# A Cross-Layer Power and Timing Evaluation Method for Wide Voltage Scaling

Wenjie Fu, Leilei Jin, Ming Ling*, Yu Zheng and Longxing Shi
*National ASIC System Engineering Technology Research Center*
Southeast University, Nanjing, Jiangsu, China
{wenjfu, leileijin, trio, yuzheng, lxshi}@seu.edu.cn

***Abstract*—Wide supply voltage scaling is critical to enable worthwhile dynamic adjustment of the processor efficiency against varying workloads. In this paper, a cross-layer power and timing evaluation method is proposed to estimate the processor energy efficiency using both circuit and architectural information in a wide voltage range. The process variations are considered through statistical static timing analysis while the voltage effect is modeled through secondary iterated fittings. The error for estimating processor energy efficiency decreases to 8.29% when the supply voltage is scaled from 1.1V to 0.6V, while traditional architectural evaluations behave more than 40% errors.**

***Keywords*—Power and timing evaluation, statistical static timing analysis, wide supply voltage, cross-layer evaluation**

## I. Introduction

Artificial Intelligence of Things (AIoT) are expected to drive the electronic market of the next decades. Such devices need to work under both ultra-low voltage for high energy efficiency and normal voltage for high performance against varying workload demands. Several architectures have been proposed to realize both high energy efficiency and performance under a wide supply voltage range [1]. These studies propose innovations in the architecture layer, but the energy efficiency is only verified after the physical design flow or even tape out. To fast explore the optimal architecture configurations under the wide voltage range, the cross-layer evaluation is needed from the circuit layer to the architecture layer.

Architecture layer performance evaluation, such as gem5 [2], takes cycles as granularity. However, the absolute time in one cycle highly depends on the supply voltage. The absolute time (i.e., the delay of the circuit) is normally estimated by the static timing analysis in the circuit layer. Equation (1) gives the relationship between power, performance and energy efficiency. The energy efficiency, $EE$, is defined as the number of operations per second, $OPS$, that can be executed with a Watt [3]. The $OPS$ can be rewritten as the number of operations per cycle, $OPC$, multiples the system frequency, $Freq$, where $OPC$ is obtained through architecture simulations and $Freq$ is limited by the supply voltage $V$.

$$EE = \frac{OPS}{Power} = \frac{OPC * Freq}{P_{leakage}(V) + P_{dynamic}(V, Freq)} \quad (1)$$

McPAT is a popular power and timing evaluation tool widely used in processor evaluations [4]. It calculates power based on scaling factors from the ITRS roadmap for different voltages and technology nodes under a given frequency. It just gives the trend other than a precise result as the frequency is not a constant when the supply voltage is scaled down. Moreover, the circuit timing at and beyond the 28 nm technology node is extremely sensitive to variations in manufacturing processes, supply voltages, and temperatures (PVT) [5]. As the supply voltage decreases, the circuit becomes increasingly susceptible to PVT variations. The circuit delay increases rapidly to maintain sufficient timing margins to avoid failures. Under such circumstances, McPAT-PVT [6] was proposed to consider the process variation (PV) influences in advanced technology nodes. However, it still has several constraints. First, it cannot provide a reasonable frequency for the target processor, which is an important metric in energy efficiency evaluation. Second, the voltage effect is scaled by factors in leakage power modeling. Third, the influence of PVs on the path delay is assumed as a constant across different supply voltages. Last but not least, the power and timing of functional units are based on Verilog simulation results, which cannot be parameterized for fast design space explorations.

In this paper, a statistical standard cell library is firstly built for quantifying the cell delay under different operating conditions (i.e., input slew, output loads, supply voltages and temperatures) and the PV effects. The effect of operating conditions is modeled by a quadratic fitting while the cell delay deviation is considered through a secondary iterated fitting, from operating conditions to PVs. Based on the library, the circuit layer analysis is realized through the block-based statistical static timing analysis (SSTA) by expanding the OpenTimer [7], an open-source deterministic static timing analysis tool. The SSTA tool is then taken as an interface for the McPAT-PVT simulator to accurately estimate the timing of functional units. The leakage power is considered as an exponential of PVs while the dynamic power is calculated using the timing results from circuit layer analysis. The main contributions of this paper can be summarized as follows:

1) Proposing a quadratic model that quantifies the standard cell delay under different operating conditions such as supply voltages and temperatures.

2) Proposing a secondary iterated fitting model for the PV effects to the cell delay deviation under different operating conditions. The iterated model of each cell is written into a library for supporting the block-based statistical timing analysis, which firstly enables voltage scaling analysis for circuit delay deviations under PV effects.

This work was supported by the National Natural Science Foundation of China under Grant No. 61974024.

*Ming Ling is the corresponding author.

978-1-7281-1085-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.

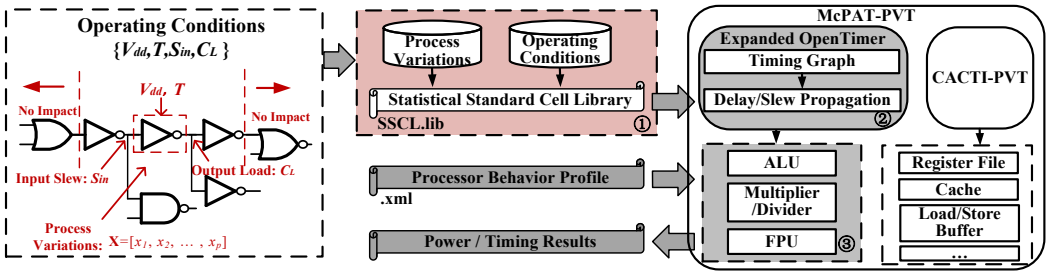


Fig. 1. Block Diagram of Proposed Evaluation Method

3) Proposing a cross-layer power and timing evaluation method. Based on the reasonable frequency estimated in the circuit layer, the architecture layer power and timing evaluation accuracy can be guaranteed under a wide supply voltage range, which is more accurate than the traditional architecture layer evaluations.

The rest of this paper is structured as follows: Section II introduces the related works. Section III describes the overall evaluation method for voltage scaling. Section IV gives the details of statistical modeling for power and timing. The experimental setup and the accuracy of evaluation are discussed in Section V. Finally, Section VI concludes this paper.

## II. RELATED WORKS

OpenTimer is a circuit static timing analysis tool that efficiently supports both block-based and path-based timing propagations [7]. The cell delay is obtained through the lookup table in the standard cell library, which is built in the single voltage node and ignores the influence of PVs. Statistical static timing analysis (SSTA) considers PV influences and treats the delay as a probability density function instead of constant values [5]. It has a high precision in modeling the circuit delay deviation under given operating conditions without considering the influence of different conditions. Therefore, existing SSTA technology only focuses on the a certain condition.

In architectural evaluations, McPAT [4] is an integrated power, area, and timing modeling framework for multi-core architectures. All the registers and caches inside a processor core are modeled by CACTI, a detailed cache model [8]. McPAT-PVT is developed for FinFET-based processors, considering PV influences at the normal supply voltage [6]. It is built based on McPAT and FinCANON [9], a delay and power modeling framework for cache and network on chip (NoC). However, both McPAT and McPAT-PVT do not have a detailed model for functional units. Their power and delay are scaled from those of existing processors and cannot ensure correctness for different supply voltages. Thus, those wide voltage designs and evaluations based on McPAT cannot give an accurate result in low voltages [10] [11]. Despite VARIUS and VARIUS-NTV provide more accurate SRAM models for McPAT-based evaluations, they focus on wide voltage caches and other storage microarchitectures, other than functional units [12] [13].

Authors of [14] presented the energy efficiency in approximate computing design of neuromorphic systems, from the algorithm level to the circuit level. It quantifies metrics in each layer independently. Authors of [15] presented a cross-layer adaptive approach for optimizing the power and performance of STT-MRAM, from device level simulations to system level simulations. Their cross-layer evaluation aims to obtain the high-level feature for determining the memory working mode, other than an optimal supply voltage node. Hence, it needs an accurate architectural energy efficiency evaluation method that flexibly models the functional units for voltage scaling.

## III. CROSS-LAYER EVALUATION FLOW

In this paper, a cross-layer power and timing evaluation method is proposed that can be used to obtain the processor energy efficiency in a wide voltage range. In the circuit layer, the PVs and operating conditions are considered during the static timing analysis based on the proposed statistical cell library. The statistical propagations of delay and slew from one cell to another are calculated by using the timing graph captured by OpenTimer. After quantifying the delay variations of functional units, the frequency is set to be the reciprocal of the circuit delay and is then used for dynamic power calculation. In the architecture layer, caches and other storage units, e.g., register files, load/store buffers, are modeled using FinCANON, which is capable of wide voltage evaluations [9]. The functional unit timing analysis is performed by the expanded OpenTimer. Fig. 1 shows the cross-layer evaluation flow, containing three parts: ① the cell layer statistical library establishment, ② the circuit layer statistical delay and slew propagation and ③ the overall energy efficiency evaluation in the architecture layer.

**Cell layer statistical library establishment**. Fig. 2 shows the construction flow of the statistical standard cell library (SSCL). Each cell is modeled under different input pins and output rise/fall transitions. The mean of cell delays is considered as a function of operating conditions while delay variations are quantified considering the PV influences. Both the cell delay and the output slew are modeled by the nonlinear fittings. Besides, the influence of the PVs to the delay at each cell is also modeled. The input fitting sets for PVs are obtained through 5,000 times Monte Carlo (MC) simulations. The coefficients of each variation which mean different contributions to the delay can be extracted from MATLAB's regression function based on the formulas given in Section IV. Afterwards, a set of coefficients can be extracted for each cell to reveal the contributions of PVs to the cell delay under a given operating condition. Finally, the extracted coefficients for each cell with different pins and transitions are stored in SSCL for block-based statistical static timing analysis.

Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.

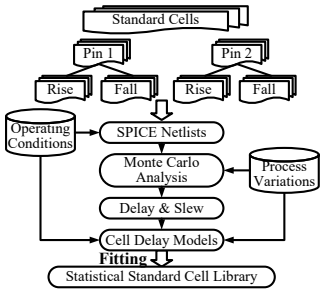


Fig. 2. Statistical Standard Cell Library Construction Flow

**Circuit layer statistical delay and slew propagation**. In the circuit layer, the cell level netlist is analyzed by OpenTimer for getting the timing graph, which contains timing nodes (cell pins) and timing arcs (from one pin to another). The statistical *SUM* and *MAX* operations are then given for the statistical static timing analysis. The procedures in detail are as follows. First, a gate-level netlist which needs to be analyzed is submitted into OpenTimer as the input file. The timing graph will be obtained secondly for delay and slew propagation. Thirdly, OpenTimer including the promoted *SUM* and *MAX* operations will give the timing results of the input netlist based on SSCL. During the propagation calculation, four operating conditions are considered: input slew, output loads, supply voltages and temperatures. To sum up, once the operating conditions and the cells' PVs are defined by the user or imported from real-fabrication data, the circuit delay and its variation can be calculated.

**Architecture layer energy efficiency evaluation**. The processor evaluation flow takes the delay information from the expanded OpenTimer as input. It specifies the temperature, the supply voltage, the technology node, and configurations of processor components, e.g., cache sizes, arithmetic-logic units (ALUs), floating-point units (FPUs), and the pipeline architecture. The evaluation for functional units includes the delay distribution estimation based on the proposed SSTA and power estimation based on the proposed power models, considering both operating conditions and PVs. The evaluation flow also contains the processor behavior profile (e.g., number of cache accesses, hit rates) obtained from cycle accurate simulators (e.g., gem5 [2]) or analytical models [16]. The evaluation flow will generate three types of information: area, power and timing. The area profile contains the area information of all the components in the processor. The power profile includes the dynamic and leakage power of each processor components. The timing information of each processor components will be characterized in the timing profile. Different from the results from conventional tools, the power and timing results are in distributions under the PV effects.

## IV. STATISTICAL ANALYSIS

### A. Cell Delay Statistical Modeling

The delay of the standard cell $D_{cell}$ is determined by operating conditions and PVs. The cell delay model is divided into two parts: $D_\theta$ and $D_{PV}$, which quantifies the effect of operating conditions and PVs, respectively. In the traditional static timing analysis, the nonlinear delay model

Table I Operating Conditions

| Operating Conditions | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Voltage (V) | 0.65 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 |
| Slew (ps) | 1 | 10 | 20 | 30 | 40 | 50 | 60 |
| Load (fF) | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 |
| Temp (°C) | -30 | -5 | 20 | 45 | 70 | 95 | 120 |

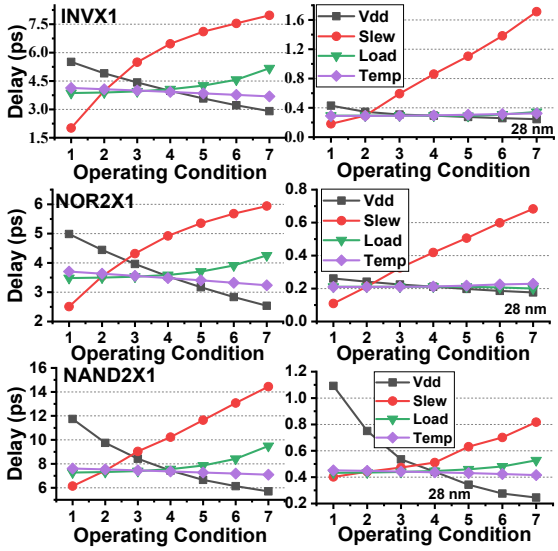


Fig. 3. Means (Left) and SDs (Right) of Three Standard Cells Under Different Operating Conditions

[17] is built that calculates the cell delay by indexing the input slew and output load under a given supply voltage and temperature. Additionally, the output slew is also captured from a lookup table based on the output capacitance and the input slew. Fig. 3 shows the influences of these conditions to the INVX1/NAND2X1/NOR2X1 delay distributions through SPICE and MC simulations. The left sub-figures show the mean values of delay distributions under different conditions while the right sub-figures show their standard deviations (SDs). The values of each condition are given in Table I. As Fig. 3 shows, it is obvious to find that the delay $D_\theta$ has a quadratic dependence with the operating conditions.

The delays under different operating conditions which are randomly generated are taken as the training data. Hence, the weight of operating condition $\theta = [V_{dd}, T, S_{in}, C_L]$ can be calculated through (2), where $\mathbf{P}$ and $\mathbf{Q}$ are the coefficient vectors of the operating conditions, which can be quantified by nonlinear regressions. Meanwhile, the output slew can be modeled in the same manner [18].

$$D_\theta = \mathbf{P} * \theta + \mathbf{Q} * \theta^2 \tag{2}$$

In [19], a quadratic model was proposed to capture the nonlinearity between cell delay deviation $D_{PV}$ and PVs under a given supply voltage and input slew, shown in (3). The $p \times 1$ Gaussian variation vector, $\mathbf{X} = [x_1, x_2, ..., x_p]$ represents process variations such as threshold voltage variations, gate channel length variations and gate oxide thickness variations.

$$D_{PV} = \sum a_i x_i + \sum b_i x_i^2 + \xi \tag{3}$$

The coefficients $a_i$, $b_i$ are the first-order and the second-order terms of $D_{PV}$, representing the contribution to the cell delay deviation. Considering the effects of the voltage

Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.

and input slew given in Fig. 3, a secondary iterated fitting-based model is proposed. It firstly quantifies the coefficients $a_i$, $b_i$ using the cell's operating conditions. As the supply voltage and the input slew are quadratic to the cell delay deviation, the coefficients $a_i$, $b_i$ are represented by a quadratic function of voltage and slew, as is shown in (4). Vectors $\mathbf{M} = [m_0, m_1, m_2, m_3, m_4]$ and $\mathbf{N} = [n_0, n_1, n_2, n_3, n_4]$ are corresponding coefficients for each $i$. Afterwards, the coefficients are taken back into (3) for calculating the deviation of the cell delay. As the mean of $D_{PV}$ is $\sum b_i + \xi$ and equals to zero, $\xi$ is equal to the opposite of $\sum b_i$.

$$\begin{cases} a_i = m_{i,0} + m_{i,1}V_{dd} + m_{i,2}S_{in} + m_{i,3}V_{dd}^2 + m_{i,4}S_{in}^2 \\ b_i = n_{i,0} + n_{i,1}V_{dd} + n_{i,2}S_{in} + n_{i,3}V_{dd}^2 + n_{i,4}S_{in}^2 \end{cases} \tag{4}$$

Ultimately, the cell delay $D_{cell}$ that equals to the sum of $D_\theta$ and $D_{PV}$ can be written as (5), where $c$ is a constant when the operating condition is given.

$$\begin{aligned} D_{cell} &= \sum a_i x_i + \sum b_i x_i^2 + \mathbf{P}\theta + \mathbf{Q}\theta^2 - \sum b_i \\ &= \sum a_i x_i + \sum b_i x_i^2 + c \end{aligned} \tag{5}$$

A thousand sets of operating conditions $\{V_{dd}, T, S_{in}, C_L\}$ are randomly generated for obtaining the coefficients $\mathbf{P}$ and $\mathbf{Q}$. All of the coefficients $\mathbf{P}, \mathbf{Q}, \mathbf{M}, \mathbf{N}$ mentioned above in (2)-(5) are calculated and stored into the statistical standard cell library in order to conduct statistical static timing analysis. After determining the operating conditions and PVs, cell delays can be calculated.

### B. Block-Based Statistical Static Timing Analysis

*SUM* and *MAX* are two fundamental operations to calculate the arrival times in the block-based SSTA. Denoting that the arrival time of the input of a cell is $D_1$ and the delay of itself is $D_2$, the arrival time of its output pin is the *SUM* operation of $D_1$ and $D_2$, represented as $D_{sum}$:

$$D_{sum} = \sum (a_{D1} + a_{D2})x_i + \sum (b_{D1} + b_{D2})x_i^2 + c_{D1} + c_{D2} \tag{6}$$

In contrast, the exact result of the *MAX* operation is computationally expensive. If there are two timing arcs pointing to the same output pin of a cell, denoted as $T_1$ and $T_2$, the arrival time of the pin is the *MAX* operation of $T_1$ and $T_2$. Using the methodology proposed by Clark [20], the result of the *MAX* operation can be written as (7) and the coefficients can be computed as (8):

$$D_{max} = \sum a_{max} x_i + \sum b_{max} x_i^2 + c_{max} \tag{7}$$

$$\begin{cases} a_{max} = \phi\ a_{T1} + (1-\phi) a_{T2} \\ b_{max} = \phi\ b_{T1} + (1-\phi) b_{T2} \\ c_{max} = \phi\ c_{T1} + (1-\phi) c_{T2} + \varphi \sigma_{T1-T2} \end{cases} \tag{8}$$

where $\phi$ and $\varphi$ are the cumulative distribution function (CDF) and the probability density function (PDF) of the standard Gaussian distribution calculated at $\mu_{T1-T2}/\sigma_{T1-T2}$. The value of $\mu_{T1-T2}$ and $\sigma_{T1-T2}$ can be obtained as (9).

$$\begin{cases} \mu_{T1-T2} = \sum (b_{T1} - b_{T2}) + c_{T1} - c_{T2} \\ \sigma_{T1-T2}^2 = \sum (a_{T1} - a_{T2})^2 + 2 * \sum (b_{T1} - b_{T2})^2 \end{cases} \tag{9}$$

### C. Power Modeling

Regarding the power model, a widely known processor power model is employed that the total power of the target processor is the sum of dynamic power and leakage power, represented by $P_{dyn}$ and $P_{leak}$, respectively. The dynamic power is shown in (10) where the term $\alpha$ is the average rate of capacitance that switching per cycle while $V_{dd}$ and $f$ are the supply voltage and frequency of the processor. $C_{tot}$ is the total load capacitance of the whole netlist. The parameter $\alpha$ is a strong function of circuit activity that differs from one application to another. The frequency is assumed to be the reciprocal of maximal delay of execution stage, which is the function of PVs ($\mathbf{X}$) and operating condition ($\theta$) analyzed through the SSTA above.

$$P_{dyn} = \alpha * C_{tot} * V_{dd}^2 * f(\mathbf{X}, \theta) \tag{10}$$

Leakage power is modeled as a lognormal random variable denoted as $\ln N(\mu, \sigma)$ because the relationship between leakage power and process variations is exponential rather than polynomial [21]. Hence, leakage power can be estimated by (11) where $\omega_0$, $\omega_i$, $\omega_V$ and $\omega_T$ are coefficients of PVs, supply voltages and temperatures.

$$P_{leak} = e^{\omega_0 + \sum \omega_i x_i + \omega_V V_{dd} + \omega_T T} \tag{11}$$

## V. EVALUATION

### A. Experimental Setup

In the circuit layer, the accuracy of timing analysis is verified using the ISCAS85 benchmark suite under SMIC 28 nm PDK [22]. The supply voltage ranges from 0.6V to 1.1V and the temperature ranges from -30°C to 125°C. In the architecture layer, the power and timing are tested based on Alpha 21364 [23]. The Verilog files of functional units are obtained from an open source processor: OpenSPARC T1 [24], which is also used in McPAT [4] and McPAT-PVT [6].

### B. Block-Based SSTA Accuracy

In the circuit layer analysis, a statistical library is built using 16 standard cells for simplification[1]. Each cell is simulated through 5,000 times SPICE Monte Carlo analysis under different operating conditions, which are randomly generated. The supply voltage of experiments decreases from 1.1V to 0.6V along with the accuracies of cell delay models deteriorating, shown in Fig. 4. As mentioned above, the cell delay is divided into $D_\theta$ and $D_{PV}$. $D_{PV}$ quantifies delay deviations using a secondary iteration form to solve the non-Gaussian cell delay distribution when the voltage is lower than 0.7V. However, the Clark's *MAX* approximation calculation in SSTA is based on the assumption that the input sources follow Gaussian distributions, which induces errors in low voltage circumstances.

Meanwhile, ISCAS85 benchmark circuits are conducted by block-based SSTA under wide voltages from 0.6V to 1.1V using the statistical standard cell library. The temperature and

[1]The standard cells modeled in this paper are: INVX1/2/4/8, NAND2X1/2/4/8, NOR2X1/2/4/8, DQ2X2/4 and LATCH2X2/4.

Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.

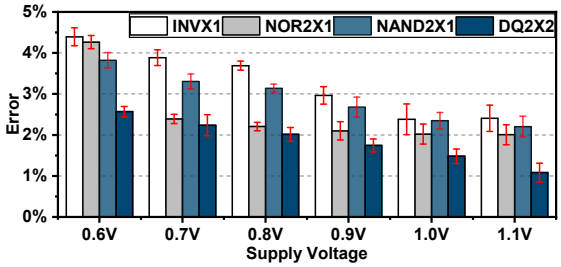


Fig. 4. Mean and Standard Deviation Errors of Cell Delay Modeling

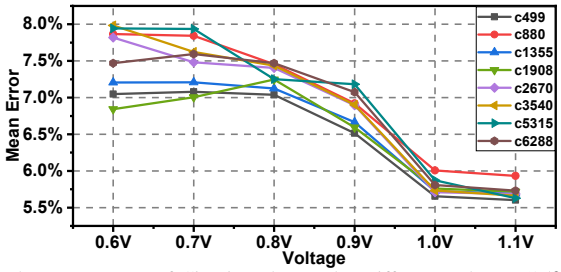


Fig. 5. Mean Errors of Circuit Delay Under Different Voltages (25°C)

the output capacitance are set to be 25°C and 0.1fF along with the 10ps input slew. Compared with the results from the SPICE Monte Carlo analysis, the mean error of our method is about 6.94% while the standard deviation error is about 2.83% shown in Fig. 5 and Fig. 6, respectively. Unlike the conventional SSTA which can only work at normal voltage, our circuit layer delay model performs well in a wide voltage range. As the error in the cell delay model increases in the ultra-low voltage, the analysis accuracy for ISCAS benchmark circuits shows a precision loss when the voltage is scaled lower. Despite that, it still has a high precision and can be used for timing analysis of architectural layer components.

### C. Power Modeling Accuracy

The frequency calculated from our circuit layer timing analysis has great effects on the dynamic power calculation as is shown in (10). Meanwhile, the leakage power is determined by operating conditions and PVs of the delay in (11). The errors of dynamic power and leakage power modeling for different benchmark circuits are shown in Fig. 7. The average error for total power evaluation is about 6.3% compared to SPICE simulations.

### D. Energy Efficiency Evaluation

As the FPU logic is the critical one in the processor pipeline, the frequency is mostly limited by the FPU, which is considered as a constant in traditional McPAT or McPAT-PVT tools. Fig. 8 shows the FPU delay distribution when the supply voltage is scaled down. When the supply voltage scales from 1.1V to 0.6V, the delay of FPU increases 5.96 times.

Based on the frequency analyzed under different supply voltage nodes, the power of ALUs and FPUs are recalculated in McPAT-PVT using the dynamic and leakage power model proposed in this work. The power consumption of different components in the processor is shown in Fig. 9, including the leakage and dynamic power of I-cache, branch predictor, D-cache, register files, ALU and FPU. The dynamic power of

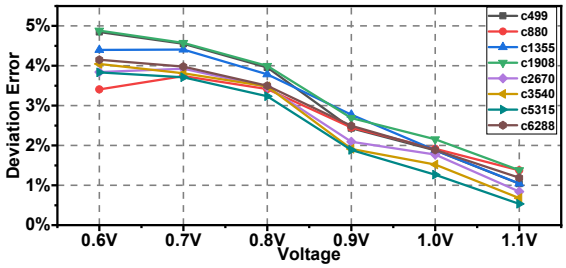


Fig. 6. Standard Deviation Errors of Circuit Delay Under Different Voltages (25°C)

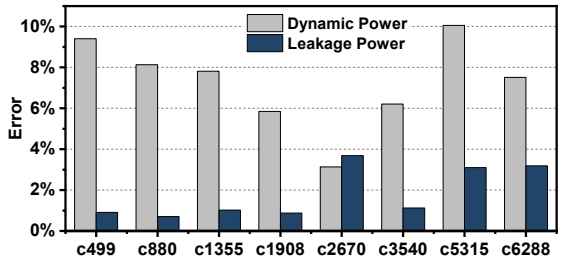


Fig. 7. Errors of Dynamic and Leakage Power Modeling for Different Benchmark Circuits (25°C)

ALU and FPU dramatically declines as the supply voltage is scaled down to 0.6V.[2]

The energy efficiency is calculated using the power and timing results from our method. The results in Fig. 10 are compared with that of SPICE, McPAT and McPAT-PVT. McPAT is the traditional method and McPAT-PVT is the improved McPAT. Note that the SPICE simulation only gives the golden results of ALU and FPU for McPAT-PVT. Other components such as I-Cache and Register Files are still evaluated using CACTI's memory models. As the traditional McPAT and McPAT-PVT cannot provide a reasonable frequency, the energy efficiency calculation is based on the constant 2.5GHz frequency, which is tested at 0.9V. The energy efficiency using our method shows an average evaluation error of 8.29% compared to the SPICE results, while traditional McPAT and McPAT-PVT both show a more than 40% estimation error under lower supply voltages.

## VI. CONCLUSION

In this paper, a cross-layer power and performance evaluation method is proposed for obtaining the processor efficiency in a wide voltage range. In the circuit layer, process variations are considered through the statistical static timing analysis while the voltage deviation effect is modeled through the secondary iterated fitting. An architecture layer evaluation tool is expanded for accurate efficiency evaluation by using the statistical delay model. The circuit layer timing and power modeling is tested by the ISCAS85 benchmark circuits with acceptable precision. The power and performance estimation of the architecture layer is verified through OpenSPARC T1 processor. Compared to using functional unit metrics from SPICE, the error for estimating processor energy efficiency decreases apparently when the supply voltage is scaled from

[2]The inputs of McPAT, McPAT-PVT and our work are xml format files in the McPAT-PVT source code directory generated by the blackscholes benchmark.

Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.

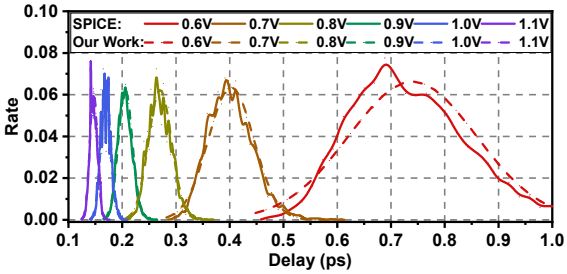


Fig. 8. Delay Distribution Estimation of FPU

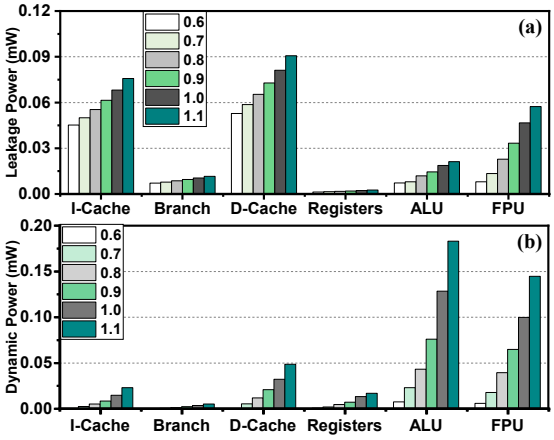


Fig. 9. Leakage and Dynamic Power Trend in Voltage Scaling

1.1V to 0.6V, which behaves better than traditional architectural evaluations. The accuracy of our proposed method becomes more attractive in low voltages.

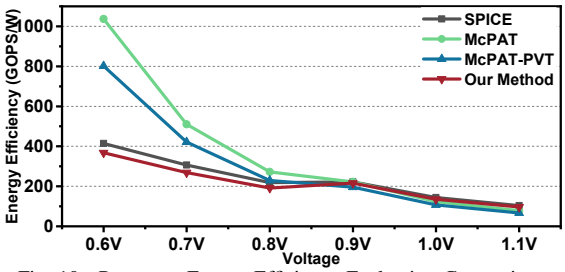


Fig. 10. Processor Energy Efficiency Evaluation Comparisons

## REFERENCES


[1] S. Yin, P. Ouyang, S. Tang, F. Tu, X. Li, S. Zheng, T. Lu, J. Gu, L. Liu, and S. Wei, "A high energy efficient reconfigurable hybrid neural network processor for deep learning applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 4, pp. 968–982, 2017.

[2] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, *et al.*, "The gem5 simulator," *ACM SIGARCH Computer Architecture News*, vol. 39, no. 2, pp. 1–7, 2011.

[3] D. Castells-Rufas, A. Saa-Garriga, and J. Carrabina, "Energy efficiency of many-soft-core processors," *arXiv preprint arXiv:1601.07133*, 2016.

[4] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 469–480, ACM, 2009.

[5] L. Cheng, J. Xiong, and L. He, "Non-gaussian statistical timing analysis using second-order polynomial fitting," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 1, pp. 130–140, 2008.

[6] A. Tang, Y. Yang, C.-Y. Lee, and N. K. Jha, "Mcpat-pvt: Delay and power modeling framework for finfet processor architectures under pvt variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 9, pp. 1616–1627, 2014.

[7] T.-W. Huang and M. D. Wong, "Opentimer: A high-performance timing analysis tool," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 895–902, IEEE Press, 2015.

[8] S. Li, K. Chen, J. H. Ahn, J. B. Brockman, and N. P. Jouppi, "Cacti-p: Architecture-level modeling for sram-based structures with advanced leakage reduction techniques," in *Proceedings of the International Conference on Computer-Aided Design*, pp. 694–701, IEEE Press, 2011.

[9] C.-Y. Lee and N. K. Jha, "Fincanon: A pvt-aware integrated delay and power modeling framework for finfet-based caches and on-chip networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 5, pp. 1150–1163, 2013.

[10] B. Gopireddy, C. Song, J. Torrellas, N. S. Kim, A. Agrawal, and A. Mishra, "Scalcore: Designing a core for voltage scalability," in *2016 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pp. 681–693, IEEE, 2016.

[11] I. Stamelakos, S. Xydis, G. Palermo, and C. Silvano, "Variation-aware voltage island formation for power efficient near-threshold manycore architectures," in *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 304–310, IEEE, 2014.

[12] U. R. Karpuzcu, K. B. Kolluru, N. S. Kim, and J. Torrellas, "Varius-ntv: A microarchitectural model to capture the increased sensitivity of manycores to process variations at near-threshold voltages," in *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)*, pp. 1–11, IEEE, 2012.

[13] S. R. Sarangi, B. Greskamp, R. Teodorescu, J. Nakano, A. Tiwari, and J. Torrellas, "Varius: A model of process variation and resulting timing errors for microarchitects," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 1, pp. 3–13, 2008.

[14] P. Panda, A. Sengupta, S. S. Sarwar, G. Srinivasan, S. Venkataramani, A. Raghunathan, and K. Roy, "Cross-layer approximations for neuromorphic computing: From devices to circuits and systems," in *2016 53nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1–6, IEEE, 2016.

[15] N. Sayed, R. Bishnoi, F. Oboril, and M. B. Tahoori, "A cross-layer adaptive approach for performance and power optimization in stt-mram," in *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 791–796, IEEE, 2018.

[16] W. Fu, M. Ling, W. Wang, and L. Shi, "Amps: Accelerating mcpat power evaluation without cycle-accurate simulations," *IEEE Embedded Systems Letters*, 2019.

[17] T. El Motassadeq, "Ccs vs nldm comparison based on a complete automated correlation flow between primetime and hspice," in *2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC)*, pp. 1–5, IEEE, 2011.

[18] W. Zhang, A. Singhee, J. Xiong, P. Habitz, A. Joshi, C. Visweswariah, and J. Sundquist, "A dynamic method for efficient random mismatch characterization of standard cells," in *2012 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 180–186, IEEE, 2012.

[19] Y. Zhan, A. J. Strojwas, X. Li, L. T. Pileggi, D. Newmark, and M. Sharma, "Correlation-aware statistical timing analysis with non-gaussian delay distributions," in *Proceedings of the 42nd annual Design Automation Conference*, pp. 77–82, ACM, 2005.

[20] C. E. Clark, "The greatest of a finite set of random variables," *Operations Research*, vol. 9, no. 2, pp. 145–162, 1961.

[21] Y. Yang and N. K. Jha, "Finprin: Finfet logic circuit analysis and optimization under pvt variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 12, pp. 2462–2475, 2013.

[22] F. Brglez, D. Bryan, and K. Kozminski, "Combinational profiles of sequential benchmark circuits," in *IEEE international symposium on circuits and systems*, vol. 3, pp. 1929–1934, IEEE, 1989.

[23] A. Jain, W. Anderson, T. Benninghoff, D. Berucci, M. Braganza, J. Burnetie, T. Chang, J. Eble, R. Faber, O. Gowda, *et al.*, "A 1.2 ghz alpha microprocessor with 44.8 gb/s chip pin bandwidth," in *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No. 01CH37177)*, pp. 240–241, IEEE, 2001.

[24] I. Parulkar, A. Wood, J. C. Hoe, B. Falsafi, S. V. Adve, J. Torrellas, and S. Mitra, "Opensparc: An open platform for hardware reliability experimentation," in *Fourth Workshop on Silicon Errors in Logic-System Effects (SELSE)*, pp. 1–6, Citeseer, 2008.


Authorized licensed use limited to: Chinese University of Hong Kong. Downloaded on November 22,2026 at 13:01:57 UTC from IEEE Xplore. Restrictions apply.