Vol.2017-MBL-84 No.9
Vol.2017-CDS-20 No.9
2017/8/30

耐タンパ PUF に対する電磁波解析の基礎検討

野崎佑典[†1]　吉川雅弥[†1]

概要： 近年，偽造半導体の脅威が深刻化している．偽造半導体は，コンシューマ機器だけでなく，インフラ設備に組み込まれることで，深刻な事故が起きる危険がある．そのため，偽造を防ぐための技術として，Physical Unclonable Function（PUF）が期待されている．一方で，PUF に対する機械学習攻撃の脅威が報告されており，これらの攻撃に耐性を持つ XOR アービターPUF や Lightweight PUF などの耐タンパ PUF が提案されている．しかし近年では，これらの PUF に対して消費電力などのサイドチャネル情報を用いた新たな機械学習攻撃が報告されている．また，今後の PUF の安全性を検討する上で，耐タンパ PUF の詳細な安全性検証を行うことは非常に重要である．そこで本研究では，代表的な耐タンパ PUF の 1 つである XOR アービターPUF に対する電磁波解析を提案する．そして，いくつかの評価実験により，提案手法の有効性と XOR アービターPUF の電磁波解析に対する安全性について検証する．

キーワード：ハードウェアセキュリティ，Physical Unclonable Function，電磁波解析，耐タンパ性

A Study of Electromagnetic Analysis for a Secure PUF

YUSUKE NOZAKI[†1]　MASAYA YOSHIKAWA[†1]

Abstract: The risk of counterfeit semiconductors is pointed out. The counterfeit semiconductors may occur the accident not only consumer electronics but also the infrastructure. Therefore, to prevent the counterfeit semiconductors, physical unclonable functions (PUFs) have been attracted attention. On the other hand, the threat of machine-learning attacks for PUFs was reported, and then secure PUFs, such as an XOR arbiter PUF or a lightweight PUF, have been proposed. However, a new machine-learning attack using side-channel information such as the power consumption for these PUFs was reported. Also, to obtain the security of PUFs, it is very important that a detailed tamper resistance verification of PUFs. Therefore, this study proposes a new electromagnetic analysis for the XOR arbiter PUF. The validity of the proposed method and the tamper resistance of the XOR arbiter PUF against electromagnetic analysis are verified by several experiments.

Keywords: Hardware security, Physical Unclonable Function, Electromagnetic analysis, Tamper resistance

1. はじめに

近年，大規模集積回路（Large Scale Integration：LSI）の解析技術の進歩によって，偽造半導体の脅威が顕在化してきた．実際に，2017 年 7 月には欧州半導体産業協会（European Semiconductor Industry Association：ESIA）が，欧州税関での捜査で 100 万点以上の偽造半導体を押収している [1]．偽造半導体が様々な電子機器に組み込まれることで，コンシューマ機器だけでなく，インフラ設備に対して，深刻な事故を引き起こす可能性がある．そのため，偽造半導体を防ぐための技術として Physical Unclonable Function（PUF）が期待されている [2], [3], [4], [5]. PUF は，半導体の製造ばらつきを，その半導体の固有の識別 ID として抽出する技術である．半導体の製造ばらつきは，人工的に制御することが難しいことから，PUF を複製することは困難である．代表的な PUF には，アービターPUF [2]などがある．

一方で，アービターPUF に対する機械学習攻撃の脅威が報告されている[6], [7]．機械学習攻撃は，アービターPUF の構造をモデル化し，機械学習を行うことで，PUF の機能を複製することが出来る．そのため，機械学習攻撃に対して耐性のある PUF として，XOR アービターPUF [2]や Lightweight PUF [3]などが提案されている．しかし近年では，XOR アービターPUF や Lightweight PUF を対象に回路動作時に発生する消費電力などのサイドチャネル情報を利用した新たな機械学習攻撃 [8]が報告されている．一方で，この機械学習攻撃では，サイドチャネル情報として消費電力を利用しており，放射電磁波を利用した解析手法についての検討は行われていない．また，今後の PUF の安全性を検討する上で，PUF の詳細な安全性の検証を行うことは非常に重要である．

そこで本研究では，代表的な PUF の 1 つである XOR アービターPUF に対する電磁波解析について検討する．提案手法では，放射電磁波の局所性に着目して，XOR アービターPUF を構成する各アービターPUF に対象とした機械学習攻撃を実施する．そして，最終的に XOR アービターPUF

†1 名城大学
Meijo University

Vol.2017-MBL-84 No.9
Vol.2017-CDS-20 No.9
2017/8/30

の出力の推定を行う．また，実デバイスを使用した評価実験を行い，提案手法の有効性と XOR アービターPUF の電磁波解析に対する安全性について検証する．

2. 準備

2.1 Physical Unclonable Function

PUF は半導体の製造ばらつきを固有の識別 ID として抽出する．具体的には，チャレンジと呼ばれる値を PUF に入力し，このときに得られるレスポンスと呼ばれる出力を利用する．これまでに遅延型の PUF [2], [3], [4]やメモリ型の PUF[5]など，多くの PUF が提案されている．遅延型の PUF には，アービターベースの PUF [2], [3]やリングオシレータ PUF [4]などがあり，本研究で対象とするアービターベースの PUF は代表的な PUF の 1 つである．

2.1.1 アービターPUF

アービターPUF [2]は，2 本の等長な配線で構成するセレクタチェーンとアービター回路で構成する．*N* 段のセレクタチェーンのアービターPUF を図 1 に示す．アービターPUF では，*N*[bit]のチャレンジ **C**（C_1，…，C_N）を各セレクタに入力し，信号の伝搬経路を決定する．このとき，チャレンジが 0 の場合は，信号は直進し，チャレンジが 1 の場合は，信号は交差する．各信号はアービター回路に接続されており，最終的にどちらの信号が早く到着するかによって，レスポンスを決定する．図 1 のアービターPUF は，アービター回路に D フリップフロップ（D Flip-Flop : DFF）を使用しており，DFF のデータ D 入力に到着する信号の方が早い場合は，レスポンスは 1 に，クロック CLK 入力に到着する信号の方が早い場合は，レスポンスは 0 となる．

一方で，アービターPUF は信号の伝搬時間とチャレンジの関係を線形モデルで表すことができ，機械学習攻撃に対して脆弱であることが知られている [6], [7]. PUF は物理的に複製することは困難であるが，機械学習攻撃は PUF の機能を複製することが出来るため，脅威とされている．

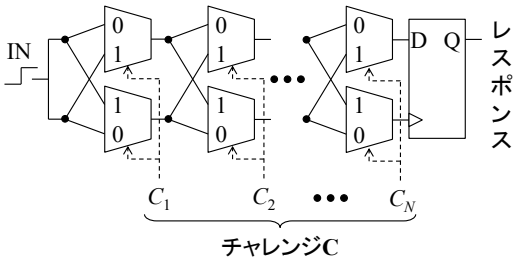

図 1　アービターPUF
Figure 1　Arbiter PUF.

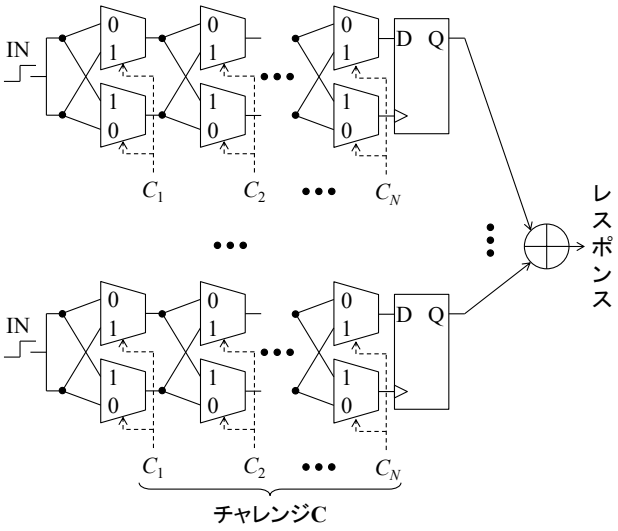

図 2　XOR アービターPUF
Figure 2　XOR arbiter PUF.

2.1.2 XOR アービターPUF

XOR アービターPUF [2]は，機械学習攻撃への対策手法として提案された PUF である．XOR アービターPUF は，複数のアービターPUF で構成する．そして，各アービターPUF に同一のチャレンジを入力し，このときに得られる各出力の XOR を計算し，これを PUF のレスポンスとする．XOR 演算を行うことで，線形モデルで表すことが難しくなるため，機械学習攻撃に対して耐性を持つ．

図 2 に *n*-XOR アービターPUF の例を示す．*n*-XOR アービターPUF は *n* 個のアービターPUF で構成し，*n*[bit]の出力を XOR し，1bit のレスポンスを計算する．

2.1.3 Lightweight PUF

Lightweight PUF [3]は，機械学習攻撃への対策手法として提案された PUF であり，その構造は XOR アービターPUF と似ている．Lightweight PUF は，XOR アービターPUF と同様に複数のアービターPUF で構成し，各アービターPUF の出力を XOR したものをレスポンスとする．ここで，Lightweight PUF では，XOR アービターPUF とは異なり，各アービターPUF に入力するチャレンジはそれぞれ異なる．具体的には，Lightweight PUF では input network と呼ばれる層で変換されたチャレンジが各アービターPUF へ入力される．

2.2 アービターPUF に対する機械学習攻撃

アービターPUF に対する機械学習攻撃 [6], [7]では，アービターPUF の構造を線形モデルでモデル化する．具体的には，アービターPUF の各セレクタチェーンにおける信号の伝搬時間の差を δ_i^0，δ_i^1（δ_i^0 はチャレンジが 0 のときの i 番目のセレクタチェーンの信号の伝搬時間の差，δ_i^1 はチャ

Vol.2017-MBL-84 No.9
Vol.2017-CDS-20 No.9
2017/8/30

レンジが 1 のときの i 番目の信号の伝搬時間の差）とすると，信号の伝搬時間のモデル $\mathbf{u}$（$\mathbf{u} = u_1, \cdots, u_{N+1}$）は式(1)で表すことが出来る．次に，式(2)に示すチャレンジのモデル $\mathbf{v}$（$\mathbf{v} = v_1, \cdots, v_{N+1}$）を用いることで，最終的なアービター回路における信号の伝搬時間の差Δは式(3)で表すことが出来る．そして，この信号の伝搬時間の差Δが正のときにレスポンスを1に，負のときにレスポンスを0と判定する．

$$\begin{cases} u_1 = \dfrac{\delta_1^0 - \delta_1^1}{2} \\ u_i = \dfrac{\delta_{i-1}^0 - \delta_{i-1}^1 + \delta_i^0 - \delta_i^1}{2} \\ u_{N+1} = \dfrac{\delta_N^0 + \delta_N^1}{2} \end{cases} \tag{1}$$

$$\begin{cases} v_1(\mathbf{C}) = \prod_{i=l}^{N} (1 - 2C_i) \\ v_{N+1}(\mathbf{C}) = 1 \end{cases} \tag{2}$$

$$\Delta = \mathbf{u}^{\mathrm{T}}\mathbf{v} \tag{3}$$

2.3 消費電力を利用した機械学習攻撃

これまでに，Lightweight PUF と XOR アービターPUF に対する消費電力を利用した機械学習攻撃が報告されている [8]．この解析では，各アービターPUF のレスポンスが全て 1 または，全て 0 になるレスポンスを出力するチャレンジを良いチャレンジとして，この良いチャレンジのみを用いて機械学習攻撃を行う．この良いチャレンジは，アービター回路動作時の消費電力を観測することで推定する．文献 [8]では，実際の消費電力を用いた解析は行われていないが，良いチャレンジを利用することで，Lightweight PUF に対する機械学習攻撃を成功させている．

一方で，XOR アービターPUF では，良いチャレンジを用いても機械学習攻撃を成功させることは難しい [8]．なぜなら，XOR アービターPUF を構成する全てのアービターPUF には，同一のチャレンジが入力されるからである．全てのアービターPUF に対して，同一のチャレンジが入力されるため，n 個のアービターPUF に対して，n 個の同一の学習モデルが生成される．そのため，良いチャレンジを用いた機械学習攻撃を成功させることは難しい．そこで，文献 [8]では，各アービターPUF の出力の 1 の個数，すなわち，ハミング重み（Hamming Weight：HW）を教師データとして学習させる手法を提案している．この HW はアービター回路動作時の消費電力を観測することで推定する．しかし，文献 [8]では，実際の消費電力を用いた解析は行われていない．

また，これまでに XOR アービターPUF に対する電磁波解析に関する研究は見当たらない．

3. 提案手法

本研究では，XOR アービターPUF に対する電磁波解析について検討する．これまでに，電磁波解析に関する研究としては，主に暗号 LSI を対象としたもの [9], [10]が行われている．これらの研究では，暗号 LSI 動作時の放射電磁波を観測し，統計処理を行うことで，内部の秘密鍵情報の解析を行っている．また，消費電力を用いた電力解析には現れない，放射電磁波の局所性を利用した解析についての研究も行われている [11], [12]．さらに，文献[13]では DFF が保持する 0 または 1 の情報が要因の静的な DFF リークが存在し，放射電磁波によりこの DFF リークの観測が可能なことが報告されている．以上のように，電磁波解析は電力解析にはない局所的な解析が可能であり，より強力な解析となり得る可能性がある．しかし，これまでに XOR アービターPUF などのアービターベースの PUF を対象とした電磁波解析の研究は，筆者らの知る限りにおいて報告されていない．

提案手法では，XOR アービターPUF に対する局所的な電磁波解析を行う．提案手法の概要を図 3 に示す．提案手法は図 3 に示すように，学習フェーズと解析フェーズの 2 つ

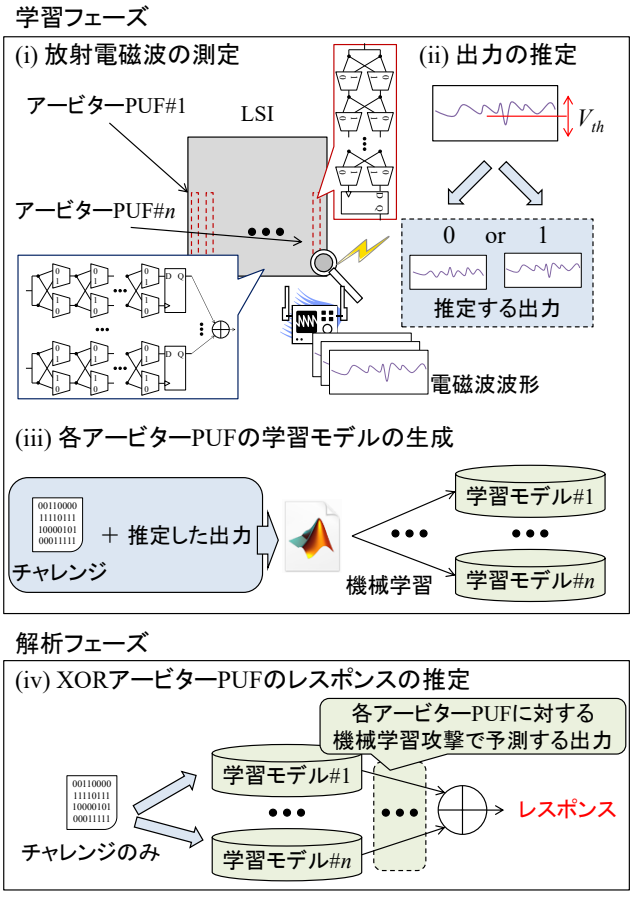

図 3　提案手法の概要

Figure 3　Outline of the proposed method.

のフェーズで構成する．まず学習フェーズでは，XOR アービターPUF のレスポンスの推定に必要な学習モデルの生成を行う．具体的には，図 3 の(i)に示すように，提案手法では XOR アービターPUF を構成する各アービターPUF の出力（各アービターPUF のレスポンス）を，局所的に測定した電磁波波形から推定する．すなわち，あるアービターPUF に注目して解析を行う．ここで，注目するアービターPUF の出力が 0 のときは，アービター回路（DFF）の値が遷移しないため，発生する放射電磁波は小さい．一方で，注目するアービターPUF の出力が 1 のときは，DFF の値が遷移するため，スイッチングによる放射電磁波が発生する．具体的には，図 3 の(ii)に示すように測定した電磁波波形がある閾値 V_{th} よりも小さい場合は，出力を 0，閾値 V_{th} よりも大きい場合は，出力を 1 と推定する．そして，図 3 の(iii)に示すように，推定した出力を用いて注目するアービターPUF に対して機械学習を行い，学習モデルを生成する．このとき，攻撃者は各アービターPUF の出力値を直接取得することは出来ないが，*n*-XOR アービターPUF のレスポンスは利用することが出来る．そのため，*n*-XOR アービターPUF に対する攻撃では，*n*-1 回の放射電磁波の測定を行うことで，解析に必要な学習モデルを生成することが出来る．なぜなら，*n*-1 回の測定で得られた電磁波波形から推定した *n*-1[bit]の各アービターPUF の出力と，*n*-XOR アービターPUF の 1[bit]のレスポンスを利用することで，残りの 1 つのアービターPUF の出力を XOR で計算することが出来るからである．

次に，解析フェーズでは XOR アービターPUF のレスポンスを推定する．具体的には，図 3 の(iv)に示すように，学習フェーズで生成した各アービターPUF の学習モデル（学習モデル#1 から学習モデル#*n*）を利用することで，各アービターPUF の出力（*n*[bit]）を予測する．そして，最終的に *n*[bit]の予測出力値の XOR 演算を行うことで，*n*-XOR アービターPUF のレスポンスを推定する．

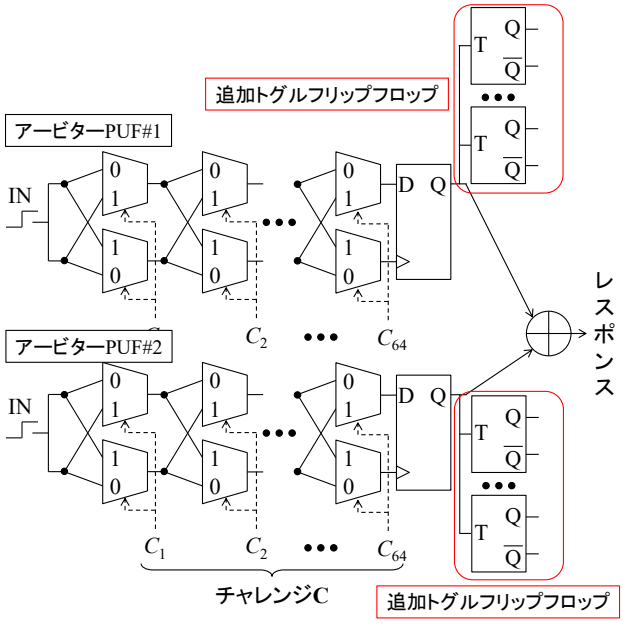

図 4　実装する XOR アービターPUF
Figure 4　The implemented XOR arbiter PUF.

4. 評価実験

4.1 実験環境

実験は，Field Programmable Gate Array（FPGA）ボードを用いて行った．FPGA ボードには，サイドチャネル攻撃標準評価ボード（Side-channel Attack Standard Evaluation Board : SASEBO-W [14]）を使用した．そして，SASEBO-W

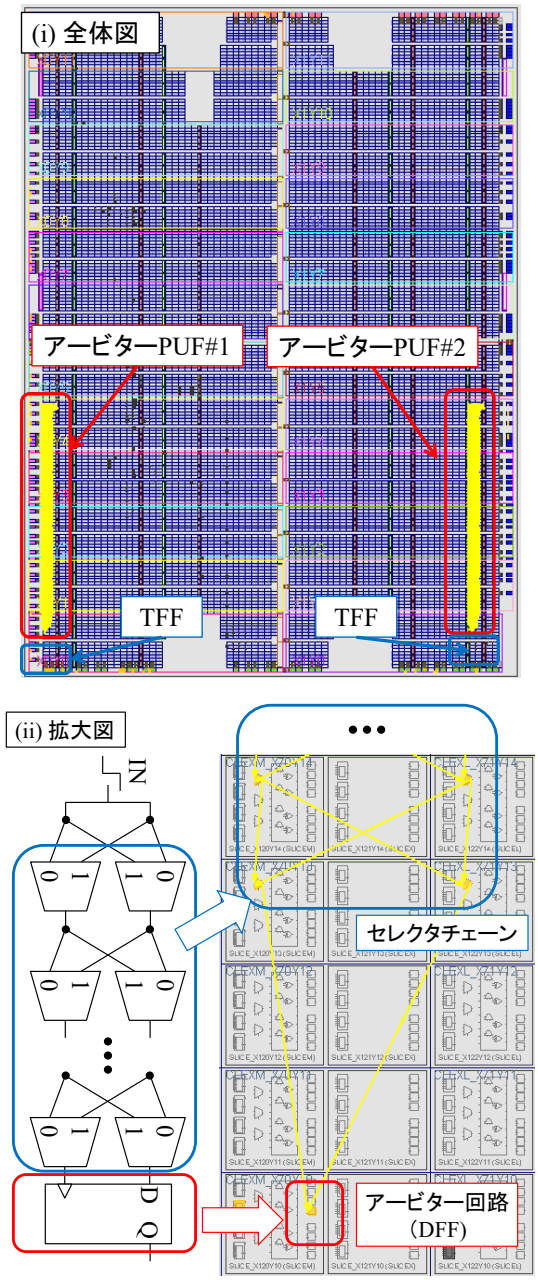

図 5　実装した XOR アービターPUF のフロアプラン
Figure 5　Floorplan of the implemented XOR arbiter PUF.

上の FPGA Spartan-6 にセレクタチェーンが 64 段の 2-XOR アービターPUF を実装した．実装した 2-XOR アービターPUF の概要を図 4 に示す．図 4 に示すように，この XOR アービターPUF では，アービター回路で生じる放射電磁波の測定を容易にするために，トグルフリップフロップ（Toggle Flip-Flop：TFF）を t 個（t = 40）追加実装した．またフロアプランでは，2-XOR アービターPUF の各アービターPUF を，一定の間隔をあけて実装した．フロアプランを図 5 に示す．図 5 の(i)はフロアプランの全体図を，図 5 の(ii)は 2-XOR アービターPUF を構成する 2 つ目のアービターPUF（アービターPUF#2）の拡大図をそれぞれ示している．図 5 の(i)に示すように，1 つ目のアービターPUF（アービターPUF#1）は，FPGA の左下の SLICE_X0Y76 から SLICE_X2Y10 の範囲に，2 つ目のアービターPUF（アービターPUF#2）は，FPGA の右下の SLICE_X120Y76 から SLICE_X123Y10 の範囲に配置した．

実験環境を図 6 に，その詳細を表 1 に示す．放射電磁波の測定では，自作したシールディッドループアンテナを使用し，アンテナのループ開口面を FPGA に対して水平になるように固定した．また，アービターPUF#2 を対象として放射電磁波を測定した．測定位置の拡大図を図 7 に示す．

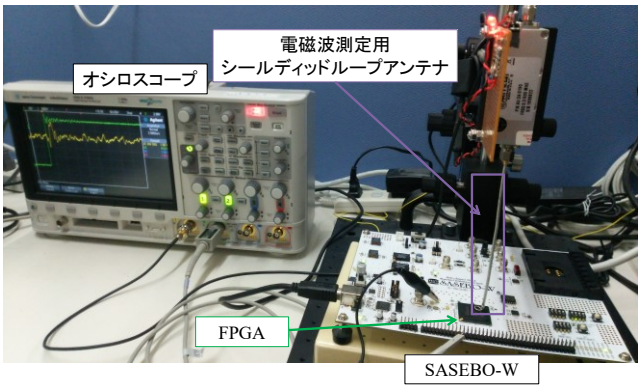

図 6　実験環境
Figure 6　Experimental environment.

表 1　測定環境の詳細
Table 1　Detail of experimental condition.

実装した PUF	2-XOR アービターPUF
セレクタチェーンの段数	64
FPGA ボード	SASEBO-W
FPGA	Spartan-6 XC6SLX150-FGG484
開発環境	Xilinx ISE Design Suite 14.7
フロアプラン	Xilinx PlanAhead v14.7
オシロスコープ	Agilent DSO-X 3104A
サンプリングレート	5 [Gsa/sec]
電磁波測定用プローブ	シールディッドループアンテナ
電磁波測定用 RF アンプ	MITEQ AU-3A-0150

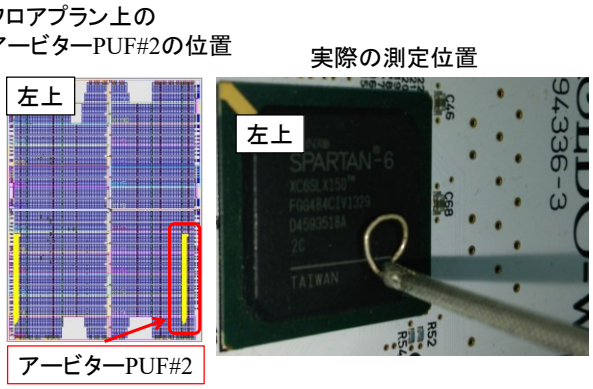

図 7　電磁波測定位置の拡大図
Figure 7　Enlarged view of the location for the measurement of electromagnetic waves.

表 2　解析環境
Table 2　Analytical condition.

SVM	LibSVM [15]
カーネル関数	線形カーネル
パラメータ	デフォルト
学習データ数	1,000
テストデータ数	10,000
PC	HP ProBook 6570b
OS	Windows7 Professional
メモリ	8.00 GB
CPU	Intel Core i7-3520M
解析ソフト	MATLAB 2013b

図 7 に示すように，プローブはフロアプラン上でアービターPUF#2 が実装されていると思われる位置（FPGA の右下部分）で固定した．そして，乱数で生成した 1,000 個のチャレンジを入力し，1,000 個の電磁波波形を測定した．このとき，電磁波波形は 1 個のチャレンジに対して，10 回の平均をとっている．

解析で用いた環境の詳細を表 2 に示す．解析では，機械学習にはサポートベクターマシーン（Support Vector Machine：SVM）を使用した．また，テストデータには 10,000 個のチャレンジレスポンスペア（Challenge Response Pairs：CRPs）を用意した．そして，10,000 個のチャレンジに対するレスポンスを，提案手法により推定し，この推定したレスポンスの正解率を予測率（予測率[%] = （推定に成功したレスポンス数[bit] / 10,000）× 100）として算出した．

4.2　実験結果

実験結果を図 8 に示す．図の縦軸はレスポンスの予測率を，横軸は学習に使用したデータの数を示している．図 8 に示すように，1,000 個の学習データを用いることで，約

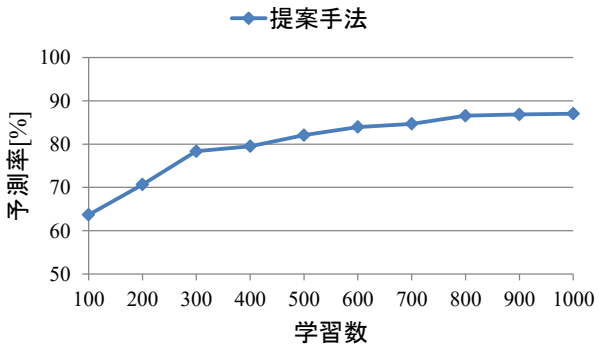

図 8　実験結果
Figure 8　Experimental results.

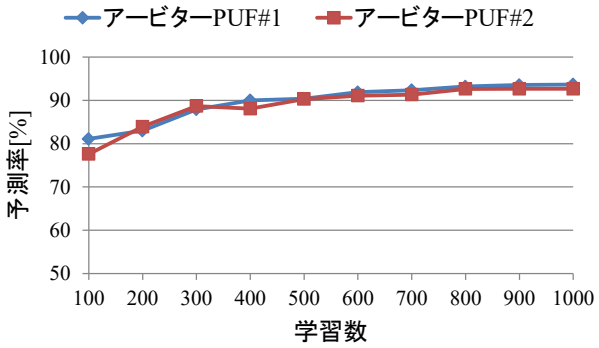

図 9　各アービターPUF に対する機械学習攻撃の結果
Figure 9　Results of machine-learning attack for each arbiter PUF.

87%のレスポンスの推定に成功した．したがって，提案手法は有効であり，XOR アービターPUF が電磁波解析に対して脆弱であることが分かる．

次に，各アービターPUF（アービターPUF#1 とアービターPUF#2）における機械学習攻撃の結果を図 9 に示す．図 9 に示すように，どのアービターPUF（#1 と#2）も 600 個以上の学習数で，90%以上の出力の予測に成功していることが確認出来る．

5. まとめ

本研究では，新たに XOR アービターPUF に対する局所的な電磁波解析について検討した．提案手法では，XOR アービターPUF を構成する各アービターPUF の出力を，放射電磁波を測定することで推定し，各アービターPUF に対して機械学習攻撃を行う．そして，予測した出力値を用いて，最終的に XOR アービターPUF のレスポンスを推定する．2-XOR アービターPUF を FPGA 実装し，実際の放射電磁波を測定した実験では，1,000 個の学習データを用いることで約 87%のレスポンスの推定に成功した．したがって，XOR アービターPUF の電磁波解析に対する新たな脆弱性を明らかにした．

今後は，XOR アービターPUF を構成するアービターPUF の数を変化させた場合など，より詳細な安全性について検証する予定である．

謝辞　この成果は，国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の助成事業の結果得られたものです．

参考文献

[1] ESIA, : Over one million counterfeit semiconductors seized – ESIA supported customs operation with expertise, http://www.eusemiconductors.eu/images/static_website/newsroom/PR/ESIA_PR_JCO-Wafers_3Jul2017.pdf
[2] Lee, J.-W., Lim, D., Gassend, B., Suh, G. E., Dijk, M. V., and Devadas, S.: A Technique to Build a Secret Key in Integrated Circuits for Identification and Authentication Applications, Proc. of the IEEE VLSI Circuits Symposium, pp.176–179 (2004).
[3] Majzoobi, M., Koushanfar, F., and Potkonjak, M.: Lightweight Secure PUFs, Proc. of IEEE/ACM Int. Conf. on Computer Aided Design (ICCAD), pp.670–673 (2008).
[4] Suh, G. E. and Devadas, S.: Physical Unclonable Functions for Device Authentication and Secret Key Generation, Proc. of 44th ACM/IEEE Design Automation Conf. (DAC), pp.9–14 (2007).
[5] Guajardo, J., Kumar, S. S., Schrijen, G. J., and Tuyls, P.: FPGA Intrinsic PUFs and Their Use for IP Protection, Proc. of 9th Int. Workshop on Cryptographic Hardware and Embedded Systems (CHES 2007), LNCS 4272, pp.63–80, Speinger-Verlag (2007).
[6] Lim, D.: Extracting Secret Keys from Integrated Circuits, M.S. thesis, MIT (2004).
[7] Rührmair, U., Sölter, J., Sehnke, F., Xu, X., Mahmoud, A., Stoyanova, V., Dror, G., Schmidhuber, J., Burleson, W., and Devadas, S.: PUF Modeling Attacks on Simulated and Silicon Data, IEEE Trans. on Information Forensics and Security, vol. 8, no. 11, pp.1876–1891 (2013).
[8] Mahmoud, A., Rührmair, U., Majzoobi, M., and Koushanfar, F.: Combined Modeling and Side Channel Attacks on Strong PUFs, IACR Cryptology ePrint Archive: Report 2013/632 (2013).
[9] Gandolfi, K., Mourtel, C. and Olivier, F.: Electromagnetic Analysis: Concrete Results, Proc. of 3rd Int. Workshop on Cryptographic Hardware and Embedded Systems (CHES 2001), LNCS 2162, pp.251–261, Springer-Verlag (2001).
[10] Meynard, O., Guilley, S., Danger, -L. J. and Sauvage, L.: Far Correlation-based EMA with a Precharacterized Leakage Model, Proc. of Design, Automation and Test in Europe Conference and Exhibition (DATE 2010), pp.977–980 (2010).
[11] 庄司陽彦，角尾幸保，板倉征男，: FPGA に対する漏洩電磁波の局所性を利用した電磁波解析，2010 年暗号と情報セキュリティシンポジウム講演論文集，3B3-2，pp.1–6 (2010).
[12] 森田秀一，松本　勉，高橋芳夫，四方順司 : 暗号ハードウェアの局所情報と電磁波解析 (その 3)，2011 年暗号と情報セキュリティシンポジウム講演論文集，2D3-2，pp.1–7，(2011).
[13] 中井綱人，汐崎　充，久保田貴也，菅原　健，鈴木大輔，藤野　毅：レジスタに値を保持しているだけで生じる静的なサイドチャネルリーク，2015 年暗号と情報セキュリティシンポジウム講演論文集，2F3-4，pp.1–7 (2015).
[14] Research Institute for Secure Systems, AIST, : Evaluation Environment for Side-channel Attacks, http://www.risec.aist.go.jp/project/sasebo
[15] A Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
6ページ右段上から6行目～8行目	**謝辞**　この成果は，国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の助成事業の結果得られたものです．	**謝辞**　この成果は，国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の委託業務の結果得られたものです．