

Probabilistic CMOS (PCMOS) Logic for Nanoscale Circuit Design

Krishna V. Palem^{1,4}, Pinar Korkmaz², Kiat-Seng Yeo^{3,4}, and Zhi-Hui Kong^{3,4}

¹*Department of Computer Science, Electrical and Computer Engineering,*

Department of Statistics and The VISEN Center, Rice University, Texas, USA

²*Intel Corporation, Oregon, USA*

³*Division of Circuits and Systems, School of Electrical and Electronic Engineering,*

Nanyang Technological University (NTU), Singapore

⁴*Institute for Sustainable Nanoelectronics (ISNE), Nanyang Technological University*

(NTU), Singapore

Many believe that as CMOS devices approach the 10 nm scale, noise and a range of other imperfections will render them unstable and hence result in probabilistic behaviour. Specifically, noise, parametric variations, defects, and tunneling effects [1-3] are inevitable impediments, resulting in probabilistic behaviors in silicon-based [3] as well as carbon-based [4] devices. These phenomena warrant a fundamental shift in the extant approaches to the design of circuits and systems—away from the deterministic approaches of the past to those that embody *probabilistic* models and methods. (Please also see Packan [1], Meindl et al. [2] and Borkar [3].) This inevitability motivates the need to extend common “rules of thumb” or rules used by circuit designers, to *actively* model and characterize the impact of probabilities of correctness of the devices as part of the design space—this is in addition to the typical design space dimensions spanning speed, energy (power) and area. In response to this need, in this paper, we provide a

design space of CMOS based devices wherein, the probability of correctness is characterized *explicitly*. Using our characterization, a circuit designer can explicitly understand the relationship between the probability of correctness of a device or switch p , and other attributes such as the energy consumed, or its switching speed.

Thus, similar to conventional approaches to modeling and characterizing the behavior of circuits based on deterministic models of semiconductor devices and switches, an explicit characterization of the behavior of *probabilistic CMOS* (PCMOS) computing logic gates and switches—the ubiquitous building blocks of circuits—is needed in the probabilistic CMOS regime. Following Kish [5], we use thermal noise as the causal factor inducing probabilistic behavior in CMOS devices throughout this paper. Based on this approach, we provide two key rules of thumb to guide circuit designers in the form of the following probabilistic laws (*i*) relating the energy consumed by an inverter as it switches, to its probability of correctness in Figures 1 and 2, and of an exclusive OR gate in Figure 3 and (*ii*) a surprising observation (Figure 4) that the probability of correctness p of an (inverter) switch is uniquely determined by σ , the noise magnitude and V_{dd} , the operating voltage of the switch, and is thus *independent of the feature size of the devices*. (While the specific probabilistic behaviors described in this paper are based on thermal noise, our methodology and approach is not limited to behaviors induced by thermal noise alone—they are applicable to other types of probabilistic phenomena as well [6].)

We continue by demonstrating the value of our laws of rules of thumb or laws to guide the design of a more elaborate circuit for realizing *hyperencryption*—the functional block diagram is shown in Figure 5 and a layout snapshot of the test chip is also included in the

sequel—through which we show that a PCMOS based approach can be more efficient by a multiplicative factor of 205 over its counterpart realized using conventional CMOS technology, through the *energy-performance product* (akin to the energy-delay product) metric. While noise susceptibility is one possible source of probabilistic behavior and is temporal, it is also anticipated that the physical parameters of the transistors will vary significantly and statistically these variations are spatial. To understand the value of our approach to characterizing probabilistic behaviors through the laws of PCMOS, we demonstrate the impact of varying of the *channel length* parameter on the first law stated above, in Figure 6. More precisely, changes in energy per switching step due to parametric variations in *channel length* become more significant for smaller technology nodes.

A *switch* is an element that computes a Boolean function, whereas a *probabilistic switch* is a switch that computes a Boolean function correctly with probability p . Building on our earlier work [7] where the details of designing such a probabilistic switch are outlined, our PCMOS inverter exhibits incorrect probabilistic switching behavior that fluctuates between the digital 1 and 0 at its output due to perturbations by thermal noise, *even when the input is held at a constant value of 0*. For such a switch, the quantitative form of the *energy-probability* law of PCMOS is shown in Figure 1, and states:

In any technology generation, the energy consumed by one switching step of an inverter switch has a lower bound set by exponential function of the probability of correctness p for a fixed noise magnitude σ .

Referred to as the E - p (for energy-probability) relationship, this law is developed in its full mathematical form in our earlier work [7] and is validated in Figure 1 using TSMC's 0.25 μm process through physical measurements for the first time. In detail, the mathematical form of the law relates the energy consumed by a switching step E to p through C , the load capacitance of the transistor and the RMS value of the noise coupled, σ . As shown in Figure 1, this mathematical form, the HSpice simulated results and the physical measurements are in agreement for noise magnitudes σ , of 0.77 V and 0.4 V; additional noise magnitude values have been measured with similar outcomes.

Additionally, as shown in Figures 2 and 3 respectively, we have also validated this law for an inverter and an XOR gate using CHRT's 0.18 μm process as well as a Predictive Technology Model (PTM) at 32 nm feature size, provided by Arizona State University. The fact that our HSpice simulated data is in agreement with our analytical model from [7] and the fact that for XOR gates, measured values are in agreement with our analytical model for [7] is the basis for this claim.

We note that the E - p relationship is technology dependent and the actual energy values decrease in a manner consistent with Moore's law, when we compare the results in Figure 1 for the TSMC 0.25 μm process to those in Figure 2 for smaller feature sizes. Specifically, the switching energy decreases in accordance with Moore's law for a fixed probability of correctness p . However, for a designer to properly understand and use a rule such as this law for trading energy for accuracy or correctness for example, or for deciding on the amount of error-correction to invest at a given energy point to improve the probability of correctness p , the designer must be aware of the details of the particular

technology generation. It would be simpler and more elegant if the probabilistic design approach could be made independent of the particular technology generation as determined by the feature size, much as circuit design rules and intuitions today are meant to span multiple technology generations—especially when quick and clever decisions have to be made about trading area for energy, speed for power and so on.

Surprisingly, we are able to demonstrate this technology independence in the PCMOS context, since it turns out that the probability of correctness p can be characterized in a technology independent manner stated qualitatively as follows for an inverter switch:

The output of any PCMOS based probabilistic (inverter) switch is computed correctly with a probability p that is determined by the ratio of the magnitude of noise σ , to that of the signal V_{dd} termed the noise-to-signal ratio or NSR. Furthermore, this relationship is invariant across technology generations represented by the feature size of the constituent transistors.

The validation of this *law of probabilistic invariance* is shown in Figure 4. Here, the probability of correct switching p is related to the NSR value quantitatively. As seen from Figure 4, illustrating the invariant behavior under different technology generations, the data points from all the physically measured and simulated switches overlap *and match the results expected of the analytical form* developed earlier.

We reiterate that the main goal of these physical characterizations is to help map out the design space with the goal of circuit design in the probabilistic regime—specifically to use probabilistic switching to help realize low-energy computing architectures. To demonstrate this fact, we use the main lesson learned from the first law (Figures 1, 2 and 3) that a “small” amount of correctness can be traded for significant savings in energy. Using the trade-off between energy and correctness, we establish that probabilistic or

PCMOS based circuit designs realizing computing architectures can *simultaneously* yield improvements not only in terms of the energy consumed but also in terms of the performance (or running-time) over conventional circuit design approaches.

Specifically, for the hyperencryption algorithm, our approach involves using a *system-on-a-chip* style design (proposed by Chakrapani et al in [8], the reader is referred to this earlier work for algorithmic details of our hyperencryption implementation) wherein the “control” part of the application is executed deterministically on a conventional CMOS style design, whereas, the part involving the probabilistic steps—these are steps that use a pseudo-random number generator in a conventional design—are realized using PCMOS as a co-processor. Earlier, we used a combination of HSpice, C based simulations and the JouleTrack tool [9] in part to establish that the PCMOS based approach is shown to yield gains of more than a multiplicative factor of 9.38, when compared to the CMOS based approach—this design was based on a design originally published by Chakrapani et al [8] and worked at a maximum speed of 1 MHz in both the PCMOS and CMOS cases. Since then, we have redesigned the circuit whose block diagrammatic representation is shown in Figure 5—this redesigned version operates at a much higher frequency of 2.5 MHz using the CHRT’s 0.18 μm process. In contrast, the counterpart CMOS based design wherein the random source labeled RESINA is replaced by a pseudo-random number generator, which operates at an effective frequency of 317 kHz to generate 32 random bits. Additionally, we demonstrate for the first time through measurement of this circuit realized using the CHRT’s 0.18 μm process that our PCMOS design is more efficient than its CMOS counterpart by a factor of 205, using the energy-performance product of the two

designs as the metric. (The energy consumption of the CMOS counterpart is derived from simulations)

The circuits described above embody probabilistic behavior at the application level naturally—hyperencryption for cryptography being the example in this paper. In contrast, applications might be traditionally designed to be deterministic even though the need for 100% correctness might not be essential. Signal processing and graphics workloads are good examples wherein, errors in the low-order bits can quite often result in tolerable degradation in video or audio quality, in return for significant energy savings based on the E-p relationships described above. A simulated proof of this concept has been published by George et al [10] wherein adders and multipliers rendered probabilistic as described above, yielded energy savings of a (multiplicative) factor 2 to a factor of 5 in the context of FIR filters and the FFT. This provides a prospective approach for using our design rules based on the PCMOS characterizations presented above in mainstream designs that were traditionally considered to be purely deterministic!

References

- [1] P. A. Packan, "Device physics: Pushing the limits," *Science*, vol. 285, no. 5436, pp. 2079-2081, Sept. 1999.
- [2] J. D. Meindl, Q. Chen, and J. A. Davis, "Limits on silicon nanoelectronics for terascale integration," *Science*, vol. 293, no. 5537, pp. 2044-2049, Sept. 2001.
- [3] Shekhar Y. Borkar, "Designing reliable systems from unreliable components: The challenges of transistor variability and degradation," *IEEE Micro*, vol. 25, issue 6, pp. 10-16, Nov. 2005.

- [4] R. I. Bahar, J. Mundy, and J. Chen, "A probabilistic-based design methodology for nanoscale computation," in *Proceedings of the International Conference on Computer-Aided Design*, pp. 480-486, Nov. 2003.
- [5] L.B. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics," in *Physical Review Letters A*, pp. 144-149, vol. 305, 2002.
- [6] K. Natori and N. Sano, "Scaling limit of digital circuits due to thermal noise," *Journal of Applied Physics*, vol. 83, no. 10, pp. 5019–5024, May 1998.
- [7] P. Korkmaz, B. E. S. Akgul, K. V. Palem, and L. N. Chakrapani, "Advocating noise as an agent for ultra low-energy computing: Probabilistic CMOS devices and their characteristics," *Japanese Journal of Applied Physics*, vol. 45, no. 4B, pp. 3307-3316, Apr. 2006.
- [8] L. N. Chakrapani, P. Korkmaz, B. E. S. Akgul, and K. V. Palem, "Probabilistic system-on-a-chip architectures," *ACM Transactions on Design Automation of Electronic Systems*, vol. 12, no. 3, article 29, Aug. 2007.
- [9] A. Sinha and A. Chandrakasan, "JouleTrack: A web based tool for software energy profiling," in *Proceedings of the 38th Conference on Design Automation*, pp. 220–225, June 2001.
- [10] J. George, B. Marr, B. E. S. Akgul, and K. V. Palem, "Probabilistic Arithmetic and Energy Efficient Embedded Signal Processing," *Proceedings of the Intl. Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES)*, Seoul, Korea, October 23-25, 2006. (Received Best Paper Award.)

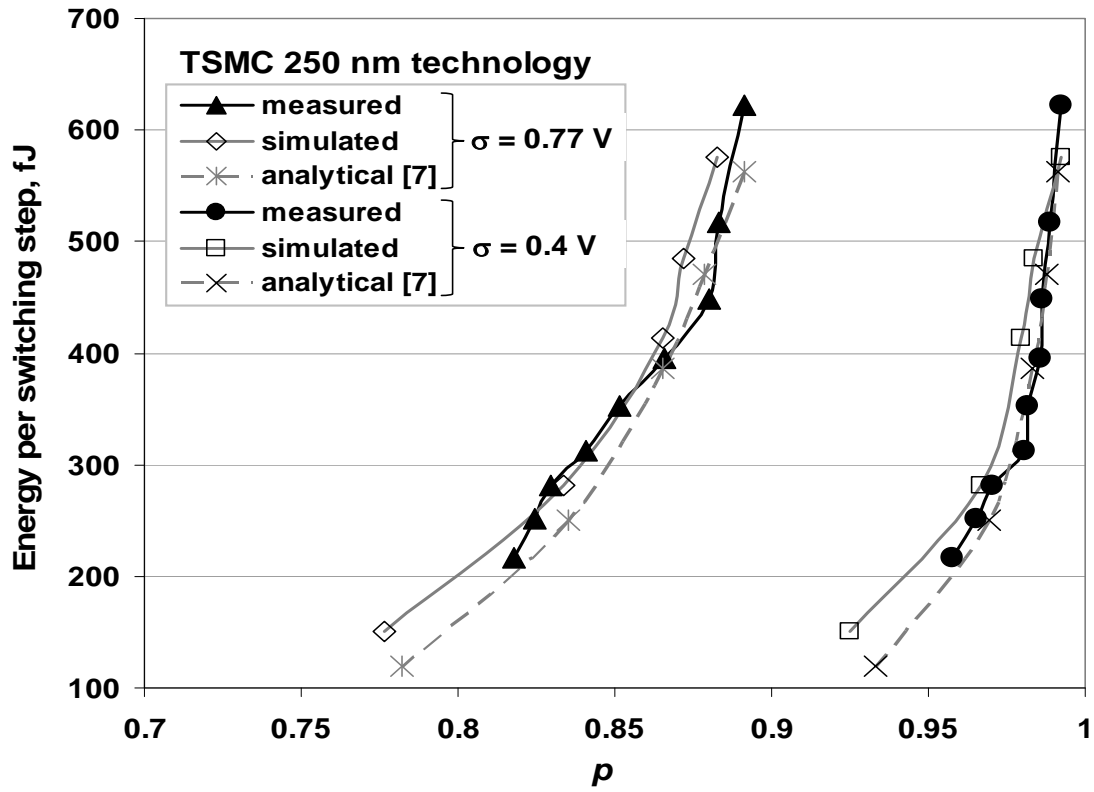


Figure 1: The E - p relationship of PCMOS inverters. Measurement, simulation, and analytical results based on TSMC 0.25 μm technology with a noise magnitude of 0.77V RMS and 0.4V RMS.

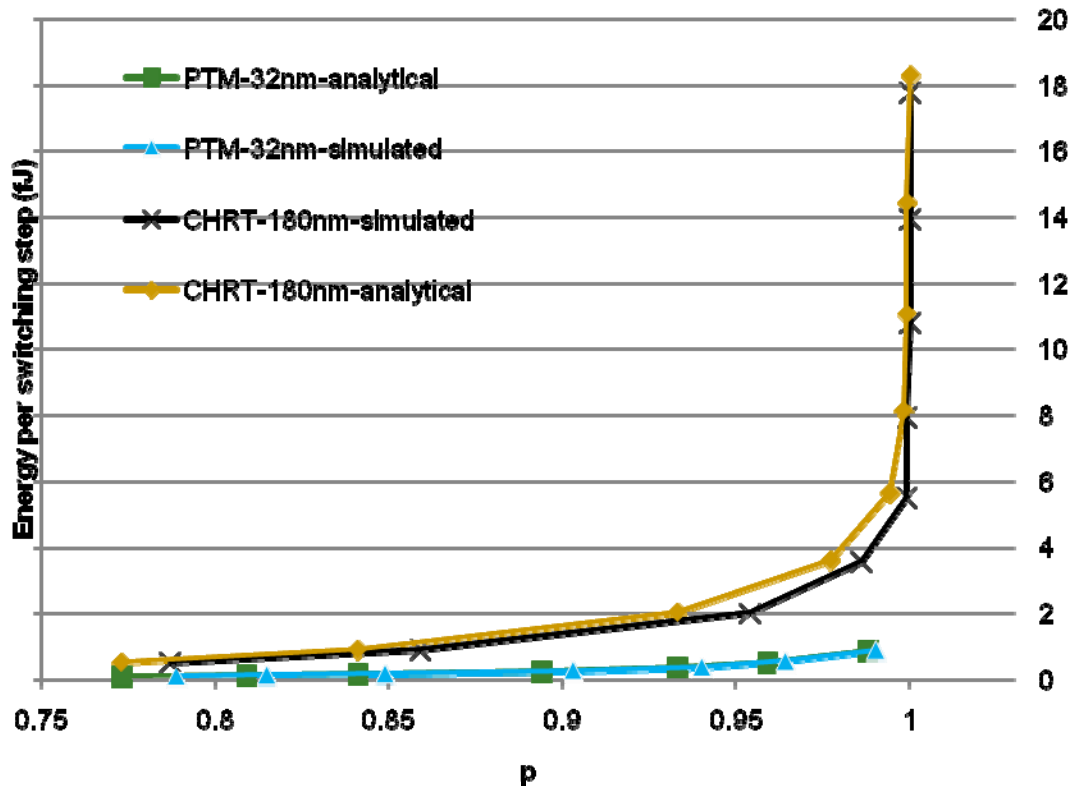


Figure 2: The E - p relationship of PCMOS inverters. Simulated and analytical results based on CHRT 0.18 μm technology and simulated and analytical results for predictive PTM 32 nm technology.

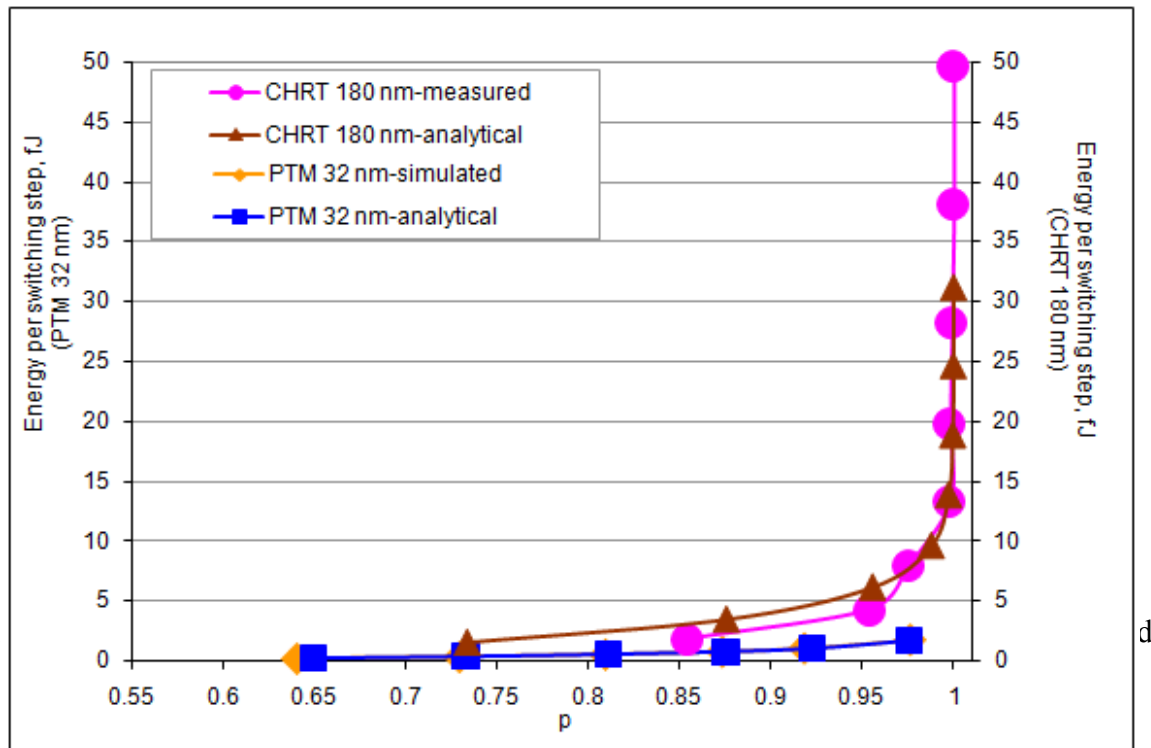


Figure 3: The E - p relationship of PCMOS XOR gates. . Measured and analytical results based on CHRT 0.18 μm technology and simulated and analytical results for predictive PTM 32 nm technology.

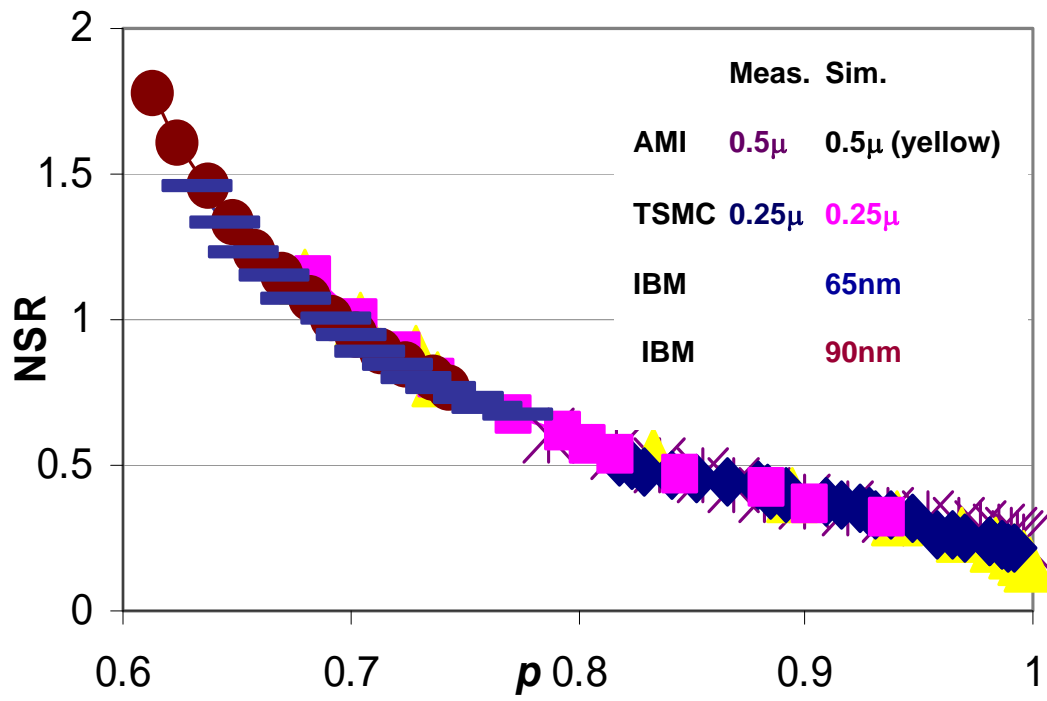


Figure 4: Validation of the law of invariance across technology generations

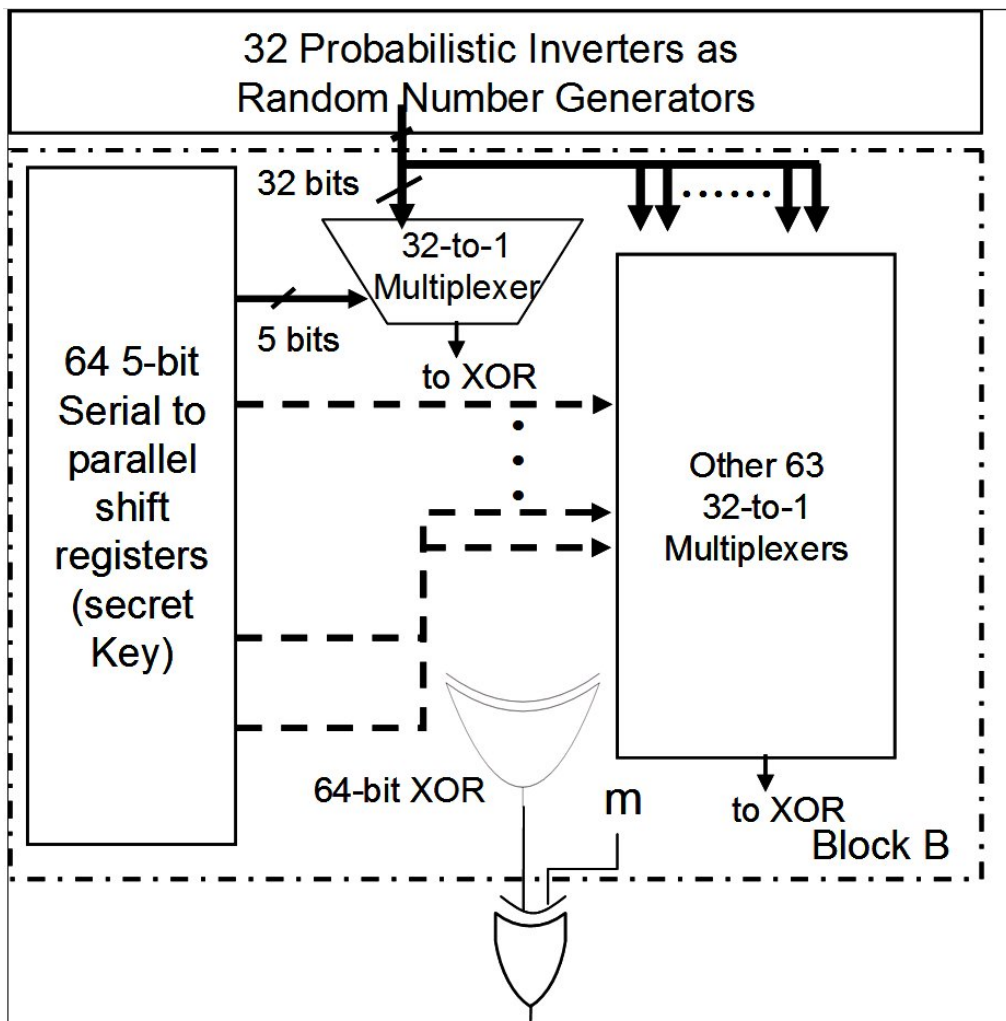
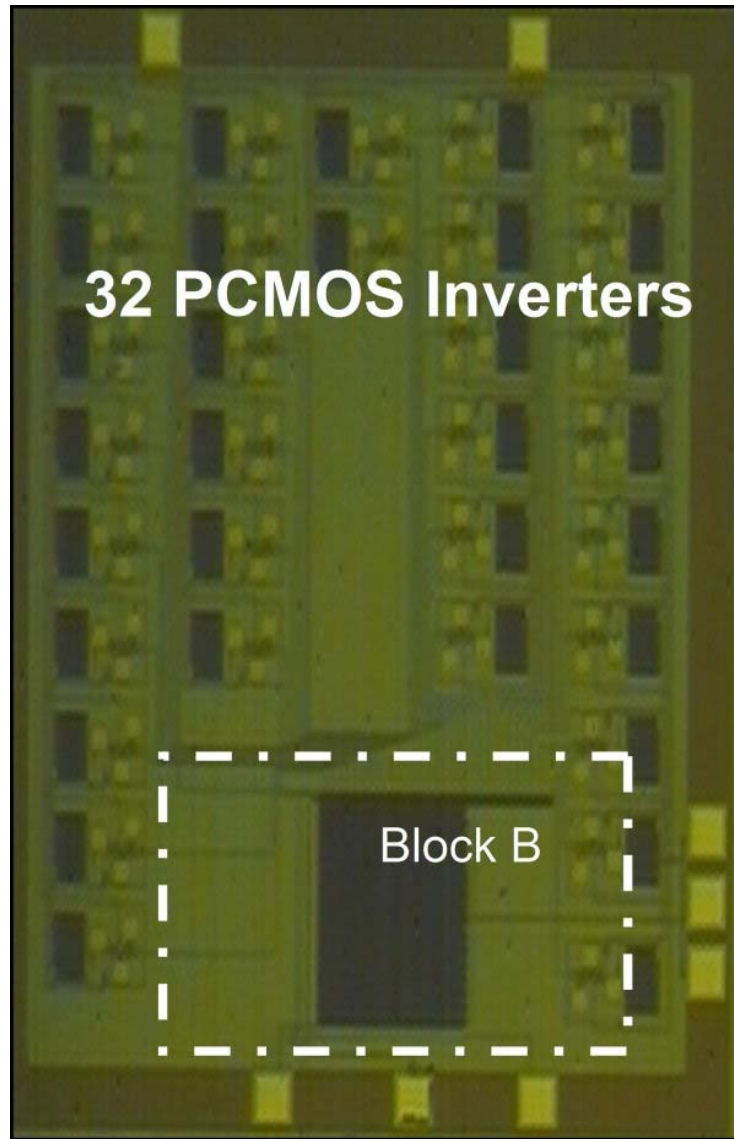


Figure 5: A block diagrammatic representation of the hyperencryption circuit



A Microphotograph of the test-chip, implementing the circuit from Figure 5, using the CHRT 0.18 μm process technology

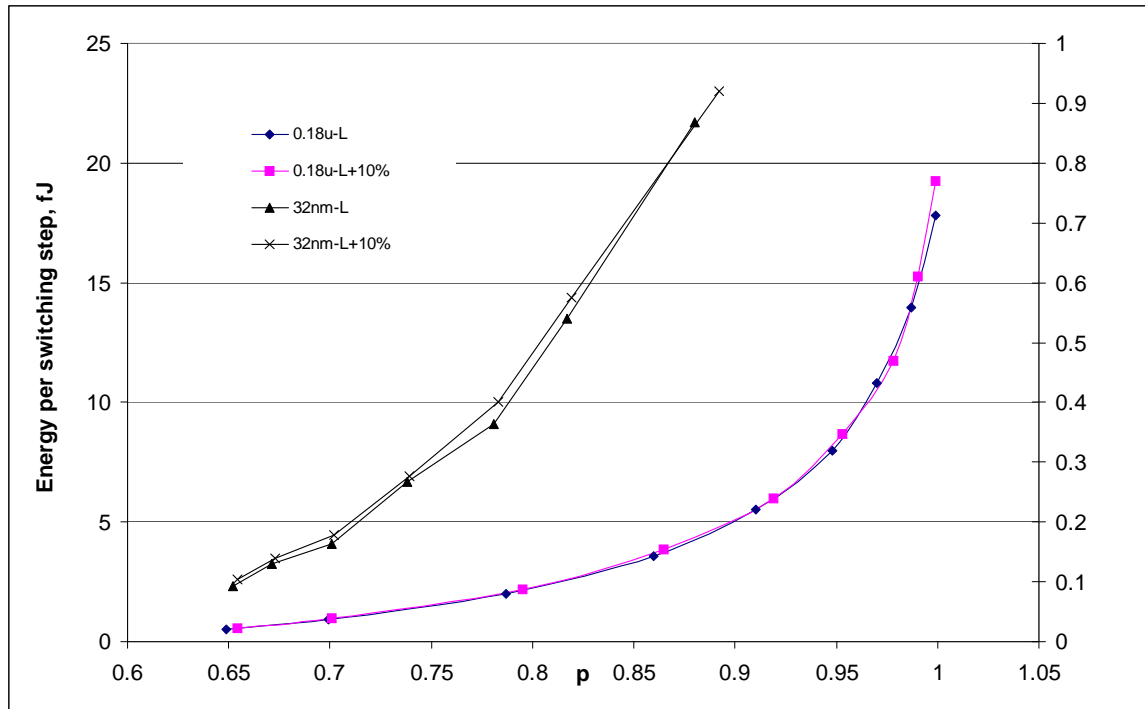


Figure 6: Effect of channel length variation on the behavior of PCMOS XOR gate based on CHRT 0.18 μm and PTM 32 nm technologies through simulations.