IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 30 April 2013; accepted 18 December 2013. Date of publication 2 February 2014; date of current version 7 May 2014.

Digital Object Identifier 10.1109/TETC.2014.2303934

On-Chip Nanoscale Capacitor Decoupling Architectures for Hardware Security

MATTHEW MAYHEW (Student Member, IEEE) AND RADU MURESAN (Member, IEEE)

School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada

CORRESPONDING AUTHOR: M. MAYHEW (mmayhew@uoguelph.ca)

ABSTRACT This paper presents new power analysis attack (PAA) countermeasures for nanoscale cryptographic devices. Specifically, three circuit level architectures called partial decoupling architecture, full decoupling architecture, and randomized switch box architecture are developed and analyzed. The architectures' primary feature is the use of on-chip nMOS gate capacitors as intermediate power storage elements to decouple the power supply from internal low-power modules processing sensitive data. The proposed countermeasures are algorithm independent and allow different tradeoffs between security protection and the incurred overheads. Test benches of the proposed architectures were simulated in 65-nm TSMC CMOS technology. A correlation PAA was performed for each test bench targeting a custom implementation of the advanced encryption standard subbytes operation. All architectures were found to resist the correlation PAA at the power supply, with the more complex architectures also offering protection against invasive attacks. The success value indicator was used to analyze the effectiveness of the countermeasures. It was found that all architectures provided a negative value at the power supply, showing protection against PAAs. We demonstrate that the use of nMOS gate capacitors can help to increase security and present a feasibility analysis focused on the needed decoupling capacitances.

INDEX TERMS Data encryption, hardware security countermeasures, power analysis attacks, security.

I. INTRODUCTION

Modern day embedded systems contain a variety of functional modules specialized for specific tasks. Due to the growing importance of data security as a design metric [1]; many systems include a cryptographic module in order to process secure data. While the inclusion of a dedicated cryptographic module serves to enhance data security, there are still ways for an attacker to compromise the security of the device by indirectly extracting the secret key used for encryption. Passive and non-invasive, or side-channel, attacks make use of leaked information obtained from real-time observation of a device as it processes secure data. Leakage targets for side-channel attacks include power consumption [2], [3], timing [4], and electromagnetic (EM) emanation [5]. These attacks are a threat to a variety of devices, such as smart cards, which implement cryptographic algorithms ranging from the simple Tiny Encryption Algorithm (TEA) to the more complex Advanced Encryption Standard (AES) algorithm. The AES algorithm is commonly used as a benchmark target for both attacks and countermeasures [3], [6]–[10]. The TEA is a lightweight cipher used for low-power systems such as Radio Frequency Identification (RFID) enabled devices and Wireless Sensor Node (WSN) network devices [11]–[13]. Furthermore, new research is being performed on side-channel attacks and countermeasures targeting complex platforms supporting cloud computing [14] and mobile computing [15].

This paper develops three circuit level architectures designed to provide protection against Power Analysis Attacks (PAA) [16] for modules handling secure data in small and large systems. PAA are a form of well-known side-channel attacks which exploit the power consumption of a device. The main technique employed by all three architectures is the use of on-chip capacitors to act as intermediate energy storage elements. The capacitors are used to decouple the power supply from internal modules handling sensitive data. The new proposed architectures in this paper are a partial decoupling architecture, a full decoupling architecture, and a randomized switch box architecture. Each architecture offers protection at the power supply pin while being simple to integrate into an existing design without the need to alter existing modules. The architectures allow a designer to balance the level of increased security with trade-offs in design complexity, area and power consumption overheads.

Section II of this paper briefly presents related background information including a brief overview of the Differential Power Analysis (DPA) attack and Correlation Power Analysis (CPA) attack procedures, explanations of the target cryptographic modules used in the simulated test benches, and existing countermeasures from literature. Section III demonstrates the vulnerability of the unprotected modules to the CPA attack. Section IV outlines the developed architectures. Section V presents the initial simulation results for all architectures from CPA attacks performed on both traces collected at the power supply and traces from the internal capacitors. Section VI contains our conclusions.

II. BACKGROUND

Statistical Power Analysis (StPA) attacks are a well-researched form of PAA. In a StPA, an attacker applies statistical evaluation methods to collected traces to determine the value of the secret key [2], [3], [17]. Both the DPA and CPA are StPA. In this section, brief outlines of the DPA and CPA are provided, followed by a description of two possible target algorithms and an overview of several related countermeasures from literature.

A. CPA AND DPA PROCEDURE

The DPA and CPA both consist of three main phases. In the first phase, an attacker collects a set of M_{i} power traces while the target device processes a set of I_i known inputs. While power traces are most commonly collected at the power supply pin [2], it is desirable to collect measurements as close as possible to the targeted module when dealing with larger systems containing multiple power and ground lines [18].

In the second phase, the attacker chooses an internal signal corresponding to a target operation. The chosen signal should be a function of both the known inputs, I_i, and the secret key k. As the internal signal is a function of the secret key, it will accurately correlate to the measured power consumption only when the correct value of k is guessed. As a result, all possible values of k must be guesssed when evaluating the internal signal.

While the DPA and CPA share the first two phases, the evaluation method used in the third phase differs. For the DPA, the attacker uses a selection function to partition the collected power traces based upon the guessed value of the targeted internal signal for each key guess. Selection functions vary from simply partitioning the traces based upon the Most Significant Bit (MSB) to custom functions meant for attacking specific systems. The selection function used by an attacker can strongly impact the effectiveness of the DPA [18]. After the partitioning has been performed, the average power consumption of each partition is then calculated using

$$A_D(j) = \frac{1}{N_D}, \sum\nolimits_{i \in S_D} PC_i(j); \quad D = 0, n-1 \tag{1}$$

where, A_D is the average power consumption of the partition, N_D is the number of traces in the partition, S_D is the set of traces belonging to the partition, and n is the total number of partition bins. The absolute difference between the average power consumption of two selected partitions is then calculated to produce a differential spike. To account for unknown implementation details, the partitions are evaluated at multiple time steps over the length of the collected traces, denoted by j. A large differential spike should be observed when the correct key value is guessed [2], [16].

While the CPA makes use of the same traces and internal signal in its evaluation phase, the procedure is different. The attacker first estimates the power dissipated by the circuit at the time the target operation is performed as shown in (2):

$$PC_i(j^*) = pc\,(g\,(I_i, k)) + error \tag{2}$$

where, j^* denotes the time step in which the targeted operation is performed, with $PC_i(j^*)$ representing the corresponding measured power consumption of the circuit, and $pc(g(I_i, k))$ is some power consumption model that is a function of the input I_i and the secret key k. While some error will be present, the model chosen should reflect the measured traces. Two common power models used to provide a quick and reasonable estimate for analysis are the Hamming Weight and Hamming Distance models, denoted by w_i [16]. Similar to how the DPA is influenced by the chosen selection function, the effectiveness of the CPA is partially determined by how well the actual power consumption of the device fits the chosen model [7].

A correlation coefficient $\rho_{WH}(j)$ between the measured power consumption and the calculated estimate is then computed for each guessed key value. Similar to the DPA, the correlation coefficient is calculated at multiple time steps to account for unknown implementation details. The calculation is performed using (3) [3], [17]:

$$\rho_{WH}(j) = \frac{N \sum W_i * PC_i(j) - \sum W_i \sum PC_i(j)}{\sqrt{N \sum W_i^2 - (\sum W_i)^2} * \sqrt{N \sum PC_i(j)^2 - (\sum PC_i(j))^2}} \tag{3}$$

where N denotes the total number of traces. In the case of a successful attack, a large spike should be observed for the correct key guess. While the CPA is typically more robust than the DPA [3], [17], similar countermeasures can be applied to reduce the effectiveness of both attacks [3]. For the purpose of this work, the CPA was used to attack the targeted modules.

B. ADVANCED ENCRYPTION STANDARD ALGORITHM

The AES algorithm supports key lengths of 128, 192, and 256 bits. In this work, the focus will be on the 128 bit implementation of the algorithm. The initial plaintext input is 128 bits, which is then represented by sets of 8-bit sub-blocks organized as a 4 X 4 matrix. The main algorithm begins with an initial XOR between the input plaintext and the secret key. This is then followed by ten rounds consisting of the following fundamental operations:

1) *AddRoundKey*: An XOR operation between the AES state and a designated round key. The key for each round is determined using the initial value of the secret key and the AES key schedule.
2) *SubBytes*: The 8-bit sub-blocks of the state are replaced according to a pre-determined look-up table. The operation is also commonly referred to as the "*Sbox*" operation.
3) *ShiftRows*: A left cyclical shift of the sub-blocks within a row. The positions shifted varies by row.
4) *MixColums*: A linear transformation is performed on one column of sub-blocks using (4):

$$\begin{bmatrix} S'_{0,C} \\ S'_{1,C} \\ S'_{2,C} \\ S'_{3,C} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} S_{0,C} \\ S_{1,C} \\ S_{2,C} \\ S_{3,C} \end{bmatrix} \quad (4)$$

Rounds one through nine of the algorithm are identical while round ten omits the *MixColumns* operation. For this work, two sensitive functional modules representing the *AddRoundKey* and *SubBytes* operations were created to implement the initial XOR between a plaintext input and the secret key followed by the *SubBytes* operation of the first round. Both of the modules are designed to operate on one eight-bit sub-block of the AES State.

The *AddRoundKey* module consists of two arrays of eight latches to hold the input values steady followed by an array of XOR gates to perform the operation. The *SubBytes* module contains a custom implementation of the Sbox used for the *SubBytes* operation. The implementation is organized like a memory unit, with each cell containing a set of pull-up and pull-down modules to generate the desired output value. The implementation is divided into four quadrants. The desired quadrant, row, and column are all determined based on the value of the input.

The *SubBytes* operation was chosen as the CPA target. A block diagram of the implemented modules may be seen in Fig. 1.

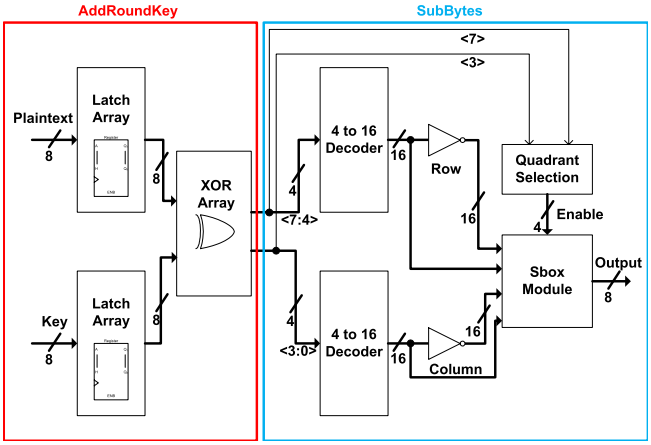

FIGURE 1. ***AddRoundKey* and *SubBytes* functional modules.**

C. TINY ENCRYPTION ALGORITHM

The TEA was designed with a focus on the development of a simple, lightweight encryption algorithm that could be easily implemented on a variety of low-power platforms. As the TEA uses only standard logic operations and lacks any pre-determined look-up tables, the algorithm has little set-up time [19]. The algorithm is designed to operate on 64-bit inputs with a 128-bit key.

While the number of encryption rounds used in the algorithm can be varied, a minimum of 16 rounds is suggested. This allows for single-bit changes in the plaintext or key to spread throughout the final result. In each individual round, the 64-bit input is split into two 32-bit halves, *Input_0* and *Input_1*, to better accommodate 32-bit data paths. A series of XOR, Add, and Shift operations are then applied to both halves to introduce non-linearity and mix the input and key bits. The datapath for a round of the TEA may be seen in Fig. 2 [19].

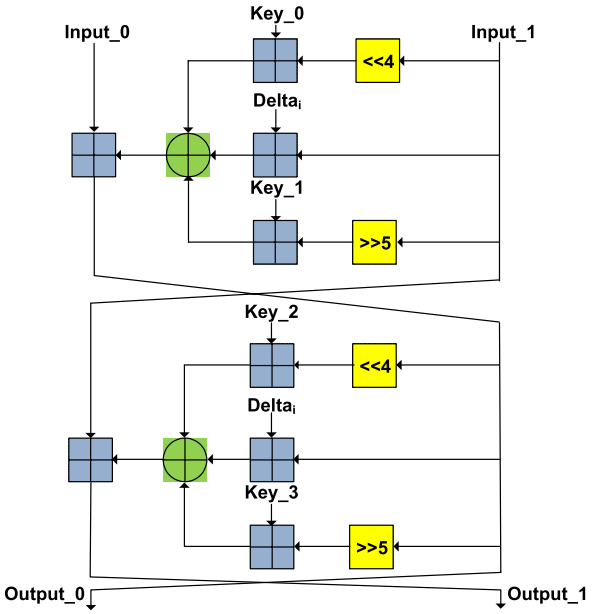

FIGURE 2. **TEA datapath.**

Key_0, *Key_1*, *Key_2*, and *Key_3* in Fig. 2 are obtained by splitting the secret key into four 32 bit sub-keys. To differentiate the rounds a running sum, denoted by $Delta_i$, is used. The value of $Delta_i$ is increased by adding a constant, *Delta*, to the sum before each round. While the value of *Delta* may be specified by the designer, a constant value of 0x9e3779b9 is used by convention.

D. PREVIOUS RESEARCH ON PAA COUNTERMEASURES

Countermeasures against PAA can be implemented in both hardware and software. The main strategies employed at the hardware level are decoupling, minimization, randomization, desynchronization, and noise insertion. In [20], Shamir proposes a capacitor based decoupling method for smart cards

using 2 capacitors of order 0.1 μF. An alternative method discussed in the patent is the use of a single capacitor for decoupling. The solution provided, however, requires support for smart card context saving and processor halting during the charging period of the single capacitor. The decoupling method in [20] might still be vulnerable to indirect attacks that monitor cumulative charge. As a result, [21] describes a voltage control method using a variable capacitor embedded into the smart card body. The integration of the variable capacitor proposed, however, can be challenging and bulky. In [22], Hubert develops a PAA countermeasure based on a current source and a buffer capacitor. The authors in [6] present a current equalizer that is implemented using an array of integrated switch capacitors that provide the supply current for the sensitive blocks of an AES engine. In [23], the authors propose a current-injection loop for eliminating the low-frequency current variations and a low-pass filter for removing high frequency current variations. In [24], a simplified decoupling architecture incorporating a switch box module is proposed. The architecture is presented in the context of a device's existing power management system. In [25], a current flattening technique is used to maintain a stable level of current consumption by sensing the current consumed by the system and injecting current as necessary.

In addition, other PAA countermeasures of varying complexity have been implemented at multiple levels of design abstraction. For example, in [10], a multiprocessor based countermeasure is used, in [26], an architecture with a reconfigurable functional unit is introduced, and in [27] a randomization module is developed. Generally, most hardware level countermeasures offer increased protection at the cost of area and power consumption trade-offs from the inclusion of additional modules or circuit elements.

PAA countermeasures at the gate and circuit level are another existing alternative. Gate level countermeasures typically rely on new logic families that attempt to build circuits with data-independent power consumption by balancing the power profile of the rising and falling transitions. Examples of balanced logic families include: Sense Amplifier Based Logic (SABL) [28], Wave Dynamic Differential Logic (WDDL) [9], dual-rail circuits [29], MOS Current Mode Logic (MCML) [30], and logic based on adiabatic and dual rail circuits [31], among others. Similar to hardware countermeasures, these altered logic styles offer an increase in security at the expense of increased area and power consumption. Approaches using customized cells also necessitate the use of either non-standard design flows or customized cell libraries in addition to any trade-offs inherent to the cells themselves.

III. ATTACKS ON UNPROTECTED MODULES

To confirm the vulnerability of the AES and TEA to the CPA, initial experiments were performed targeting unprotected versions of the modules shown in Figs. 1 and 2. The selected *AddRoundKey* and *SubBytes* modules were implemented for the AES algorithm while the full TEA datapath was created. All modules were developed in Cadence using 65 nm CMOS TSMC technology and simulated using the Cadence Analog Environment. Traces were collected directly at the power supply. A fixed key was used for all encryptions with randomly generated plaintext values. The results for the AES modules are presented first followed by the results for the TEA. To our knowledge, no previous CPA procedure has been presented in literature for the TEA.

A. UNPROTECTED AES MODULES

The *SubBytes* module was found to be highly susceptible to the CPA, with a strong positive correlation for the correct key value calculated using only 500 traces. The results are presented in Fig. 3. It can be seen that a high, clear spike is observed for the correct key value with no other competing guesses.

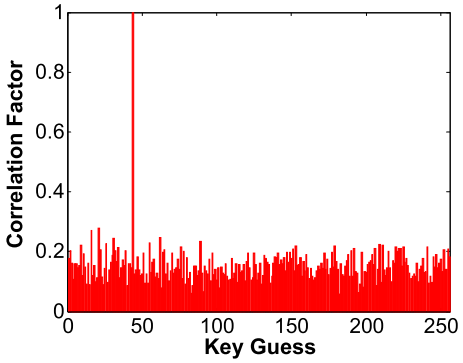

FIGURE 3. **CPA on unprotected *SubBytes* module.**

B. UNPROTECTED TEA DATAPATH

A slightly more complex procedure was needed to attack the TEA datapath due to the increased size of 32 bits for each key block. Due to the time needed to cycle through the collected traces, an exhaustive search of all 42,9496,7296 possibilities for a 32-bit key is impractical. As an alternative, it was found that by starting at the Least Significant Bit (LSB) and dealing with the key in 8-bit chunks, it was possible to conduct a feasible CPA using a relatively small number of traces. As the TEA datapath contains a number of 32-bit additions, more accurate knowledge of the lower bits allows for a better estimation of the higher bit values resulting from propagating carries. For cases where this approach does not produce a distinct spike for the correct value of the bits being evaluated, a set of candidate values is generally obtained. Using the candidate values, the procedure can then be repeated to obtain more accurate results. An overview of the general procedure used may be seen in Fig. 4, where g_0 contains bits 7 to 0, g_1 contains bits 15 to 8, g_2 contains bits 23 to 16, and g_3 contains bits 31 to 24.

Using the outlined approach, it was possible to perform a successful attack using 2000 traces. The initial results following the first round of the approach are shown in Fig. 5. The results for g_0, g_1, g_2, and g_3 are shown in Fig. 5(a)–(d) respectively. While the correct values were determined

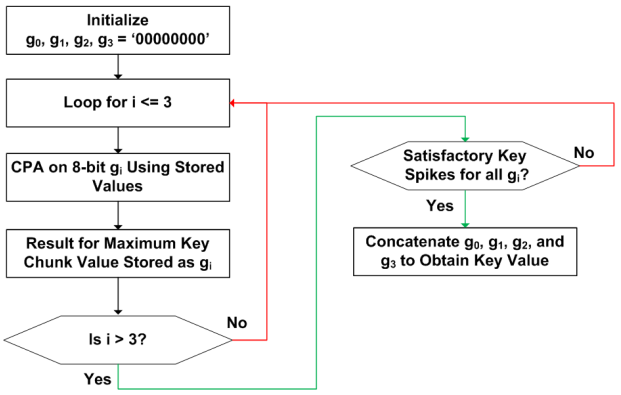

FIGURE 4. Generalized CPA procedure for unprotected TEA data path.

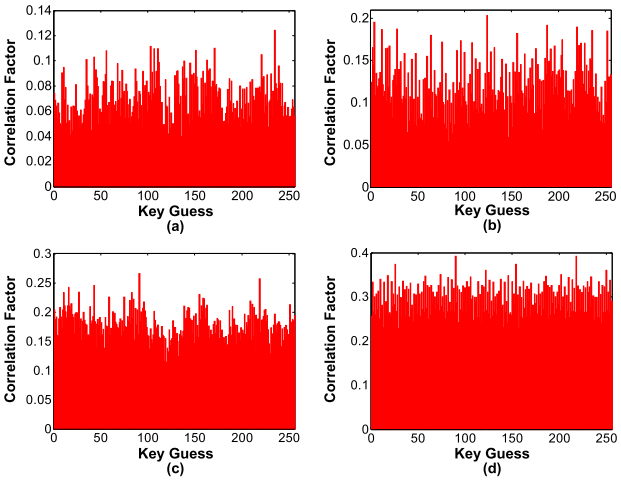

FIGURE 5. CPA results for the TEA datapath, first pass. (a) Results for g_0. (b) Results for g_1. (c) Results for g_2. (d) Results for g_3.

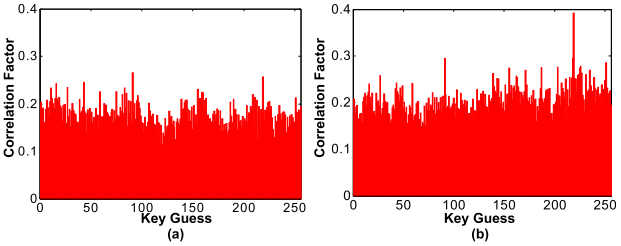

FIGURE 6. CPA results for g_2. (a) Results from first pass. (b) Results using correct candidate key value for g_3.

initially for g_0 and g_1, high competing peaks for g_2 and g_3 necessitated a second pass. The results from the initial pass for g_2 and the results when using the correct candidate value for g_3 are presented in Fig. 6(a) and (b) for comparison. A spike is observed for the correct value in Fig. 6(b), which also helps confirm the candidate value selected for g_3. From this, it can be seen that the proposed procedure allows for a CPA attack on a block of the TEA key while only searching a small portion of the key space.

Based on the presented results it can be seen that while a CPA attack is possible on both unprotected modules, more decisive results are possible with a simpler procedure and fewer traces for the AES *SubBytes* module. As the proposed protective architectures are not algorithm dependent, this work presents test benches and results using the more vulnerable *SubBytes* module.

IV. PROTECTIVE ARCHITECTURES

The main challenges for hardware based PAA countermeasures are increased chip area, increased power consumption, and increased design complexity. In this section, we present several on-chip PAA countermeasures which use capacitive elements to decouple the power supply from internal modules handling sensitive data. The proposed countermeasures incorporate the principles of decoupling, desynchronization, and randomization. Each countermeasure is also designed to be simple to implement and integrate with existing modules. The countermeasures may also be operated using a high decoupling frequency to reduce the needed capacitance values, allowing for the capacitive elements to be implemented using on-chip NMOS gate capacitors.

The proposed architectures are presented in order of increasing complexity, beginning with the partial decoupling architecture followed by the full decoupling architecture and finishing with the randomized switch box architecture. This section concludes with a feasibility analysis of the switches and decoupling capacitors needed by the proposed architectures.

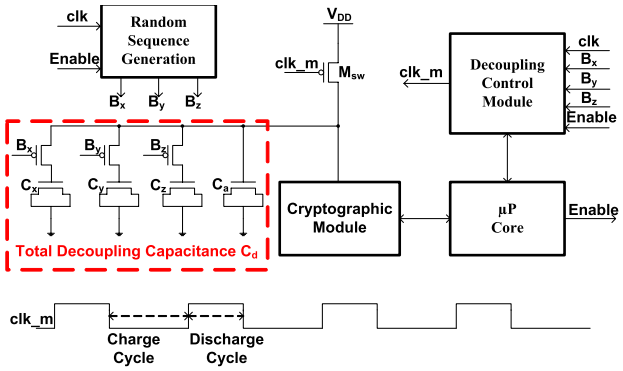

FIGURE 7. General partial decoupling architecture using NMOS gate capacitors.

A. PARTIAL DECOUPLING

A block diagram of the proposed partial decoupling architecture using NMOS gate capacitors is shown in Fig. 7. The main components of the proposed architecture are a protected cryptographic module, a switching PMOS transistor, M_{SW}, a total decoupling capacitance, C_d, and a decoupling control module.

The total decoupling capacitance may be a constant or randomized from cycle to cycle. If a randomized capacitance value is desired, a random sequence of bits, $B_xB_yB_z$, may be

generated such that:

$$C_d = B_x * C_x + B_y * C_y + B_z * C_z + C_a \quad (5)$$

where, C_a is a fixed capacitor to ensure uninterrupted operation of the cryptographic module and capacitors C_x, C_y, and C_z are used to generate a randomized capacitance value added to C_a. An enable signal may also be included for the purpose of saving power. When the countermeasure is enabled, the decoupling module will generate a periodic clock signal, *clk_m*, that controls M_{SW}. As *clk_m* is toggled, the system passes through alternating charge and discharge cycles. When the countermeasure is disabled, M_{SW} may be held either high or low. If the signal is held low, the cryptographic module will operate without decoupling. If the signal is held high, the cryptographic module will be disabled.

During the discharge phase of operation, *clk_m* is high and M_{SW} is turned off, decoupling the power supply from the functional module. In this phase, the cryptographic module is powered solely by the total decoupling capacitance C_d. Only a small amount of directly leaked information will be present at traces collected at the power supply pin when the system is in this state.

When *clk_m* becomes low, M_{SW} is turned on and the system enters the charge phase. The dominant feature of traces collected at this point will be the charging profile of the currently connected capacitance forming C_d. The act of randomly varying C_d serves to mix and hide any information regarding the data processed during previous discharge phases that may be indirectly leaked through uneven charging profiles.

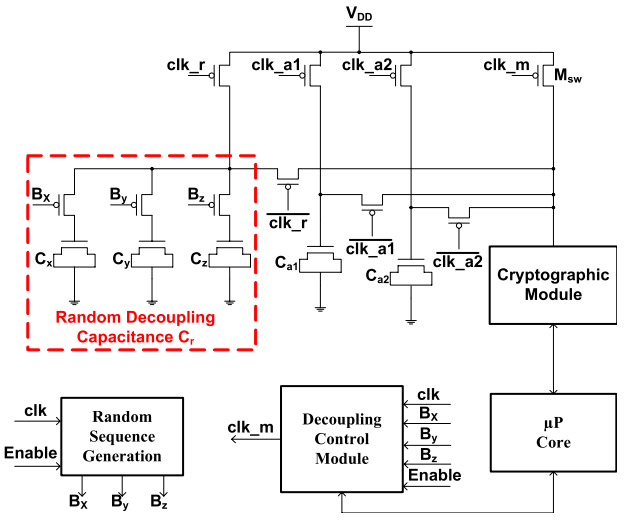

FIGURE 8. General full decoupling architecture using NMOS gate capacitors.

B. FULL DECOUPLING

The partial decoupling countermeasure presented can be modified to create a full decoupling architecture as shown in Fig. 8. While the full decoupling architecture incurs increased area overhead from additional capacitive elements and greater design complexity, it also offers increased data security by removing the direct connection between the power supply and cryptographic module previously present during the charge phase.

The full decoupling architecture contains two primary capacitors, C_{a1} and C_{a2}, which pass through complementary charge and discharge phases. This allows for the full decoupling of the cryptographic module from V_{DD}. The charging and discharging of C_{a1} and C_{a2} are handled by control signals *clk_a1* and *clk_a2*. Similar to the partial decoupling countermeasure, a small randomized capacitance, denoted by C_r, may also be included to distort any cumulative leaked information related to data processed during previous charge and discharge phases. The addition of C_r to C_{a1} or C_{a2} is managed by *clk_r*.

Note that *clk_a1* and *clk_a2* should share a short overlap period or *clk_m* should briefly establish a direct connection when switching between charge and discharge phases to support uninterrupted operation of the cryptographic module.

C. RANDOMIZED SWITCH BOX ARCHITECTURE

The last proposed architecture extends the principle of using multiple decoupling elements demonstrated by the full decoupling architecture. The randomized switch box architecture uses multiple on-chip decoupling capacitors as intermediate power storage elements. The capacitors are charged by the power supply and then in turn provide power to multiple functional modules. The capacitors are organized into multiple sets. While one set is charged by the power supply, another set provides power, decoupling the internal functional modules.

Connections between the capacitors and internal functional modules are made through a switch box element with randomized connections. Both cryptographic modules handling sensitive data and non-sensitive modules may be included in the switch-box's outputs. The inclusion of non-sensitive modules increases the number of possible pairings between capacitors and modules. This in turn adds switching noise to the system and decreases the probability that traces collected by an attacker at more invasive points, such as the decoupling power capacitor terminals, will contain leaked information related to the operations being performed by sensitive modules. A general block diagram of the architecture is shown in Fig. 9.

The general architecture presented contains a charge cycle control module responsible for handling connections. The roles of the generated signal groups (L, N, and M) are to connect capacitor sets to the power supply, discharge capacitors, and manage connections inside the switch box respectively. For the case presented, L corresponds to the number of desired capacitor sets, N corresponds to the number of individual capacitors, and M is determined by both the number of capacitors grouped in a set and the number of modules included as switch box outputs. Performing a fixed discharge of the capacitors serves to prevent the leakage of side-channel information from previous cycles through uneven charging profiles and charge sharing between capacitors [6]. Similar

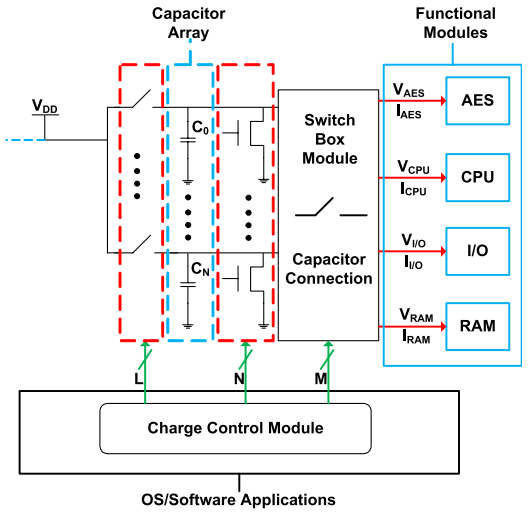

FIGURE 9. General randomized switch box architecture.

to the full decoupling architecture, a slight overlap should be in place when switching control signals to prevent a loss of power to the functional modules.

In practice, the functionality of the charge cycle control module may be handled at either the operating system level or through dedicated hardware. The switch box and multiple capacitors give this architecture the flexibility needed to implement multiple decoupling schemes.

D. FEASIBILITY ANALYSIS

In this section we consider the feasiblity of implementing the proposed decoupling architectures in the 65 nm TSMC CMOS process. For this analysis, a standard value of 1.0 V is assumed for V_{DD}. The practicality of the proposed countermeasures are related to the maximum, I_m, and average, I_a, current consumption of the cryptographic module to be protected, the size of the PMOS switches, and the implementation and size of the on-chip decoupling capacitive elements.

First, we address the size of the PMOS switches by deriving a width equation. One of the primary concerns is the voltage drop resulting from the equivalent on resistance of the PMOS transistors. The resistance of an individual switch can be modeled as [32]:

$$R_{SW} = 2R_u/W_{SW} \tag{6}$$

where, R_{SW} is the resistance of the PMOS switch, R_u is the unit resistance of the technology assuming a transistor with minimum dimensions, and W_{SW} is the width of the PMOS switch. Designing for a minimum voltage value of $V_{DDmin} = 0.8$ V, the maximum allowable voltage drop becomes $\Delta V_{SW} = 0.2$ V. The actual desired voltage drop, δ_V, however, should be lower to allow for a full charge on the decoupling capacitors (i.e. $\Delta V_{SW} = \delta_V << 0.2$ V.). The desired width of the switch can then be approximated by:

$$\frac{W_{SW}}{W_{MIN}} = \frac{(2R_u * I_m)}{\delta_V} \tag{7}$$

For ease of calculation, we consider the unit resistance $R_u = 10$ kΩ, which is a high margin for the 65 nm process. Furthermore, assuming $\delta_V = 0.2$ V and a maximum current value of $I_m = 1$ mA, the corresponding width of the switch can be calculated as 12 μm. Protecting small cryptographic modules with low peak power consumption will further reduce the size of the PMOS switch.

Secondly, we consider the implementation of the on-chip decoupling capacitive elements. While several options exist for implementing capacitors in conventional CMOS processes, NMOS capacitors have been chosen due to their comparitively high ratio of capacitance to area [32]. As we desire to keep the voltage range defined by V_{DD} and V_{DDmin} small, the non-linear gate controlled capacitance is not a primary concern. The slightly fluctuating capacitance, in actuality, helps to flatten the current consumption at the power supply pin in certain circumstances [21]. As the current consumption of connected modules varies, the voltage drop across the PMOS switches also changes accordingly. This in turn changes the voltage at the terminals of the NMOS gate capacitors, resulting in a fluctuating capacitance which helps flatten the current observed at the power supply pin. An example is shown in Fig. 10 for the partial decoupling system.

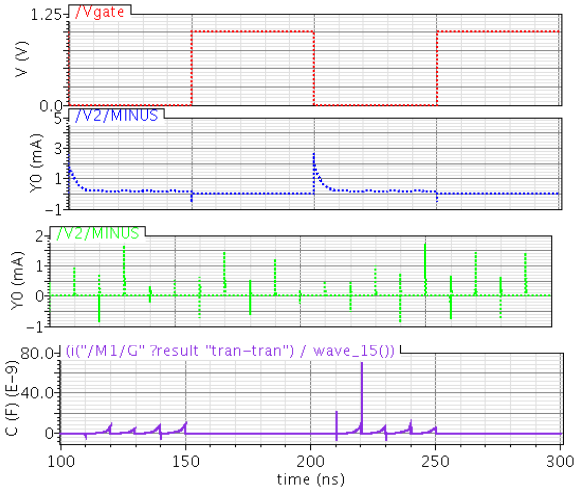

FIGURE 10. Waveforms for *clk_m*, V_{DD} flattened, V_{DD} standard, and C_d to show flattening effect of NMOS gate capacitors.

The waveforms in Fig. 10 show the switch control signal *clk_m*, the flattened current observed at the power supply with the NMOS gate capacitors in place, the current from an unprotected power supply, and the fluctuating capacitance resulting from the changing voltage level at the NMOS gate. Note that sharp peaks are observed for the unprotected power supply waveform.

To determine the size of the required decoupling capacitance, C_d, (8) can be used, where τ_d is the discharge time from V_{DD} to V_{DDmin}. As an example, assuming an even duty cycle and a target switching frequency of 10 MHz, τ_d would

be 50 ns.

$$C_d = \frac{I_a * \tau_d}{V_{DD} * \ln(\frac{V_{DD}}{V_{DDmin}})} \quad (8)$$

Assuming that the NMOS gate capacitors will be operating in the strong inversion region, their gate capacitance can be calculated as [32]:

$$C_d = C_g = C_{permicron} * W \quad (9)$$

where, C_g is the overall gate capacitance of the NMOS transistor, $C_{permicron}$ is the capacitance per micron of width with the minimum process length, and W is the gate width in microns.

As the *SubBytes* module was selected as the CPA target, initial simulations were performed to examine the module's expected current consumption. It was found that a value of $I_a = 0.1$ mA was sufficient to account for both the average current consumption of the module and the transient spikes observed on transitions. Using (8) with a target decoupling frequency of 10 MHz for the clock controlling the charge and discharge phases and assuming an even duty cycle, a required capacitance value of 22 pF was calculated for each primary decoupling capacitor. Based on (9) with a value of 1 fF/μm for $C_{permicron}$ and a minimum process length of 65 nm, the corresponding area is a square NMOS structure of 37.8 μm X 37.8 μm. To account for the process limits on width and length, the dimensions were adjusted to 72 μm X 20 μm.

Based on the experimental calculations and considering (7)–(9), it can be seen that the proposed countermeasure architectures are best suited to protecting individual cryptographic modules with low average current consumption. This reduces the size of both the needed switches and decoupling capacitances. The size of the decoupling capacitors may also be reduced by using a fast switching frequency to decrease τ_d. Note that when dealing with a large module exhibiting high current consumption, the randomized switch box architecture could be used to isolate individual operations as sub-modules for protection to reduce the needed decoupling capacitance.

V. RESULTS

In order to validate the proposed system architectures, initial results were obtained through simulation. Test benches were developed for each architecture protecting the vulnerable *SubBytes* module. This section first provides a brief outline of the simulated test benches followed by an analysis of the collected traces. A brief summary of the results concludes the section.

To evaluate the performance of each proposed architecture, the Success Value Indicator (SVI) proposed in [25] is used. The SVI is calculated using (10), where i_{pc} is the height of the spike corresponding to the correct value of the secret key and i_{pw} is the maximum spike for all other possible, incorrect key values. Note that a negative SVI value indicates an unsuccessful attack.

$$SVI = i_{pc} - i_{pw} \quad (10)$$

A. ARCHITECTURE TEST BENCHES

All test bench systems were simulated using a decoupling frequency of 10 MHz, as outlined in the feasibility analysis. Traces were collected at the power supply, V_{DD}, as well as the capacitor terminals to account for the possibility of more invasive attacks. A set of 2000 known plaintexts were processed for each test bench with a fixed key value. The same inputs were used for all test benches to provide an equal comparison. The *SubBytes* module was run at a frequency of 100 MHz to process multiple inputs per charge/discharge cycle. A fixed strobe period was used to generate exact results at each sampling point.

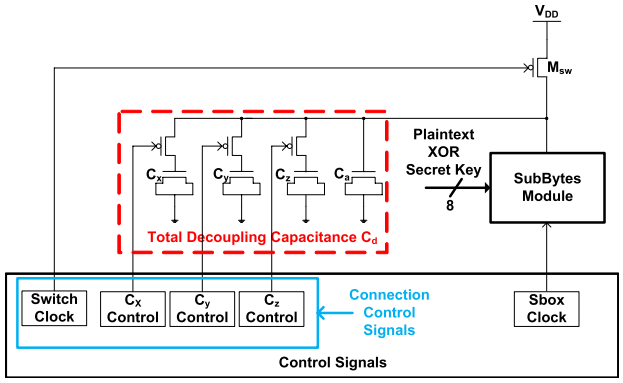

FIGURE 11. Test bench system for proposed partial decoupling architecture.

The test benches used for the partial and full decoupling architectures are similar to the general architectures shown in Figs. 7 and 8 respectively. The test bench for the partial decoupling architecture is shown in Fig. 11 as an example. The control signals for the random capacitors, C_x_*Control*, C_y_*Control*, and C_z_*Control,* were randomly generated. The inputs provided to the *SubBytes* module were the results of a software XOR between the initially generated plaintexts and the chosen key value. The test bench for the full decoupling architecture was configured in an identical fashion.

The test bench for the randomized switch box architecture is slightly more complex. The test bench contains a total of eight decoupling capacitors numbered from C_0 to C_7 and four functional modules. The capacitors are grouped into two sets (C_0 to C_3 and C_4 to C_7). The sets alternate between being charged and providing power to the functional modules. In the set providing power, one capacitor is connected to each individual functional module through the switch box. A block diagram of the switch box architecture test bench is presented in Fig. 12.

Two of the functional modules, *AddRoundKey* and *SubBytes,* handle sensitive information. The initial plaintext and key values are provided to the *AddRoundKey* module which creates the inputs for the *SubBytes* module as shown previously in Fig. 1. *Load1* and *Load2* are simple switching loads meant to represent the inclusion of modules handling non-sensitive information among the switch box's outputs.

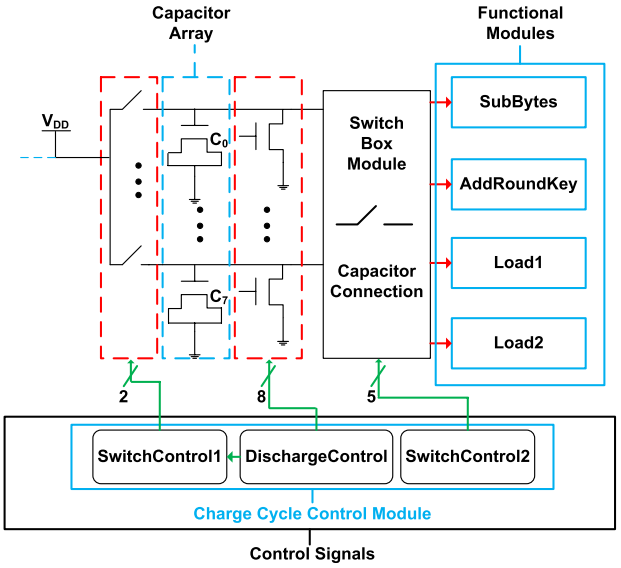

FIGURE 12. Test bench system for proposed randomized switch box architecture.

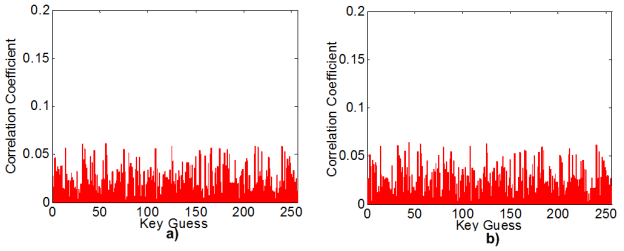

FIGURE 13. CPA results for partial decoupling test bench. (a) Power supply. (b) Terminal of C_a, NMOS gate capacitors.

Periodic switching inputs are used to generate activity on the loads over the simulation run-time.

Connections in the randomized switch box architecture test bench are managed by a charge cycle control module through three distinct signal groups identified as *SwitchControl1*, *SwitchControl2*, and *DischargeControl*. *SwitchControl2* controls the connections made between individual capacitors and functional modules through the switch box module. *DischargeControl* is used to introduce a short discharge period when switching between capacitor sets. Lastly, *SwitchControl1* is used to toggle between the capacitor set powering the functional modules and the capacitor set being charged by the power supply. Note that the *DischargeControl* signal and discharge phase may be omitted if desired to remove some power overhead from the system.

B. PARTIAL DECOUPLING RESULTS

The results at the power supply pin are considered first. It was found that the partial decoupling architecture was able to defeat the CPA when attacked at the power supply pin using the 2000 collected traces. If performing a more invasive attack and assuming access to the on-chip capacitors, however, we found that an attack was still possible. At 1900 traces, the correct key value was briefly determined. The CPA results at the power supply and capacitor terminal are presented in Fig. 13(a) and (b) respectively. Note that the spike for the correct key observed at the capacitor terminal is still very small, which indicates that the noise from a physical implementation may help obscure the correct value.

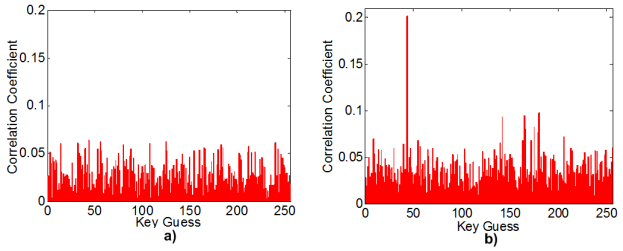

FIGURE 14. CPA results for partial decoupling test bench at capacitor C_a. (a) NMOS gate capacitors. (b) Standard capacitive elements.

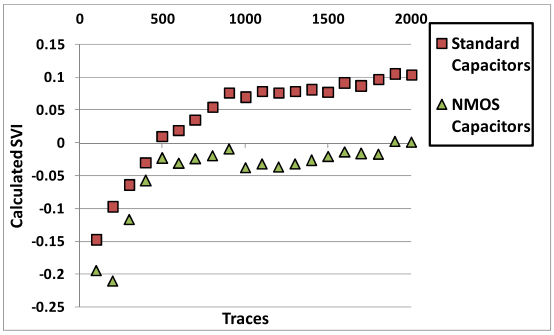

FIGURE 15. Calculated SVI values as traces are varied at the terminal of C_a for partial decoupling test bench.

In order to observe the benefits of the fluctuating capacitance value of the NMOS gate capacitors, an additional simulation was performed for the partial decoupling test bench using standard capacitive elements. It was found that the invasive attack at the terminal of C_a was more successful, with a noted spike observed for the correct key value using 2000 traces. The standard capacitors were consistently found to be more vulnerable than the NMOS gate capacitors for this architecture as the number of traces used in the analysis was varied. A comparison between the CPA results using NMOS gate capacitors and standard capacitive elements is shown in Fig. 14. The corresponding SVI chart comparing the two capacitor types as the number of traces used in the analysis is varied is shown in Fig. 15.

Overall, the proposed partial decoupling architecture provides a low complexity and low overhead means of adding PAA resistance, successfully defending against a CPA attack performed at the power supply. If more invasive attacks are possible, the partial decoupling architecture still exhibits a potential vulnerability at the capacitor terminal due to the

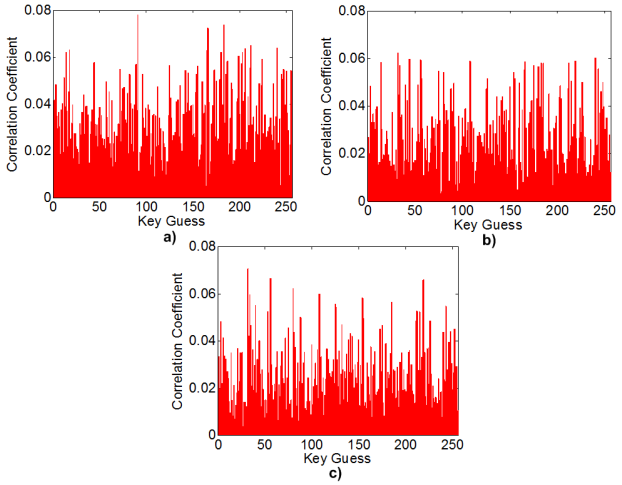

FIGURE 16. CPA results for the full decoupling test bench. (a) Results at power supply. (b) Results at terminal of capacitor C_{a1}. (c) Results at terminal of capacitor C_{a2}.

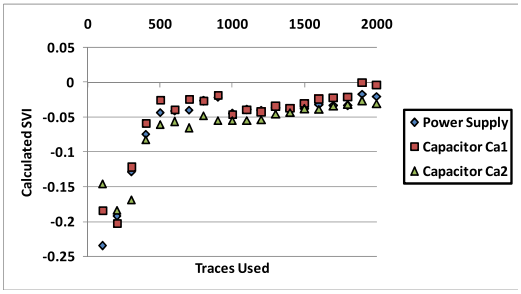

FIGURE 17. Calculated SVI values as traces are varied for the full decoupling test bench.

presence of only a single decoupling element. The full decoupling and switch box architectures, however, address this point at the cost of increased design complexity and overhead.

C. FULL DECOUPLING RESULTS

The results obtained for the full decoupling architecture, were, in general, improved from those observed for the partial decoupling architecture. The CPA at the power supply pin was prevented with a greater negative SVI calculated, indicating increased data security. The resistance to more invasive attacks targeting the terminals of the on-chip decoupling capacitors was improved as well. While a CPA was still possible at the terminal of C_{a1} at 1900 traces, the CPA was not successful at C_{a2}. The CPA results at C_{a2} were found to show resistance largely equivalent to that observed at the power supply. The CPA results for the full decoupling architecture using 2000 traces may be seen in Fig. 16. The results at the power supply, the terminal of C_{a1}, and the terminal of C_{a2} are shown in Fig. 16(a)–(c) respectively. The calculated SVI as traces are varied is presented in Fig. 17 for all three attacked points.

Overall, the full decoupling architecture offers greater data security at the cost of a slight increase in complexity and area overhead from the additional NMOS gate capacitor. The results also indicate that the addition of a larger number of on-chip decoupling capacitors with more frequently adjusted connections helps to reduce the effectiveness of more invasive attacks conducted at the capacitor terminals. This observation is consistent with the results presented in the following section for the randomized switch box architecture.

D. RANDOMIZED SWITCH BOX RESULTS

As an initial note, it was determined through experimentation that the decoupling capacitance values could be reduced to 15 pF for the randomized switch box test bench. To account for the process limits, each capacitor was sized to be 48.75 μm X 20 μm.

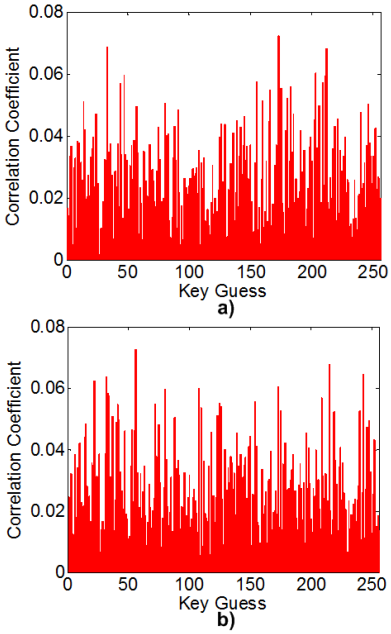

FIGURE 18. CPA results for the randomized switch box test bench. (a) Results at power supply. (b) Results at terminal of capacitor C_4.

Similar to the partial and full decoupling architectures, the CPA was unsuccessful using the 2000 traces collected at the power supply. An improvement in security at the decoupling capacitors was also noted compared to both of the previous simulated architectures. While the CPA was previously successful at some of the capacitor terminals for the partial and full decoupling test benches, the randomized switch box architecture resisted the CPA at all capacitor terminals. The additional decorrelation added from the shuffled connections decreases the probability that traces collected at the terminal of an individual capacitor will contain information related to sensitive data. For brevity and clarity, results are only presented for selected capacitors. The CPA results for both the power supply and the terminal of capacitor C_4 are presented in Fig. 18. It can be seen that the calculated correlation coefficients are fairly scattered at both attacked points with no strong individual spike.

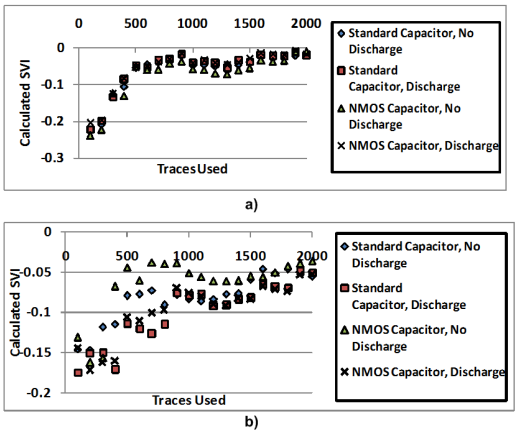

FIGURE 19. Calculated SVI for varying traces for randomized switch box test bench. (a) Power supply. (b) Terminal of capacitor C_4.

The next aspect considered was the effect of the fixed discharge at 10 MHz. While some difference between the results with and without the fixed discharge were noted, the effect of the more frequently changed connections was more dominant. Additionally, it was also found that there was not a sharp difference between the test bench implemented using NMOS gate capacitors and the test bench implemented using standard capacitive elements. The SVI trends comparing the effects of the fixed discharge and the capacitor type are presented in Fig. 19. The results at the power supply and at the terminal of C_4 are shown Fig. 19(a) and (b) respectively. Similar SVI trends were observed at the terminals of the other decoupling capacitors.

E. RESULTS SUMMARY

Several key comparisons between the results obtained for the proposed architectures are outlined in Table 1. We focus on the area overhead associated with the decoupling capacitors, the SVI for each architecture at the power supply, and the peak SVI observed for an invasive attack at the decoupling capacitor terminals. The values in Table 1 assume NMOS gate capacitors and the use of all 2000 collected traces in the CPA attack.

TABLE 1. Results summary using 2000 traces.

Architecture Type	Individual Capacitor Size (pF)	Number of Capacitors	Capacitor Area @ 10 MHz (μm^2)	SVI at Power Supply	Peak SVI at Capacitors
Partial Decoupling	22	1	1428.84	-0.0069	-0.0016
Full Decoupling	22	2	2857.68	-0.0202	-0.0027
Randomized Switch Box	15	8	7800.00	-0.0156	-0.0149

It can be seen that both the full decoupling architecture and randomized switch box architecture offer noticeably increased data security at the power supply pin compared to the partial decoupling architecture. Based on the SVI, security benefits of approximately 65% and 55% were achieved when compared to the partial decoupling architecture. As a trade-off, the area required for the decoupling NMOS gate capacitors also increased by approximately 50% and 81% respectively.

In addition, it was noted that increasing the number of possible pairings between the decoupling NMOS gate capacitors and functional modules helped to reduce the effectiveness of more invasive attacks. An improvement of approximately 80% was achieved for the randomized switch box architecture over both the partial and full decoupling architectures in this category. As an additional consideration, the randomized switch box architecture also allows a designer to split a module into smaller, protected sub-operations. This allows the countermeasure to be applied to larger modules with higher current consumption.

VI. CONCLUSION

This paper has presented three on-chip nanoscale circuit level architectures meant to increase data security to PAA. All of the architectures presented are designed to be algorithm independent. As PAA have been shown to be effective on both algorithms designed for constrained platforms, such as the TEA, to the more widely adopted AES, the proposed architectures could be included in a variety of platforms as a low complexity means of adding data security. The architectures presented consist of a partial decoupling architecture, a full decoupling architecture, and a randomized switch box architecture. Each architecture uses NMOS gate capacitors as decoupling elements to isolate internal modules handling sensitive data from the external power supply commonly targeted by PAA.

Based on the feasibility analysis it can be seen that the proposed architectures are best suited to protecting individual modules with low current consumption to reduce the size of the decoupling capacitors. The different architectures allow a designer to adjust the trade-off between the added security and the incurred overheads depending on the target platform. Based on the calculated SVI values, the full decoupling architecture and randomized switch box architecture offered data security improvements of approximately 65% and 55% when compared to the partial decoupling architecture, with corresponding capacitor area increases of 50% and 81% respectively. In regards to more invasive attacks, the randomized switch box architecture was found to be far more secure, with SVI improvements of over 80% compared to the other two architectures. It was also the only architecture to prevent the CPA at all capacitor terminals.

With this in mind, the partial decoupling architecture is likely best suited for systems with strict area constraints. Otherwise, the full decoupling architecture offers large security benefits with comparitively low additional area overhead and design complexity. Should more invasive PAA be a concern, the randomized switch box architecture should be used. It should also be noted that the randomized switch box architecture allows a designer greater flexibility as multiple sub-modules may be included in the switch box outputs.

REFERENCES

[1] P. Kocher, R. Lee, G. McGraw, A. Raghunathan, and S. Ravi, "Security as a new dimension in embedded system design," in *Proc. Des. Autom. Conf.*, 2004, pp. 753–760.

[2] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology-CRYPTO*. Berlin, Germany: Springer-Verlag, Jan. 1999, pp. 388–397.

[3] M. Alioto, M. Polie, and S. Rocchi, "Power analysis attacks to cryptographic circuits: A comparitive analysis of DPA and CPA," in *Proc. ICM*, Dec. 2008, pp. 333–336.

[4] P. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Proc. Adv. Cryptol.*, 1996, pp. 104–113.

[5] K. Gandolfi, C. Mourter, and F. Oliver, "Electromagnetic analysis: Concrete results," in *CHES LNCS*, Berlin, Germany: Springer-Verlag, 2001, pp. 251–261.

[6] C. Tokunaga and D. Blaauw, "Security encryption systems with a switched capacitor current equalizer," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 23–31, Jan. 2010.

[7] F.-X. Standaert, B. Gierlichs, and I. Verbauwhede, "Partition vs. Comparison side-channel distinguishers: An empirical evaluation of statistical tests for univariate side-channel attacks against two unprotected CMOS Devices," in *Proc. ICISC*, 2008, pp. 253–267.

[8] E. Oswald, S. Mangard, C. Herbst, and S. Tillich, "Practical second-order DPA attacks for masked smart card implementations of block ciphers," in *Proc. Cryptograph. Track RAS Conf.*, 2006, pp. 192–207.

[9] D. D. Hwang, K. J. V. Tiri, A. Hodjat, L. Bo-Cheng, Y. Shengline, P. Schaumont, *et al.*, "AES-based security coprocessor IC in 0.18 μm CMOS with resistance to differential power analysis side-channel attacks," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 781–792, Apr. 2006.

[10] J. A. Ambrose, R. G. Ragel, S. Parameswaran, and A. Ignjatovic, "Multiprocessor information concealment architecture to prevent power analysis based side channel attacks," *IET Comput. Digital Tech.*, vol. 5, no. 1, pp. 1–15, Jan. 2011.

[11] P. Israsena, "Design and implementation of low power hardware encryption for low cost secure RFID using TEA," in *Proc. 5th Int. Conf. Inf., Commun. Signal Process.*, 2005, pp. 1401–1406.

[12] M. B. Abdelhalim, M. El-Mahallawy, M. Ayyad, and A. Elhen-nawy, "Implementation of a modified lightweight cryptographic TEA algorithm in RFID system," in *Proc. ICITST*, Dec. 2011, pp. 509–513.

[13] J.-P. Kaps, "Chai-tea, cryptographic hardware implementation of xTEA," in *Proc. 9th Int. Conf. Cryptol.*, 2008, pp. 363–375.

[14] Y. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Cross-VM side channels and their use to extract private keys," in *Proc. ACM Conf. Comput. Commun. Sec.*, 2012, pp. 305–316.

[15] J.-W. Lee, S.-C. Chung, H.-C. Chang, and C.-Y. Lee, "Efficient power-analysis resistant dual-field elliptic curve cryptographic processor using heterogeneous dual-processing-element architecture," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 22, no. 1, pp. 49–61, Jan. 2014.

[16] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks, Revealing the Secrets of Smart Cards*. New York, NY, Springer-Verlag, 2007.

[17] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Proc. CHES 6th Int. Workshop*, 2004, pp. 16–29.

[18] P. Kocher, J. Jaffe, B. Jun, and P. Rohatgi, "Introduction to differential power analysis," *J. Cryptograph. Eng.*, vol. 1, no. 1, pp. 5–27, Mar. 2011.

[19] D. J. Wheeler and R. M. Needham, "TEA, a tiny encryption algorithm," in *Proc. 2nd Int. Workshop Fast Softw. Encryption*, 1994, pp. 363–366.

[20] A. Shamir, "Protecting smart cards from power analysis with detachable power supplies," U.S. Patent 6 507 913, Jan. 14, 2003.

[21] K. Seo-Kyu, "Smart cards having protection circuits therein that inhibit power analysis attacks and methods of operating same," U.S. Patent 7 620 823, Nov. 17, 2009.

[22] G. T. M. Hubert, "Current source for cryptographic processor," U.S. Patent 7 571 491, Aug. 4, 2009.

[23] G. B. Rantapal, R. D. Williams, and T. N. Blalock, "An on-chip signal suppression countermeasure to power analysis attacks," *IEEE Trans. Dependable Secure Comput.*, vol. 1, no. 3, pp. 179–189, Jul./Sep. 2004.

[24] M. Mayhew and R. Muresan, "Integrated capacitor switchbox for security protection," in *Proc. IEEE ISCAS*, May 2012, pp. 1452–1455.

[25] R. Muresan and S. Gregori, "Protection circuit against differential power analysis attacks for smart cards," *IEEE Trans. Comput.*, vol. 57, no. 11, pp. 1540–1549, Nov. 2008.

[26] S. A. Seyyedi, M. Kama, H. Noori, and S. Safari, "Securing Embedded processors against power analysis based side channel attacks using reconfigurable architecture," in *Proc. EUC*, 2011, pp. 255–260.

[27] D. B. Shu, L.-W. Chow, and W. Clark, "Cryptographic architecture with instruction masking and other techniques for thwarting differential power analysis," U.S. Patent 8 095 993, Jun. 2, 2012.

[28] I. M. Verbauwhede and K. J. V. Tiri, "Dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis," U.S. Patent 7 417 468, Feb. 12, 2009.

[29] D. Sokolov, J. Murphy, A. Bystrov, and Y. Yakovlev, "Design and analysis of dual-rail circuits for security applications," *IEEE Trans. Comput.*, vol. 54, no. 4, pp. 449–460, Apr. 2005.

[30] Z. Toprak and Y. Leblebici, "Low-power current mode logic for improved DPA-resistance in embedded systems," in *Proc. IEEE ISCAS*, May 2005, pp. 1059–1062.

[31] P. K. Sana and M. Satyam, "An energy efficient secure logic to provide resistance against differential power analysis attacks," in *Proc. ISED*, 2010, pp. 61–65.

[32] N. H. E. Weste and D. M. Harris, *CMOS VLSI Design, A Circuits and Systems Perspective*, 4th ed. Reading, MA, USA, Addison Wesley, 2011.

MATTHEW MAYHEW (S'12) received his M.A.Sc. and B.S. degrees in engineering systems and computing from the University of Guelph, Canada, in 2009 and 2007 respectively. He is currently a PhD candidate in engineering systems and computing at the University of Guelph. His research interests include integrated circuit design, cryptography, and the development of cryptographic circuits that resist power analysis attacks.

RADU MURESAN (M'03) received both his Ph.D. and M.A.Sc. degrees in electrical and computer engineering from the University of Waterloo, Canada. Currently he is an associate professor with the School of Engineering at the University of Guelph, Canada. His research interests include system-on-chip design, real-time embedded systems, the power aspects of cryptographic implementations at system and circuit level, and the development of hardware countermeasures for cryptographic devices against power analysis attacks.