Detecting and Mitigating Hardware Security Attacks in Emerging Technologies and Applications

A Dissertation Presented

by

Krithika Dhananjay

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

August 2022

Stony Brook University

The Graduate School

Krithika Dhananjay

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Dr. Emre Salman - Advisor of Dissertation
Associate Professor, Department of Electrical and Computer Engineering

Dr. Petar Djuric - Chairperson of Defense
Chair, Distinguished Professor, Department of Electrical and Computer Engineering Department

Dr. Milutin Stanacevic - Defense Committee Member
Associate Professor, Department of Electrical and Computer Engineering

Dr. Amir Rahmati - Defense Committee Member
Assistant Professor, Department of Computer Science

This dissertation is accepted by the Graduate School

Celia Marshik
Interim Dean of the Graduate School

Abstract of the Dissertation

Detecting and Mitigating Hardware Security Attacks in Emerging Technologies and Applications

by

Krithika Dhananjay

Doctor of Philosophy

in

Department of Electrical and Computer Engineering

Stony Brook University

2022

Starting from the famous Van Eck Phreaking attack (performed in 1985) that picks up electromagnetic emissions for spying on a computing device to the more recent Spectre and Meltdown attacks (performed in 2018) that allow programs to steal data being processed by a computer, hardware vulnerabilities have emerged as significant threat to computing devices. This thesis broadly encompasses two types of such dangerous hardware attacks in emerging technologies: (1) power side-channel attacks in resource-constrained IoT devices and (2) covert-channel attacks in modern 3D integrated multicore processors.

In the first part of the thesis, a novel charge-based attack methodology is developed to perform a power side-channel attack on an ultra-low power encryption core with two orders of magnitude reduction in required power samples. This reduction

decreases the execution time of the attack by 2×. Furthermore, two circuit-level countermeasures are proposed to enhance the side-channel attack resistance of IoT encryption hardware. The proposed technique to secure RF-powered IoT devices increases the attack resistance by 52×, while lowering the energy per encryption by 15.6%. The second technique to protect energy-efficient and secure IoT devices reduces the energy consumption of two-input logic gates by up to 35% while maintaining similar security characteristics.

The second part of this thesis involves the study of thermal covert-channel communication in modern 3D multicore architectures. A 2015 study on an Intel Xeon server platform has demonstrated that 16-digits of credit card information can be illegally extracted within five seconds by exploiting the heat propagation in multicore systems. In this research, we demonstrate that by leveraging 3D vertical integration technologies in processors, it is sufficient to execute low power applications to transfer 200 bits of secret data in one second via such thermal covert-channels. We also show that the bandwidth of this thermal communication in 3D ICs is more resilient to thermal interference caused by applications running in other cores. Furthermore, a technique is proposed to detect such low power thermal covert-channels with a detection accuracy of 100% with no false-positives, for up to 100 Hz of transmission bit rate. To summarize, the proposed methodologies in this thesis enable

resource constrained IoT devices and 3D multicore processors that are more resistant to side-channel and covert-channel attacks.

Table of Contents

ACKNOWLEDGEMENTS

The completion of this thesis has truly been a life-changing experience for me, and it would not have been possible without the encouragement and support from many people across the globe.

I would like to express my deepest gratitude to my advisor Dr. Emre Salman for his incredible support, patience, and guidance throughout my research. I cannot thank him enough for all the brainstorming sessions and his immense insight while writing research papers. He not only gave me the freedom to explore ideas on my own, but also steered me in the right direction whenever my steps faltered. This Ph.D would not have been a memorable journey without his immense support and understanding.

I would also like to sincerely thank our collaborators Dr. Vasilis Pavlidis from the University of Manchester and Dr. Ayse Coskun from Boston University, for providing their insightful ideas and discussions during our weekly meetings. I am also grateful to Dr. Gianluigi De Geronimo and Dr. Milutin Stanacevic for their patient guidance during the beginning of my Ph.D.

My sincere thanks to my defense committee: Dr. Petar Djuric, Dr. Milutin Stanacevic and Dr. Amir Rahmati for their invaluable suggestions, comments, and

questions.

Special thanks to my friends from Stony Brook University: Tutu, Manav, Ivan, Brian, Mallika, Emerson and Wenbin and, from Boston University: Prachi and Zihao for their constant support and co-operation.

Last but certainly not the least, I am deeply indebted to my family for making so many sacrifices during this process. Without their endless support and encouragement, this Ph.D. would not have been accomplished. I dedicate this thesis to them.

Chapter 1

Introduction

In an online survey conducted in 2021, at least 63% of organizations have reported data breaches due to hardware or silicon-level attacks (1). Unlike the more common software attacks, it is typically not possible to fix powerful hardware attacks with subroutines or patches (2; 3). Once a hardware is compromised, an adversary can gain access to sensitive personal information such as location, medical records, and credit card information. Therefore, the field of hardware security has gained significant attention during the past decade. In 2015, 43% of the software assisted hardware vulnerabilities were contributed by information leakage (4). Detection and mitigation of information leakage via side-channel and covert-channel attacks is the focus of this thesis.

Side-channel and covert communication channel attacks represent an open threat to exfiltrate sensitive/secret information from victim devices. Side-channel attacks pose a serious threat particularly to encryption hardware since the secret key can be retrieved by observing the physical characteristics of the hardware such as power, execution time, electromagnetic emissions and temperature for many different in-

put traces and key guesses. Via various statistical analyses on these data, secret keys can be quickly retrieved. Alternatively, covert communication channel attacks exchange secret information between compromised parties by communicating through a covert medium such as time delays or temperature. The first part of this thesis focuses on power-based side-channel attacks and corresponding protection techniques for resource-constrained applications such as RF-powered IoT devices. The second part of the thesis focuses on establishing a high bandwidth thermal covert-channel communication in modern multicore processors by leveraging vertical integration technologies and finally describes enhanced techniques to detect such covert-channel attacks.

The design of lightweight hardware for resource-constrained IoT applications and ensuring its security against such attacks is significantly challenging due to highly limited resources in terms of compute capability, power consumption, and physical area. The traditional low power design techniques such as supply voltage scaling (5), frequency scaling, clock and power gating (6) are not sufficient for low power IoT applications where the power budgets are in the range of several micro Watts. Adiabatic (also referred to as charge-recycling) circuit technology offers more than an order of magnitude reduction in power consumption compared to the traditional static CMOS technology (7). Adiabatic circuits were proposed in early 90s as part of the significant efforts on developing reversible computing (8). Contrary to conventional static CMOS, adiabatic circuits rely on variable/AC power supply signal in the form of a trapezoidal or sinusoidal waveform, typically referred to as as power-clock signal. Despite offering significantly high energy efficiencies, the practical application of these circuits has remained limited due to primary challenges such as performance limitations, generation of power-clock signal, and lack

of design automation. More recently, the application of adiabatic circuits to wirelessly powered lightweight devices (such as RFIDs and wireless sensor nodes) was proposed, referred to as AC computing methodology (7; 9). When adiabatic circuits are used for RF-powered applications via AC computing, the harvested signal is already in the form of AC, thus eliminating the need for power-clock generation. Furthermore, these applications typically do not need GHz range frequencies and consist of circuits with lower complexity in terms of transistor count. Thus, the primary limitations related to adiabatic circuits are mitigated in RF-powered applications. The application of adiabatic circuits to RF-powered devices introduces new challenges, as discussed in (10). Multiple solutions to these challenges were recently proposed as part of the AC computing methodology, demonstrating significant improvements in energy efficiency (11). For example, in (12), it was demonstrated that the encryption efficiency (evaluated in bits per second per Watt) of an RF-powered adiabatic encryption core is 27.5$\times$ higher than conventional static CMOS based implementation, making it highly suitable for lightweight IoT applications. In addition to ultra low power consumption, adiabatic circuits and AC computing methodology exhibit unique hardware security characteristics, as investigated in this thesis.

The primary contributions of this research related to side-channel attacks on adiabatic/AC computing based encryption core are listed below:

- A novel charge-based methodology for mounting a power-based side-channel attack on an adiabatic lightweight encryption cipher (based on SIMON algorithm) is proposed. The proposed methodology significantly reduces the attack complexity by reducing the required number of power samples by two orders of magnitude.

- The inherent resistance provided by charge-recycling adiabatic operation to power-based side-channel attacks is evaluated on SIMON encryption cipher. It is shown that an adiabatic SIMON cipher exhibits $4\times$ more side-channel resistance and $10\times$ higher encryption efficiency (kilobits/sec/W) when compared to an unprotected static CMOS based SIMON. The effect of target signal capacitance on the the CPA attack resistance is studied for both the adiabatic and static SIMON implementations.

- A novel, protected adiabatic logic gate, referred to as SEcure Adiabatic Logic for Wirelessly-Powered IoT Devices (SEAL-RF), is proposed. SEAL-RF exhibits enhanced resistance to power-based side-channel attacks.

- A protected adiabatic SIMON core is developed with the proposed secure logic gate. The power side-channel attack resistance is increased by $52\times$, while lowering the energy per encryption by 15.6% as compared to an unprotected adiabatic SIMON implementation.

Recently, temperature-based covert-channel communication has gained attention, where an adversary uses heat to communicate sensitive data between two unauthorized compute elements (13; 14). For example, in a multicore processor, the thermal covert-channel communication (henceforth referred to as TCC) is established between two cores of the processor by encoding sensitive information within the temperature profile of the transmitting core (13). Specifically, an attacker application transmits a bit '1' by executing a program in the transmitting core to raise its temperature. In order to transmit a bit '0', the attacker stops program execution to lower the temperature of the transmitting core. Due to thermal coupling among the cores, an application in a receiving core can retrieve the information by reading

its temperature sensor that is accessible to user applications (15). Several studies have been published during the past decade about TCC modeling (14; 16), detection (17; 18), and countermeasures (17; 18; 19) in multicore processors. Long *et al.* (20) and Huang *et al.* (17) demonstrate that a successful TCC with low error rates can be established provided that the transmission frequency of the channel, i.e., TCC rate, is higher than the frequency band of power consumption of the other active cores. In conventional 2D ICs, high TCC rates can be achieved by executing high power programs (such as CPU stress tests). Thus, a majority of the existing works rely on high power programs to sufficiently raise the temperature of the transmitting core, thereby reducing the error rates of covert communication. These programs, however, are likely to cause overheating, thus enabling the attack to be detected relatively easily. In this work, we demonstrate that a high-bandwidth TCC can be established with relatively low power benchmark applications by leveraging vertical integration technologies, such as through-silicon via (TSV) based die stacking (21) and monolithic 3D (Mono3D) integration (22). Unlike TCC in conventional 2D integration, where heat flows between the cores of a processor in lateral fashion, the close proximity of tiers in 3D ICs increases the vertical thermal coupling among the inter-tier functional blocks. Thus, TCC attacks are potentially more dangerous in 3D integrated multicore systems because larger blocks of sensitive data can be communicated at faster rates. Thus, the following items represent the primary contributions of this research related to thermal covert communication in modern multicore processors:

- The TCC established in 3D processors is shown to achieve negligible error rates ($< 1\%$) with transmission rates of upto 250 bps by executing commonly used SPLASH-2 benchmark applications. Therefore, the average power con-

sumed during the attack is significantly reduced, making the attack more difficult to detect.

- It is demonstrated that the TCC bandwidth in Mono3D and TSV3D processors is, respectively, 3.4× and 3.1× greater than the TCC bandwidth in a conventional 2D processor.

- The effect of thermal interference from applications running on other cores is also investigated. We observe that the TCC bandwidth in Mono3D and TSV3D of, respectively, 200 *bps* and 182 *bps*, remains unaffected in the presence of interference from one of the cores. Alternatively, for a 2D integrated processor, the bandwidth degrades by 12% with a minimum achievable error rate of 3%. Therefore, the TCC in 2D processors is shown to be highly sensitive to the heat generated by other active cores whereas TCC in 3D processors is comparatively more robust.

- A novel detection metric is proposed to detect a high bandwidth TCC established by executing low power programs. Five low power applications from SPLASH-2/PARSEC benchmark suites are considered. It is shown that all of the covert-channels are detected with 0% false positive rate.

The rest of the thesis is organized as follows. Relevant background for both power-based side-channel attacks for AC computing and thermal covert-channel communication is provided in Chapter 2. A novel power-based side-channel attack methodology for an adiabatic cipher is presented in Chapter 3. The proposed protected adiabatic logic for AC computing is described in Chapter 4. A comparative study to evaluate the security and energy of logic gates designed with the proposed logic family is also provided to evaluate the proposed approach. The power-based

side-channel attack results on protected adiabatic SIMON core (using the proposed method) are presented in Chapter 5. Establishing high-bandwidth TCC by leveraging 3D integration technologies is described in Chapter 6. A novel detection metric to detect such low power TCCs is presented in Chapter 7. Finally, the thesis is concluded in Chapter 8.

Chapter 2

Background

Side-channel and covert-channel attacks have remained an open threat to the flow of secure information in modern computing devices. Designing secure hardware with increased resistance to these attacks is the primary goal of this thesis. Therefore, a brief background about side-channel and covert-channel attacks is provided in Section 2.1 of this chapter.

In the first part of this research, side-channel resistant adiabatic circuits for AC computing applications are developed. Designing lightweight encryption circuits that are also resistant to side-channel attacks, however, is highly challenging due to scarcity of available power. The application of adiabatic/AC computing based circuits to wirelessly powered RF devices offers significant benefits, not only for higher energy efficiency, but also for lightweight (yet effective) security. Therefore, Section 2.2 in this chapter provides background information on some of the primary topics covered in this part of the thesis. Specifically, a detailed background on adiabatic circuits and AC computing methodology is provided in Section 2.2.1. A lightweight encryption cipher, SIMON, is introduced in Section 2.2.2. This ci-

pher is used to demonstrate the proposed methodologies in this thesis. Finally, an overview of power-based side-channel attacks is provided in Section 2.2.3.

For the second part of this thesis, a high bandwidth and low power thermal covert communication is established by leveraging the strong vertical thermal coupling between the tiers of a 3D integrated processor. Since the existing detection techniques fail to capture this type of vulnerability, an enhanced technique is proposed in this research. The required background for this study on thermal covert-channel attacks in multicore processors is presented in Section 2.3. A brief background about the vertical integration technologies discussed in this thesis is presented in Section 2.3.1. Finally, a detailed background on thermal covert channel attacks in modern multicore processors is provided in Section 2.3.2.

2.1 Side-channel and Covert-channel Attacks

Hardware back-doors exist at various stages during the design cycle of a chip. A computer hardware that was once assumed as an untouched brick wall, can now be cracked by much simpler means because of the affordable threat models and the vulnerabilities at several steps of the semiconductor supply chain.

While cryptography is the study of securing or encrypting sensitive information, cryptanalysis is the art of breaking an encryption hardware. Cryptanalysis can be broadly classified as software attacks that make use of the weakness in the algorithm and hardware attacks that exploit the actual hardware implementation of an encryption cipher. Fig. 2.1 shows a broad classification of cryptanalysis and specifically hardware attacks, as explained below:

- **Invasive attacks:** Invasive attacks are the strongest type of attacks and typi-

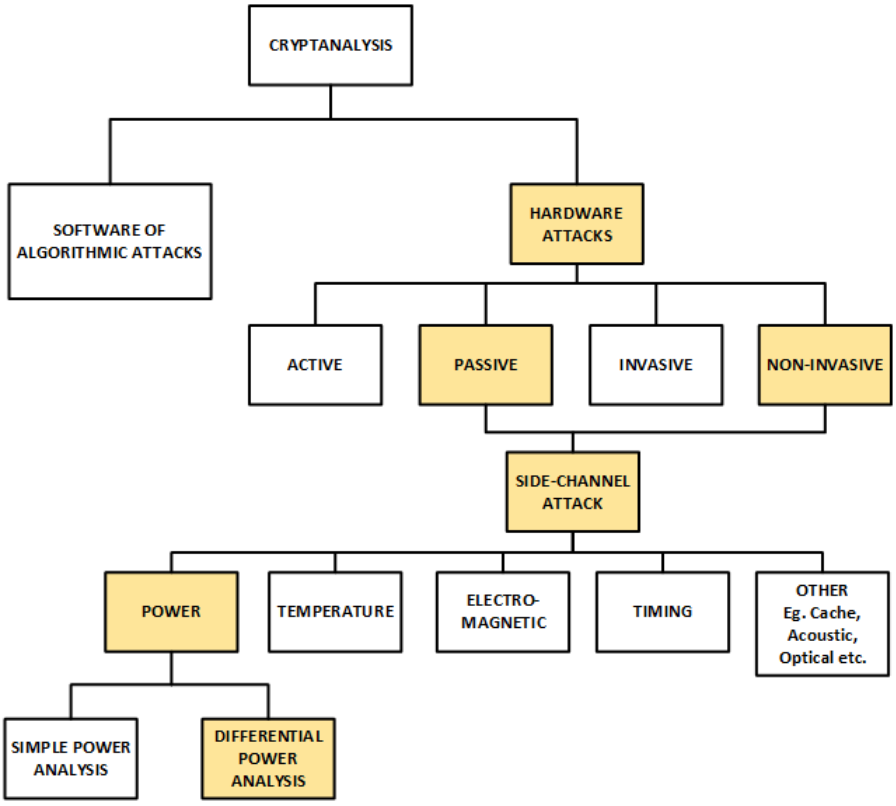


Figure 2.1: Classification of cryptanalysis - the art of deciphering the key.

cally involve de-packaging the entire device (23; 24). This type of attack can be used to extract any type of information (not specifically the secret key) depending on the level of probing. However, these attacks require relatively expensive probing equipment.

- **Non-invasive attacks**: A non-invasive attack exploits the externally available information without altering the device. Only externally accessible ports are used for this attack and hence is less expensive.

- **Active attacks:** Active attacks tamper with the proper functioning of a device by varying its inputs or the environment on-the-fly in order to extract the sensitive information.

- **Passive attacks:** Passive attacks only observe the behaviour of the device without affecting the functionality.

Side-channel attacks are passive and non-invasive hardware attacks that extract the sensitive information by observing the physical properties of a system such as power consumption, electromagnetic emissions, execution time, temperature and sound. Since these attacks can typically be performed with available equipment, they pose as a common and dangerous form of hardware attacks.

In 1973, Lampson defined the term 'covert-channel' as a channel that is not intended for information transfer (25). Unlike a side-channel attack where data is ex-filtrated from an unsuspecting victim, in a covert-channel communication, both the transmitting compute element and receiving compute element are compromised. A simple block diagram of covert-channel communication is shown in Fig. 2.2. Some examples of covert-channels (or medium) are temperature (13), network delays (26), shared processor cache states (27), I/O devices (28) and sound produced

by mobile devices (29), as highlighted in the figure. Typically, the attacker embeds a program in the transmitter, that encodes the secret data on these shared channels. Similarly, this shared resource is read by the receiver to decode the confidential information. Because of the secretive nature of these attacks, detecting their existence in real-time poses a significant challenge to the computing hardware.

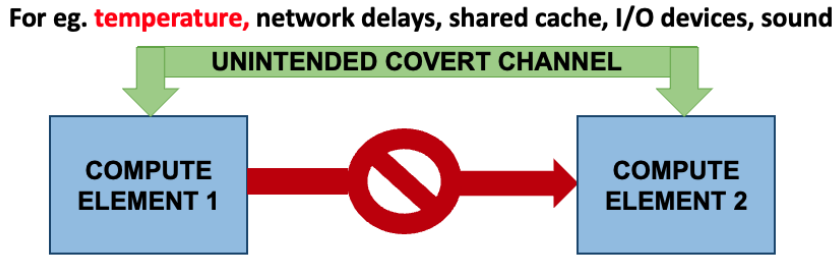


Figure 2.2: Covert-channel communication.

In this research, a side-channel attack resistant ultra-low power encryption core is designed for RF-powered IoT applications. A detailed background for this research is provided in Section 2.2. Furthermore, a high-bandwidth thermal covert-channel attack by leveraging 3D integration technologies is studied. The required background for this research is detailed in Section 2.3.

2.2 Hardware Security for RF-powered Applications

2.2.1 Ultra-low power adiabatic computing

There has been a proliferation of low power design strategies implemented over the past three decades in order to satisfy the increasing compute demand of applications and the growing need for on-site processing. The concept of adiabatic and

reversible computing was proposed as an alternative strategy in order to reduce the energy consumption by an order of magnitude or more (7).

The introduction of adiabatic circuits dates back to 1960s when physicist Landauer discussed the concepts of irreversibility and heat generation for computing systems (30). Landauer demonstrated that energy is dissipated in each computation only when the information is erased (30). For each bit erased, the theoretical lower bound of energy dissipation was derived to be $kT\ln(2)$. Since then, multiple works (31; 32; 33) have proposed to conserve maximum amount of energy by maintaining logical reversibility. Logical reversibility is achieved when the outputs can be retraced back to the inputs in order to avoid information erasure. However, the methods proposed thus far to achieve this objective suffer from excessive area usage and/or loss of performance (7). In parallel to developments in logically reversible computing, it was concluded that in order to achieve dramatic energy savings, it is not necessary for the gate to be logically reversible. Instead, it can be sufficient for the gate to be *energetically reversible*, where the charge stored on the load capacitance is recovered or restored back to the power supply in order to conserve energy. The transfer of charge between power supply and capacitor is achieved with an AC signal rather than the conventional DC supply voltage. This technique is referred to as adiabatic computing.

While achieving a truly adiabatic operation may be challenging and impractical (due to extremely slow movement of current and unavoidable static losses in conventional CMOS processes), various logic families have been proposed to lower power consumption by leveraging adiabatic operation (34; 35; 36; 37; 38). Even though these logic families exhibit significant differences in terms of how close they get to fully adiabatic operation, the primary characteristic is the presence of a

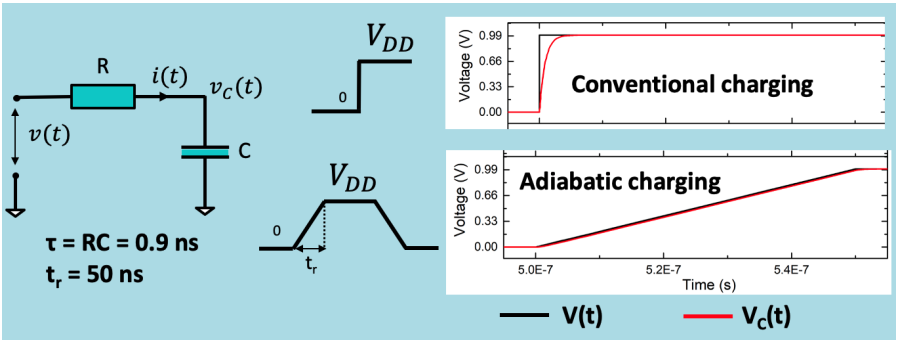


Figure 2.3: Illustration of adiabatic charging with a trapezoidal power supply signal as compared to traditional charging with a constant DC voltage.

variable/AC power supply signal in the form of a trapezoidal or sinusoidal waveform. This signal also behaves as a clock signal for the adiabatic circuit since it synchronizes the flow of data and typically referred to as power-clock signal.

A trapezoidal or sinusoidal power supply signal maintains a small voltage difference between the power supply and output nodes during charging (7). As such, adiabatic operation reduces the power consumption by minimizing the current to charge the output node. Furthermore, as the power supply signal falls, the charge stored at the output node is recycled back to the power supply.

Consider the equivalent circuit of an adiabatic operation shown in Fig. 2.3. R represents the on-resistance of the transistor and the interconnect resistance of the output wire and C represents the output load capacitance. The power supply signal is a trapezoidal waveform with a transition time of t_r. If t_r is sufficiently long as compared to the RC time constant, then $v_c(t)$ approximately follows $v_{dd}(t)$, thereby minimizing the power loss across R. The constant charging current i is approximated as

$$i = C\frac{dv_c(t)}{dt} \approx \frac{CV_{DD}}{t_r}. \tag{2.1}$$

Each time the load capacitor is adiabatically charged, the energy that is dissipated across the resistor R is given by

$$E_{ad} = \int_0^{t_r} v_R(t) i(t) dt = i^2 R t_r = \frac{RC}{t_r} C V_{dd}^2. \tag{2.2}$$

Since one cycle consists of adiabatic charging (when the power supply signal is rising) and recovery of the charge back to power supply (when the power supply signal is falling), the overall energy dissipated per cycle is

$$E_{ad} = 2\frac{RC}{t_r} C V_{dd}^2. \tag{2.3}$$

Unlike conventional static CMOS based operation where energy does not depend upon transition time, in adiabatic operation, a larger transition time reduces the overall energy, as described by (2.3). Critical transition time t_r^{crit} at which the energy consumed by static CMOS operation ($E_{st} = \frac{1}{2}\alpha C V_{dd}^2$) is equal to the energy consumed by adiabatic operation can be determined by comparing E_{st} with (2.3) and is given by,

$$t_r^{crit} = 4\frac{RC}{\alpha}, \tag{2.4}$$

where α is the switching activity factor. Thus, if t_r is greater than t_r^{crit}, adiabatic circuits consume less energy than conventional circuits. As such, applications that operate at relatively low frequencies and with moderate to high activity factors are good candidates for adiabatic operation. Note however that the absolute value of the t_r^{crit} is highly technology dependent due to R and C. In advanced nanoscale technologies, adiabatic operation can save considerable power even at frequencies in the range of several hundred megahertz (10).

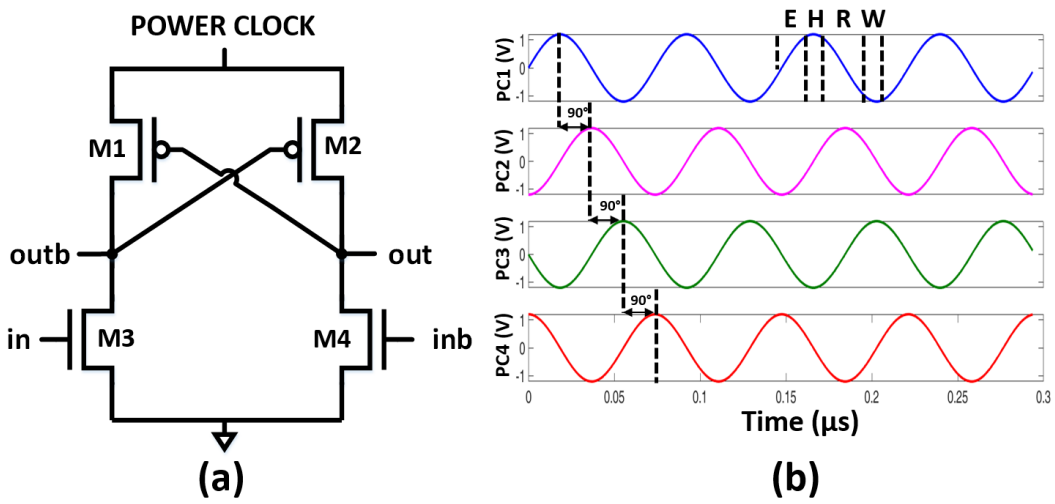


Figure 2.4: Operation of an ECRL buffer: (a) transistor-level schematic, (b) four-phase sinusoidal power-clock inputs, each with a 90° phase shift.

Since adiabatic logic gates are based on conventional CMOS based MOSFET transistors, the energy can be recovered only until the output reaches the threshold voltage of the pMOS devices, after which the pMOS transistor stops conduction. Due to this reason, MOSFET based adiabatic logic that relies on pMOS devices for energy recovery is typically referred to as quasi-adiabatic. Several quasi-adiabatic logic topologies have been proposed over the years. Efficient charge recovery logic (ECRL) is one such topology that is applicable to modern low voltage technologies due relatively robust operation (34). The transistor-level implementation of an ECRL buffer is shown in Fig. 2.4(a) and four sinusoidal power-clock inputs are illustrated in Fig. 2.4(b).

ECRL utilizes four power supply signals, each with a 90° phase shift. Specifically, there is a 90° phase difference in the power supply signal of adjacent logic gates. There are four stages of operation, depending upon the power supply signal:

- Evaluate (E): In this stage, the power supply signal rises and the inputs *in* and *inb* are stable. If $in = 1$, $outb = 0$, $M2$ turns on once power supply reaches the threshold voltage. Thus, *out* follows power supply signal.

- Hold (H): Power signal and the outputs remain stable for the subsequent gate to evaluate.

- Recover (R): Both inputs are discharged by the previous gate. The power supply falls and *out* follows power supply signal until it reaches the threshold voltage of $M2$. The charge is partially recovered back to the power supply during this stage.

- Wait (W): The gate waits for the next evaluation stage.

The multi-phase operation in an ECRL gate enables the outputs to be evaluated only during the $evaluate$ stage when the inputs remain stable (since the preceding gate is at $hold$ stage). Thus, adiabatic logic is inherently pipelined where each gate acts as a sequential circuit and consumes a quarter of a cycle.

2.2.1.1 Primary limitations of conventional adiabatic circuits

Although adiabatic circuits enable ultra-low power operation, the industrial adoption of these circuits is limited, since the commercial electronic devices rely on DC voltage. Some of the primary limitations of adiabatic circuits that have prevented them from becoming a popular design strategy are as follows:

- Lack of design automation since existing tools do not fully support adiabatic logic

- Performance limitation since the transition time for the power-clock signal should be sufficiently long, as shown by (2.4)
- Efficient generation of the multi-phase power-clock signals, which typically requires high quality passive devices, and reliable distribution of these global signals throughout the chip.

These challenges are partially mitigated when adiabatic circuits are used for RF-powered applications, as discussed in the following subsection.

2.2.1.2 Application of adiabatic circuits to RF-Powered devices: AC computing

Adiabatic circuits exhibit a highly encouraging opportunity for IoT devices that harvest RF power. Some examples to these applications include RFID-based systems and wireless sensor nodes that traditionally have highly limited computing capabilities. An existing digital logic within these RF-powered devices can be adiabatically driven since the wirelessly harvested signal is already in the form of a sinusoidal waveform as shown in Fig. 2.5. This approach of harvesting the ambient or dedicated RF energy to power an adiabatic core is referred to as AC computing methodology (9; 10; 11; 39; 40; 41; 42).

AC computing has several significant benefits in enhancing energy efficiency of the RF-powered logic: (1) the challenges related to the generation of the power-clock signal are partially mitigated, (2) significant power loss related to rectification process in conventional methods is eliminated, (3) digital logic runs more efficiently due to adiabatic operation. An important consideration for this approach is that the carrier frequency becomes the power-clock frequency for the logic. For example,

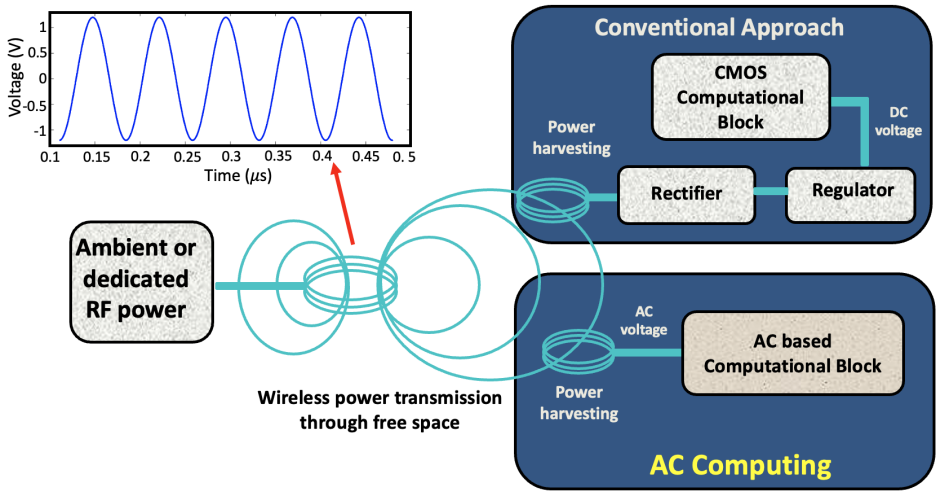


Figure 2.5: AC computing methodology for RF-powered devices (11).

for an RF-powered application in the HF RFID band, the wirelessly powered adiabatic logic needs to run at 13.56 MHz. Thus, the energy-performance requirements of the application should match with the carrier frequency of power harvesting, which also affects the antenna size.

In (11), a near-field inductive coupling based wireless link was developed to adiabatically power an 8-bit arithmetic logic unit (ALU) designed in various adiabatic logic families using 65 nm technology. Simulation results for pass transistor adiabatic logic (PAL) that requires two out-of-phase power-clock signals demonstrate up to 30× reduction in power consumption as compared to a static CMOS based ALU powered via a DC voltage obtained after rectification (11). However, an important disadvantage of PAL is that the output nodes remain floating for a short period of time during operation, which degrades robustness. Alternatively, ECRL exhibits higher robustness and permits low voltages (AC signal amplitude) as shown in (11). However, since ECRL requires a phase shifter with passive *LC*

components (39), the overall size increases, particularly at low frequencies. Thus, existing adiabatic logic families exhibit interesting tradeoffs for RF-powered applications (43).

In this thesis, the primary emphasis is on the hardware security aspects of adiabatic circuits with application to AC computing. An important security aspect for these applications is lightweight encryption hardware and side-channel resistance, as discussed in the following section.

2.2.2 SIMON: A flexible lightweight cipher

Ensuring the security and data privacy for lightweight applications (such as RFID based systems, wireless sensor nodes and energy harvesting IoT devices) is significantly challenging due to highly limited resources in terms of compute capability, power consumption, and physical area. The robust general-purpose encryption algorithms such as the AES are not suitable candidates because of the high hardware cost. For example, the area specification for typical lightweight applications cannot exceed 2,000 gate equivalents (GE) (44). However, the smallest hardware implementation of AES encryption utilizes 2,400 GE and there are limits to how far these algorithms can be optimized for resource-constrained platforms (44).

Back in the 1990s, several non-standard cryptography algorithms such as A5/1 and A5/2 in cell phones and KeeLoq in car locks were used due to the lack of more sophisticated lightweight primitives (45). These algorithms were prone to hardware attacks and therefore could be compromised easily to leak secret data. Thus, the past decade has witnessed a major increase in the number of new lightweight ciphers and their standardization by organizations such as NIST and ISO. Several lightweight block ciphers have been proposed recently including PRESENT-

80 (46), PRINCE (47), CLEFIA (48), CAMELLIA (49), SIMON and SPECK (50).

SIMON and SPECK are two sister algorithms developed by the National Security Agency and internationally standardized by ISO/29167-21 (51) as part of RFID air interface standard for use by commercial entities. SIMON was optimized specifically for hardware performance and SPECK for software implementations. The flexibility and simplicity of the SIMON algorithm makes it suitable for diverse lightweight applications based on the power, performance, area, and security requirements. Specifically, the hardware implementation of the smallest configuration of SIMON (with 32-bit plaintext and 64-bit key) achieves an area utilization of only 523 GE, thus enabling encryption for ultra-low area and low power applications, where it is highly challenging to afford integrated encryption circuitry (50).

The SIMON algorithm caters to a wide range of block and key sizes that can be chosen depending upon the application and required level of security. A SIMON block cipher with n-bit word plaintext ($2n$-bit block) and m-word key (mn-bit block) is typically referred to as SIMON $2n/mn$ (50). The configuration adopted for this work is 32-bits of plaintext and 64-bits of key (SIMON 32/64), and 32 rounds of encryption. A typical SIMON algorithm is comprised of a round function and key expansion, as summarized below.

2.2.2.1 Round function

The SIMON round function uses a two step Fiestel mapping, as shown in Fig. 2.6 and is given by

$$R(L_{i+1}, R_{i+1}) = (R_i \oplus f(L_i) \oplus K_i, L_i), \tag{2.5}$$

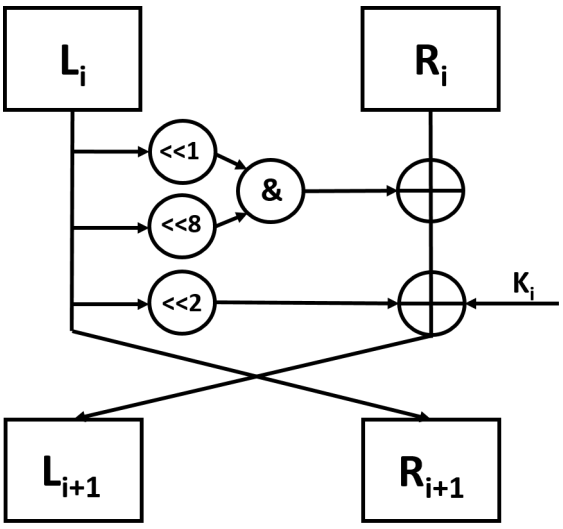


Figure 2.6: SIMON round function algorithm.

where i is the current round and $i+1$ is the next round, R is the right word and L is the left word of a block, and K is the key generated by the key expansion module. Function $f(L_i)$ is given by

$$f(L_i) = ((L_i << 1) \,\&\, (L_i << 8)) \oplus (L_i << 2), \tag{2.6}$$

where $a << b$ refers to a left-shifted by b bits. This round function is iterated until the desired number of rounds is reached.

2.2.2.2 Key expansion

The strength of the input key determines the level of security for any cryptography hardware. In the SIMON key expansion module, an input key is used to generate a unique key for each round of encryption. Unlike the round function, the key expansion functions vary depending upon the width of the key word m, which

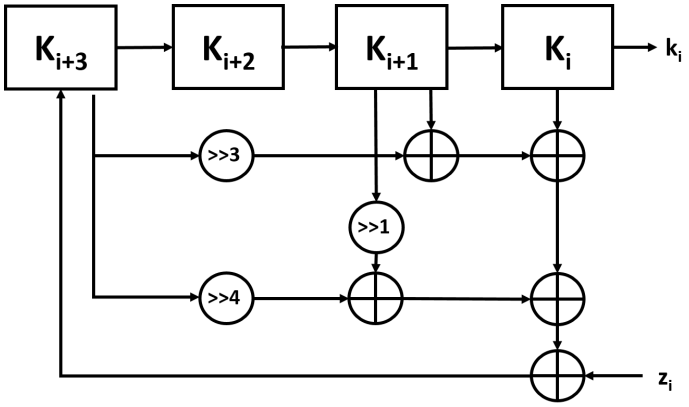


Figure 2.7: SIMON key expansion algorithm for m=4.

can be 2, 3 or 4. Since the configuration used in this paper is SIMON $32/64$, the key expansion algorithm for $m = 4$ is chosen, as illustrated in Fig. 2.7. The first four rounds use the four words of 64-bit key input and the key used from the fifth round, K_{i+4}, is generated by using the following function,

$$K_{i+4} = (K_i \oplus K_{i+1} \oplus (K_{i+3} >> 3)) \oplus$$
$$(K_{i+1} >> 1) \oplus (K_{i+3} >> 4) \oplus z_i, \tag{2.7}$$

where $1 <= i <= 28$ and z_i is referred to as the round constant that is used to eliminate slide properties and circular shift symmetries (44).

A key feature of SIMON algorithm is that there is a scope for serialization at every level, unlike *s*-box based algorithms. Depending upon the area constraint and throughput requirement of an application, SIMON algorithm can have a bit-level, round-level or encryption-level parallelism. Since the primary objective of this work is to design and analyze the side-channel resistance of SIMON hardware with minimal area and power constraints, the lowest level of parallelism *i.e.* the bit-serial implementation is adopted.

2.2.3 Correlation power analysis side-channel attack

The power-based side-channel attack is the focus of this thesis, which can be further classified into Simple Power Analysis (SPA) side-channel attack and Differential Power Analysis (SPA) based side-channel attack. SPA is the method of inspection of the transient power consumption measurements (referred to as traces) to gain insight into device operation. However, SPA is practical only when the data dependencies on the power consumption is apparent. Alternatively, DPA involves more sophisticated statistical techniques to analyze the power/current consumption in order to identify the data dependent correlations and recover the sensitive information (52). It is the most common type of power side-channel attack, primarily because it does not require a detailed knowledge about the device (24). Correlation power analysis (CPA) attack is a type of DPA that exploits the statistical theory of Pearson correlation between a chosen hypothetical power model and the actual current consumption for various random plaintexts to reveal the secret key and is the primary focus of this work.

The steps involved in a CPA attack are outlined by the flowchart shown in Fig. 2.8. These steps are qualitatively explained below:

- **Selection of target hardware:** The target encryption hardware is chosen by the adversary to mount the CPA attack. In this work, the SIMON32/64 implemented as an Application Specific Integrated Circuit (ASIC) is chosen as the target hardware.

- **Selection of intermediate signal:** CPA attacks are based on a divide and conquer strategy where portions of the key are retrieved/attacked separately, which significantly reduces the complexity as compared to brute force at-

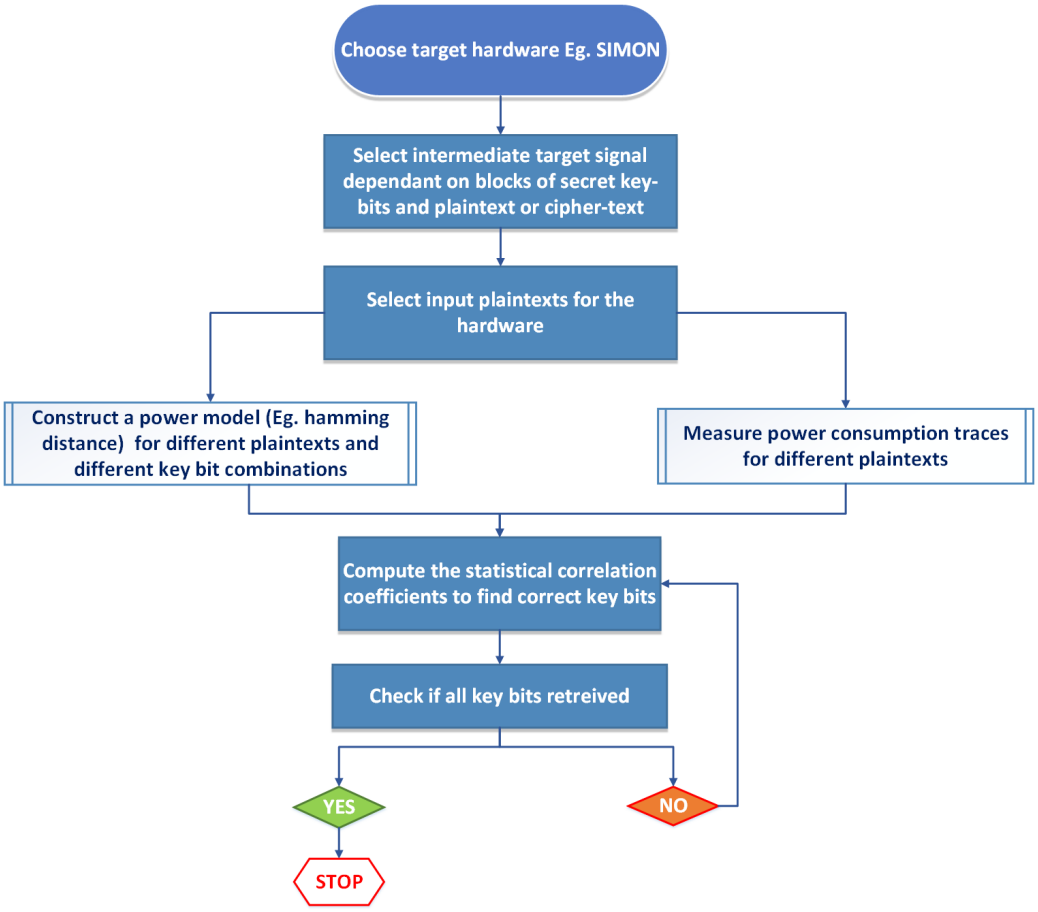


Figure 2.8: Steps involved in mounting a CPA side-channel attack.

tacks. To achieve this, a target intermediate signal is chosen in the circuit, such that it is dependent on a portion of the key bits and the plaintext or the ciphertext.

- **Selection of input plaintexts:** In this step, the type of the inputs to be fed to the hardware is selected in order to mount the attack. For example, random inputs are fed to the SIMON32/64 hardware in this work.

- **Construction of hypothetical power/current model:** The hypothetical power model is constructed to map the intermediate signal values to the power/current consumption. For example, let the intermediate signal be s. If s is dependent on two key bits and the Hamming distance of s has a direct correlation to the power consumption, then a hamming distance power model matrix is constructed for all the random plaintexts and the four possible combinations of the key bits (also referred to as key hypothesis).

- **Measurement of the power/current consumption:** In this step, all the chosen random inputs are fed to the actual hardware in order to measure the current/power consumption traces.

- **Statistical correlation to recover the secret key**: The statistical Pearson's correlation coefficients are measured to analyze the correlation between the above hypothetical power model and the measured power traces. Those key bits that render the maximum correlation coefficient correspond to the correct key bits.

The above steps are repeated until all the key bits of the encryption hardware are recovered. The mathematical formulation of a CPA attack is described below.

Let $h(n,k)$ be the hypothetical power model matrix with $n = 1,2..N$, where N is the overall number of random plaintexts and $k = 1,2,..K$, where K is the overall number of key hypotheses for a portion of the input key. Let $i(n,t)$ be the measured current traces with $t = 1,2,..T$, where T is the length of the trace. The correlation coefficient $r(k,t)$ is given as,

$$r(k,t) = \frac{\sum_{n=1}^{N}(h_{n,k} - \overline{h_k}).(i_{n,t} - \overline{i_t})}{\sum_{n=1}^{N}(h_{n,k} - \overline{h_k})^2.(i_{n,t} - \overline{i_t})^2}, \tag{2.8}$$

where $\overline{h_k}$ and $\overline{i_t}$ refer to the average of columns in, respectively, $h_{n,k}$ and $i_{n,t}$. The correct key hypotheses is the row value k, for which the correlation coefficient $r(k,t)$ is maximum. This algorithm is repeated for several key hypotheses until all of the key bits are recovered.

The topics presented so far establish the foundation for the first part of the thesis (Chapters 3, 4, 5). The following sections of this chapter will provide a detailed background about thermal-covert channel communication studied in the second part of this thesis (Chapters 6, 7 and 8).

2.3 Hardware Security for 3D Integrated Multicore Processors

2.3.1 An introduction to 3D integration technologies

As two-dimensional geometry scaling of conventional transistors is coming to an end as predicted by the International Roadmap for Devices and Systems

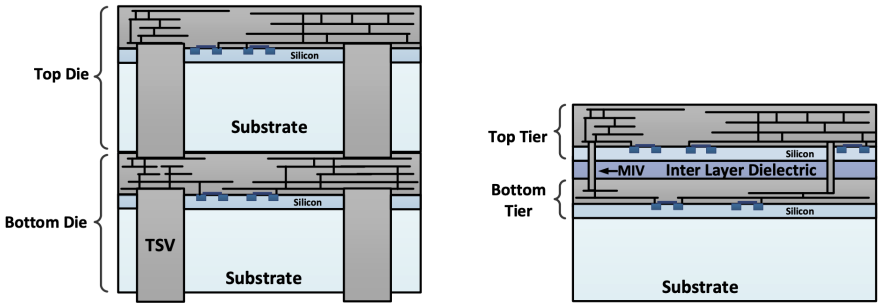


Figure 2.9: Two vertical integration technologies: (a) TSV3D integration and (b) Monolithic 3D integration (58)

(IRDS) (53), the semiconductor industry has witnessed significant improvements in various forms of vertical integration technologies in the past two decades. Through silicon via (TSV) based 3D integration (also referred to as chip stacking) and sequential monolithic 3D (Mono3D) integration are two of the major 3D integration technologies. In TSV based 3D integration, tiers originate from separately fabricated wafers and are interconnected using TSVs that are several micrometers in thickness as shown in Fig. 2.9(a) (54). Some of the commercial applications that utilize these TSV-based integration are memory arrays, such as the Hybrid Memory Cube (HMC) (55) and High Bandwidth Memory (HBM) (56), which are multi-layer DRAM chips. Most recently, commercial integration of multiple logic chiplets has also been demonstrated in a face-to-face configuration with TSVs (57).

Although TSV-based 3D technologies enable significant benefits in system-level performance, power consumption, and form factor, compared to typical 2D integration, these technologies suffer from a noticeable asymmetry between the transistor dimensions and the dimensions of the TSVs (59). The channel length of modern transistors has reached sub-10 nm dimensions, whereas the diameter of modern

TSVs is in the range of several micrometers. This large gap is a significant limitation on the density/granularity of TSV-based die stacking (60; 61). Mono3D technology mitigates these problems by reducing the dimension of vertical interconnects, referred to as monolithic inter-tier vias (MIVs), down to nanometers, thereby enabling unprecedented levels of integration density and granularity (62), as shown in Fig. 2.9(b). Unlike other vertical integration technologies, manufacturing multiple transistor layers on a single substrate in MONO3D technology exhibits unique opportunities for providing extremely dense ICs. Very recently, AMD announced the world's fastest gaming multicore processor using TSV3D technology and there has been recent research works that leverage the unprecedented benefits offered by Mono3D for multicore processors (63; 64; 65; 66; 67).

Hardware security challenges and opportunities related to 3D technologies have also received attention (68; 69; 70; 71; 72; 73; 74; 75; 76; 77). Conventional multicore processors are vulnerable to temperature-based covert-channel attacks that secretly ex-filtrate the secure information handled by them. The extent of the danger imposed by these attacks on the emerging 3D integrated multicore processors, as discussed in this section, is a primary focus of this thesis. The following sections provide background about different types of existing covert-channel attacks, followed by a detailed review of the works on thermal covert-channel attacks in multicore processors.

2.3.2 Thermal covert-channel attacks in multicore processors

Covert-channel communication using heat as a carrier has been identified as a significant threat for several hardware platforms such as cloud-based FPGAs (78; 79), IoT devices (80) and desktop (14; 17; 18), mobile (14; 81) and server proces-

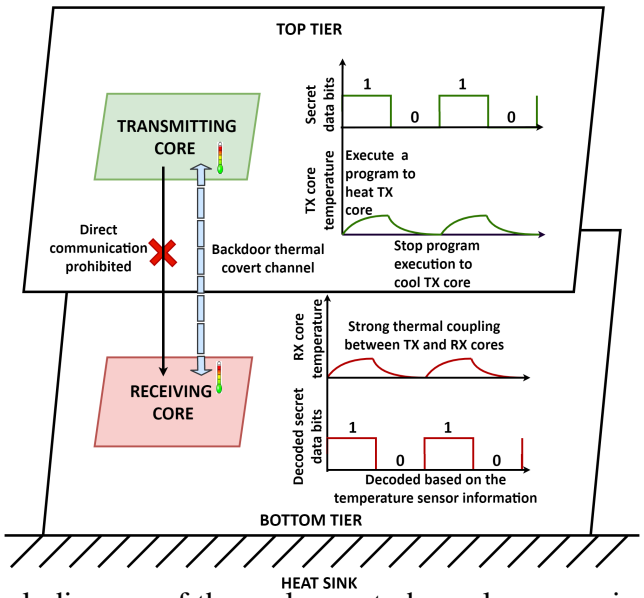


Figure 2.10: Block diagram of thermal covert-channel communication in 3D multicore processors.

sors (13). Masti *et al.* were the first to show that the accessibility of temperature sensors in modern processors enables thermal covert communication between its cores (13). Specifically, to send a bit '1' in a TCC, a program is executed in the transmitting core to raise its temperature and to transmit a bit '0' the execution of the program is stopped, thereby reducing its temperature as shown in Fig. 2.10 . A neighboring receiver core will have similar changes in its temperature profile due to the thermal coupling between the cores and therefore can decode the information by reading its temperature sensors.

Masti *et al.* demonstrated a TCC with maximum bandwidth of 1.33 *bps* with 11% bit-error rate (BER) via measurement results for an Intel Xeon processor (13). Bartolini *et al.* proposed an enhanced communication scheme that uses Manchester encoding and bit-wise decoding with Naive-Bayes classifier. They demonstrated via measurements that a TCC bandwidth of 5 *bps* with less than 1% BER can be

achieved between neighboring cores. They also proposed spectral techniques to characterize the maximum capacity of thermal covert channels for mobile and laptop platforms (14). Both of these works assume that cores other than the transmitter and receiver are idle to minimize thermal interference. Long *et al.* considered the thermal interference from other cores and showed that the BER can be reduced by 75% and the transmission rate can be increased by 370% via two techniques: (1) by selecting a higher TCC transmission frequency than the frequency of the power consumption caused by applications running in other cores and (2) by adopting a return-to-zero encoding scheme (20).

Several works focused on detection and mitigation of TCC. Huang *et al.* proposed techniques for thermal covert channel detection based on scanning the frequency spectrum of temperature profiles (18) and instructions per cycle (IPC) (17) of each processor core. Furthermore, Huang *et al.* also proposed countermeasures based on dynamic voltage and frequency scaling (DVFS) to mitigate an active TCC attack (18). Wang *et al.* proposed a channel-aware noise jamming technique to mitigate a TCC that dynamically changes its transmission frequency (82). Furthermore, Wang *et al.* developed analytic models to efficiently determine the critical TCC parameters (16). A majority of these works leverage the lateral thermal coupling in 2D technologies to establish TCC.

Chen *et al.* exploited the close proximity of SoC to the DRAM chip (fabricated using package-on-package technology) to transmit secret information using heat (83). This temperature-based communication was achieved by generating heat patterns in one core of the SoC and indirectly decoding them at another core by measuring the decay rate of the DRAM cells. Finally, Huang *et al.* presented TCC detection and DVFS-based countermeasure techniques for 2D and TSV-based 3D

integrated processors (17).

However, all of the previous works have established TCC with *computationally intensive programs*, which are relatively easier to detect. In this research, we demonstrate that by leveraging the 3D technologies, a moderate power SPLASH-2 benchmark application can be sufficient to establish a TCC attack with significant communication bandwidth. The bandwidth and BER of TCC are quantified in both Mono3D and TSV-based 3D processors. The robustness of TCC in the presence of thermal interference from applications running in other cores is also investigated for both 2D and 3D systems. Finally, we propose a novel detection metric to detect such TCC attacks established using low-power benchmark programs.

In this chapter, a detailed background information about adiabatic circuits with its application to AC computing, SIMON lightweight encryption algorithm, power-based CPA side-channel attacks, an introduction to two major 3D integration technologies and a detailed background information on thermal covert-channel attacks in multicore processors were presented . These concepts establish the foundation for the remaining chapters of this thesis.

Chapter 3

Power Analysis based Side-Channel Attack for an Unprotected Adiabatic Lightweight Cipher

In this chapter, the correlation power analysis (CPA) based side-channel attack is mounted on an unprotected adiabatic SIMON and the inherent resistance provided is evaluated against static CMOS based SIMON. Innovative methods that leverage specific adiabatic circuit characteristics are also presented to reduce attack complexity. Finally, the effect of target signal capacitance on the resistance to CPA attacks is studied. Several interesting implications of this study are discussed.

The organization of the chapter is as follows. The adiabatic SIMON32/64 hardware architecture is explained in Section 3.1. CPA attack methodology for an adiabatic SIMON is described in Section 3.2. The results of a successful attack on adiabatic SIMON and its comparison to static CMOS based implementation is pre-

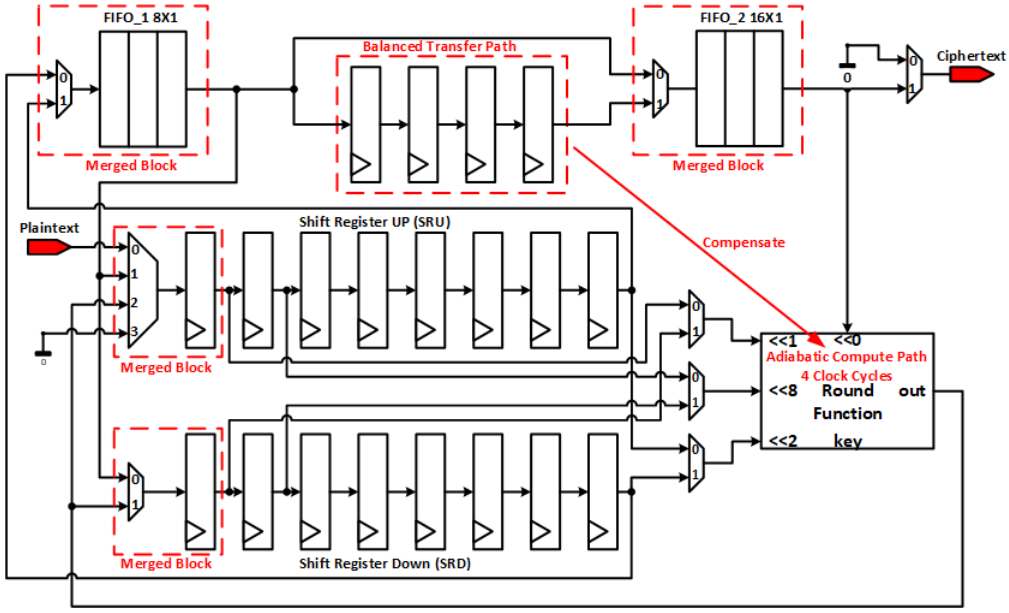


Figure 3.1: Implementation of round function in the adiabatic SIMON architecture, illustrating the merged blocks and balanced transfer paths.

sented in Section 3.3. The study of effect of target signal capacitance on the CPA attack resistance for both static and adiabatic implementations with results are shown in Section 3.4. Finally the chapter is summarized in Section 3.5

3.1 Ultra-low Power Adiabatic SIMON Architecture

The bit-serial static CMOS based SIMON consists of compute and transfer paths in the round function and key expansion modules (84). In the round function, a compute path is comprised of logical operations that compute each bit of the left word of a round operation and a transfer path consists of logic that shifts bits from the left word of a round operation to the right word of the successive round operation. The ping-pong shift registers, shift register up (SRU) and shift register

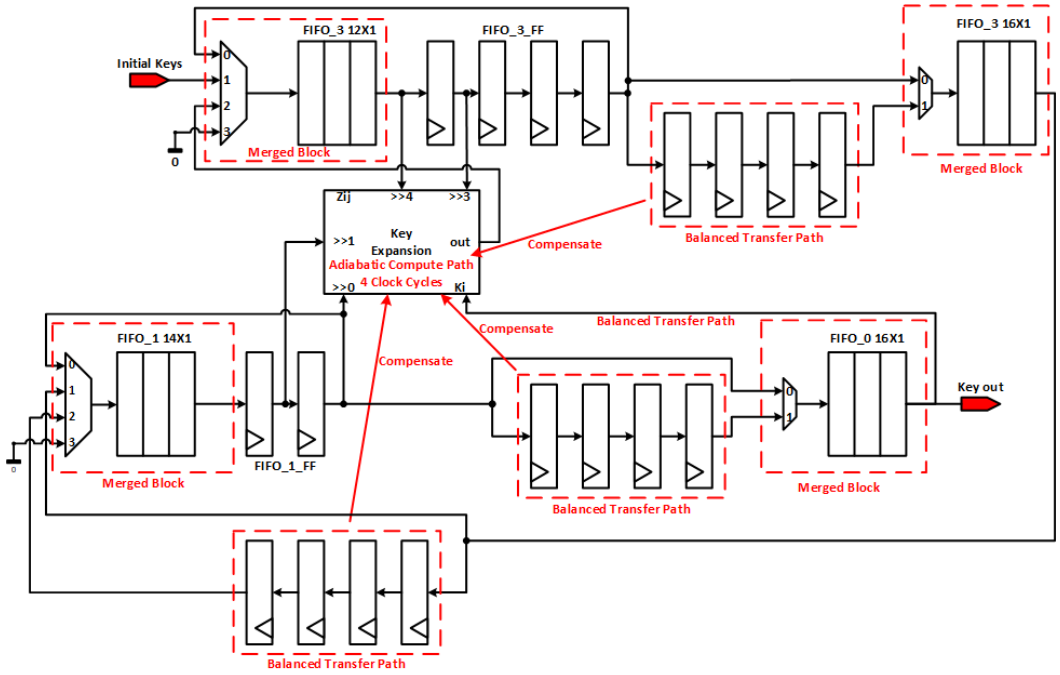


Figure 3.2: Implementation of the key expansion function in the adiabatic SIMON architecture, illustrating the merged blocks and balanced transfer paths.

down (SRD), are used to store the upper half left block output L_{i+1} and to perform the circular left shift operations, alternating their roles in each round (84). Adapting this static CMOS-based architecture for adiabatic operation requires several innovations to ensure timing synchronization. These innovations, illustrated in Fig. 3.1 (adiabatic round function) and Fig. 3.2 (adiabatic key expansion), are described below.

3.1.1 Merged blocks

Due to inherent pipelining in adiabatic logic (see Section 2.2.1), each multiplexer (designed as a single complex gate) in the adiabatic implementation adds an

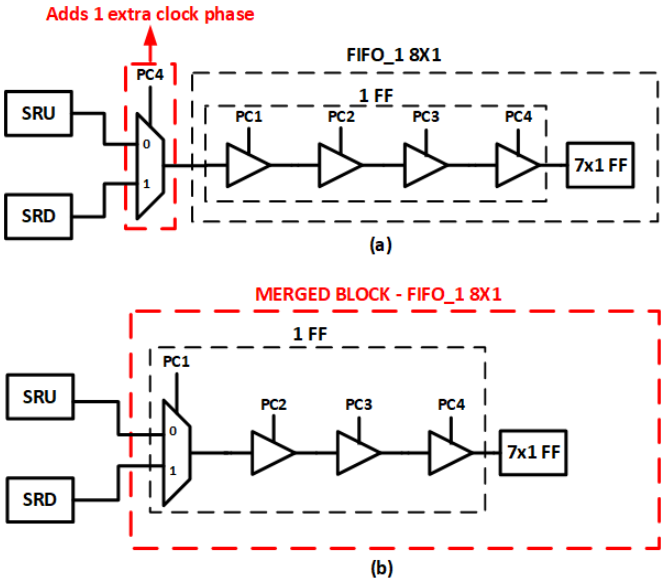


Figure 3.3: Example of a merged block in the round function: (a) multiplexer and $FIFO\_1$ 8×1 before merging, (b) multiplexer and $FIFO\_1$ 8×1 after merging.

additional clock phase. To compensate for this, multiplexers are merged with the following FIFO blocks, referred to as merged block in Figs. 3.1(b) and 3.2(b). For example, consider the 2-bit multiplexer driving the $FIFO\_1$, as shown in Fig. 3.3. The first flip-flop (FF) in the FIFO is a chain of 4 buffers with the respective power-clock signals, as shown in Fig. 3.3(a). Since the multiplexer adds an additional clock (PC4) phase delay, the input of the $FIFO\_1$ cannot be updated in every cycle, thus affecting the left shift operation. Therefore, the multiplexer is merged with the first FF, as shown in Fig. 3.3(b) to ensure that the bit-wise operation is consecutive. In this case, the merged block functions as a multiplexing flip-flop.

3.1.2 Balanced transfer paths

In the conventional static CMOS based bit-serial SIMON (84), four additional look-up table registers (LUT_FF) are used to store the output of the key expansion in the first four cycles so that the four MSB in the input $FIFO$ can be used for circular right shift operation at the same time. Starting from the fifth cycle, the output is stored back in the $FIFO$. Since adiabatic circuits are inherently pipelined, these four cycles of pipelining are integrated in the combinational logic within the key expansion block. The logic depth of this compute path is chosen according to the maximum number of bits to be shifted, which in this case is 4, thus eliminating the use of the LUT_FF. As a result, each computation takes four additional cycles and therefore the compute and transfer paths are not synchronized. For example, 20 cycles are consumed to compute a new word in the key expansion, whereas only 16 cycles are used to transfer the bits to the next word. In order to bridge this gap, four additional registers are added to balance each transfer path in both round function and key expansion modules. These additional registers are referred to as balanced transfer paths, as shown in Figs. 3.1(b) and 3.2(b). Note that due to the multi-phase operation of the adiabatic logic where each gate consumes 90° of the power-clock signal, four buffers [see Fig. 2.4(b) for a single buffer] are cascaded to realize the function of a flip-flop for data synchronization.

3.2 CPA Attack Methodology for Adiabatic SIMON

The process of mounting a CPA attack is comprised of choosing an intermediate target signal, mapping the intermediate values to a hypothetical power model, measuring the actual current traces of the circuit under attack and finally calculating the

correlation coefficients by statistically comparing the hypothetical power model to the actual current consumption to reveal the secret key (24). Each of these steps is elaborated in the following subsections for the proposed adiabatic SIMON implementation. The CPA attack methodology in this work assumes that the plaintexts to be encrypted are known to or chosen by the attacker.

3.2.1 Power model

The hypothetical power model maps the key-dependant data to the actual power/current consumption. For static CMOS logic gate, a commonly used power model representation is the Hamming distance (HD) model (85; 86) and it makes an assumption that current is drawn by a gate for $0 \rightarrow 1$ and $1 \rightarrow 0$ output transitions, while in reality the current is drawn from the power supply only to charge the output from $0 \rightarrow 1$.

Alternatively, for an adiabatic circuit implementation, the HD model is a more accurate power model representation for correlation measurement, as illustrated in Fig. 3.4. In this figure, output voltage simulation of an ECRL buffer with transitions $0 \rightarrow 1 \rightarrow 1 \rightarrow 0 \rightarrow 0$ and the corresponding power supply current are depicted. Note that the output voltage is discharged during the *recovery* phase irrespective of the input since the power-clock signal falls. Unlike static CMOS, the output transition occurs during the $evaluate$ stage of consecutive clock cycles. As indicated, whenever there is a change in the output voltage (*i.e.* $0 \rightarrow 1$ or $1 \rightarrow 0$), the charging current increases and $HD = 1$. However, when the output remains the same ($0 \rightarrow 0$ or $1 \rightarrow 1$), $HD = 0$ and the current decreases. This one-to-one correlation between the charging current and the HD renders this model to be a suitable choice for CPA on adiabatic SIMON implementation.

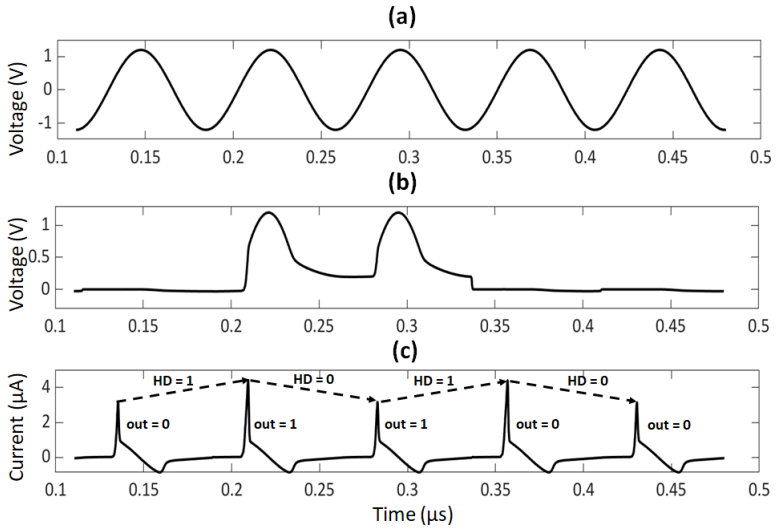


Figure 3.4: Use of Hamming distance as the power model for adiabatic ECRL circuits: (a) power-clock signal, (b) output voltage of the ECRL buffer, (c) current drawn from the supply by the buffer for output transitions $0 \rightarrow 1 \rightarrow 1 \rightarrow 0 \rightarrow 0$.

3.2.2 Intermediate signal for attack

In order to establish a successful CPA attack, an intermediate signal should be chosen such that the signal should be a function of a non-constant data value and a portion of the key (24). An immediate choice in SIMON algorithm is the output of a round function since the output of each round operation depends both on the key K_i and the computed output of the previous round for each random plaintext input, as expressed by (2.5). The output of the first round operation is a function of the first round key and the known plaintext, thus exhibiting a linear dependency with the key bits. For the attack to be more efficient, the intermediate result should have a non-linear dependency with the key and the key bits should get *diffused* with the state (86). Therefore, output of the second round operation is chosen as the target intermediate result.

For the proposed adiabatic SIMON implementation, the output of second round operation is stored in shift register SRU starting from the fifth cycle because of the four additional cycles added by the balanced transfer path, as shown in Fig. 3.5. Consequently, the HD model is constructed starting from L_0^2 and L_1^2 and is given by,

$$HD(L_0^2, L_1^2) = fn(K_8^1, K_{14}^1, K_{15}^1, K_0^2, K_9^1, K_0^1, K_1^2), \tag{3.1}$$

where, L_0^2 and L_1^2 are the first and second bit of the second round operation output. From (3.1), it can be seen that the HD is a function of seven bits of the 64-bit input key, $K_8^1, K_{14}^1, K_{15}^1, K_0^2, K_9^1, K_0^1, K_1^2$. Using this model, the matrix $HD(p,k)$ is constructed where $1 <= p <= P$ for P different random plaintexts and $1 <= k <= 128$ for the 128 hypotheses of the seven key bits in (3.1). This process is repeated for the consecutive cycles until the entire sample space of the 64 key bits is covered, as listed in Table 3.1. The table is divided into three sub-sections listing the power model for each successive round starting from the second round until all of the key bits are recovered. The total number of hypothesis complexity for the adiabatic SIMON32/64, as seen from the table, is 324.

Alternatively, for the static CMOS based SIMON32/64 implementation, the HD power model can be constructed starting from the sixteenth bit of the plaintext (L_{15}^0), as depicted by Fig. 3.6. The contents of the shift register SRU at three consecutive cycles starting from the last cycle of first round and the first cycle of the second round are shown in the figure. From (2.5), the HD of L_{15}^0 and L_0^2 is given by,

$$HD(L_{15}^0, L_0^2) = fn(K_8^1, K_{14}^1, K_{15}^1, K_0^2), \tag{3.2}$$

where L_{15}^0 is the sixteenth bit of the plaintext and L_0^2 is the first bit of the second

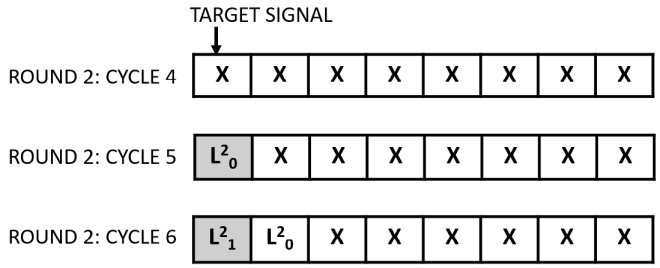


Figure 3.5: Contents of the 8-bit SRU loading the target signal at three cycles starting from the fourth cycle of second round for the proposed adiabatic SIMON.

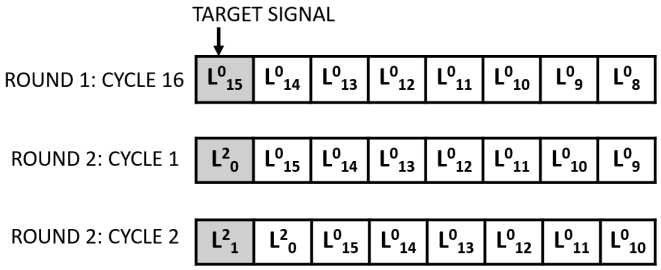


Figure 3.6: Contents of the 8-bit SRU loading the target signal at three cycles starting from last cycle of first round for the static CMOS based SIMON.

round output. The power model matrix is constructed for 16 key hypotheses in order to find the 4 bits $K^1_8, K^1_{14}, K^1_{15}, K^2_0$. Similarly, $HD(p,k)$ is constructed for each key hypotheses, as listed in Table 3.1, in order to find the correct 64 bits of the secret input key. The total number of key hypothesis complexity for the static CMOS based SIMON32/64 is reduced by approximately half (from 324 to 156) because of the change in the construction of the power model, as listed in Table 3.1.

3.2.3 Obtaining current traces

The current traces are typically obtained as samples and stored as a matrix $I(p,n)$, where $1 <= p <= P$ and $1 <= n <= N$ for P random plaintexts and N number of samples. For example, the latency for one encryption in an adiabatic SI-

Table 3.1: Complexity of the CPA attack for static CMOS based SIMON32/64 and adiabatic SIMON32/64 implementations: power model and number of key hypotheses required. L_n^m refers to the n^{th} bit of the left block output of the m^{th} round and K_n^m refers to the n^{th} bit of the m^{th} word of the input 64-bit key.

	Static SIMON			Adiabatic SIMON		
Hamming distance between	Bits of the input key	Number of key bits	Number of key hypotheses	Bits of the input key	Number of key bits	Number of key hypotheses
L_{15}^0 and L_0^2	K_8^1, K_{14}^1, K_{15}^1,K_0^2	4	16			
L_0^2 and L_1^2	K_9^1, K_0^1, K_1^2	3	8	K_8^1, K_{14}^1, K_{15}^1,K_0^2,K_9^1, K_0^1, K_1^2	7	128
L_1^2 and L_2^2	K_{10}^1, K_1^1, K_2^2	3	8	K_{10}^1, K_1^1, K_2^2	3	8
L_2^2 and L_3^2	K_{11}^1, K_2^1, K_3^2	3	8	K_{11}^1, K_2^1, K_3^2	3	8
L_3^2 and L_4^2	K_{12}^1, K_3^1, K_4^2	3	8	K_{12}^1, K_3^1, K_4^2	3	8
L_4^2 and L_5^2	K_{13}^1, K_4^1, K_5^2	3	8	K_{13}^1, K_4^1, K_5^2	3	8
L_5^2 and L_6^2	K_5^1, K_6^2	2	4	K_5^1, K_6^2	2	4
L_6^2 and L_7^2	K_6^1, K_7^2	2	4	K_6^1, K_7^2	2	4
L_7^2 and L_8^2	K_7^1, K_8^2	2	4	K_7^1, K_8^2	2	4
L_{15}^1 and L_0^3	K_{14}^2, K_{15}^2,K_0^3	3	8			
L_0^3 and L_1^3	K_9^2, K_1^3	2	4	K_{14}^2, K_{15}^2,K_0^3,K_9^2, K_1^3	5	32
L_1^3 and L_2^3	K_{10}^2, K_2^3	2	4	K_{10}^2, K_2^3	2	4
L_2^3 and L_3^3	K_{11}^2, K_3^3	2	4	K_{11}^2, K_3^3	2	4
L_3^3 and L_4^3	K_{12}^2, K_4^3	2	4	K_{12}^2, K_4^3	2	4
L_4^3 and L_5^3	K_{13}^2, K_5^3	2	4	K_{13}^2, K_5^3	2	4
L_{15}^2 and L_0^4	K_8^3, K_{14}^3, K_{15}^3,K_0^4	4	16			
L_0^4 and L_1^4	K_9^3, K_1^4	2	4	K_8^3, K_{14}^3, K_{15}^3,K_0^4,K_9^3, K_1^4	6	64
L_1^4 and L_2^4	K_{10}^3, K_2^4	2	4	K_{10}^3, K_2^4	2	4
L_2^4 and L_3^4	K_{11}^3, K_3^4	2	4	K_{11}^3, K_3^4	2	4
L_3^4 and L_4^4	K_{12}^3, K_4^4	2	4	K_{12}^3, K_4^4	2	4
L_4^4 and L_5^4	K_{13}^3, K_5^4	2	4	K_{13}^3, K_5^4	2	4
L_5^4 and L_6^4	K_6^4	1	2	K_6^4	1	2
L_6^4 and L_7^4	K_6^3, K_7^4	2	4	K_6^3, K_7^4	2	4
L_7^4 and L_8^4	K_7^3, K_8^4	2	4	K_7^3, K_8^4	2	4
L_8^4 and L_9^4	K_9^4	1	2	K_9^4	1	2
L_9^4 and L_{10}^4	K_{10}^4	1	2	K_{10}^4	1	2
L_{10}^4 and L_{11}^4	K_{11}^4	1	2	K_{11}^4	1	2
L_{11}^4 and L_{12}^4	K_{12}^4	1	2	K_{12}^4	1	2
L_{12}^4 and L_{13}^4	K_{13}^4	1	2	K_{13}^4	1	2
L_{13}^4 and L_{14}^4	K_{14}^4	1	2	K_{14}^4	1	2
L_{14}^4 and L_{15}^4	K_{15}^4	1	2	K_{15}^4	1	2
TOTAL		64	**156**		64	**324**

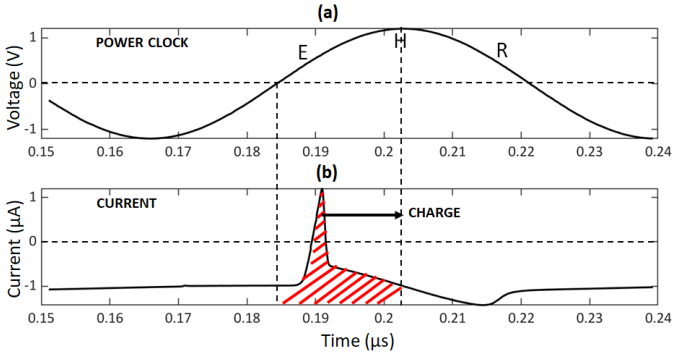


Figure 3.7: Proposed power sampling method in adiabatic SIMON core: (a) power-clock signal, (b) charge analysis with respect to the *evaluate* phase of the power-clock signal.

MON32/64 is 704 cycles. Assuming that an attacker measures around 200 current samples in each cycle (87), 140.8K samples (it takes 704 cycles for one encryption) should be collected. In a four-phase adiabatic logic such as ECRL, since charging occurs only during the *evaluate* phase (which is one fourth of the full cycle), as described in Section 2.2.1, the number of collected traces can be reduced by a factor of 4. Thus, it is sufficient to collect 35.2K current samples during the *evaluate* phase for an adiabatic SIMON.

A charge-based method of acquiring the traces is proposed in this work to further reduce the number of samples in adiabatic circuits. Specifically, in this paper, the traces are measured as an *integral* of current waveform over each *evaluate* stage of the power-clock signal, as illustrated in Fig. 3.7. The shaded portion in this figure indicates the charge obtained in one *evaluate* phase of a clock cycle. The charge traces acquired for the first plaintext can be expressed as,

$$Q(1,n) = \int_{[(n-1)T]}^{[(n-1)T+\frac{T}{4}]} I(t)dt, \tag{3.3}$$

where T is the time period of the power-clock signal. The lower and upper integration limits of the integral are determined based on the start and end times of the *evaluate* phase, which are known by the attacker via the power-clock signal. Note that this approach is not feasible in conventional static CMOS based operation since the current is drawn from the supply voltage based on the timing characteristics of the input of the target signal, which is typically not accessible to the attacker. In this work, the charge traces are obtained based on the simulated results using high performance Spectre APS simulation platform in Cadence Virtuoso environment (88). Power models are constructed and correlated with the charge traces to establish a CPA attack in MATLAB (89). Since modern digital oscilloscopes have the option of using mathematical functions such as calculating the integral in real-time, the number of collected samples can be reduced by approximately two orders of magnitude by integrating the current and effectively measuring the charge in each *evaluate* phase. Therefore, in the above example, the overall number of required samples to be collected can be reduced to only 704 (by measuring one charge sample in each clock cycle).

3.2.4 Correlation computation

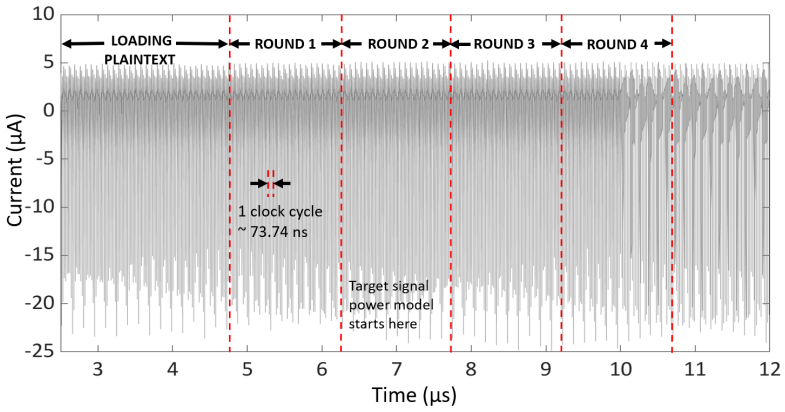
The Pearson correlation coefficient matrix $r(k,n)$ is calculated between the hypothetical power model, $HD(p,k)$ and the charge traces, $Q(p,n)$ in order to establish a CPA attack. The correct key hypotheses are given by the row number k with the maximum value of correlation coefficient. Measurements-to-disclosure (MTD) is the metric used to determine the resistance of the proposed hardware implementation against CPA attack (90). MTD is the number of current (charge for adiabatic) traces measured at the crossover point between the correlation coefficient

of the correct key and the maximum correlation coefficient of all of the incorrect key hypotheses. Higher MTD implies greater resistance to the attack.

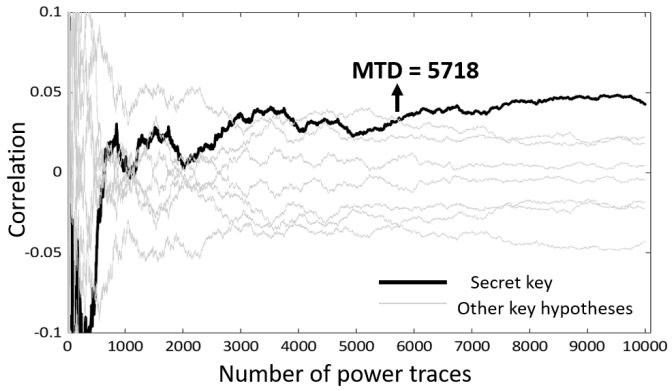
3.3 Results of the CPA Attack on Unprotected Adiabatic SIMON

The static CMOS and adiabatic ECRL SIMON core were implemented using a commercial 65 nm CMOS technology, operating at RFID frequency of 13.56 MHz. In order to establish a CPA attack, the methodology described in Section 3.2 was utilized. Current traces were measured for a large number of encryption scenarios with randomly generated input plaintexts with a key value `16'h 1918 1110 0908 0100`. The bit-serial adiabatic implementation of SIMON32/64 takes 32 rounds (with 20 cycles in each round) to encrypt one plaintext. A sample trace of the overall current consumption starting from loading the plaintext until the fourth round is depicted in Fig. 3.8(a).

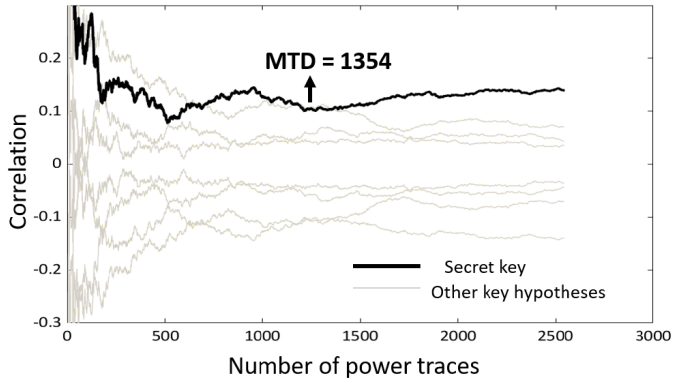
The CPA algorithm was built in MATLAB (89). The Hamming distance power model was constructed based on Table 3.1 for each key hypotheses. All of the key bits were successfully retrieved for both implementations. The correlation coefficient vs. number of power traces for static CMOS based SIMON for the key bits K_8^3, K_{14}^3, K_{15}^3,K_0^4 is illustrated in Fig. 3.8(b). The black curve shows the correlation coefficient for the correct key hypotheses `4'b 1000` and the grey curves are the correlation for the other key guesses. As observed from this figure, the MTD to retrieve all of the 64 bits of the key is determined as 1,354 power traces. Alternatively, for adiabatic ECRL based SIMON, the maximum MTD is 5,718 power traces, as de-



(a)



(b)



(c)

Figure 3.8: Correlation power analysis (CPA) attack: (a) sample current trace starting from the first round until the fourth round, (b) correlation vs. number of traces for static CMOS based SIMON with MTD = 1,354, (c) correlation vs. number of traces for adiabatic ECRL based SIMON with MTD = 5,718.

picted in Fig. 3.8(c). Thus, the SIMON block cipher implemented using adiabatic logic is approximately 4× less vulnerable to power side-channel attack as compared to the conventional static CMOS counterpart. It is important to note that, despite achieving enhanced inherent CPA resistance, the MTD of an unprotected adiabatic SIMON core is not sufficiently high to achieve full protection. As a comparison, in (91), a static CMOS based SIMON128/128 has been implemented for various levels of serialization. The MTD of the bit-serial implementation was reported to be 1,300, which is similar to the MTD of static CMOS based SIMON in this work. Therefore, the proposed adiabatic implementation is also 4× less susceptible to CPA when compared to (91).

3.4 Effect of Load Capacitance on CPA

The overall current consumption of a circuit during CPA attack can be expressed as,

$$I_{total} = I_{signal} + I_{noise}, \tag{3.4}$$

where I_{signal} is the current drawn to charge the CPA target signal capacitance and I_{noise} is the current consumed to charge all of the other nodes within the circuit. For an adiabatic circuit, the overall current consumption is given by,

$$I_{total} = \frac{C_{target} V_{dd}}{t_r} + \frac{C_{rem} V_{dd}}{t_r}, \tag{3.5}$$

where C_{target} is the capacitance of the target CPA signal including the interconnect capacitance, the gate capacitance of the load gate, and intrinsic capacitance. C_{rem} refers to the capacitance of other nodes in the circuit and t_r is the transition time

of the power-clock signal. According to (3.5), an increase in C_{target} amplifies the required target I_{signal}, isolating it from I_{noise}. This behavior can be observed in Fig. 3.9, where an increase in the width of the load gate increases the signal current without significantly affecting the noise current. The noise current is relatively independent of this change in the C_{target} in adiabatic operation since the load transistors are only n-type (due to the absence of a complementary pull-up network in ECRL circuits). Thus, increasing the width of the nMOS load transistor does not change the current consumed by the load gate. The measured current I_{total} is increased due to an increase in target I_{signal}. Based on (2.8), this increase contributes to a higher correlation coefficient of the correct key when compared to the incorrect coefficients. This improved correlation of the correct key results in a lower MTD and therefore, lesser resistance to CPA attack. An adversary typically has access to the interface ports of a system. Therefore, if the output ciphertext is chosen as the target signal, the load capacitance at the port can be modified by the attacker and the effect discussed here can cause the encryption core to be more vulnerable to the CPA attack. Note that in (92), the authors have studied the effect of an intentional load capacitance added to the output of a low-dropout (LDO) regulator (that powers an encryption core) on CPA attack and observed a similar result. In this work, however, the effect of the parasitic capacitance at the CPA target signal (within the encryption core) on CPA results are analyzed and compared for both static and adiabatic implementations of the SIMON core.

The dependence of MTD on the size of the load gate is shown in Fig. 3.10 for both static CMOS and adiabatic implementations. According to these results, for static CMOS implementation, if the size of the load transistor is increased by $6\times$ (thereby increasing the capacitance seen by the target signal), MTD is reduced by a

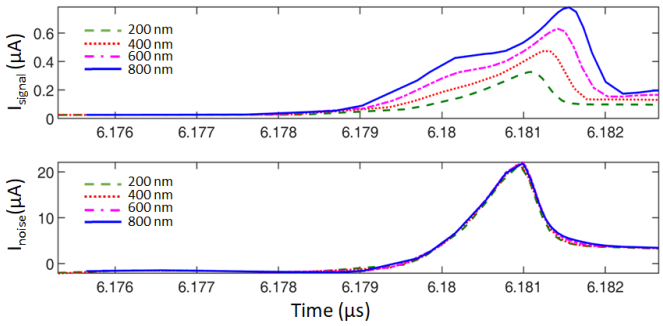


Figure 3.9: Signal and noise currents drawn from the power supply for different gate widths of the target signal load in adiabatic SIMON core.

factor of 2. Alternatively, for adiabatic implementation, the same reduction in MTD is observed when the size of the load transistor is increased by only $2\times$. Thus, the CPA attack on adiabatic SIMON is more sensitive to the changes in the capacitance seen by the target signal. The primary reason for this difference is related to the method of analysis of the current traces. Since the integral of current is used for adiabatic SIMON CPA attack, as explained in Section 3.2.3, the effect of increased load amplifies the charge at a higher rate than the peak current samples used in static CMOS based SIMON. This behavior is depicted in Fig. 3.11 where the dependence of charge and current on the size of load is shown. When the width of the load gate is increased by $4\times$, the charge consumed by the adiabatic ECRL is doubled whereas the peak current consumed by the static CMOS logic increases by approximately $1.2\times$. Thus, the correlation is higher for ECRL based SIMON for the same increase in load size, thereby reducing the MTD more.

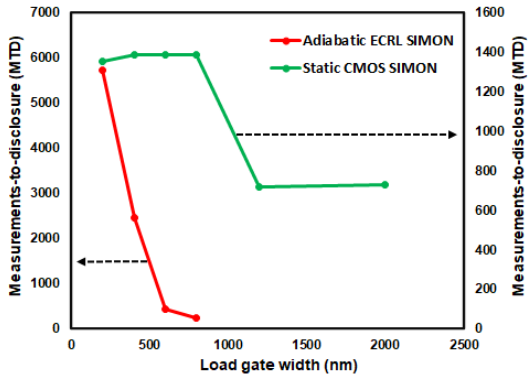


Figure 3.10: CPA target signal load size vs. MTD for static CMOS based SIMON core and adiabatic SIMON core.

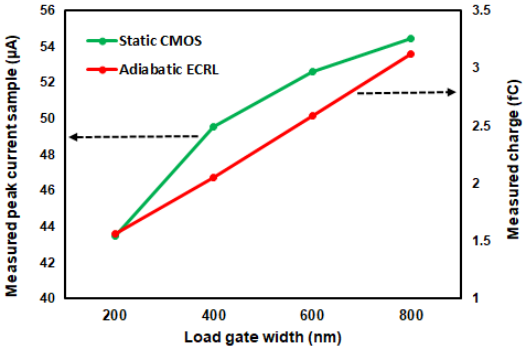


Figure 3.11: Dependence of peak current and charge drawn by the driving gate on target load capacitance for static CMOS and adiabatic ECRL.

3.5 Summary

A correlation power analysis (CPA) attack was established on an adiabatic SIMON block cipher. It was demonstrated that adiabatic operation exhibits approximately $4\times$ higher inherent CPA resistance as compared to static CMOS based SIMON implementation. A charge-based method to measure the current traces in adiabatic operation was proposed. This method significantly reduces the CPA attack complexity in adiabatic circuits by reducing the required number of samples by two orders of magnitude. The effect of increasing the target load capacitance on the side-channel resistance was also investigated. The results demonstrate that doubling the capacitance seen by the target signal in the proposed adiabatic SIMON implementation can reduce the MTD by $5\times$.

Chapter 4

SEAL-RF: SEcure Adiabatic Logic for Wirelessly-Powered IoT Devices

From the previous chapter, it was concluded that ECRL is a robust adiabatic logic with promising characteristics that inherently provide $4\times$ resistance to power side-channel attacks as compared to conventional static CMOS based SIMON. However, the amount of protection provided is not sufficient to develop a side-channel resistant cipher.

In order to improve the resistance to power side-channel attacks, several countermeasures have been developed to reduce the correlation between data and power consumption. Firstly, some techniques reduce the signal-to-noise ratio of the leaked information by incorporating noise injecting circuits such as ring oscillators, that purely inject noise to the supply current traces and thereby reduce the correlation (93). Moreover, there are also circuits that isolate the power supply from the encryption core by using switched capacitor and integrated voltage regulator techniques

(92). Finally, there are several circuit-based countermeasures, that aim at reducing the dependence of power consumption on data transitions at the input of every logic gate (94; 95) and is the primary focus of this chapter.

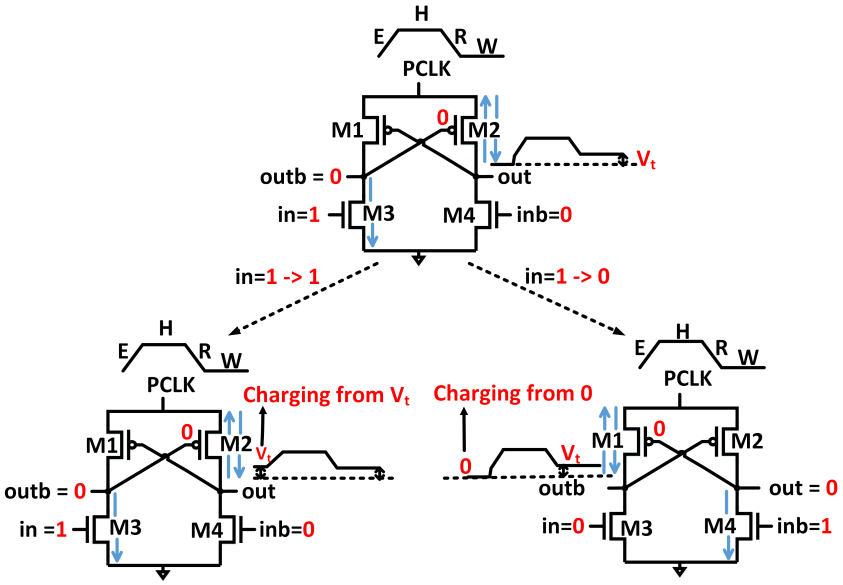
In this chapter, the problems associated with the unprotected adiabatic ECRL are discussed in Section 4.2, followed by the existing works on adiabatic circuit-based countermeasures in Section 4.2. A novel protected adiabatic logic for AC computing, called SEcure Adiabatic Logic for Wirelessly-Powered IoT Devices (SEAL-RF) is presented in Section 4.3. Security metrics that quantify the amount of power side-channel attack resistance provided by SEAL-RF based logic gates are evaluated in Section 4.4. The power and energy metrics are evaluated for SEAL-RF based logic gates in Section 4.5 and the chapter is finally summarized in Section 4.6.

4.1 Side-Channel Leakage in Unprotected Adiabatic Logic

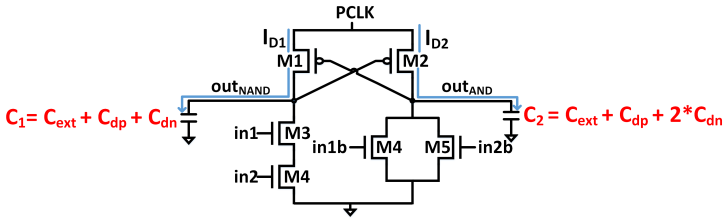
Efficient charge recovery logic (ECRL) is an example of a relatively robust adiabatic logic family, as shown in Fig. 4.1(a) where an ECRL based inverter/buffer is illustrated. In ECRL, a 4-phase *PCLK* signal is used where the phase difference in the *PCLK* signal of two adjacent gates is 90°. As such, an ECRL logic gate has four phases of operation: "Evaluate (E)", "Hold (H)", "Recover (R)" and "Wait (W)", as indicated in the figure. For example, when an ECRL gate is in the "Evaluate" phase, the preceding gate is in the "Hold" phase and the next gate is in the "Wait" phase. An ECRL inverter/buffer operates as follows. When $in = 1$, $outb = 0$, thus turning on the pMOS $M2$ (in Fig. 4.1(a)) and charging *out* signal. Because of the

MOSFET characteristics, $M2$ turns on only after the *PCLK* reaches its threshold voltage (V_t). Similarly, during the "Recover" phase, the output voltage recovers and closely follows *PCLK* signal, only until it reaches the V_t of $M2$, as depicted in Fig. 4.1(a). Thus, some of the charge at the output cannot be recovered since $M2$ (pMOS) cannot fully pass logic-low. This characteristic makes conventional adiabatic logic vulnerable to power side-channel attacks. Specifically, consider the behaviour of an ECRL buffer for two scenarios: *in* switching from '1' to '0' and *in* remaining at '1', as shown in Fig. 4.1(a). When *in* remains the same in the following cycle, *out* charges through $M2$, starting from V_t. However, if $in = 0$ in the next cycle, $inb = 1$, therefore $out = 0$ and $outb$ charges starting from '0' through $M1$. Thus, based on the successive input signals, the output nodes start charging with an initial voltage of either '0' or V_t as shown in the figure. Therefore, the charging current drawn by an adiabatic ECRL buffer is correlated to the input transitions, making ECRL susceptible to power side-channel attacks.

The second side-channel leakage mechanism is due to capacitance imbalance at the complementary output nodes, as described with the help of an ECRL based AND/NAND gate shown in Fig. 4.1(b). Due to the asymmetry of the two differential paths, the capacitance at the output nodes is not equal. When both inputs are high, $out_{NAND} = 0$ turning $M2$ on and charging C_2 by the supply current I_{D2}. For all of the other input combinations, $M1$ turns on and C_1 is charged through I_{D1}. To reduce power side-channel susceptibility, the current consumption for all of the input combinations should be the same (i.e. I_{D1} should be equal to I_{D2}). However, due to imbalance in the output capacitors, $I_{D2} > I_{D1}$. This current imbalance exists in all of the ECRL gates with multiple inputs, resulting in increased correlation between the power consumption and switching inputs. Several secure adiabatic logic gates have



(a)



(b)

Figure 4.1: Security drawbacks of existing ECRL logic family showing: (a) different initial output voltages for two input transitions $in = 1 \rightarrow 1$ and $in = 1 \rightarrow 0$ for ECRL buffer and (b) output load capacitance imbalance in ECRL AND/NAND gate where C_{ext} includes the output interconnect capacitance and the gate capacitance of the load transistor, C_{dp} is the drain capacitance of the pMOS transistor and C_{dn} is the drain capacitance of the nMOS transistor.

been proposed in the literature to mitigate these side-channel leakage mechanisms, as summarized in the following section.

4.2 Existing Secure Adiabatic Logic Families and Limitations

In order to mitigate the issue of voltage imbalance at the output nodes, explained in Fig. 4.1(a), a majority of the existing secure adiabatic logic families use additional transistors to balance or discharge the output nodes before each evaluation. Furthermore, to overcome the imbalance in output capacitances, illustrated in Fig. 4.1(b), existing secure logic gates utilize symmetric circuit structures for computing the complementary outputs. These existing secure adiabatic families and their limitations are summarized below.

Symmetric Adiabatic Logic (SyAL) (96) is an ECRL based secure logic that uses charge sharing transistors with external inputs to balance the initial voltages at the output nodes and internal nodes. Charge-Sharing Symmetric Adiabatic Logic (CSSAL) (97; 98) is an enhancement of SyAL for security that not only balances the initial voltages at the outputs, but also ensures that the equivalent *RC* to charge the outputs and internal nodes is approximately the same for each input combination. This enhancement further reduces the input-dependent current consumption. However, CSSAL requires 12 trapezoidal external voltage sources, which significantly increases the overhead.

Secure Quasi Adiabatic Logic (SQAL) (96) was proposed that has a similar structure to SyAL, but with significantly fewer transistors (e.g. an SQAL XOR

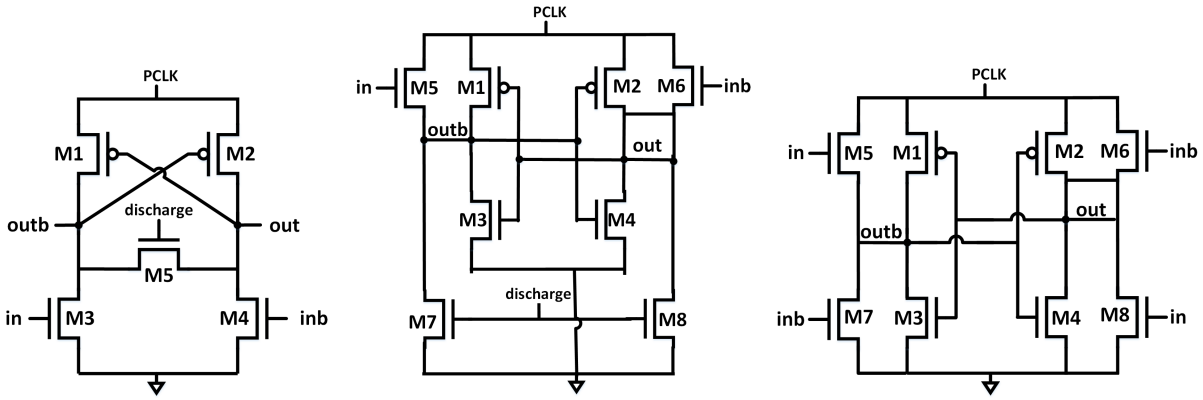


Figure 4.2: Existing secure adiabatic logic buffers: (a) SQAL (96), (b) EE-SPFAL (102), and (c) WCS-QuAL (103) / EQUAL (104).

gate has 9 transistors whereas SYAL XOR has 15 transistors and CSSAL XOR has 18 transistors) and enhanced security against power side-channel attacks. The schematic of an SQAL buffer is illustrated in Fig. 4.2(a). However, SQAL logic suffers from non-adiabatic energy loss (7), which was mitigated by Secure Pass Gate Adiabatic Logic (SPGAL) (95; 99; 100; 101) and its enhancement Energy-Efficient Secure Positive Feedback Adiabatic Logic (EE-SPFAL) (102). SPGAL and EE-SPFAL outperform all of the previous logic families in terms of energy-efficiency and security. The schematic of an EE-SPFAL buffer is shown in Fig. 4.2(b). Both of these logic families mitigate the non-adiabatic energy loss via connecting evaluating nMOS transistors ($M5$ and $M6$ in Fig. 4.2(b)) in parallel to the pMOS transistors. Compared to SPGAL, EE-SPFAL logic has increased robustness due to additional cross-coupled nMOS transistors ($M3$ and $M4$).

Successively, Without Charge Sharing Quasi Adiabatic Logic (WCS-QuAL) (103; 105) was proposed, where the output nodes of a logic gate have equal capac-

itance for all of the input transitions. This symmetric structure is achieved by using a dual-duplicate evaluation-network (transistors $M5-M8$), as shown in Fig. 4.2(c). Furthermore, this approach avoids the use of an external charge sharing/discharge signal to balance the output voltages before evaluation. Instead, the output nodes of the WCS-QuAL buffer are discharged through the transistors $M7$ or $M8$ since at least one of them is active before each evaluation. An extension of the WCS-QuAL, referred to as Efficient QUasi Adiabatic Logic (EQUAL) (104), was recently proposed. EQUAL-based NAND/NOR gates reduce the number of transistors by up to 40% (therefore improve energy efficiency) while achieving comparable security characteristics to WCS-QuAL.

All of the secure adiabatic logic gates discussed above have four-phase operation. A 3-phase adiabatic logic family (106) has also been proposed, that has a similar structure to the EE-SPFAL, but the *PCLK* signal does not have the "Hold" phase. This is achieved by reducing the slope of the *PCLK* signal and by optimizing the design of the *PCLK* generator. As such, this logic family outperforms all of the previously discussed secure adiabatic logic structures in terms of CPA attack resistance.

Despite these innovations in the design of secure adiabatic logic gates, application of these existing structures to AC computing suffers from several major drawbacks. In AC computing methodology, as described in Section 2.2.1.2, the harvested RF signal is a bipolar sinusoidal wave with both positive and negative components (see Fig. 2.5). This AC signal is directly used as the *PCLK* signal for the adiabatic logic based computation blocks. Therefore, the 3-phase adiabatic logic family (106) cannot be used because of the specific requirement for a 3-phase *PCLK* signal generator. Among the other existing secure adiabatic logic families,

SPGAL (100), EE-SPFAL (102), WCS-QuAL (103) and EQUAL (104) cannot satisfactorily operate with a bipolar *PCLK* signal. The reason is that the bipolar *PCLK* is connected to the source (or drain) terminal of the evaluating nMOS transistors (for example, $M5$ and $M6$ in Fig. 4.2(b) and (c)). These nMOS transistors fail to operate during the negative half of the *PCLK* signal since the body-source junction diode becomes forward biased. Thus, significant amount of body current (in the range of milli amps) flows through both evaluating nMOS transistors, irrespective of the input transitions. This large body current not only increases the power consumption by several orders of magnitude, but it can also lead to reliability issues such as latchup (107). Therefore, these secure logic families can be operated with AC computing only if the harvested signal is converted to a unipolar signal, which further increases the overhead.

Due to this limitation, only SyAL (96), CSSAL (98), and SQAL (96) can operate with AC computing methodology since these logic families do not have evaluation nMOS transistors where the source/drain terminal is connected to the *PCLK* signal. However, these logic families rely on external input signals to discharge the output nodes. Since these signals are 4-phased (similar to the *PCLK* signal), the generation and distribution of these additional input signals introduce significant energy overhead (43; 108; 109; 110).

Due to these limitations of the existing secure adiabatic logic families, a novel secure adiabatic logic for RF-powered AC computing applications, referred to as SEAL-RF, is proposed in this work. SEAL-RF can reliably work with a bipolar sinusoidal *PCLK* signal while exhibiting high security characteristics and low power consumption, as detailed in the following sections.

4.3 Proposed SEcure Adiabatic Logic for RF-powered Devices (SEAL-RF)

SEcure Adiabatic Logic (SEAL-RF) is described in this section. In SEAL-RF, output capacitances are discharged in each cycle by *reusing* the negative phases of the *PCLK* signal. Note that discharging these capacitances are important to enhance power side-channel resistance, as described in the previous section. Since SEAL-RF exploits the negative phase of the *PCLK* signal for this task, the need for external inputs is eliminated. The schematic of the proposed SEAL-RF buffer/inverter gate and its operation for each phase of the *PCLK* signal are depicted in Fig. 4.3. The five phases of operation, "Evaluate (E)", "Hold (H)", "Recover (R)", "Discharge (D)" and "Wait (W)" are detailed below:

- *Evaluate (E):* In this phase, evaluation of the buffer is performed using nMOS transistors $M3$ and $M4$ based on the input voltages. In Fig. 4.3(a), *in* is assumed to be '0' and *inb* is assumed to be '1', turning $M4$ ON. Therefore, the capacitance at *out* is discharged through $M4$. The *PCLK* signal in this phase increases higher than the threshold voltage of the pMOS transistors (V_{tp}), as indicated by the red shaded region in the figure. Since the source-gate voltage difference for $M1$ (V_{sg-M1}) is greater than $|V_{tp}|$, $M1$ turns ON, charging the capacitance at $outb$. Node $outb$ follows the *PCLK* signal in this phase, thus ensuring adiabatic operation. The discharge transistors, $M5$ and $M6$, are OFF.

- *Hold (H):* During the "Hold" phase, the evaluation transistors $M3$ and $M4$ turn OFF and the outputs are held stable to enable the evaluation of the successive gates. Fig. 4.3(b) shows the operation of the SEAL-RF buffer during

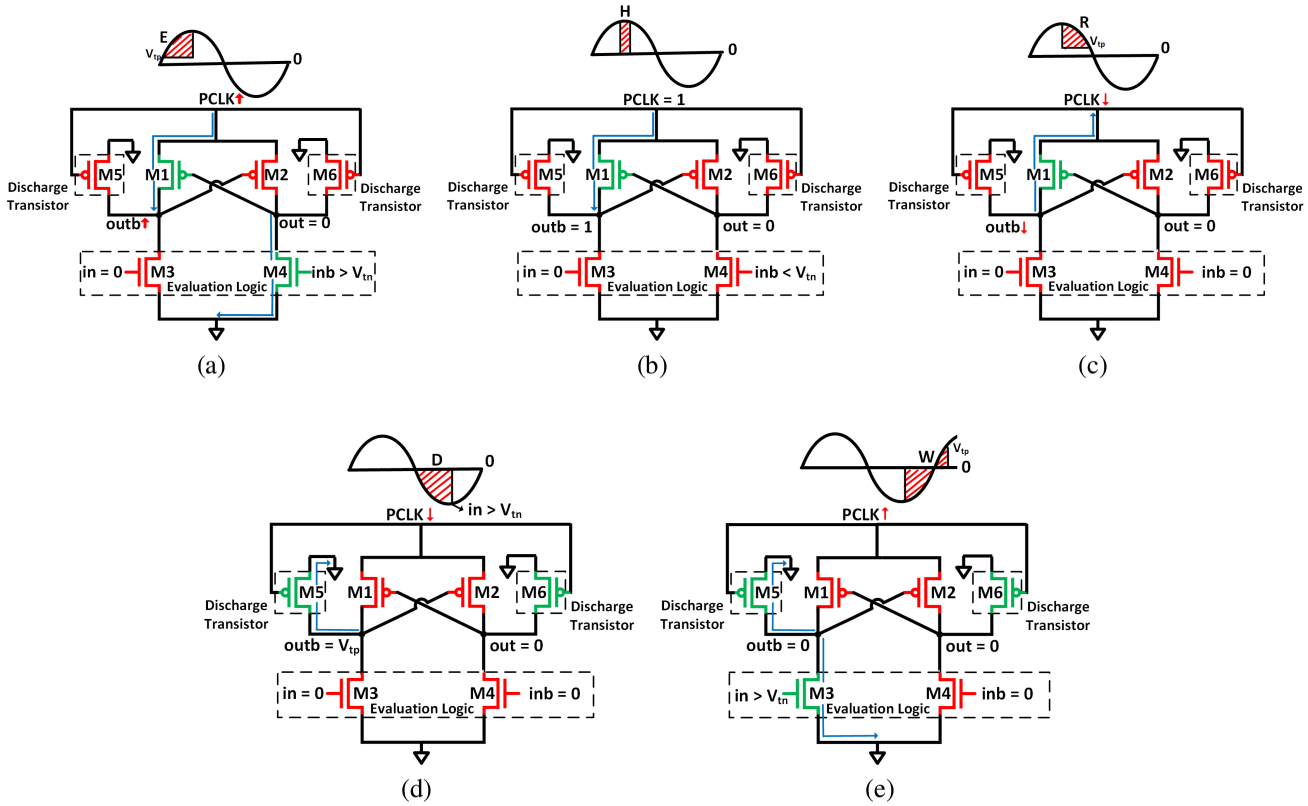


Figure 4.3: Schematic of the proposed SEcure Adiabatic Logic (SEAL-RF) buffer and its operation during (a) "Evaluate" phase, (b) "Hold" phase, (c) "Recover" phase, (d) "Discharge" phase and (e) "Wait" phase of the *PCLK* signal. The green transistors are ON and the red transistors are OFF. The blue arrows indicate the current path.

this phase. Note that the "Hold" phase of *PCLK* (shaded in red) is very small because of its sinusoidal nature. The input *inb* falls less than its threshold voltage V_{tn}, thus turning $M4$ OFF. Since both the evaluation transistors are not conducting and the *PCLK* signal is at maximum amplitude, both outputs are held stable by the cross-coupled pMOS transistors $M1$ and $M2$. The discharge transistors, $M5$ and $M6$, remain OFF.

- *Recover (R):* Charge recovery occurs in this phase, where the *PCLK* signal starts to decrease, as shown in Fig. 4.3(c). The transistors $M3$ and $M4$ remain OFF and since V_{sg-M1} is still greater than $|V_{tp}|$, $outb$ is discharged through $M1$ until it reaches $|V_{tp}|$. Thus, $outb$ follows the *PCLK* signal, recycling the charge back to the power supply, as indicated by the direction of current in the figure. Note that charge recovery occurs in this phase, only until *PCLK* signal reaches $|V_{tp}|$, after which $M1$ turns OFF. During this phase, the discharge pMOS transistors, $M5$ and $M6$, remain OFF.

- *Discharge (D):* In this phase, the output capacitances of the SEAL-RF buffer are completely discharged to ensure input independent current consumption. The *PCLK* signal in this phase falls below '0', as shown by the red shaded region in Fig. 4.3(d). Since V_{sg-M1} is less than $|V_{tp}|$, $M1$ turns OFF. In the proposed SEAL-RF buffer, the negative polarity of the *PCLK* signal is exploited to discharge $outb$ completely before the next "Evaluation" phase. Consequently, *PCLK* signal is connected to the gate terminal of two pMOS discharge transistors ($M5$ and $M6$), as shown in the figure. The other two terminals of the transistors are connected to the output nodes and ground. When *PCLK* signal falls below 0 V, the source-gate voltage difference of

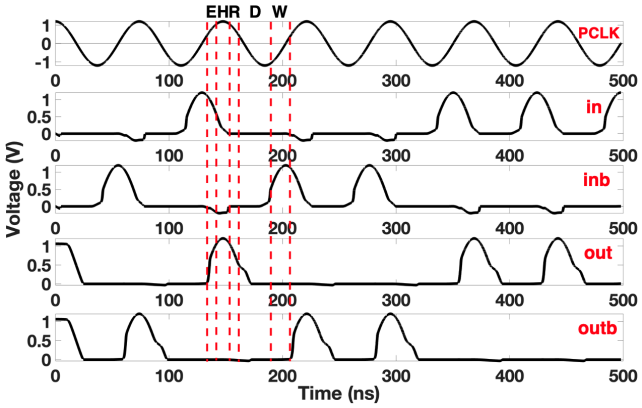


Figure 4.4: Waveforms of the bipolar *PCLK* signal marked with the five phases of operation, inputs (*in* and *inb*) and outputs (*out* and *outb*) of the proposed SEAL-RF buffer.

$M5$, $V_{sg-M5} > |V_{tp}|$, thus completely discharging the capacitance at *outb*, as indicated by the current direction in the figure. Similarly, any charge accumulation at *out* is completely discharged through $M6$.

- *Wait (W):* The operation of the buffer in this phase is shown in Fig. 4.3(e). Since the preceding gate starts the next evaluation in this phase, one of the inputs (for example *in*) starts to increase above V_{tn}. Therefore, the corresponding evaluation nMOS transistor ($M3$) turns on. However, since *PCLK* signal is still less than $|V_{tp}|$ in this phase, $M1$ and $M2$ are OFF and the discharge transistors $M5$ and $M6$ remain ON. Thus the outputs are still completely discharged, waiting for the next evaluation phase.

The waveforms of the bipolar *PCLK* signal, inputs *in* and *inb* and outputs *out* and *outb* of a SEAL-RF buffer gate are depicted in Fig. 4.4. The five phases of operation are marked for one of the cycles of the *PCLK* signal. The discharging of both outputs when *PCLK* signal is less than 0 V can be observed from the figure.

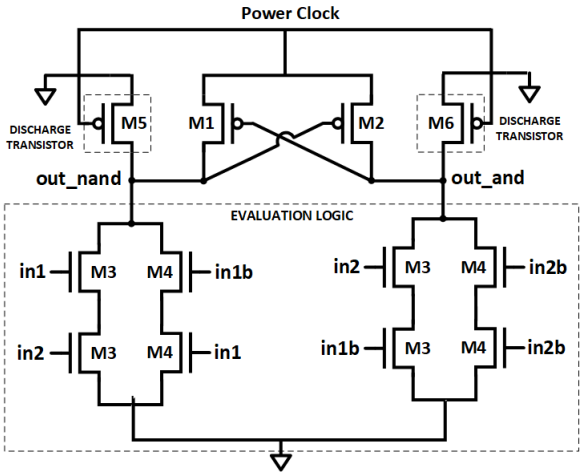


Figure 4.5: Schematic of the proposed SEAL-RF AND/NAND gate.

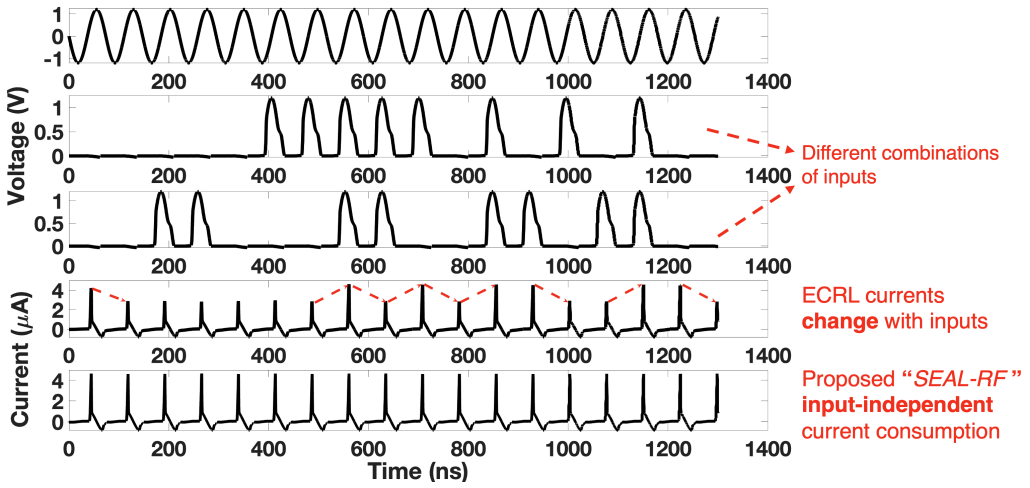


Figure 4.6: Waveforms of the bipolar *PCLK* signal, in1, in2, ECRL supply current and proposed SEAL-RF supply current for AND/NAND gate for all combinations of the input transitions. The dotted lines indicate the variation of ECRL gate current consumption with inputs.

Therefore, SEAL-RF exploits the negative polarity of the harvested *PCLK* signal to discharge the output capacitances (thereby reducing the correlation between inputs and current consumption) instead of relying on external inputs. Furthermore, SEAL-RF balances the output capacitances to further minimize the dependence of supply current on input signals, as shown in Fig. 4.5 for a SEAL-RF AND/NAND gate. The evaluation nMOS transistors for the AND/NAND gate logic are $M3$, $M5$, $M9$ and $M10$. The additional transistors $M4$, $M6$, $M7$ and $M8$ are added to make the evaluation network symmetric and thus balance the capacitance at the complementary output nodes. Therefore, both drawbacks described in Section 4.1 are mitigated by the proposed SEAL-RF logic.

The input-independent nature of the supply current can be observed for the SEAL-RF AND/NAND gate in Fig. 4.6. The current waveforms are illustrated for both the conventional ECRL and proposed SEAL-RF for 16 different input transitions for the 2-input AND/NAND gate. As observed in this figure, the current peaks of the ECRL gates vary depending upon the input signals. Alternatively, the current peaks for the proposed SEAL-RF AND/NAND gate is independent of the input transitions, making it highly resistant to CPA attacks. Furthermore, since the additional discharge transistors do not draw any current from the power supply, the high energy efficiency is maintained. Both of these metrics (CPA attack resistance and energy efficiency) are quantified in the following section.

4.4 Security Evaluation of SEAL-RF Based Logic Gates

The circuits are designed and simulated using a commercial 65 nm technology node. The harvested *PCLK* signal is simulated as a bipolar sinusoidal signal with

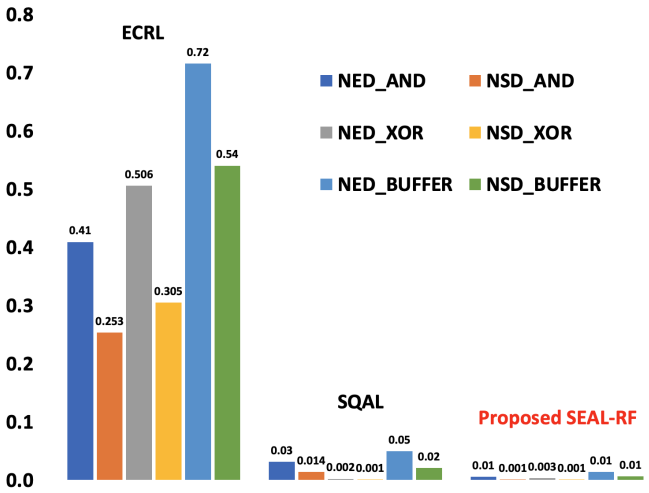


Figure 4.7: Comparison of the security metrics NED and NSD of the proposed SEAL-RF AND/NAND, OR/NOR and BUFF/INV gates with the NED and NSD of the corresponding adiabatic ECRL and SQAL gates for all possible input transitions.

Adiabatic logic family	**Average energy per transition (fJ)**			**Number of transistors**		
	AND/NAND	XOR/XNOR	BUFF/INV	AND/NAND	XOR/XNOR	BUFF/INV
ECRL (111)	3.7	4.63	5.12	6	6	4
SQAL (112)	4.07	5.27	3.83	13	9	5
Proposed SEAL-RF	**3.79**	**3.58**	**3.1**	**12**	**10**	**6**

Table 4.1: Comparison of average energy per transition and number of transistors of the proposed adiabatic logic family (SEAL-RF) with conventional adiabatic logic (ECRL) and secure adiabatic logic (SQAL).

RFID frequency of 13.56 MHz and a maximum amplitude of 1.2 V. At the gate-level, the security of the proposed SEAL-RF is evaluated with the help of two commonly used metrics: Normalized Energy Deviation (NED) and Normalized Standard Deviation (NSD) (113). The metric NED is defined as, $NED = \frac{E_{max} - E_{min}}{E_{max}}$, where E_{max} and E_{min} are the maximum and minimum energy consumed over all the combinations of input transitions. The metric NSD is defined as $NSD = \frac{\sigma}{E_{avg}}$, where σ is the standard deviation and is defined as $\sqrt{\frac{\sum_{i=1}^{N}(E_i - E_{avg})^2}{N}}$, where E_i is the energy consumption per input transition, E_{avg} is the average energy for all combinations of input transitions and N is the total number of possible combinations of input transitions. For a two-input adiabatic logic gate, if input A is assumed to transition from A_i to A_f and input B is assumed to transition from B_i to B_f in consecutive *PCLK* cycles, there can be 16 total combinations of A_i, A_f, B_i and B_f. Therefore, NED and NSD are calculated for all of the 16 different input transitions for a two-input adiabatic gate, and similarly for 4 different input transitions for a single-input adiabatic buffer. Lower NED and NSD signify higher resistance to power side-channel attacks. These results are compared with ECRL (conventional unprotected adiabatic logic) and SQAL (96) (existing secure adiabatic logic) since these are existing adiabatic families that can operate with bipolar *PCLK* signal. The comparison results are shown in Fig. 4.7 for BUFF/INV, AND/NAND and XOR/XNOR gates. As observed from the bar plot, unprotected ECRL has the maximum NED and NSD values. The NED and NSD of SEAL-RF based logic gates is up to 180× and 195× lower than ECRL and is up to 5.5× and 10.8× lower than SQAL, respectively. Thus, SEAL-RF gates achieve significantly higher protection against power side-channel attacks and outperform existing secure adiabatic logic.

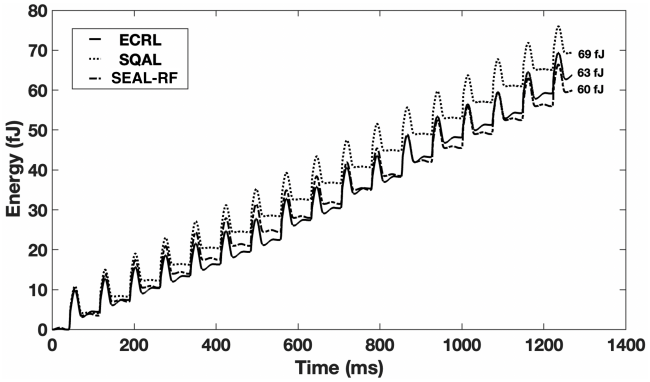


Figure 4.8: Comparison of the transient energy consumption of ECRL, SQAL and proposed SEAL-RF based AND/NAND logic gate.

4.5 Energy Evaluation of SEAL-RF Based Logic Gates

The energy consumption of ECRL, SQAL and the proposed SEAL-RF based AND/NAND adiabatic logic gate is illustrated in Fig. 4.8 as a function of time for all of the 16 possible input transitions. The low energy consumption and recovery characteristic expected from adiabatic operation can be observed for each logic family. The overall energy consumption at the end of 17 *PCLK* cycles is indicated as 69 fJ for SQAL, 63 fJ for ECRL, and 60 fJ for the proposed SEAL-RF AND/NAND gate. The proposed SEAL-RF based logic gate has the least energy consumption, primarily because both output nodes are completely discharged/recovered before each evaluation. Furthermore, the additional discharge transistors in SEAL-RF do not consume any current from the power supply. Alternatively, for SQAL and ECRL, additional energy loss is due to unwanted charge sharing that occurs during a short duration within the "Evaluation" phase (before the pMOS transistors start conducting). Note that the energy consumption of ECRL gate is less than SQAL since the output capacitance is charged starting from different initial voltages (V_{tp} or '0'), re-

sulting in reduced current consumption for several cycles, as can be observed from Fig. 4.6.

The average energy consumed per transition for BUFF/INV, AND/NAND and XOR/XNOR gates is listed in Table 4.1 for each logic family. The average energy per transition for SEAL-RF based logic gates is up to 39% lower compared to unprotected ECRL and up to 32% lower compared to SQAL-based gates. Thus, SEAL-RF enhances security characteristics while also lowering the overall energy consumption. The table also lists the number of transistors for each gate for the three logic families. Both SQAL and SEAL-RF based gates have higher number of transistors than unprotected ECRL. The number of transistors is comparable for SQAL and SEAL-RF.

4.6 Summary

In this chapter, the shortfalls of ECRL with respect to security is identified and the existing works on secure adiabatic logic families that overcome these problems have been discussed. The design of a novel secure adiabatic logic gate that operates with a bipolar sinusoidal wave for AC computing applications, is proposed. The operation of the proposed logic has been explained, in addition to the schematics of the individual gates and the relevant waveforms. The chapter is concluded by comparing the security metrics in the gate level between the state-of-the art bipolar adiabatic logic designs. Specifically, at the gate-level, the normalized energy deviation (NED) of SEAL-RF is up to 180$\times$ lower than conventional (unprotected) adiabatic logic, while consuming up to 39% less average energy per transition. Furthermore, the NED of SEAL-RF is up to 5.5$\times$ lower than an existing secure adiabatic logic,

while consuming up to 32% less average energy per transition.

Chapter 5

Power Analysis Attack on SEAL-RF SIMON Core

A novel adiabatic logic for AC computing called SEAL-RF, has been proposed in the previous chapter and the gate-level security characteristics have been compared with the state-of-the-art. However, it is important to evaluate the resistance of a secure adiabatic logic to power side-channel attack, in a system level and hence in this chapter, a CPA attack has been mounted on the SIMON encryption core been designed with SEAL-RF.

The chapter is organized as follows. The functional verification of SEAL-RF based SIMON is performed in Section 5.1. The results of the CPA attack on SEAL-RF SIMON are presented in Section 5.2. The power and energy characteristics are compared between ECRL and SEAL-RF based SIMON blocks in Section 5.3. and the chapter is finally concluded in Section 5.4.

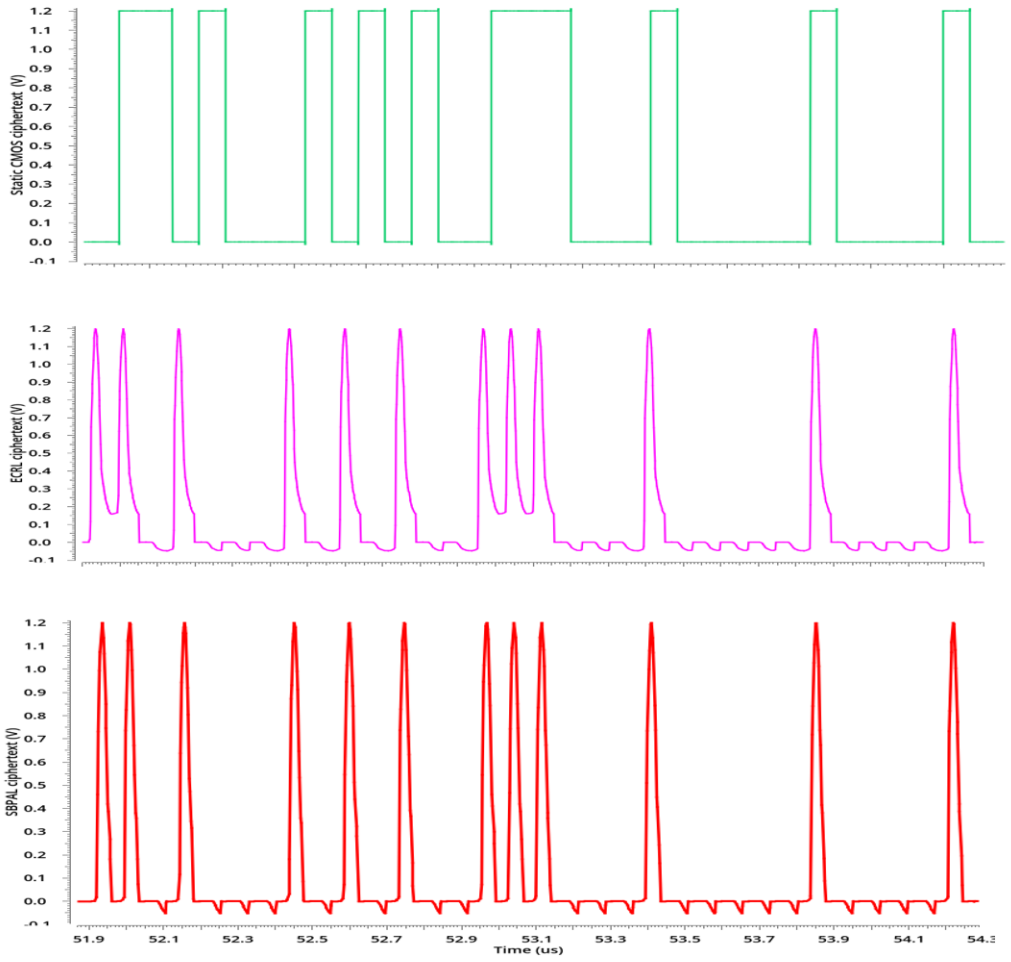


Figure 5.1: Ciphertext outputs of Static CMOS, ECRL and the proposed SEAL-RF based SIMON for functional verification

5.1 Functional Verification of SEAL-RF-based SIMON

The SIMON encryption core was designed with the proposed SEAL-RF logic, in 65nm using Cadence Virtuoso, with a bipolar power supply of 1.2 V and RFID clock frequency of 13.56 MHz. The techniques of merged blocks and balanced transfer paths discussed in Chapter 3, were also adopted for the SEAL-RF based SIMON, in order to ensure the proper synchronization.

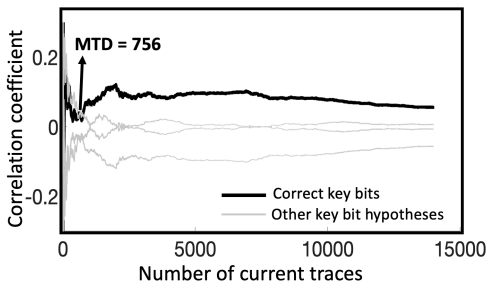
Before proceeding to the analysis of the resistance to CPA attack, the functionality of the SEAL-RF SIMON is verified with the output of the ECRL and static CMOS based SIMON for the same plaintext, with a key value `16'h 1918 1110 0908 0100`.

The ciphertext output for all the three implementations is shown in the Fig. 5.1. It can be seen that all the three implementations produce the same output bits, thus verifying the correctness in the logical operation of SEAL-RF based SIMON.

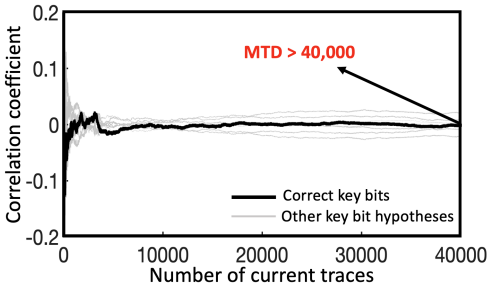
It can also be observed from the figure that the output is charged from '0' for some bits and from V_t for the remaining bits for the ECRL based SIMON. However, the outputs for the SEAL-RF based SIMON is consistently charged from '0', hence reducing the data dependant switching power and increasing the resistance to CPA attacks, as will be shown in Section 5.2.

5.2 Results of CPA Attack on SEAL-RF SIMON

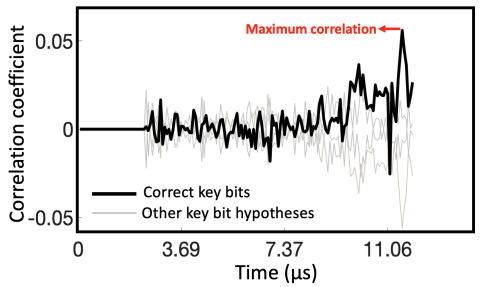
CPA side-channel attacks use Pearson's correlation coefficient to extract sensitive information such as the secret key of an encryption core (24; 114). Specifically, the secret key bits are retrieved by statistically correlating the actual power consumption or supply current traces of a hardware, measured for various random input plaintexts, and, a power model constructed for the same random plaintexts and different key-bit hypotheses. In this work, a CPA attack is mounted on the SEAL-RF based SIMON implementation based on the methodology described in (115). The supply current traces for 40,000 random input plaintexts (each 32 bit) were obtained. A Hamming distance power model was constructed for different key-bit hypotheses and for the same random plaintexts. The results of the CPA attack



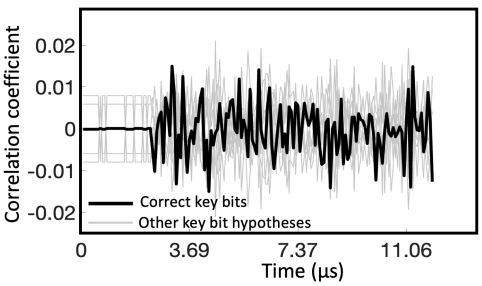
(a) Correlation vs. number of power traces for unprotected ECRL based SIMON core with MTD = 756.



(b) Correlation vs. number of power traces for the proposed SEAL-RF based SIMON core with MTD $>$ 40,000.



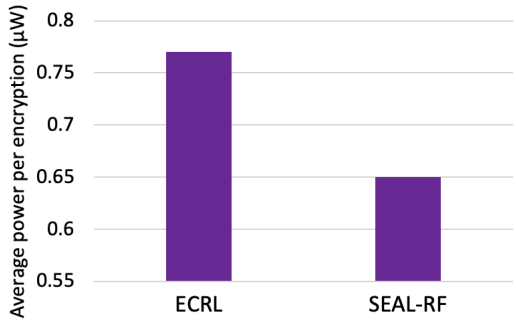
(c) Correlation vs. time (in cycles) for unprotected ECRL based SIMON core.



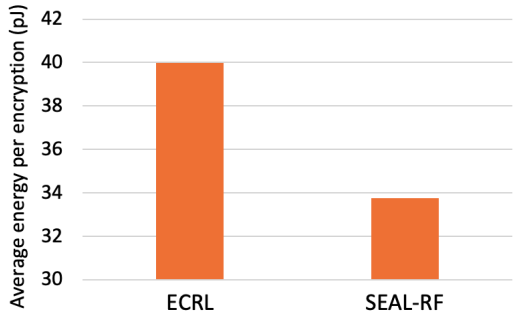
(d) Correlation vs. time (in cycles) for the proposed SEAL-RF based SIMON core.

Figure 5.2: Correlation power analysis (CPA) based power side-channel attack results.

are shown for both unprotected ECRL based SIMON and the proposed SEAL-RF based SIMON core in Fig. 5.2. The correct key bits are highlighted in black and the other key bit hypotheses are shown in grey. Figs. 5.2 (a) and (b) show the correlation coefficient with respect to the number of current traces for unprotected ECRL based SIMON and proposed SEAL-RF based SIMON, respectively. Fig. 5.2 (a) is shown for 2 key bits (4 hypotheses) out of the total 64 key bits since these bits require the maximum number of current traces to be retrieved, indicated as measurements-to-disclosure (MTD) equal to 756 in the figure. The other 62 key bits could be retreived with lesser number of current traces. Therefore, all of the 64 bits of the secret key were successfully retrieved for the unprotected ECRL based SIMON with a maximum of 756 current traces. Fig. 5.2 (b) is shown for 3 key bits (8 hypotheses) since these bits could not be retreived even with 40,000 input traces, achieving an MTD greater than 40,000. Thus, SEAL-RF achieves at least $52\times$ higher CPA resistance than the unprotected ECRL based SIMON core. This increased resistance offered by SEAL-RF is due to the low correlation between the current consumption and the input switching. This behavior can be distinctly observed in Figs. 5.2(c) and 5.2(d). These figures depict the correlation coefficient with respect to time for both ECRL and SEAL-RF based SIMON cores. For ECRL based SIMON core, the correlation coefficient is maximum at 11.6μs for the correct key bits with a correlation coefficient value of 0.056. However, for SEAL-RF based SIMON core, the correlation coefficient for all of the key hypotheses is similar and varies in the range of -0.02 and 0.02. Therefore, the correct key bits are *masked* from the other key hypotheses for SEAL-RF based SIMON core.



(a)



(b)

Figure 5.3: Power (a) and energy (b) results of the bit-serialized SIMON cipher core implemented in adiabatic ECRL and proposed adiabatic SEAL-RF logic.

5.3 Performance and Energy Analysis of SEAL-RF SIMON

Average power and energy per encryption of bit-serial SIMON architecture designed with conventional adiabatic logic (ECRL) and the proposed SEAL-RF are illustrated in Fig 5.3. According to this bar chart, SEAL-RF does not introduce any power and energy overhead. In fact, the average power and energy dissipated by the SEAL-RF SIMON core is 15.6% lower than the unprotected ECRL SIMON core. This reduction in energy is achieved due to two reasons: (1) unlike ECRL,

the charge at the output capacitances are fully recovered/recycled, (2) additional transistors introduced to enhance side-channel resistance do not draw any current from the supply. Also, note that both ECRL and SEAL-RF based implementations of the SIMON core have the same architecture and therefore the number of clock cycles to complete one encryption is the same for both scenarios. Since they also both operate at the same frequency of 13.56 MHz, the throughput is 616 Kbps for both implementations.

5.4 Summary

The SIMON encryption core is designed with the proposed SEcure Adiabatic Logic for Wirelessly-Powered IoT Devices (SEAL-RF) logic in this chapter. The design is functionally validated to ensure the correctness of operation and the performance characteristics are evaluated. The CPA attack results is presented and compared with the static CMOS and the ECRL based SIMON implementations. It was observed that the SEAL-RF SIMON requires more than 52$\times$ of traces in order to retrieve the secret key when compared to the its unprotected adiabatic counterpart and yet consumes 15.6% lower energy.

Chapter 6

EQUAL: Efficient QUasi Adiabatic Logic for Enhanced Side-Channel Resistance

A novel secure adiabatic logic targeted for AC computing applications was proposed and the security and energy characteristics were evaluated in the last two chapters. In this chapter, another novel secure adiabatic logic called "efficient quasi adiabatic logic" (EQUAL) is proposed that achieves the lowest energy with comparable security characteristics.

The rest of the chapter is organized as follows. Some of the existing secure adiabatic logic families are discussed in Section 6.1. Operation principle of the proposed EQUAL logic is explained in Section 6.2. Simulation results are presented in Section 6.3. Finally, the chapter is concluded in Section 6.4.

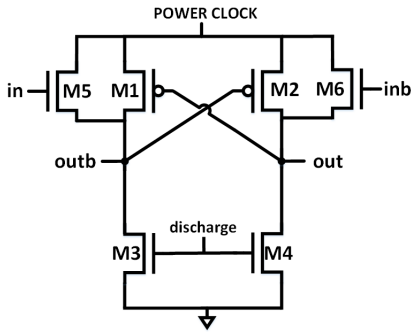


Figure 6.1: Schematic of an adiabatic SPGAL buffer (100).

6.1 Drawbacks of existing secure adiabatic logic for energy efficiency

As discussed in Section 4.2, the secure adiabatic logic families that outperformed earlier approaches (either in terms of security or energy consumption) are the "secure pass-gate adiabatic logic" (SPGAL) (100) and "without charge-sharing quasi-adiabatic logic" (WCS-QuAL) (112), both of which operate with four-phase power supply signal. Thus, this section focuses on these two highly relevant prior work.

The schematic of an SPGAL buffer is illustrated in Fig. 6.1. A four-phase discharge signal is used to discharge the output load capacitance before each evaluation phase in order to reduce the dependence of supply current on input data. The major advantage of this logic is that the non-adiabatic energy loss is prevented during the evaluate phase, thus enhancing overall energy efficiency while also achieving better security characteristics. A practical limitation of this approach is the requirement for four-phase discharge signals, which need to be distributed throughout the chip. This issue exacerbates the already challenging task of generating and distributing

the four-phase power-clock signals in adiabatic logic (116).

Recently, this issue was mitigated by the WCS-QuAL adiabatic logic family (112), where the weaker dependence of current consumption on input was achieved without employing external discharge or charge-sharing signals.

The schematics of WCS-QuAL based buffer and NAND/AND gates are shown in Fig. 6.2. A dual-duplicate evaluation network is employed to mitigate the issue of additional input pins required for the *discharge* signal. Since one of the duplicate evaluation networks conducts during the discharge phase (when the power-clock signal is zero), the output nodes are automatically discharged before the beginning of each evaluation phase, without any additional discharge logic. Furthermore, there are no non-adiabatic losses during the evaluate phase of the power-clock signal, unlike a majority of the existing secure adiabatic logic families. However, the number of transistors (and hence the area of two input logic gates) in WCS-QuAL is significantly higher. For example, in the NAND/AND implementation, in order to balance the capacitance at the output node for all of the input transitions, a symmetric logic implementation is connected at the outputs to ensure equal *RC* delays for all of the input combinations, resulting in 20 transistors per logic gate. Although preventing the non-adiabatic losses results in a reduction in the energy consumption when compared to other existing secure adiabatic gates, this large increase in the number of transistors is a significant drawback for resource-constrained applications with small form factors.

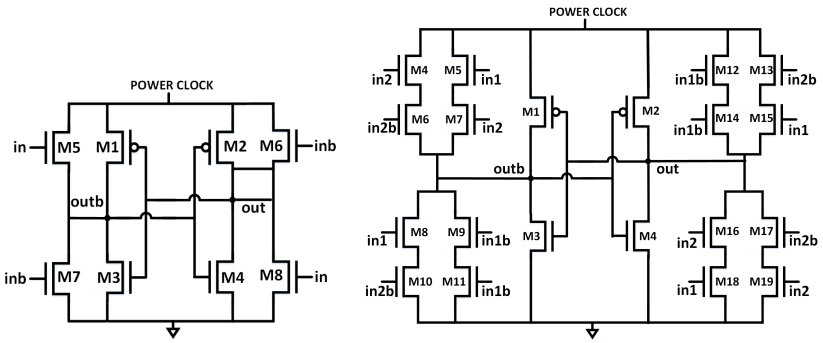


Figure 6.2: Schematic of an adiabatic WCS-QuAL (112): (a) buffer, (b) NAND/AND gate.

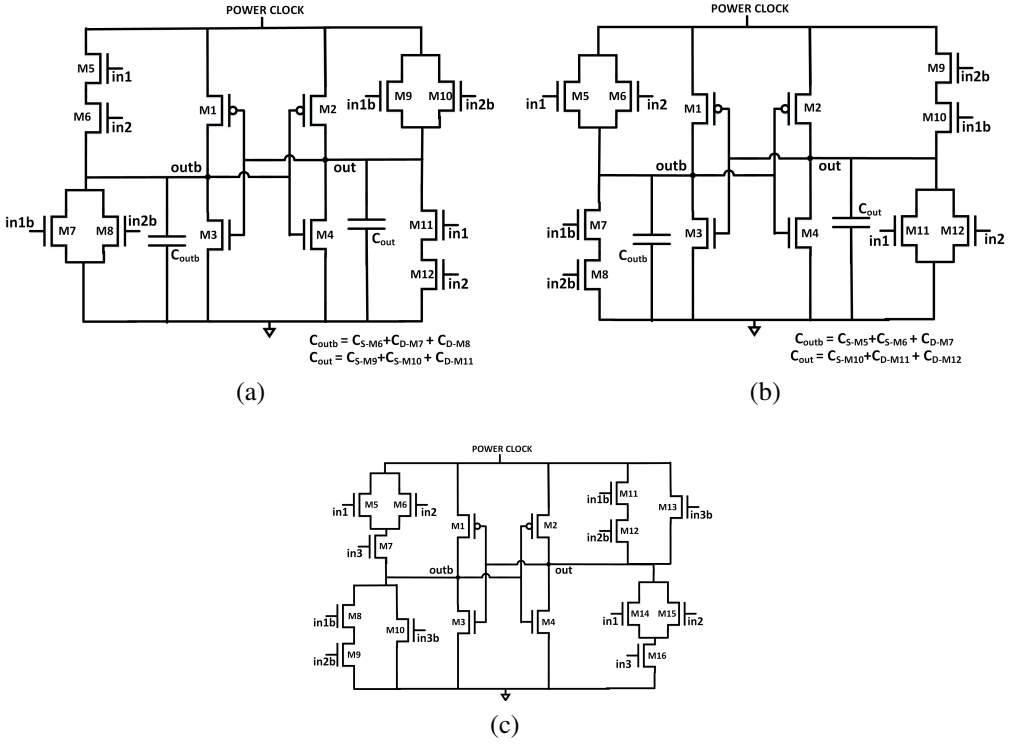


Figure 6.3: Transistor-level schematics of the proposed secure adiabatic logic (EQUAL): (a) NAND/AND gate, (b) NOR/OR gate, and (c) an example OAI3 gate.

6.2 Proposed EQUAL Logic

An enhancement of the WCS-QuAL logic family referred to as "Efficient QUasi Adiabatic Logic" (EQUAL), is proposed in this work. The schematics of the EQUAL NAND and NOR gates are depicted, respectively, in Figs. 6.3(a) and (b). An example OAI3 gate is also depicted in Fig. 6.3(c). In this logic, the evaluation network consists of a dual-complimentary evaluation logic, E1 to E4, where, E1/E3 and E2/E4 are identical. Therefore, the capacitance at the output nodes is equal to $C_{out} = C_{S-M9} + C_{S-M10} + C_{D-M11}$ and $C_{outb} = C_{S-M6} + C_{D-M7} + C_{D-M8}$, where $C_{S/D-Mx}$ is the capacitance at the source or drain terminals of transistor M_x. Since all of the transistors have the same size, the capacitance at the source/drain terminals of all of the nMOS transistors is equal (neglecting the dependence of junction capacitance on terminal voltages). Thus, the output capacitance is balanced for any input combination as indicated in the figure, thereby improving the resistance to power side-channel attacks. Note that the proposed enhancement is for the logic gates that have asymmetric implementations of the complementary logic (such as NAND/AND and NOR/OR gates). The schematic of EQUAL based logic gates that have symmetric implementations of the complementary operations (such as inverter and XOR/XNOR) is identical to the WCS-QuAL counterpart. The four-phase operation of the EQUAL NAND/AND gate shown in Fig. 6.3(a) is described below:

- *Discharge*: The inputs ($in1$,$in2$) rise and the power-clock signal *PC* is at ground potential for a short period of time (for a sinusoidal power-clock signal). Thus, the output nodes are discharged through $E1$ and $E3$ or $E2$ and $E4$, depending upon the input combinations.

- *Evaluate*: Inputs are stable and power-clock signal *PC* rises and *out* follows

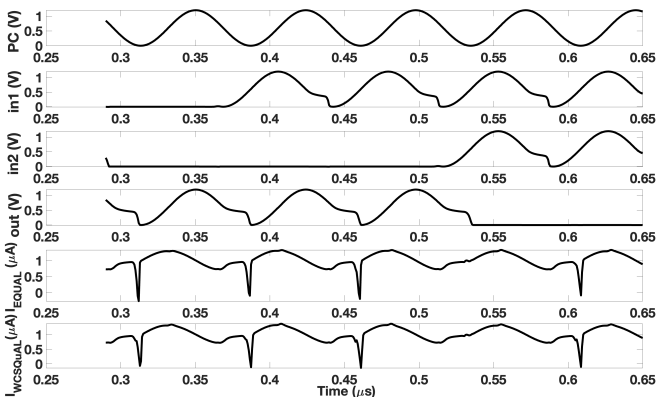


Figure 6.4: Waveforms of EQUAL NAND/AND gate, illustrating (from top) power-clock signal *PC*, input signals *in*1 and *in*2, output signal, and power supply current of EQUAL and WCS-QuAL.

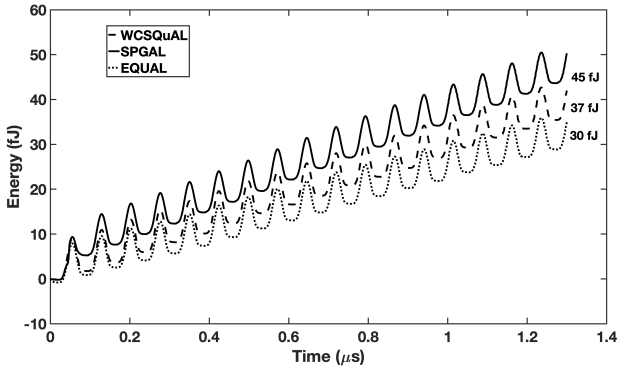
PC through $E1$ (assuming $E1$ and $E3$ are conducting). When the power-clock signal *PC* crosses the threshold voltage, $M1$ starts to conduct until *PC* reaches V_{DD}. Since *out* continuously follows power-clock signal through $E1$ or $M1$, the non-adiabatic losses are mitigated during the evaluation phase.

- *Hold*: Inputs start to fall and the outputs are held stable by the cross-coupled latch for the evaluation of the successive gate .

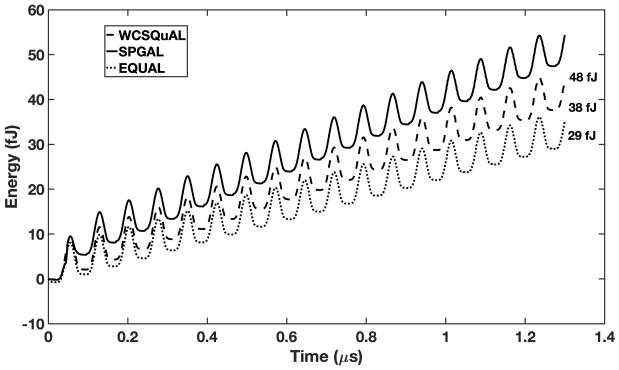
- *Recover*: Power-clock signal *PC* starts to fall and *out* continues to follow *PC* through $M1$, until *PC* reaches the threshold voltage of $M1$.

Although the dependence of power supply current on input data is significantly weakened, the dependence still exists in a small timing window between the *recovery* and *discharge* phases, when the input is still less than the threshold voltage of the evaluating transistors, resulting in difference in the charge at the output nodes, as identified in (112).

The input, output, and power supply current waveforms of an EQUAL NAND/AND



(a)



(b)

Figure 6.5: Comparison of the energy consumption of secure EQUAL (proposed) with energy consumption of adiabatic SPGAL and WCS-QuAL for all possible input transitions: (a) NAND/AND gates and (b) NOR/OR gates.

gate are depicted in Fig. 6.4. As observed in this figure, despite the difference in the resistance of the conduction paths within the evaluation networks E1-E4 (with respect to the inputs), the difference in the current waveforms of EQUAL and WCS-QuAL NAND/AND gates is negligible. Thus, EQUAL has the potential to maintain the security characteristics of WCS-QuAL while significantly reducing the area overhead and energy dissipation,a s quantified in the following section.

Adiabatic logic family	Average energy per transition (fJ)			Number of transistors		
	NAND/AND	NOR/OR	OAI3	NAND/AND	NOR/OR	OAI3
Unprotected ECRL (111)	1.82	2	2.41	6	6	8
Unprotected PAL (31)	1.98	2.01	3.09	6	6	8
WCS-QuAL (112)	2.1	2.22	2.21	20	20	20
SPGAL (100)	2.4	2.63	2.32	12	12	12
EQUAL (proposed)	**1.75**	**1.75**	**2.11**	**12**	**12**	**16**

Table 6.1: Comparison of Average energy per transition and number of transistors of the proposed adiabatic logic family (EQUAL) with several other adiabatic logic families.

6.3 Simulation Results

The proposed secure adiabatic logic gates were designed and simulated using a commercial 65 nm technology node and a sinusoidal power-clock signal with an amplitude of 1.2 V. The frequency of operation is 13.56 MHz, targeting RFID based applications and wireless sensor nodes.

The commonly used security metrics, normalized energy deviation (NED) and normalized standard deviation (NSD), are used to quantify and compare the power side-channel attack resistance of EQUAL logic gates with the existing approaches. NED is given by

$$\text{NED} = \frac{E_{max} - E_{min}}{E_{max}},$$

where E_{max} and E_{min} are, respectively, the maximum and minimum energy consumed over 16 input transitions for a 2-input gate and 4 input transitions for a single

input gate. NSD is determined by

$$\text{NSD} = \frac{\sigma}{E_{avg}},$$

where the standard deviation σ is

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(E_i - E_{avg})^2}{N}},$$

where E_i is the energy consumption that corresponds to input transition i and E_{avg} is the average energy consumption for N input transitions. The goal of both metrics is to evaluate the sensitivity of energy consumption to different input transitions to determine the level of power-based side-channel attack resistance offered by a logic gate. Therefore, lower NED and NSD signify higher resistance to the attacks.

For a 2-input logic gate, there are overall 16 possible input transitions. The comparison of transient energy consumption of the proposed EQUAL NAND/AND gate with other existing adiabatic logic families is shown in Fig. 6.5 as the input signals vary to cover all of the 16 possible transitions. According to this figure, the overall energy consumption of the proposed NAND/AND gate is 30 fJ, which is 30% lower than SPGAL and 18% lower than WCS-QuAL. The total energy consumption of the NOR/OR gate is 29 fJ, which is 39% lower than SPGAL and 24% lower than WCS-QuAL. Note that for SPGAL, since the discharge phase of the sinusoidal power-clock signal is very small, the switching power contributed by the discharge input signal is non-negligible.

The average energy per transition and the number of transistors required for NOR/OR, NAND/AND and a complex 3-input OAI3 logic gate are listed in Ta-

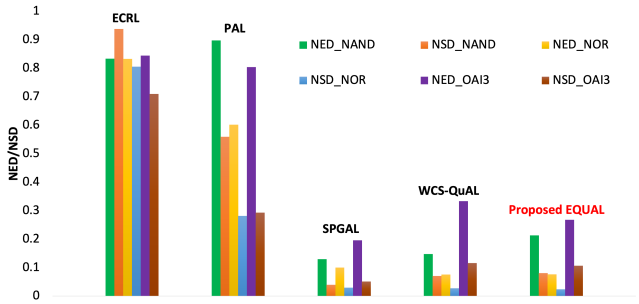


Figure 6.6: Comparison of the security metrics NED and NSD of the secure EQUAL NAND/AND, NOR/OR and OAI3 gates with the NED and NSD of adiabatic ECRL, PAL, SQAL, SPGAL and WCS-QuAL NAND/AND and NOR/OR gates for all possible input transitions.

ble 6.1 for the proposed and existing approaches. The proposed EQUAL NAND/AND and NOR/OR consume the lowest energy of 1.75 fJ, while having 40% less number of transistors than the WCS-QuAL counterpart. The proposed complex OAI3 gate also consumes the lowest energy of 2.11 fJ and has 20% less number of transistors than WCS-QuAL based OAI3. The energy consumption of the proposed EQUAL gates is even lower than the unprotected adiabatic ECRL and PAL (due to mitigating non-adiabatic loss) at the expense of an increase in the number of transistors.

The NED and NSD security metrics achieved by the proposed EQUAL NAND/AND and NOR/OR gates and a complex 3-input OAI3 gate are compared with the existing secure adiabatic families in Fig. 6.6. Note that for the complex gate, these metrics are computed for all of the 64 possible input transition combinations (6 transition levels for 3 inputs, producing a total of 2^6 possible transition combinations). According to the bar plots shown in the figure, although the area/transistor count and energy consumption of the proposed logic are significantly lower than SPGAL and WCS-QuAL based gates, NED and NSD are only marginally degraded (by a maximum of 8% and 4%). Furthermore, for the NOR/OR implementation, NED

of the proposed approach is 2% lower than SPGAL and the same as WCS-QuAL whereas NSD is 1% lower than both approaches. For the complex OAI3 gate, both NED and NSD are degraded by 7% as compared to SPGAL and lower than WCS-QuAL by 6% and 1%, respectively. This marginal degradation in the security metrics for the proposed EQUAL logic implementation is due to the imbalance in the resistances among the evaluation networks E1 to E4, as explained in Section 6.2.

As mentioned in Section 6.2, there is a small timing window during the discharge phase when the supply current is input dependent. Since the resistance of the path of this discharge current is lower for NOR/OR gate (as compared to NAND/AND gate), the dependence of supply current on input has a diminishing effect, thereby resulting in better security characteristics for EQUAL NOR/OR gate.

6.4 Summary

A novel energy and area efficient secure adiabatic logic family, referred to as "Efficient QUasi Adiabatic Logic" (EQUAL), was proposed for resource-constrained IoT applications. The average energy consumption per switching of the proposed logic is 34% lower than SPGAL and 21% lower than the WCS-QuAL secure adiabatic logic families. The number of transistors is equal to the SPGAL, whereas 40% lower than the WCS-QuAL NAND/AND and NOR/OR implementations. The NED and NSD security metrics were compared with the existing adiabatic logic families to evaluate power-based side-channel attack resistance. The increase in the energy efficiency is achieved at the expense of an average increase of 3% in the NED and 1.5% increase in the NSD for EQUAL based NAND/AND and NOR/OR logic gates.

Chapter 7

High Bandwidth Thermal Covert Channel in 3D-Integrated Multicore Processors

Covert-channel attacks are another source of information leakage in emerging hardware applications similar to side-channel attacks. Specifically, covert-channel communication using heat, are a tremendous threat to modern multicore processors and is the focus of this and the next chapters of this thesis. A thermal covert-channel (TCC) attack is established in a multicore processor by encoding secret data bits on the temperature profile of a processor core. To encode a bit '1', a program is executed to raise the temperature of the core and to encode a bit '0', the program execution is stopped to cool it down. In existing works, a computation intensive program such as a cpu stress-test is used for the encoding. Such covert channels with high-power programs are typically easier to detect as they cause significant rise

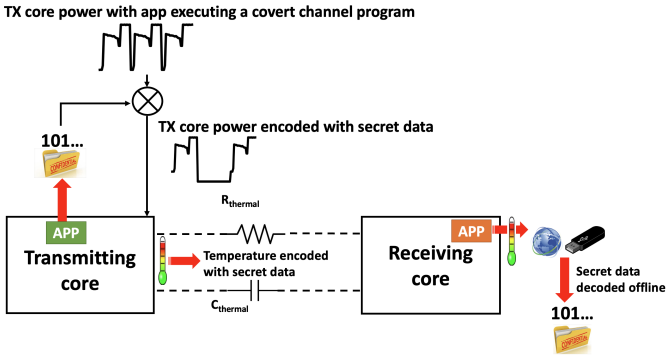


Figure 7.1: Attack model of TCC between two cores of a multicore processor.

in temperature. In this work, we demonstrate that by leveraging vertical integration, it is sufficient to execute typical SPLASH-2/PARSEC benchmark applications to establish a high bandwidth thermal covert channels.

This chapter is organized as follows. The attack methodology for TCC analysis, characterization framework, design models, and covert communication protocol (encoding and decoding) are detailed in Section 7.1. The extensive results for both 2D and 3D systems are described in Section 7.2. Finally, the chapter is summarized in Section 7.3.

7.1 Methodology

TCC attack model and analysis framework are described in this section. The processor architecture, 3D floorplans, layer stack models, covert channel application and communication protocol are also detailed.

7.1.1 Attack model

In this work, we consider that TCC is established between two physical cores of a multicore processor that execute compromised software applications (henceforth, referred to as *apps*) concurrently, as shown in Fig. 7.1. Let us assume that the app executed on the transmitting core has access to confidential information. Some examples include a contact app on a mobile phone that has access to private list of contacts or personal finance management apps with access to confidential monetary data. In modern multicore processors, data packets from these secure applications can be protected by special enclaves using technologies such as Intel Software Guard Extensions (SGX) (117) and Arm TrustZone (118). These technologies prevent sensitive data managed by these apps from being accessed by outside world. However, thermal coupling between the physical cores can be leveraged to leak sensitive data by bypassing these security measures (13; 17). In order to achieve this covert channel, the transmitting app controls the execution of a program to raise and lower the power consumption of the transmitting core. A sample power profile of the transmitting core executing a covert channel program is shown in Fig. 7.1. Consequently, temperature of the transmitting core is encoded with the sensitive information and is coupled to the receiving core through the thermal resistance ($R_{thermal}$) and capacitance ($C_{thermal}$) of the medium, as shown in the figure.

We assume that the app on the receiving core is not security enforced and hence does not have direct access to any confidential information. However, for this app to read the temperature profile of the receiving core, we assume that it has access to the temperature sensor of the core. This assumption is based on commonly used thermal management policies, where the user-installed apps can access temperature sensor data without special permissions (14). The app either decodes the data

bits on the receiving core or sends the temperature data for offline decoding, as illustrated in Fig. 7.1.

7.1.2 TCC analysis framework

The simulation framework of TCC for both 2D and 3D ICs is depicted in Fig. 7.2. Each step of this flowchart is described in this section.

7.1.2.1 Processor architecture

Our target system is a 4-core CPU based on Intel Haswell architecture (119). The 22 nm processor operates at 3.4 GHz frequency with a supply voltage of 1.2 V. The specific architectural configurations, such as performance models of each core, L1, L2 and L3 caches, translation buffer and reorder buffers are adapted from published data for the processor (120; 121), as listed in Table 7.1. The workloads used for TCC are simulated on this architecture using SNIPER (122), which is an interval based timing simulator designed specifically for multicore Intel processors. The transient power traces of the covert channel applications are obtained via the multicore power simulator, McPAT (123), which is integrated within SNIPER. The power consumption obtained from McPAT is calibrated with the real power measurements of a similar processor architecture (124).

7.1.2.2 Encoding the secret data

The transmitting core encodes the secret data bits in its transient temperature profile before communicating to the receiver core. In this work, the encoding is performed using the NRZ encoding technique. To transmit a bit '1', the compro-

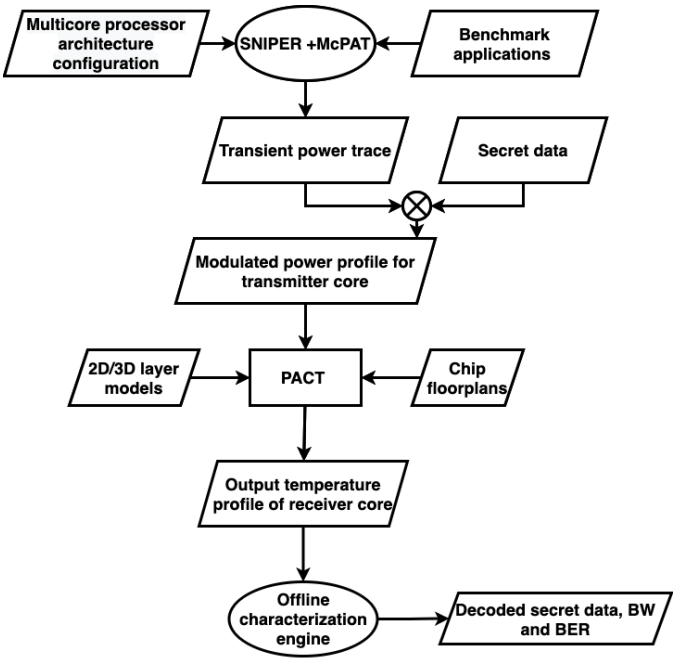


Figure 7.2: Flowchart showing the simulation framework for characterizing TCC in 2D and 3D multicore processors.

Instruction set architecture	x86-64
Clock frequency	3.4 GHz
Technology node	22 nm
Supply voltage	1.2 V
Issue width	4
Reorder buffer entries	192 entries
TLB entries	ITLB: 128, DTLB: 64, STLB: 1024
L1 cache	32 KB 8-way set associative L1I cache and L1D cache
L2 cache	256 KB 8-way set associative
L3 cache	8 MB shared

Table 7.1: Architectural configurations of 4-core Haswell processor.

mised app on the transmitting core continuously executes a program to increase the power consumption (and, hence, the temperature) of that core until the next bit to be transmitted is '0'. Similarly to transmit bit '0', the app stops the program execution and remains idle such that the temperature of the core decreases. The pseudo-code of this encoding process is shown in Algorithm 1. We define the bit-width in the pseudo-code as the duration during which a bit '1' or '0' is transmitted.

Algorithm 1 Generation of modulated power trace encoded with secret data

```
number_of_ones ← 0
next_bit:
for bit in secret data do
  if bit == 1 then
    if number_of_ones == 0 then
      run program for bit-width duration
      update power trace
    else
      continue program execution for bit-width duration
      update power trace
    end if
    number_of_ones ← number_of_ones + 1
    goto nextbit
  else
    number_of_ones ← 0
    stop program if executing
    update power trace
    goto next_bit
  end if
end for
```

To transmit bit '1', we execute programs from the common SPLASH-2 and PARSEC benchmark application suites (125). The applications *freqmine*, *ferret* and *blackscholes* are from PARSEC and the other applications are from SPLASH-2. The total simulated power (obtained from McPAT) consumed by the Haswell

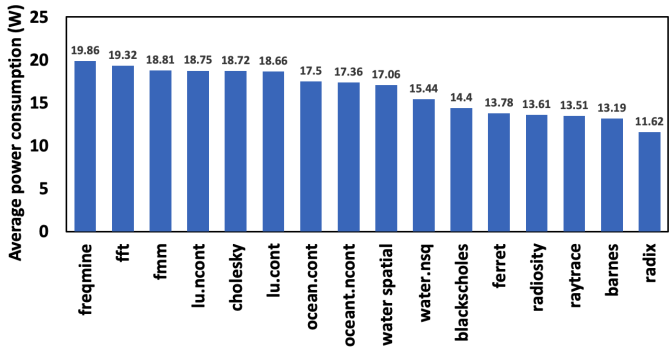


Figure 7.3: Average power consumption of Haswell 4-core processor running different applications from SPLASH-2 and PARSEC benchmark suites.

processor executing these single-threaded applications ranges from 11 to 20 W, as illustrated in Fig. 7.3. From this figure, it can be observed that *FFT* from SPLASH-2 and *freqmine* from PARSEC consume the highest power and, therefore, are capable of producing relatively significant variations in the temperature profile of the transmitting core. Therefore, *FFT* is chosen as the target application program and is continuously executed in a single-threaded fashion for a bit '1', until the following bit is '0'. The transient power trace of the transmitting core executing the *FFT* program is obtained from McPAT with a time-step of 0.5 ms.

To synchronize the start of the communication, the app in the transmitting core prefixes the beginning of every secret block of data with preamble bits. The preamble consists of a sequence of alternate bits of ones and zeros. This pattern is used since it ensures a symmetric and well-correlated temperature profile between the two communicating cores (13). Sample secret bits prefixed with preamble bits are plotted in Fig. 7.4 along with the modulated transmitting core power profile generated based on Algorithm 1.

In previous works, other encoding techniques, such as return-to-zero amplitude-shift keying (ASK) or Manchester encoding have been exploited, even though the

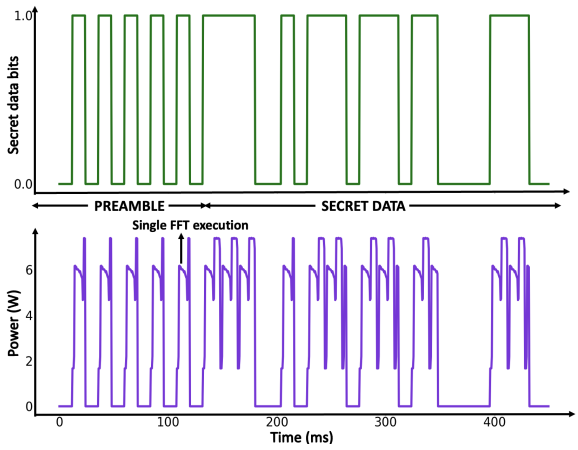
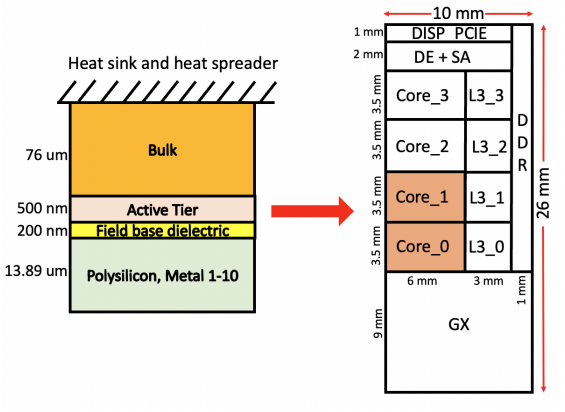


Figure 7.4: Sample secret data bit-stream and the corresponding TCC application power profile of the transmitting core.

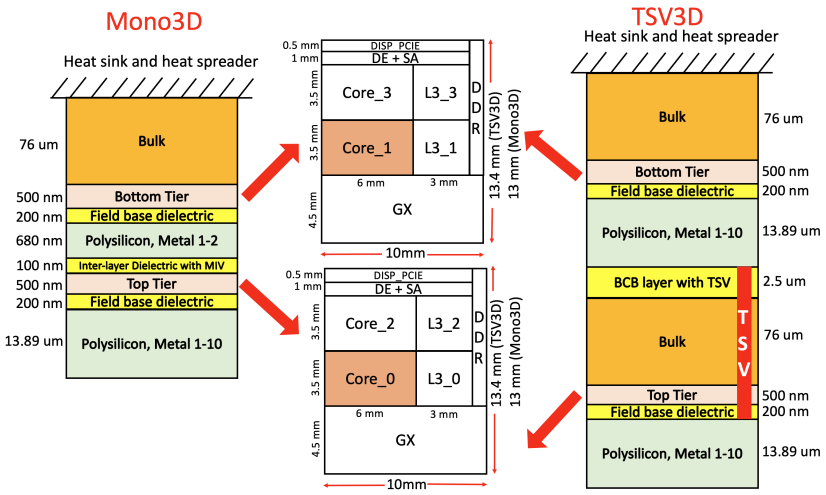
communication bandwidth is reduced compared to the NRZ encoding technique (14; 20). The primary reason for using these alternative encoding schemes is, when encoding a continuous stream of bit '1's in NRZ technique, the CPU stress test has to be executed continuously. Thus, the temperature of the transmitting core can increase excessively, leading to overheating issues. However, this problem is mitigated in our work since we execute a normal *FFT* program that does not consume such high power and, hence, does not cause overheating even when transmitting continuous '1's. Therefore, the need for alternative encoding schemes that reduce the communication bandwidth is eliminated.

7.1.2.3 Thermal covert communication analysis

To characterize the impact of 3D integration on thermal covert channel attacks, the modulated power trace of the transmitting core encoded with the secret message is given as an input to a thermal simulator. We perform the simulations in PACT, a modern parallel thermal simulator capable of efficiently handling multi-layered



(a)



(b)

Figure 7.5: Cross-sectional layers and active layer floorplan for Haswell processor integrated in: (a) 2D and (b) Mono3D and TSV3D technologies.

chips with fine-granularity of layer thickness (126). Yuan *et al.* have validated the transient simulations of PACT and have demonstrated it to be 186× faster than the well-known thermal simulator HotSpot (126), while exhibiting a maximum error of 2.77% for steady-state and 3.28% for transient thermal simulations compared to COMSOL, a state-of-the-art finite-element method (FEM) based simulator (127). We model the cross-section of 2D, Mono3D and TSV3D chip, as shown in Fig. 7.5. The 2D chip floorplan for the quad-core processor is adapted from the published work on the Intel Haswell processor (119). After analyzing the existing types of tier partitioning strategies for 3D multicore processors (64; 128; 129), the 2D floorplan in this work is partitioned into two tiers for both Mono3D and TSV3D systems where each tier has two cores, as shown in Fig. 7.5(b). We characterize the bandwidth and bit-error rates of TCC among the cores of a flip-chip two-tier Mono3D and TSV3D (face-to-back bonding) integrated processor (58). The TCC transmitting core ($CORE\_0$) and receiver core ($CORE\_1$) are highlighted in both floorplans in the figure. For TSV based 3D IC, TSVs are modeled as a 250×5 array with a diameter of 10 μm (60; 130; 131; 132) and center-to-center pitch of 40 μm (21). TSVs cross field-base dielectric layer, top-tier, adhesive benzocylcobutene (BCB) layer and the bulk layers, as shown in the figure. Alternatively, for Mono3D, the monolithic inter-tier via (MIV) is modeled with a diameter of 50 nm and center-to-center pitch of 170 nm on the inter-layer dielectric (ILD) layer (133; 134). For our thermal simulations, we set the same grid size for 2D, TSV3D and Mono3D and we use the default steady-state and transient solver options in PACT. The heat sink for all of the systems is modeled as a conventional pin fin heat sink with a heat spreader and a fan to mimic practical cooling mechanisms in processors (126; 135). The heat transfer coefficient is assumed the same for the three systems to ensure

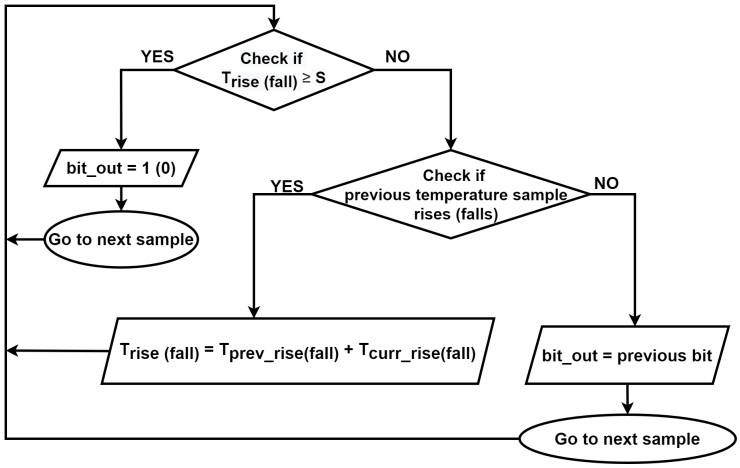
that the same amount of heat is removed by the heat sink per unit area.

A linear model for the temperature dependant leakage power for the target processor is adapted and scaled based on the published data for Intel 22 nm processors (136). We derive the leakage power model for every core of the Intel Haswell processor as $P_{leak} = 0.0137 \times T - 0.055$, where T is the temperature profile of the core from PACT in °C. The thermal simulations are typically performed in multiple iterations until the temperature variation is within 1°C. In our experiments, we obtain this convergence in two iterations. The final output temperature of the receiver core is analyzed to characterize the covert communication channel and to decode the secret data.

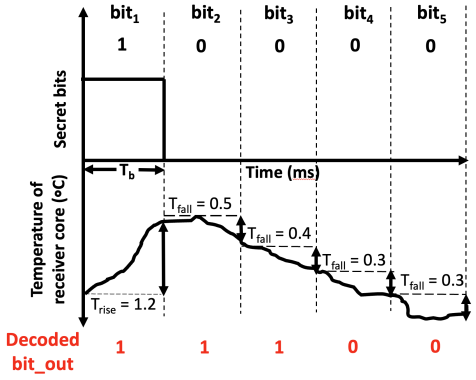
7.1.2.4 Decoding the secret data

The secret data is embedded in the temperature profile of the transmitting core and is communicated to the receiving core through the thermal coupling between them. Since both of the cores are compromised by the attacker, the time of communication, the bit-width, preamble bits are agreed upon as part of the communication protocol. Therefore, the receiving core records the temperature sensor information as soon as the transmitting core starts the encoding process. These data bits can either be decoded within the receiving core or sent offline for remote decoding since the receiving core has access to the network, as described in Section 7.1.1. The decoding process used in this work is explained through the flowchart in Fig. 7.6(a) and the example in Fig. 7.6(b).

The transient temperature profile of the receiving core is sampled at the end of every bit-width. The temperature rise (T_{rise}) or fall (T_{fall}) at every sample is recorded. The minimum difference in the temperature detectable by the sensor,



(a)



(b)

Figure 7.6: Decoding process: (a) algorithm explained using flowchart, (b) example, where sensor resolution, S = 1°C and $T_{rise(fall)}$ is the rise (fall) time in the temperature of the receiver core.

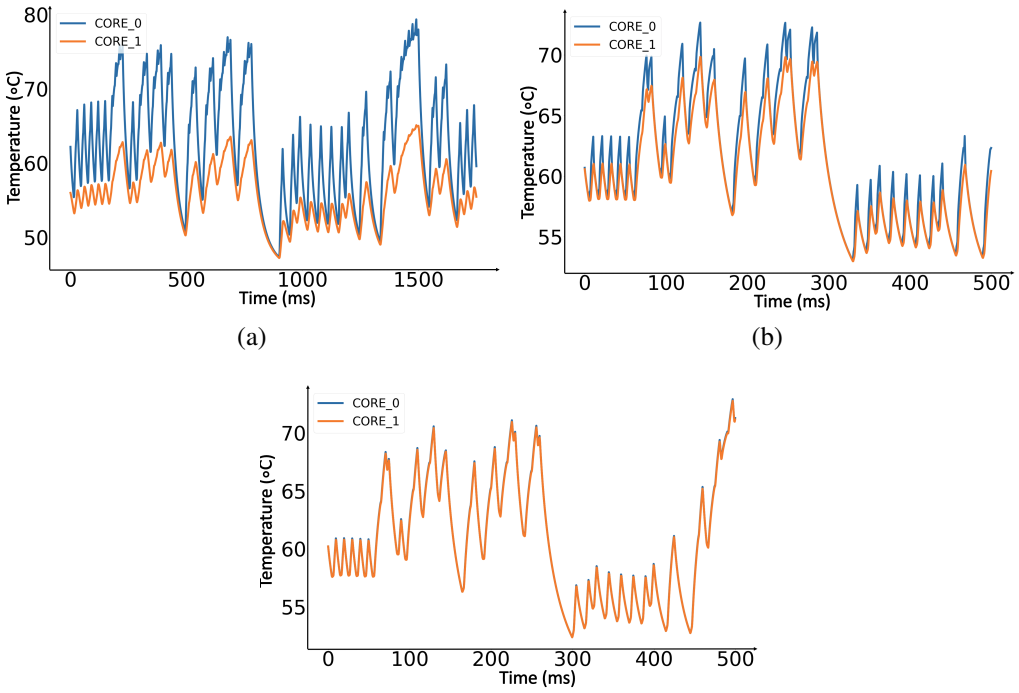


Figure 7.7: Transient temperature profiles of the transmitting core (*CORE*_0) and receiver core (*CORE*_1) of the Haswell processor for transmitting one block of secret data for (a) 2D for bit-width = 17 ms, (b) TSV3D for bit-width = 5.5 ms and (c) Mono3D for bit-width = 5 ms. The time-scale is different for (a) since the *FFT* application is executed for longer durations in 2D to obtain reduced error rates.

also called the sensor resolution, is referred to as S. Since most modern processors have temperature sensors with a resolution of approximately 1°C (137; 138), S is assumed to be 1°C in this work. In Fig. 7.6(b), the first secret bit, bit_1, is '1' and the corresponding rise in the receiver core temperature is 1.2°C. Based on the algorithm in Fig. 7.6(a), since this temperature rise is greater than $S = 1°C$, decoded output bit is also '1'. However, for the following two bits (bit_2 and bit_3), the decoded output bit remains as '1', since the cumulative fall in temperature is less than S. For bit_4, cumulative fall $T_{fall} = 0.5 + 0.4 + 0.3 = 1.2°C$ and is greater than S and therefore the output bit is '0'.

Based on the detected *bit_out*, the error rate of communication is characterized as

$$BER = \frac{n_{corr}}{N} \times 100, \tag{7.1}$$

where BER is the bit-error rate (in %), n_{corr} is the number of correct bits in *bit_out* that matches the transmitted bits and N is the total number of bits received. In the example discussed above, $n_{corr} = 3$ and $N = 5$ and, hence, $BER = 60\%$. The transient temperature simulations are performed extensively for various transmission rates ($\frac{1}{bit\_width}$). The effective bandwidth of the communication channel is estimated as the highest bitrate that can yield less than 1% BER. Note that this assumption for BER is based on previous works that show similar or higher TCC error rates (13; 14).

7.2 TCC Simulation Results

The experiments are performed for 2D, Mono3D and TSV3D based Haswell processor by encoding 10 blocks of secret data into the transient power profile of

the transmitting core (Fig. 7.4). Each block comprises 10 bits of preamble and 100 random bits of data and the same set of random bits are encoded in 2D, Mono3D and TSV3D systems to ensure a fair comparison. The covert channel application is executed on the transmitting core, *CORE*_0, located in the active tier of the 2D system and the top tier of the Mono3D and TSV3D based systems. The output transient temperature of the receiver core from the PACT thermal simulator is used to characterize the TCC bandwidth. The characterization of TCC is performed for six different scenarios: (1) without thermal interference from other cores (noise-free) to isolate the effect of thermal coupling between the transmitting and receiving cores on TCC, (2) with thermal interference to study the effect of other active cores on TCC bandwidth, (3) with partial and non-overlapping placement of transmitter and receiver cores in 3D systems to analyze the effect of 0% and 50% overlap on TCC, (4) when the transmitting core is placed closer to the heat sink to analyze the effect of changing heat flow on TCC bandwidth, (5) with a lower power TCC program to analyze the effect of transient power variations on TCC, and (6) with a four-tier 3D system to study the effect of more than two tiers on TCC. These results are described in the following subsections.

7.2.1 TCC characterization without thermal interference

In this scenario, the cores other than the transmitter and the receiver are assumed to be in sleep state. The transmitting core (*CORE*_0) and the receiving core (*CORE*_1) are placed adjacent to each other on the 2D system and are placed on top of each other on Mono3D and TSV3D based systems, as depicted in Fig. 7.5. The output transient temperature profiles of all the cores for transmitting one block of data is shown in Fig. 7.7(a) for 2D processor for a bit-width of 17 ms, (b) for

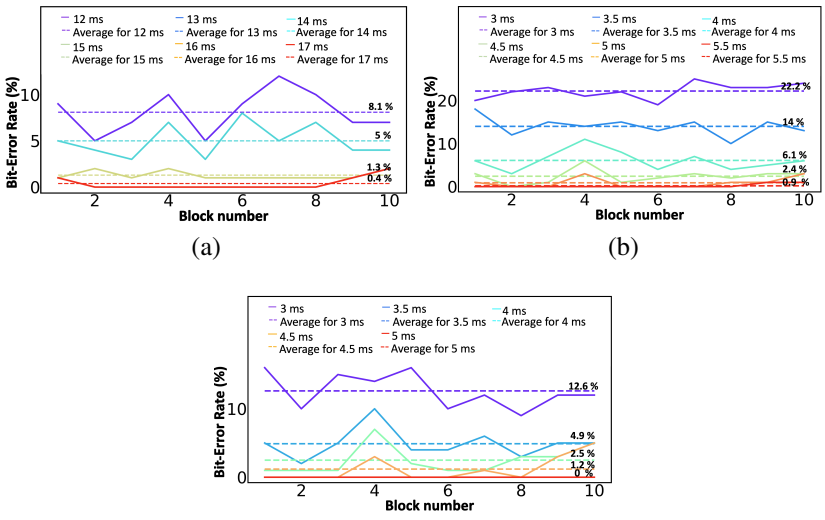


Figure 7.8: Bit-error rate versus number of blocks of secret data for different bit-widths of time for (a) 2D, (b) TSV3D and (c) Mono3D based systems.

TSV3D processor for a bit-width of 5.5 ms and (c) for Mono3D processor for a bit-width of 5 ms. Transmitting ten blocks of secret data at these bit-widths yields a BER of less than 1%.

The differences in the thermal coupling between transmitter and receiver cores for 2D, Mono3D, and TSV3D based systems can be observed from the figure. For the 2D system, a 10°C increase in the *CORE*_0 temperature produces an increase of 2°C in the *CORE*_1 temperature due to the lateral coupling. However, for TSV3D, a 10°C increase in the *CORE*_0 temperature produces an increase of 8°C because of the stronger vertical thermal coupling between the transmitting core located on top tier and receiver core located on the bottom tier. Alternatively, the temperature profile of the transmitter and the receiver cores overlap for the Mono3D system because of the highest inter-tier thermal coupling enabled by the sufficiently small thickness of the ILD layer. Note that the temperature range of the transmitting core in 2D

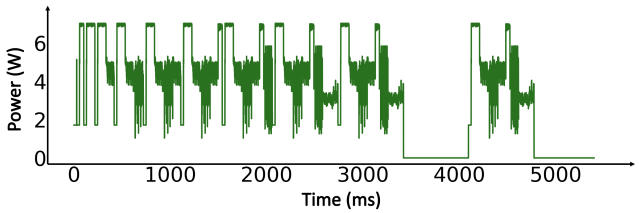


Figure 7.9: Power profile of the noise application executing in cores other than the transmitter and the receiver cores.

Integration technology	With 100% overlap				With 50% overlap				With 0% overlap			
	Without noise		With noise		Without noise		With noise		Without noise		With noise	
	BW (*bps*)	BER (%)	BW (*bps*)	BER (%)	BW (*bps*)	BER (%)	BW (*bps*)	BER (%)	BW (*bps*)	BER (%)	BW (*bps*)	BER (%)
2D	59	<1	52	3	NA	NA	NA	NA	NA	NA	NA	NA
Mono3D	200	<1	200	<1	167	<1	142	<1	77	<1	67	<1
TSV3D	182	<1	182	<1	154	<1	100	<1	77	<1	67	6

Table 7.2: Comparison of TCC bandwidth (BW) for 2D, Mono3D, and TSV3D based multicore processor for three scenarios with and without thermal interference (noise): (1) with 100% vertical overlap, (2) with 50% vertical overlap, and (3) with 0% overlap between transmitting and receiving cores.

IC in Fig. 7.7(a) is higher than the corresponding temperatures for Mono3D and TSV3D systems because the *FFT* application is executed for an extended period of time in the 2D system (bit-width of 17 ms) to ensure sufficiently low *BER*.

The *BER* of the covert communication is characterized for various bit-widths in order to determine the effective bandwidth. The variation of *BER* for different bit-widths for each block of data and the average *BER* of all the blocks are illustrated in Fig. 7.8. As observed from the figure, the error rate varies for each block due to the randomness of the secret data. The bit-widths at which the average *BER* is less than 1%, determines the effective bandwidths for the three systems, as listed in columns 2 and 3 of Table 7.2.

From the table, it can be observed that the bandwidth of a Mono3D based TCC attack is 200 bps and is 3.5$\times$ greater than the bandwidth achieved by a 2D integrated processor. Although Mono3D based systems lead to the highest bandwidth, it is still limited by the transmitting core temperature and the minimum detectable temperature of the sensor on the receiving core. Furthermore, the bandwidth achieved by Mono3D and TSV3D technologies is similar, even though the vertical thermal coupling is stronger in Mono3D. The primary reason for this similarity is the higher thermal resistance between the top tier (where the transmitting core is located) and the heat-sink for the TSV3D based system, as observed from Fig. 7.5. Therefore, the steady-state and the peak-to-peak values of the temperature of the *transmitting core* for TSV3D based system is greater than that of Mono3D. Consequently, even though the inter-tier coupling is lesser in TSV3D, since the transmitting core temperatures are greater, the bandwidth of TSV3D is comparable to Mono3D, making both technologies more vulnerable to a TCC attack than a 2D based many-core processor.

7.2.2 TCC characterization with thermal interference

In Section 7.2.1, we show that a typical benchmark program, such as *FFT* with a nominal power profile, is sufficient to transfer at 200 bps in a 3D many-core processor. However, this bandwidth was achieved in a lightly loaded scenario when all of the other cores are in sleep state. In this section, we investigate the effect of an active core (referred to as noise core) other than the transmitter and receiver cores, on TCC bandwidth.

During TCC between the transmitting and the receiving cores, the noise core sequentially executes random applications from SPLASH-2 benchmark suite, with

dispersed instants of idle time. These noise applications are executed in *CORE*_3 for the 3D systems and in *CORE*_2 for the 2D system because of the proximity of these cores to the receiving core (as shown in Fig. 7.5). The transient power consumption of the noise core executing the noise applications is shown in Fig. 7.9. Thermal simulations in PACT are performed with the noise power trace. TCC error rates are estimated using a similar analysis as in Fig. 7.8 in order to characterize the bandwidth. These results are tabulated in columns 4 and 5 of Table 7.2. As observed from the table, TCC bandwidth in Mono3D and TSV3D remains, respectively, as 200 *bps* and 182 *bps*, which is similar to the results without thermal interference. Therefore, TCC in 3D integrated processors exhibits increased resistance to noise. The reason for this higher robustness can be explained with the help of the transient temperature profiles of the transmitter, receiver and noise cores, as illustrated in Fig. 7.10. First, the variation in the temperature profile of the noise core (*CORE*_3) executing SPLASH-2 applications is sufficiently slower compared to the temperature profile of the transmitting core (*CORE*_0) encoded with the secret data, as observed from Figs. 7.10(b) and 7.10(c). Furthermore, the temperature of the receiver core (*CORE*_1) also has sharp rise and fall times compared to the temperature of the noise core, because of the strong vertical thermal coupling between the transmitter and the receiver cores for Mono3D and TSV3D, as seen in the figures. Since TCC bandwidth depends upon these steep rise and fall times of the receiver core, the 3D integrated processors exhibit lower sensitivity to the thermal interference from the noise core. Note that the noise applications from SPLASH-2 are executed for the total execution time whereas the TCC program is executed only for the duration of the bit-width shown in Fig. 7.10. Therefore, the rise in temperature for the transmitting core (*CORE*_0) and the receiving core (*CORE*_1) is lower

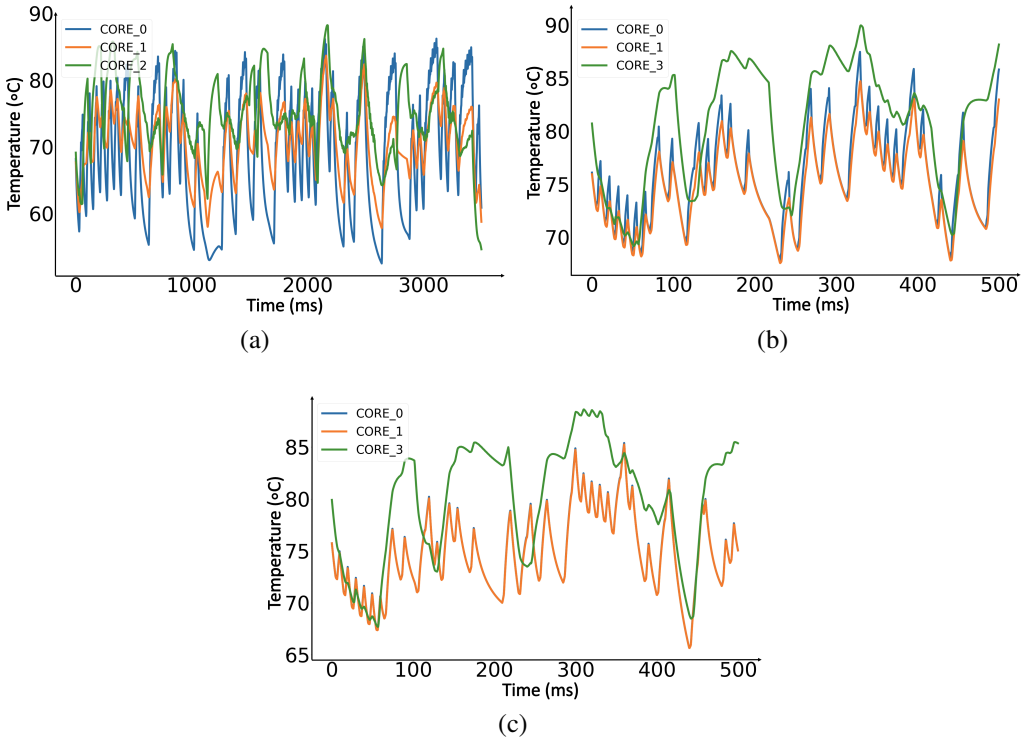


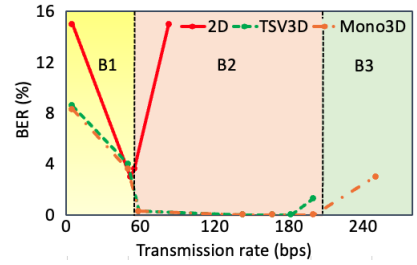
Figure 7.10: Transient temperature profiles of the transmitting core (*CORE*_0), receiver core (*CORE*_1) and noise core of the Haswell processor for transmitting one block of secret data for (a) 2D for a bit-width of 30 ms, (b) TSV3D for a bit-width of 6 ms and (c) Mono3D for a bit-width of 5 ms. The time-scale is different for (a) because, the *FFT* application is executed for longer duration in 2D to obtain reduced error rates.

than the rise in temperature of the noise core (*CORE*_3).

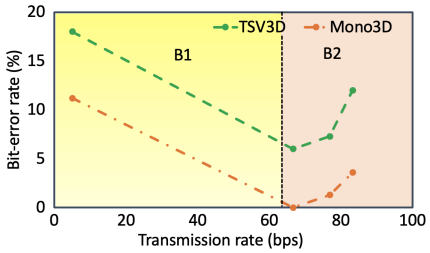
Alternatively, in 2D based systems with thermal interference, the TCC bandwidth degrades to 52 *bps* with *BER* of 3%, as listed in column 5 of Table 7.2. This degradation is due to similar rate of variation in the temperature profile of the noise core (*CORE*_2) and transmitting core for the 2D system, as shown in Fig. 7.10(a). In other words, the transmission rate of the covert channel application overlaps with the rate at which the temperature of the noise core varies.

The results are further described by Fig. 7.11(a) that depicts the variation of the *BER* with TCC transmission rates in 2D and 3D processors in the presence of the noise core. The transmission rates are divided into three regions: B1 (yellow), B2 (pink), and B3 (green). B1 refers to the range of bitrates that overlap with the frequency of the temperature profile of the noise core. B2 refers to the range for which the *BER* is negligible for TSV3D and Mono3D processors. The *BER* in 3D systems starts to increase with increase in the transmission rates in region B3. In B2, *BER* at 83 *bps* is 16% for the 2D processor. When the transmission rate is decreased, *BER* also decreases, as expected. However, the *BER* does not decrease below 3% (at 52 *bps*). Reducing the TCC rate below 52 *bps* interferes with the frequency range of the temperature of noise core in B1 and this interference increases the BER monotonically. However, a TCC attack can be mounted with negligible error rates in 3D systems at transmission rates up to 182 *bps* for TSV3D and 200 *bps* for Mono3D, without interfering with the frequency of the noise core temperature in B1 (as observed in the figure). Thus, for 3D systems, the bandwidth of covert communication with thermal interference is the same as the bandwidth achieved in noiseless scenario, as discussed in Section 7.2.1.

Please note that the execution of the noise application in *CORE_2* (closer to the transmitting core) instead of *CORE_3* does not make a difference in TCC bandwidth results presented for the 3D systems. However, when the noise application is executed in both *CORE_2* and *CORE_3*, the TCC bandwidth of the Mono3D and TSV3D systems are 182 *bps* and 133 *bps*, respectively. Alternatively, for 2D, the maximum TCC bandwidth is 50 *bps* with a *BER* of 7%. Although there is some degradation in the bandwidth for Mono3D and TSV3D systems, the effect of thermal interference is significantly weaker compared to TCC in 2D system.



(a)



(b)

Figure 7.11: BER versus transmission rate for (a) 2D, TSV3D and Mono3D integrated processor in the presence of a noise core (*CORE*_3 for 3D processors and *CORE*_2 for the 2D processor), (b) TSV3D and Mono3D integrated processor with 0% overlapping transmitter *CORE*_0 and receiver *CORE*_3 and a noise core *CORE*_1.

7.2.3 Non-overlapping transmitting and receiving cores

As shown in the previous sections, the TCC bandwidth in 3D processors is significantly higher than 2D processors due to the strong thermal coupling between the transmitter (*CORE*_0) and receiver (*CORE*_1) cores that are located in different tiers, as shown in Fig. 7.5. In this section, we investigate the effect of reducing the overlap between the transmitter and receiver cores on TCC bandwidth. Two different placement scenarios are considered, one with 50% overlap and the other one with 0% overlap. For the first scenario, the 2D floorplan of the Haswell processor (see Fig. 7.5) is partitioned into two tiers such that there is 50% overlap between the cores on each tier, as depicted in Fig. 7.12. The overlap is highlighted in blue dotted lines in the figure. The transmitting core is *CORE*_0 and receiving core is *CORE*_1, as highlighted in the figure. The TCC bandwidth with $BER < 1\%$ is analyzed following the same method described in Section 7.1. The results are listed in columns 6 and 7 of Table 7.2. The TCC bandwidth yielding $BER < 1\%$ is 167 *bps* for Mono3D and 154 *bps* for TSV3D. It can be observed that these results are 17% and 15% lower than the bandwidths obtained with 100% overlap. The primary reason for this degradation is the reduced thermal coupling between the two cores in this placement scenario. Furthermore, the noise application was also executed in *CORE*_1 to study the effect of thermal interference and the results are tabulated in columns 8 and 9 of Table 7.2. The TCC bandwidths degraded further, by 15% and 35% for, respectively, Mono3D and TSV3D technologies.

For the second scenario, a 0% overlap is considered between transmitting and receiving cores. The same floorplan shown in Fig. 7.5 is assumed. The transmitting core is still *CORE*_0 whereas, the receiving core is *CORE*_3. According to TCC results, when the transmitter and receiver cores do not have any overlap, TCC

bandwidth degrades by 62% for Mono3D and 58% for TSV3D, as listed in columns 10 and 11 of Table 7.2. Furthermore, the thermal interference from *CORE*_1 (the core closest to the receiver *CORE*_3) was also considered, similar to Section 7.2.2. Different applications from SPLASH-2 suite are executed sequentially in *CORE*_1. In the presence of the noise power profile shown in Fig. 7.9, the TCC bandwidth in Mono3D and TSV3D is further reduced by approximately 13%, as listed in columns 12 and 13 of Table 7.2. Note that the degradation in bandwidth in the presence of thermal interference for both of the scenarios is in contrast to the unaffected TCC bandwidth when the cores fully overlap. The reason for the degradation of bandwidth for this scenario can be explained with the help of Fig. 7.11(b) that illustrates the variation of *BER* with the transmission rates for 0% overlapping cores. The plots resemble the variation for 2D processors observed in Fig. 7.11(a), and the transmission rate is similarly divided into regions B1 (yellow) and B2 (pink). Due to the weaker thermal coupling between the non-overlapping cores, the minimum BER for TSV3D and Mono3D is achieved only at a transmission rate of 67 *bps*, as seen in region B2. For less than 67 *bps*, the transmission rates start overlapping with the frequency of temperature profile of the noise core, thus increasing the sensitivity to heat generated by other cores. However, when the thermal interference is considered from *CORE*_2, a significant degradation in the TCC bandwidth was observed due to strong coupling between the noise core and receiving core. Specifically, the TCC bandwidth for Mono3D and TSV3D was determined as 50 *bps* with a *BER* of 12%. Therefore, in 3D ICs, it is preferable for apps that have access to sensitive information to execute on cores that are not fully overlapping with cores executing the insecure apps.

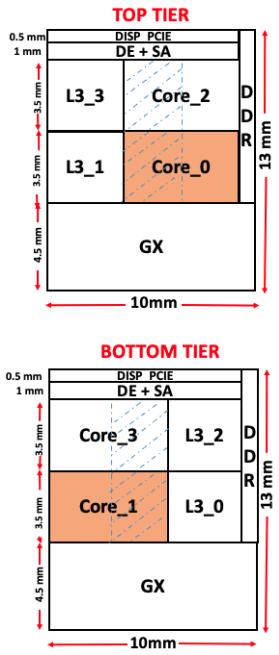


Figure 7.12: Top tier and Bottom tier floorplan for the scenario where the transmitting and receiving cores overlap by 50%.

7.2.4 Placement of transmitting core closer to heat sink

All of the simulations in the previous sections considered the transmitting core to be on the top-tier, away from the heat sink, as shown in the Mono3D and TSV3D layer stack in Fig. 7.5. Alternatively, in this section, we consider a scenario where the transmitting core (*CORE*_1) is located on the bottom tier, closest to the heat sink and the receiving core (*CORE*_0) is located on the top tier. The temperature profile of the transmitting core (*CORE*_1) and the receiving core (*CORE*_0) are shown in Fig. 7.13 for both Mono3D and TSV3D-based processors. Since the bottom tier is located closer to the heat sink, the majority of the heat flows from the transmitting core on the bottom tier to the heat sink. Thus, the core temperatures in both tiers overlap in both of the systems, as illustrated in the figure. This behaviour is in con-

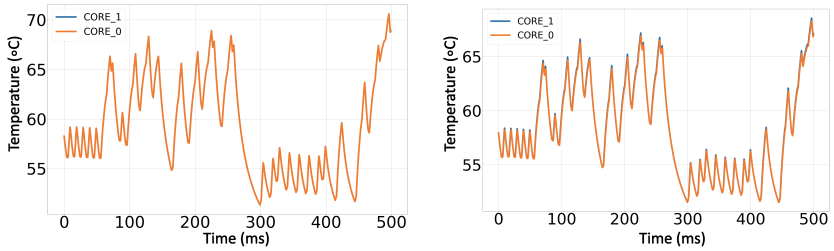
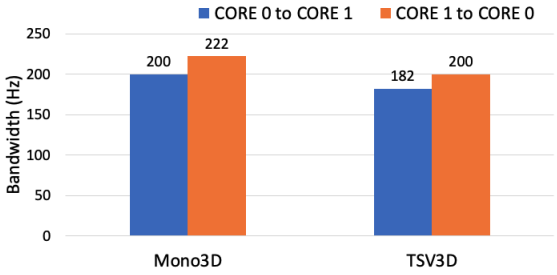


Figure 7.13: Transient temperature profiles when the transmitting core is *CORE*_1 (on the bottom tier) and the receiving core is *CORE*_0 (on the top tier) for (a) Mono3D and (b) TSV3D systems.



(a)

Figure 7.14: Comparison of TCC bandwidth (with BER $< 1\%$) for Mono3D and TSV3D systems with both *CORE*_0 to *CORE*_1 and *CORE*_1 to *CORE*_0 communication.

trast to the transient temperature profiles for *CORE*_0 to *CORE*_1 communication in Fig. 7.7, particularly for TSV3D where there is significant difference between the transmitting and receiving core temperatures.

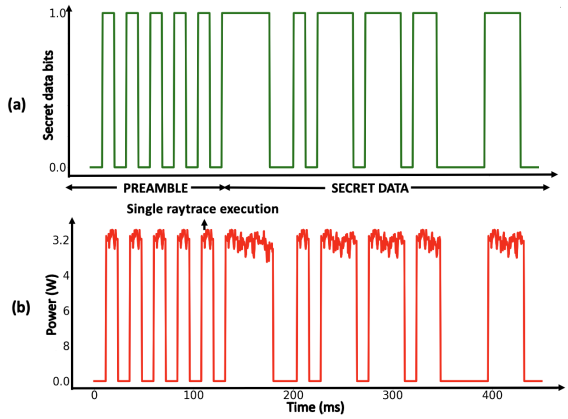
The bandwidth for less than 1% BER is characterized for this scenario, as shown in Fig. 7.14. In the figure, the bandwidths are compared for both *CORE*_0 (in top tier) to *CORE*_1 (in bottom tier) and *CORE*_1 to *CORE*_0 scenarios. We can observe that TCC bandwidths for Mono3D and TSV3D are greater for the *CORE*_1 to *CORE*_0 scenario by, respectively, 11% and 9.9%.

7.2.5 Effect of transient power variations on TCC bandwidth

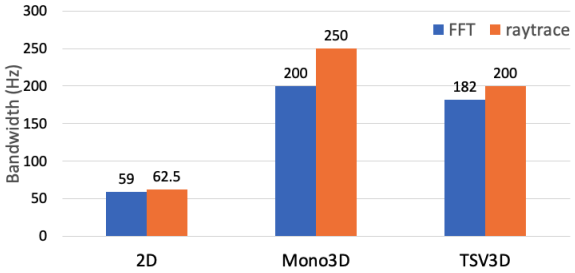
As discussed in Section 7.1.2.2, *FFT* program from the SPLASH-2 suite is used to establish a TCC attack for the results presented thus far. In this section, the effect of executing a lower power program on TCC bandwidth is explored. As seen in Fig. 7.3, *raytrace* is one of the low power applications within SPLASH-2. When this application is used to encode the sensitive data bits, the encoded power profile of the transmitting core is shown in Fig. 7.15(a). The TCC bandwidth that yields a $BER < 1\%$ is illustrated in Fig. 7.15(b) for 2D, Mono3D and TSV3D systems. Although the peak power consumption of the transmitting core executing the *FFT* program is greater than the peak power consumed when executing the *raytrace* program, TCC bandwidth achieved by *raytrace* is greater by 6%, 25% and 10%, respectively, for 2D, Mono3D, and TSV3D systems. This increase in TCC bandwidth can be explained through the transient power profile. Specifically, when executing *FFT*, even though the peak power is higher, the power variation during execution is also high, as seen in Fig. 7.4. Alternatively, the transient power profile of *raytrace* is relatively more stable during execution (even though at lower power levels), as shown in Fig. 7.15(a). This transient stability of power profile results in higher TCC bandwidth since the corresponding temperature profile exhibits less noise.

7.2.6 TCC in 3D processors with more than two tiers

In this section, the TCC bandwidth is analysed for a scenario with more than two tiers in Mono3D and TSV3D processors. The 2D floorplan of the Intel Haswell processor in Fig. 7.5 is partitioned into four tiers and the floorplan of each tier



(a) TCC encoding showing (a) sample secret data bit-stream and (b) the encoded transmitting core power profile.



(b) Comparison of communication bandwidths with *FFT* and *raytrace* as the TCC programs.

Figure 7.15: TCC using lower power *raytrace* program from SPLASH-2.

is shown in Fig. 7.16. The layer model shown in Fig. 7.5 is extended to include the two additional tiers. We perform two experiments with the four tier processor. First, TCC between cores on non-adjacent tiers is studied, where *CORE*_3 on Tier 3 is the transmitting core and *CORE*_1 on Tier 1 is the receiving core. The objective is to quantify the impact of having an additional tier between transmitting and receiving cores. According to the results of this scenario, as shown in Fig. 7.18, the TCC bandwidth for Mono3D system is the same as the two-tier scenario where the communicating cores are located on adjacent tiers. Alternatively, the bandwidth of TSV3D is 10% greater for the four-tier scenario, despite the fact that transmitting and receiving cores are located farther apart on non-adjacent tiers. This behavior can be explained with the help of the transient temperature profiles of the transmitting and receiving cores, as shown in Fig. 7.17. For TCC between non-adjacent tiers, the thermal coupling is still sufficiently strong for Mono3D (due to thin cross-sectional layers). However, the peak temperature of *CORE*_3 for TSV3D is 75°C whereas for Mono3D, it is 70°C. Furthermore, the rise and fall times are also steeper for *CORE*_3 in TSV3D because the thermal resistance of the upper most tier (Tier 3) is the highest for TSV3D (due to thick bulk and BCB layers between consecutive tiers). Thus, due to higher temperatures and steeper rise/fall times, the TCC bandwidth for TSV3D is greater for the four-tier scenario with non-adjacent transmitting and receiving cores as compared to the two-tier scenario with adjacent transmitting and receiving cores.

For the second scenario with the four tier model, *CORE*_3 is considered to be the transmitting core, *CORE*_2 is the receiving core, and noise application is executed in *CORE*_1. In this scenario, the maximum TCC bandwidth is calculated as 100 *bps* for both Mono3D and TSV3D with *BER* of 2% and 3%, respectively. When the

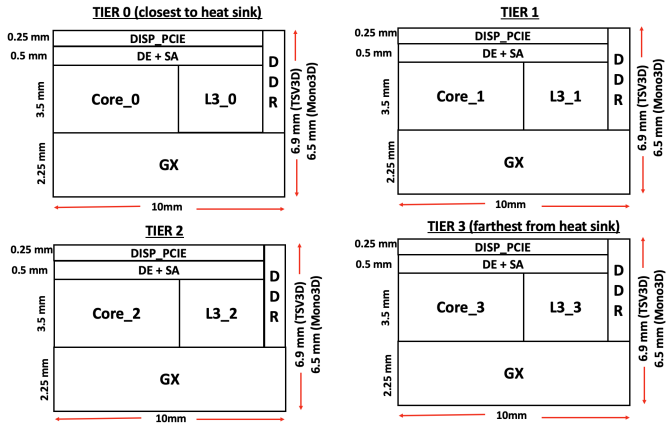
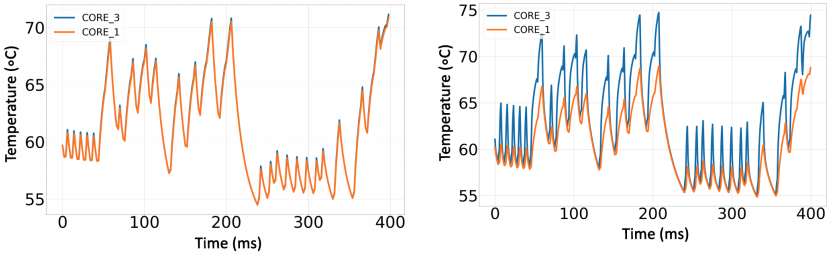


Figure 7.16: Floorplan of the Haswell processor partitioned into 4-tiers using Mono3D and TSV3D integration.



(a) (b)

Figure 7.17: Transient temperature profiles of the transmitting core (*CORE*_3) on Tier 3 and receiving core (*CORE*_1) on Tier 1 for (a) Mono3D and (b) TSV3D based processor.

noise application is executed within the tier directly beneath the receiving tier, the TCC bandwidth is significantly degraded due to strong coupling of the interference caused by the noise application.

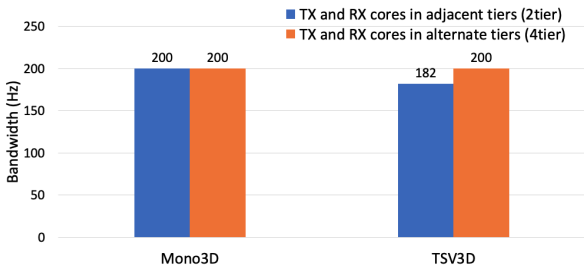


Figure 7.18: Comparison of TCC bandwidth (with BER $<$ 1%) for Mono3D and TSV3D systems with transmitting and receiving cores on adjacent tiers and on alternate tiers.

7.3 Summary

In this chapter, we demonstrated that the thermal covert-channel bandwidth between cores of a Mono3D and TSV3D based Intel processor is 3.2$\times$ and 4$\times$ greater than the bandwidth achieved in a conventional 2D processor. Furthermore, unlike previous works that typically rely on computationally intensive CPU stress applications to encode the secret data, in this work, a high bandwidth channel is established by using common SPLASH-2 benchmark applications such as *FFT* and *raytrace* to transfer upto 250 *bps* and 200 *bps* of secret data in Mono3D and TSV3D systems respectively. Additionally, thermal covert-channel characterization is also performed in the presence of thermal interference due to neighboring active cores and it is shown that the covert-channel bandwidth in 3D systems is mostly unaffected from heat generated by the other cores while still achieving less than 1% bit-error rate. However, for 2D systems, the thermal interference increases the minimum bit-error rate to 7% and the bandwidth degrades by 13%. The significant increase in covert-channel communication bandwidth in vertically integrated processors is due to the overlapping transmitter and receiver cores, which maximizes the thermal coupling between them. Thus, the effect of reducing or eliminating the overlap between

the cores on covert-channel communication is also investigate and the bandwidth is shown to degrade by up to 62% for Mono3D processor and 58% for TSV3D processor.

Chapter 8

Enhanced Detection of Thermal Covert Channel Attacks in 3D-Integrated Multicore Processors

In the previous chapter, 3D integration technologies were leveraged to establish a high bandwidth TCC attack by executing nominal or low power benchmark programs. In this chapter, we show that the existing detection techniques fail to detect TCC attacks established by low power benchmark programs. Therefore, a novel enhanced metric for detection of low power TCC is proposed in this chapter.

The organization of this chapter is as follows. The existing TCC detection techniques and their drawbacks are presented in Section 8.1. The proposed metric for detection is presented in Section 8.2 and results for quantifying the new detection metric are shown in Section 8.3. Finally, the chapter is concluded in Section 8.4.

8.1 Drawbacks of Existing Works on TCC Detection

Huang et al. recently proposed two techniques for TCC detection (17; 18). The first technique analyzes the temperature profile of each core in the frequency domain (18). Specifically, power spectrum of temperature profile of each core is scanned at high frequencies by comparing them against a fixed threshold. However, this frequency scanning technique requires the use of band pass filter at several frequency steps and therefore increasing overhead in every detection cycle. Moreover, this technique fails to detect TCC between applications in the same physical core.

Subsequently, the same team proposed another detection metric that analyzes the frequency spectrum of the CPU workload of each logical core to detect a TCC (17). The CPU workload is measured in terms of the Instructions Per Cycle (IPC). In this technique, the FFT spectrum of IPC of each core is obtained and the maximum amplitude of the spectrum is compared against a set threshold to determine if it is contributed by a TCC. Typically, the power consumption of benchmark applications from SPLASH-2 or PARSEC suites occupy a low frequency band of about 0-10 Hz. Therefore, a TCC attack is established at greater frequencies to avoid any interference. Consequently, the above detection technique focuses on a frequency range from 10 Hz up to 500 Hz. Previously, we had shown in Section 7.2.5 that a high bandwidth TCC attack can be established with a low power application such as *raytrace*. However, the IPC-based technique fails to detect this TCC attack as shown in Fig. 8.1.

Figs. 8.1(a) and (c) show the time domain and frequency domain IPC spectrum of a Haswell processor core without TCC (executing random benchmark applications from the SPLASH-2 and PARSEC suite sequentially). Figs. 8.1(b) and (d) depict the time domain and frequency domain IPC spectrum of a core with *raytrace*

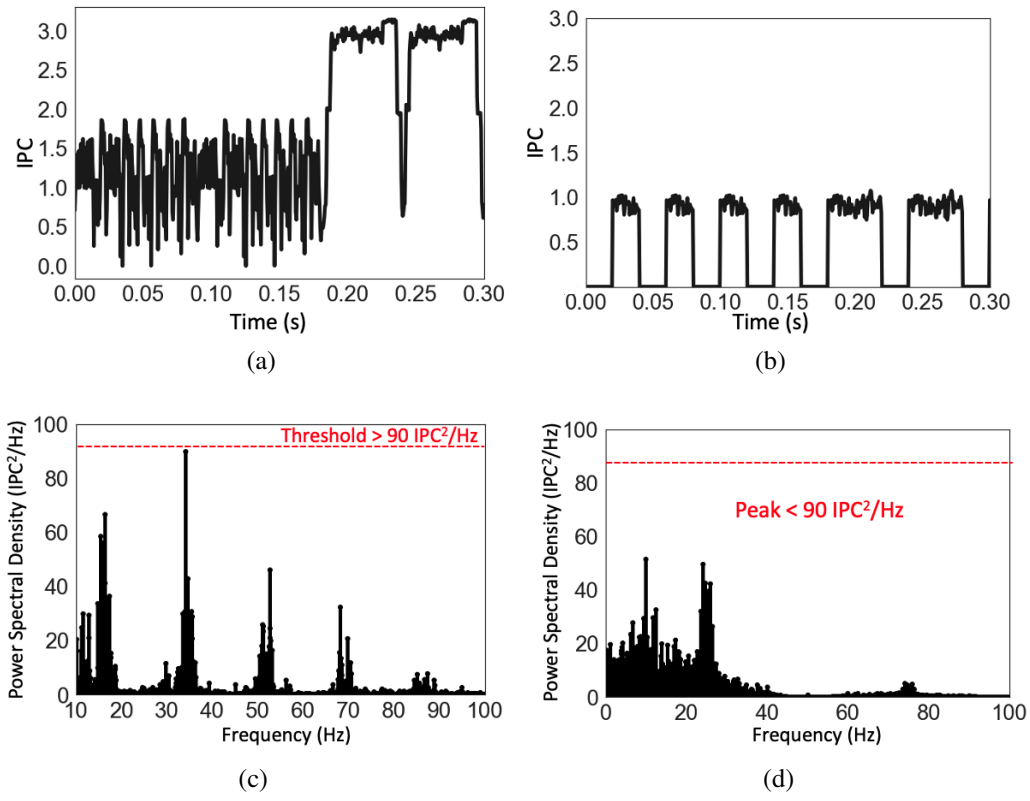


Figure 8.1: IPC profiles showing (a) time domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites, (b) time domain spectrum a core executing *raytrace* program encoded with the secret data, (c) frequency domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites and (d) frequency domain spectrum a core executing *raytrace* program encoded with the secret data.

based TCC program, respectively. From the time domain spectrum, it can be observed that the IPC of the processor core with raytrace based TCC is much lower than the IPC of the core without TCC. Furthermore, in the frequency domain, it can be observed that the peak power spectral density (PSD) of IPC spectrum without TCC is greater than the PSD of IPC spectrum with TCC, even in the range of 10 Hz to 100 Hz. In (17), the threshold for detection is set to be greater than the average IPC without TCC. For example, in Fig. 8.1(c), the threshold is approximately greater than the average peaks of 50 IPC^2/Hz. However, the PSD of the IPC spectrum with TCC in Fig. 8.1(c) is much lower than 50 IPC^2/Hz. Therefore, this type of TCC cannot be detected by looking only at the IPC metric of each core. The following section describes the enhanced detection technique proposed in this work that can overcomes this drawback.

8.2 Proposed Technique for Detecting Low-power and High Bandwidth TCC

8.2.1 Enhanced detection metric

In this section, we propose two enhancements to the above technique that also leverages CPU workload to detect a TCC established by a low power program. First, the metric Giga Instructions Per Second (GIPS) of each processor is calculated, as $GIPS = IPC \times f$, where f is the frequency of the processor. In a TCC, a bit '1' is encoded on the temperature profile of the transmitting core by executing a program to raise its temperature. If an attacker chooses to additionally increase the frequency of the core to significantly raise the core temperature, the IPC met-

ric falls short of detecting this situation since IPC is *independent* of the processor frequency. Therefore, we propose GIPS as a better measure for detection.

However, GIPS is only a scaled version of IPC and the peak frequency spectrum of GIPS without TCC is still greater than the peak spectrum with TCC, for a *ray-trace* based attack as discussed in Section 8.1. Therefore, in this work, we propose to leverage the transitions in the GIPS profile to detect the low power program based TCC. In other words, a GIPS spectrum with TCC has several extreme variations in its magnitude because of the frequent turning ON (for encoding a bit '1') and OFF (for encoding a bit '0') of the application, unlike the GIPS spectrum without TCC. Therefore, we propose to calculate the difference between the consecutive samples of GIPS (henceforth referred to as $\Delta GIPS$) to quantify these variations.

Fig. 8.2(a) and Fig. 8.2(b) depict the GIPS spectrum of the processor core without TCC and with TCC at 100 Hz, respectively. The corresponding $\Delta GIPS$ spectrum is calculated as the difference of the GIPS spectrum between consecutive samples for the processor core without TCC and shown in Fig. 8.2(c). Similarly, $\Delta GIPS$ for the core with TCC is shown in Fig. 8.2(d). Please note that even though the magnitude of variations is larger in Fig. 8.2(a) and hence resulting in higher peaks in Fig. 8.2(c), the occurrence of the peaks is random and less frequent compared to Fig. 8.2 (d) where the occurrence of the peaks are consistent and more frequent. Specifically, Fig. 8.2 (d) is a dirac comb function and the frequency response of this type of dirac comb function is also a dirac comb with peaks at odd multiples of the fundamental frequency, as shown in Fig. 8.2(f) (139). It can be directly observed that the peaks in the frequency spectrum occur at odd multiples of the transmission frequency of 100 Hz and therefore can be leveraged to detect a TCC. However, to capture this effect, we calculate the sum of these peaks and determine to see if the

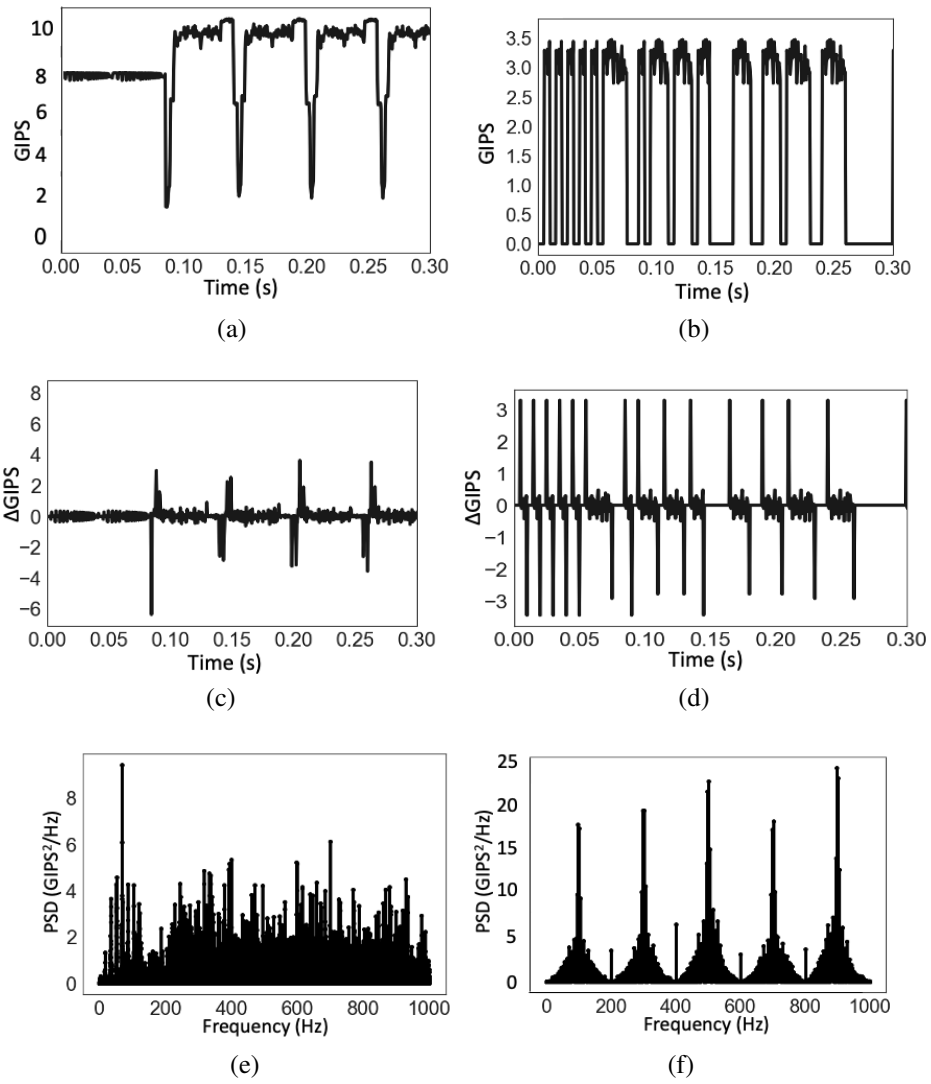


Figure 8.2: GIPS profiles showing (a) time domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites, (b) time domain spectrum a core executing *raytrace* program encoded with the secret data, (c) frequency domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites and (d) frequency domain spectrum a core executing *raytrace* program encoded with the secret data.

resultant magnitude is greater than the peak in the power spectrum without TCC in Fig. 8.1(e). The algorithm for detection using this enhanced detection metric is described in the following section.

8.2.2 Detection algorithm

Algorithm 2 $\Delta GIPS$ based detection of each core

1: **Inputs:** $\Delta GIPS_i, T_{det}, N_{cores}$
2: **Initialization:** $f = 10:10:1000Hz$
3: **for** $1 \leq i \leq N_{cores}$ **do**
4: $\quad \mathcal{F} \leftarrow$ Fast Fourier Transform (FFT) spectrum of $\Delta GIPS_i$
5: $\quad$ **if** $F \in$ C(f) **then**
6: $\quad\quad P_{\Delta GIPS_i} \leftarrow$ Sum of all peaks in $\mathcal{F}$
7: $\quad$ **else**
8: $\quad\quad P_{\Delta GIPS_i} \leftarrow$ Peak in $\mathcal{F}$
9: $\quad$ **end if**
10: $\quad$ **if** $P_{\Delta GIPS_i} > T_{det}$ **then**
11: $\quad\quad detect_i \leftarrow 1$
12: $\quad$ **end if**
13: **end for**

The TCC detection algorithm is shown in Alg. 2. We assume that TCC detection can be configured for the secure cores. Based on the configuration, the above algorithm is executed as an asynchronous thread on these cores having access to secure information. The detection cycle is assumed to execute every 1 second in these cores. In each detection cycle, each secure core extracts the GIPS spectrum and calculates the frequency response of $\Delta GIPS\ (F)$ for that core as shown in line 4 of Alg. 2. If the frequency response resembles that of a dirac comb function as discussed in the previous section, the sum of all the peaks in F is stored in $P_{\Delta GIPS_i}$ (see lines 5 to 9 in Alg. 2). If $P_{\Delta GIPS_i}$ crosses a set threshold, a $detect$ flag is set as shown

in lines 10 to 12 in Alg. 2. The threshold is statistically calculated based on an exhaustive set of simulations of common applications from SPLASH-2/PARSEC as described in the following subsection.

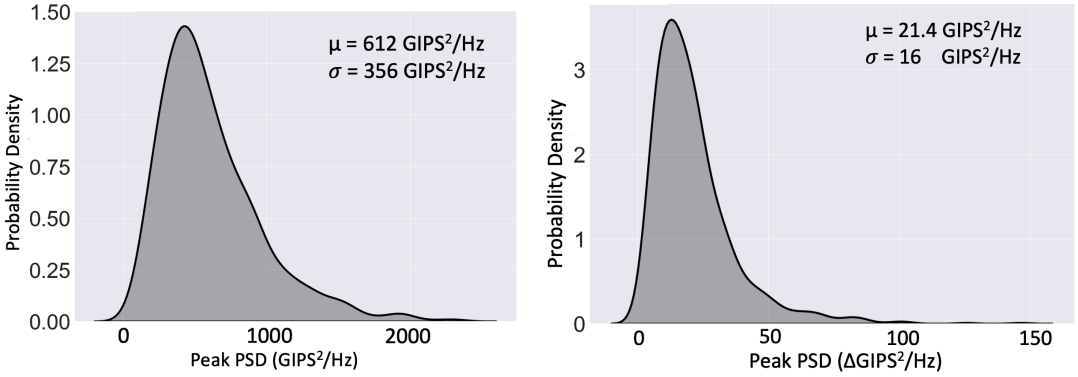
8.2.3 Threshold determination

In existing IPC based detection technique (17), the threshold for detection is estimated based on the average of the IPC power spectrum for SPLASH-2/PARSEC applications without a TCC and the deviation in the threshold is not considered extensively. Therefore, in this work, we execute random applications from SPLASH-2/PARSEC in a core sequentially with dispersed instants of idle time to mimic a processor core without TCC. In order to statistically model the threshold for detection, the peak $GIPS$ and $\Delta GIPS$ values are recorded for 1000 such simulations. These amplitudes follow a Gaussian distribution as shown in Fig. 8.3(a) for $GIPS$ and Fig. 8.3(b) for $\Delta GIPS$.

In order to consider 95% of variation in the amplitudes, the maximum threshold for detection is calculated based on the 4σ variation as 1324 $GIPS^2/Hz$ and 53 $\Delta GIPS^2/Hz$. Therefore, if $P_{\Delta GIPS_i}$ in Alg. 2 is greater than the threshold calculated above, a TCC is said to be detected.

8.3 Simulation Results

To simulate a TCC transmitting core, 10 blocks of secret data with 100 random bits in each block are encoded in the power profile and consequently on the GIPS profile of the applications. To evaluate the effectiveness of the proposed detection technique, we simulate 100 such communications and define a metric referred to as



(a) (b)

Figure 8.3: Determination of the detection threshold for (a) time domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites, (b) time domain spectrum a core executing *raytrace* program encoded with the secret data, (c) frequency domain spectrum of core sequentially executing random applications from SPLASH-2 and PARSEC suites and (d) frequency domain spectrum a core executing *raytrace* program encoded with the secret data.

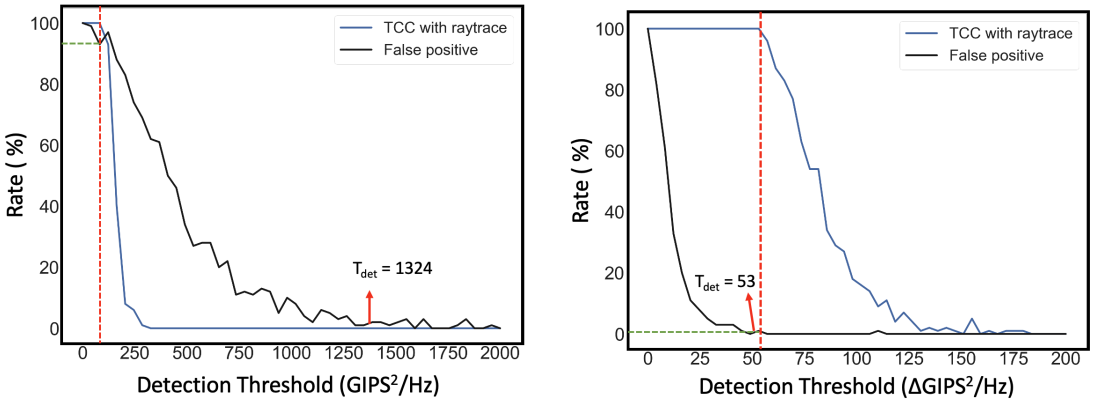
detection accuracy or TCC detection rate (R_{det}), similar to (17),

$$R_{det}(in\%) = \frac{N_{detected}}{N_{TCC}}, \tag{8.1}$$

where, $N_{detected}$ is the number of TCCs that can be detected with the sum of peak of $\Delta GIPS_i$ spectrum crossing the threshold (see lines 10 to 12 in Alg. 2) and N_{TCC} is the total number of TCCs established.

Furthermore, to compare this detection against cores executing nominal SPLASH-2/PARSEC applications, we execute 100 simulations of a sequence of such applications and we define a metric referred to as false positive rate (R_{fp}),

$$R_{fp}(in\%) = \frac{N_{fp}}{N}, \tag{8.2}$$

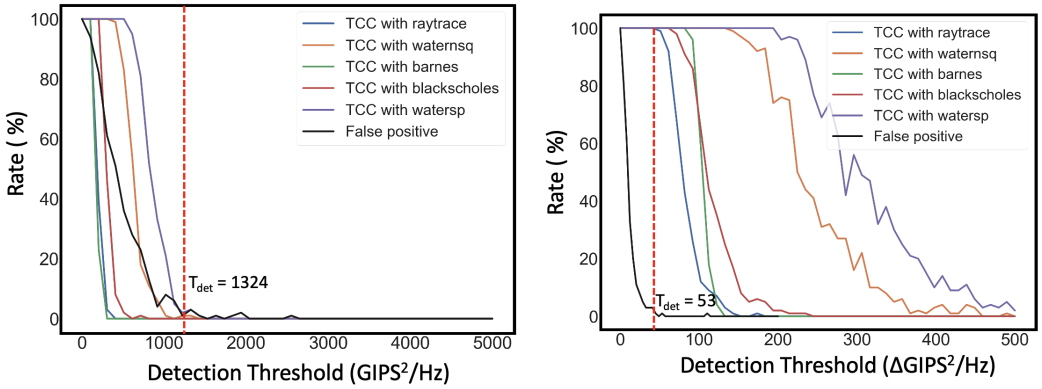


(a) (b)

Figure 8.4: TCC detection rate and false positive rate vs detection threshold for (a) GIPS based detection and (b) proposed $/DeltaGIPS$ based detection.

where, N_{fp} is the total number of typical simulations without TCC that get detected as a TCC (also referred to as false positives) and N is the total number of such simulations.

Fig. 8.4(a) and Fig. 8.4(b) show the variation of R_{det} (blue plot) and R_{fp} (black plot) with different detection thresholds with GIPS (or IPC) based detection and with proposed $\Delta GIPS$ based detection. R_{det} is calculated for *raytrace* based TCC. The detection threshold (T_{det}) determined in the previous section with 4σ variation is marked in both plots. First, it can be observed that the false positive rate is 0% when the threshold is increased beyond T_{det} in both of the plots and thus verifying the calculation of T_{det}. Second, it can be observed from Fig. 8.4(a) that at the marked T_{det}, the TCC detection rate is also 0% and therefore the GIPS based threshold fails to detect any of the TCCs encoded by executing *raytrace* application. However, the rate of detecting the same *raytrace* based TCC with $\Delta GIPS$ at T_{det} is 100%. Furthermore, the red and green lines in the plots show that for



(a) (b)

Figure 8.5: TCC detection rate and false positive rate vs. detection threshold for (a) GIPS based detection and (b) proposed $/DeltaGIPS$ based detection.

the thresholds where R_{det} is 100%, R_{fp} is 95% in Fig.8.4(a) whereas R_{fp} is 0% in Fig.8.4(b). Therefore, based on the specification needed for the rate of detection and false positives, the threshold can also be determined from this plot.

The proposed algorithm is extended for TCC encoded by executing four more low power applications, *water.nsq*, *barnes*, *blackscholes* and *water.spatial*. Similar plots for the variation of R_{det} and R_{fp} with the detection threshold are shown in Figs. 8.5(a) and (b) with *GIPS* and $\Delta GIPS$. It can be observed in Fig. 8.5(b) that with the proposed algorithm, the TCC detection rate with all of the applications is 100% at $T_{det} = 53$ with 0% false positive rate. However, R_{det} and R_{fp} are 0% for all of the applications at T_{det} in Fig. 8.5(a).

8.4 Summary

In this chapter, a novel detection metric using Giga Instructions Per Second (GIPS) is proposed to detect a high bandwidth TCC established by executing low

power programs. The threshold for detection is statistically modelled by considering 95% of the variations. At the determined threshold, 100% of the TCCs were detected with 0% false positive rate and the results are shown for five low power applications from SPLASH-2/PARSEC benchmark suites.

Chapter 9

Conclusion And Future Work

A secure, power-based side-channel attack resistant SIMON encryption core was developed in this research with application to RF-powered devices. A thermal covert-channel leveraging 3D integration technologies to establish a high bandwidth communication was identified and an enhanced detection technique was also presented. The contributions of this thesis are summarized in Section 9.1 and possible future directions are presented in Section 9.2.

9.1 Thesis Summary

A novel charge-based methodology was developed to mount a power side-channel attack on a charge-recycling lightweight SIMON encryption core. A correlation power side-channel attack was mounted on an adiabatic ECRL based SIMON core with the proposed methodology. It was demonstrated that adiabatic operation enhances encryption efficiency (kilobits/sec/W) by approximately 10× while also exhibiting approximately 4× higher CPA resistance as compared to static CMOS

based SIMON implementation.

Despite achieving higher CPA resistance, an unprotected adiabatic SIMON is still susceptible to CPA attacks since the resistance offered is not sufficiently high. Consequently, a novel adiabatic logic structure called SEcure Adiabatic Logic for Wirelessly-Powered IoT Devices (SEAL-RF) was proposed. The SIMON encryption core was designed with the proposed adiabatic logic and it was concluded that the proposed SIMON implementation provides 52× higher CPA resistance while reducing the energy per encryption by 15.6%, compared to the unprotected adiabatic SIMON.

In the second part of the research, a high bandwidth thermal covert communication was established by leveraging 3D integration technologies. It is demonstrated that by leveraging vertical integration, it is sufficient to execute typical SPLASH-2 benchmark applications to transfer 200 bits per second of secret data via thermal covert-channels. Therefore, the probability of detecting such an attack is low and hence increasing the danger posed by it. The strong vertical thermal coupling among the cores of a 3D multicore processor is shown to increase the rates of covert communication by 3.4× compared to covert communication in conventional 2D ICs. Furthermore, the bandwidth of this thermal communication in 3D ICs is shown to be less affected by applications running in other cores and the effect of reducing inter-tier overlap between colluded cores is investigated. Furthermore, a novel detection metric is proposed that can detect a thermal covert-channel attack established by low power applications with 100% detection accuracy and 0% false positive rate for up to 100 Hz of transmission frequency.

9.2 Possible Future directions

9.2.1 A generic secure adiabatic logic gate

The novel adiabatic logic (SBPAL) proposed in this work was designed with a bipolar power-clock supply and compared with the existing adiabatic logic that can operate with a similar supply. Even though this logic has significant potential in terms of security and energy consumption for AC computing, the possibility of extending it to other generic applications should be studied. Specifically, the effects of generating a bipolar power-clock signal should be evaluated and the security metrics should also be compared with the existing secure adiabatic designs operating with unipolar power-clock supply in order to develop a *generic* secure adiabatic logic family. Such logic family can be used not only for RF-powered applications, but also for more conventional battery powered DC-based devices.

9.2.2 Dynamic Frequency Scaling (DFS) based thermal covert channel attacks

A thermal covert communication channel is established by encoding the secret data bits on the temperature profile of the transmitting core. In all of the existing works, a program is executed to raise the temperature of the core to transmit a bit '1' and the program execution is stopped to transmit a bit '0'. Modern multicore processors have different types of CPU frequency governors that define the frequency of the CPU. Specifically, in *user space* governor, the frequency of the core can be set by the user or a user space program. In (140), a DFS based covert-channel is shown using this user space governor where the transmitter is a user space program

that can change the processor frequency and the receiver is a program that can read the frequency register information to decode the secret bits. Similarly, a thermal covert-channel can be established using this DFS governor. Specifically, to transmit a bit '1', the frequency of the transmitting core can be increased (using the above CPU governor) to increase its temperature. Similarly, to transmit a bit '0', the frequency of the transmitting core can be decreased. This type of DFS-based thermal covert communication is an interesting problem to investigate for future work.

9.2.3 Monolithic 3D power and performance models for multi-core processors

In order to establish a proof-of-concept of the covert-channels in a monolithic 3D based multicore processor, thermal simulations should be performed at the architecture level. A monolithic 3D system enables unprecedented levels of integration density and granularity by reducing the dimension of vertical interconnects, referred to as monolithic inter-tier vias (MIVs). These vertical interconnects should be modeled for performance and power estimation in order to have a reasonable estimation of the bandwidth and error rates of covert-channels. Even though there are various architecture-level power and thermal simulators, developing standard performance and power models for monolithic 3D ICs remains as an open problem.

Bibliography

[1] "News article," https://www.techrepublic.com/article/63-of-organizations-face-security-breaches-due-to-hardware-vulnerabilities/#:~:text=Hardware\%2Dlevel\%20breaches\%20are\%20one,to\%20a\%20hardware\%20security\%20vulnerability., accessed: 2021-06-16.

[2] P. . C. John Neuffer, "Patching hardware vulnerabilities is harder than you may think," Feb 2019. [Online]. Available: https://www.semiconductors.org/patching-hardware-vulnerabilities-is-harder-than-you-may-think

[3] D. Mukhopadhyay and R. Chakraborty, *Hardware Security: Design, Threats, and Safeguards*. Taylor & Francis, 2014. [Online]. Available: https://books.google.com/books?id=oY3aBAAAQBAJ

[4] F. Farahmandi, Y. Huang, and P. Mishra, *System-on-Chip Security Vulnerabilities*. Cham: Springer International Publishing, 2020, pp. 1–13. [Online]. Available: https://doi.org/10.1007/978-3-030-30596-3_1

[5] R. Gonzalez, B. M. Gordon, and M. A. Horowitz, "Supply and threshold

voltage scaling for low power cmos," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 8, pp. 1210–1216, 1997.

[6] J. Rabaey, *Low Power Design Essentials*, ser. Integrated Circuits and Systems. Springer US, 2009. [Online]. Available: https://books.google.com/books?id=A-sBy_nmQ8wC

[7] P. Teichmann, *Adiabatic Logic: Future Trend and System Level Perspective*. Springer Publishing Company, Incorporated, 2011.

[8] M. Morrison, "Theory, synthesis, and application of adiabatic and reversible logic circuits for security applications," in *2014 IEEE Computer Society Annual Symposium on VLSI*. IEEE, 2014, pp. 252–255.

[9] T. Wan, Y. Karimi, M. Stanacevic, and E. Salman, "Energy efficient ac computing methodology for wirelessly powered iot devices," in *IEEE Int. Symp. on Circuits and Systems*, May 2017.

[10] T. Wan, Y. Karimi, M. Stanaćević, and E. Salman, "Perspective paper—can ac computing be an alternative for wirelessly powered iot devices?" *IEEE Embedded Systems Letters*, vol. 9, no. 1, pp. 13–16, 2017.

[11] T. Wan, Y. Karimi, M. Stanaćević, and E. Salman, "Ac computing methodology for rf-powered iot devices," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 5, pp. 1017–1028, 2019.

[12] T. Wan and E. Salman, "Ultra low power simon core for lightweight encryption," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–5.

[13] R. J. Masti, D. Rai, A. Ranganathan, C. Müller, L. Thiele, and S. Capkun, "Thermal covert channels on multi-core platforms," in *24th {USENIX} Security Symposium ({USENIX} Security 15)*, 2015, pp. 865–880.

[14] D. B. Bartolini, P. Miedl, and L. Thiele, "On the capacity of thermal covert channels in multicores," in *Proceedings of the Eleventh European Conference on Computer Systems*, 2016, pp. 1–16.

[15] L. Brown and H. Seshadri, "Cool hand linux* handheld thermal extensions," in *Linux Symposium*, 2007, p. 75.

[16] S. Wang, X. Wang, Y. Jiang, A. Singh, L. Huang, and M. Yang, "Modeling and analysis of thermal covert channel attacks in many-core systems," *IEEE Transactions on Computers*, 2022.

[17] H. Huang, X. Wang, Y. Jiang, A. K. Singh, M. Yang, and L. Huang, "Detection of and countermeasure against thermal covert channel in many-core systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021.

[18] ——, "On countermeasures against the thermal covert channel attacks target-

ing many-core systems," in *2020 57th ACM/IEEE Design Automation Conference (DAC)*, 2020, pp. 1–6.

[19] Q. Wu, X. Wang, and J. Chen, "Defending against thermal covert channel attacks by task migration in many-core system," in *2021 IEEE 3rd International Conference on Circuits and Systems (ICCS)*. IEEE, 2021, pp. 111–120.

[20] Z. Long, X. Wang, Y. Jiang, G. Cui, L. Zhang, and T. Mak, "Improving the efficiency of thermal covert channels in multi-/many-core systems," in *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2018, pp. 1459–1464.

[21] P. Emma, A. Buyuktosunoglu, M. Healy, K. Kailas, V. Puente, R. Yu, A. Hartstein, P. Bose, and J. Moreno, "3d stacking of high-performance processors," in *2014 IEEE 20th International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 2014, pp. 500–511.

[22] K. Dhananjay, P. Shukla, V. F. Pavlidis, A. Coskun, and E. Salman, "Monolithic 3d integrated circuits: Recent trends and future prospects," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 3, pp. 837–843, 2021.

[23] F.-X. Standaert, *Introduction to Side-Channel Attacks*. Boston, MA: Springer US, 2010, pp. 27–42. [Online]. Available: https://doi.org/10.1007/

978-0-387-71829-3_2

[24] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards (Advances in Information Security)*. Berlin, Heidelberg: Springer-Verlag, 2007.

[25] B. W. Lampson, "A note on the confinement problem," *Communications of the ACM*, vol. 16, no. 10, pp. 613–615, 1973.

[26] S. Cabuk, C. E. Brodley, and C. Shields, "Ip covert channel detection," *ACM Transactions on Information and System Security (TISSEC)*, vol. 12, no. 4, pp. 1–29, 2009.

[27] Z. Wu, Z. Xu, and H. Wang, "Whispers in the hyper-space: high-bandwidth and reliable covert channel attacks inside the cloud," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 603–615, 2014.

[28] G. Shah, A. Molina, M. Blaze *et al.*, "Keyboards and covert channels." in *USENIX Security Symposium*, vol. 15, 2006, p. 64.

[29] L. Deshotels, "Inaudible sound as a covert channel in mobile devices," in *8th {USENIX} Workshop on Offensive Technologies ({WOOT} 14)*, 2014.

[30] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development*, vol. 5, no. 3, pp. 183–191, 1961.

[31] V. G. Oklobdzija, D. Maksimovic, and Fengcheng Lin, "Pass-transistor adiabatic logic using single power-clock supply," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 10, pp. 842–846, 1997.

[32] V. Anantharam, M. He, K. Natarajan, H. Xie, and M. P. Frank, "Driving fully-adiabatic logic circuits using custom high-q mems resonators." in *ESA/VLSI*, 2004, pp. 5–11.

[33] S. G. Younis and T. F. Knight Jr, "Practical implementation of charge recovering asymptotically zero power cmos," in *Proceedings of the 1993 symposium on Research on integrated systems*, 1993, pp. 234–250.

[34] Y. Moon and D.-K. Jeong, "An efficient charge recovery logic circuit," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 4, pp. 514–522, 1996.

[35] V. Oklobdzija, D. Maksimovic, and F. Lin, "Pass-transistor adiabatic logic using single power-clock supply," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 10, pp. 842–846, 1997.

[36] Y. Ye and K. Roy, "Qserl: Quasi-static energy recovery logic," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 2, pp. 239–248, 2001.

[37] C.-S. A. Gong, M.-T. Shiue, C.-T. Hong, and K.-W. Yao, "Analysis and de-

sign of an efficient irreversible energy recovery logic in 0.18-m cmos," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 9, pp. 2595–2607, 2008.

[38] D. Maksimovic, V. Oklobdzija, B. Nikolic, and K. W. Current, "Clocked cmos adiabatic logic with integrated single-phase power-clock supply: experimental results," in *Int. Symp. on Low Power Electronics and Design*, 1997, pp. 323–327.

[39] T. Wan, E. Salman, and M. Stanacevic, "A new circuit design framework for iot devices: Charge recycling with wireless power harvesting," in *IEEE Int. Symp. on Circuits and Systems*, May 2016.

[40] E. Salman, M. Stanaćević, S. Das, and P. M. Djurić, "Leveraging rf power for intelligent tag networks," in *Proceedings of the 2018 on Great Lakes Symposium on VLSI*. ACM, 2018, pp. 329–334.

[41] Y. Huang, T. Wan, E. Salman, and M. Stanaćević, "Signal shaping at interface of wireless power harvesting and ac computational logic," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2019, pp. 1–5.

[42] T. Wan, E. Salman, and M. Stanaćević, "Ac computing methodology for rf powered iot security," in *Government Microcircuit Applications & Critical Technology Conference*, 2018, pp. 939–944.

[43] K. Dhananjay and E. Salman, "Special session: Adiabatic circuits for energy-efficient and secure iot systems," in *2020 IEEE 38th International Conference on Computer Design (ICCD)*. IEEE, 2020, pp. 17–20.

[44] R. Beaulieu, S. Treatman-Clark, D. Shors, B. Weeks, J. Smith, and L. Wingers, "The simon and speck lightweight block ciphers," in *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2015, pp. 1–6.

[45] A. Biryukov and L. Perrin, "State of the art in lightweight symmetric cryptography," *Cryptology ePrint Archive*, 2017.

[46] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. Robshaw, Y. Seurin, and C. Vikkelsoe, "Present: An ultra-lightweight block cipher," in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2007, pp. 450–466.

[47] J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavun, M. Knezevic, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, P. Rombouts, S. S. Thomsen, and T. Yalçın, "Prince – a low-latency block cipher for pervasive computing applications," in *Advances in Cryptology – ASIACRYPT 2012*, X. Wang and K. Sako, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[48] T. Akishita and H. Hiwatari, "Very compact hardware implementations of the

blockcipher clefia," in *International Workshop on Selected Areas in Cryptography*. Springer, 2011, pp. 278–292.

[49] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita, "Camellia: A 128-bit block cipher suitable for multiple platforms—design and analysis," in *International Workshop on Selected Areas in Cryptography*. Springer, 2000, pp. 39–56.

[50] R. Beaulieu, D. Shors, J. Smith, S. Treatman-Clark, B. Weeks, and L. Wingers, "The simon and speck families of lightweight block ciphers." *IACR Cryptology ePrint Archive*, vol. 2013, no. 1, pp. 404–449, 2013.

[51] ISO Security services for RFID air interfaces, "Information technology — automatic identification and data capture techniques," International Organization for Standardization, Geneva, CH, Standard ISO/IEC TR 29167-21:2018, 2018. [Online]. Available: https://www.iso.org/standard/70388.html

[52] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Annual international cryptology conference*. Springer, 1999, pp. 388–397.

[53] "International roadmap for devices and systems." [Online]. Available: https://irds.ieee.org

[54] "Compact model to efficiently characterize tsv-to-transistor noise coupling

in 3d ics," *Integration*, vol. 47, no. 3, pp. 296–306, 2014.

[55] T. Pawlowski, "Hybrid memory cube (hmc)," in *Proceedings of the IEEE Hot Chips Symposium*, August 2011.

[56] D. U. Lee *et al.*, "25.2 a 1.2v 8gb 8-channel 128gb/s high-bandwidth memory (hbm) stacked dram with effective microbump i/o test methods using 29nm process and tsv," in *Proceedings of the IEEE International Solid-State Circuits Conference*, February 2014.

[57] "Embedded multi-die interconnect bridge," https://www.intel.com/content/www/us/en/foundry/emib.html, accessed: 2020-11-02.

[58] P. Shukla, A. K. Coskun, V. F. Pavlidis, and E. Salman, "An overview of thermal challenges and opportunities for monolithic 3d ics," in *Proceedings of the 2019 on Great Lakes Symposium on VLSI*, 2019, pp. 439–444.

[59] A. Todri-Sanial and C. S. E. Tan, *Physical Design for 3D Integrated Circuits*. Boca Raton, Florida: CRC Press, 2016.

[60] H. Wang, M. H. Asgari, and E. Salman, "Compact Model to Efficiently Characterize TSV-to-Transistor Noise Coupling in 3D ICs," *Integration, the VLSI Journal*, vol. 47, no. 3, pp. 296–306, June 2014.

[61] S. M. Satheesh and E. Salman, "Power distribution in TSV-based 3D

processor-memory stacks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, pp. 692–703, Dec. 2012.

[62] S. Wong *et al.*, "Monolithic 3d integrated circuits," in *Proceedings of the International Symposium on VLSI Technology, Systems and Applications*, April 2007.

[63] "News article," https://www.amd.com/en/technologies/3d-v-cache., accessed: 2021-06-16.

[64] B. Gopireddy and J. Torrellas, "Designing vertical processors in monolithic 3d," in *2019 ACM/IEEE 46th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 2019, pp. 643–656.

[65] P. Vivet, E. Guthmuller, Y. Thonnart, G. Pillonnet, C. Fuguet, I. Miro-Panades, G. Moritz, J. Durupt, C. Bernard, D. Varreau *et al.*, "Intact: A 96-core processor with six chiplets 3d-stacked on an active interposer with distributed interconnects and integrated power management," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 1, pp. 79–97, 2020.

[66] C. Yan and E. Salman, "Routing congestion aware cell library development for monolithic 3d ics," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, 2017, pp. 1–4.

[67] I. Miketic and E. Salman, "Energy-efficient adiabatic circuits using

transistor-level monolithic 3d integration," in *2020 IEEE 33rd International System-on-Chip Conference (SOCC)*, 2020, pp. 191–194.

[68] J. Dofe, Q. Yu, H. Wang, and E. Salman, "Hardware security threats and potential countermeasures in emerging 3d ics," in *International Great Lakes Symposium on VLSI (GLSVLSI)*, 2016, pp. 69–74.

[69] Y. Xie, C. Bao, C. Serafy, T. Lu, A. Srivastava, and M. Tehranipoor, "Security and vulnerability implications of 3D ICs," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 2, pp. 108–122, 2016.

[70] P. Gu, D. Stow, R. Barnes, E. Kursun, and Y. Xie, "Thermal-aware 3d design for side-channel information leakage," in *IEEE International Conference on Computer Design*, 2016, pp. 520–527.

[71] J. Knechtel and O. Sinanoglu, "On mitigation of side-channel attacks in 3d ics: Decorrelating thermal patterns from power and activity," in *ACM/EDAC/IEEE Design Automation Conference*, 2017, pp. 1–6.

[72] J. Dofe, C. Yan, S. Kontak, E. Salman, and Q. Yu, "Transistor-level camouflaged logic locking method for monolithic 3d ic security," in *IEEE Asian Hardware-Oriented Security and Trust (AsianHOST)*, 2016, pp. 1–6.

[73] J. Dofe, Z. Zhang, Q. Yu, C. Yan, and E. Salman, "Impact of power distribution network on power analysis attacks in three-dimensional integrated

circuits," in *Proceedings of the on Great Lakes Symposium on VLSI*, 2017, p. 327–332.

[74] C. Yan, J. Dofe, S. Kontak, Q. Yu, and E. Salman, "Hardware-efficient logic camouflaging for monolithic 3-d ics," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 6, pp. 799–803, 2018.

[75] C. Yan and E. Salman, "Physical design of monolithic 3d ics with applications to hardware security," in *Government Microcircuit Applications & Critical Technology Conference*, 2018, pp. 702–707.

[76] I. Miketic and E. Salman, "Monolithic 3d integrated encryption core," in *Government Microcircuit Applications & Critical Technology Conference*, 2019.

[77] ——, "Power and data integrity in monolithic 3d integrated simon core," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2019, pp. 1–5.

[78] S. Tian and J. Szefer, "Temporal thermal covert channels in cloud fpgas," in *Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 2019, pp. 298–303.

[79] I. Giechaskiel, K. B. Rasmussen, and J. Szefer, "C 3 apsule: Cross-fpga covert-channel attacks through power supply unit leakage," in *2020 IEEE*

Symposium on Security and Privacy (SP). IEEE, 2020, pp. 1728–1741.

[80] T. Claeys, F. Rousseau, B. Simunovic, and B. Tourancheau, "Thermal covert channel in bluetooth low energy networks," in *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, 2019, pp. 267–276.

[81] P. Rahimi, A. K. Singh, and X. Wang, "Selective noise based power-efficient and effective countermeasure against thermal covert channel attacks in multi-core systems," *Journal of Low Power Electronics and Applications*, vol. 12, no. 2, p. 25, 2022.

[82] J. Wang, X. Wang, Y. Jiang, A. K. Singh, L. Huang, and M. Yang, "Combating enhanced thermal covert channel in multi-/many-core systems with channel-aware jamming," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 39, no. 11, pp. 3276–3287, 2020.

[83] S. Chen, W. Xiong, Y. Xu, B. Li, and J. Szefer, "Thermal covert channels leveraging package-on-package dram," in *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, 2019, pp. 319–326.

[84] E. Gulcan, A. Aysu, and P. Schaumont, "A flexible and compact hardware architecture for the simon block cipher," in *Lightweight Cryptography for*

Security and Privacy, T. Eisenbarth and E. Öztürk, Eds. Cham: Springer International Publishing, 2015, pp. 34–50.

[85] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Cryptographic Hardware and Embedded Systems - CHES 2004*, M. Joye and J.-J. Quisquater, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 16–29.

[86] D. Shanmugam, R. Selvam, and S. Annadurai, "Differential power analysis attack on simon and led block ciphers," in *Security, Privacy, and Applied Cryptography Engineering*, R. S. Chakraborty, V. Matyas, and P. Schaumont, Eds. Cham: Springer International Publishing, 2014, pp. 110–125.

[87] S. Bhasin, T. Graba, J. Danger, and Z. Najm, "A look into simon from a side-channel perspective," in *2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, May 2014, pp. 56–59.

[88] "Cadence spectre simulation platform," https://www.cadence.com/en_US/home/tools/custom-ic-analog-rf-design/circuit-simulation/spectre-simulation-platform.html, accessed: 2020-10-27.

[89] MATLAB, *9.9.0.1467703 (R2020b)*. Natick, Massachusetts: The MathWorks Inc., 2020.

[90] D. D. Hwang, K. Tiri, A. Hodjat, B. C. Lai, S. Yang, P. Schaumont, and

I. Verbauwhede, "AES-based security coprocessor IC in 0.18-μm CMOS with resistance to differential power analysis side-channel attacks," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 781–790, 2006.

[91] A. Singh, N. Chawla, J. H. Ko, M. Kar, and S. Mukhopadhyay, "Energy efficient and side-channel secure cryptographic hardware for iot-edge nodes," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 421–434, Feb 2019.

[92] D. Das, S. Maity, S. B. Nasir, S. Ghosh, A. Raychowdhury, and S. Sen, "High efficiency power side-channel attack immunity using noise injection in attenuated signature domain," in *2017 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE, 2017, pp. 62–67.

[93] P.-C. Liu, H.-C. Chang, and C.-Y. Lee, "A low overhead dpa countermeasure circuit based on ring oscillators," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 7, pp. 546–550, 2010.

[94] D. D. Hwang, K. Tiri, A. Hodjat, B. . Lai, S. Yang, P. Schaumont, and I. Verbauwhede, "Aes-based security coprocessor ic in 0.18-*muhboxm*cmos with resistance to differential power analysis side-channel attacks," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 781–792, 2006.

[95] H. Thapliyal, T. S. S. Varun, and S. D. Kumar, "Adiabatic computing based low-power and dpa-resistant lightweight cryptography for iot devices," in *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July

2017, pp. 621–626.

[96] M. Avital, H. Dagan, I. Levi, O. Keren, and A. Fish, "Dpa-secured quasi-adiabatic logic (sqal) for low-power passive rfid tags employing s-boxes," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 1, pp. 149–156, Jan 2015.

[97] C. Monteiro, Y. Takahashi, and T. Sekine, "Dpa resistance of charge-sharing symmetric adiabatic logic," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2013, pp. 2581–2584.

[98] ——, "Robust secure charge-sharing symmetric adiabatic logic against side-channel attacks," in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2013, pp. 732–736.

[99] S. D. Kumar, H. Thapliyal, A. Mohammad, V. Singh, and K. S. Perumalla, "Energy-efficient and secure s-box circuit using symmetric pass gate adiabatic logic," in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2016, pp. 308–313.

[100] S. Kumar, H. Thapliyal, A. Mohammad, and K. Perumalla, "Design exploration of a symmetric pass gate adiabatic logic for energy-efficient and secure hardware," *Integration, the VLSI Journal*, vol. 58, 09 2016.

[101] S. D. Kumar, H. Thapliyal, A. Mohammad, V. Singh, and K. S. Perumalla,

"Energy-efficient and secure s-box circuit using symmetric pass gate adiabatic logic," in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2016, pp. 308–313.

[102] S. Kumar, H. Thapliyal, and A. Mohammad, "Ee-spfal: A novel energy-efficient secure positive feedback adiabatic logic for dpa resistant rfid and smart card," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, pp. 1–1, 12 2016.

[103] H. S. Raghav, V. A. Bartlett, and I. Kale, "Robustness of power analysis attack resilient adiabatic logic: Wcs-qual under pvt variations," in *2017 27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)*. IEEE, 2017, pp. 1–8.

[104] K. Dhananjay and E. Salman, "Equal: Efficient quasi adiabatic logic for enhanced side-channel resistance," in *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2021, pp. 332–337.

[105] H. S. Raghav, V. A. Bartlett, and I. Kale, "Investigating the effectiveness of without charge-sharing quasi-adiabatic logic for energy efficient and secure cryptographic implementations," *Microelectronics journal*, vol. 76, pp. 8–21, 2018.

[106] B. Fadaeinia and A. Moradi, "3-phase adiabatic logic and its sound sca evaluation," *IEEE Transactions on Emerging Topics in Computing*, 2020.

[107] B. Gregory and B. Shafer, "Latch-up in cmos integrated circuits," *IEEE Transactions on Nuclear Science*, vol. 20, no. 6, pp. 293–299, 1973.

[108] N. Jeanniot, G. Pillonnet, P. Nouet, N. Azemard, and A. Todri-Sanial, "Synchronised 4-phase resonant power clock supply for energy efficient adiabatic logic," in *IEEE Int. Conf. on Rebooting Computing*, 2017.

[109] A. Bargagli-Stoffi et al., "Resonant 90 degree shifter generator for 4-phase trapezoidal adiabatic logic," *Advances in Radio Science*, vol. 1, no. D. 2, pp. 243–246, 2003.

[110] D. Maksimovic and V. G. Oklobdzija, "Integrated power clock generators for low energy logic," in *IEEE Power Electronics Specialist Conference*, 1995, pp. 61–67.

[111] Y. Moon and D.-K. Jeong, "An efficient charge recovery logic circuit," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 4, pp. 514–522, 1996.

[112] H. S. Raghav, V. A. Bartlett, and I. Kale, "Investigating the effectiveness of without charge-sharing quasi-adiabatic logic for energy efficient and secure cryptographic implementations," *Microelectronics Journal*, vol. 76, pp. 8–21, 2018.

[113] K. Tiri, M. Akmal, and I. Verbauwhede, "A dynamic and differential cmos logic with signal independent power consumption to withstand differential

power analysis on smart cards," in *Proceedings of the 28th European solid-state circuits conference*. IEEE, 2002, pp. 403–406.

[114] P. Kocher, J. Jaffe, B. Jun, and P. Rohatgi, "Introduction to differential power analysis," *Journal of Cryptographic Engineering*, vol. 1, no. 1, pp. 5–27, 2011.

[115] K. Dhananjay and E. Salman, "Charge based power side-channel attack methodology for an adiabatic cipher," *Electronics*, vol. 10, no. 12, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/12/1438

[116] K. Dhananjay and E. Salman, "Special session: Adiabatic circuits for energy-efficient and secure iot systems," in *IEEE International Conference on Computer Design (ICCD)*, 2020, pp. 17–20.

[117] F. McKeen, I. Alexandrovich, A. Berenzon, C. V. Rozas, H. Shafi, V. Shanbhogue, and U. R. Savagaonkar, "Innovative instructions and software model for isolated execution." *Hasp@ isca*, vol. 10, no. 1, 2013.

[118] "Arm trustzone," https://developer.arm.com/ip-products/security-ip/trustzone.

[119] N. Kurd, M. Chowdhury, E. Burton, T. P. Thomas, C. Mozak, B. Boswell, P. Mosalikanti, M. Neidengard, A. Deval, A. Khanna *et al.*, "Haswell: A family of ia 22 nm processors," *IEEE Journal of Solid-State Circuits*, vol. 50,

no. 1, pp. 49–58, 2014.

[120] "Haswell - microarchitectures - intel." https://en.wikichip.org/wiki/intel/microarchitectures/haswell_(client), accessed: 2021-10-21.

[121] "Intel's haswell cpu microarchitecture." https://www.realworldtech.com/page/3/, accessed: 2021-10-12.

[122] T. E. Carlson, W. Heirman, S. Eyerman, I. Hur, and L. Eeckhout, "An evaluation of high-level mechanistic core models," *ACM Transactions on Architecture and Code Optimization (TACO)*, 2014.

[123] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, 2009, pp. 469–480.

[124] T. Rauber, G. Rünger, and M. Stachowski, "Performance and energy metrics for multi-threaded applications on dvfs processors," *Sustainable Computing: Informatics and Systems*, vol. 17, pp. 55–68, 2018.

[125] S. C. Woo, M. Ohara, E. Torrie, J. P. Singh, and A. Gupta, "The splash-2 programs: Characterization and methodological considerations," *ACM SIGARCH computer architecture news*, vol. 23, no. 2, pp. 24–36, 1995.

[126] Z. Yuan, P. Shukla, S. Chetoui, S. Nemtzow, S. Reda, and A. K. Coskun, "Pact: An extensible parallel thermal simulator for emerging integration and cooling technologies," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–1, 2021.

[127] C. Multiphysics, "Introduction to comsol multiphysics®," *COMSOL Multiphysics, Burlington, MA, accessed Feb*, vol. 9, p. 2018, 1998.

[128] G. H. Loh, Y. Xie, and B. Black, "Processor design in 3d die-stacking technologies," *Ieee Micro*, vol. 27, no. 3, pp. 31–48, 2007.

[129] X. Zhou, J. Yang, Y. Xu, Y. Zhang, and J. Zhao, "Thermal-aware task scheduling for 3d multicore processors," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 1, pp. 60–71, 2009.

[130] J. Meng, K. Kawakami, and A. K. Coskun, "Optimizing energy efficiency of 3-d multicore systems with stacked dram under power and thermal constraints," in *DAC Design Automation Conference 2012*. IEEE, 2012, pp. 648–655.

[131] S. K. Samal, D. Nayak, M. Ichihashi, S. Banna, and S. K. Lim, "Monolithic 3d ic vs. tsv-based 3d ic in 14nm finfet technology," in *2016 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*. IEEE, 2016, pp. 1–2.

[132] Y.-H. Gong, J. Kong, and S. W. Chung, "Quantifying the impact of monolithic 3d (m3d) integration on l1 caches," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 854–865, 2019.

[133] C. Yan and E. Salman, "Mono3d: Open source cell library for monolithic 3-d integrated circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 3, pp. 1075–1085, 2017.

[134] C. Yan, S. Kontak, H. Wang, and E. Salman, "Open source cell library mono3d to develop large-scale monolithic 3d integrated circuits," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1–4.

[135] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware microarchitecture," *ACM SIGARCH Computer Architecture News*, vol. 31, no. 2, pp. 2–13, 2003.

[136] "A comparison of intel's 32nm and 22nm core i5 cpus: Power, voltage, temperature, and frequency." http://blog.stuffedcow.net/2012/10/intel32nm-22nm-core-i5-comparison/, accessed: 2021-10-12.

[137] J.-J. Horng, S.-L. Liu, A. Kundu, C.-H. Chang, C.-H. Chen, H. Chiang, and Y.-C. Peng, "A 0.7v resistive sensor with temperature/voltage detection function in 16nm finfet technologies," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, 2014, pp. 1–2.

[138] T. Oshita, J. Shor, D. E. Duarte, A. Kornfeld, and D. Zilberman, "Compact bjt-based thermal sensor for processor applications in a 14 nm tri-gate cmos process," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 3, pp. 799–807, 2015.

[139] A. Córdoba, "Dirac combs," *letters in mathematical physics*, vol. 17, no. 3, pp. 191–196, 1989.

[140] M. Alagappan, J. Rajendran, M. Doroslovački, and G. Venkataramani, "Dfs covert channels on multi-core platforms," in *2017 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC)*. IEEE, 2017, pp. 1–6.