Power-gated Static Random-Access Memory-based Physically Unclonable Function

Yujin Zheng, Alex Bystrov and Alex Yakovlev
Microsystems Group, Newcastle University, NE1 7RU, UK

An architecture and implementation of a Physically Unclonable Function (PUF) based on Static Random-Access Memory (SRAM) is developed to accommodate stochastic processing of PUF *Response*. An 8T-PUF cell is built to eliminate data retention and maximise physical mismatch. The power gating method is used to perform measurement on a subset of PUF cells repeatedly at a high rate. Two-phase power gating is applied selectively to parts of PUF matrix and is designed for minimising EMI and crosstalk amongst the cells during metastability resolution whilst maintaining high performance. The design is meant to serve as a platform for extraction of a variety of stochastic metrics for subsequent inclusion into PUF *Responses*, which comprises the novelty of the approach.

Keywords: Physically Unclonable Function, PUF, power gating, SRAM, hardware security, metastability, reliability.

Introduction: Internet of Things (IoT) devices, such as wearables, bank tokens, medical implants, smart home gadgets or self-driving vehicles, are facing security attacks. To build a strong physical cyber-defence capability from the very start of the Integrated Circuit (IC) design [1], PUFs are promising hardware security primitives because they are easy to evaluate and physically hard to duplicate [2]. PUFs are functions which map *Challenge* to *Response* through physically unclonable devices [2, 3], and this function is the sole existence for each individual PUF device.

The unclonable physical parameter mismatch comes from the process variation during the fabrication of semiconductor devices. This unique characteristic can be utilised to generate *True Random Numbers* or derive *Secret Key* in cryptography. SRAM is an indispensable part in mainstream embedded designs because of its symmetric structure and mass entropy. SRAM-PUF was first proposed by Guajardo et al. [4] and Holcomb et al. [5] who discovered the intrinsically random start-up values of SRAM cells. Accordingly, these repeatable start-up values are the raw data for creating the *Response* of SRAM-PUF; and the address used to read them are the *Challenge* of the PUF as shown in Fig. 1. However, random logical states inevitably appear after every power-up in a small amount of SRAM cells due to the negligible mismatch between cross-coupled inverters. Consequently, there are some requisite techniques for identifying those unstable cells, such as Multiple Evaluation [6], Temporal Majority Voting (TMV) [7] or dark-bit masking [8]. Moreover, SRAM cells are sensitive to environmental changes which cause bit errors that *Secret Key* cannot tolerate, so bit error correction techniques are vital, such as BCH codes [9] or Hamming codes [10] etc.

Novelties: Our aim is to perform sequential measurements with stochastic processing on those unstable PUF cells, which are discarded in the reference approaches. An architecture of power-gated 8T PUF is presented with the following novelties:

- Custom PUF: The 8T-PUF cell discards the writing process of normal SRAM and only keeps the reading behaviour. Two NMOS transistors are added to the standard 6T-SRAM cell to eliminate data retention so as to protect the PUF cell from security attacks. Moreover, the smallest transistor sizes are chosen to maximise physical mismatch for PUF application.
- No extra fabrication process required: The power-gated 8T PUF does not require a special process for high-density SRAM manufacturing, and it can be implemented in the same process as MCUs.
- Improve security whilst saving energy: Originally, the power-gating technique is to reduce leakage which is independent of the switching activity of transistors, and thus lessen the power dissipation. In our approach, the 8T-PUF cells are partitioned into clusters and only the chosen clusters will be powered up during reading process. Those PUF clusters without power supply cannot be read out and are protected from security attacks.
- Facilitating future stochastic measurements: The power gating architecture can facilitate stochastic measurements by repeatedly applying power on-off cycles to the chosen PUF clusters. Then, the bias probability of each PUF cell can be extracted and marked on bit map to identify unstable PUF cells. Instead of discarding those unstable readings, we intend to use them as part of the PUF *Response*. The high-speed stochastic measurements require suppression of data retention, which is a part of our approach.
- Two-phase power gating method: The power-switching process includes reset stage, phase I power-up and phase II power-up. The reset stage quickly drains the remaining current, drops the virtual power supply *vddv* to *0V* and eliminates retained data. The aim of the phase I power-up is to prolong the metastability resolving process thus to reduce EMI and the crosstalk amongst PUF cells. Then the phase II speeds up the voltage ramp-up process. Moreover, the two-phase method enables different combinations of power-gating parameters in hardware implementation, so the in-rush current can be curbed by adjusting those parameters. In addition, the test circuit will facilitate a thorough investigation into the metastability behaviour of bistable device which results in the unique reading of SRAM-based PUF.

Related PUF Works with Power-gating Method: Previously, there have been several PUF applications [11, 12, 13] applying the power-gating method but their purposes and implementations are different from this approach. In 2016, Holcomb et al. [11] utilised the random duration of multiple power gating to replace voltage control so as to induce failures for finding out the Data Retention Voltage (DRV). Although with different purpose, this research and its predecessor [14] clarified that "a strong DRV fingerprint is correlated to power-up tendency". It substantiates that the stability or instability degrees of PUF cells comes from their innate physical mismatches. However, the shortest single evaluation cycle is more than 2μs which is about 40 times more than our approach. Afterwards, a 2-D power-gating scheme to relieve enhancement-enhancement (EE) SRAM-PUF from short-circuit current and also to protect PUF data from attacks was presented by Liu et al. [12]. However, there was no consideration of the major power-gating parameters, such as *SLEEP* transistor design [15] or power distribution network [16] etc. This scheme also has a half-selected cell problem which requires additional peripheral circuits to lessen extra energy consumption.

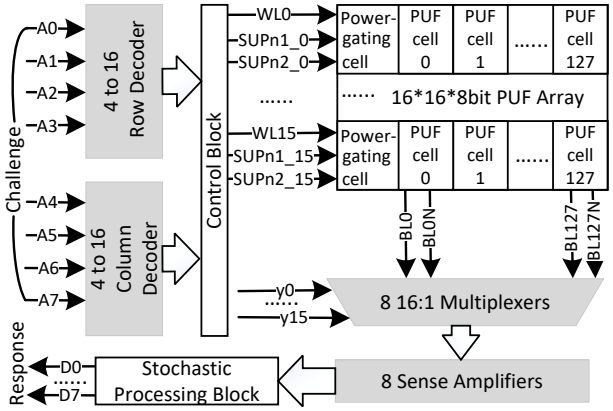

Fig. 1. *A 2kb Power-gated PUF Architecture*

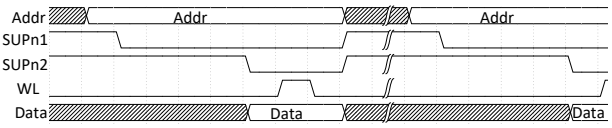

Fig. 2. *Protocol of Two-phase Power Gating*

Power-gated PUF Architecture: An example architecture of the power-gated PUF with 8-bit address as input and 8-bit data as output is illustrated in Fig. 1. It consists of some general SRAM function blocks in grey, a 2kb power-gated PUF array and two functional blocks, i.e. a Control block, a Stochastic Processing block. In the PUF array, the power supply of each PUF row is controlled by a power-gating cell. Since our main purpose of power gating is facilitating stochastic experiments by the switching activity of the power supply, the general term *SLEEP* for normal power gating is replaced by *SUPPLY* in our design. The switching activity is controlled

by the Control block which correlates *SUPn* signals with *WL* (Word Line) signals decoded from address. The power supply will be switched on with the *SUPn* signal discharging to *'0'* for a reading process and off after the reading while the *SUPn* signal asserting to *'1'*. The protocol of two-phase power gating is shown in Fig. 2. The Stochastic Processing block is to evaluate and mark the raw read-out data for improving PUF entropy. The input vector and the output vector can be concatenated between multiple tiles to match the dimension of the primary *Challenge* and *Response*, e.g. 128 bits. The parameters of this architecture, such as the type of PUF cell, the number of rows or columns, or the word width etc., can be altered for various purposes or implemented in different fabrication techniques.

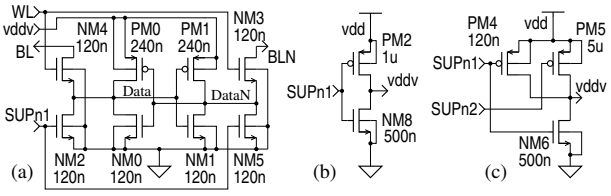

Fig. 3 *Schematics of (a) 8T-PUF cell, (b) Single-phase Power-gating Cell and (c) Two-phase Power-gating Cell*

Design of 8T-PUF Cell: Our newly developed 8T-PUF cell shown in Fig. 3(a) stems from conventional 6T SRAM [17]. The PUF characteristic comes from the threshold voltage variations of four transistors, namely *PM0*, *PM1*, *NM0*, and *NM1*. These transistors compose two cross-coupled inverters and store data in them. After powering up, the mutual input and output nodes, i.e., *Data* and *DataN*, cross the metastable state [18] at which time the outputs of two cross-coupled inverters are lingering between logical '0' and '1', and then settle in one of these stable states. Different from 6T SRAM, two NMOS transistors *NM2* and *NM5* are added for rapidly discharging the *Data* and *DataN* nodes during reset stage with *SUPn* signal asserting. As a result, the reset time can be shortened by magnitude, i.e. from microsecond to nanosecond, so as to facilitate the high-rate operation of PUFs. In addition, no retention data can be exploited by attackers [19].

Design of Power-gating Cells: In our architecture Fig. 1, each power-gating cell controls the power supply of a cluster of 128 PUF cells. There are two kinds of newly designed power-gating cells in our experiment, namely the single-phase power-gating cell shown in Fig. 3(b) and the two-phase power-gating cell illustrated in Fig. 3(c). The single-phase power-gating cell consists of a PMOS as *SUPPLY* transistor to switch power on and off, and an NMOS for discharging currents quickly at reset stage. In contrast, the two-phase power-gating cell has two *SUPPLY* transistors. Phase I utilise the smallest *SUPPLY* transistor *PM4* to curb the *vddv* output. On the other hand, Phase II employs a relatively larger *PM4* to release the *vddv* quickly.

Several aspects have been considered for the trade-off amongst performance, area overhead and energy efficiency whilst ensuring enough PUF entropy for security. Note, those effects vary in different technologies and processes, so they should be evaluated for each individual design [15].

The conventional power gating is conducted via *SLEEP* transistor to enable power or ground connection. Our power-gating design employs PMOS as *SUPPLY* transistor for switching power connection. Although NMOS transistor of the same width has smaller on-resistance and causes less voltage drop in power supply chain [20], its layout implementation needs an extra WELL to segregate virtual ground from P-substrate. Since our design utilises the lower voltage, NMOS transistor for ground connection is not our preference.

Secondly, the body (substrate) bias of the *SUPPLY* transistor should be determined. There are three kinds of body bias, namely normal body bias, forward body bias (FBB) and reversed body bias (RBB). The reversed body bias connects the substrate of the *SUPPLY* transistor to a voltage lower than the source voltage so as to increase the threshold voltage V_T. This brings some benefits such as the reduction of current leakage and the improvement of switch efficiency etc. [15]. In our case, the normal body bias is applied by connecting the substrate and the source of the *SUPPLY* transistor to *vdd* directly, because the FBB or RBB hardware implementation needs extra voltage supply or a separated WELL for ground connection. Meanwhile, the standard process high threshold voltage *SPHVT* transistor in 90nm CMOS technology is chosen as the *SUPPLY* transistor to achieve relatively higher V_T.

As a final point, the size of the *SUPPLY* transistor is a crucial parameter for SRAM-based PUF application because it not only affects the voltage level of *vddv* but also dictates the voltage ramp-up time and the in-rush current. Since smaller transistor curbs the drain current and the *vddv* output, our two-phase power-gating exploits this to generate the gentle voltage incline with the smallest *SUPPLY* transistor in phase I and create a steep slope of *vddv* using a larger *SUPPLY* transistor in phase II. Consequently, the metastability resolving time in phase I is lengthened in the hope of minimum mutual disturbance amongst PUF cells. On the other hand, if all PUF cells start metastability in a very short period of time, the in-rush current will be large because the two pairs of cross-coupled transistors in an 8T-PUF cell are both in saturation mode. This prolonged phase I is also able to flatten the current peak. Afterwards, phase II will take control and raise up *vddv* swiftly. The sizes of two *SUPPLY* transistors and the time durations of two phases can be manipulated to get the best trade-off. This will be discussed in simulation part.

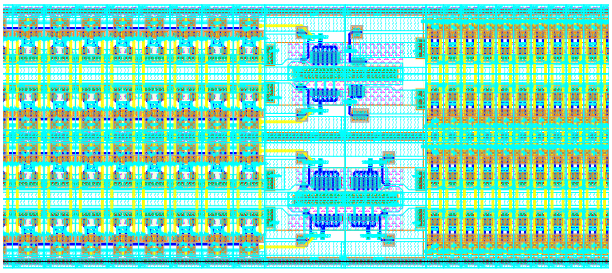

Fig. 4. *Part of the Power-gated PUF Test Circuit Layout*

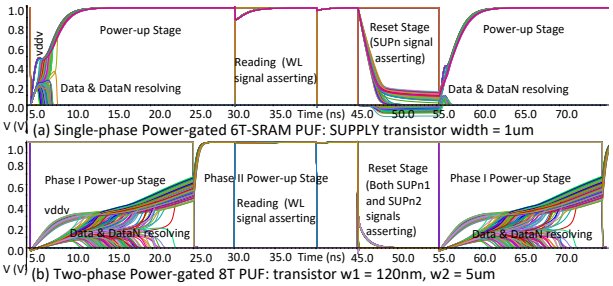

Fig. 5. *Waveforms of a 100-Run Post-layout Monte-Carlo Simulation*

Simulation of Power-gated PUF: Our test circuit of power-gated PUF without peripheral blocks is implemented in UMC 90nm technology as shown in Fig. 4. It includes rows of single-phase power-gated 6T-SRAM PUF cells, two-phase power-gated 6T-SRAM PUF cells and two-phase power-gated 8T PUF cells with different sets of power-gating parameters. Each row has 128 PUF cells. This circuit enables the post-layout simulation which investigate the relationship amongst the outputs of PUF cells, their supply voltages and time duration of three power-up stages.

To determine the size of *SUPPLY* transistors and evaluate the behaviour of SRAM-based PUF cells, the post-layout Monte-Carlo simulations with the Gaussian distribution of transistor threshold voltages have been carried out under the supply voltage of *1V*, ambient temperature of *27 °C* and *TT* (Typical-Typical) process corner without external noise. Although the test circuit can work at the temperature range from *-55 °C* to *125 °C* or at different process corners, the evaluation at different PVT (process, voltage, temperature) corners will be the future work.

In Fig. 5, 100 runs of post-layout Monte-Carlo simulation results illustrate different power-up and reset behaviour associated with 6T-SRAM or 8T-PUF arrays, the *SUPPLY* transistor width and resolving time. Fig. 5 (a) demonstrates that the *vddv* of single-phase power-gating with *1um* width *SUPPLY* transistor reaches *1V* in roughly *7ns* whilst resolving metastability of inside node pairs, i.e., *Data* and *DataN*. In comparison, Fig. 5 (b) illustrates that *vddv* reaches around *352mV* at *9ns* and lingers for about *2ns* then gradually climbs up towards *1V* at phase I power-up stage with the *SUPPLY* transistor width *120nm*. Following this, phase II starts from *25ns* and swiftly raises *vddv* to *1V* in around *1ns*. On the other hand, these low start-up voltages in phase I lengthen the metastability resolving time of PUF cells. It can be seen that *Data* and *DataN* start wrestling while *vddv* is ramping up slowly, then escape out of metastability and tend to their

distinct logical status in various resolving times due to their intrinsically various physical parameters. These opposite tendencies of *Data* and *DataN* resemble the random PUF behaviour in real circuits. After powering up, a short plunge of *vddv* appears in Fig. 5 (a) whilst *WL* signal is asserting for reading. By contrast, Fig.5 (b) displays only slight fluctuation in two-phase power-gating circuit. Finally in the reset stage, PUF cells are powered down and drain the remaining currents away to prepare for the next cycle. However, unlike Fig. 5 (b) resetting within *10ns*, Fig. 5 (a) shows that the traditional 6T-SRAMs have obvious retention data which not only affect the initial states of *Data* and *DataN* pairs but also are targets of attackers.

The linear relationship between the *SUPPLY* transistor width and the approximate value of *vddv* plateau before the visible metastability resolving and the quadratic dependence of the *SUPPLY* transistor width and the metastability resolving time duration is shown in Fig. 6. The PUF cells with the starting *vddv* lower than *400mV* resolve metastability above *10ns* whereas the others reach to their stable states much quicker with the relatively higher *vddv* supplied by larger *SUPPLY* transistors. Hence, by varying the size of *SUPPLY* transistor, the metastability resolving time can be manipulated. Meanwhile, Fig. 7 illustrates the comparison amongst the current peaks of different signal edges in various combinations of *SUPPLY* transistors whilst power gating 128-bit PUF. It can be seen that the higher power-up current corresponds to larger *SUPPLY* transistor whilst the phase II current also associates with the *vddv* value at the start of phase II. The energy consumption of a PUF cell for each reading cycle is roughly from *3fJ* to *4.2fJ*. Those differences are negligible provided that higher entropy can be achieved, because the principal aim of our trade-off amongst area, time and energy consumption is for stronger physical cyber-defence ability.

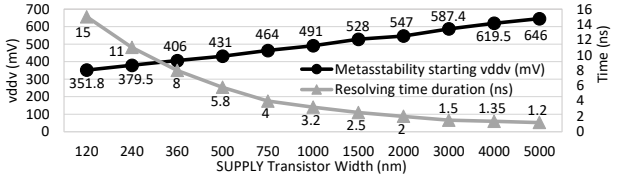

Fig. 6. *SUPPLY Transistor Width vs* vddv *and Resolving Time*

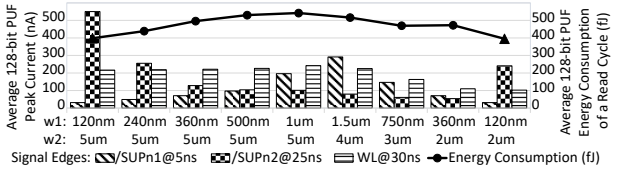

Fig. 7. *SUPPLY Transistor Size vs Current and Energy Consumption*

Conclusion and Future Work: In the paper, an example of power-gated SRAM-based PUF design is presented, whose purpose is to enhance PUF entropy. It can repeatedly apply on-off power cycles to PUF clusters to facilitate stochastic experiments for extracting the bias probability of PUF cells. The extracted information of unstable cell will be added to PUF *Response* to increase entropy. This design can also control the power-up process via varying power gating methods and parameters so as to decrease the interference amongst sensitive PUF cells and limit the in-rush current during power-up. In addition, the post-layout simulation demonstrates the metastability resolving process and each PUF cell reaching their implicit logical levels which resemble the real-world PUF behaviour. In the design, an SRAM-based 8T PUF cell with the ability to eliminate data retention is built, and a two-phase power-gating method is devised and evaluated. Apart from switching power supply, the two-phase power gating cell prolongs the phase I to avoid PUF cells interfering each other and speeds up the power-up by the phase II.

Our future work will include more simulations on various PVT corners to further evaluate their behaviour under different power-gating methods. Numerous post-layout Monte-Carlo simulations on the test circuits are ongoing to quantitatively assess the entropy improvement of the two-phase power-gated 8T PUF comparing to its single-phase power-gated 6T-SRAM PUF counterpart. Those results can also be used to study metastable behaviour. Moreover, the Control block and the Stochastic Processing block will be implemented, and then be aggregated with the general SRAM function blocks, the power-gated 6T-SRAM PUF arrays and the two-phase power-gated 8T-PUF arrays with various parameters in a test chip. Afterwards, the fabricated test chips will enable on-chip stochastic experiments which will be employed to map the biased cells, evaluate the bias percentage and then extract analogue secrets. Finally, the PUF entropy differences with various power-gating methods and settings will be assessed and reported.

References

1 Verbauwhede, I.: 'Security Adds an Extra Dimension to IC Design: Future IC Design Must Focus on Security in Addition to Low Power and Energy', *IEEE Solid-State Circuits Magazine, 9(4)*, 2017, p. 41–45.

2 Gassend, B., Clarke, D., Van Dijk, M. and Devadas, S.: 'Silicon physical random functions', *Proceedings of the 9th ACM conference on Computer and communications security*, 2002, p. 148–160.

3 Gassend, B., Clarke, D., Van Dijk, M., and Devadas, S.: 'Controlled physical random functions', *In 18th Annual Computer Security applications Conference, Proceedings*, 2002, p. 149-160.

4 Guajardo, J., Kumar, S.S., Schrijen, G.-J. and Tuyls, P.: 'FPGA Intrinsic PUFs and Their Use for IP Protection', *International workshop on cryptographic hardware and embedded systems*, 2007, p. 63–80.

5 Holcomb, D.E., Burleson, W.P. and Fu, K.: 'Initial SRAM State as a Fingerprint and Source of True Random Numbers for RFID tags', *Conference on RFID Security, Vol. 7, No. 2*, 2007, p. 01.

6 Bhargava, M., Cakir, C., and Mai, K.: 'Reliability enhancement of bi-stable PUFs in 65nm bulk CMOS', *In 2012 IEEE International Symposium on Hardware-Oriented Security and Trust*, 2012, p. 25-30.

7 Mathew, S.K., Satpathy, S.K., Anders, M.A., Kaul, H., Hsu, S.K., Agarwal, A., Chen, G.K., Parker, R.J., Krishnamurthy, R.K. and De, V.: '16.2 A 0.19pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22nm CMOS', *In 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, p. 278–279.

8 Satpathy, S., Mathew, S., Li, J., Koeberl, P., Anders, M., Kaul, H., Chen, G., Agarwal, A., Hsu, S. and Krishnamurthy, R.: '13fJ/bit probing-resilient 250K PUF array with soft dark-bit masking for 1.94% bit-error in 22nm tri-gate CMOS', *In ESSCIRC 2014-40th European Solid State Circuits Conference (ESSCIRC)*, 2014, p. 239-242.

9 Bose, R.C. and Ray-Chaudhuri, D.K.: 'On a class of error correcting binary group codes', *Information and control, 3(1)*, 1960, p.68-79.

10 Hamming, R.W.: 'Error Detecting and Error Correcting Codes', *Bell System Technical Journal, 29 (2)*, 1950, p.147–160.

11 Xu, X. and Holcomb, D.E.: 'Reliable PUF design using failure patterns from time-controlled power gating', *2016 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, 2016, p. 135–140.

12 Li, G., Wang, P., Ma, X., Shi, Y., Chen, B., and Zhang, Y.: 'A multimode configurable physically unclonable function with bit-instability-screening and power-gating strategies', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 29(1)*, 2020, p. 100-111.

13 Liu, K., Min, Y., Yang, X., Sun, H. and Shinohara, H.: 'A 373-F^2 0.21%-Native-BER EE SRAM Physically Unclonable Function With 2-D Power-Gated Bit Cells and V_{ss} Bias-Based Dark-Bit Detection', *IEEE Journal of Solid-State Circuits, 55 (6)*, 2020, p. 1719–1732.

14 Holcomb, D.E., Rahmati, A., Salajegheh, M., Burleson, W.P. and Fu, K.: 'DRV-Fingerprinting: Using Data Retention Voltage of SRAM Cells for Chip Identification', *International Workshop on Radio Frequency Identification: Security and Privacy Issues*, 2013, p. 165–179.

15 Keating, M., Flynn, D., Aitken R., Gibbons, A., and Shi, K.: 'Sleep Transistor Design', *Low power methodology manual: for system-on-chip design*, 2007, p. 249-265.

16 Keating, M., Flynn, D., Aitken R., Gibbons, A., and Shi, K.: 'Design of the Power Switching Network', *Low power methodology manual: for system-on-chip design*, 2007, p. 225-247.

17 List, F.J.: 'The static noise margin of SRAM cells', *In ESSCIRC'86: Twelfth European Solid-State Circuits Conference*, 1986, p. 16-18.

18 Kleeman, L., and Cantoni, A.: 'Metastable behavior in digital systems', *IEEE Design & Test of Computers, 4(6)*, 1987, p.4-19.

19 Halderman, J.A., Schoen, S.D., Heninger, N., Clarkson, W., Paul, W., Calandrino, J.A., Feldman, A.J., Appelbaum, J. and Felten, E.W.: 'Lest we remember: cold-boot attacks on encryption keys', *Communications of the ACM, 52(5)*, 2009, p.91-98.

20 Roy, K., Mukhopadhyay, S. and Mahmoodi-Meimand, H.: 'Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits', *Proceedings of the IEEE, 91(2)*, 2003, p.305-327.