Using Emerging Technologies for Hardware Security Beyond PUFs

An Chen, X. Sharon Hu, Yier Jin, Michael Niemier, Xunzhao Yin

(1) ITRS ERD, (2) University of Notre Dame, (3) University of Central Florida

Abstract—We discuss how the unique I-V characteristics offered by emerging, post-CMOS transistors can be used to enhance hardware security. Different from most existing work that exploits emerging technologies for hardware security, we (i) focus on transistor characteristics that either do not exist in MOSFETs, or are difficult to duplicate with MOSFETs, and (ii) aim to move beyond hardware implementations of physically unclonable functions (PUFs) and random number generators (RNGs).

I. INTRODUCTION

Like performance, power, and reliability, hardware security is becoming a critical design consideration. Hardware security threats in the IC supply chain, include counterfeiting of semiconductor components, side-channel attacks, invasive/semi-invasive reverse engineering, and IP piracy. A rapid growth in the “Internet of Things” (IoT) only exacerbates problems. While hardware security enhancements and circuit protection methods can mitigate security threats in protected components, they often incur a high cost with respect to performance, power and/or cost. Raising the resilience of hardware systems with minimal compromise to other metrics is a daunting challenge.

Advances in emerging, post-CMOS technologies may provide hardware security researchers with new opportunities to change the passive role that CMOS technology currently plays in security applications. While many emerging technologies aim to sustain Moore’s Law-based performance scaling and/or to improve energy efficiency [1], emerging technologies also demonstrate unique features that could drastically simplify circuit structures for protection against hardware security threats. Security applications could not only benefit from the non-traditional I-V characteristics of some emerging devices, but also help shape research at the device level by raising security measures to the level of other design metrics.

At present, most emerging technologies being studied in the context of hardware security applications are related to designing physically unclonable functions (PUFs). Post-CMOS devices such as domain-wall memories (DWM) [2], memristors [3], carbon nanotubes (CNTs) [4], etc. have all been suggested as a pathway to a PUF design. While intriguing, these approaches (i) only cover a small part of the hardware security landscape, and (ii) PUF designs often depend on device characteristics that a designer would like to eliminate when considering utility for logic or memory (e.g., pinning in domain wall memory [2]). Given the many emerging devices being studied [1] and that few if any devices were proposed with hardware security as a “killer application”, we explore how the unique I-V characteristics of emerging transistors that are not found in traditional MOSFETs could benefit hardware security applications.

The first category of I-V characteristics of interest (Sec. III) are those exhibiting tunable polarity, which appear in devices such as carbon nanotube [5], graphene [6], silicon nanowire (SiNW) transistors [7], and most recently transition metal dichalcogenide (TMD) tunnel FETs (TFETs) [8], all of which have already been fabricated experimentally. Tunable polarity could be exploited for IP protection and hence would help prevent the counterfeiting of IC components.

The second category of I-V characteristics (Sec. IV) can be broadly classified as devices with atypical switching behaviors. We consider I-V curves that either have tunable hysteresis or are bell shaped. Devices such as the negative capacitance FETs (NCFETs) [9] and ionic FETs [10] display tunable hysteresis behavior. Bell-shaped I-V characteristics have been observed experimentally in double-layer graphene FETs [11], [12] and ThinTFETs [13]. Such atypical switching behaviors are candidates for implementing novel circuits to prevent tampering via power supply or other side-channel attacks.

Specific security threats addressed in this paper include IC counterfeiting, side channel attacks (e.g., differential power analysis), memory leakage, and unauthorized access.

II. BACKGROUND

Here, we review hardware security needs and challenges, and the transistor technologies that form the basis of our work.

A. Hardware security needs and challenges

To reduce design costs and increase profits, IC manufacturers are continuing to outsource low-profit services (manufacturing, assembling, etc.) to offshore vendors. Only higher profit endeavors (design, service, etc.) are likely to remain state-side. Ironically, while globalization has helped to reduce total cost, it has exacerbated security concerns. Below, we briefly review areas in which emerging technologies might help to alleviate security concerns/threats.

Hardware fingerprinting and authentication can protect hardware intellectual property (IP) cores against reverse engineering toolsets [14]. Hardware authentication can create a piracy-proof design flow, where only the authorized end-user can activate IP designs. Researchers have proposed using PUFs for the authentication process [15]. The challenge-response pair based protocols are verified between the manufacturer and end-users, limiting utility of the device to the authentic user. However, modeling attack methods have been developed to predict the PUF responses that diminishes the security level of PUF-based authentication [16].

Camouflaging [17] relies on layout-level obfuscation that makes it difficult to decipher a circuit’s structure via reverse
engineering [18]. However, the overhead of CMOS camouflaging gates is often significant—especially as the level of protection increases. (A XOR+NAND+NOR camouflaging gate has 5.1X-5.5X higher power, 1.1X-1.6X higher delay, and 4X higher area compared to a conventional NAND or NOR gate [18].) Furthermore, SAT-based methods have shown to be able to “de-camouflage” a circuit under protection in minutes [19]. Design-level obfuscation or logic encryption is the other well-studied solution that could prevent attackers from easily recovering/reproducing circuit designs without the authentication key [20]. While these methods have proven to be robust to attacks (IP piracy could only occur if attackers know both the netlist and the keys), performance overhead and layout re-design present significant challenges.

Counterfeit ICs—i.e., recycled, remarked, cloned, tampered, overproduced, or out-of-spec integrated circuits—have recently found their way into safety-critical and military applications [21]. Solutions for detecting counterfeit products are limited. While PUFs and aging sensors have been proposed as solutions [21], drawbacks include high power and area costs.

Other threats include side-channel analysis and fault injections. Without physical intrusion, attackers can recover internal signals leveraging static/differential analyses on side channels such as timing, power consumption, and electromagnetic emissions. Cryptographic circuits are also vulnerable to power supply-based fault injections. To counter these attacks, researchers have developed various logic circuitry and on-chip sensors to balance the side-channel signals and to detect signal anomalies [22]. However, even with design optimization methods, the existing MOSFET based countermeasures still incur high performance overhead [23].

B. Device characteristics of interest

In Fig. 1 we present an initial mapping from post-CMOS devices to hardware security needs. The bottom-level identifies device technologies of interest, the next level up illustrates unique I-V characteristics that are specific to particular devices, the next level up details what security-centric hardware primitives may be enabled by said I-V characteristics, while the top level indicates what security-centric need and/or threat a given hardware primitive might address. Per Fig. 1, a given device may ultimately address more than one security need or threat. Furthermore, in addition to addressing typical needs/threats, new transistor technologies also introduce other “added value,” e.g., in the form of ultra low-power lightweight ciphers to support IoT. While a detailed discussion is beyond the scope of this paper, below we introduce the I-V characteristics and the devices that form the basis of this paper. We emphasize designs based on device technologies that exhibit tunable tunable polarity, “bell-shaped” I-V curves, and hysteresis. Device specific paths to said characteristics are described in Secs. II-B1, II-B2, and II-B3 respectively.

1) Tunable Polarity: In many nanoscale FETs (45nm and below), the superposition of n-type and p-type carriers is observable under normal bias conditions. The phenomenon—ambipolarity—exists in various materials such as silicon [24], carbon nanotubes [25] and graphene [26]. By controlling ambipolarity, device polarity can be adjusted/tuned post-deployment. Transistors with a configurable polarity—e.g., carbon nanotubes [5], graphene [6], silicon nanowires (SiNWs) [7], and transition metal dichalcogenides (TMDs) [8]—have already been experimentally demonstrated. While this work primarily focuses on SiNW FETs and TFETs built with TMDs, both devices may serve as a “proxy” for other device concepts.

SiNW FETs have an ultra-thin body structure and lightly-doped channel which provides the ability to change the carrier type in the channel by means of a gate. FET operation is enabled by the regulation of Schottky barriers at the source/drain junctions. The control gate (CG) acts conventionally by turning the device on and off via a gate voltage. The polarity gate (PG) acts on the side regions of the device, in proximity to the source/drain (S/D) Schottky junctions, switching the device polarity dynamically between n- and p-type. The input and output voltage levels are compatible, enabling directly-cascadable logic gates [27].

Ambipolarity is an inherent property of TFETs due to the use of different doping types for drain and source if an n/p doping profile is employed [28]. By properly biasing the n-doped and p-doped regions as well as the gate, a TFET can function either as an n- or p-type device. No polarity gate is needed in this case. Furthermore, as the magnitude of ambipolar current can be tuned (i.e., reduced) via doping or by increasing the drain extension length [28] one can envision fabricating devices that could be better suited for logic as well as security-related applications. Given that screening length in TMD devices scales with their body thickness, one can achieve substantial tunneling currents.

2) Bell-Shaped I-Vs: Emerging transistor technologies may also exhibit bell-shaped I-V curves. Symmetric graphene FETs (SymFETs) and ThinTFETs are representatives of this group and will be considered in our design work. In a SymFET, tunneling occurs between two, 2-D materials separated by a thin insulator. The $I_{DS-V_{GS}}$ relationship exhibits a strong, negative differential resistance (NDR) region (Fig. 1a). The I-V characteristics of the device are “bell-shaped,” and the device can remain off even at higher values of $V_{DS}$. Per Fig. 1a, the magnitude of the current peak and the position of the peak are tunable via the top gate ($V_{TG}$) and back gate ($V_{BG}$) voltages of the device. Such behavior has been observed experimentally [11], [12]. More specifically, $V_{TG}$ and $V_{BG}$ change the carrier type/density of the drain and source graphene layers by electrostatic field, which can modulate $I_{DS}$. ITFETs or ThinTFETs may exhibit similar I-V characteristics [13].

3) Tunable Hysteresis: Tunable hysteresis suggests that (i) an I-V curve may contain a hysteresis loop, and (ii) the hysteresis loop can either be moved to different locations along the applied voltage levels by some mechanisms, or be made to disappear altogether. The NCFET is representative of this group and will be considered as a design target (see Fig. 1b). NCFETs are made by adding a ferroelectric (FE) material in the gate stack of a MOSFET. The high polarizability of the FE material provides a nonlinear capacitance...
which becomes negative under certain electric field values, which enables step-up voltage conversion of the applied gate bias to the surface potential leading to switching slope (SS) steeper than 60 mV/decade. NCFETs can be made with or without hysteretic behavior by varying the gate stack material composition. Changing the thickness of the FE material can also make NCFETs to either have or not have hysteresis. Theoretical analyses have revealed that the hysteresis loop of drain current vs. gate voltage can be altered by changing the drain voltage [29] as electrostatic coupling of the channel to the drain changes the FE capacitance as well as the FET capacitance. Thus, the position of the hysteresis loop in an NCFET can be dynamically tuned. Ionic FETs may also exhibit tunable hysteresis.

III. HARDWARE SECURITY BASED ON TUNABLE POLARITY

Here, we present our recent work on exploiting the tunable polarity property for hardware security. The ability to dynamically change the polarity of a transistor opens the door to define the functionality of a layout or a netlist post fabrication. Though one may use field programmable gate arrays (FPGAs) to achieve the same goal, FPGAs cannot compete with ASICs in terms of performance and power, and an FPGA’s reliance on configuration bits being stored in memory introduces another vulnerability. Our work has resulted in novel security primitives by leveraging the tunable polarity property of SiNW FETs and TFETs to provide logic and layout obfuscation [30], [31]. These primitives can serve as building blocks for IP protection, IP piracy prevention, and to counter hardware Trojan attacks.

A. Polymorphic logic gates

Polymorphic logic circuits provide an effective way for logic encryption such that attackers cannot easily identify circuit functionality even though the entire netlist/layout is available. However, polymorphic logic gates have never been widely used in CMOS circuits mainly due to the difficulties in designing such circuits using CMOS technology.

In [30], [31], we introduced SiNW FET based polymorphic gates to prevent IP piracy. If the control gate (CG) of a SiNW FET is connected to a normal input, while the polarity gate (PG) is treated as the polymorphic control input, through different configurations on the polymorphic control inputs, we can easily change the circuit functionality without a performance penalty. For example, per Fig. 2a-b), a SiNW FET based NAND gate can be converted to a NOR gate, whereas a CMOS-based NAND cannot be converted to a fully functioning NOR by switching power and ground.

We have recently designed TFET-based polymorphic logic circuits as well. Fig. 2c shows a 2-input polymorphic NAND/NOR gate. As shown in Fig 2c, by properly biasing the gate, the n-doped region, and the p-doped region, a TFET device can function either as an n-type transistor or p-type transistor. (The small circle on the transistor designates the p-doped region.) For the schematic in Fig. 2c, if the n-doped region of the two parallel TFETs is connected to $V_{DD}$, and the p-doped region of the bottom TFET is connected to $GND$, the circuit behaves like a NAND gate. If the n-doped region of the two parallel TFETs is connected to $GND$ and the p-doped region of the bottom TFET is connected to $V_{DD}$, the circuit behaves as a NOR gate. The simulation results based on a 1D ballistic QCL (quantum capacitance limit) model (representative of TMD devices) shows the expected
polymorphic functionality. By using two MUXes (one at the top and the other at the bottom) to select between the two types of connections, the circuit then functions as a polymorphic gate where the control to the MUXes forms a 1-bit key.

Using the low-cost polymorphic logic gates built from either SiNW FETs or TFETs, we can readily design polymorphic functional modules which only perform a desired computation if properly configured. If some key components (e.g., the datapath) in an ASIC is designed in this manner, the chip is thus encrypted such that a key, i.e., the correct circuit configuration, is required to unlock the circuit functionality. Invalid users or attackers cannot use the circuit without the key. Thus, IP cloning and IP piracy can be prevented with extremely low performance overhead. As a preliminary result, a 32-bit polymorphic adder using SiNW FETs was designed and simulated. Fig. 3a shows the schematic of a 1-bit full-adder using polymorphic gates as well as the block diagram of a 32-bit polymorphic adder relying on the full-adder. In this 32-bit polymorphic adder, two pairs of configuration bits (with up to 32-bits in length) are introduced and the adder can only perform addition functionality if the correct configuration bits are provided.

B. Camouflaging Layout

Split manufacturing and IC camouflaging are used to secure the CMOS fabrication process, albeit with high overhead and decreased circuit reliability. With CMOS camouflaging layouts, both power and area would increase significantly in order to achieve high levels of protection [18]. A CMOS camouflaging layout that can function either as an XOR, NAND or NOR gate requires at least 12 transistors. Emerging technologies help reduce the area overhead. For example, most 2-input logic gates including NAND, NOR, XOR and XNOR only consume four SiNW FETs and share a similar layout.

In our recent work [30], [31], we have demonstrated that only 4 SiNW FETs with tunable polarity are required to build a camouflage layout that can perform NAND, NOR, XOR or XNOR functionality. (Fig. 3b depicts a camouflaging layout that can perform any one of the four functions given.) Again, the SiNW FET based camouflage layout has more functionality and requires less area than CMOS counterparts – and could offer higher levels of protection to circuit designs.

C. Security Analysis

Logic obfuscation is subject to brute-force attacks. If there are $N$ polymorphic gates incorporated in the design, it would take $2^N$ trials for an attacker to determine the exact functionality of the circuit. As the value of $N$ increases, the probability of successfully mounting a brute-force attack becomes extremely low. In our preliminary implementation of 32-bit adder, the incorporated key size is 32 bit. The probability that an attacker can retrieve the correct key becomes $1/2^{32}$ ($2.33 \times 10^{-10}$). Obviously, polymorphic based logic obfuscation techniques are resistant to a conventional brute-force attack. With respect to camouflage layouts, given that our proposed SiNW based camouflage layout can perform four different functions, the probability that an attacker can retrieve the correct layout is 25%. Therefore, if $N$ SiNW FET camouflage layouts are incorporated in a design, the attacker has to compute up to $4^N$ times to resolve the correct layout design. Compared to polymorphic gates based logic obfuscation, camouflage layout embraces higher security level but with larger area overhead.

IV. HARDWARE SECURITY LEVERAGING ATYPICAL SWITCHING BEHAVIORS

Many post-CMOS transistors aim to achieve steeper sub-threshold swing, which in turn enables lower operating voltage and power. Many devices in this space also exhibit I-V characteristics that that are not representative of a conventional MOSFET. Here, we begin to consider how to exploit said characteristics for designing hardware security primitives. We will focus on devices with (i) bell-shaped I-V curves and (ii) tunable hysteresis.

A. SymFET-based Protector Circuits:

Side-channel analysis, such as fault injection, power and timing analysis, allows attackers to learn about internal circuit signals without destroying the fabricated chips. Countermeasures have been proposed to balance the delay and power consumption when performing encryption/decryption at either the algorithm or circuit levels [32]. These methods often cause higher power consumption and longer computation time in order to balance the side-channel signals under different conditions. Thus, an important goal is to prevent fault injection...
and to counter side-channel analysis by introducing low-cost, on-chip voltage/current monitors and protectors. Graphene SymFETs, which have a voltage-controlled unique peak current (see Fig. 1a) can be used to build low-cost, high-sensitivity circuit protectors through supply voltage monitoring.

In our recent work, we have developed a SymFET-based power supply protector [30], [31]. With only two SymFETs, the power supply protector (Fig. 1a) can easily monitor the supply voltage to ensure that the supply voltage to the circuit-under-protection is within a predefined range [30]. In the event of a fault injection, the decreased supply voltage will power down the circuit rather than injecting a single-bit fault, and can thus protect the circuit from fault injection attacks (Fig. 4b). If we use $V_{out}$ as the power supply to a circuit under protection (e.g., an adder), due to the bell-shaped I-V characteristic of the SymFET, an intentional lowering of $V_{DD}$ cuts off the power supply. Thus, the sum (see Fig. 4b) and carry-out of the full adder output ‘0’, and no delay related faults are induced. A similar CMOS power supply protector would require op-amps for voltage comparison. As a result of the voltage/current monitors developed thus far, voltage/current based fault injections can be largely prevented. By inserting the protectors in the critical components of a given circuit design, the power supply to these components can be monitored and protected. (See [31] for more detail.)

B. NCFET based logic-in-memory:

The tunable hysteresis property found in post-CMOS devices such as NCFETs and ionic FETs allows the device to be dynamically configured as either a switch or a non-volatile storage element. An immediate benefit is the ability to design simpler and more power efficient logic-in-memory (LiM) cells. This could reduce communication between a CPU and memory (reducing vulnerability to memory attacks during communication) [33]. LiM cells can also help to reduce overhead incurred by key access and verification via storing the keys in LiM cells. However, designing superior LiM cells that can be used at large scales has proven to be elusive.

We have recently designed several different LiM cells using NCFETs. Fig. 5a shows an example of a LiM cell (performing a NAND function) based on the pseudo-NMOS logic style. The circuit has two modes – update mode and hold mode. In the update mode (i.e., $Write\_EN = 1$), the $Y$ input is written into the NCFET, and the output realizes the logic function of $Z = \overline{Y}$. In the hold mode (i.e., $Write\_EN = 0$), the circuit outputs $Z = \overline{X} \cdot S$, where $S$ is the bit value stored in the NCFET which remains unchanged. The pseudo-NMOS design may lead to relatively large leakage, but similar CMOS-like designs can also be obtained. Fig. 5b shows an AND/NAND-LiM cell design based on the dynamic current mode (DyCML) style. Similar to the circuit in Fig. 5a, this circuit also has an update and hold mode. In the hold mode (i.e., $Write\_EN = 0$), the circuit outputs $Z = \overline{X} \cdot S$ and $Z = X \cdot S$ where $S$ is the bit value stored in the NCFET. In the update mode (i.e., $Write\_EN = 1$), $Y$ and $\overline{Y}$ are written into the two NCFETs, respectively, while output does the same evaluation as in the hold mode. (Note that a NAND example was used for illustrative purposes. More complex functions are also possible.)

Looking forward, devices with tunable hysteresis offer unique functionality that is difficult to obtain with MOSFETs. (i) Such a device can be readily changed from being a non-volatile storage element to a switch. This property could help achieve logic obfuscation. (ii) With three terminals, such a device can be used as a “more capable” storage element (e.g.,
comparing with a ferro-electric capacitor). This opens the door for simpler LiM cells, which could lead to efficient memory protection strategies. (iii) The retention time of such a device as a non-volatile storage element can also be tuned which may be exploited for tamper resistant circuitry.

To exploit the capability of NCFETs being either a storage element or a switch, in future work we will investigate NCFET-based efficient design obfuscation on both combinational logic and sequential logic. For example, in [34], the authors have proposed techniques that use finite state machines (FSMs) together with some combinational logic to help obfuscate an IP design. NCFETs could be employed to implement such FSMs. If we leverage the design concepts of the LiM cells discussed in Sec. IV-B to construct a FSM, the FSM behavior can be tuned, hence providing another level of obfuscation. The control mode (between being a storage element or a switch) then becomes the encryption key.

Finally, NCFET-based logic could also help to resist denial-of-service (DoS) attacks and differential power analysis (DPA). DoS attacks can be deployed on energy constrained systems (e.g., mobile phones). LiM cells, if used both to store the security key and for authentication, could provide an extremely energy efficient authentication process through close integration of memory and logic.

V. CONCLUSIONS

Many post-CMOS devices “naturally” exhibit unique I-V characteristics that may not be immediately considered to be useful when looking for a drop-in replacement for MOSFETs. However, our preliminary work has demonstrated that they can be extremely effective in implementing certain security functions. Given the importance of hardware security, more research is needed to fully understand the potential of these I-V characteristics for security as well as new attack models that these new security primitives may have to deal with.

Acknowledgement: This work was supported in part by the Center for Low Energy Systems Technology (LEAST), an SRC STARnet center sponsored by MARCO and DARPA.

REFERENCES