}'_i = (\mathsf{IDT} \oplus V_{1,i})$, $A_{1,i} = h(K, V_{1,i})$, and $A_{2,i} = h(K, V_{2,i})$. An adversary which is able to intercept this tuple for all $t < t_i$ has two possible approaches to attack the security of this protocol. The first is to just guess the arguments of the intercepted hash functions in an attempt to derive the key and the authentication values. Similar to the analysis in HMAP, this approach is practically infeasible. The other approach is to XOR the intercepted IDT'_is together. This reveals the XOR of all the timer values. Nevertheless, since the timer values are random, the adversary gets no advantage of the XORing result to guess any of the timer values at any time $0 < t < T_{Life}$. In other words, the result of $\mathsf{IDT}'_i \oplus \mathsf{IDT}'_{i-1} = V_{1,i} \oplus V_{1,i-1}$, neither gives any advantage to the adversary in finding $V_{1,i}$ nor $V_{1,i-1}$. Therefore, the proposed HPMAP is as secure as the cryptographic hash functions.

In our previous section [39], we extensively analyzed the security implications of using such timers in dynamic authentication of tags. We proved that using these timers provides immunity against most kinds of well-known attacks such as replay, de-synchronization and traceability attacks. Table 4.4 compares the security ability of the proposed protocol to some state-of-the-art protocols proposed in literature.

Performance Analysis: To evaluate the performance of the proposed protocol we analyze the design from two main aspects: storage and efficiency. Since tags are typically very resource constrained, this analysis is extremely important to evaluate and compare different designs. Generally, the tag is the part of the system with the least storage and power resources. Therefore, since the reader is assumed to be powerful and has sufficient storage, we focus on the tag's resources.

Storage and area overhead: In terms of storage, the tag is only required to store the static key K. The area overhead on chip for the two self-powered timers is in orders of micrometers. Therefore, the tag area overhead as a result of exploiting the pair of self-powered timers is almost negligible.

Efficiency: At the tag side, H-MAC requires only two hash function computations, while HPMAC requires two hash functions and an XOR computation. In terms of communications, both H-MAC and HPMAC require only two communication sessions between the tag and the reader.

In Table 4.5, we present a cost analysis comparison between the proposed protocol and some of the state-of-the-art protocols. We note that even the protocols that provide a close security and privacy functionality to our protocols, achieve this at a remarkably higher computation and communication costs.

Table 4.5: Cost comparison

	# of communications	Computation cost for the Tag
Weis et al.[89]	1	$C_C + C_H + C_R$
Ohkubo et al.[63]	1	$2C_H$
Song et al. [81]	3	$6C_X + 3C_H + C_R$
Fu et al.[33]	4	$5C_C + 2C_X + 4C_H + 2C_R$
Zhou et al. [107]	3	$6C_X + 5C_H$
H-MAP	2	$2C_R + 2C_H$
HPMAP	2	$2C_R + 2C_H + C_X$

C_C: Concatenation cost, C_R: Random number generation cost,
C_X: XOR cost, C_F: Flip operation cost,
C_H: Hash function cost, C_S: Circular shift cost.

4.4 Summary

In this chapter, a dynamic authentication protocol for passive IoT devices is proposed based on the self-powered timing device. By employing a pair of synchronized timers on the tag and server, dynamic authentication tokens that are random can be generated for authentication purpose. The protocol is lightweight and consumes comparable low-computation cost with other protocols, therefore is feasible for passive IoT devices. The dynamic and random feature improves the security performance and it is proven to be robust to most attacks.

Based upon the proposed authentication protocol, a mutual authentication protocol could be achieved by employing two pairs of synchronized timers. The computation cost is also analyzed and is lightweight enough for passive devices. The protocol takes advantage of the dynamic timer response and the security of hash functions to achieve security improvement. The protocol can be modified to accommodate the privacy preserving requirement at the cost of computation complexity at the server end.

Chapter 5

Self-powered Sensing of Time-of-Occurrence

Conventional self-powered sensors cannot record the time-of-occurrence due to the lack of system clocks. In this chapter, a SoC system that combines the self-powered FN timers with a self-powered mechanical events data logger is demonstrated to achieve time-of-occurrence sensing of mechanical events. The architectures of mechanical events detection are first discussed and characterized [97, 98, 102], and then the topology of the self-powered time-of-occurrence sensor is introduced in detail [99].

5.1 Architecture of Self-powered Time-stamping Sensors

Self-powering of integrated circuits is attractive for sensing applications where availability of energy is scarce, for instance, embedded or implantable applications [18, 22, 36, 53, 66,

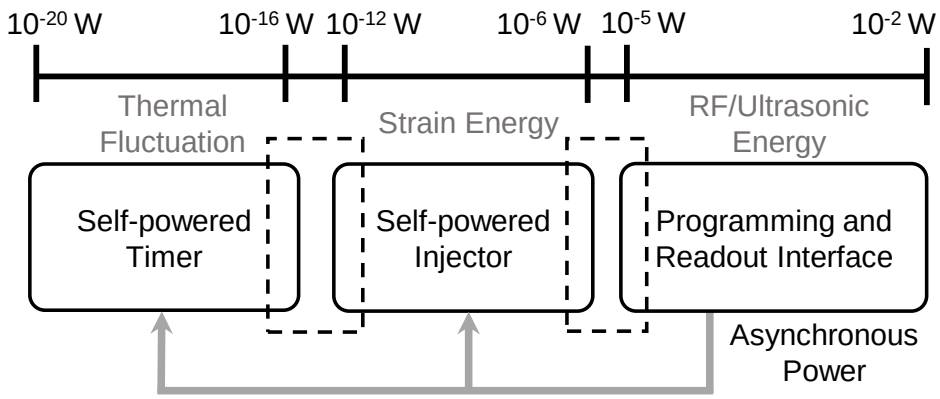


Figure 5.1: Multi-scale power requirements for self-powered sensing of event time-of-occurrence.

77] or for applications where large-scale remote powering is not practical, such as passive Internet of Things [7, 12, 67]. Unlike conventional energy scavenging systems, self-powered sensors scavenge their operational energy directly from the signal being sensed, thus obviating the need for periodic sampling or wake-up interrupts. For example, the self-powered sensor-data-logger which was proposed in [44, 97] harvested the energy from a mechanical event to store and update the event-statistics on a nonvolatile memory. Unfortunately, compared to a conventional interrupt-based paradigm, the asynchronous self-powering paradigm is unable to sense the time-of-occurrence for an event. This is due to the unavailability of a time-reference or a watch-dog clock [11, 55, 59, 80] both of which require continuous powering. In previous chapter, the long-term dynamics (up to three years) of FN tunneling based time-keeping device was demonstrated to be stable and robust to device mismatch and temperature variations.

In this chapter we exploit the reliable time-keeping ability of an FN-timer device to design a fully integrated self-powered system that is capable of sensing time-of-occurrence of an event. Even though the proposed self-powered architecture is applicable for different sensing modalities, the focus will be on time-stamping of mechanical events, for example physical impact [9, 23, 32, 34] or structural strain [38, 43]. The architecture of the proposed system is

shown in Fig. 5.1 and comprises three system modules:

Self-powered Timer — will serve as a source of a precision time-reference over a desired monitoring period. Since the monitoring period could be more than three years, the power-levels at which a chip-scale time-keeping device needs to operate (using only the energy stored on an integrated capacitor) have to be below an atto-watt (or 10^{-18} W). The FN-timer device can operate at this power-level because it exploits a self-compensating physics of electron tunneling through a triangular FN barrier to implement a precision time-reference.

Self-powered Injector — while thermodynamic and quantum fluctuations could potentially be used to implement system clocks and timers, the energy is too scarce for sensing, computation or data-storage. In the proposed system we combine the operational primitives of the FN-timers with the physics of a piezoelectricity driven impact-ionized hot-electron injector (p-IHEI) device [9, 45, 46, 97, 101] to time-stamp mechanical events and for non-volatile storage within a pico-watt to nano-watt power budget.

Programming and Read-out Interface — while FN-tunneling and strain-based self-powering might be sufficient for time-stamping, sensing, computation, and non-volatile storage, it is not sufficient for initialization, calibration or transfer of the stored measurements. In this regard, asynchronous powering, calibration, and read-out could be achieved using a dedicated plug-and-play interface [9, 32] or a wireless radio-frequency [44] or ultrasound [51] interface.

5.2 P-IHEI Self-powered Sensors

Piezoelectricity driven hot-electron injectors (p-HEI) have been shown to be attractive for long-term, autonomous and self-powered structural health monitoring (SHM) applications [53, 65] where the use of batteries or remote powering is considered to be impractical. Some biomedical applications where the p-HEI sensors have been successfully demonstrated include

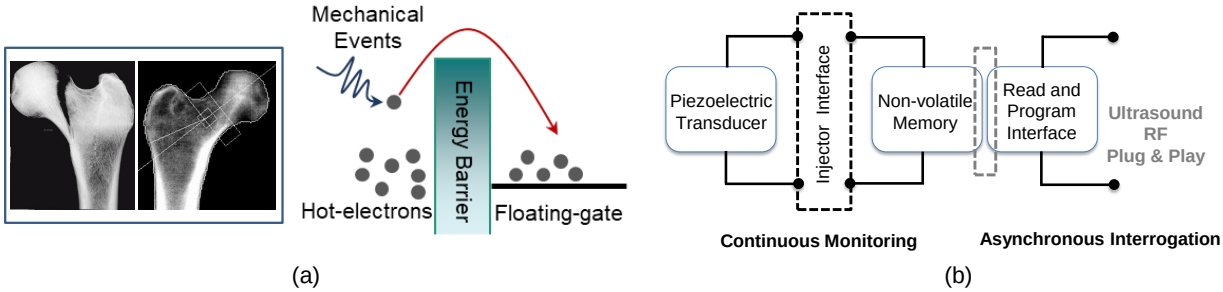


Figure 5.2: Illustration of p-HEI operation: (a) principle of operation and applications of p-HEI device; and (b) system level architecture of a sensor incorporating the p-HEI device.

in-vivo usage monitoring of orthopedic implants [18, 53] and monitoring of mechanical impacts in helmeted sports [32]. Another potential application for self-powered p-HEI sensors is for monitoring the activity of load-bearing and biomechanical structures (the musculoskeletal system) in ambulatory animals. The activity statistics recorded by the sensor could potentially be used for understanding the degenerative pathologies of these load-bearing tissues, such as osteoporosis and muscular dystrophy [16, 19, 68, 90], and allow the longitudinal assessment of the efficacy of therapeutics and clinical treatment strategies.

Since detailed explanation and verification of the p-HEI device physics can be found in literature [45], here we briefly introduce the generic working principle of a p-HEI device in Fig. 5.2(a) using a simplified energy-band diagram. A piezoelectric transducer harvests energy from mechanical strain variations to generate high-energy electrons (or hot-electrons) in the channel of a MOSFET transistor [45]. When the energy of some of these electrons (with the right momentum vector) exceeds the energy barrier (3.2 eV) of the silicon, silicon-di-oxide interface (as shown in Fig. 5.2(a)), these electrons surmount the barrier and get trapped onto a floating-gate. Because the floating-gate is electrically isolated by a high quality insulating oxide, the injected electrons remain trapped for a long period of time. As the piezoelectric transducer is periodically excited, more electrons are injected onto the floating-gate and the total amount of charge stored on the floating gate is thus a function of

the duration and the magnitude of the mechanical excitation. This approach directly couples the physics of piezoelectric energy harvesting with the physics of hot-electron injection to sense, compute and store mechanical usage statistics and hence can be used to push the fundamental limits of self-powered sensing of mechanical strain. The floating-gate also serves as a non-volatile storage memory from which the mechanical usage data could be retrieved at a later stage for offline analysis. This leads to a "sense-now analyze-later" paradigm which has been the basis of p-HEI based sensor architecture [44] as shown in Fig. 5.2(b). The differentiator between different p-IHEI sensor designs is the interface circuit that connects the piezoelectric transducer with the floating-gate memory. By modifying the topology of the interface circuit, the response of the sensor can be changed and the sensor can also be designed to record different statistics of the strain signal [45, 46]. Fig. 5.2(b) also shows the read-out and initialization interface of the p-HEI sensor. Because read-out and initialization of the floating-gates require a more precise calibration and control, which requires more energy than what can be harvested from mechanical strain variations, the functionality is powered by a different energy source, as shown in Fig. 5.2(b). This source of energy could be delivered either using a plug-and-play interface [32], or using a radio-frequency telemetry link [44] or using an ultrasonic telemetry link [31]. In this work we will use a plug-and-play interface for read-out, calibration and programming, keeping in mind that the approach can be integrated with other remote powering techniques.

In literature, two topologies of p-HEI interface circuitry have been proposed and are shown in Fig. 5.3(a) and (b). For both these topologies, the energy harvested from the piezoelectric transducer (whose strain-mode equivalent circuit is shown in Fig. 5.3) is rectified and then used for generating current and voltage references labeled as "R" in Fig. 5.3(a) and (b). The rate at which electrons are injected on the floating-gate is determined by the transistor source current and by the channel-to-drain voltage [45, 46]. In the topology shown in Fig. 5.3(a), the source

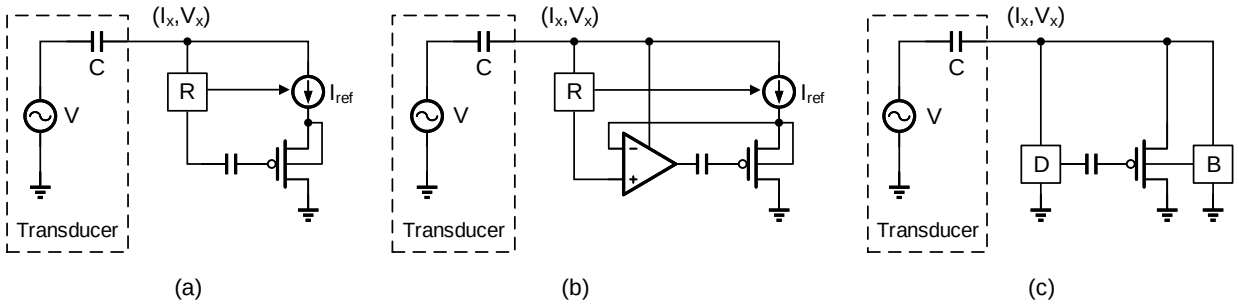


Figure 5.3: Configuration of p-HEI injectors: (a) constant current injector reported in [45]; (b) linear injector reported in [46]; and (c) direct-coupling injector. The modules R,D and B in the figures refer to bias generation circuits.

current is limited by a constant current reference which yields an electron injection rate that exponentially decreases with respect to time [45] . The compressive response of the injector with respect to time makes it challenging to discriminate between different mechanical usage statistics accumulated over a relatively short duration. The topology shown in Fig. 5.3(b) overcomes this problem by using a negative feedback approach to linearize the hot-electron injection process [46]. A feedback amplifier and a current reference maintains the source current and the channel-to-drain voltage constant which ensures a constant injection current. This yields a linear response which has been exploited for designing self-powered strain-gauges [75] and self-powered mechanical impact counters [32]. However, the disadvantage of both these approaches is that most of the transducer energy is dissipated in the rectification and reference generation circuitry. In [45] it was shown that the p-HEI process is functional even at picoampere current levels, which implies that a better interface circuitry connecting the piezoelectric transducer to the floating-gate transistor could significantly reduce the power dissipation of the p-HEI sensor. Lowering the power dissipation of the p-HEI sensor will make it attractive for monitoring applications that require ultra-low activation levels of mechanical strain and mechanical acceleration. For example in systems monitoring seismic activity, the frequency range of the oscillations is usually below 1 Hz, and acceleration level is less than 1 m/s^2 [82]. An embedded piezoelectric accelerometer with a tip-mass of 1 mg

and an oscillation speed 1 cm/s could generate electrical power only in the range of a few nanowatts.

Fig. 5.3(c) illustrates the third topology that directly connects the transducer to the floating-gate transistor and exploits the limited driving capability of the energy transducer [97]. The additional biasing modules labeled as "D" and "B" shown in Fig. 5.3(c) are not in the direct path of energy transfer between the transducer and the floating-gate transistor. As a result, the p-HEI sensor could in principle operate at energy levels that is an order of magnitude lower than previously reported injector topologies and we describe the operational principles and measurement results in the following sections.

5.2.1 Mathematical Model

The mathematical model presented in this section is based on the schematic shown in Fig. 5.3 where it is assumed that the output of the energy transducer is characterized by its output voltage and output current $(V_x(t), I_x(t))$. Since the focus of this paper is to determine the limits of energy-efficiency (low levels of input power), the floating-gate pMOS transistor in Fig. 5.5(c) will be assumed to be always biased in the weak-inversion regime. Also, for the sake of simplicity we will assume that the drain of the pMOS transistor is connected to ground. The transistor source current $I_s = I_x$ can then be expressed in terms of the floating-gate voltage V_{fg}, the substrate voltage V_B and the source voltage V_x as [86]

$$I_x = I_0 \exp\left[\frac{(1-\kappa)V_B}{\kappa U_T}\right] \exp(\frac{-V_{fg}}{\kappa U_T}) \left[\exp(\frac{V_x}{U_T}) - 1\right] \tag{5.1}$$

where I_0 is the characteristic current, κ is the gate-efficiency factor and $U_T = \frac{kT}{q}$ is the thermal voltage (≈26 mV at 300 K) which is directly proportional to the ambient temperature. In weak-inversion regime, the hot-electron injection current can be expressed by the following

empirical model [41] as

$$I_{inj} = -\beta I_x \exp(V_x/V_{inj}) \tag{5.2}$$

where β and V_{inj} are injection parameters which are a function of the transistor size and process parameters. The injection current changes the charge on the floating-gate Q_{fg} and the floating-gate voltage V_{fg} according to

$$C_T \frac{dV_{fg}}{dt} = I_{inj} \tag{5.3}$$

where $C_T = C_{cg} + C_{tun} + C_{gs}$ denotes the total capacitance at the floating-gate node with C_{cg} being gate-coupling capacitor, C_{tun} being the tunneling capacitor and C_{gs} being the gate-to-source capacitance. To complete the model of the injector circuit in Fig. 5.3, the voltage V_B is chosen such that it tracks the highest of the source or the drain voltages. For the sake of simplicity we will use a smooth logistic function that interpolates V_B between V_x and ground as

$$\frac{V_B}{U_T} = \log\left(1 + \exp\left(\frac{V_x}{U_T}\right)\right). \tag{5.4}$$

The equations 5.1-5.4 are coupled which makes it difficult to infer a closed-form solution of $V_{fg}(t)$ for a general form of the input $V_x(t), I_x(t)$. Fortunately, closed-form expressions for $V_{fg}(t)$ can be obtained under special conditions which will provide the justification for the design of the proposed injector. The first special case is when the output current of the transducer $I_x(t) = I_x$ is kept constant; and the second case is when the output voltage of the transducer $V_x(t) = V_x$ is kept constant.

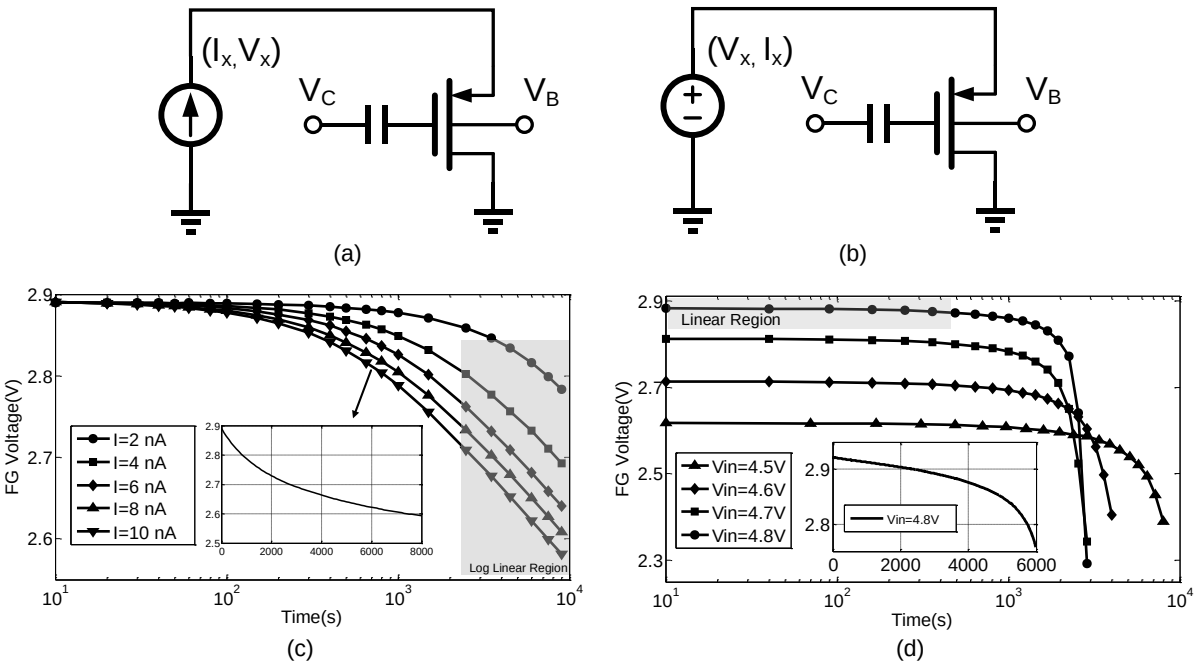


Figure 5.4: Specific operational modes of the injector:(a) constant current mode; and (b) constant voltage mode. Measured results showing (c) a saturating response corresponding to the constant current mode (a); and (d) exponential response corresponding to the constant voltage mode (b).

Constant current mode

When the injector is driven by a constant current I_x as shown in Fig. 5.4(a) the equations 5.1-5.4 reduces to

$$V_{fg}(t) = V_{fg0} - V_{inj}\log(1 + K_1 t) \tag{5.5}$$

where V_{fg0} is the initial floating-gate voltage and the values of K_1 is given by

$$K_1 = \frac{\beta I_x}{V_{inj} C_T}\left(\frac{I_x}{I_0}\right)^{\frac{\kappa U_T}{V_{inj}}} \exp\left(\frac{V_{fg0}}{V_{inj}}\right) \tag{5.6}$$

Note that the value of K_1 is positive. Therefore, the injection process, according to equation 5.5 is a self-stabilizing negative-feedback process. Fig. 5.4(c) shows the $V_{fg}(t)$ measured

from fabricated injector prototypes showing consistency with the equation 5.5. The inset in Fig. 5.4(c) shows the response on a linear time-scale clearly showing the self-stabilizing response. Note for large t, equation 5.5 can be approximated by a log-linear form as

$$V_{fg}(t) \approx V_{inj} \log\{\frac{V_{inj}C_T}{\beta I_x}(\frac{I_x}{I_0})^{-\frac{\kappa U_T}{V_{inj}}} - V_{inj}\log(t). \tag{5.7}$$

Constant voltage mode

The coupled differential equations 5.1-5.4 can also be solved in closed form when the output voltage of the transducer is assumed to be constant $V_x(t) = V_x$, as shown in Fig. 5.4(b). In this operating mode, the hot-electron process shows a positive feedback response, where decrease in the floating-gate voltage V_{fg} leads to an increase in the pMOS source current which further increases the hot-electron injection rate. The form of $V_{fg}(t)$ can be obtained from equation 5.1-5.4 as

$$V_{fg}(t) = V_{fg0} + \kappa U_T \log(1 - K_2 t) \tag{5.8}$$

where

$$K_2 = \frac{\beta I_0}{\kappa C_T U_T} \exp(\frac{V_x}{\kappa U_T} + \frac{V_x}{V_{inj}} - \frac{V_{fg0}}{\kappa U_T}) \tag{5.9}$$

Fig. 5.4(d) shows the injector's response when the transducer output voltage is varied from 4.5 V to 4.8 V. At the start, when the injection duration t is small, the approximation $ln(1 - x) \approx -x$ can be used in equation 5.8 to obtain

$$V_{fg}(t) = V_{fg0} - \kappa K_2 U_T t \tag{5.10}$$

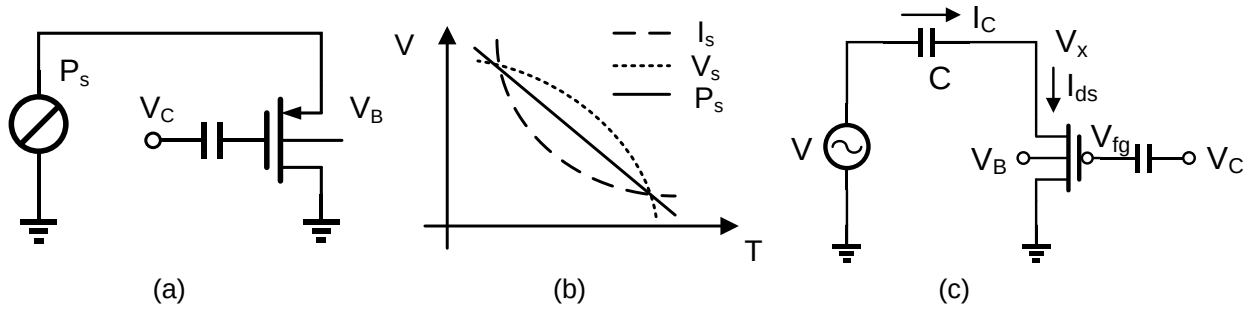


Figure 5.5: (a) A low-powered injector driven by constant-power input. (b)Interpolation of constant-current injection and constant-voltage injection leading to the proposed quasi-linear constant-power injection. (c) Simplified model of an injector driven by a piezoelectric transducer.

which is linear with respect to t. However, as more electrons are injected and the floating-gate voltage V_{fg} reduces, the injection rate increases to the point where the change in V_{fg} is exponential which is also illustrated in the inset of Fig. 5.4(d).

The operation of the proposed hot-electron injector is illustrated in Fig. 5.5 and is based on the following two observations: (a) constant current mode hot-electron injection is a negative-feedback process and $V_{fg}(t)$ shows a saturating response as shown in Fig. 5.4(c); and (b) constant voltage mode hot-electron injection is a positive-feedback process where $V_{fg}(t)$ decreases exponentially as shown in Fig. 5.4(d). Thus, when the input power is constrained to be constant (or product of input voltage and input current is constant) the hot-electron injection should interpolate between both the negative and positive feedback process, as illustrated in Fig. 5.5, leading to a quasi-linear response.

Constant power injector model and numerical study

However, when the input power to the injector is held constant as shown in Fig. 5.5(a), equations 5.1-5.4 can not be solved in closed form and we resort to numerical analysis using a specific type of energy transducer model. Since the target application for the injector device

is for monitoring mechanical activity, a piezoelectric cantilever model has been chosen for this study.

A simplified equivalent circuit model of a piezoelectric transducer operating in strain-mode (non-resonance mode) and driving the process of hot-electron injection in a pMOS transistor is shown in Fig. 5.5(c). The voltage source V models the electrical signal transduced by the strain variations and the capacitor C models the mechanical stiffness of the transducer. Both these variables are a function of the dimensions, the material properties and the mechanical configuration of the piezoelectric transducer. For a cantilever configuration and with a transducer with dimensions $w \times l \times h$, the open-source voltage (V) generated can be expressed in terms of the mechanical force (F) perpendicular along the length as [21]:

$$V = \frac{F}{w} g_{31} = S(t) Y^E h g_{31} = \frac{S(t) Y^E d_{31} h}{\epsilon_p} \tag{5.11}$$

where g_{31} and d_{31} are piezoelectric constants, $S(t)$ is the time-varying mechanical strain, Y^E is the short circuit elastic modulus and ϵ_p is the electrical permittivity of the material. The capacitance in the equivalent circuit model is given by

$$C = \epsilon_p \frac{w \times l}{h} \tag{5.12}$$

A change in mechanical strain induces a displacement current through C which is expressed as

$$I_C = C(\frac{dV}{dt} - \frac{dV_X}{dt}) \tag{5.13}$$

where V_X is the output voltage of the transducer.

The next step in the numerical analysis is to determine the values of the model parameters in equations 5.1-5.4 using experimental data. The measured injector response in either constant

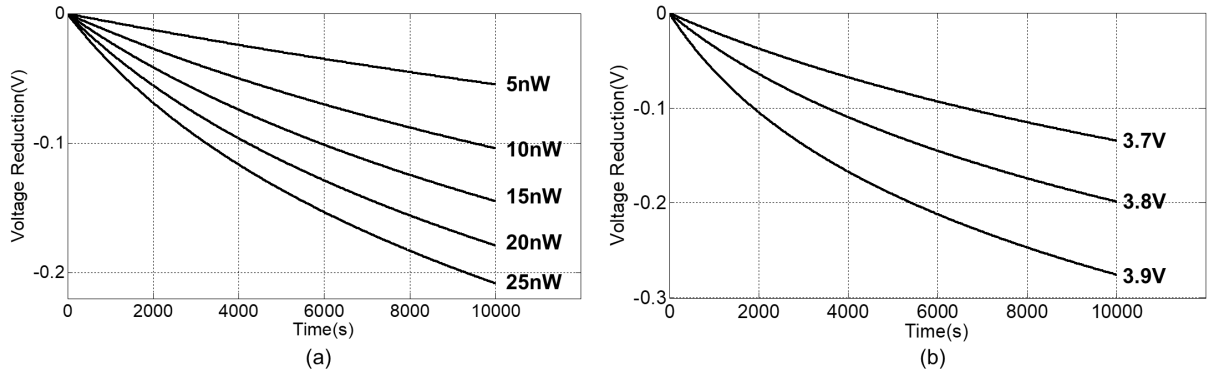


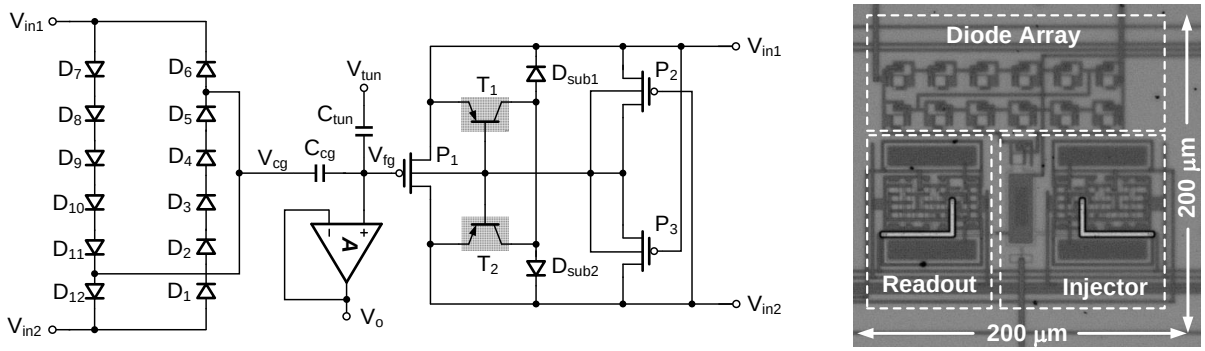
Figure 5.6: (a) Simulated results showing injector responses at different levels of input power. (b) Simulated results showing injector responses with different initial V_{fg} for constant input power ($P_{in} = 15nW$).

current mode or constant voltage mode could be used to estimate the parameters K_1 and K_2 which can then be used to estimate the parameters of the model in 5.1-5.4. The estimated model parameters are summarized in Table 5.1 which also summarizes the equivalent circuit parameters for the piezoelectric cantilever obtained from [74].

Table 5.1: Measured model parameters

Parameter	Value
C	10 nF
I_0	3.5×10^{-24}A
κ	1.1
U_T	0.026 V
V_{inj}	0.16 V
β	1.8×10^{-21}
C_T	1 pF

For the sake of simplicity, $V(t)$ was chosen to be a ramp function in a manner that the average input power to the injector can be varied. Fig. 5.6(a) shows the simulated response of the injector where the reduction in floating-gate voltage is plotted for different levels of input power. As predicted, the response is quasi-linear which is indeed an interpolation of the constant current and constant voltage mode responses. The simulation study also shows



(a) (b)

Figure 5.7: Implementation of the low-power injectors: (a) schematic and (b) micro-photograph of the die.

that the injector can be operated at power levels in the range of a few nano-watts. In another simulation study shown in Fig. 5.6(b), the response of the injector is plotted for different initial values of the floating-gate voltage and at a fixed input power level of 15 nW. The result shows that the duration of the injector can be programmed using different initial values of the floating-gate voltages. This implies that the injector can be used for applications with different requirements on the duration of monitoring.

5.2.2 System Implementation

Fig. 5.7(a) shows the complete schematic of the proposed injector implemented on a standard 0.5-μm bulk CMOS process. The transistor pair P_2 and P_3 implements the interpolation function as in equation 2.14 and ensures that the substrate of P_1 is maintained at the highest potential. The diodes D_1 to D_{12} serve two functions: (a) they limit the input voltage V_x to be within a compliant operating range; and (b) they generate the bias for the control-gate voltage $V_{cg} \approx \frac{V_{in}}{m}$, where m is the number of the diodes in series. T_1 and T_2 represent the parasitic transistors formed by the source and drain terminals with the bulk and p-substrate. The diodes D_{sub1} and D_{sub2} maintain the substrate at the lowest potential.

One of the considerations for implementing the injector in a bulk CMOS process is that the nwell and p-implant of a normal *p-n* diode can form a parasitic *p-n-p* transistor with the p-type substrate. This configuration draws a significant current even when the transducer output voltage V_{in} is small. This issue can be resolved using a diode-connected pMOS with an isolated and shielded nwell.

A unity-gain buffer is used for reading out the voltage on the floating-gate node as shown in Fig. 5.7(a). To prevent the floating-gate charge from leaking to the oxide through the vias, the FG node is formed using the first polysilicon layer and is connected to the input of the amplifier through the same poly layer. To avoid hot-electron injection through the gate of the input transistors of the read-out amplifier, the voltage of the FG is maintained below 3.3 V. Also, when the injector is self-powered through the energy transducer, the power to the read-out buffer is disabled.

The injector was prototyped in a standard 0.5-μm CMOS process using circuit components whose form factors are summarized in the Table 5.2. The circuit occupies an area of 200 μm $\times$ 200 μm and its die photograph is shown in Fig.5.7(b).

Table 5.2: Specifications of the prototyped circuit

Parameter	Value
Technology	0.5-μm CMOS
Die Size	$1.5 \times 1.5\ mm^2$
Layout Size	$200 \times 200\ \mu m^2$
Minimum Current	< 1 nA
Diode Current	< 0.5 nA
Power Limit	< 5 nW

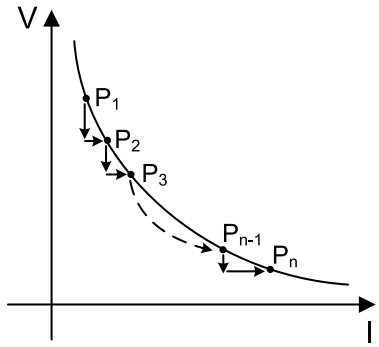


Figure 5.8: Measurement technique used to maintain a constant input power level.

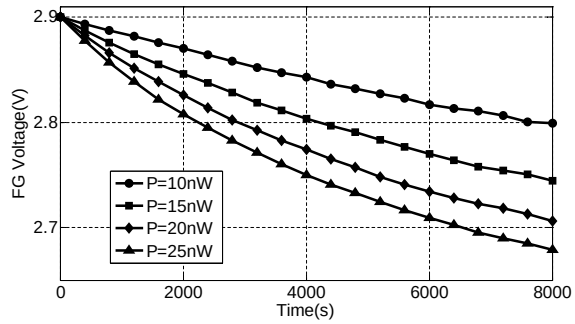


Figure 5.9: Measured injector responses for different levels of input power.

5.2.3 Measurement Results

The fabricated prototypes were used to measure the injector responses for different levels of input power. The floating-gate of the injector was calibrated to a fixed initial voltage. A Keithley 2400 source meter was then used to directly power the injector. A constant input power was maintained by continuously adapting the source current as illustrated using a constant power curve in Fig. 5.8. As more electrons are injected onto the floating-gate, its voltage reduces leading to a reduction in the source voltage. After each 10 second duration, the reduction in source voltage is measured using the source-meter and the source current is increased so that the resulting input power P_{in} remains constant. Fig. 5.9 shows the measured injection responses when the input power is varied from 10 nW to 25 nW. The measured results validate the quasi-linear responses as predicted using simulations shown in Fig. 5.6.

Note that for the fabricated injector, the diode chain also consumes a portion of the input power. This can be estimated using the expression of the diode current as

$$I_{dio} = I_0 \exp(\frac{V_{in}}{m\kappa U_T}) \tag{5.14}$$

with m being the number of diode in the chain. For this implementation $m = 6$ which implies that for V_{in} less than 4.5 V, the diode leakage current is smaller than 500 pA. The actual power consumed by the injector was obtained by subtracting the power consumed by the diode chain estimated using the measured results shown in Fig. 5.10. The response of the injector was obtained at an input power level of P_{in}= 5 nW which results in a longer monitoring duration. The measured result again validates the quasi-linear response, which is consistent with the numerical analysis.

5.2.4 Verification of the Injector

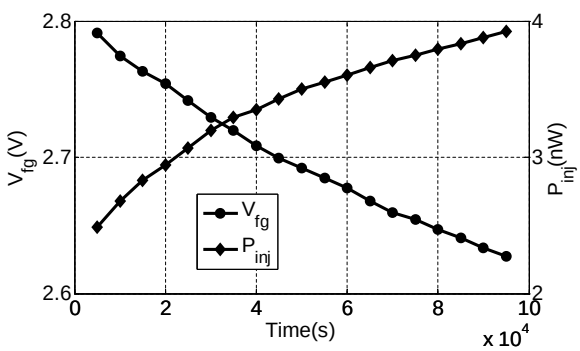


Figure 5.10: Injector response and power consumption of the injector with fixed input power=5 nW.

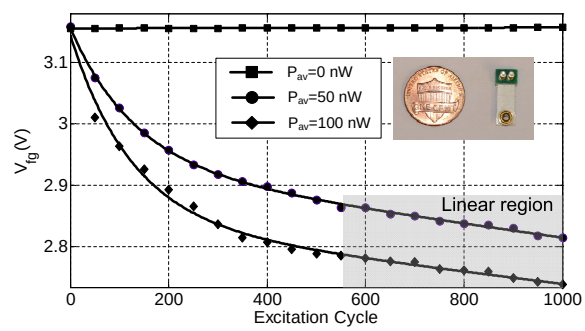


Figure 5.11: Measured injector responses for different levels of power generated by the piezoelectric cantilever.

The fabricated injector prototype was integrated with a miniature commercial off-the-shelf (COTS) piezoelectric cantilever (dimensions: 10mm×5mm×0.1mm, material:PZT-5H) as shown in Fig. 5.11 (inset). The cantilever was subject to periodic mechanical excitation (step displacement) such that it generated different levels of output power. To measure the power delivered by the transducer, a resistive load was connected to the output of the cantilever. For a specific mechanical excitation pattern, the voltage was measured across the resistive load which was then used to determine the output power level. The resistor

was then replaced by the fabricated injector and transducer was again subjected to similar mechanical excitation. The measured injector responses are shown in Fig. 5.11 for two levels of output power which again validates the predicted quasi-linear responses. In another set of experiment, the mechanical excitation of the transducer was reduced to a level such that the output was around 5 nW. The measured response for this condition is shown in Fig. 5.12 and clearly demonstrates a linearized response, as predicted by the simulation model.

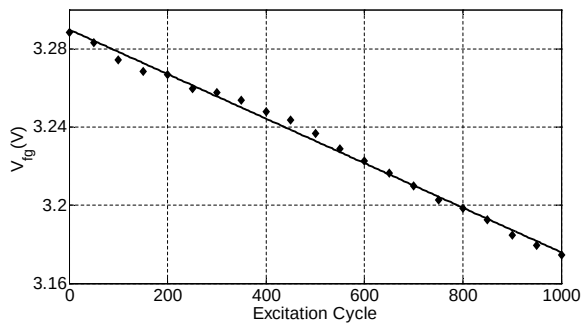


Figure 5.12: Measured injector response when the power generated by the piezoelectric cantilever is pushed down to 5nW.

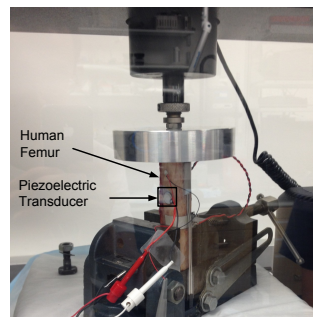


Figure 5.13: An experimental setup using human cadaveric femur with the proposed injector and a ceramic piezoelectric transducer.

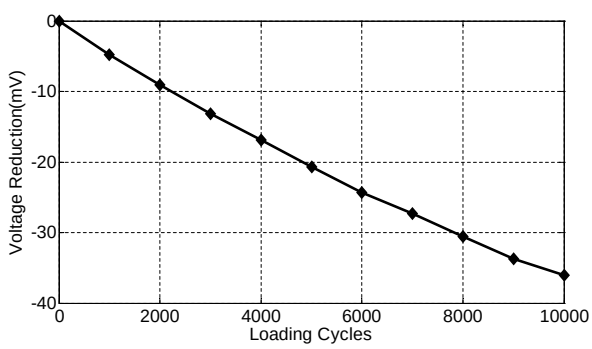


Figure 5.14: Measured response of the injector when the cadaver setup is subjected to cyclic loading.

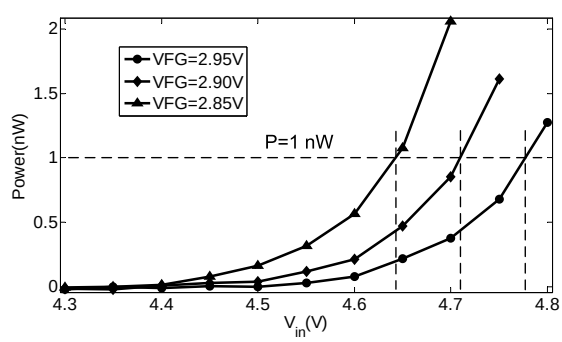


Figure 5.15: Measured effect of biasing and diode-chain on the activation threshold of the injector.

In the last set of experiments the fabricated p-HEI sensor and a ceramic piezoelectric transducer (material: PZT-5H) was attached to the surface of a cadaver femur. The experimental setup is shown in Fig. 5.13. A force of 800 N was applied cyclically to the top of the bone with a frequency of 1 Hz using an Instron 5866 mechanical testing system. The compressive force induces strain-variations in the piezoelectric transducer which then powers the p-HEI sensor. For this experiment, the injector sensitivity was initialized to operate for loading cycles greater than 10000. The measured response is shown in Fig. 5.14 which also verifies the quasi-linear response of the proposed injector. Note, that the magnitude and the type of loading applied to the bone is much lower than what would be expected under typical service loads. Therefore, the injector should also be responsive under normal operating conditions.

Table 5.3: Comparison of different self-powered mechanical activity monitors

work	power
[45]	800 nW
[44]	540 nW
[43]	200 μW
This work	5 nW

5.2.5 Summary

In this section, we presented the design of an ultra-low power CMOS hot-electron injector that can be used for long-term, self-powered monitoring of mechanical strain variations. When compared to the previously reported hot-electron injectors, the proposed injector can operate at 5 nW which is an order of magnitude lower, as is shown in the comparison Table 5.3. This improvement in energy efficiency has been achieved by combining the physics of an energy transducer with the injection response of a pMOS floating-gate transistor when

operating in a constant power mode. The quasi-linear response achieved by the injector is important because the circuit could be used for accurate counting and recording of events of interest [46]. On the other hand, the 5 nW operating limit makes the injector suitable for designing structural health monitoring sensors that can be embedded and implanted inside structures [64]. Another attractive feature of the proposed injector is the ability to adjust the minimum power that can activate the injection process. As illustrated by equation 2.11, the injection efficiency is an exponential function of the source voltage V_s, which implies larger input voltage results in a higher injection efficiency. Therefore, by programming the charge stored on the FG node, we can adjust the injection efficiency and the power threshold. As shown in Fig. 5.15, for a fixed input power, a larger initial V_{fg} corresponds to a larger source voltage required for activation.

5.3 Linearization of Low-power Injectors

Continuous monitoring of variations in mechanical strain presents a significant challenge when the sensor is operating in an embedded and an implanted environment due to limitations on the size of the sensor and due to unavailability of a continuous and reliable energy source (e.g. solar or radio-frequency) [30, 66, 71, 78]. Self-powered sensors based on piezoelectricity driven hot-electron injectors (p-HEI) can overcome these limitations by scavenging its operational energy directly from the strain-variations without requiring any additional source of energy [45, 46, 53]. As a result these devices can continuously monitor and store the statistics of the mechanical strain variations without experiencing any loss of data. The stored statistics could then be retrieved offline, using a radio-frequency [44] or an ultrasonic interface [31]. However, the applicability of a p-HEI sensor to different biomedical applications is determined by the power harvested by the transducers which in turn is determined by the level of ambient strain-variations, the volume of the transducer and the transducer's material properties. Fig. 5.16

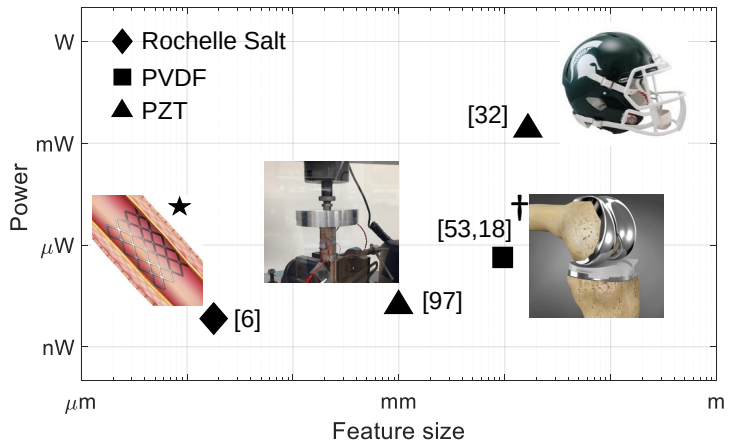


Figure 5.16: Comparison of harvestable strain energy from different biomechanical structures and using different piezoelectric materials. (Image source: ⋆ from American Vein & Vascular Institute, † from Conformis).

compares different levels of electrical power that can be harvested from strain-variations in different biomechanical structures (ranging from stents in blood-vessels to a knee-implant) and by using different types of piezoelectric material. For example, using a Rochelle salt crystal (with $d_{21} = 700$) as an energy transducer, approximately a few nanowatts of power can be harvested from strain-variations at the surface of a blood-vessel (feature size in μm). The harvested energy could be used to monitor the mechanical usage of a stent implanted in a vessel [6]. On the other end of the energy/size spectrum, a PZT-5H ceramic embedded inside a sports helmet could be used to harvest milliwatts of power and used for monitoring head-impacts in contact sports [32]. In all these diverse applications, a p-HEI sensor could sense, compute and store historical strain-level statistics which can then be used either for diagnosis or for predicting any impending mechanical failure. Example of usage statistics include the total duration for which the structure experienced strain-levels that exceeded a pre-determined threshold level [44].

In previous section, a p-HEI sensor that could operate at power levels as small as 5 nW was proposed and discussed. Compared to the previous generation of p-HEI sensors [45, 46], this

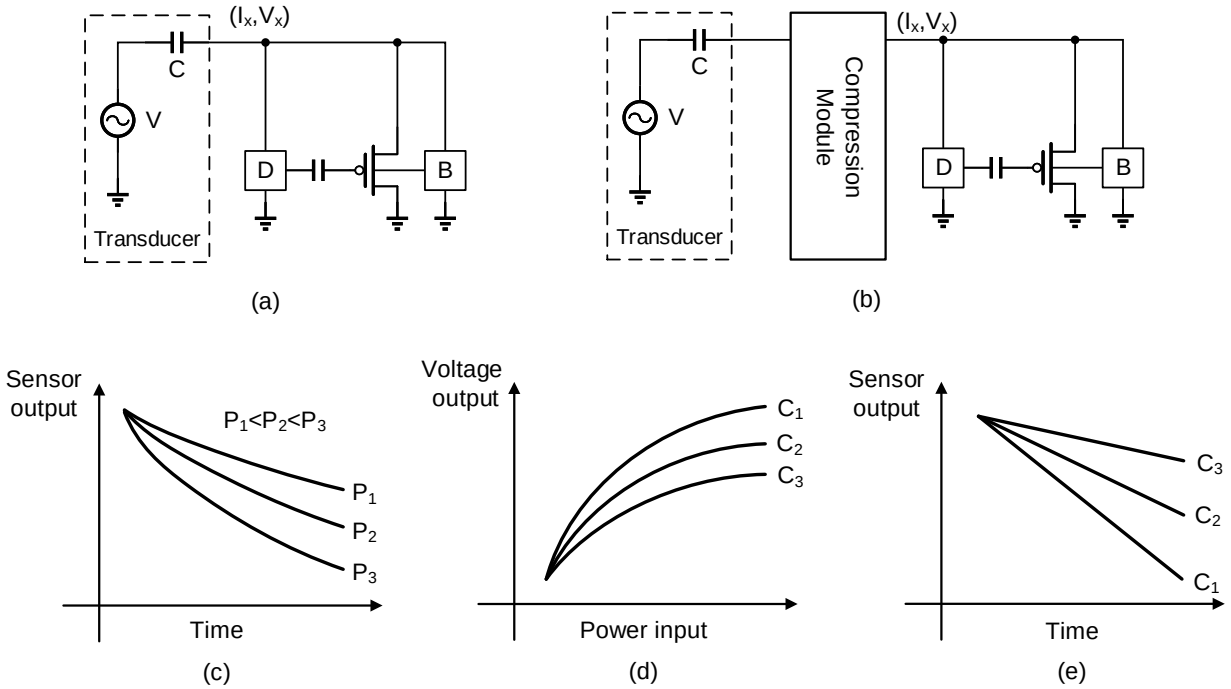


Figure 5.17: Architecture of a p-IHEI sensor: (a) as reported in [97] without pre-compensation; and (b) proposed using a compensation circuit. The modules "D" and "B" in (a) denote biasing circuits for gate and bulk terminals. (c) Quasi-linear response of the injector shown in (a) as a function of mechanical loading; (d) desired response of the compensation circuit for different compression parameters C_i; and (e) linearization effect by combining the responses in (c) and (d).

was an order of magnitude improvement in energy-efficiency. As a result, the sensor could be miniaturized and could be used for implanted monitoring of strain-variations on the surface of a bone [97]. While the response of the injector was demonstrated to be linear for input power levels less than 50 nW, its response become quasi-linear as the level of input power increased. This obviates the use of the technique for statistical monitoring of strain-levels in a range of applications shown in Fig. 5.16, where the harvestable power could vary from 1 nW to 1 mW. Addressing this limitation and extending the linear operating range is the main focus of this section.

Fig. 5.17 shows the schematic of the p-HEI sensor circuit that was reported in [97]. A piezoelectric transducer operating in strain-mode has been modeled by its low-frequency equivalent circuit which directly drives the input of a floating-gate transistor. By eliminating the biasing and regulation circuitry from the direct signal path and by coupling the physics of piezoelectric energy transduction with the dynamics of hot-electron injection on the floating-gate transistor, the circuit is able to reduce its activation energy. A typical response of such injector as reported in [97] is illustrated in Fig. 5.17(c), where each curve corresponds to a different magnitude of input power. The response is quasi-linear with respect to time and the degree of non-linearity increases with magnitude of the input power. In this section, we propose a compensation circuit with an input-output response as illustrated in Fig. 5.17(b). The objective will be to combine the injector response shown in Fig. 5.17(c) with the compressive response shown in Fig. 5.17(d) to obtain a linearized response shown in Fig. 5.17(e) using the compressive response to pre-compensate and linearize the p-HEI injector reported in [97].

5.3.1 Mathematical Model of Compressive Low-power Injectors

In a p-HEI sensor, the flow of electrons generated by a piezoelectric transducer as a result of a mechanical loading event is used to generate hot-electrons in the channel of a MOSFET transistor. Some of the hot-electrons generated in the channel surmount the gate-oxide energy barrier and manifest themselves as injection or gate current. If the gate terminal is terminated only by a capacitive load (also called floating-gate), any charge that is injected into the gate cannot leak out and results in a permanent change in the floating-gate voltage. This change in voltage is measured and then can be used to infer the statistics of the mechanical excitation. A compressive response circuit is introduced to modulate the interface between the piezoelectric transducer and the floating-gate transistor. An equivalent circuit model combining the model of a piezoelectric transducer, the compressive response circuit and the equivalent circuit of a

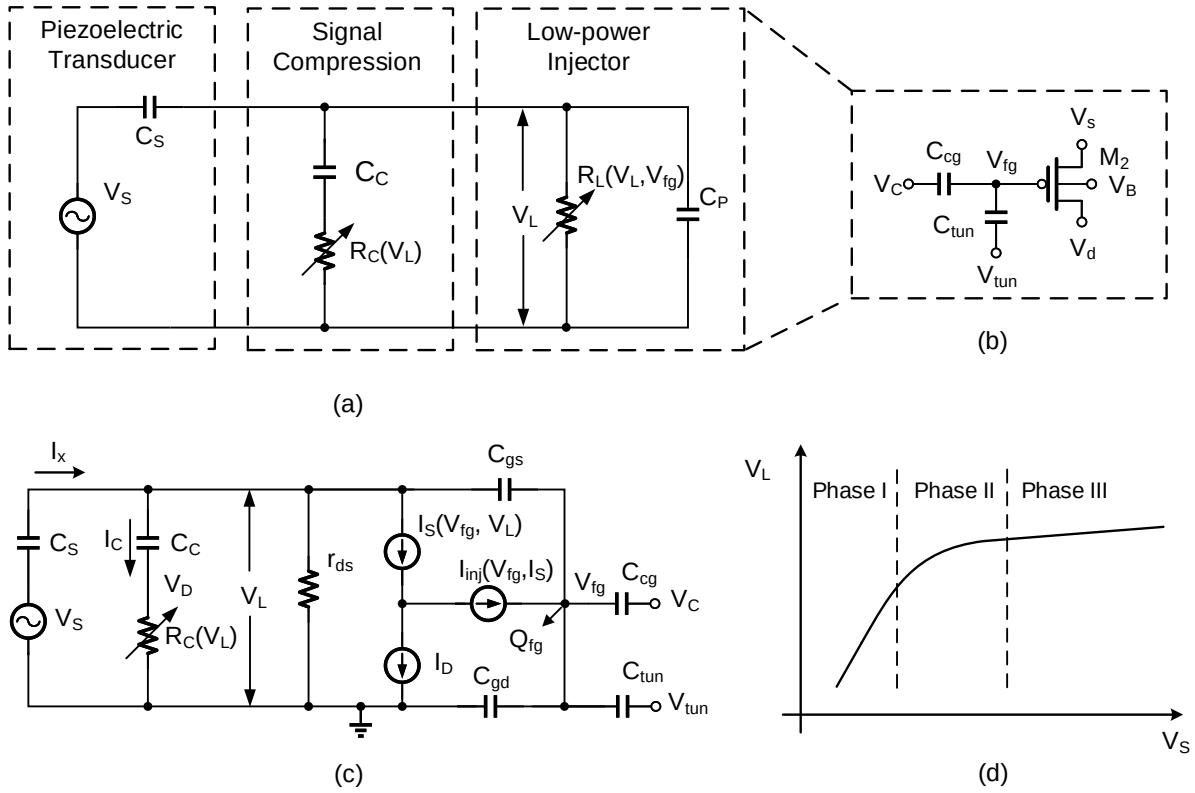


Figure 5.18: Operation of the proposed p-IHEI sensor:(a) Simplified circuit of the system, (b) core circuit of the injector, (c) small-signal equivalent circuit of the sensor and (d) different phases of operation induced by the compression module.

floating-gate pMOS transistor is shown in Fig. 5.18. A piezoelectric transducer operating in strain-mode (non-resonance mode) is modeled as a series connection of voltage source and a capacitor [21]. The voltage source V_S models the electrical signal transduced by the strain variations and the capacitor C_S models the mechanical stiffness of the transducer. Both these variables are a function of the dimensions, the material properties and the mechanical configuration of the piezoelectric transducer. For a cantilever configuration of a transducer with dimensions $w \times l \times h$, the open-source voltage (V_S) and the capacitance in the equivalent circuit model is given by (5.11) and (5.12) respectively.

The injection current depends on the channel current I_S and source-to-drain voltage V_L in the form of:

$$I_{inj} = -\beta I_S \exp(\frac{V_L}{V_{inj}}) \tag{5.15}$$

and the channel current of a PMOS floating-gate transistor biased in weak-inversion region can be expressed in terms of its source voltage V_L, floating-gate voltage V_{fg} as [86]

$$I_S = I_0 \exp(\frac{-V_{fg}}{\kappa U_T}) \exp(\frac{V_L}{U_T}) \tag{5.16}$$

where I_0 is the characteristic current which is a function of process parameters, U_T is thermal voltage (≈26 mV at 300 K) and κ is the gate efficiency factor. Here we assume the bulk V_B is connected to the source terminal. The transducer driving current I_x flowing through the source capacitance C_S can be expressed as:

$$I_x = Cs\frac{dV_S - dV_L}{dt} \tag{5.17}$$

By exploiting the small signal model we have

$$I_C = \frac{V_D}{R_C} \tag{5.18}$$

Notice that I_C charges and discharges C_c, we have

$$I_C = C_C\frac{dV_L - dV_D}{dt} \tag{5.19}$$

Since the injection current is relatively small, we have

$$I_x = I_C + \frac{V_L}{r_{ds}} + I_S + C_{gs}\frac{dV_L}{dt} \tag{5.20}$$

The change in floating-gate charge due to the injection current I_{inj} can be expressed by the following first order differential equation:

$$C_T \frac{dV_{fg}}{dt} = I_{inj} \tag{5.21}$$

where $C_T = C_{cg} + C_{tun} + C_{gs}$ is the total capacitance with C_{cg} being the gate-coupling capacitor, C_{tun} being the tunneling capacitor and C_{gs} being the gate-to-source capacitance. Given the form of V_S, the dependence of V_{fg} with respect to t can be derived by solving the differential equations 5.16-5.21. However, coupled form of the differential equations makes it difficult to derive a closed-form dependence of V_{fg} with respect to t. Instead we evaluate the response of the equivalent circuit for two boundary conditions and then extrapolate between the two responses.

The first boundary condition occurs when the magnitude of the input signal is small enough that $R_C(V_L) \approx \infty$. This is depicted by phase I in Fig. 5.18(d) where the piezo-capacitance C_S and the parasitic capacitance of the injector C_P act as a voltage divider. V_L and V_S are related as

$$\Delta V_L \approx \frac{C_S}{C_S + C_P} \Delta V_S \tag{5.22}$$

Note that since $C_P \ll C_S$, $\Delta V_L \approx \Delta V_S$, and the response of V_L follows the changes of V_S.

The other boundary condition occurs when the magnitude of the input signal is large enough that $R_C(V_L) \approx 0$ and is depicted by phase III in Fig. 5.18(d). During this condition V_L and V_S are related as

$$\Delta V_L = \frac{C_S}{C_S + C_C} \Delta V_S \tag{5.23}$$

If we choose $C_C \gg C_S$, then the input-output response is $\Delta V_L \approx 0$ as illustrated in the phase III of Fig. 5.18(d). In the intermediate phase II the response can be extrapolated between

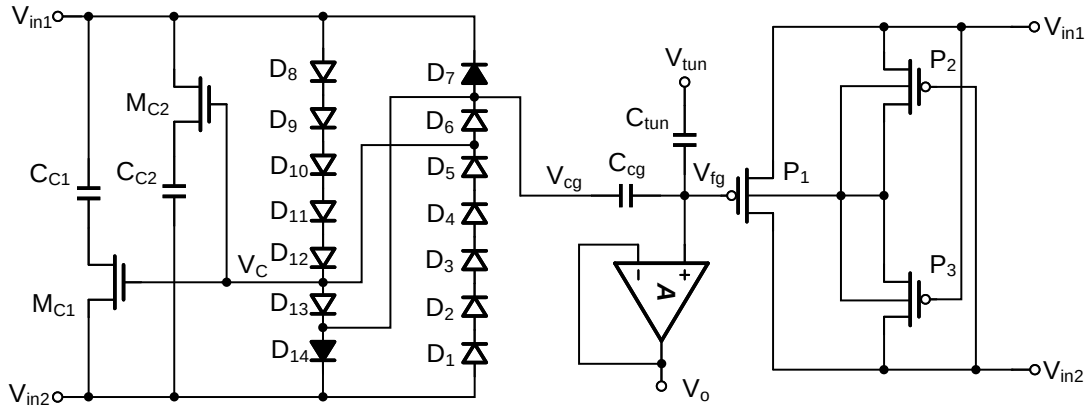


Figure 5.19: Circuit schematic of a complete fully-differential p-IHEI injector with a fully differential compensation module.

phase I and phase III as

$$\Delta V_L = \left(\gamma \frac{C_S}{C_S + C_P} + \theta \frac{C_S}{C_S + C_C}\right) \Delta V_S \tag{5.24}$$

where γ and θ are weighting coefficients that can be adjusted to achieve the desired pre-compensation.

5.3.2 Circuit Implementation

Fig. 5.19 shows the complete schematic of the proposed injector implemented in a standard bulk CMOS process. The design is fully differential as a result of which the circuit can be directly driven by a piezoelectric transducer without any rectification. P_1 is the floating-gate transistor and C_{fg} is the control-gate that couples the control-gate voltage V_{cg} with the floating-gate voltage V_{fg}. C_{tun} is a junction capacitance used for tunneling electrons out of the floating-gate according to the procedure described in our previous papers [45] to program the initial charge on the FG node. The cross-coupled transistor pair P_2 and P_3 ensures that the substrate of P_1 is maintained at the highest potential irrespective of the polarity of the

input voltages V_{in1} and V_{in2}. The diode chain D_1 to D_{14} serve three functions: (a) they limit the input voltage amplitude $|V_{in1} - V_{in2}|$ to be within a compliant operating range; (b) they generate the bias for the control-gate voltage $V_{cg} \approx V_{dio}$, where V_{dio} is the voltage drop on a *p-n* diode (D_7 and D_{14} in Fig. 5.19); and (c) they modulate the gate voltages of transistors M_{C1} and M_{C2} based on the difference in the input voltages V_{in1} and V_{in2}. Thus, M_{C1}, C_{C1} and M_{C2}, C_{C2} implements a differential version of R_C and C_C shown in Fig. 5.18(a). Because the sizes of the capacitors C_{C1} and C_{C2} are comparable to the source capacitance of the piezoelectric transducer, they have been implemented off-chip. The transistor sizes of M_{C1} and M_{C2} were chosen to be $\frac{15\ \mu m}{6\ \mu m}$. To avoid the parasitic p-n-p transistor formed by the p-n diode with p-substrate, D_1-D_6 and D_7-D_{13} are realized using PMOS transistor whose bulk voltage is interpolated by a circuit similar to a configuration formed by P_2 and P_3.

A unity-gain buffer with DC open-loop gain larger than 75 dB is used for reading out the voltage on the floating-gate node as shown in Fig. 5.19. To prevent the floating-gate charge from leaking to the oxide through the vias, the FG node is formed using the first polysilicon layer and is connected to the input of the amplifier through the same poly layer. To avoid hot-electron injection through the gate of the input transistors of the read-out amplifier, the voltage of the FG is maintained below 3.4 V during readout phase. Also, when the injector is self-powered through the energy transducer, the power to the read-out buffer is disabled.

5.3.3 Measurement Results

The injector was prototyped in a standard 0.5-μm CMOS process using circuit components whose form factors are summarized in Table 5.4. The micrograph of the fabricated prototype is shown in Fig. 5.20 and the injector circuit occupies an area of 200 μm $\times$ 150 μm.

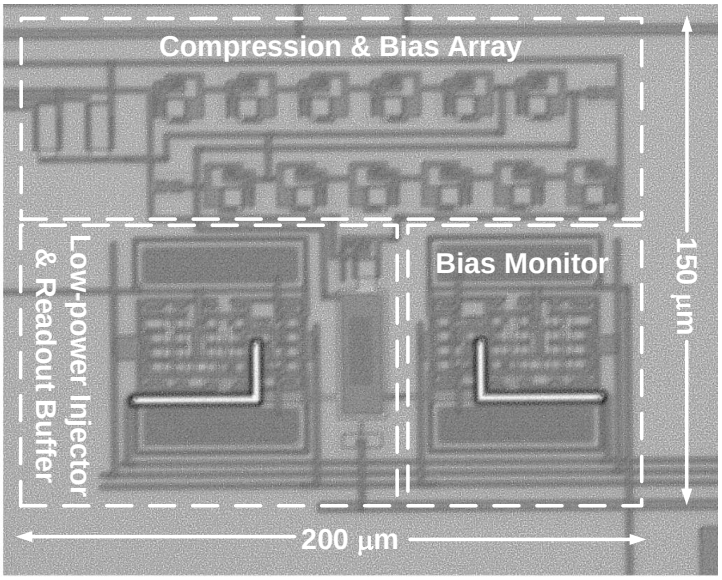


Figure 5.20: Circuit schematic of a complete fully-differential p-IHEI injector with a fully differential compensation module.

Table 5.4: Specifications of the prototyped circuit

Parameter	Value
Technology	0.5-μm CMOS
Die Size	$1.5 \times 1.5\ mm^2$
Layout Size	$200 \times 150\ \mu m^2$
Minimum Current	< 1 nA
Diode Current	< 0.2 nA
Power range	5 nW $\sim$ 1.5 μW

The first group of experiments was designed to verify the dynamic response of the compensation circuit. A signal generator was used for generating a periodic ramp signal with different rise-times and at a duty-cycle of 10% and the signal is directly applied to the input of the injector. For this experiment, the circuit in Fig. 5.19 was configured in a single-ended mode with one of the input terminals connected to ground. The compression capacitor C_C is chosen as 10 nF. Fig. 5.21(a) shows the measured V_D when the input rate is set to 50 V/s. The measured response show that as the input signal ramps up, the transistor M_{C1} turns ON. As a result, the pull-down current through the transistor is larger than the displacement current

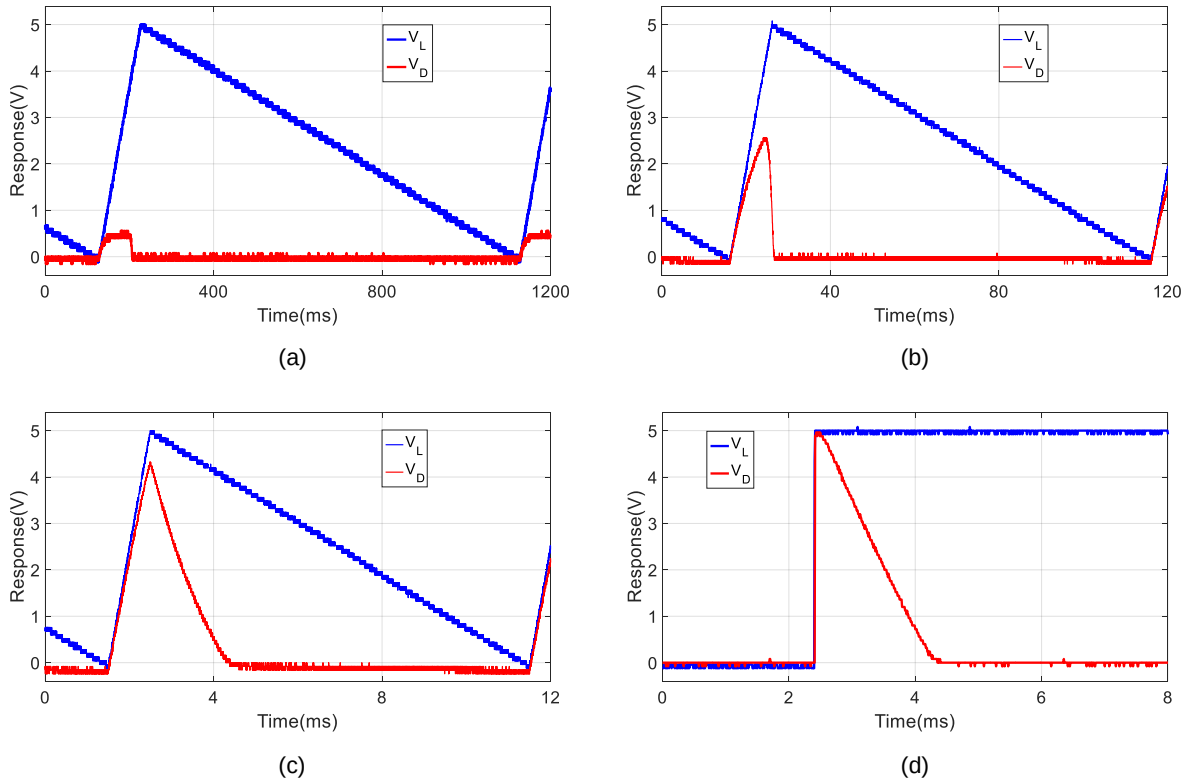


Figure 5.21: Measured response of the compensation circuit for different signal-rates: (a) 50 V/s, (b) 500 V/s, (c) 5000 V/s, and (d) pulse stimulus.

flowing through the capacitor (which is a function of the signal ramp-rate). As a result, the capacitor is fully engaged which will result in the response illustrated in phase III in Fig. 5.18. However, when the ramp-rate is increased, as shown in Fig. 5.21(b)-(d), V_D is able to track V_L implying the capacitor is initially disengaged (unconnected) before being connected to the network. Thus, the dynamic response of the compensation circuit will be able to dynamically interpolate between the phases I and III of the desired response. This dynamic response will be a function of the capacitance C_{C1} and the gate voltage of M_{C1} which is determined by the diode network in Fig. 5.19.

The second experiment was designed to verify the compressive response of the compensation circuit. For this setup the piezoelectric transducer was emulated using a current source I_p

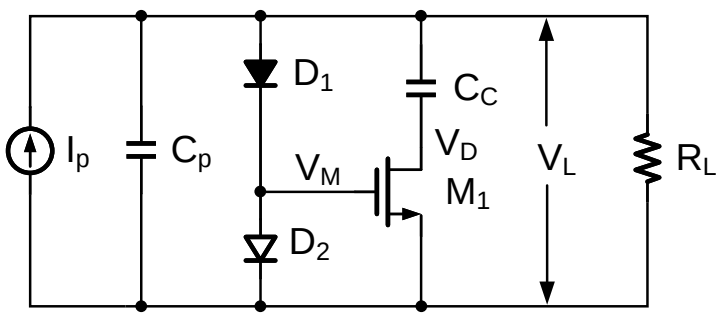


Figure 5.22: Schematic of the test setup used for verifying the compressive response of the compensation circuit.

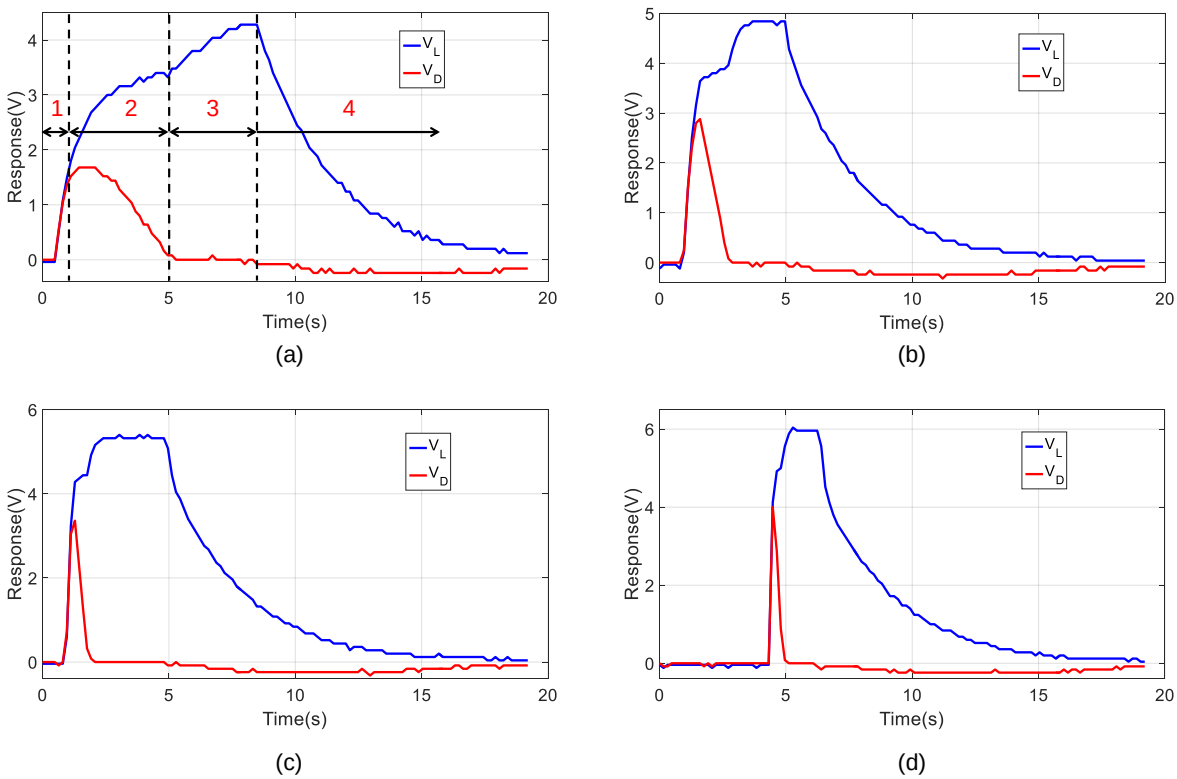


Figure 5.23: Measured response of the compensation circuit for different magnitudes of source currents : (a) 5 μA, (b) 10 μA, (c) 20 μA, and (d) 40 μA.

parallel with a source capacitor C_p as illustrated in Fig. 5.22. C_p was chosen to be 1 μF, and C_C was chosen to be 2 μF. The load resistor R_L was set to be 10 MΩ. A Keithley 2400 source-meter was used to generate 5 μA, 10 μA, 20 μA, and 40 μA of current as shown in

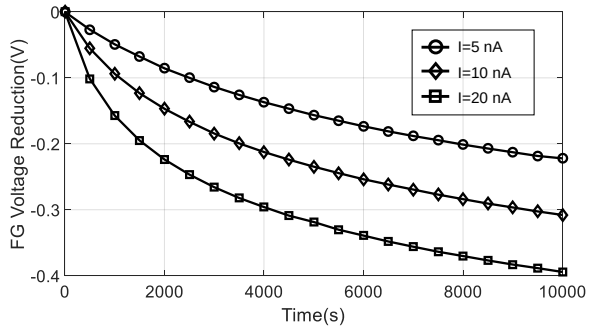


Figure 5.24: Measured injector response when driven by a constant current source.

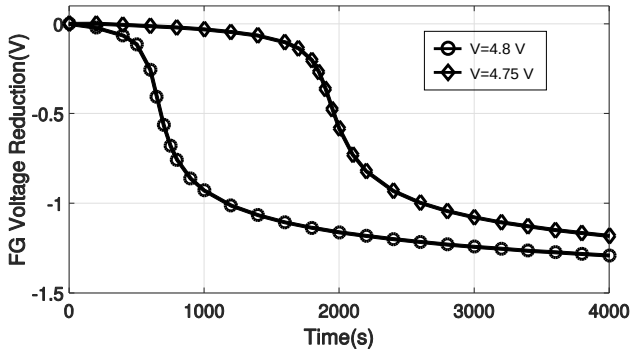


Figure 5.25: Measured injector response when driven by a constant voltage source.

Fig. 5.23(a), (b), (c), and (d) respectively. The measured response, shown in Fig. 5.23(a), can be explained using four temporal phases. In phase 1, V_D follows the change in V_L because the capacitor C_C is disengaged. During this stage, most of the power is consumed by the load R_L. In phase 2, the transistor M_1 starts drawing current and most of the source current is used to charge the energy storage capacitor C_C. As a result V_L starts to show a compressive response. When M_1 is completely on, the circuits enter into stage 3, namely the compensation mode, the capacitor C_C is fully engaged. During the last stage, current source is turned OFF and the energy stored in the capacitor C_p and C_C is discharged through the load R_L. Similar to the previous experiment, the dynamic response of the compensation circuit are verified for different magnitudes of the source current. The results also imply that with large input power, large $C_{Ci}(i = 1, 2)$ is required for effective compression.

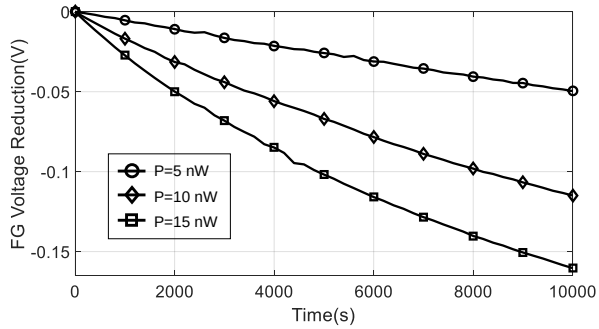


Figure 5.26: Measured injector response when driven by a constant power source.

The next set of experiments verified the operation of the injector under three different source models: constant current source, constant voltage source and constant power source. A Keitheley 2400 source meter was used to bias the injector in different modes. In constant current mode, a current source was applied directly to the injector in Fig. 5.18(b). Fig 5.24 shows the measured response of the injector. Because the floating gate voltage V_{fg} decreases, the source voltage will also decrease implying a slower injection rate. Therefore, constant current injection demonstrates a saturation behavior as shown in Fig. 5.24.

In contrast, for constant voltage mode, where the source-to-drain voltage of transistor M_2 in Fig. 5.18 is kept constant, the injection rate will first increase because the source current increases, as shown by equation 5.15. However, when the current is large, the overdrive voltage dropped in the pinch-off region decreases which implies the energy of hot electrons will decrease at the same time. When this negative effect overcomes the positive feedback brought by the increasing channel current, the injection speed starts to decrease. This saturation process is also illustrated by the measured results shown in Fig. 5.25.

In constant power mode, we used the Keitheley source meter to periodically update the voltage and current to ensure the input power to the injector is constant. The measured responses with constant input power from 5 nW to 15 nW are shown in Fig. 5.26, which shows a quasi-linear response. The constant-power operation of the proposed hot-electron injector

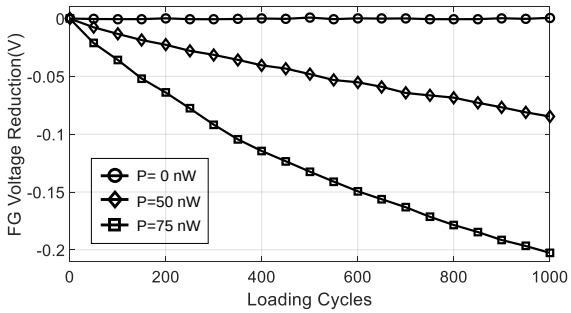


Figure 5.27: Measured injector responses for different magnitude of input power (C_{TD} = 1.5 nF).

can be understood intuitively based on the following two observations: (a) constant current mode hot-electron injection is a negative-feedback process and $V_{fg}(t)$ shows a saturating response as shown in Fig. 5.24; and (b) constant voltage mode hot-electron injection is a positive-feedback process where $V_{fg}(t)$ decreases exponentially as shown in Fig. 5.25. Thus, when the input power is constrained to be constant (or product of input voltage and input current is constant) the hot-electron injection should interpolate between both the negative and positive feedback process, leading to a quasi-linear response. The measured results of the injector are consistent with the conclusions in [97].

For the fourth set of experiments the functionality of the proposed injector was validated by interfacing with a miniature commercial off-the-shelf (COTS) piezoelectric transducer (dimensions: 10mm×5mm×0.1mm, material: PZT-5H) operating in a strain-mode. Since the response of the piezoelectric transducer have already been verified in our previous papers using different biomedical phantoms [97] (as shown in the inset of Fig. 5.29), we have emulated the response using a benchtop set up which can be configured to generate different levels of input power. Note that due to the diversity of applications that the proposed injector could be applied to (shown in Fig. 5.16), a controlled phantom / test-bed experiment seemed more appropriate. This way we are be able to verify the response of the sensor while avoiding

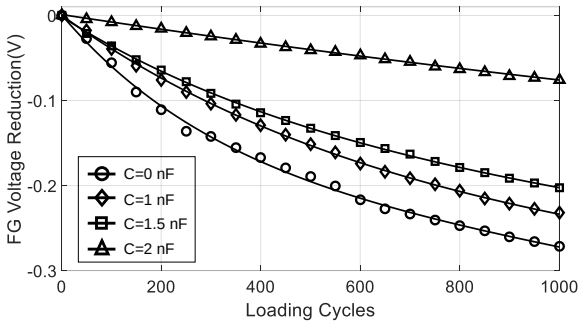


Figure 5.28: Measured injector responses for different values of compensation capacitances (P_{in}=75 nW).

issues related with packaging and sensor attachment with different bio-mechanical structures, like bone, stents and wearables. For the benchtop experiments, the transducer was subject to periodic mechanical excitation (step displacement) such that it generated different levels of output power. To measure the power delivered by the transducer, a resistive load was connected to the output of the cantilever. For a specific mechanical excitation pattern, the voltage was measured across the resistive load which was then used to determine the output power level. The resistor was then replaced by the fabricated injector and the transducer was again subjected to similar mechanical excitation. A pair of off-chip capacitors with 1.5 nF capacitance were connected as the compressive capacitor. The injectors' response to the cantilever with different input power levels are shown in Fig. 5.27, which validate the quasi-linear response.

The next set of experiments were designed to determine the effect of compensation circuit parameters on the linearization of the proposed injector. Different compensation capacitances with values ranging from 0 nF to 2 nF were used and the measured responses are shown in Fig. 5.28. $C_C = 0$ equals to the case that there is no compensation circuit and the circuit degenerates to the case described in [97]. With the increase of capacitance, the charge injected to the FG node is decreased due to the compressive response of the compensation

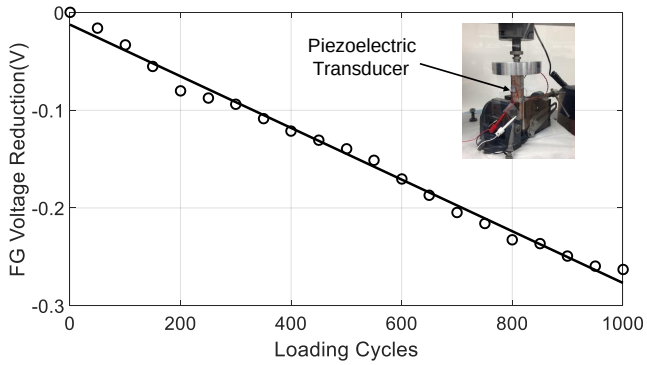


Figure 5.29: Measured injector responses when the p-IHEI sensor is directly driven by a ceramic PZT-5H transducer for C_C=100 nF and average input power P_{in}=1500 nW. The benchtop setup emulates the biomechanical phantom shown in the inset which was reported in our previous work [97]

circuit. As a result, the linearity of the injector is improved validating the functionality of the compensation circuit.

The last experiment was designed to verify that the proposed injector could be adapted to different applications with different input power by adjusting the magnitude of the compensation capacitor. A ceramic piezoelectric transducer (Outer diameter: 20 mm, inner diameter:15 mm, material: PZT) was integrated with the injector to generate larger power output. The compression capacitor was chosen as 100 nF to accommodate power ranging from 1 μW to 2 μW. The piezoelectric transducer was subject to periodic mechanical excitations. The floating gate voltage was initialized to 3.4 V and injector's response was measured and recorded as shown in Fig. 5.29, which displays an almost linear dependence with respect to the number of mechanical loading cycles.

Fig. 5.30 summarizes the response of the linearized injectors for different input power levels which also includes the error residue estimated after a linear regression over the measured data. The linearity of the injection results can be estimated by the ratio between the injection range and the largest regression error, and the measured results show that the proposed

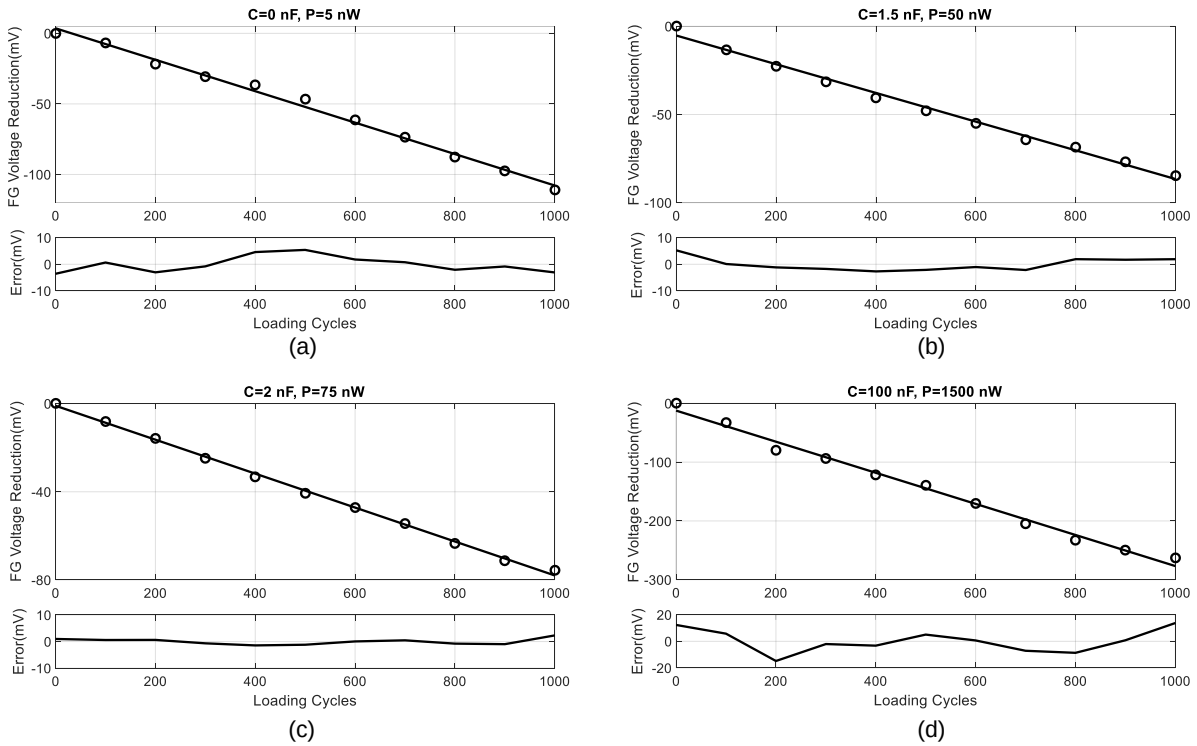


Figure 5.30: Linearization of the injector response for different levels of input power by choosing different values of the compensation capacitances.

method can achieve a linearity of more than 25 dB, which is significantly better than [97]. Note that in [46] we had reported a linear injector with more than 40 dB linear range; however, the injector could only be activated at power levels greater than 100 nW. This design therefore represents a significant improvement over the prior work. Also note that the linearity of the proposed injector can be further improved by choosing a larger value of the compensation capacitance, however at the expense of lower sensitivity. For instance, when the input power is reduced to as low as 5 nW, the injector shows a linear response without any compensation. However, when the input power level was increased the compensation circuit requires a larger capacitance for linearization. As an example, a 2 nF capacitor is sufficient to achieve a linearity of 25dB for an average input power of 75 nW applied over a duration of 1000 loading cycles. When the input power is increased to 1500 nW, the capacitance had

Table 5.5: Minimum compensation capacitance for linearization

Power	Capacitance
5 nW	0 nF
50 nW	1.5 nF
75 nW	2 nF
1.5 μW	100 nF
1 mW	100 μF(extrapolated)

to be increased to 100 nF. In principle the compensation circuit should be able to linearize the response of the injector beyond 1500 nW by appropriately choosing a larger value of the capacitance. Table 5.5 summarizes a prescription of a minimum value of the compensation capacitor that needs to be chosen to linearize the response of the injector. Note that a larger compensation capacitance would enhance the linearity of the injector, however, the sensitivity of the measurement will also reduce, as was demonstrated in Fig. 5.28. Conversely, for compensation capacitances greater than the minimum value shown in Table 5.5, the linearity would be enhanced for power levels lower than what is specified in Table 5.5. This implies that the value of the compensation capacitor is a design parameter that needs to be chosen based on the target application and the size and material properties of the piezoelectric transducer, as shown in Fig. 5.16. Also shown in Table 5.5 is the extrapolated minimum value of compensation capacitor required to linearize the injector at mW of input power. Thus, the proposed injector could potentially replace the linear injector that was previously used for self-powered monitoring of head-impacts in helmeted sports [32].

5.3.4 Summary

In this section, we proposed an improvement to our previously reported ultra-low power CMOS hot-electron injector by linearizing its response to different magnitudes of input power. As a result the proposed injector can be used in a variety of applications that require

continuous and implanted monitoring of biomechanical strain. For each application and for each input power range, the magnitude of the compensation capacitor needs to be chosen to be larger than a minimum experimentally determined value, to achieve the target linear response. Future work in this area will focus on adaptively choosing the value of the compensation capacitor based on historical statistics or by using a well defined deployment protocol.

5.4 Time-of-Occurrence Sensing

After introducing the details of the self-powered timers and low-power injectors, now we introduce the interface that combines the two for self-powered sensing of time-of-occurrence.

5.4.1 Operation Principle of the Timer-Injector

The principle of operation for the proposed time-of-occurrence sensing circuit is illustrated using an equivalent circuit model as shown in Fig. 5.31(a). The time-reference is implemented using the circuit a1 where a current sink I_{timer} discharges a pre-charged capacitor C_{timer}. As a result, the change in voltage V_{timer} serves as an indicator of the elapsed time. V_{timer} then modulates another current sink I_{inj}, as shown in the circuit a2 of Fig. 5.31(a), which

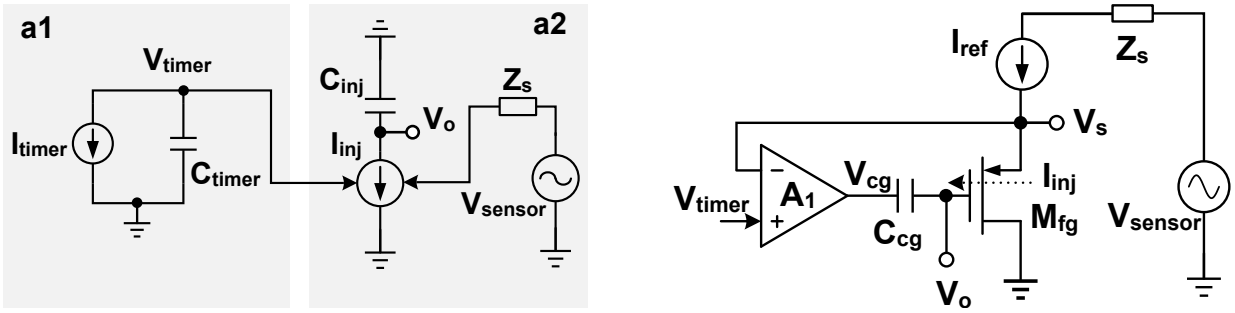


Figure 5.31: Principle of operation of the proposed time-injector circuit: (a) equivalent timer and injector circuits; (b) structure of a linear impact-ionized hot-electron injector.

discharges another capacitor C_{inj}. Because the current I_{inj} is supplied by the energy from a sensor transducer, the change in voltage V_{o} across C_{inj} encodes the time-of-occurrence of the event when the sensor was activated.

In chapter 2 [105] we demonstrated that the response of a basic FN-timer device is reliable over durations up to three years and is robust to fabrication mismatch. Here we use this basic FN-tunneling device to implement a chip-scale precision time-reference and design a self-powered system for sensing event time-of-occurrence. The relationship between timer output $V_{\text{timer}}(t)$ and time t has been derived in chapter 2 and we show here again as

$$V_{\text{timer}}(t) = \frac{k_2}{\ln\left(k_1 t + k_0\right)} + V_{\text{sub}}, \tag{5.25}$$

The circuit a2 in Fig. 5.31(a) can be implemented using an impact-ionized hot-electron injection on a floating-gate, which is a polysilicon strip that is completely insulated by high-quality silicon-di-oxide. This mechanism is more suitable than other possible implementations of a2 because: (a) the floating-gate acts as a non-volatile storage element, therefore, any injected charge is retained during the state when the timer-injector is unpowered; and (b) the physics of impact-ionized hot-electron injection matches the physics of a piezoelectric transducer [45], that is, at micro-strain levels of mechanical excitation, the transducer can easily generate large voltages ($\geq$4.2 V for 0.5 µm CMOS process) necessary to activate hot-electron injection [45]. Also, out of different types of p-IHEI that have been reported in literature [9, 45, 46, 97, 101], a linear variant, as shown in Fig. 5.31(b), is suitable for the proposed timer-injector circuit because it uses a feedback amplifier A_1 to isolate V_{timer} and the injection current I_{inj}. Also the negative feedback due to A_1 and the constant current reference I_{ref} ensures that all the floating-gate parasitics are held to a constant potential, as a result, the response of the injector is modulated only by the timer output V_{timer}. This can

be expressed as a function of V_{timer},

$$I_{\text{inj}}(t) = \beta I_{\text{ref}} \exp\left(\frac{V_{\text{timer}}(t)}{V_{\text{inj}}}\right), \tag{5.26}$$

where parameters β and V_{inj} are coefficients determined by the device form factors and process parameters. By substituting the expression for V_{timer} in (5.26), the output of the timer-injector circuit can be described by the following differential equation

$$C_{\text{inj}} \frac{dV_{\text{out}}}{dt} = -\beta I_{\text{ref}} \exp\left(\frac{k_2'}{\ln(k_1 t + k_0)} + \frac{V_{\text{sub}}}{V_{\text{inj}}}\right), \tag{5.27}$$

where $k_2' = k_2/V_{\text{inj}}$ and C_{inj} is the total floating-gate capacitance. Note that even if the timer discharge characteristics start saturating for long monitoring durations, the exponential nature of the hot-electron injection process partially compensates for the saturation effect. The closed-form solution to (5.27) is not readily expressible using straightforward mathematics, thus we utilize a linearization technique.

Under an assumption that the injection process, i.e. duration of an event, Δt is much smaller than the monitoring time t:

$$\Delta V_{\text{out}} = -\lambda \exp(\frac{\gamma_2}{\ln(k_1 t + k_0)})\Delta t, \tag{5.28}$$

where

$$\lambda = \frac{\gamma_1 I_{\text{ref}}}{C_{\text{inj}}}, \quad \gamma_1 = \beta \exp\left(\frac{V_{\text{sub}}}{V_{\text{inj}}}\right), \quad \gamma_2 = \frac{k_2}{V_{\text{inj}}}. \tag{5.29}$$

Thus, given an estimate of the event duration Δt and the measured change in the timer-injector output voltage ΔV_{out}, the time-of-occurrence of an event can be determined from

(5.28) as

$$t \approx \frac{1}{k_1} \exp\left(\frac{\gamma_2}{\log\left[-\frac{1}{\lambda}\frac{\Delta V_{\text{out}}}{\Delta t}\right]}\right) - \frac{k_0}{k_1}. \tag{5.30}$$

The approximation of (5.30) could be used to further estimate the sensitivity of the prediction when the measurement ΔV_{out} is noisy. The first-order error, Δt_{n}, in prediction can be approximated by

$$\Delta t_{\text{n}} \approx \left[\left(t + \frac{k_0}{k_1}\right) \frac{\gamma_2}{\log^2\left[-\frac{1}{\lambda}\frac{\Delta V_{\text{out}}}{\Delta t}\right]} \frac{1}{\Delta V_{\text{out}}}\right] \Delta V_{\text{n}}, \tag{5.31}$$

where ΔV_{n} is the measurement error. It can be seen from (5.31) that the prediction error monotonically increases with respect to the predicted time; furthermore, the random fluctuations induced by ΔV_{n} can be mitigated by either increasing ΔV_{out} to enhance the signal-to-noise ratio, or averaging out ΔV_{n}, which is assumed to be a zero-mean signal. Note that in this model, we have assumed the temperature is constant. In practice, temperature variations will introduce systematic noise into the model and compensation techniques will be necessary to maintain the prediction accuracy.

5.4.2 Circuit Implementation and System Design

FN Timer Array

The circuit level implementation of the FN timer array is shown in Fig. 5.32. The array comprises of several timer cells each of which could be selected using a switch S_1. The schematic of a timer cell is shown in Fig. 5.32 and comprises of two floating-gate transistors M_{fg} and M_{fgr}. One of the floating-gate transistors M_{fg} acts as an FN-tunneling junction and the other floating-gate transistor M_{fgr} is used for read-out in a source follower configuration. In addition to the source-follower the output V_{s} can also be accessed through a unity-gain

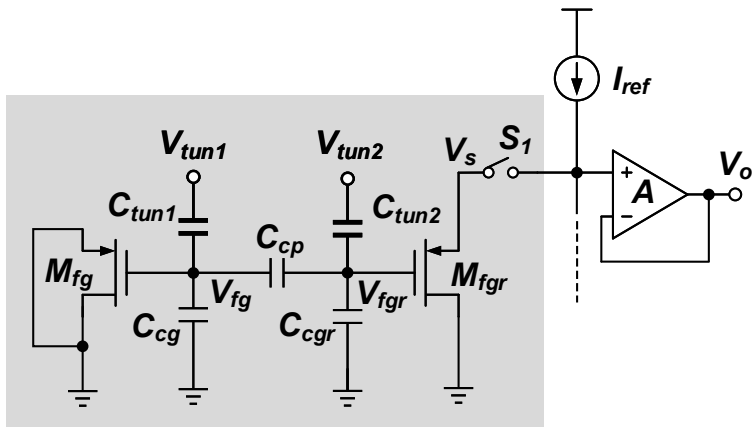


Figure 5.32: Schematic implementation of the FN timer array with a dual floating-gate read-out.

buffer A. The dual read-out architecture is useful to accurately calibrate the effect of the source follower.

Note that the dual-floating-gate architecture has been used because the FN-tunneling of electrons into the floating-gate M_{fg} occurs at relatively high-voltages >7 V which is well beyond the nominal supply voltage range of read-out circuits. Therefore, the dual floating-gate acts as a capacitive divider formed by C_{cp} and C_{cgr} to step-down the floating-gate voltage to V_{fgr}. Thus, V_{fgr} can be expressed in terms of the floating-gate charge Q_{fgr} as

$$V_{\mathrm{fgr}} = \frac{Q_{\mathrm{fgr}}}{C_{\mathrm{Tr}}} + \frac{C_{\mathrm{cp}}}{C_{\mathrm{Tr}} + C_{\mathrm{cp}}} V_{\mathrm{fg}}. \tag{5.32}$$

where C_{Tr} is the total capacitance seen from the gate node of M_{fgr} including C_{cgr} and gate capacitance of M_{fgr}. Note that the initial charge Q_{fgr} and hence V_{fgr} can be programmed to be within the nominal input dynamic range of the read-out circuits where the leakage is negligible [46]. Since Q_{fgr} can be assumed to be constant, V_{fgr} shows linear dependence on V_{fg}.

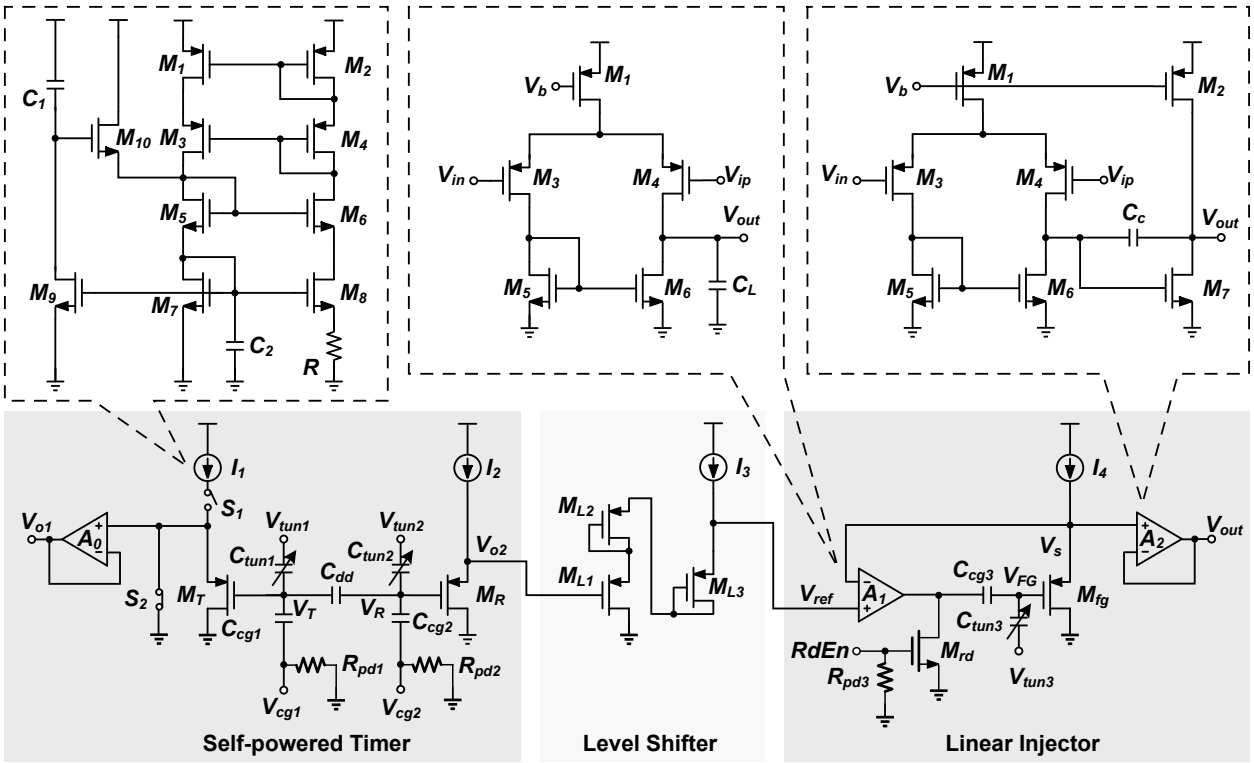


Figure 5.33: Integration of the timer and injector circuit shown with respective programming switches, pull-down resistors and schematic implementation of basic components (shown in the inset).

Timer-Injector Circuit Implementation

The integration of the timer and the p-IHEI circuit is shown in Fig. 5.33 along with their respective programming, read-out, and calibration switches. M_{T} and C_{cg1} form the floating-gate transistor which functions as the timer device and the timer read-out circuit (capacitive divider and source-follower) is formed by M_{R}, I_2, C_{dd} and C_{cg2} which provides access to the timer voltage V_{T} through V_{R} and V_{o2}. A unity-gain buffer A_0 is used for calibrating the response curve between the V_{o1} and V_{o2}, hence between V_{T} and V_{R}, based on (5.32) as

$$V_{\mathrm{o1}} = c_1 \cdot V_{\mathrm{o2}} + c_0. \tag{5.33}$$

Here c_1 and c_0 are determined by the initial conditions and value of the capacitors C_{cg1}, C_{dd} and C_{cg2}. The circuit also shows the pull-down resistors R_{pd1} and R_{pd2} which ensure that the nodes are in a well-defined state when unpowered. A level shifter formed by pMOS transistors M_{L1} to M_{L3} is used to match the output dynamic range of the timer to the input range of the linear injector, which is three times that of a pMOS source-to-gate voltage higher than V_{o2}. The linear injector feedback amplifier A_1 sets the source voltage of the injection transistor M_{fg} as $V_{\mathrm{s}} \approx V_{\mathrm{ref}} \approx V_{\mathrm{o2}} + 3V_{\mathrm{sg}}$. The change in injection voltage V_{FG} is therefore a function of timer output V_{o2}. V_{FG} is read-out by breaking the feedback loop using a pull-down transistor M_{rd}. In the read-out configuration, M_{fg} and I_4 act as a source-follower with V_{s} providing a measure of the stored floating-gate voltage V_{FG}. A unity-gain buffer A_2 is used for isolating different linear injector cells and for driving the external read-out circuits. The respective circuit level implementation of the current sources I_1–I_4 and the buffers A_0–A_2 are shown in the top portion of Fig. 5.33. Programming the floating-gate nodes V_{T}, V_{R} and V_{FG} employs a combination of FN tunneling and hot-electron injection, which is described in chapter 2 in detail and is neglect here for sake of brevity.

System Implementation

A complete implementation of the self-powered timer-injector system is shown in Fig. 5.34 which contains the harvesting, programming, and configuration modules. The chip-scale time-reference is comprised of a 4×4 timer cell array, each of which could be selected using a decoder. A decoder module is used for initializing and reading the timer cells and for isolating defective cells and clustering the output of valid cells. Two charge pumps are integrated on the time-reference chipset to enable independent programming of V_{T} and V_{R} to different initialization range. Note that for a 0.5 µm CMOS process, V_{T} needs to be initialized to voltages greater than 8 V whereas V_{R} needs to be initialized to voltages less than 3.3 V. The

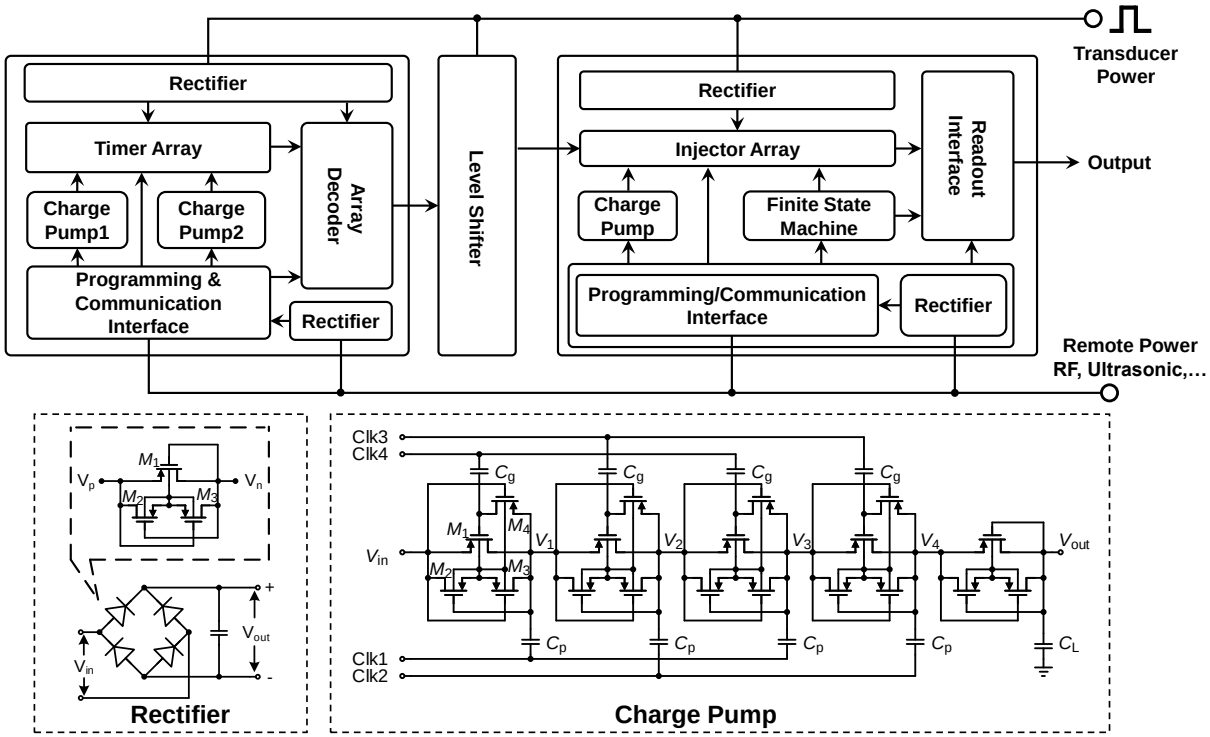


Figure 5.34: System level architecture and integration of the time-of-occurrence sensing system.

circuit-level implementation of the charge-pumps have been previously reported in [44] and is shown in Fig. 5.34's inset.

In the self-powered timing mode, the timer array continuously operates without any external power source whilst during the mechanical event, the transducer signal provides power (through the full-wave rectifier shown in Fig. 5.34 inset) for the timer read-out, level shifter, and linear injector. In this work, we have emulated the mechanical event using an electrical pulse with a well-controlled pulse duration. Note that this approximation is valid as long as the duration of the event is significantly shorter than the total monitoring period. The system block of the linear injector is also illustrated in Fig. 5.34, where an on-chip charge pump is integrated for initializing the floating-gate values. A finite-state-machine (FSM) functions as the control logic for channel read-out and programming together with the programming

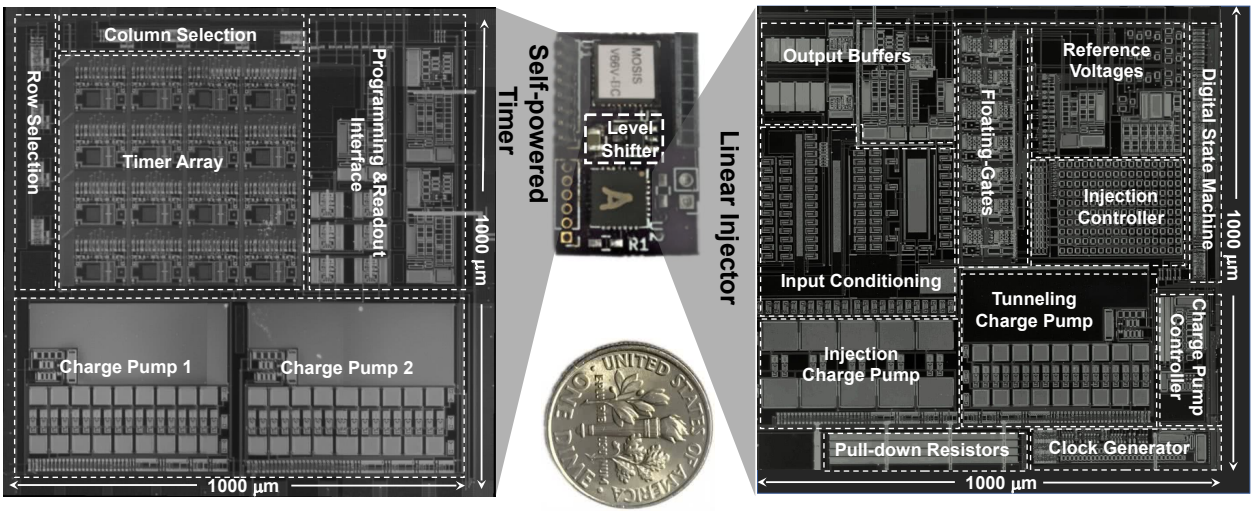


Figure 5.35: Integration of the self-powered time-stamped sensor on a PCB and die micro-photograph of the fabricated prototypes.

and communication interface. The stored value on each of the injector channels can be accessed through a read-out module. The energy required for programming and data transfer is provided through an external plug-and-play, radio-frequency or an ultrasound interface. Circuit details of the programming, FSM and remote power-transfer have been reported in [9, 32, 44] and are omitted here for the sake of brevity.

5.4.3 Measurement Results

Due to the difference in operating voltage requirements, the timer and the linear injector have been implemented on two different silicon dies. The micrographs of both silicon dies are shown in Fig. 5.35 and the chipsets have been integrated onto a printed circuit board (PCB) with a level-shifter acting as the interface circuitry. Separating the two substrates reduces the effect of substrate noise introduced by the power harvesting and charge-pump circuits in the linear injector. Also, isolating the timer and injector modules enables us to measure, calibrate and program the two modules independent of each other. Both the modules occupy

Table 5.6: Specifications of the Time-stamped Sensor

Parameter	Value	
Technology	0.5 µm CMOS	
PCB Size	$1.4 \times 2.2\,\mathrm{cm}^2$	
Die Sizes	$2 \times 1.5 \times 1.5\,\mathrm{mm}^2$	
Timer Dynamic Range	0.3 V	
Injector Linear Range	1.2 V for 1.8 V supply	
Monitoring Period	up to 3 years	
Power for 6 V supply	Timer Read-out	300 nW
	Level Shifter	120 nW
	Linear Injector	420 nW
	Total	840 nW

a $1000 \times 1000\,\mu\mathrm{m}^2$ die area and in Table 5.6 we summarize key specifications of the prototype system.

Timer Array Response

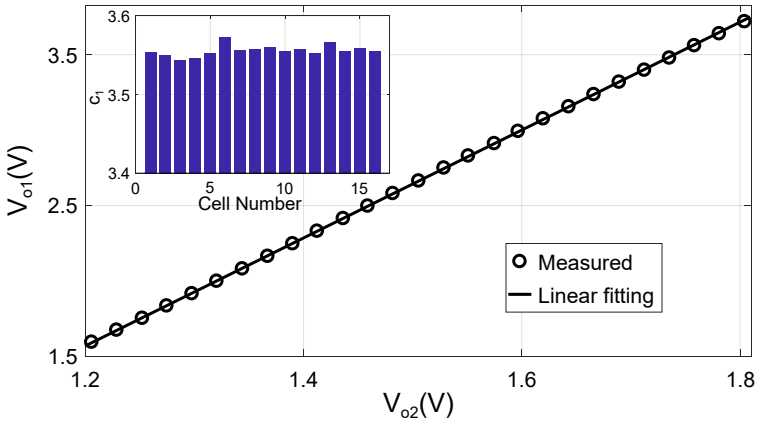


Figure 5.36: Measured read-out response of the timer: linear dependence between V_{o1} and V_{o2}, and distribution of coefficient c_1.

The first set of experiments characterized the response of the timer array. Because each element of the array is accessed using a dual-floating-gate capacitive divider, as depicted

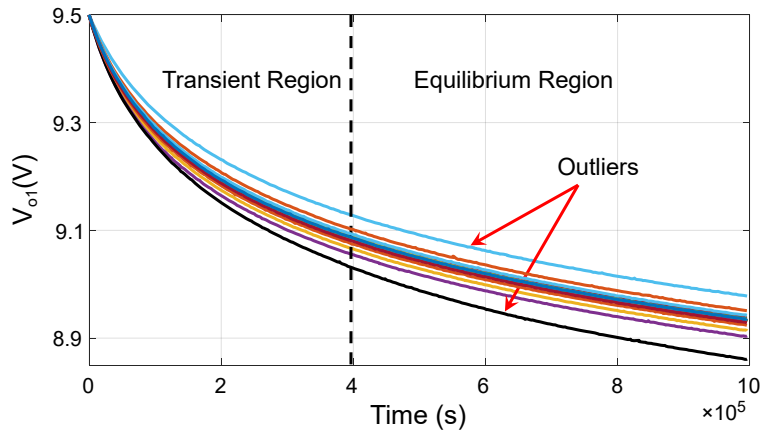


Figure 5.37: Measured temporal response of the array with 16 cells.

in Fig. 5.33, we characterized the dependence between the voltages V_{T} and V_{R} for different timer cells. By biasing V_{T} and V_{R} within a small read-out range, we obtained the relationship between V_{o1} and V_{o2} as shown in Fig. 5.36. The response shows a linearity larger than 50 dB, indicating that V_{o1} can be accurately estimated by measuring the voltage V_{o2} which modulates the injector current. The inset of Fig. 5.36 shows the estimated calibration coefficient c_1 for each cell in the timer array. c_1 was estimated to be 3.55±0.02 and the variation was found to be relatively small, therefore the average value of c_1 could be used for calibrating the entire array.

The second group of experiments were conducted to measure the time-keeping ability of the timer cells in the array. We first programmed the V_{T} of each timer-cell to around 8.5 V implying an output of V_{o1} close to 9.5 V to initiate FN tunneling through the gate-oxide barrier. Fig. 5.37 plots the temporal responses of each of the 16 timer cells in the array and the discharge characteristics conform to the model given in (5.25). Due to the mismatch in the threshold voltage of M_{R} and the FN tunneling junction, the measured timer responses (and hence the parameters $k_1, k_2, k_3, V_{\mathrm{sub}}$) would vary. As shown in Fig. 5.37, after the initial transient response, where the effects of mismatch are more prominent, each of the timer cells enter an equilibrium region where the change in timer voltages are relatively robust with

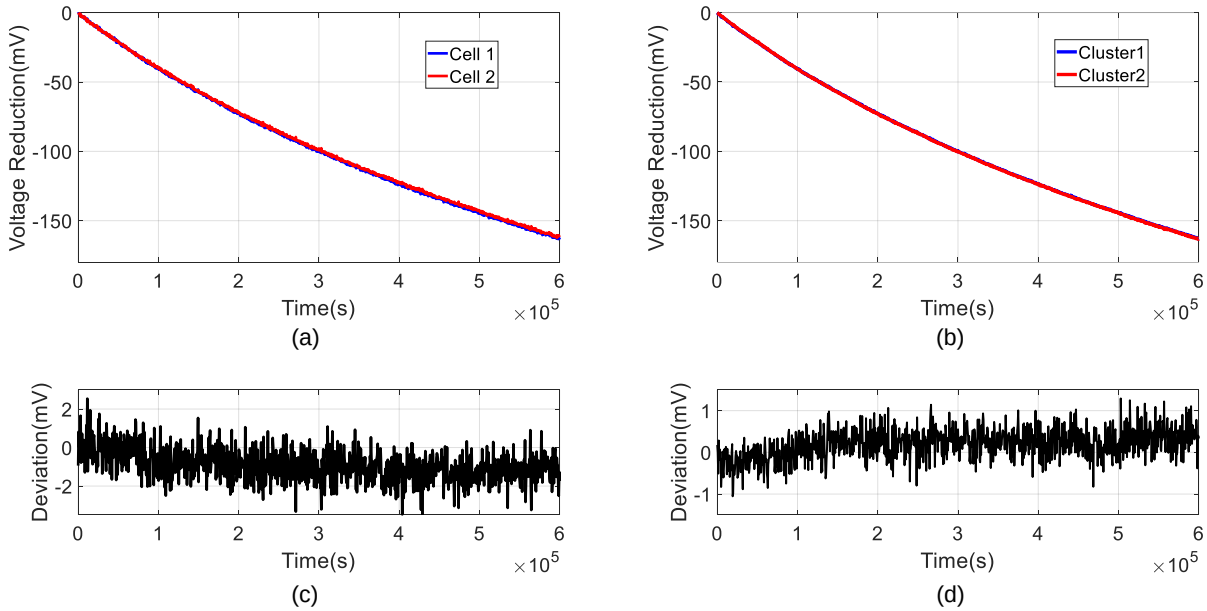


Figure 5.38: Comparison of the timer's response with and without clustering: (a) response of two timer cells in the equilibrium region, (b) deviation between the two cells response, (c) response of two clusters derived from the 16 cells, and (d) deviation between the response of the two clusters.

respect to one another. The physics of the equilibrium region are described in detail in [105] and are omitted here for the sake of brevity. As shown in Fig. 5.37, the measured responses is used to identify outliers of the timers. These timer cells are marked defective and disabled using the decoder.

The responses of multiple timer-cells can be combined to improve the synchronization accuracy. To demonstrate this we split the timers in the array into two clusters with equal number of cells, namely eight in this case. We chose four timers from each cluster such that the eight timers in each cluster were sorted according to their timer values, and only the middle four timers were kept and the remaining four were discarded. The timer responses corresponding to two separate cells operating in the equilibrium region are shown in Fig. 5.38(a) and the relative deviation with respect to each other is plotted in Fig. 5.38(c). The responses of the two timer clusters are plotted in Fig. 5.38(b) and the deviation with respect to the clusters

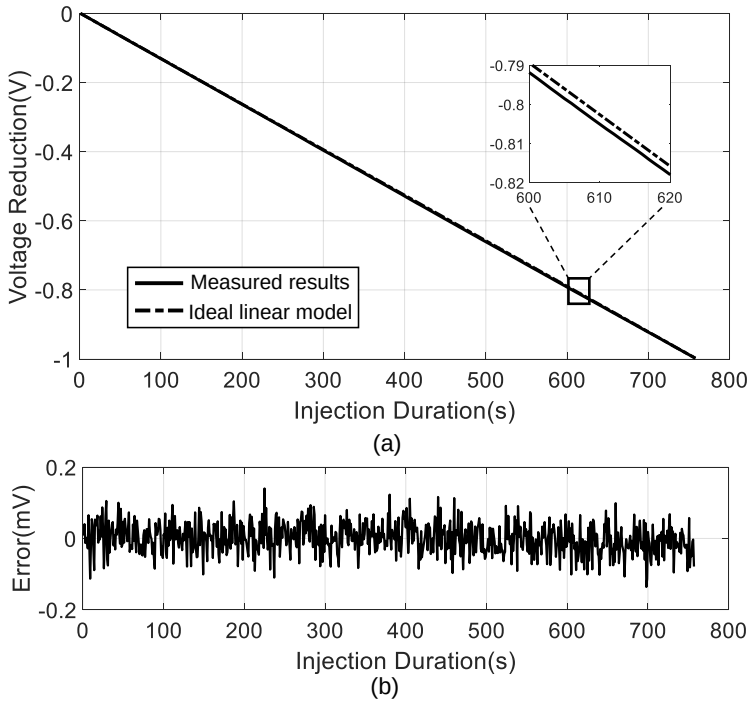


Figure 5.39: Measured linear injector's response showing: (a) the linear dependence of the injection voltage on the programming duration at V_{ref} =4.8 V and I_{s} =20 nA, and (b) difference between the measured results and the linear model.

is plotted in Fig. 5.38(d). The results show that the synchronization accuracy of the timer cluster is superior compared to individual timer cell by a factor of two. The maximum deviation for the two clusters is measured to be less than 1.5 mV across a dynamic range of 150 mV, implying a synchronization accuracy better than 1 % over a duration larger than 160 hours, or approximately $6{\cdot}10^5$ s.

Linear Injector Response

Fig. 5.39 shows the measured linear response of the injector across a 1 V dynamic programming range. As shown in Fig. 5.39(a), the reduction of the floating-gate voltage shows a linear dependence with the injection duration and Fig. 5.39(b) plots the programming accuracy of the linear injector. The maximum amplitude of the programming variation is less than 0.15 mV,

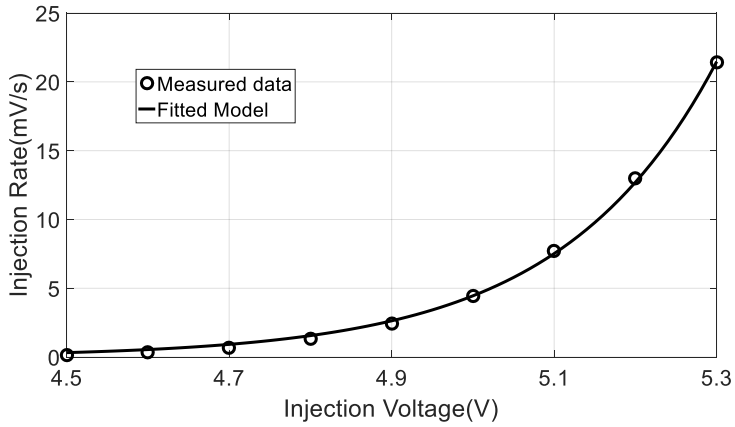


Figure 5.40: Dependence of injection rate on source to drain voltage.

which effectively gives 12.5 bit accuracy. We characterized the linearity for programming voltage V_s ranging from 4.5 V to 5.3 V and in worst case, the linearity is around 11 bits, which is consistent with the results in [46] when considering the additional supporting circuitry, such as output buffers utilized in this design.

Besides the linearity, another key feature of the injector is the accuracy of the model expressed by (5.26). The recovery of the time depends on the accuracy of the model. To characterize it, we measured the injection rate at different injection reference voltages, and the results are plotted in Fig. 5.40. The fitted data using the model of (5.26) is plotted as the solid line and the measured data are denoted with a circle. As evident, the measured data is strongly correlated to the fit model across the injection voltage reference range 4.5 V–5.3 V.

Time-of-occurrence Measurement

The next group of experiments were designed to validate the ability of the system to time-stamp events. As shown in Fig. 5.33, the reference terminal V_{ref} of the injector was connected to the level-shifted timer output. To emulate a mechanical event, the transducer input to the system (as shown in Fig. 5.34) was powered every 10 minutes for a duration of one second.

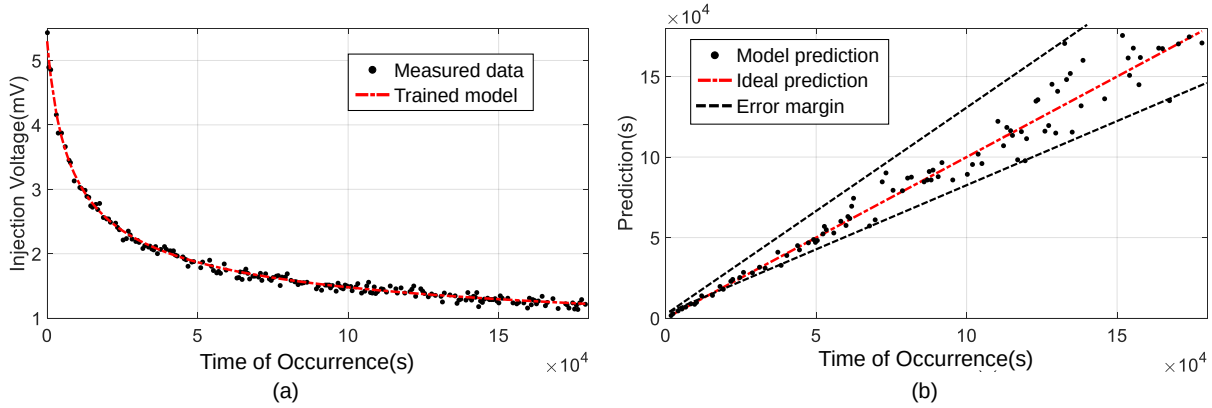


Figure 5.41: Training and testing using the measured response from the time-stamped injector: (a) data used for training marked as dot and the trained model plotted as red line; (b) predicted time using testing data marked with a dot.

The choice of the event duration is reasonable and has been validated previously for different real-world mechanical event monitoring scenarios [18, 32, 44, 97]. This process of powering was repeated for 3,000 minutes (or $1.8{\cdot}10^5$ s). The measured data were randomly split into two groups, 200 points for calibrating the model parameters in (5.28) and the remainder of the 100 points as a test set for verifying the model accuracy. Fig. 5.41(a) shows the training data points marked with black dots, and the red line tracks the trained model expressed by (5.28). This model was used to predict when a specified event occurred for the data in the test set. Fig. 5.41(b) shows the predicted time with respect to the true time-of-occurrence for the test set, where the red dashed line indicates the ideal prediction. Based on the measured results, the average relative time recovery accuracy is calculated to be 6.9 % across 50 hours ($1.8{\cdot}10^5$ s) of operation. The error margin plotted as dashed line indicates the worst case accuracy in predicting time-of-occurrence. Here the average accuracy is defined as:

$$\text{Average\_accuracy} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\tilde{t}_i - t_i}{t_i}\right| \tag{5.34}$$

where $\tilde{t}_i$ is the recovered time at time instant t_i.

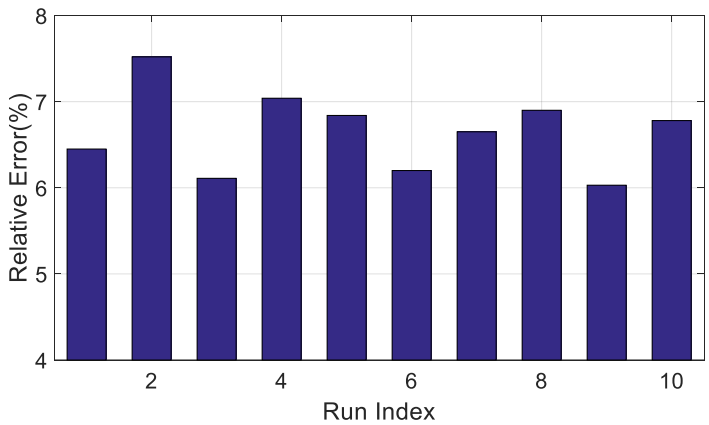


Figure 5.42: Relative time recovery accuracy across multiple runs.

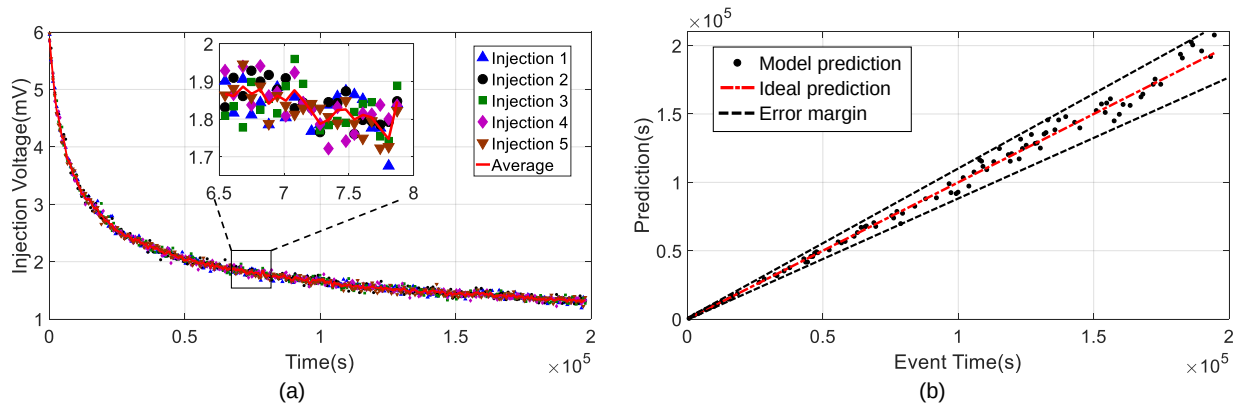


Figure 5.43: Using averaging to improve the time-of-occurrence estimation: (a) measured response of five injections at the same time instant and average response across the five measurements, and (b) predicted time using the average data for higher accuracy (the average relative accuracy is 3.2 %).

Another group of experiments was conducted to verify the robustness and reliability of the proposed time-stamping procedure. Using the aforementioned process, with the same event activation frequency and duration, the average relative error for each test set is shown in Fig. 5.42. The relative accuracy across 10 independent runs falls within 6–8 %, demonstrating reliable recovery performance.

The model of (5.28) indicates that the sensed value is a function of source current I_{s}, injection reference V_{ref} and the injection duration Δt, which implies that any noise introduced into

these three parameters will affect the reconstruction accuracy. For instance, the source current shows dependence on the temperature and power supply variations. The output of the level-shifter is also a function of environmental variations such as temperature and RF coupling. The injection duration shows random variations due to the intrinsic circuit start-up time, the pulse duration error introduced by the programming interface and so forth. For these small random fluctuations, techniques such as averaging, enlarging the injection duration, and tighter program control could help improve the reconstruction accuracy.

To validate the impact of random noise and variations, we conducted the time-stamping experiment using an averaging technique. For each time instant, we activate five of injector channels and recorded the measured output. Fig. 5.43(a) plots the measured data points for each injection and the average value across all the injections. As can be seen, the variance of each injector output is larger than that of the average, as expected. We further used the average output for parameter estimation and for time-of-occurrence prediction, and the result is shown in Fig. 5.43(b). The error margin is much closer to the ideal prediction than that shown in Fig. 5.41(b) and the relative accuracy is measured to be 3.2 %, which is an improvement over the 6.9 % for the individual injection. This finding corroborates the hypothesis that one is able to improve the time recovery accuracy by employing array implementations with multiple channels.

Another experiment was conducted to verify that the injection duration also affects time recovery accuracy. With a larger injection duration, the variations such as the start-up time will be proportionally small compared to the injection duration. The time-stamping experiment was conducted with event durations from one second to five seconds and the predicted accuracy for each duration is shown in Fig. 5.44. The results indicate that the relative error monotonically decreases with the increase of the injection duration, which validates the assumption regarding accuracy and injection duration. This result also suggests

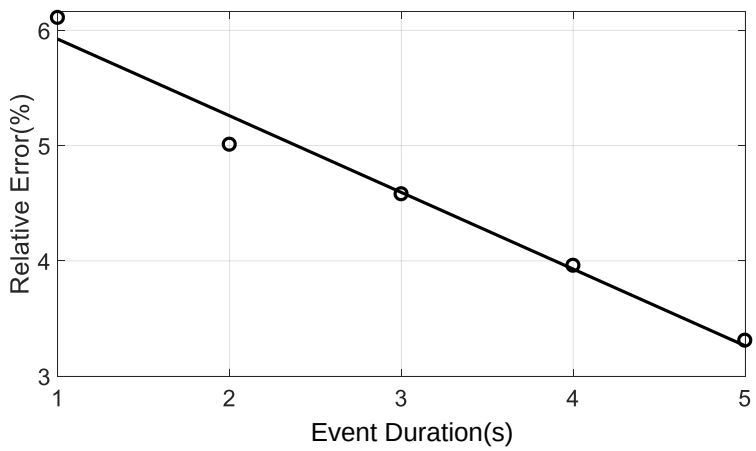


Figure 5.44: Dependence of relative time-of-occurrence estimation accuracy on the event duration.

that we can improve the time recovery accuracy by extending the duration of the event, possibly using our previous reported time-dilation technique [32, 101].

5.4.4 Accuracy Analysis

Because the total current consumption of our circuit (when fully activated) is limited by the PTAT current reference to 140 nA, the load to the mechanical (piezoelectric) transducer can be considered as a current sink. Under this condition, the voltage generated at the output of the transducer is indicative of the input power. Also, hot-electron injection in this CMOS process requires a minimum of 4.2 V drop across the source and the drain terminal of the floating-gate transistor. Thus, there is a minimum voltage requirement for the proposed sensor to log the occurrence of a mechanical event. As long as this minimum power and additional power to activate the measurement circuitry (total of 840 nW) is available from the mechanical event, the nature of the mechanical event need not be repeatable. For a single event with duration of one second, the total energy required would be 840 nJ — significantly smaller than what are available from mechanical impacts. When the available energy is much higher than the energy that can be consumed by the sensor circuitry, an energy buffering

front-end is required to accurately measure the magnitude of the impact. An example of this is our previously reported time-dilation circuitry used for harvesting energy from head impacts[32]. Note that once the energy is successfully buffered, a wide variety of data logging options is available including the use of a low power ADC and non-volatile memory. This can help improve the accuracy of the time-of-occurrence measurement.

With regards to estimating the time-of-occurrence, four types of errors could affect the accuracy. The first is due to the thermal-noise in the measurement circuits. This can be mitigated by choosing a larger injection duration that averages the errors and the sensitivity analysis summarized in (5.31) highlights the degree of mitigation. The second source of error is due to the start-up which is determined to be in the range of a few milliseconds. While we have assumed an event duration of one second, the actual time of injection is less than one second due to the inherent delay in the start-up circuits.

The systematic part of the startup time will be absorbed by the Δt term of (5.30), it does not affect the accuracy; nonetheless, the random part will degrade the time recovery accuracy. As discussed in previous section, this type of noise is random and can be mitigated by a larger injection duration or averaging technique. The third type of error source is model accuracy. Both models developed for the timer and injector consider only first-order effects and neglect the high-order effects' impact. For the timer model, the image charge barrier effect is neglected for model derivation. For the linear injector model, the finite gain of the feedback amplifier will result in a model mismatch in the second order. As a result, the model will introduce a systematic error to the time recovery process. Although beyond the scope of this paper, leveraging machine learning could allow for more precise and accurate models to improve the time recovery accuracy. The last factor that could affect the time recovery accuracy is temperature. In this work, the current generated by the PTAT current source has a strong temperature dependence, which will affect the timer readout and injection rate. This

issue can be resolved either by employing a bandgap reference or using temperature dependent R in Fig. 5.33 to compensate the temperature effect. The impact of the temperature on timer response was analyzed in [105] and Fig. 2.20 shows the dependence of timing accuracy of the timer on temperature. Since the timer is completely self-powered, active temperature compensation is infeasible here. Our future work will focus on exploiting differential response in an array of timers to compensate for temperature.

The lifetime of self-powered sensors can span several years, which requires a timing device that can operate on the same time scale. Literature [105] verified that the proposed timer device can operate as long as three years. However, there is a tradeoff between the operational lifetime and the time accuracy. As illustrated in Fig. 5.37, the timer output has a nonlinear dependence with respect to time, and exhibits a saturating behavior. This leads to weaker dependence of the injection voltage on time, thus a lower accuracy in time. A potential solution to this issue is to employ multiple timers and use some ensemble statistics of their outputs which will improve the accuracy, albeit at the cost of chip area and power. Also in this implementation, the capacitor divider used in the read-out circuit, reduces the dynamic range of the timer by a factor of 0.28. Thus, the dynamic range can be improved by exploring a different capacitor divider topology or by exploring other potential read-out circuits that can reduce the effect of signal attenuation. Note that the calibration time (3,000 minutes in this work, or $1.8 \cdot 10^5$ s) can be reduced to a few minutes by biasing the timer at a higher level (> 10 V). However, this will degrade the timing accuracy due to the error incurred in the parameter estimation.

5.5 Discussions

The topology of time-of-occurrence sensing in this chapter takes the event duration as a prior known parameter, which is not true in practical scenarios. Therefore, measurement circuit is necessary to extract the duration of event. This could be easily done by employing another linear injector which uses a constant injection voltage and current. The potential change in the floating-gate node will be proportional to the duration of the event. The error introduced by the measurement of duration may also degrade the time recovery accuracy, but this can also be mitigated by a cluster of techniques such as enlarging the duration using time-dilation circuit, averaging circuit.

Another topology that has not been discussed in this chapter is to employ a Analog-to-Digital Converter (ADC) to digitize the timer output and then use this output to select channel for injection. The whole monitoring period will be divided into several consecutive time intervals and each injection channel will correspond to one time intervals. This time sensitive channel selection architecture can also stamp time-of-occurrence into the sensing results. Unlike the previous mentioned time-stamping sensor which can recover the time in a continuous manner, this topology trades off the timing accuracy with sensing capability of multiple events. This topology will be future work to extend this research.

Although in this chapter time-stamping of mechanical events is extensively discussed and characterized, the timer can also be integrated with other self-powered sensing systems provided the sensor output can be used to modulate the sensing process. The generic stamping mechanism is either through a direct modulation or through a channel selection mechanism. This will be another research aspect of future work.

Chapter 6

Conclusions

The target of this dissertation is to tackle one fundamental problem: is it feasible to integrate self-powered time-keeping devices on standard CMOS process? Starting from the fundamental physical processes existing in CMOS technology, potential mechanisms were explored and analyzed to rationalize the architecture of the time-keeping device. It was shown that the device combining floating-gate structures with FN tunneling demonstrates necessary attributes for time-keeping.

The proposed timing devices demonstrates dynamic behavior which provides extra dimension of design freedom for passive and self-powered systems. By taking advantage of this dynamic process, several goals can be achieved: (a) time-stamping in self-powered sensing systems; (b) dynamic footprint detection based on the near-ideal synchronization performance and (c) security enhancement in passive IoT systems.

Prototyped timing devices and optimized system-level integration were implemented to validate the concept and design target using mainstream CMOS processes. Specifications of the proposed device including but not limited to power, footprint, accuracy, robustness

and duration of operation were fully characterized. Systems that use the timer for specific applications were also successfully implemented.

6.1 Summary of Contributions

The major contributions of this dissertation are summarized as follows.

• Exploring and designing self-powered time-keeping devices on CMOS process.

The leakage processes in CMOS process were explored and analyzed, and it was demonstrated that only FN tunneling satisfies the performance requirements on duration and robustness. Timing device based upon FN tunneling was fabricated on standard CMOS process and its dynamic behavior was fully characterized. A mathematical model that can dedicatedly capture the dynamics of the FN tunneling timers was proposed and it was shown that the model accuracy is better than 40 dB. The timer's robustness to mismatch and ambient variations were also investigated and it was shown that the timer is highly robust to non-ideal factors. Long-term study was also conducted to verify that the timer can operate as long as several years, validating the feasibility for passive systems. Scalability of the timer device was also investigated by fabricating and optimizing the device on more advanced CMOS technology nodes, and the model still remains accurate.

• Investigating passive synchronization and desynchronization in FN tunneling timers.

The proposed self-powered timers demonstrate perfect synchronization behavior when they experience identical environment footprint. However, they can be desynchronized due to multiple mechanisms, which were studied and analyzed. A mathematical model that can capture the dynamic desynchronization behavior was proposed and demonstrates good

accuracy. The application of using this dynamic footprint detection was also explored and discussed.

• Proposing dynamic authentication protocols for security enhancement based on FN timers.

The synchronization attribute of the FN timers enables dynamic authentication in passive devices and systems. Protocols similar to SecureID type of authentication were studied and proposed for passive IoT devices using the synchronized timers. The security performance and computation cost of the protocol were analyzed to validate its usage in passive devices whose resources are limited. An enhanced version of the protocol namely HPMAP was further proposed to preserve the privacy during authentication process by employing two pairs of synchronized timers. Simulation results verified the performance and feasibility of the proposed protocols.

• Implementing self-powered sensing of time-of-occurrence using FN timers.

A sensing modality that uses the output of the timer to modulate the behavior of a mechanical sensor was proposed and characterized to achieve self-powered sensing of time-of-occurrence of mechanical events. The sources of error were analyzed in detail and techniques that can improve the performance were also proposed and verified. Measurement results show that the average time reconstruction accuracy can be improved from 6.9% to 3.2% for a monitoring period of 50 hours.

6.2 Directions of Future Work

Future work will mainly focus on the following aspects.

• Compensating temperature dependence of FN tunneling for timing accuracy improvement.

While the timer is robust to mismatch, the temperature dependence could degrade the absolute timing accuracy. Therefore, passive temperature compensation techniques will be studied to improve the accuracy. One possible technique is to employ an array of timers with unique form factors to extract the common-mode temperature noise and compensate it.

• Exploring fundamental limits of the timing accuracy.

The tunneling behavior of the FN timer depends on the bias of the floating-gate node. When it approaches the limit of tunneling where single electron behavior shows up, statistical behavior will dominate the timing accuracy. We will explore the statistical properties of the tunneling behavior in the timer device settings. This behavior might show some intrinsic connections with the prime number distributions.

• Extending self-powered sensing of time-of-occurrence.

The self-powered time-of-occurrence sensing of mechanical events was already studied in this dissertation. However, this architecture can be extended to much more broad areas such as RF or solar exposure events, mechanical tampering events and so forth. Moreover, a self-consistent and complete sensing system is preferred for practical implementation.

• Implementation of dynamic authentication in passive systems.

The dynamic authentication protocol has been verified to be effective and efficient in passive and self-powered IoT devices. Hardware implementation of the protocol will be the next step to verify the system-level performance. The token generation algorithms will be implemented using hardware to verify the efficacy.

References

[1] URL: http://www.bbc.com/news/world-asia-china-35878624..

[2] URL: http://www.emc.com/security/rsa-securid.htm..

[3] https://www.cryptopp.com/benchmarks.html.

[4] E Abad, F Palacio, M Nuin, A Gonzalez De Zarate, A Juarros, JM Gómez, and S Marco. "RFID smart tag for traceability and cold chain monitoring of foods: Demonstration in an intercontinental fresh fish logistic chain". In: *Journal of food engineering* 93.4 (2009), pp. 394–399.

[5] Syed A Ahson and Mohammad Ilyas. *RFID handbook: applications, technology, security, and privacy*. CRC press, 2008.

[6] Yarub Alazzawi, Chunqui Qian, and Shantanu Chakrabartty. "Feasibility of non-contact ultrasound generation using implanted metallic surfaces as electromagnetic acoustic transducers". In: *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*. IEEE. 2015, pp. 1–4.

[7] Sara Amendola, Rossella Lodato, Sabina Manzari, Cecilia Occhiuzzi, and Gaetano Marrocco. "RFID technology for IoT-based personal healthcare in smart spaces". In: *IEEE Internet of Things Journal* 1.2 (2014), pp. 144–152.

[8] Rebecca Angeles. "RFID technologies: supply-chain applications and implementation issues". In: *Information systems management* 22.1 (2005), pp. 51–65.

[9] Kenji Aono, Tracey Covassin, and Shantanu Chakrabartty. "Monitoring of Repeated Head Impacts using Time-dilation based Self-powered Sensing". In: *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. 2014, pp. 1620–1623.

[10] Kenji Aono, Nizar Lajnef, Fred Faridazar, and Shantanu Chakrabartty. "Infrastructural health monitoring using self-powered Internet-of-Things". In: *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. IEEE. 2016, pp. 2058–2061.

[11] Hiroki Asano, Tetsuya Hirose, Keishi Tsubaki, T Miyoshi, Toshihiro Ozaki, Nobutaka Kuroki, and Masahiro Numa. "A 1.66-nW/kHz, 32.7-kHz, 99.5 ppm/° C Fully Integrated Current-mode RC Oscillator for Real-time Clock Applications with PVT Stability". In: *European Solid-State Circuits Conference, ESSCIRC Conference 2016: 42nd*. IEEE, 2016, pp. 149–152.

[12] Luigi Atzori, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey". In: *Computer networks* 54.15 (2010), pp. 2787–2805.

[13] Mihir Bellare and Phillip Rogaway. "Collision-resistant hashing: Towards making UOWHFs practical". In: *Advances in Cryptology - Crypto'97, Lecture Notes in Computer Science* 1294 (1997), pp. 470–484.

[14] Charles H Bennett. "The thermodynamics of computation—a review". In: *International Journal of Theoretical Physics* 21.12 (1982), pp. 905–940.

[15] H.C. Berg. *Random walks in biology*. Princeton University Press, 1993.

[16] Bruce D Beynnon and Braden C Fleming. "Anterior cruciate ligament strain in-vivo: a review of previous work". In: *Journal of biomechanics* 31.6 (1998), pp. 519–525.

[17] Rafik H Bishara. "Cold chain management-an essential component of the global pharmaceutical supply chain". In: *American Pharmaceutical Review* 9.1 (2006), pp. 105–109.

[18] Wassim Borchani, Kenji Aono, Nizar Lajnef, and Shantanu Chakrabartty. "Monitoring of Postoperative Bone Healing Using Smart Trauma-Fixation Device With Integrated Self-Powered Piezo-Floating-Gate Sensors". In: *IEEE Transactions on Biomedical Engineering* 63.7 (2016), pp. 1463–1472.

[19] J Borodkin, N Caldwell, and S Hollister. "The influence of mechanical strain magnitude on bone adaptation to porous-coated implants". In: *[Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meetring of the Biomedical Engineering Society] BMES/EMBS Conference, 1999. Proceedings of the First Joint*. Vol. 2. IEEE. 1999, 768–vol.

[20] Richard Andrew Bullen, TC Arnot, JB Lakeman, and FC Walsh. "Biofuel cells and their development". In: *Biosensors and Bioelectronics* 21.11 (2006), pp. 2015–2045.

[21] Shantanu Chakrabartty and Nizar Lajnef. "Infrasonic power-harvesting and nanowatt self-powered sensors". In: *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*. IEEE. 2009, pp. 157–160.

[22] Hong Chen, Ming Liu, Wenhan Hao, Yi Chen, Chen Jia, Chun Zhang, and Zihua Wang. "Low-power circuits for the bidirectional wireless monitoring system of the orthopedic implants". In: *IEEE transactions on biomedical circuits and systems* 3.6 (2009), pp. 437–443.

[23] Joseph J Crisco et al. "Frequency and location of head impact exposures in individual collegiate football players". In: *Journal of athletic training* 45.6 (2010), pp. 549–559.

[24] Canan Dagdeviren et al. "Conformal piezoelectric systems for clinical and experimental characterization of soft tissue biomechanics". In: *Nature materials* 14.7 (2015), p. 728.

[25] Simone Dalola, Marco Ferrari, Vittorio Ferrari, Michele Guizzetti, Daniele Marioli, and Andrea Taroni. "Characterization of thermoelectric modules for powering autonomous sensors". In: *IEEE Transactions on Instrumentation and Measurement* 58.1 (2009), pp. 99–107.

[26] Ivan Damgard. "Collision free hash functions and public key signature schemes." In: *EUROCRYPT* (1987), pp. 203–216.

[27] Mohsen Darianian and Martin Peter Michael. "Smart home mobile RFID-based Internet-of-Things systems and services". In: *Advanced Computer Theory and Engineering, ICACTE'08*. 2008, pp. 116–120.

[28] Supriyo Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[29] Michael Dunbar. "Plug-and-play sensors in wireless networks". In: *Instrumentation & Measurement Magazine, IEEE* 4.1 (2001), pp. 19–23.

[30] Niell G Elvin, Nizar Lajnef, and Alex A Elvin. "Feasibility of structural monitoring with vibration powered sensors". In: *Smart materials and structures* 15.4 (2006), p. 977.

[31] Biyi Fang, Tao Feng, Mi Zhang, and Shantanu Chakrabartty. "Feasibility of B-mode diagnostic ultrasonic energy transfer and telemetry to a cm 2 sized deep-tissue implant". In: *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 782–785.

[32] Tao Feng, Kenji Aono, Tracey Covassin, and Shantanu Chakrabartty. "Self-powered Monitoring of Repeated Head Impacts Using Time-dilation Energy Measurement Circuit". In: *IEEE transactions on biomedical circuits and systems* 9.2 (2015), pp. 217–226.

[33] J. Fu, C. Wu, X. Chen, R. Fan, and L. Ping. "Scalable pseudo random RFID private mutual authentication." In: *2nd IEEE International Conference on Computer Engineering and Technology (ICCET)* (2010), pp. 497–500.

[34] Luke M Gessel, Sarah K Fields, Christy L Collins, Randall W Dick, and R Dawn Comstock. "Concussions among United States high school and collegiate athletes". In: *Journal of athletic training* 42.4 (2007), pp. 495–503.

[35] Elizabeth Gibney. "Atomic clocks face off: Next generation of hyper-precise timekeepers can only be tested against each other". In: *Nature* 522.7544 (2015), pp. 16–18.

[36] Victor Giurgiutiu, Andrei Zagrai, and Jing Jing Bao. "Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring". In: *Structural Health Monitoring* 1.1 (2002), pp. 41–61.

[37] Rupert Thomas Gould and Lt Commander. *The marine chronometer: its history and development*. JD Potter London, 1923.

[38] Yang Guo, Christopher Aquino, David Zhang, and Boris Murmann. "A Four-Channel, +-36 V, 780 kHz Piezo Driver Chip for Structural Health Monitoring". In: *IEEE Journal of Solid-State Circuits* 49.7 (2014), pp. 1506–1513.

[39] M H Afifi, Liang Zhou, Shantanu Chakrabartty, and Jian Ren. "Dynamic Authentication Protocol Using Self-powered Timers for Passive Internet of Things". In: PP (Sept. 2017).

[40] R. L. T. Hampton. "A hybrid analog-digital pseudo-random noise generator". In: *Proceedings of the spring joint computer conference, ACM* (1964), pp. 287–301.

[41] Paul Edward Hasler. "Foundations of learning in analog VLSI". PhD thesis. California Institute of Technology, 1997.

[42] N Hinkley et al. "An atomic clock with 10–18 instability". In: *Science* 341.6151 (2013), pp. 1215–1218.

[43] Yingzhe Hu, Liechao Huang, Warren SA Rieutort-Louis, Josue Sanz-Robinson, James C Sturm, Sigurd Wagner, and Naveen Verma. "A self-powered system for large-scale strain sensing by combining CMOS ICs with large-area electronics". In: *IEEE Journal of Solid-State Circuits* 49.4 (2014), pp. 838–850.

[44] Chenling Huang and Shantanu Chakrabartty. "An asynchronous analog self-powered CMOS sensor-data-logger with a 13.56 MHz RF programming interface". In: *IEEE Journal of Solid-State Circuits* 47.2 (2012), pp. 476–489.

[45] Chenling Huang, Nizar Lajnef, and Shantanu Chakrabartty. "Calibration and characterization of self-powered floating-gate usage monitor with single electron per second operational limit". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 57.3 (2010), pp. 556–567.

[46] Chenling Huang, Pikul Sarkar, and Shantanu Chakrabartty. "Rail-to-rail, linear hot-electron injection programming of floating-gate voltage bias generators at 13-bit resolution". In: *IEEE Journal of Solid-State Circuits* 46.11 (2011), pp. 2685–2692.

[47] Ari Juels. "RFID security and privacy: A research survey". In: *IEEE journal on selected areas in communications* 24.2 (2006), pp. 381–394.

[48] H-U Kim, W-H Lee, HV Rasika Dias, and Shashank Priya. "Piezoelectric microgenerators-current status and challenges". In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.8 (2009).

[49] John Frank Charles Kingman. *Poisson processes*. Vol. 3. Clarendon Press, 1992.

[50] Svenja Knappe, Vishal Shah, Peter DD Schwindt, Leo Hollberg, John Kitching, Li-Anne Liew, and John Moreland. "A microfabricated atomic clock". In: *Applied Physics Letters* 85.9 (2004), pp. 1460–1462.

[51] Sri Harsha Kondapalli, Yarub Alazzawi, Marcin Malinowski, Tomasz Timek, and Shantanu Chakrabartty. "Multi-access In-vivo Biotelemetry Using Sonomicrometry and M-scan Ultrasound Imaging". In: *IEEE Transactions on Biomedical Engineering* (2017).

[52] Ju-Chia Kuo and Mu-Chen Chen. "Developing an advanced multi-temperature joint distribution system for the food cold chain". In: *Food Control* 21.4 (2010), pp. 559–566.

[53] Nizar Lajnef, Niell G Elvin, and Shantanu Chakrabartty. "A piezo-powered floating-gate sensor array for long-term fatigue monitoring in biomechanical implants". In: *IEEE transactions on biomedical circuits and systems* 2.3 (2008), pp. 164–172.

[54] Man Kay Law, Amine Bermak, and Howard C Luong. "A Sub-uW Embedded CMOS Temperature Sensor for RFID Food Monitoring Application". In: *IEEE journal of solid-state circuits* 45.6 (2010), pp. 1246–1255.

[55] Yoonmyung Lee, Bharan Giridhar, Zhiyoong Foo, Dennis Sylvester, and David B Blaauw. "A sub-nW multi-stage temperature compensated timer for ultra-low-power sensor nodes". In: *IEEE Journal of Solid-State Circuits* 48.10 (2013), pp. 2511–2521.

[56] M Lenzlinger and EH Snow. "Fowler-Nordheim Tunneling into Thermally Grown SiO2". In: *Journal of Applied physics* 40.1 (1969), pp. 278–283.

[57] Hélène Lhermet, Cyril Condemine, Marc Plissonnier, Raphaël Salot, Patrick Audebert, and Marion Rosset. "Efficient power management circuit: From thermal energy harvesting to above-IC microbattery energy storage". In: *IEEE Journal of solid-state circuits* 43.1 (2008), pp. 246–255.

[58] C. Lim and T. Kwon. "Strong and robust RFID authentication enabling perfect ownership transfer." In: *Conference on Information and Communications Security (ICICS)* (2006), pp. 1–20.

[59] Yu-Shiang Lin, Dennis Sylvester, and David Blaauw. "A sub-pW timer using gate leakage for ultra low-power sub-Hz monitoring systems". In: *Custom Integrated Circuits Conference, 2007. CICC'07. IEEE*. IEEE, 2007, pp. 397–400.

[60] Aikaterini Mitrokotsa, Melanie R Rieback, and Andrew S Tanenbaum. "Classifying RFID attacks and defenses". In: *Information Systems Frontiers* (2010).

[61] Moni Naor and Moti Yung. "Universal one-way hash functions and their cryptographic applications." In: *STOC* (1989), pp. 33–43.

[62] Micah O'Halloran and Rahul Sarpeshkar. "A 10-nW 12-bit accurate analog storage cell with 10-aA leakage". In: *IEEE journal of solid-state circuits* 39.11 (2004), pp. 1985–1996.

[63] M. Ohkubo, K. Suzuki, and S. Kinoshita. "Hash-Chain Based Forward Secure Privacy Protection Scheme for Low-Cost RFID". In: *Proc. of the Symposium on Cryptography and Information Security* (2004), pp. 719–724.

[64] Gyuhae Park, Tajana Rosing, Michael D Todd, Charles R Farrar, and William Hodgkiss. "Energy harvesting for structural health monitoring sensor networks". In: *Journal of Infrastructure Systems* 14.1 (2008), pp. 64–79.

[65] Brent M Phares, Terry J Wipf, Lowell F Greimann, and Yoon-Si Lee. *Health Monitoring of Bridge Structures And Components Using Smart-Structure Technology. VOLUME I.* Tech. rep. 2005.

[66] Stephen R Platt, Shane Farritor, Kevin Garvin, and Hani Haider. "The use of piezoelectric ceramics for electric power generation within orthopedic implants". In: *IEEE/ASME transactions on mechatronics* 10.4 (2005), pp. 455–461.

[67] Radislav A Potyrailo, Nandini Nagraj, Zhexiong Tang, Frank J Mondello, Cheryl Surman, and William Morris. "Battery-free Radio Frequency Identification (RFID) Sensors for Food Quality and Safety". In: *Journal of Agricultural and Food Chemistry* 60.35 (2012), pp. 8535–8543.

[68] Rabindra L Pradhan, Eiji Itoi, Yuji Hatakeyama, Masakazu Urayama, and Kozo Sato. "Superior Labral Strain during the Throwing Motion A Cadaveric Study". In: *The American Journal of Sports Medicine* 29.4 (2001), pp. 488–492.

[69] R Ramprasad. "Phenomenological theory to model leakage currents in metal–insulator–metal capacitor systems". In: *physica status solidi (b)* 239.1 (2003), pp. 59–70.

[70] NM Ravindra and Jin Zhao. "Fowler-Nordheim tunneling in thin SiO2 films". In: *Smart Materials and Structures* 1.3 (1992), p. 197.

[71] E Romero, RO Warrington, and MR Neuman. "Energy scavenging sources for biomedical sensors". In: *Physiological measurement* 30.9 (2009), R35.

[72] Shad Roundy, Paul K. Wright, and Jan Rabaey. "A study of low level vibrations as a power source for wireless sensor nodes". In: *Computer Communications* 26.11 (2003). Ubiquitous Computing, pp. 1131–1144.

[73] Luis Ruiz-Garcia and Loredana Lunadei. "The role of RFID in agriculture: Applications, limitations and challenges". In: *Computers and Electronics in Agriculture* 79.1 (2011), pp. 42–50.

[74] Pikul Sarkar and Shantanu Chakrabartty. "Compressive Self-Powering of Piezo-Floating-Gate Mechanical Impact Detectors". In: *Circuits and Systems I: Regular Papers, IEEE Transactions on* 60.9 (2013), pp. 2311–2320.

[75] Pikul Sarkar, Chenling Huang, and Shantanu Chakrabartty. "An ultra-linear piezo-floating-gate strain-gauge for self-powered measurement of quasi-static-strain". In: *Biomedical Circuits and Systems, IEEE Transactions on* 7.4 (2013), pp. 437–450.

[76] S. E. Sarma, S. A. Weis, and D. W. Engels. "RFID Systems and Security and Privacy Implications". In: *CHES, Springer-Verlag* (2003), pp. 454–469.

[77] Edward Sazonov, Haodong Li, Darrell Curry, and Pragasen Pillay. "Self-powered sensors for monitoring of highway bridges". In: *IEEE Sensors Journal* 9.11 (2009), pp. 1422–1429.

[78] Loren Schwiebert, Sandeep KS Gupta, and Jennifer Weinmann. "Research challenges in wireless networks of biomedical sensors". In: *Proceedings of the 7th annual international conference on Mobile computing and networking*. ACM. 2001, pp. 151–165.

[79] Dongjin Seo, Ryan M Neely, Konlin Shen, Utkarsh Singhal, Elad Alon, Jan M Rabaey, Jose M Carmena, and Michel M Maharbiz. "Wireless recording in the peripheral nervous system with ultrasonic neural dust". In: *Neuron* 91.3 (2016), pp. 529–539.

[80] Aatmesh Shrivastava and Benton H Calhoun. "A 150nW, 5ppm/° C, 100kHz On-Chip Clock Source for Ultra Low Power SoCs". In: *Custom Integrated Circuits Conference (CICC), 2012 IEEE*. IEEE, 2012, pp. 1–4.

[81] Boyeon Song and Chris J. Mitchell. "RFID Authentication Protocol for Low-cost Tags". In: *Proceedings of the First ACM Conference on Wireless Network Security*. 2008, pp. 140–147.

[82] Seth Stein and Michael Wysession. *An introduction to seismology, earthquakes, and earth structure*. John Wiley & Sons, 2009.

[83] Ian StJohn and Robert M Fox. "Leakage effects in metal-connected floating-gate circuits". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 53.7 (2006), pp. 577–579.

[84] PS Taoukis and TP Labuza. "Applicability of time-temperature indicators as shelf life monitors of food products". In: *Journal of Food Science* 54.4 (1989), pp. 783–788.

[85] Theofania Tsironi, Efimia Dermesonlouoglou, Maria Giannakourou, and Petros Taoukis. "Shelf life modelling of frozen shrimp at variable temperature conditions". In: *LWT-Food Science and Technology* 42.2 (2009), pp. 664–671.

[86] Eric Vittoz and Jean Fellrath. "CMOS analog integrated circuits based on weak inversion operations". In: *Solid-State Circuits, IEEE Journal of* 12.3 (1977), pp. 224–231.

[87] Fei-Yue Wang, Daniel Zeng, and Liuqing Yang. "Smart cars on smart roads: an IEEE intelligent transportation systems society update". In: *IEEE Pervasive Computing* 5.4 (2006), pp. 68–69.

[88] Ning Wang, Naiqian Zhang, and Maohua Wang. "Wireless sensors in agriculture and food industry—Recent development and future perspective". In: *Computers and electronics in agriculture* 50.1 (2006), pp. 1–14.

[89] S.A. Weis, S.E. Sarma, R.L. Rivest, and D.W. Engels. "Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems". In: *Proc. of the 1st International Conference on Security in Pervasive Computing* (2003), pp. 201–212.

[90] Thomas W Wright, Frank Glowczewskie, David Cowin, and Donna L Wheeler. "Ulnar nerve excursion and strain at the elbow and wrist associated with upper extremity motion". In: *The Journal of hand surgery* 26.4 (2001), pp. 655–662.

[91] Yee Chia Yeo, Qiang Lu, Wen Chin Lee, Tsu-Jae King, Chenming Hu, Xiewen Wang, Xin Guo, and TP Ma. "Direct tunneling gate leakage current in transistors with ultrathin silicon nitride gate dielectric". In: *IEEE Electron Device Letters* 21.11 (2000), pp. 540–542.

[92] Mingquan Yuan, Evangelyn C Alocilja, and Shantanu Chakrabartty. "A novel biosensor based on silver-enhanced self-assembled radio-frequency antennas". In: *IEEE Sensors Journal* 14.4 (2014), pp. 941–942.

[93] Mingquan Yuan, Evangelyn C Alocilja, and Shantanu Chakrabartty. "Self-powered wireless affinity-based biosensor based on integration of paper-based microfluidics and self-assembled RFID antennas". In: *IEEE transactions on biomedical circuits and systems* 10.4 (2016), pp. 799–806.

[94] Mingquan Yuan, Qisheng Jiang, Keng-Ku Liu, Srikanth Singamaneni, and Shantanu Chakrabartty. "Towards an Integrated QR Code Biosensor: Light-Driven Sample Acquisition and Bacterial Cellulose Paper Substrate". In: *IEEE Transactions on Biomedical Circuits and Systems* (2018).

[95] Chao Zhang et al. "Time–Temperature indicator for perishable products based on kinetically programmable Ag overgrowth on Au nanorods". In: *ACS nano* 7.5 (2013), pp. 4561–4568.

[96] L. Zhou and S. Chakrabartty. "Self-powered continuous time-temperature monitoring for cold-chain management". In: *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. Aug. 2017, pp. 879–882. DOI: 10.1109/MWSCAS.2017.8053064.

[97] Liang Zhou, Adam C Abraham, Simon Y Tang, and Shantanu Chakrabartty. "A 5 nw quasi-linear CMOS hot-electron injector for self-powered monitoring of biomechanical strain variations". In: *IEEE transactions on biomedical circuits and systems* 10.6 (2016), pp. 1143–1151.

[98] Liang Zhou, Adam C Abraham, Simon Y Tang, and Shantanu Chakrabartty. "Approaching the Limits of Piezoelectricity driven Hot-electron Injection for Self-powered in vivo Monitoring of Micro-strain Variations". In: *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. IEEE. 2016, pp. 1810–1813.

[99] Liang Zhou, Kenji Aono, and Shantanu Chakrabartty. "A CMOS Timer-Injector Integrated Circuit for Self-Powered Sensing of Time-of-Occurrence". In: *IEEE Journal of Solid-State Circuits* 53.5 (2018), pp. 1539–1549.

[100] Liang Zhou and Shantanu Chakrabartty. "A 7-transistor-per-cell, high-density analog storage array with 500μV update accuracy and greater than 60dB linearity". In: *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE. 2014, pp. 1572–1575.

[101] Liang Zhou and Shantanu Chakrabartty. "Linearization of CMOS Hot-Electron Injectors for Self-Powered Monitoring of Biomechanical Strain Variations". In: *IEEE Transactions on Biomedical Circuits and Systems* (2016).

[102] Liang Zhou and Shantanu Chakrabartty. "Linearization of CMOS Hot-Electron Injectors for Self-Powered Monitoring of Biomechanical Strain Variations". In: *IEEE Transactions on Biomedical Circuits and Systems* 11.2 (2017), pp. 446–454.

[103] Liang Zhou and Shantanu Chakrabartty. "Secure dynamic authentication of passive assets and passive iots using self-powered timers". In: *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*. IEEE. 2017, pp. 1–4.

[104] Liang Zhou and Shantanu Chakrabartty. "Self-powered sensing and time-stamping of rare events using CMOS fowler-nordheim tunneling timers". In: *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. IEEE. 2016, pp. 2839–2842.

[105] Liang Zhou and Shantanu Chakrabartty. "Self-Powered Timekeeping and Synchronization Using Fowler–Nordheim Tunneling-Based Floating-Gate Integrators". In: *IEEE Transactions on Electron Devices* 64.3 (2017), pp. 1254–1260.

[106] Liang Zhou, Pikul Sarkar, and Shantanu Chakrabartty. "Scavenging thermal-noise energy for implementing long-term self-powered CMOS timers". In: *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE. 2013, pp. 2203–2206.

[107] Shijie Zhou, Zhen Zhang, Zongwei Luo, and Edward C. Wong. "A lightweight anti-desynchronization RFID authentication protocol". In: *Information Systems Frontiers* 12.5 (2010), pp. 521–528.

[108] Adamu Murtala Zungeru, Li-Minn Ang, S Prabaharan, and Kah Phooi Seng. "Radio frequency energy harvesting and management for wireless sensor networks". In: *Green mobile devices and networks: Energy optimization and scavenging techniques* (2012), pp. 341–368.