Methoden zur Verbesserung von CMOS integrierten Arbiter-PUFs

Andreas Herkle, Joachim Becker, Maurits Ortmanns
Institut für Mikroelektronik, Universität Ulm
Albert-Einstein-Allee 43, 89081 Ulm
andreas.herkle@uni-ulm.de

Kurzfassung

In diesem Beitrag wird eine Auswahl von Methoden zur Verbesserung integrierter CMOS PUFs vorgestellt. Am Beispiel der Arbiter-PUF zeigen wir, dass sowohl die PUF-definierenden Zufallsquellen erheblich optimiert werden können, als auch die ergänzende Messschaltung. Wir zeigen, dass insbesondere die Sensitivität der Gesamtschaltung gegenüber elektrischem Rauschen stark reduziert werden kann.

1 Einführung

Das Internet der Dinge (IoT, Internet-of-things) ist ein stark wachsender Forschungssektor. Vernetzte Haushaltsgeräte, intelligente Saugroboter und selbst-auffüllende Kühlschränke sind eine kleine Auswahl dessen, was möglich scheint. Jedoch hat sich bereits mehrfach in der Vergangenheit gezeigt, dass diese Vollvernetzung dieselben Schwächen aufweist, wie ans Internet angebundene Arbeitsplatzrechner: Aufgrund eklatanter Sicherheitsmängel konnten bereits etliche IoT-Geräte gehackt werden. Daher gilt die Annahme, dass vernetzte Geräte nicht implizit sicher sein können, weshalb insbesondere Hardware-Embedded-Security ein wichtigeres Thema als je zuvor ist. Um sichere Kommunikation und Authentifizierung gewährleisten zu können, muss ein IoT-Gerät eine geheime Information be(in)halten oder (re)generieren können. Für diese Aufgabenstellung eignet sich eine “Physical Unclonable Function” (PUF) hervorragend. Integrierte PUFs sind elektrische Schaltungen, die basierend auf herstellungsbedingten Abweichungen einen gerätespezifischen Fingerabdruck erzeugen. Diese, üblicherweise als Bitvektoren dargestellten Fingerabdrücke, sind nicht vorhersagbar oder beeinflussbar, auf dem gefertigten Chip jedoch sehr leicht reproduzierbar. Im Gegensatz zu sicheren Speichern benötigen PUFs keine durchgehende Energieversorgung und können zusätzlich ihre Ausgabe (response) in Abhängigkeit von einem Steuereingang (challenge) variieren (Challenge-Response-Pairing, CRP). Aufgrund dieser hervorragenden Eigenschaften bieten PUFs eine sichere Möglichkeit zur Generierung von hardware-kryptographischen Schlüsseln u.a. für IoT-Geräte. PUFs haben sich daher als fester Bestandteil hardware-kryptographischer Konferenzen und Workshops etabliert.

In unserem Beitrag stellen wir zwei Ansätze zur Verbesserung von PUFs auf Schaltungsebene vor. Kapitel 2 zeigt Möglichkeiten zur Verbesserung an den Zufallsquellen von PUFs. Kapitel 3 zeigt Möglichkeiten zur Verbesserungen an den zugehörigen Messschaltungen. Kapitel 4 fasst unseren Beitrag zusammen.

1.1 Zeitbasierte PUFs - Arbiter-PUF

Eine der ersten integrierten PUFs ist die Arbiter-PUF [1], exemplarisch dargestellt in Abbildung 1. Diese Implementierung nutzt die konstanten Abweichungen von der nominalen Verzögerungszeit von CMOS-Invertern (in rot als Buffer), um eine Unterscheidbarkeit zweier konkurrierender Signale zu ermöglichen. Bei der Arbiter-PUF wird eine steigende Flanke gleichzeitig auf zwei identische Verzögerungspfade geleitet, deren Verlauf durch eine Challenge (switches in grün) modifiziert werden kann. Am Ende dieser Verzögerungskette evaluiert ein Arbiter (RS-Latch in blau), welches der beiden Signale zuerst ankommt (Vorzeichen der Phasendifferenz Δt) und definiert darüber die Bitantwort dieser PUF-Zelle. Eine Systemantwort mit längeren Bitvektoren kann durch die Verwendung mehrerer solcher PUF-Zellen erzeugt werden (r_1 bis r_8).

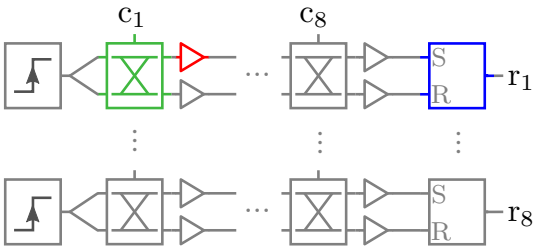

Abbildung 1 Aufbau einer 8-bit Arbiter-PUF [3]

2 Quellenmodifikation durch angepasste Strukturgrößen

Die Qualität einer PUF bezogen auf Einzigartigkeit und Unterscheidbarkeit wird maßgeblich von der Eignung der zugrundeliegenden Zufallsquelle definiert. Das Maß für diese Eignung ist die inter-Hammingdistanz (inter-HD), welche die mittlere Anzahl unterschiedlicher Bits zweier Chip-Instanzen angibt (ideal 50%). Obwohl die Zufallsquellen meist gut gewählt sind und von vornherein gute statistische Eigenschaften aufweisen können, verbleibt immer Raum für Verbesserungen, beispielsweise

durch die Verstärkung des Einflusses der herstellungsbedingten Abweichungen [5]. Insbesondere die Robustheit gegenüber schwankenden Umgebungsbedingungen (Rauschen, Temperatur- und Spannungsschwankungen) stellt bei PUFs eine Herausforderung dar [4]. Die Anfälligkeit einer PUF gegenüber diesen Einflüssen wird als intra-Hammingdistanz (intra-HD) bezeichnet und gibt die Anzahl an Bit-Flips über mehrere Messungen an (ideal 0%).
Eine Möglichkeit zur Verbesserung der statistischen Eigenschaften einer Arbiter-PUF sind angepasste Strukturgrößen [3]. Wir haben die einzelnen Komponenten einer Arbiter-PUF analysiert; im speziellen das RS-Latch als Arbiter, die inverter-basierten Verzögerungselemente und die Transmission-Gate basierten Schalter. Um eine bessere Vergleichbarkeit zur erreichen und den Einfluss der Verbesserungen hervorzuheben, haben wir unsere Implementierung entsprechend der ursprünglichen Arbiter-PUF gewählt [1]. Mit Hilfe von Monte Carlo Simulationen und Variation der Transistor-Dimensionierungen lässt sich die Qualität der Messungen durch angepasste Strukturgrößen signifikant verbessern, ohne die Grundstruktur der Schaltung zu verändern. Im folgenden wird für drei Teile der Schaltung gezeigt, wie die Performanz gesteigert werden kann.

2.1 RS-Latch als Arbiter

Die Nichtidealität und Rauschanfälligkeit des Arbiter-Elementes hat maßgeblichen Einfluss sowohl auf die Qualität der Systemantwort als auch auf die Stabilität deren Messung. Die Präzision der Messung wird als die Phasendifferenz Δt beschrieben, unterhalb der mit einer Wahrscheinlichkeit von 20% eine fehlerhafte Antwort generiert wird. Für ein RS-Latch, bestehend aus zwei kreuzgekoppelten NANDs, in einer 40nm Technologie mit minimal dimensionierten Transistoren liegt dieser Wert bei 0.17ps. Durch eine Vergrößerung der Gatebreite W_n der NMOS-Transistoren beider NANDs auf 1µm konnte dieser Wert um über 76% auf 0.04ps reduziert werden. Dieser Effekt erklärt sich dadurch, dass die Transkonduktanz der Transistoren erhöht wird und damit das eingangsbezogene Rauschen verringert wird, während die Verstärkung des Latch insgesamt nur wenig verringert wird.
Diese Modifikation der Schaltung hat ebenso positive Auswirkungen auf die Nichtidealität des RS-Latches, da der relative Einfluss herstellungsbedingter Abweichungen im Latch selbst minimiert wird. Das RS-Latch mit minimalen Größen (W_n = 120nm) verschiebt die Entscheidungsgrenze um bis zu 1ps, d.h. im Vergleich zu einem idealen Entscheider werden Signale, die gleichzeitig ankommen, entschieden, als hätten sie einen Versatz von 1ps. Abbildung 2 stellt diese Abweichungen vom idealen Entscheider dar. Durch die Verbreiterung von W_n auf 1µm reduziert sich dieser Effekt um 62.4% auf 0.38ps. Durch einen Vergleich zweier verschiedener Flankenanstiegszeiten am Eingang des Arbiters in Relation zur Phasendifferenz zeigt sich auch, dass eine schnellere Anstiegszeit mit einer geringeren Fehlerwahrscheinlichkeit (BER) korreliert. Im Detail verglichen wir Flankenanstiegszeiten von 10ps und 1ns, wobei erstere eine BER von 0% für eine Phasendifferenz von 0.5ps

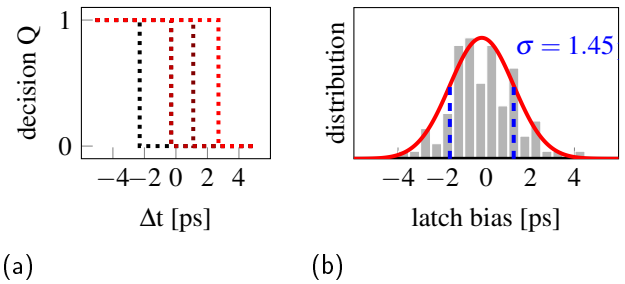

Abbildung 2 Verschiebung der Entscheidungsgrenze eines RS-Latches durch herstellungsbedingte Abweichungen, (a) exemplarische Darstellung, (b) Verteilung

aufweist, während die Flanke mit 1ns dort noch eine deutliche erhöhte BER von 37.5% aufweist. Folglich sollten also die letzten Verzögerungsstufen vor dem Arbiter eine schnell ansteigende Flanke erzeugen und dementsprechend entworfen werden.

2.2 Inverter als Verzögerungselemente

Für die inverter-basierten Verzögerungselemente wird normalerweise angenommen, dass eine minimale Kanalbreite sowie eine vergrößerte Kanallänge die herstellungsbedingten Abweichungen verstärken. Es zeigt sich aber auch, dass der umgekehrte Fall die Rauschimmunität der Schaltung verbessern kann. Vergrößert man, wie beim RS-Latch, die Kanalbreite W_n der NMOS-Transistoren, erhöht sich die Varianz der einzelnen Verzögerungszeiten um 228%, während die Varianz des Rauschens nur um 70% ansteigt. Dieser Effekt erklärt sich dadurch, dass die PMOS-Transistoren einer Stufe eine höhere parasitäre Kapazität laden müssen, während ihre eigene Treiberstärke unverändert bleibt. Folglich verstärkt sich der Einfluss herstellungsbedingter Abweichungen deutlich.

2.3 Transmission-Gates als Schalter

Zuletzt haben wir auch den Einfluss der veränderten Kanalbreiten bei den als Schalter fungierenden Transmission-Gates untersucht. Es zeigte sich jedoch, dass die Auswirkungen auf beide erwähnten Varianzen proportional zueinander sind und daher an dieser Stelle keine Verbesserung erreicht werden kann.

2.4 Bewertung

In der abschließende Auswertung wurden die vorgestellten Verbesserungen, also die Verbreitung der Kanalbreiten W_n der Verzögerungselemente und des Arbiters, zusammen evaluiert. Die abschließende Bewertung zeigte, dass dadurch der Einfluss von Rauschen auf die Stabilität um mehr als den Faktor von 2 reduziert werden konnte: die intra-HD sank von 6.57% auf 2.59%, bei einer unveränderten inter-HD von nahezu idealen 50%. Für alle Elemente der Schaltung wurde auch der Einfluss der Kanalbreite W_p untersucht. Erwartungsgemäß zeigten sich hierbei dieselben Effekte mit geringerer Auswirkung, weshalb die Kanalbreite W_p minimal dimensioniert belassen wurde.

3 Verbesserte Messung von PUFs

Da die informationstragenden Kenngrößen (also die elektrischen Signale der Zufallsquellen) in einer PUF oft sehr kleine Differenzen aufweisen, ist eine hochqualitative Messschaltung absolut essentiell für die Reproduzierbarkeit der PUF-Signatur. Insbesondere sollte die Messschaltung die ursprüngliche Information nicht beeinflussen (Bias) oder anfällig gegenüber zeitlich veränderlichen Einflüssen sein (Rauschen, Temperatur, Spannungsschwankungen). Beispielsweise wurde in [7] ein Buffer als Messschaltung verwendet, dann aber durch einen genaueren und stabileren Sense-Amplifier ersetzt. Oft kommen ergänzend auch post-processing Methoden zur Anwendung, beispielsweise Dark-Bit Masking und Temporal Majority Voting [6].

3.1 Eye-Opening Oscillator als Arbiter

Wie bereits erwähnt, ist ein RS-Latch als Arbiter das kritischste Element in Bezug auf die Stabilität der Messung. Insbesondere für kleine Phasendifferenzen erhöht sich die Einschwingzeit des RS-Latches drastisch und ist währenddessen noch sensitiver gegenüber elektrischem Rauschen. Es zeigt sich zudem, dass die Gesamtfehlerrate einer Arbiter-PUF stark davon abhängt, ob die jeweilige angelegte Challenge zu solchen kleinen Phasendifferenzen führt. Daraus resultierende extrem hohe Fehlerraten dominieren die Gesamtfehlerrrate, obwohl sie vergleichsweise seltener auftreten.

Unser Ansatz zur Vesserungs hierbei ist die Anwendung des Konzepts eines Eye-Opening Oscillators [8] auf eine Arbiter-PUF [2], die in Abbildung 3 als "multiplexed delay-lines- dargestellt ist und der originalen Schaltung aus Kapitel 2 entspricht. Dabei wird das RS-Latch durch zwei D-Master-Slave-Flip-Flops (D-MS-FF) ersetzt, bei denen jeweils die Eingänge vertauscht angebunden werden: entsprechend Abbildung 3 stellt der obere Ausgang der Verzögerungskette den D-Eingang des einen D-MS-FFs und das Taktsignal des anderen D-MS-FFs dar.

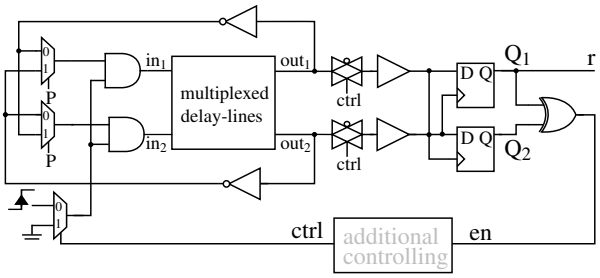

Abbildung 3 Arbiter-PUF mit D-MS-FF als Eye-Opening Oscillator und invertiertem Feedback

Für den unteren Ausgang der Verzögerungskette und das zweite D-MS-FF gilt die umgekehrte Verschaltung. Durch diese Verknüpfung der jeweiligen Setup-Zeiten der einzelnen Flip-Flops entsteht eine sogenannte Deadzone. Um eine gültige Messung zu erhalten, muss für die Phasendifferenz zwischen den beiden konkurrierenden Signalen gelten, dass sie betragsmäßig nicht innerhalb dieser Deadzone liegen darf. In einem solchen Fall würden beide Flip-Flops an ihrem digitalen Ausgang Q den Bit-Wert 0 ausgeben. Durch eine XOR-Verknüpfung kann dieser Fall erkannt werden und damit eine fehleranfällige Entscheidung dieses neuen Arbiters verhindert werden. Führt man dann die beiden Ausgänge der Verzögerungskette invertiert zurück auf dessen Eingänge, schickt man damit die beiden Signale erneut durch die Verzögerungskette (versetzt um die Phasendifferenz Δt) und erhält ein oszillierendes Verhalten, ähnlich einem Ringoszillator. Hierbei muss darauf geachtet werden, dass die Kontinuität des Feedbacks eingehalten wird: Der Ausgang, welcher zu dem oberen Eingangspfad korrespondiert, muss auch auf den oberen Eingang zurückgeführt werden. Da sich diese Zuordnung über verschiedene Challenges ändern kann, übernehmen Multiplexer diese Zuordnung anhand der Parität der Challenge.

Bei korrekter Rückführung wird die Phasendifferenz zwischen den beiden Signale stetig um denselben Betrag vergrößert: Es findet also eine Integration der herstellungsbedingten Abweichungen über Zeit statt, während zeitlich veränderliche Einflüsse wie Rauschen unterdrückt werden. Nach mehreren Oszillationen kann die Phasendifferenz so groß werden, dass ihr Betrag die Deadzone verlässt, eines der beiden Flip-Flops einen Bit-Wert von 1 ausgibt und damit die Messung abgeschlossen werden kann. Abbildung 4 zeigt beispielhafte Verläufe einer solchen Integration über mehrere Oszillationen, dabei ist die Deadzone als transparente rote Fläche dargestellt.

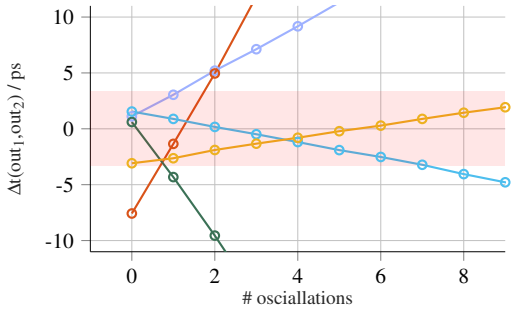

Abbildung 4 Beispielhafte Verläufe der Phasendifferenzen Δt einzelner PUF-Instanzen über mehrere Oszillationen, deadzone als rote Fläche dargestellt

3.2 Diskussion der Oszillationsdauer

Da sich die Phasendifferenz bei manchen Konfigurationen nur sehr langsam aus der Deadzone bewegt und durch Rauschen auch kurzzeitig zurückfallen kann, empfiehlt es sich, die Abbruchbedingung mehrfach zu prüfen. Hierfür verwenden wir eine Kette von Flip-Flops (Abbildung 5). Das Signal *en* aus Abbildung 3 dient hierbei als active-low Reset-Signal, während über die UND-Verknüpfung von out_1 und out_2 das Taktsignal erzeugt wird. Die Oszillation wird dadurch nur dann abgeschaltet, wenn die Messung mehrfach hintereinander eindeutig entschieden werden konnte, d.h. die Flip-Flops in dieser Zeit nicht zurückgesetzt wurden.

Anhand der Simulation über 100 verschiedene Instanzen hat sich auch gezeigt, dass sich etliche Phasendifferenzen erst relativ spät in die Deadzone bewegen, während andere sich sehr nah an deren Grenze bewegen. Es kann also vor-

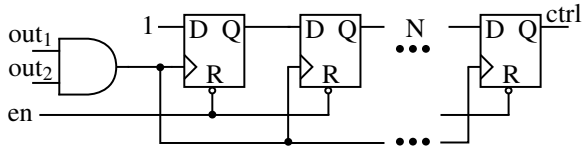

Abbildung 5 Flip-Flop Kette zur Erzwingung mehrerer Oszillationen und eindeutiger Entscheidungen

teilhaft sein, eine Mindestanzahl an Oszillationen vorzugeben, welche diese problematischen Konstellationen abfängt. Eine solche Mindestanzahl kann durch eine Verlängerung der Flip-Flop Entscheidungskette eingestellt werden. Jedoch darf diese Mindestanzahl nicht zu groß gewählt werden, da andernfalls ein Sprung in der Phasendifferenz auftreten kann: wenn der Phasenversatz also so groß ist, dass Flanken am Arbiter verglichen werden, die zu unterschiedlichen Evaluationsdurchgängen gehören.
Zusätzlich sollte eine maximale Anzahl an Oszillationen bedacht werden. Einerseits kann damit die maximale Dauer der Evaluation begrenzt werden, andererseits geht eine lange Oszillationszeit mit einer sehr geringen initialen Phasenverschiebung einher, d.h. die Unterschiedlichkeit der beiden Pfade ist sehr gering. Solche Konstellationen können damit als unlösbar erkannt und später maskiert werden.

3.3 Bewertung

Eine abschließende Simulation in einer 40nm Technologie bestätigt die Effizienz unseres Ansatzes: Die Gesamtfehlerrate (intra-HD) konnte auf $9.2 \cdot 10^{-5}$ reduziert werden, während die inter-HD unverändert bei 49.3% blieb. Ohne erzwungene Mindestanzahl an Oszillationen betrug deren maximale Anzahl 17, während sich diese Anzahl bei erzwungener Oszillation deutlich auf 39 erhöhte. Dieser Anstieg hebt noch einmal hervor, dass bei manchen Instanzen zu früh entschieden wurde und diese daher absichtlich in die Deadzone gezwungen werden sollten.
Abschließend kann das vorgestellte Konzept auch insofern positiv bewertet werden, dass dessen Implementierung einfach umgesetzt werden kann: Transistoren können minimal dimensioniert werden und der zusätzliche Flächenbedarf im Vergleich zu längeren Verzögerungsketten fällt minimal aus.

4 Zusammenfassung

Wir haben in diesem Beitrag verschiedene Möglichkeiten vorgestellt, die Rauschimmunität einer CMOS integrierten Arbiter-PUF zu verbessern. Eine dieser Möglichkeiten ist die Anpassung der Strukturgrößen von ursprünglich minimal dimensionierten Transistoren in verschiedenen Schaltungsteilen, wie dem Arbiter und den Verzögerungselementen. Dadurch kann die Bitfehlerrate bereits um einen Faktor von 2 verringert werden. Zusätzlich kann die Bitfehlerrate durch Einsetzen einer neuer Messschaltung für Arbiter-PUFs fast vollständig eliminiert werden. Diese als Eye-Opening Oscillator bekannte Ergänzung ist leicht zu implementieren und erhöht den Flächenbedarf der Schaltung nur minimal.

5 Literatur

[1] J. W. Lee, Daihyun Lim, B. Gassend, G. E. Suh, M. van Dijk and S. Devadas, „A technique to build a secret key in integrated circuits for identification and authentication applications," 2004 Symposium on VLSI Circuits. Digest of Technical Papers (IEEE Cat. No.04CH37525), 2004, pp. 176-179.

[2] A. Herkle, J. Becker and M. Ortmanns, „An Arbiter PUF employing eye-opening oscillation for improved noise suppression," 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 2018, pp. 1-5.

[3] A. Herkle, M. Schuster, J. Becker and M. Ortmanns, „Enhanced Arbiter PUFs using custom sized structures for reduced noise sensitivity," 2016 IEEE International Conference on Electronics, Circuits and Systems (ICECS), Monte Carlo, 2016, pp. 568-571.

[4] R. Kumar, H. K. Chandrikakutty and S. Kundu, „On improving reliability of delay based Physically Unclonable Functions under temperature variations," 2011 IEEE International Symposium on Hardware-Oriented Security and Trust, San Diego CA, 2011, pp. 142-147.

[5] R. Kumar, V. C. Patil and S. Kundu, „Design of Unique and Reliable Physically Unclonable Functions Based on Current Starved Inverter Chain," 2011 IEEE Computer Society Annual Symposium on VLSI, Chennai, 2011, pp. 224-229.

[6] S. K. Mathew et al., „A 0.19pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22nm CMOS," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, 2014, pp. 278-279.

[7] A. B. Alvarez, W. Zhao and M. Alioto, „Static Physically Unclonable Functions for Secure Chip Identification With 1.9–5.8% Native Bit Instability at 0.6–1 V and 15 fJ/bit in 65 nm," in IEEE Journal of Solid-State Circuits, vol. 51, no. 3, pp. 763-775, March 2016.

[8] K. Yoshioka and H. Ishikuro, „A 13b SAR ADC with eye-opening VCO based comparator," ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC), Venice Lido, 2014, pp. 411-414.