*Proc. International Symp. on Circuits and Systems,* May 2012

# Impact of Process Variations on Computers Used for Image Processing

Suraj Sindia*, Fa Foster Dai†, Vishwani D. Agrawal‡
Department of Electrical and Computer Engineering
Auburn University, Alabama, AL 36849, USA
*szs0063@auburn.edu, †daifa01@auburn.edu, ‡agrawvd@auburn.edu

Virendra Singh
Department of Electrical Engineering
Indian Inst. of Tech. Bombay, Mumbai 400076, India
Email: viren@ee.iitb.ac.in

***Abstract*—Manufacturing process variations (PV) of transistors in the deep-submicron regime present the single biggest design challenge for large die size VLSI circuits such as processor arrays, GPUs, and FPGAs. However, there are a few applications in signal processing, such as image processing, and speech processing, where errors in computation by the underlying hardware could be tolerated or corrected off-line with readily available image restoration algorithms. In this paper, we qualitatively and quantitatively evaluate the effect of process variation in the underlying hardware (for different technology nodes) on a high level application program such as image processing. We rely on gate level simulation, of the data-path of an image processor comprising of a dedicated multiply-accumulate (MAC) array of size $256 \times 256$, with individual gate delays of the processor sampled from a delay distribution as appropriate for each technology node. Our results show that processing images with PV degraded hardware in technologies beyond 65nm is discernible to the human eye; image quality degrades further at 45nm, and is of unacceptable quality at 32nm and beyond. We also use image restoration algorithms to restore the images corrupted due to processing on PV degraded hardware. Our results show that with standard restoration algorithms, even images processed with high levels of PV (as in 32nm) can be restored to almost the same quality as the image processed on fault-free hardware.**

## I. Introduction

Scaling of MOS transistor dimensions (thanks to Moore's law) has led to a steady increase in functions offered by microprocessor chips. Additionally, the performance (or speed) offered by these scaled devices has also been exponentially increasing. The unprecedented growth in performance of computers, however, has come at a price of an exponential increase in power density (power per unit area). After a point, roughly starting from the later half of the last decade, manufacturers have restrained from increasing the operating frequency of microprocessor chips. This stalling in frequency has prompted microprocessor industry to shift to an alternative computing paradigm such as parallel computing, where individual computers perform at a slower rate, but manage to accomplish functional tasks concurrently to be counted as an individual computer operating at a much faster rate (being roughly equal to number of parallel processors × operating frequency of individual processor).

Another possible route to mitigate the increase in power density with successive generations of a microprocessor chip, without stalling frequency scaling, is to downscale the supply voltage. Such a scaling model is popularly referred to as constant electric field scaling [1].

Regardless of the route taken to minimize power density to keep up the performance gains, the semiconductor industry is beginning to hit a road-block attributed to increased manufacturing process related variations. Reference [2] presents an insightful discussion on the trends in frequency and voltage scaling in the face of increased process variation in advanced CMOS technology nodes. Table I (reproduced from [2]) shows scaling trends in CMOS and its impact on energy, and speed in the advanced CMOS nodes. It predicts variability in transistor characteristics, both within-die and across dice as the single most important impediment to performance gains in highly scaled CMOS nodes. Variability in transistor characteristics within the chip has led to a few gates (also referred to in literature as "outliers"), located at spatially disjoint locations to offer delays that are significantly higher, and in many cases, these "outlier" gates lie on the critical path, or paths that would nominally (without any process variation) have had delays that are comparable to critical path delay. Presence of an "outlier" on critical path or close to critical-path leads to an abrupt increase in the delay offered by these paths, consequently reducing the maximum operable frequency at which functionality of the circuit is guaranteed to be correct. However, if one can trade functionality for speed, that is, under the assumption that only a few paths may have these "outliers," then we should still be able to operate the circuit at its maximum speed (as if there were no process variation) alibi with a few errors.

TABLE I
Technology scaling predictions for the end-of-CMOS era [2]. Manufacturing process variation is projected as the single biggest impediment for performance and energy improvement with device scaling.

| High Volume Manufacturing | 2006 | 2008 | 2010 | 2012 | 2014 | 2016 | 2018 |
|---|---|---|---|---|---|---|---|
| Technology node | 65 | 45 | 32 | 22 | 16 | 11 | 8 |
| Integration capacity | 4 | 8 | 16 | 32 | 64 | 128 | 256 |
| Delay= $\frac{CV}{I}$ scaling | ≈0.7 | >0.7 | Delay scaling will slow down | | | | |
| Energy/Logic Op scaling | >0.5 | >0.5 | Energy scaling will slow down | | | | |
| Variability | Medium | | High | | Very High | | |

The objective of this work is to evaluate the impact of such an "un-guaranteed performance" when we operate circuits at a rate faster than they assuredly can. We use applications in image processing to evaluate the degradation in quality of output images processed using hardware afflicted with random process variation. Human eyes can be quite forgiving on the quality of images that it sees, and image processing, we believe, is among the suitable candidates which can be tried implementing on hardware that no longer guarantees functional correctness.

The paper is organized as follows. Section II describes the

978-1-4673-0219-7/12/$31.00 ©2012 IEEE

impact of process related threshold voltage variation on CMOS gate delay. Details of the signal processing fabric built to model process variation and a numerical model to capture functional violations due to process variation is presented in section III. Section IV presents a comparison of image quality obtained by performing common image processing tasks on this signal processing fabric–with, and without, PV degradation. Section V discusses the restoration of images processed on hardware perturbed with process variation noise. We conclude in section VI.

## II. Modeling Process Variation

### A. Threshold Voltage Variation

Process variation is a term used in the very large scale integration (VLSI) literature to refer to random local variation of characteristics of two or more transistors that are on the same die that are ideally expected to have identical characteristics. Sources of random variation in an integrated circuit, include (but are not limited to) fluctuation in the number of dopant atoms in the channel of metal oxide semiconductor (MOS) transistor, and edge roughness of the laser beam used in lithography. These manufacturing process variations are effectively captured at the MOS transistor device level as variation in the threshold voltage ($V_{th}$) of the MOS transistors. Recent research [3] has shown that the threshold voltage variation can be modeled as a normal distribution, $N\left(\mu_{Vth}, \sigma^2_{Vth}\right)$ with variance ($\sigma^2_{Vth}$), normalized with respect to its mean value is given by

$$\frac{\sigma_{Vth}}{\mu_{Vth}} = \frac{K}{\sqrt{WL}} \tag{1}$$

where $K$ is proportionality constant that depends on oxide thickness and doping concentration.
$W, L$ are width and length of MOS transistor.
Plugging in typical numbers for all quantities above, assuming 90nm technology, we have $K = 8.797 \times 10^{-9}$m, $L = 45$nm, $W = 130$nm, which results in a normalized threshold voltage variance of $\frac{\sigma_{Vth}}{\mu_{Vth}}$ = **6.31%**.

### B. Impact of Threshold Voltage Variation on Circuit Functionality and Performance

Delay $t_d$, offered by a logic gate constructed using MOS transistors is related to threshold voltage of its constituent transistors as follows

$$t_d = \frac{V_{DD}}{(V_{DD} - V_{th})^{\alpha}} t_{D0} \tag{2}$$

where $t_{D0}$ = delay (ns) offered by a gate constructed using zero threshold voltage transistors
$V_{DD}$ = supply voltage (volt)
$V_{th}$ = threshold voltage of MOS transistor (volt)
$\alpha$ = MOS device parameter, value is between 1 and 2. For more advanced technologies, this parameter is closer to 1

Now, if there is variation in threshold voltage, that can be modeled by a Gaussian random variable as described in the earlier section II-A, then the distribution in $t_d$ follows a distribution, that can be obtained by using the random variable transformation specified in equation (2). Plots in Figure 1 show the histogram of threshold voltage ($V_{th}$) and maximum delay ($t_d$) offered by 1000 instances of an inverter built in $L$ = 45nm and 32nm. It can be noticed from these plots that the variation in delay is progressively degrading as the technology is advancing. This variation in threshold voltage and delay as a function of technology is summarized in the plot shown in Figure 2. With the passage of technology, variability in transistors is consistently worsening the delay distribution of transistors. That is to say a bigger fraction of transistors will fall outside the margins of delay, than the circuit was designed to handle. To ensure that the circuit functions correctly, we have to reduce the maximum frequency at which the circuit is operated. This leads to a loss in performance/speed offered by the circuit.

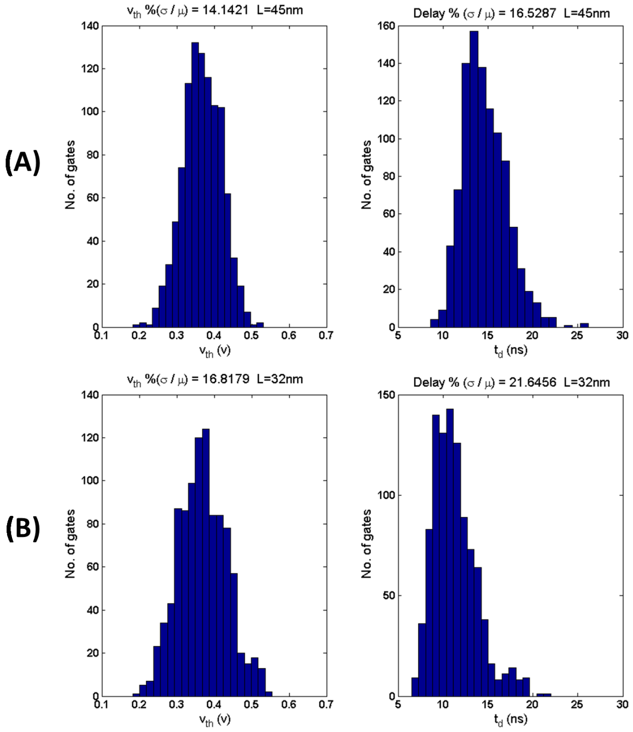

Fig. 1. Histogram of threshold voltage and delay for channel lengths (A) L=45nm, and (B) L=32nm.

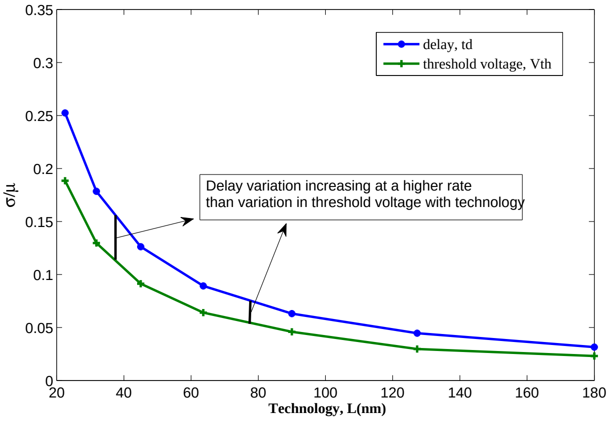

Fig. 2. Variation of delay and threshold voltage as a function of MOS channel length. Notice that as transistors get smaller, the normalized variability in delay offered by a logic gate (here an inverter) is increasing.

## III. Less than Accurate Computing

### A. Related work

As we saw in the previous section, manufacturing variations in the device tends to slow down a circuit/gate. However, if we choose to operate transistors at the same speed disregarding the prevalent process variation, then the functionality of the circuit in question is no longer guaranteed to be correct. In literature [4], [5], we find references to a paradigm where reliable computing is suggested on unreliable hardware. "Soft DSP" techniques such as algorithmic noise tolerance (ANT) [6] have been proposed to alleviate the impact of degradation in results arising from process variations in advanced CMOS technologies. Another interesting model used for emulating probabilistic switching is due to Palem [7]. None of the prior research has considered image processing as a specific application for computing with process variation degraded hardware. The work presented in this paper is an investigation of the impact of process variation on common image processing tasks. We first model the process parameter variations in the underlying hardware. To do this, we assume a hardware structure comprised of an array of multiply-and-accumulate (MAC) units. Each of these MAC units consists of gates with variable delays, whose values are sampled from the distribution in the plots shown in Figure 1 for technology corresponding to 45nm.

### B. Putting Together the Basic Building Blocks

Addition and multiplication are two basic operations needed for most image processing tasks. For example, convolution, which is a commonly used operation involves a series of multiply-accumulate operations of its two inputs. Therefore we build routines in MATLAB for 8 bit addition and multiplication, while incorporating faulty behavior as illustrated in Figure 3. Each bit location is assumed to have been computed through a cascaded chain of logic gates. The delay of the gates in each of these gates is sampled from Figure 1, depending on the technology used for hardware emulation. Process variation noise added output, $y$, can be related to its input, $x$, by a non-linear function $f$ that is defined as in equation 3.

$$y = f\left(x\right) \tag{3}$$

Where

$$x = b_7\, b_6\, b_5\, b_4\, b_3\, b_2\, b_1\, b_0 \qquad y = b'_7\, b'_6\, b'_5\, b'_4\, b'_3\, b'_2\, b'_1\, b'_0 \tag{4}$$

and in general, for all $i = 1 \cdots 7$

$$b'_i = \begin{cases} b_i & \text{for} \quad t_d \le t_{d,th} \\ 0\, or\, 1 & \text{for} \quad t_d > t_{d,th} \qquad \text{with equal probability} \end{cases} \tag{5}$$

where $t_d$ is the actual delay offered by the bit line and $t_{d,th}$ is the threshold delay beyond which the bit line enters a meta-stable state, and the final value that the line settles to, can be either 0 or 1 with equal likelihood. The number chosen for $t_{d,th}$ is usually the delay offered by a gate that has its threshold voltage at 3 times the $\sigma_{Vth}$ away from the mean value $V_{tho}$ given by

$$t_{d,th} = t_d|_{Vth=Vtho+3\sigma_{Vth}} \tag{6}$$

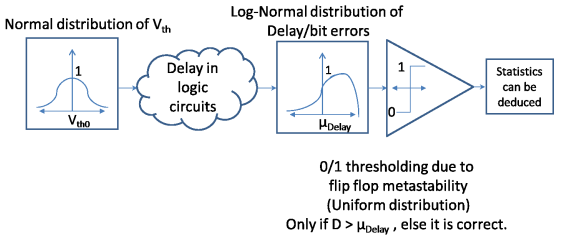
Normal distribution of V_th
1
V_tho
Delay in logic circuits
Log-Normal distribution of Delay/bit errors
1
μ_Delay
1
0
Statistics can be deduced
0/1 thresholding due to flip flop metastability (Uniform distribution)
Only if D > μ_Delay , else it is correct.

Fig. 3. Block diagram showing the PDFs of different random variables as we traverse different levels of abstraction, starting from transistors to software (or algorithmic level).

### C. Software Emulation

We build a signal processing fabric to deal with images of size 256×256 pixels. PV is captured in software as follows: We emulate a 256 × 256 array of 8-bit *add-store* units. Each *add-store* unit in turn consists of cascaded chain of 1 bit full adders to function as 8-bit adder with their outputs connected to 8-bit register as shown in Figure 4. The full adder consists of AND-OR logic gates, whose delays are sampled from delay distribution described in section III-B. We use these arrays of *add-store* units to perform eight bit arithmetic operations such as addition and subtraction. Multiplication is achieved by repeated addition. Further, if any arithmetic operation results in a value in excess of 255, the *add-store* elements are designed to saturate to 255, thus mimicking real-world scenario where 8-bit gray scale images have a maximum intensity level of 255. For simplicity, we do not use color images and restrict ourselves to the use of gray scale images in all our experiments, as we shall see in the sequel.

## IV. Experimental Results and Discussion

We conducted a low pass filter operation on two test images "cameraman" and "baby face". For both the test images, we repeated the experiment with and without process variation. For low pass filtering, we used the mask:

$$\xi = .25 \times \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{7}$$

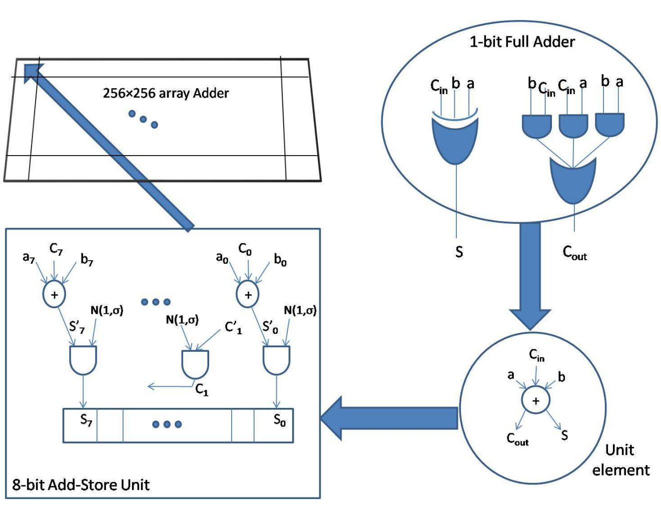

Fig. 4. Synthesis of 256×256 MAC array, with PV noise added for software emulation of PV degraded hardware.

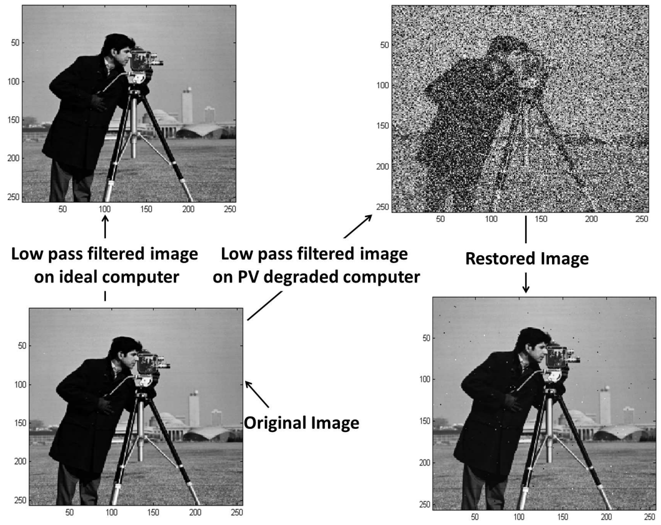
Low pass filtered image on ideal computer
Low pass filtered image on PV degraded computer
Restored Image
Original Image
50
100
150
200
250

Fig. 5. "Camera-man" images with processing (low pass filtering with mask $\eta$) carried out on hardware with and without process variation. Added process variation noise is equivalent to the model developed for 45nm technology node (delay $\sigma/\mu = 6.5\%$). The restored image using a median filter is also shown in the figure.

Cameraman images are shown in Figure 5. We find that the resulting images tend to have an increased concentration of scattered white and black spots (similar to salt and pepper noise). We also notice that the concentration of white saturated points is more than black points. This can be reasoned out as follows: Low pass filtering which is averaging tends to increase the low level intensities. If there are any MSB bit flips to zero (making the value smaller than what it should be), in the early portions of the processing, their effect gets alleviated over subsequent steps of filtering. However, if there is a MSB bit flip to a 1, this effect tends to be cumulative, often giving rise to an increased concentration of bright spots.

Figure 6 shows high pass filtered babyface images processed without and with process variation. Notice the pronounced accumulation of white spots in the high pass filtered image is due to the cumulative effect of bits flipping to 1. For high pass filtering, we used the mask:

$$\psi = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{8}$$

## V. Image Restoration

With the use of available median filters published in literature [8], [9], we were able to filter out the salt-pepper noise resulting from computation performed on PV degraded hardware. The filtered images are shown in figures 5 and 6.

## VI. Conclusion and Future Work

In this paper we developed a high level functional model for transistor manufacturing process variations, and used this model to study its impact on common image processing tasks such as spatial low pass filtering and high pass filtering. We found that process variation induced delay variability in computing elements used for image processing tasks will usually result in "salt and pepper" type of noise at the image output. Further, we find enough examples in literature [8], [9], [10] on robust salt and pepper noise filters. Such filters, were used to reliably recover images processed on PV degraded hardware. Being able to resiliently compute on process variation degraded hardware will allow the extension of Moore's law into the late CMOS era, and its consequent benefits in continued frequency upscaling, while keeping dynamic power density almost constant. An important direction of future research is in developing filters and algorithms that are *aware* of the fact that underlying hardware is "bad," and factors this in as it accomplishes its intended function.

## References

[1] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS Scaling for High Performance and Low Power - the Next Ten Years," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 595–606, Apr. 1995.

[2] S. Borkar, "Design Perspectives on 22nm CMOS and Beyond," in *Proc. 46th ACM/IEEE Design Automation Conference*, Jul. 2009, pp. 93–94.

[3] X. Yuan, T. Shimizu, U. Mahalingam, J. S. Brown, K. Z. Habib, D. G. Tekleab, T.-C. Su, S. Satadru, C. M. Olsen, H. Lee, L.-H. Pan, T. B. Hook, J.-P. Han, J.-E. Park, M.-H. Na, and K. Rim, "Transistor Mismatch Properties in Deep-Submicrometer CMOS Technologies," *IEEE Transactions on Electron Devices*, vol. 58, no. 2, pp. 335–342, Feb. 2011.

[4] 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. 2005.

[5] P. Bose, "Designing Reliable Systems With Unreliable Components," *IEEE Micro*, vol. 26, no. 5, pp. 5–6, Sep. 2006.

[6] R. Hegde and N. R. Shanbhag, "Energy-Efficient Signal Processing via Algorithmic Noise-Tolerance," in *Proc. International Symposium on Low Power Electronics and Design*, 1999, pp. 30–35.

[7] K. V. Palem, "Energy Aware Computing Through Probabilistic Switching: A Study of Limits," *IEEE Transactions on Computers*,, vol. 54, no. 9, pp. 1123–1137, Sep. 2005.

[8] G. Pok, J.-C. Liu, and A. S. Nair, "Selective Removal of Impulse Noise Based on Homogeneity Level Information," *IEEE Trans. on Image Processing*, vol. 12, no. 1, pp. 85–92, Jan. 2003.

[9] R. H. Chan, C.-W. Ho, and M. Nikolova, "Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Detail-Preserving Regularization," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1479 –1485, Oct. 2005.

[10] E. Abreu, M. Lightstone, S. K. Mitra, and K. Arakawa, "A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images," *IEEE Trans. on Image Processing*, vol. 5, no. 6, pp. 1012–1025, Jun. 1996.

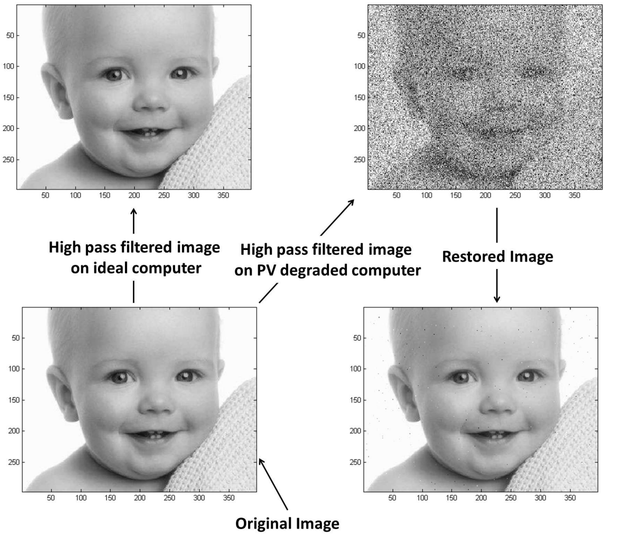
High pass filtered image on ideal computer
High pass filtered image on PV degraded computer
Restored Image
Original Image
50
100
150
200
250
300
350

Fig. 6. "Baby-face" images with processing (edge enhancement) carried out on hardware with and without process variation. Process variation noise added is equivalent to the model developed for 45nm technology node (delay $\sigma/\mu = 6.5\%$). The restored image using a median filter is also shown in the figure.