جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Approximate Computing With Stochastic Transistors' Voltage Over-Scaling

Item Type	Article
Authors	Li, Ren;Naous, Rawan;Fariborzi,Hossein;Salama, Khaled N.
Citation	Li R, Naous R, Fariborzi H, Salama KN (2019) Approximate Computing With Stochastic Transistors' Voltage Over-Scaling. IEEE Access 7: 6373–6385. Available: http://dx.doi.org/10.1109/ACCESS.2018.2889747.
Eprint version	Publisher's Version/PDF
DOI	10.1109/ACCESS.2018.2889747
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Journal	IEEE Access
Rights	This is under the open access.
Download date	2024-04-17 06:02:07
Link to Item	http://hdl.handle.net/10754/631278

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 17, 2018, accepted December 19, 2018, date of publication December 27, 2018,
date of current version January 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889747

Approximate Computing With Stochastic Transistors' Voltage Over-Scaling

REN LI, (Student Member, IEEE), RAWAN NAOUS, (Student Member, IEEE), HOSSEIN FARIBORZI, (Member, IEEE), AND KHALED NABIL SALAMA, (Senior Member, IEEE)

Department of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Corresponding authors: Ren Li (ren.li@kaust.edu.sa), Rawan Naous (rawan.naous@kaust.edu.sa), Hossein Fariborzi (hossein.fariborzi@kaust.edu.sa), and Khaled Nabil Salama (khaled.salama@kaust.edu.sa)

ABSTRACT Ubiquitous computing and the ever-rising need for energy efficiency pose challenges in terms of the processing requirements and the corresponding machine complexity. Nonetheless, the nature of the underlying applications, particularly dealing with real-world data, offers alternative paradigms toward the efficient utilization of the available design resources. In this paper, approximate computing is addressed as an accommodating technique that can benefit from the inherent resilience of the current applications to build low-power and low-complexity architectures. This paper proposes an alternative way of attaining approximation based on transistor dynamic variability. Furthermore, it presents a comprehensive study using voltage scaling scheme, starting from the impact of variation on the circuit-level output and investigating cascaded logic gates, storage elements, arithmetic building blocks, and on the application level, with an image compression outcome using 2-point discrete Fourier transform as a proof of concept. This paper addresses design analysis metrics and the efficiency of the proposed technique with respect to the technology node, operating frequency, energy and delay, process corner, and temperature. The configurable designs are shown to be possible with adaptive voltage scaling and energy-quality scalability. The proposed technique offers compromises in terms of the circuit design metrics with savings of up to 90% on energy for image compression application, in comparison with running at deterministic nominal value, while preserving the relative quality and accuracy of the output.

INDEX TERMS Approximate computing, ultra-low power circuit design, stochastic electronics, thermal noise, transistor variability.

I. INTRODUCTION

Applications such as speech analysis, image capturing and compression, real-world data sensing, and multimedia processing have relative and approximate accuracy quantification. That is, where levels of error in the output are tolerated without any tangible effect on the perceived quality depends mainly on human perception. This mode of operation paves the way for a wide and innovative design space for the hardware implementations of the underlying applications [1]–[5]. Hence, this mode primarily benefits from the accuracy relaxation to achieve savings in alternative metrics such as energy. Thereby, energy-quality scalable designs surface and capture the essence of approximate computing approaches. In contrast to the relaxed hardware approaches [6], [7], where levels of software and hardware support for fault recovery are added to ensure correct operation, approximate computing techniques tolerate the generated errors and relax the overhead of error correcting schemes.

Different design levels can be investigated in the realm of approximate computing. Application and algorithm design- level techniques include skipping computations and relaxing global synchronization and communication [8]. At the architectural level, specific accelerators and approximate programmable processors are some of the techniques used [9]. In terms of storage elements, particularly with emerging technologies, approximate memories are tackled from the perspective of associative memories and ternary content addressable memories. These novel architectures are used to accelerate the GPU based on resistive memories and an online learning framework [10], [11]. An alternative approach in approximate memory design also builds on dynamic management of the bit cell allocation for the energy-quality trade-off. Additionally, when considering the circuit level, intriguing insights are gained with respect to the hardware implementations and the level of attained savings [12], [13]. In this regard, mainly two broad concepts

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

are applied. The first includes the scaling of the driving voltage, thereby reducing the overall energy consumption [14], [15]. The second approach builds on reducing the number of transistor elements and redesigning the circuit blocks accordingly [16].

These circuit design techniques are confined to a certain limit for the transistor sizing, number of elements, and voltage levels to sustain a deterministic operation of the underlying components. Nonetheless, when dealing with 32nm and smaller technology, the attributes of Moore's law, which have been steering the needs and requirements of chip and hardware designs [17], [18], are not applicable due to the physical limits affecting the scaling of the transistor devices [19]. The nanoscale dimensions of the transistors put forward the sub-threshold leakage current and the atomic interactions, thereby abating the reliability and usability of these components [20]. As a result, aside from the static variation that originates from the manufacturing process, the dynamic variability, with operational variations over time, is becoming an increasing concern and a crucial issue for the corresponding circuit operation. The internal mechanisms are affected by stochastic ionic effects involving state variables and leading to output variations [21], [22].

The variability of the transistor devices has been traditionally regarded as a source of concern for hardware designs where correcting schemes needed to be applied to ensure accuracy [19], [23]. However, building on this intrinsic variability, and dealing with it as a source of performance shaping instead of as an impediment to correct operation, is the basis of the analysis presented in this paper. Furthermore, this paper explores the right design space and improved energy efficiency, in the presence of the variability of transistors.

The variable device characteristics are modeled within a SPICE environment, offering an easy form of temporal variability. Moreover, the stochastic transistor is used as the core building block for approximate arithmetic applications. This operational mode is highly useful for error-tolerant applications that constitute a major part of the internet of things (IoT) operations. In particular, this mode targets the devices in which real physical signals are involved, such as wireless sensor nodes. Moreover, the integration of the transistor variability into the circuit-level simulator allows for the emulation in which the circuits are subject to extensive scaling endeavors. All in all, this study provides the design framework and the voltage scaling limits for the energy-quality scalability of resilient processing circuits. Thus, the proposed technique offers an alternative approximation method for the approximate computing circuits that build on unreliable elements. Thereby, four major contributions are presented in this paper:

1) A dynamic time-dependent transistor model using thermal noise to characterize stochastic behavior within the transistor, this scheme can be easily adopted into larger circuit simulations along with extensions to various statistical distributions;

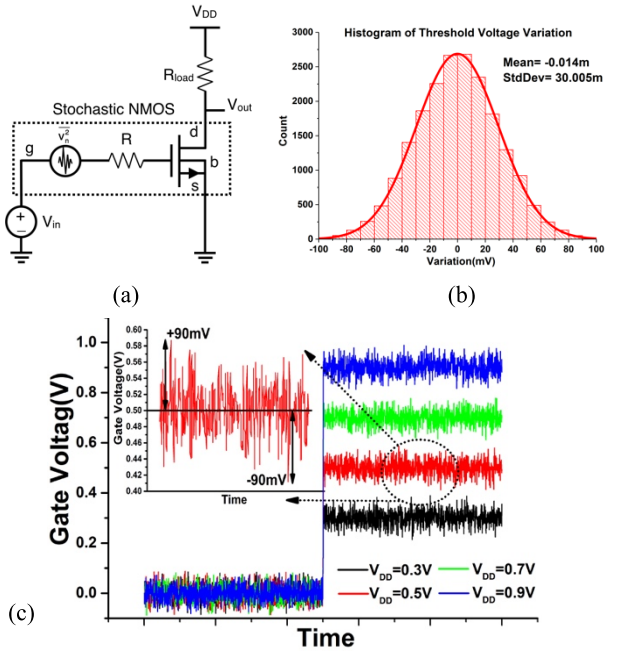


FIGURE 1. (a) Stochastic transistor model with added variability to the gate voltage. (b) The PDF of the threshold voltage variation for a Gaussian distribution and $\sigma = 30$mV matched with [34]. (c) The transistor gate voltage with the incorporated variability at V_{GS}. The inset is a zoom into the output at 0.5V showing the added variation of maximum ±90mV to induce the threshold variability.

2) An approximate computing circuit design based on stochastic components demonstrated through the analysis of N-bit adders;

3) The efficiency of the proposed solution is addressed with an analysis of the transistor size, operational frequency, energy and delay, process corner and temperature; and

4) An image compression application with approximate arithmetic blocks for energy-efficient operation with full SPICE demonstration of the application.

The rest of the paper is organized as follows. Section II discusses the physical attributions of the inherent time-dependent variability of the transistor, along with the modeling principles used for the circuit simulations. Next, Section III elaborates the implications for the approximate computing realm of study and explores the effect on the logic operators. Section IV further investigates arithmetic blocks with varying numbers of bits using simulations and explores possible trade-offs involving accuracy, energy saving, technology node, delay, process corner, and temperature. Using full SPICE verification, section V demonstrates image compression with approximate adders and discusses the effects on performance and analysis. Finally, the conclusion comprises the summary and remarks on the presented principle.

II. STOCHASTIC TRANSISTOR MODEL

The miniaturization of transistor sizes will force the underlying physical characteristics to have a more prominent effect on the output behavior. Static and dynamic forms of

variability will suffice and affect the reliability and degradation of the corresponding devices [19].

With random dopant fluctuation (RDF) playing a crucial role in the static variations, dynamic variability is mainly dominated by the quantum-level effects imposed by temperature and voltage-operating conditions [24]. The direct impact of this temporal variability is reflected in the stability of the transistor's threshold voltage (V_{th}). This section discusses the details regarding the physical origins of stochasticity; the modeling proposed to incorporate threshold variability into the simulation environment, and the different modes of stochasticity.

A. THRESHOLD VARIABILITY

The cumulative and singular effects of Bias Temperature Instability (BTI) [25], [26], Hot Carrier Injection (HCI) [27] and Random Telegraph Signal (RTS) [28] are a form of dynamic stochasticity that is translated into temporal variations in the threshold voltage of the transistor [29]. With a certain gate voltage, the switching event of the transistor is considered probabilistic. The preset bounds of the transistor's regions of operation are no longer tight nor deterministic; instead, they vary temporally depending on the threshold voltage value at each instant in time [30].

To assess the effect of these variations on the performance and energy efficiency of the circuits, a stochastic transistor model that includes its stochastic behavior needs to be established. Ideally, the physical equations governing the transistor operation need to be altered to include the effects of the noise and the non-idealities present in the device. A technique is needed that closely captures the experimental behavior but adds to the complexity of the model and the simulation process. In [31], Monte Carlo, 3D scaling, and transient noise simulations are used to conduct simulations within the SPICE environment. However, these techniques require a large amount of simulations and computation resources to capture the dynamic effects of variability. This paper introduces a model that incorporates the physical variability in a statistical manner, mainly by adding a noise source at the gate input to have the device behave probabilistically [32], [33]. The oversampling scheme allows the simulator to capture most of the variations during a single period and provides enough data points for the analysis. Hence, the paper presents three major concepts for the induction of the variation into the transistor models and for their easy integration within the circuit simulation platforms:

1) The physical variation effects can be summed up into the temporal threshold voltage variability;

2) The modeling of the variation is implemented by adding a thermal noise to the gate voltage; and

3) The added temporal variability ensures enough data points within a single transient simulation run.

Figure 1a shows the proposed stochastic transistor model with the added noise source at the gate input. The thermal noise is characterized by the standard deviation, and is calculated as follows:

$$\bar{v}_n^2 = 4k_B TRB \tag{1}$$

where k_B is the Boltzmann constant of 1.38×10^{-23} J/K, T is the temperature in Kelvin, R and B are the resistor value and the bandwidth of operation, to be specified at runtime to regulate the standard deviation. The stochastic transistor model is implemented by the SPICE netlist in Cadence Spectre simulator. Figure 1b displays the histogram of the variation based on Cadence Spectre simulations, indicating the model accurately captures the intended variation with the mean value of 0 and the standard deviation of 30mV, matching with the experimental results reported in [34].

B. STOCHASTIC DISTRIBUTIONS

The natural phenomenon affecting the behavior of the transistor elements can be fit into different distributions. Several studies have addressed this dynamic variability and provided models for the underlying noise and instability effects. A separate analysis was conducted for each contributing component, building on extracted experimental measurements. The corresponding atomic/ionic scale stochastic processes are mainly fit to the Lognormal distribution, as is the case for Negative Bias Temperature Instability (NBTI) [26]. Alternatively, with abrupt variations imposed by random telegraph signals, Uniform, Exponential, and Poisson distributions [35] are more generally used approximations. When taking into consideration all of the affecting factors on the variation of the threshold voltage, a general multivariate model was proposed in [34]. Moreover, an approximation was provided for the threshold voltage to follow a Normal distribution with a standard deviation (σ) holding an average value of 30mV and, according to the experimental number reported in [34], reaching up to 50mV in some extreme cases. Figure 1c shows the variation of the gate voltage V_{GS} over time with a max/min of $V_{DD} \pm 90$mv achieved by tuning the noisy resistance value and the noise frequency. The simulation results of the model show a well alignment with the measurement results in [34]. The Gaussian distribution of the thermal noise could easily be transformed into other various distributions using a probability transformation function. The SPICE code for the general transformation is as follows.

III. APPROXIMATE COMPUTING

Conventionally, having the transistors behave in a non- deterministic manner is considered a drawback to the circuit performance. Costly error-correcting schemes are applied to counter the induced deviations from the original operation. On the other hand, increasing the supply voltage is also considered, to a certain extent, to overcome the noise effects and ensure correct output. However, with aggressive scaling, variability is now an inevitable feature that needs to be addressed in unconventional concepts [36]. Approximate computing benefits from the probabilistic behavior of the underlying circuitry to shape the performance in error-tolerant applications. Complete accuracy of the desired output is not considered

Algorithm Stochasticity Transformation to Alternative Probability Distributions

.Param Ka = 1.38e-23, T=300, Rg =resistance value, Bg =bandwidth, pi= 3.14, lambda= 2
.Param var = 4∗K8∗T∗Rg∗Bg
∗∗∗∗*Noisy Resistor*
Rn n 0 Rg noiseon=yes
∗∗∗∗*Normalizing the Gaussian distribution values*
Enorm 10 value = 'V(n)/sqrt(var)'
∗∗∗∗*Transforming to lognormal Distribution*
Elognormal 2 0 value = 'exp(V(n))'
∗∗∗∗*Transforming to Uniform Distribution*
Euniform 3 0 value= '1-sqrt(l-exp(-(2∗V(l)"2/pi)'
∗ ∗ ∗∗*Transforming to Exponential Distribution*
Eexponentia14 0 value = '-(1/lambda)∗ln(l-V(3))'

a priority, but instead a complementary feature, depending on the available resources [14], [37]. This section proposes an alternative approach to approximate computing based on the stochastic operation of the underlying transistors. The conventional design structure of the logical operators is kept intact, while the transistors act in a non-deterministic manner. To verify the operation principles and the gains achievable with these devices, a 20nm predictive technology model (PTM) is used for circuit-level simulations [38] with Cadence Spectre at an operation frequency of 500MHz and a nominal voltage of 0.9V. The impact on Boolean operators is discussed with emphasis on the inverter and the different logic gates where each circuit output is kept as fan-out of 4 to emulate its behavior in full circuits.

A. STOCHASTIC INVERTER

The inverter is composed of two vertically concatenated transistors, PMOS and NMOS. When applying nominal voltage V_{DD} on gate input, the expected output is the inverse of the input. However, with the intrinsic stochasticity of the transistors, especially with the variation of the threshold voltage, some glitches can be observed at the output. The effect of this variability will be apparent at lower voltages, particularly at levels closer to the original threshold voltage of the device. This hypothesis is tested with the simulation of the stochastic inverter at different V_{DD} levels. The simulations cover the range from 0.2V up to the nominal value of 0.9V. Fine step sizes for simulation are taken up to 0.4V, as the variability has a considerable impact on this level, whereas it is considered to be minor at higher voltages. Figure 2 shows the output of the inverter for different input voltages with respect to enough amount of time to capture the full variation spectrum. Aside from the sub-threshold leakage, the effect of the dynamic variability is distinct mostly at lower input voltages, with deep crossing into the opposite regions for the high and low bits respectively.

The accuracy of the logic operation is specified as the number of correct samples for a digital bit divided by the

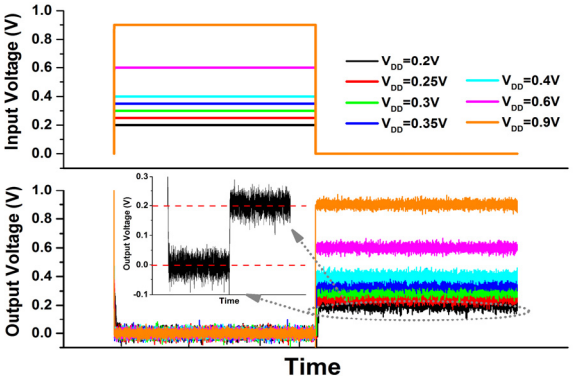


FIGURE 2. The input and output signals of the stochastic inverter at different input voltage levels.

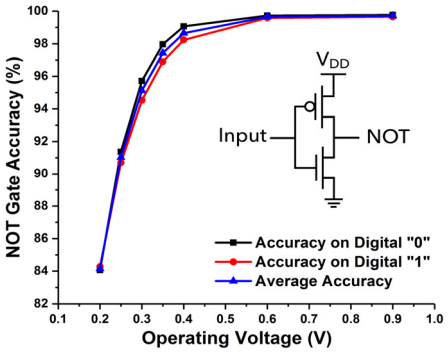


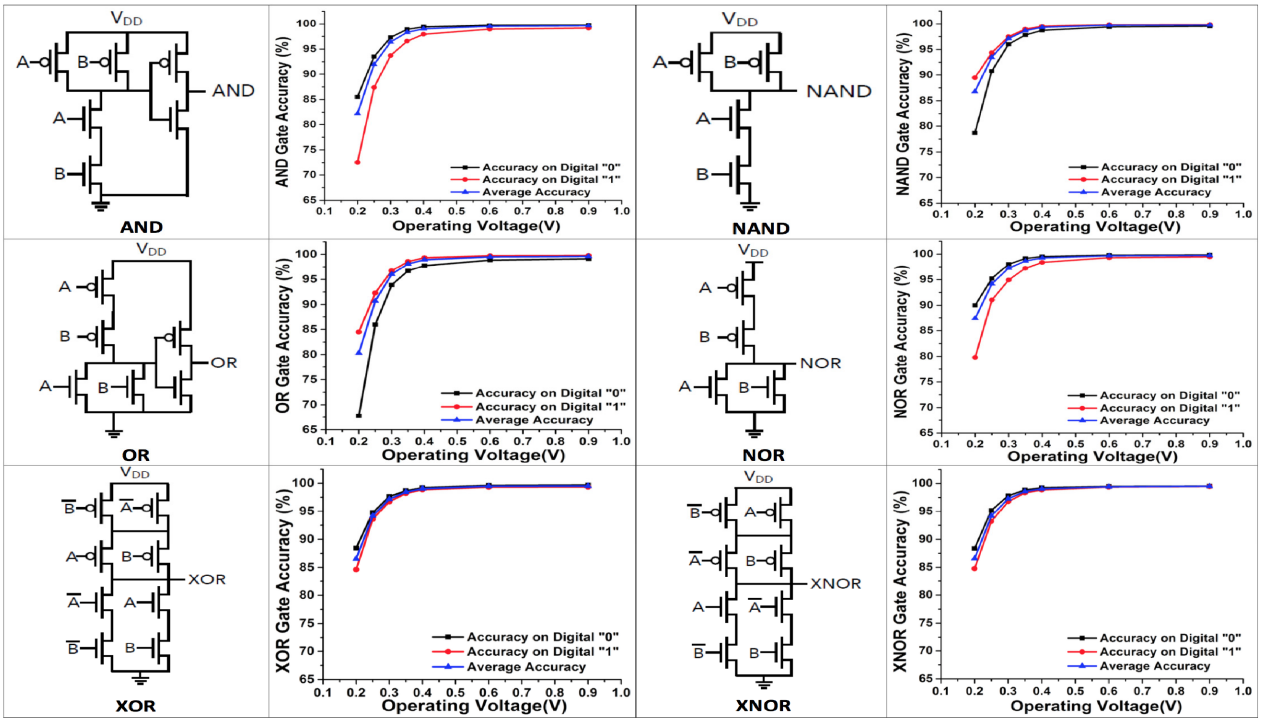
FIGURE 3. The output accuracy of the inverter in terms of getting an accurate '0' and '1' along with the average performance.

total number of samples (Accuracy = $(N0_{Correct} + N1_{Correct})$ / N_{Total}), where N0 and N1 correspond to the number of samples of correct '0' and '1' respectively. To assess the accuracy of this operation, 16,000 samples are considered for each period. An output value is considered to be correct only if it is within 10% of the ideal output value to keep with the typical noise margin convention. Namely, $0 \sim 0.1 \times V_{DD}$ will be considered as correct digital '0' if the ideal value should be 0V, and $1 \sim 0.9 \times V_{DD}$ will be counted as correct digital '1' if the output is supposed to be V_{DD}. A conservative approach is adopted for the setting of the undetermined values 'x', guarantee the worst-case scenario by assigning the "x" value to be the counterpart of the correct value, namely if the correct value is "1", then if a value falls into "x" region, it would be evaluated as "0" to create worst-case interpretation.

$$Digital\ Value = \begin{cases} 0, & if\ V_{out} < 0.1 \times V_{DD} \\ 1, & if\ V_{out} > 0.9 \times V_{DD} \\ x, & if\ 0.1 \times V_{DD} < V_{out} < 0.9 \times V_{DD} \end{cases} \quad (2)$$

The sampling focus solely on the stabilized part of each period while excluding the rising and falling transition time, Figure 3 shows the accuracy of the inverter over different levels of V_{DD}. The accuracy of both bits '0' and '1' are

TABLE 1. The variability effect of the underlying transistor elements on the logic operators. Output accuracy of the bits '0' and '1' and the average are shown with respect to different levels of input voltage. With higher V_{DD}, the variability will have a smaller impact and the gate will behave in a deterministic manner.



considered along with the average overall accuracy of the inverter or NOT gate. The average is set by taking the mean of the accuracy values of the bits '0' and '1' accordingly. Interestingly, the accuracy of the inverter reaches a minimum of 84% at an operating voltage around the original threshold of the transistor elements. Hence, a large space for energy saving is possible while maintaining high levels of accuracy as well.

B. LOGIC GATES

Extending the stochastic operation into the logic domain, the stochastic transistor devices are used to build logic operators. In this mode, the effect of added cascades is studied with respect to the expected output accuracy in a similar fashion to the inverter. Table I depicts the circuit structures for the main Boolean operators. Moreover, the accuracy of the output behavior is analyzed in terms of the percentage of having a correct '0', correct '1', and overall expected correct output for all the different combinations of the truth table.

In general, the output accuracy of the gates shows similar performance to that of the inverter. However, a slight degradation is encountered, where the minimum average accuracy reached is around 80%. Hence, the added number of cascaded blocks has only a minor effect on the output performance. Alternatively, the logic operation does show an impact on the output accuracy. For instance, the AND gate shows the lowest accuracy for the bit '1'. This could be because the gate is biased to have more zeros as it is clear from the truth table entries for this operation. Similarly, the OR operator shows the lowest performance accuracy for the bit '0', as it only appears once in its truth table. Thus, this interesting feature needs to be further investigated with larger arithmetic blocks, to determine whether the degradation is propagated or suppressed by the particular structure.

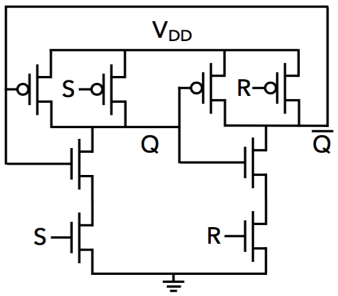


FIGURE 4. The structure of the latch circuit [39] composed of stochastic transistor elements.

C. STORAGE ELEMENTS

The transistor is the constituting component in storage elements such as SR latch [39], a level sensitive positive latch. Hence, the stochasticity inherent within the transistor operation and the scaling of the input voltage both have a large impact on these non-static structures. The primary impact of the variation is reflected on the failure rate and the delay for the correct data within the storage cells. Figure 4 presents the internal structure of the latch and the

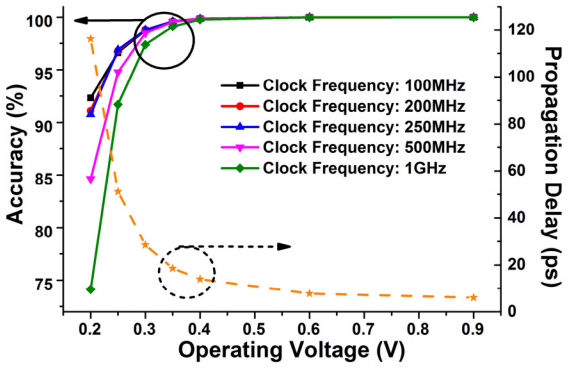


FIGURE 5. The impact of the transistor variability on the operation of the latch. Measures of delay and accuracy under different operating frequencies are highlighted.

corresponding interconnections; Figure 5 depicts the simulation of the latch under different operating frequencies. The accuracy of the output and the propagation delay are the affected parameters with the scaling of the voltage and frequency. However, the operation of the latch could achieve almost 100% accuracy at a voltage as low as 0.4V, corresponding to almost half the nominal value, with a propagation delay of 10 ps and operating frequency of 1GHz.

Hence, the transistor variability allows for more efficient operation of the storage element, in particular where complete accuracy is not a paramount requirement, as in the case of real-world signal processing applications.

IV. APPROXIMATE ADDER

Adders are the principal building blocks of arithmetic operations, and their reliability and accuracy profoundly affect simple computations and more complex processing [40]. Several approaches for implementation of approximate adders, along with other arithmetic blocks are evaluated and classified in [16] and [41]–[43]. In particular, comparisons on the error; characterizations on the circuit; and discussions on image processing applications are reviewed across various designs in [43]. In this work, however, our primary purpose is to investigate the impact of device variations, so we have selected the standard, optimized ripple carry structure for the multi-bit adder. Furthermore, this section studies and quantifies the performance of approximate adders. Starting with a detailed analysis of a single-bit adder, the section presents the accuracy of the sum and the carry generation blocks. More reflective error quantifying metrics are used to assess the behavior of N-bit adders up to 16 bits. Error distance, mean error distance, relative error distance, and mean relative error distance are calculated for the different adders [44], [45].

A. FULL ADDER

The mirror adder circuitry is adopted for improved carry generation [39]. Most importantly, the dimensions for the carry block transistors are set to ensure a more optimized operation for the output carry bit. Moreover, the number of transistors used for generating the carry bit is much fewer than the sum generation, as depicted in Figure 6. These features make the carry bit more stable and less susceptible to errors.

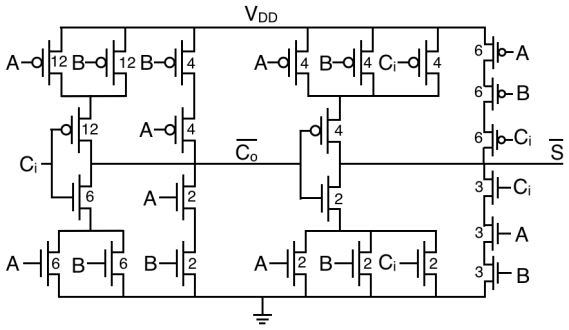


FIGURE 6. The structure of the mirror adder circuit [39] composed of stochastic transistor elements. Transistor sizes are configured to ensure optimized carry generation.

The expected sum value for the 1-bit adder reaches up to 3. Hence, any error within the generated sum or carry bit will be reflected in a substantial change in the overall output value. The performance analysis of the 1-bit adder can, therefore, be portrayed using the accuracy of these bits. Figure 7 shows the corresponding accuracy values for all input combinations for the high and low bits, respectively. As expected, the added stochasticity has a more significant impact on the sum bit than on the carry bit, due to the nature of the used structure. However, the considerably reliable operation is attained even at low voltage levels. The accuracy reaches a minimum of 70% at voltages as low as 200mV.

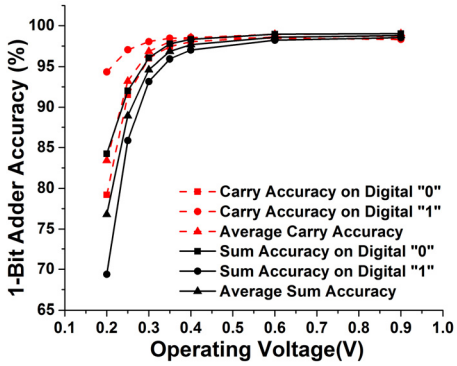


FIGURE 7. The output accuracy for the sum and carry bits of the 1-bit full adder. Better output characteristics of the carry bit is due to the nature of the used structure to optimize the carry generation.

B. N-BIT ADDERS

N-bit operation is needed to perform computations and process operations. Hence, investigating a higher number of bits for the adder offers more insights into the applicability of the approximate computing approach, particularly in the logic domain. A ripple carry adder (RCA) is used for the analysis. It is composed of cascaded blocks of 1-bit full adders with the

carry propagating between consecutive blocks. This structure is used for the analysis because it shows the effect of the probabilistic behavior of the carry-in bit to the subsequent full adder blocks, and consequently the overall output value. Figure 8 shows the block diagram for the N-bit ripple carry adder.

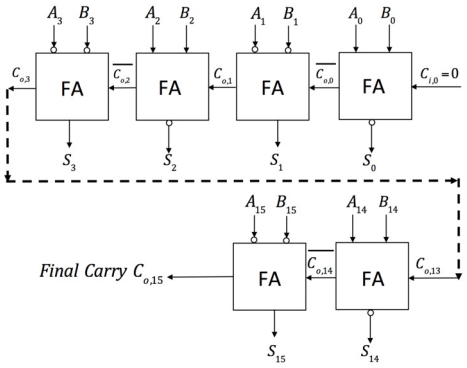


FIGURE 8. The block diagram of the N-bit adder with the propagation of the carry among the individual 1-bit adders.

To quantify the effect of the transistor variability on the adder's output behavior, the 2-bit, 4-bit, 8-bit, and 16-bit adders are simulated using Cadence Spectre. The accuracy metric used throughout this paper can efficiently reflect the accuracy of operations involving single-bit output, as it assesses whether the bit is correct or not. However, with a larger number of bits for the addition, the accuracy does not provide enough information on the effect of the probabilistic behavior. An error in any of the bits within the output sum values is considered to reduce the accuracy, wherein the binary domain, different weights are given depending on the location of the bits. The least significant bit has a lower impact on the sum than the most significant bit. This divergence further increases when a larger number of bits is used for calculation. Therefore, the metric such as the error distance (ED) is considered to be more informative [42] and is calculated as (ED(a,b) = |a-b| = $\left|\sum_i a[i] \times 2^i - \sum_j b[j] \times 2^i\right|$) where a and b are the expected and probabilistic sums, respectively [42]. The ED quantifies how far the output sum value is from the expected values. Figure 9 shows a 3D plot for the error distance with respect to the operating voltage V_{DD} and the expected sum values for a 4-bit adder. The peak ED is found at the middle addition value. This is because this particular range has the highest probability of occurring out of all the different addition values, and because of the conservative setting of the undermined values 'x'. The worst-case error is assumed in the setting process of the 'x' values that lie between the two analog values of '0' and '1'. Moreover, as expected, the maximum ED occurs for lower operating voltages and steadily decreases to 0 for operating voltages of 0.4V and higher, thus providing a highly reliable operation with at least 80% energy savings.

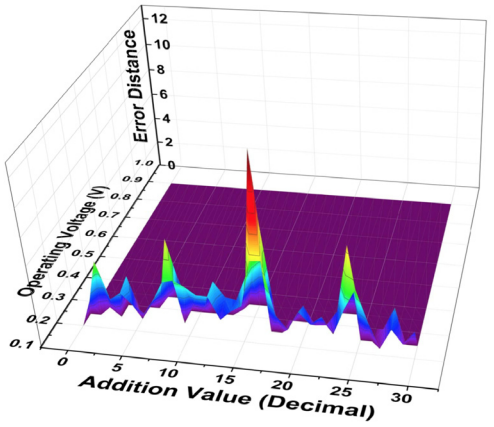


FIGURE 9. 3D plot for the Error Distance of a 4-bit adder with respect to the expected addition value and the operating voltage as well.

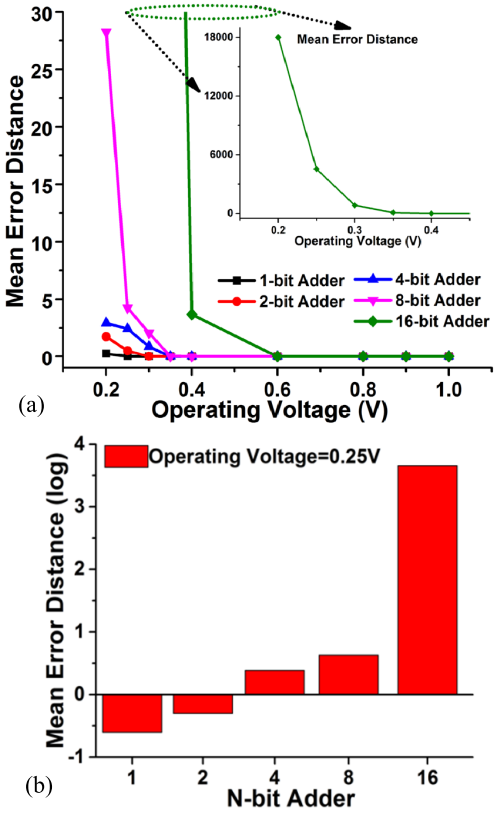


FIGURE 10. (a) The mean error distance for the adders at different operating voltages. (b) Log-scale of the MED at an operating voltage of 0.25V showing very low error for the lower bit adders.

Further quantifying metrics are calculated to provide a more elaborate view of the performance of the adders at different operating voltages. The mean error distance (MED) is calculated as (MED = $\sum_i ED_i \div M$) where all instances of the ED are summed up and divided by the total number of samples M. Figure 10a shows the simulation results for the MED over a range of operating voltages (V_{DD}). As depicted, the N-bit adders up to 16-bits have a similar performance

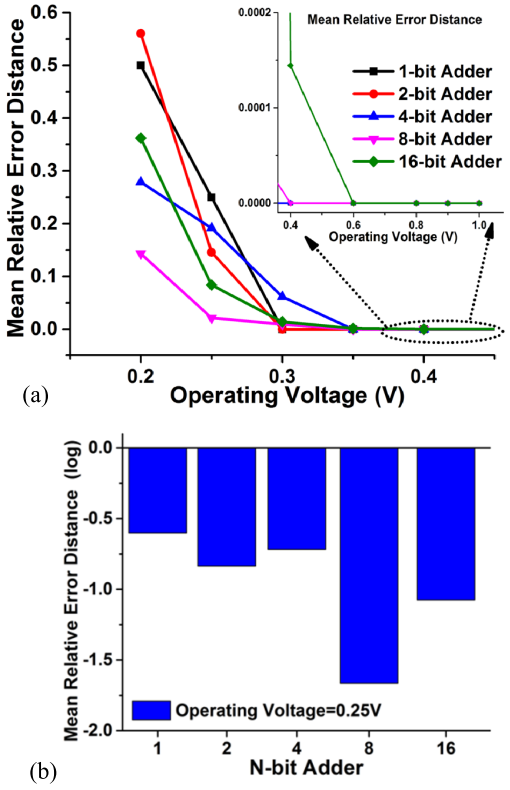


FIGURE 11. (a) The Mean Relative Error Distance for the adders at different operating voltages. (b) The log of the mean relative error distance for different N-bit adders at an operating voltage of 0.25V.

with regard to achieving almost zero error at voltages higher than 0.4V. Moreover, regarding the lower operating voltages, as shown in the inset of Figure 10a, the adders up to 8-bit have a very low MED, reaching up to 2% of the total sum. A further quantifying example is shown in Figure 10b, where the logarithmic (base10) scale is used to plot the absolute MED values for different N-bit adders at an operating voltage of 250mV.

Although the results show a high peak for the 16-bit adder, this is due to the measure that the MED portrays: it shows the absolute value of the error with no relation to the actual sum value. Therefore, the mean relative error distance (MRED) is calculated as well [42]. The relative error distance (RED) takes the actual sum value (R) into account by dividing each ED value by the corresponding expected sum ($RED = ED \div R$) and the mean relative error distance (MRED) is then calculated as ($MRED = \sum_i RED_i \div M$). Figure 11a shows the mean relative error distance for the different N-bit adders across the range of operating voltage. As depicted, the 1- and 2-bit adders show the worst characteristics at the lowest voltage of 0.2V. This is because the sum values in these adders are small and any deviation away from the expected value results in a large MRED. However, the highly close operation is shown for all of the other adders with MRED values reaching as low as 0.15 or 15% of the actual sum. A closer inspection of the error characteristics at a single operating voltage of 0.25V reflects the small errors attained for various N-bit adders in Figure 11b.

C. DISCUSSION

The efficiency of the proposed scheme under diverse process variations is addressed for the different approximate adders. The comparison is based solely on the current scheme; as to the best of the authors' knowledge, this technique serves as the first proposition of circuit approximation based on unreliable components and across the technology nodes. The evaluation is conducted with respect to the operating frequency, the technology node or transistor size, the energy and delay, the process corner, and temperature.

1) OPERATING FREQUENCY

The accuracy of the adder output is directly affected by the operational frequency. The higher the frequency, the larger the chances are of obtaining erroneous results. Hence, the adders are simulated under different frequencies, and the MED and MRED are respectively measured. Figure 12 shows the performance of a 4-bit adder, based on a 20nm PTM model, at different frequencies of operation. As depicted, to achieve full accuracy, the operating frequency should not exceed 1 GHz in case the operating voltage is scaled down to 0.4V, which corresponds to less than half the nominal value. Larger operating frequencies are still feasible but at smaller scaling levels. That is, with around 30% scaling of the nominal voltage frequency, up to 2 GHz is attainable with completely accurate results. Hence, as the IoT operation dictates, several operating points could be available for the attainment of the performance metric required under the resource constraints.

2) TECHNOLOGY NODE

The impact of the technology node is taken into consideration with the simulation of the N-bit adders performed for predictive technology models of 10nm and 20nm, and for the actual device model of TSMC 65nm. Table II provides an overall comparison between the different N-bit adders in terms of the achieved MED and MRED for the scaling of the nominal voltage for each technology node. The voltage levels are shown to reflect the same percentage of scaling for each technology node considering the differences in the absolute nominal voltages. Moreover, as the log-scale is used to represent the values, the dashes represent an achieved error of 0. As depicted, with smaller transistor sizes, higher levels of scaling of the nominal voltage are feasible, reaching down to around 0.4V while maintaining the full accuracy of the output results. However, once the technology size increases, the variability starts to degrade the performance and limit the scaling possibility. The minimum operating voltage applied to start achieving full accuracy is set to 0.8V for the 65nm technology node. Whereas configurable accuracy and energy savings are possible by choosing the level of operating voltage. A compromise arises, with lower voltage levels achieving energy savings of up to 95% but having error levels

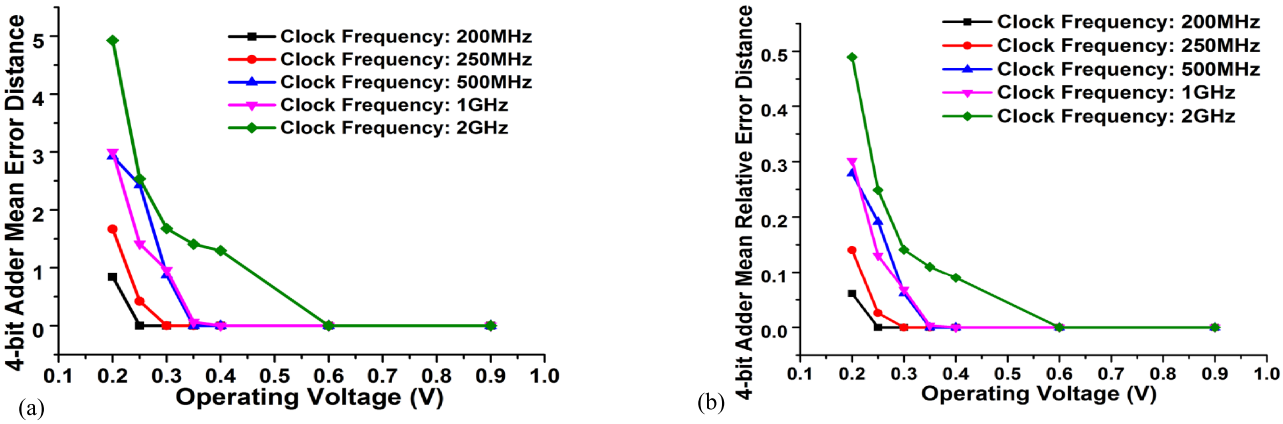


FIGURE 12. (a) The mean error distance for the 4-bit adder at different frequencies. (b) The mean relative error distance of the 4-bit adder at different frequencies.

TABLE 2. The overall comparison for the performance of the N-bit adders in terms of the MED and MRED for different technology nodes.

10mn									20nm									65nm								
	MED(log)				MRED(log)					MED(log)				MRED(log)					MED(log)				MRED(log)			
V_{in}	2b	4b	8b	16b	2b	4b	8b	16b	V_{in}	2b	4b	8b	16b	2b	4b	8b	16b	V_{in}	2b	4b	8b	16b	2b	4b	8b	16b
0.17	0.12	0.43	1.61	4.44	-0.42	-0.53	-0.71	-0.26	**0.20**	0.24	0.47	1.45	4.26	-0.25	-0.55	-0.84	-0.44	**0.27**	0.18	0.94	1.98	4.52	-0.38	-0.24	-0.36	-0.12
0.20	-0.60	0.10	0.85	3.93	-1.18	-0.89	-1.42	-0.79	**0.25**	-0.30	0.39	0.63	3.66	-0.84	-0.72	-1.67	-1.08	**0.34**	0.24	0.80	2.10	4.48	-0.26	-0.26	-0.23	-0.15
0.25	-	-1.51	0.43	3.12	-	-2.68	-1.88	-1.60	**0.30**	-	-0.06	0.32	2.91	-	-1.21	-2.02	-1.85	**0.40**	0.27	0.82	2.09	4.43	-0.23	-0.19	-0.26	-0.18
0.29	-	-	-2.19	1.71	-	-	-4.60	-3.00	**0.35**	-	-	-2.19	1.97	-	-	-4.60	-2.78	**0.48**	0.19	0.84	1.89	4.45	-0.18	-0.19	-0.45	-0.16
0.34	-	-	-	-	-	-	-	-	**0.40**	-	-	-	0.57	-	-	-	-3.84	**0.53**	0.18	0.60	1.71	4.52	-0.28	-0.39	-0.64	-0.13
0.50	-	-	-	-	-	-	-	-	**0.60**	-	-	-	-	-	-	-	-	**0.80**	-	-0.04	0.65	3.59	-	-1.19	-1.65	-1.14
0.75	-	-	-	-	-	-	-	-	**0.90**	-	-	-	-	-	-	-	-	**1.20**	-	-	-	0.75	-	-	-	-3.85

reaching around 30%. On the other hand, larger operating voltages also offer an almost accurate level of outputs, with savings reaching more than 60% at the 16-bit output at the technology node of 20nm.

3) ENERGY AND DELAY

The energy and delay parameters are important but contradictory design metrics. The lower the operating voltage, the lower the energy consumption, but the larger the propagation delay. However, with the underlying stochasticity of the transistor, several operating points are feasible depending on the technology node and the available resources. For instance, scaling the voltage down by 30% achieves energy savings of around 50% and delay in the order of fewer than 50 ps for technology nodes of 10nm and 20nm, and fewer than 150 ps for 65nm. Figure 13 shows the simulation results for a 4-bit approximate adder operating at 500MHz. The intersection points between energy consumption and the delay plots represent the optimal points of operation for the different technology nodes. As depicted, the technology nodes of 10nm and 20nm have their optimum operating point at 45% of the nominal value, which provides an accurate output result,

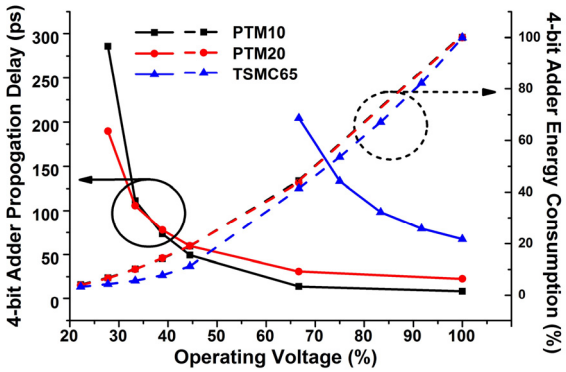


FIGURE 13. The analysis of a 4-bit adder with respect to the delay and energy consumption. The percentage scaling of the input voltage results in the scaling of the delay and the energy consumption as well.

as shown in the performance table. Similarly, the optimum operating point for the 65nm resides at 75% of the nominal value, which also provides accurate output values.

4) PROCESS CORNER

The simulation of the process corners serves as a measure of how the process and the environmental stimuli affect the

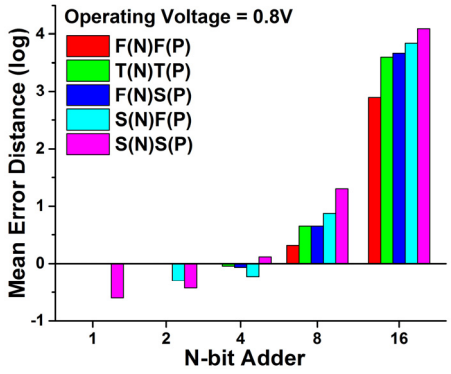


FIGURE 14. The performance, in terms of the mean error distance, of the approximate adder with different number of bits at an operating voltage of 0.8V, corresponding to 70% of the nominal voltage for 65nm technology.

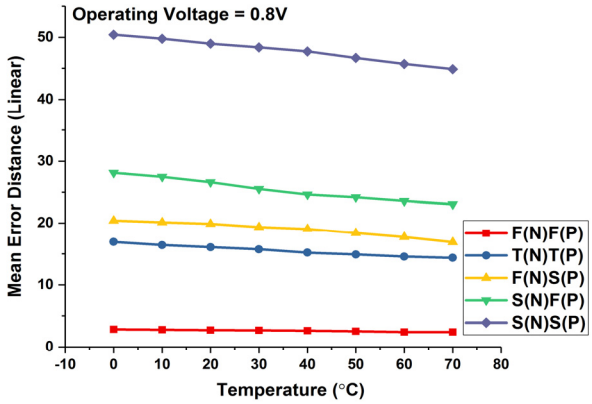


FIGURE 15. The mean error distance with temperature variations of the 8-bit approximate adder at 0.8V operating voltage and different process corners.

circuit in an extreme situation [39]. The impact on the performance of the adder is measured in terms of the MED. In that regard, Figure 14 shows the behavior of N-bit approximate adders operating at 0.8V, which corresponds to 70% of the nominal voltage of the 65nm technology node. An interesting feature that is apparent in these results is the impact of the NMOS. The Fast-Slow (FS) and Slow-Fast (SF) do not show similar behaviors. The slow operation of the NMOS degrades the performance and leads to more errors in the operations. Moreover, as depicted in the figure, a completely accurate performance is achieved for the Fast-Fast (FF) case up to 4 bits, with less than 1% error rate. Another interesting observation is that the simulation results of the Slow-Slow (SS) corner show a relatively low error rate that is barely noticeable when dealing with error-tolerant applications.

5) TEMPERATURE VARIATIONS

The temperature variations have a strong influence on the transistor parameters, which correspondingly affects the circuit behavior. Figure 15 shows the impact of the temperature on the 8-bit approximate adder for different process corners. The operating voltage is kept at 0.8V, and the MED is measured and plotted. Since the threshold voltage decreases with the temperature [46], the overdrive voltage and the on-state current increase. This leads to a decreasing MED and higher accuracy with temperature.

V. IMAGE COMPRESSION

With an elaborate investigation of the approximate adder with stochastic components, characterization of performance could be visually assessed through digital signal processing applications. In this context, this paper presents image compression using 2-point Discrete Fourier Transform (DFT). The calculation requires addition and subtraction of pixel values of the image to be compressed. With x[i] and x[i+1] being two consecutive pixels, the DFT output is calculated using the butterfly operations.

$$\begin{cases} y[i] = x[i] + x[i+1] \\ y[i+1] = x[i] - x[i+1 \end{cases} \tag{3}$$

The output values y[i] and y[i + 1] are a direct function of the input values. The subtraction is transformed into addition by using the 2's complement of the corresponding subtrahend.

A. SIMULATION SETUP

A 200×200 pixel JPEG image is used for the compression application. Each pixel is transformed into an 8-bit binary number, and the image is then mapped into two sets of 8-bit input voltage samples that are fed in parallel to Cadence Spectre. Circuit-level simulations for the corresponding additions are performed using the 8-bit RCA with mirror full adders, implemented with stochastic transistors in 20nm technology with Gaussian distribution variability. The simulations are all performed at the 500MHz frequency with a pre-layout analysis framework. The DFT additions and subtractions are performed for several operating voltages to assess the output characteristics and energy savings for each voltage. The circuit simulations cover the nominal value of 0.9V to account for the original or deterministic addition operation. The values of the addition operations for different operating voltages, ranging from 0.2V to 0.9V, are then extracted from the circuit simulator and fed into MATLAB to reconstruct the images. This reconstruction is achieved using an error-free 2-point Inverse Discrete Fourier Transform (IDFT).

B. MULTIPART FIGURES

The quality of the compression is characterized using the peak-signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [47].

1) PEAK-SIGNAL-TO-NOISE-RATIO

PSNR stands for the peak value of the pixel; with 8 bits used to represent the pixel, the peak value is considered to be 255. MSE is the mean square error between the reference value of the pixel in the original image and the value of the pixel after

reconstruction. Hence, PSNR is calculated as follows

$$PSNR = 10\log_{10}\frac{PV^2}{MSE} \tag{4}$$

Figure 16 shows the original image and two reconstructed versions of the image after the application of the DFT using approximate 8-bit adders. As can be seen, scaling the voltage down to 0.3V, which results in more than 90% energy savings, and allows for compression with a minor effect on the quality with a PSNR value of 24.17. On the other hand, more aggressive voltage scaling adversely affects the quality of the compressed images. Hence a compromise lies between the required level of quality and the energy savings based on the application requirements and available design resources.

FIGURE 16. The output image with different level of arithmetic error in the DFT operation. The Quality of the output is relative to peak signal to noise ratio and the perception of the image. Compromises arise between the energy savings and the level of quality sought for.

2) STRUCTURAL SIMILARITY INDEX

SSIM is an assessment of the perceived image quality based on the quantification of the visibility of errors [47]. This measure builds on the adaptability of the human visual system in extracting structural information. Considering two image signals x and y from the original and the reconstructed image respectively, the SSIM (S(x,y)) comprises three components

$$\begin{cases} S(x,y) = f(l(x,y), c(x,y), s(x,y)) \\ S(x,y) = \dfrac{(2\mu_x\mu_x + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \end{cases} \tag{5}$$

where l(x,y) is the luminance comparison, c(x,y) is the contrast comparison, s(x,y) is the structural comparison, μ and σ are the mean and the standard deviation, and C1 and C2 are added constants to avoid instability. The SSIM is measured for the reconstructed images over different operating voltages. The higher the value of the SSIM index, the closer the reconstructed image is to the original one until reaching a value of 1 that corresponds to complete similarity. As depicted in Figure 17 comprising the minimum, mean, and maximum values of the index within the image, increasing similarity is achieved with larger input voltage where values larger than 0.3V provide identical structures and fully accurate reconstruction. The accuracy of the approximation can be defined during runtime with the input and the supply voltage level as depicted before.

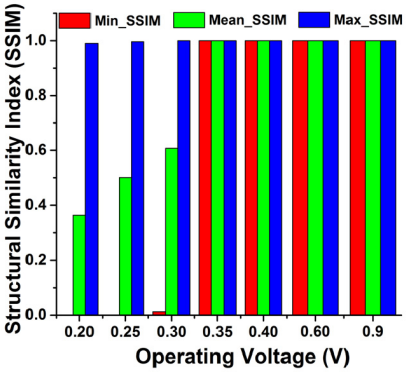


FIGURE 17. The structural similarity index for the reconstructed images at different operating voltages.

C. DISCUSSION

When compared with the alternative deterministic devices that operated at the nominal value, more than 90% on energy saving was achieved while maintaining high PSNR values of the compressed image. Voltage over-scaling schemes investigating timing path in the sequential circuits, bit error rate in different operating conditions, and error in various mathematical functions are discussed in [48]–[50] respectively. In this work, we used voltage over-scaling schemes to examine the accuracy and output characteristics of the logic and arithmetic blocks by incorporating the inherent variability of the transistors for performance shaping, in particular, energy saving in the image compression application. Multimedia and digital signal processing applications that build on these configurable arithmetic units have shown improvements in the output characteristics along with the utilization of the resources [51]. The energy-quality scalability is considered to be a control knob for the level of operation required for error-resilient applications, such as wireless sensor nodes that need to capture images, compress them, and send or even stream them in the most efficient manner to the source [52], [53]. A trade-off is apparent within the different design metrics, but it can, in fact, be sufficient and satisfactory regarding the current system requirements.

VI. CONCLUSION

Error-resilient applications offer relaxation of the mapping of design specifications regarding the corresponding hardware implementation. In this study, the variability of the nanoscale transistor devices was embraced and modeled in a statistical manner. Thermal noise was used to induce variations into the transistor elements, which allowed for the stochastic setting of the threshold voltage. Adopting this inherent stochasticity in the approximate computing concept showed the attainable benefits of doing so in terms of performance metrics savings. Analysis and simulations on simple and large arithmetic computing blocks maintained a high level of accuracy while offering savings on energy. A case study of an image compression reflected the benefit of adopting approximate adders on the application level. All in all, this

approach to transistor stochasticity provides the right design space and improved energy efficiency, in the presence of the variability of transistors. It allows for the development of configurable schemes that are adaptively controlled based on the communication channel and environment, to increase or decrease the corresponding accuracy,

ACKNOWLEDGMENT

Ren Li and Rawan Naous contributed equally to this work.

REFERENCES

[1] *A Nanotechnology-Inspired Grand Challenge for Future Computing*. Accessed: Oct. 2016. [Online]. Available: http://www.nano.gov/futurecomputing

[2] J. Liang, L. Chen, J. Han, and F. Lombardi, "Design and evaluation of multiple valued logic gates using pseudo N-type carbon nanotube FETs," *IEEE Trans. Nanotechnol.*, vol. 13, no. 4, pp. 695–708, Jul. 2014.

[3] M. Sharad, D. Fan, K. Aitken, and K. Roy, "Energy-efficient non-Boolean computing with spin neurons and resistive memory," *IEEE Trans. Nanotechnol.*, vol. 13, no. 1, pp. 23–34, Jan. 2014.

[4] A. Alaghi and J. P. Hayes, "Survey of stochastic computing," *ACM Trans. Embedded Comput. Syst.*, vol. 12, no. 2s, pp. 1–19, 2013.

[5] M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. N. Salama, "Memristors empower spiking neurons with stochasticity," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 5, no. 2, pp. 242–253, Jun. 2015.

[6] M. De Kruijf, S. Nomura, and K. Sankaralingam, "Relax: An architectural framework for software recovery of hardware faults," *ACM SIGARCH Comput. Archit. News*, vol. 38, no. 3, pp. 497–508, 2010.

[7] D. Ernst *et al.*, "Razor: Circuit-level correction of timing errors for low-power operation," *IEEE Micro*, vol. 24, no. 6, pp. 10–20, Nov./Dec. 2004.

[8] S. T. Chakradhar and A. Raghunathan, "Best-effort computing: Re-thinking parallel software and hardware," in *Proc. Design Automat. Conf.*, Jun. 2010, pp. 865–870.

[9] V. K. Chippa, S. Venkataramani, K. Roy, and A. Raghunathan, "StoRM: A stochastic recognition and mining processor," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design*, La Jolla, CA, USA, Aug. 2014, pp. 39–44.

[10] M. Imani, D. Peroni, A. Rahimi, and T. Rosing, "Resistive CAM acceleration for tunable approximate computing," *IEEE Trans. Emerg. Topics Comput.*, p. 1, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7792225/citations?tabFilter=papers#citations

[11] M. Imani, A. Rahimi, and T. S. Rosing, "Resistive configurable associative memory for approximate computing," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2016, pp. 1327–1332.

[12] J. Han, E. R. Boykin, H. Chen, J. Liang, and J. A. B. Fortes, "On the reliability of computational structures using majority logic," *IEEE Trans. Nanotechnol.*, vol. 10, no. 5, pp. 1099–1112, Sep. 2011.

[13] R. Ashraf, M. Chrzanowska-Jeske, and S. G. Narendra, "Functional yield estimation of carbon nanotube-based logic gates in the presence of defects," *IEEE Trans. Nanotechnol.*, vol. 9, no. 6, pp. 687–700, Nov. 2010.

[14] J. Han and M. Orshansky, "Approximate computing: An emerging paradigm for energy-efficient design," in *Proc. 18th IEEE Eur. Test Symp. (ETS)*, May 2013, pp. 1–6.

[15] A. Pirbadian, M. S. Khairy, A. M. Eltawil, and F. J. Kurdahi, "State dependent statistical timing model for voltage scaled circuits," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 1432–1435.

[16] Z. Yang, J. Han, and F. Lombardi, "Transmission gate-based approximate adders for inexact computing," in *Proc. IEEE/ACM Int. Symp. Nanosc. Archit. (NANOARCH)*, Jul. 2015, pp. 145–150.

[17] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, Mar. 2001.

[18] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-threshold computing: Reclaiming Moore's law through energy efficient integrated circuits," *Proc. IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.

[19] S. Borkar, "Designing reliable systems from unreliable components: The challenges of transistor variability and degradation," *IEEE Micro*, vol. 25, no. 6, pp. 10–16, Nov./Dec. 2005.

[20] R. Perricone, X. S. Hu, J. Nahas, and M. Niemier, "Can beyond-CMOS devices illuminate dark silicon?" in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2016, pp. 13–18.

[21] Y. V. Pershin and M. Di Ventra, "Memory effects in complex materials and nanoscale systems," *Adv. Phys.*, vol. 60, no. 2, pp. 145–227, 2011.

[22] S. H. Jo, K.-H. Kim, and W. Lu, "Programmable resistance switching in nanoscale two-terminal devices," *Nano Lett.*, vol. 9, no. 1, pp. 496–500, 2009.

[23] S. Das *et al.*, "RazorII: *In situ* error detection and correction for PVT and SER tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.

[24] K. Bernstein *et al.*, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. Res. Develop.*, vol. 50, no. 4.5, pp. 433–449, Jul. 2006.

[25] M. Alam, "Reliability- and process-variation aware design of integrated circuits," *Microelectron. Rel.*, vol. 48, nos. 8–9, pp. 1114–1122, 2008.

[26] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *IEEE Electron Device Lett.*, vol. 26, no. 8, pp. 560–562, Aug. 2005.

[27] A. Asenov *et al.*, "Modeling and simulation of transistor and circuit variability and reliability," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2010, pp. 1–8.

[28] K. Sonoda, K. Ishikawa, T. Eimori, and O. Tsuchiya, "Discrete dopant effects on statistical variation of random telegraph signal magnitude," *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 1918–1925, Aug. 2007.

[29] K. Ito, T. Matsumoto, S. Nishizawa, H. Sunagawa, K. Kobayashi, and H. Onodera, "The impact of RTN on performance fluctuation in CMOS logic circuits," in *Proc. Int. Rel. Phys. Symp.*, Apr. 2011, pp. CR.5.1–CR.5.4.

[30] L. Gerrer *et al.*, "Modelling RTN and BTI in nanoscale MOSFETs from device to circuit: A review," *Microelectron. Rel.*, vol. 54, no. 4, pp. 682–697, 2014.

[31] X. Wang, A. R. Brown, N. Idris, S. Markov, G. Roy, and A. Asenov, "Statistical threshold-voltage variability in scaled decananometer bulk HKMG MOSFETs: A full-scale 3-D simulation scaling study," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2293–2301, Aug. 2011.

[32] S. Cheemalavagu, P. Korkmaz, K. V. Palem, B. E. S. Akgul, and L. N. Chakrapani, "A probabilistic CMOS switch and its realization by exploiting noise," in *Proc. IFIP Int. Conf. VLSI*, 2005, pp. 535–541.

[33] G. Gielen and E. Maricau, "Stochastic degradation modeling and simulation for analog integrated circuits in nanometer CMOS," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2013, pp. 326–331.

[34] P. Weckx *et al.*, "Defect-centric perspective of combined BTI and RTN time-dependent variability," in *Proc. IEEE Int. Integr. Rel. Workshop (IIRW)*, Oct. 2015, pp. 21–28.

[35] R. Naous, M. Al-Shedivat, and K. N. Salama, "Stochasticity modeling in memristors," *IEEE Trans. Nanotechnol.*, vol. 15, no. 1, pp. 15–28, Jan. 2016.

[36] S. Ghosh and K. Roy, "Parameter variation tolerance and error resiliency: New design paradigm for the nanoscale era," *Proc. IEEE*, vol. 98, no. 10, pp. 1718–1751, Oct. 2010.

[37] R. Naous and K. N. Salama, "Approximate computing with stochastic memristors," in *Proc. 15th Int. Workshop Cellular Nanosc. Netw. Appl. (CNNA)*, Aug. 2016, pp. 1–2.

[38] *Predictive Technology Model*. Accessed: Oct. 2016. [Online]. Available: http://ptm.asu.edu/

[39] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*. London, U.K.: Pearson Education, 2003.

[40] M. Alioto and G. Palumbo, "Impact of supply voltage variations on full adder delay: Analysis and comparison," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 12, pp. 1322–1335, Dec. 2006.

[41] Z. Yang, A. Jain, J. Liang, J. Han, and F. Lombardi, "Approximate XOR/XNOR-based adders for inexact computing," in *Proc. IEEE 13th Int. Conf. Nanotechnol.*, Aug. 2013, pp. 690–693.

[42] J. Liang, J. Han, and F. Lombardi, "New metrics for the reliability of approximate and probabilistic adders," *IEEE Trans. Comput.*, vol. 62, no. 9, pp. 1760–1771, Sep. 2013.

[43] H. Jiang, C. Liu, L. Liu, F. Lombardi, and J. Han, "A review, classification, and comparative evaluation of approximate arithmetic circuits," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 4, p. 60, 2017.

[44] V. Camus, J. Schlachter, and C. Enz, "Energy-efficient inexact speculative adder with high performance and accuracy control," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 45–48.

[45] Y. Kim, Y. Zhang, and P. Li, "An energy efficient approximate adder with carry skip for error resilient neuromorphic VLSI systems," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2013, pp. 130–137.
[46] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. Reading, MA, USA: Addison-Wesley, 2010.
[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[48] D. May and W. Stechele, "Voltage over-scaling in sequential circuits for approximate computing," in *Proc. Int. Conf. Des. Technol. Integr. Syst. Nanosc. Era (DTIS)*, Apr. 2016, pp. 1–6.
[49] R. Ragavan, B. Barrois, C. Killian, and O. Sentieys, "Pushing the limits of voltage over-scaling for error-resilient applications," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2017, pp. 476–481.
[50] D. Mohapatra, V. K. Chippa, A. Raghunathan, and K. Roy, "Design of voltage-scalable meta-functions for approximate computing," in *Proc. Design, Automat. Test Europe*, Mar. 2011, pp. 1–6.
[51] R. Tessier and W. Burleson, "Reconfigurable computing for digital signal processing: A survey," *J. VLSI Signal Process.*, vol. 28, nos. 1–2, pp. 7–27, May 2001.
[52] K. K. Hasan, U. K. Ngah, and M. F. M. Salleh, "Efficient hardware-based image compression schemes for wireless sensor networks: A survey," *Wireless Pers. Commun.*, vol. 77, no. 2, pp. 1415–1436, 2014.
[53] T. Wark, P. Corke, J. Karlsson, P. Sikka, and P. Valencia, "Real-time image streaming over a low-bandwidth wireless camera network," in *Proc. 3rd Int. Conf. Intell. Sensors, Sensor Netw. Inf.*, Dec. 2007, pp. 113–118.

REN LI (S'17) received the B.S. degree in electrical engineering and its automation (advanced class) from North China Electric Power University, Beijing, China, in 2014, and the M.S. degree in power engineering from Nanyang Technological University, Singapore, in 2015. He is currently pursuing the Ph.D. degree with the King Abdullah University of Science and Technology, Saudi Arabia. His research interests include ultra-low power integrated circuits (IC) and systems design, micro-electro-mechanical-systems IC design and fabrication, and novel devices for the next generation computing.

RAWAN NAOUS (S'13) received the B.E. degree in electrical engineering from the American University of Beirut, Beirut, Lebanon, in 2005, the M.Sc. degree (Hons.) in communications electronics from the Technical University of Munich, Munich, Germany, in 2007, and the Ph.D. degree in electrical engineering from the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, in 2017.

She is currently a Postdoctoral Research Fellow of electrical engineering and computer science with the University of California at Berkeley, Berkeley, CA, USA. Her research interests include mixed analog–digital IC design, memristor, approximate and stochastic computing, hardware–software co-optimization, and embedded machine learning.

HOSSEIN FARIBORZI (M'13) received the B.S. degree in electrical engineering from the Sharif University of Technology, Iran, in 2006, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2013. In 2014, he joined the King Abdullah University of Science and Technology, Saudi Arabia, where he is currently an Assistant Professor of electrical engineering. He has been with Oracle Labs, Nokia, and Intel, as an Electronic Circuit Designer and as a System Designer.

His research interests include ultra-low-power digital/mixed-signal IC design, circuit design for beyond-CMOS technologies, and hybrid NEMS/CMOS circuits and systems. In 2011, he received the ISSCC Jack Raper Best Paper Award for outstanding technology direction.

KHALED NABIL SALAMA (S'97–M'05–SM'10) received the B.S. degree (Hons.) from the Department Electronics and Communications, Cairo University, Cairo, Egypt, in 1997, and the M.S. and Ph.D. degrees from the Department of Electrical Engineering, Stanford University, Stanford, CA, USA, in 2000 and 2005, respectively. He was an Assistant Professor with the Rensselaer Polytechnic Institute, NY, USA, from 2005 to 2009.

He joined the King Abdullah University of Science and Technology, in 2009. He was the Founding Program Chair, until 2011. His work on CMOS sensors for molecular detection has been supported by the National Institutes of Health. He has authored 200 papers. He holds fifteen patents on low-power mixed-signal circuits for intelligent fully integrated sensors and non-linear electronics, especially memristor devices. He received the Stanford-Berkeley Innovators Challenge Award in biological sciences from the Defense Advanced Research Projects Agency and he was acquired by Lumina Inc.