Real-time Minimum Energy Point Tracking Using a Predetermined Optimal Voltage Setting Strategy

***Abstract*—Minimizing the energy consumption of processors for a given computational workload is highly desired for matured and energy efficient, information oriented society. In this paper, we refer to a pair of the supply voltage ($V_{\rm DD}$) and threshold voltage ($V_{\rm TH}$), which minimizes the energy consumption of the processor under a given computational workload, as a minimum energy point (MEP in short). Since always running at the MEP largely reduces the energy consumption of processors without fundamental degradation of the performance, a lot of methods for tracking the MEP at runtime have been investigated over the past several years. However, to the best of our knowledge, all the previous methods are based on time-consuming power measurement to identify the MEP at runtime, which prevents the real-time tracking of the MEP. This paper proposes a real-time MEP tracking method based on a predetermined MEP-curve which is characterized as a linear model for each chip at a boot phase. Experimental results obtained using a 50-stage fanout-4 inverter chain designed to reflect the behavior of a microprocessor pipeline demonstrate that the energy loss introduced by the linear approximation MEP model is only 3.1% at the worst case.**

I. INTRODUCTION

One of the most effective approaches for reducing the energy consumption of microprocessors is dynamic supply and threshold voltage scaling. Techniques for dynamically scaling the supply voltage ($V_{\rm DD}$ in the following) and/or threshold voltage ($V_{\rm TH}$ in the following) under dynamic workloads of microprocessors are thus widely investigated over the past 20 years [1], [2], [3], [4]. We refer to the best pair of $V_{\rm DD}$ and $V_{\rm TH}$, which minimizes the energy consumption of the processor under a given operating condition, as a minimum energy point (MEP in the following). $V_{\rm TH}$ of transistors can be dynamically changed by tuning the back-gate bias ($V_{\rm BB}$) of the gates. Although the dynamic scaling of $V_{\rm DD}$ and $V_{\rm TH}$ has been studied for many years, no effective method for finding MEP at runtime has been established yet. There are the following three major reasons why finding MEP at runtime is not trivial;

1) There are two tuning knobs, $V_{\rm DD}$ and $V_{\rm BB}$, to be optimized simultaneously at runtime to find the MEP.
2) Some applications and situations require a real-time response for seeking and identifying the MEP.
3) The MEP depends on the operating condition such as chip temperature, activity factor, process variation, aging status, and performance required for the processor [5].

For example, if the temperature of the chip gets higher the MEP shifts toward the upper right as shown in Fig. 1. If the operating point (i.e. $V_{\rm DD}$ and $V_{\rm TH}$) remains unchanged against the change of the chip temperature, a considerable amount of energy loss is involved. However, since the time constants for chip-level thermal behavior are typically on the order of milliseconds to seconds [6], slow response on the order of milliseconds is also allowed for identifying the MEP. Unlike the temperature change, if the performance requirement for the processor gets more strict, the processor needs to immediately change its operating point to meet the requirement. Since the request of the performance change is typically sent from a real-time operating system asynchronously, finding the MEP and the performance change also have to be done in a real-time manner. However, to the best of our knowledge, all the previous algorithms [1], [2], [3], [4] for finding the MEP at runtime are based on time-consuming power measurement to find the MEP, which prevents the real-time tracking of the MEP. This paper proposes a real-time MEP tracking method using a predetermined MEP-curve which is characterized for a given chip. The key idea of the method is first separating the response to the change of the required performance and the response to the others such as process variation and temperature change. Then, for the change of the required performance, which needs the real-time response, the method skips the power measurement when finding the MEP. This can be done using a predetermined MEP-curve which is characterized at the boot phase for a given processor chip. For adapting the chip-to-chip process variation, where slow response is allowed, we use the time-consuming traditional algorithm for slowly but accurately tracking the MEP.

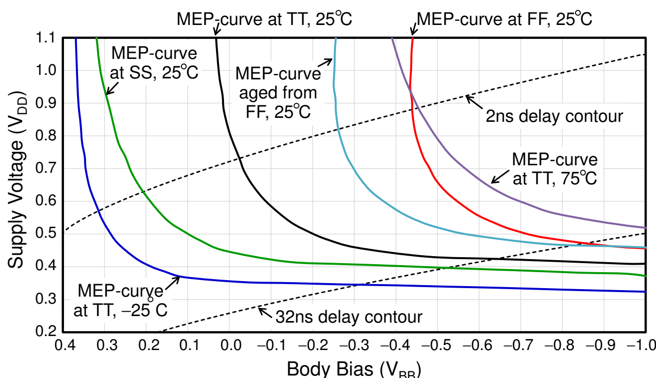

Fig. 1. Minimum energy point curves.

This paper is organized in the following way. Section II shows related work and our contributions. Several important properties of CMOS circuits on the MEP and a real-time MEP

tracking algorithm exploiting the properties are presented in Section III. Section IV validates the MEP tracking algorithm presented in Section III using a commercial 55 nm DDC process technology. Section V concludes the paper.

II. Related Work and Contributions of This Work

A. Minimum Energy Point Tracking

The best pair of $V_{\rm DD}$ and $V_{\rm TH}$ at a given delay, which we refer to as a minimum energy point (MEP), is where the delay and energy contours are tangent [7]. Figure 2 shows constant contours of delay and energy obtained with a 50-stage fanout-4 inverter chain designed to reflect the behavior of a microprocessor pipeline. The energy consumption of the circuit consists of dynamic energy $E_{\rm d}$ and static energy $E_{\rm s}$. $E_{\rm d}$ is a quadratic function of $V_{\rm DD}$ as shown in (1). $E_{\rm s}$ is exponential to $V_{\rm TH}$ and linear to the delay D and $V_{\rm DD}$ as shown in (2). k_1 and k_2 are fitting coefficients. $N_{\rm s}$ is $n_{\rm i} \cdot \phi_{\rm t}$, where $n_{\rm i}$ is ideal factor of MOSFET, which is typically between 1 and 2, and $\phi_{\rm t}$ is thermal voltage which is 26 mV at a room temperature. Assuming the value of $n_{\rm i}$ is 2.0, the value of $N_{\rm s}$ in (2) is 52 mV at a room temperature.

$$E_{\rm d} = k_1 {V_{\rm DD}}^2. \qquad (1)$$

$$E_{\rm s} = k_2 D V_{\rm DD} e^{-\frac{V_{\rm TH}}{N_{\rm s}}}. \qquad (2)$$

When $V_{\rm DD} \gg V_{\rm TH}$, the delay can be accurately modeled using alpha power law MOSFET model [8] as shown in (3), where $V_{\rm DT} = V_{\rm DD} - V_{\rm TH}$. The value of α is around 1.3 in nanometer technologies. When $V_{\rm DD}$ is near-threshold ($V_{\rm DD} \simeq V_{\rm TH}$) or sub-threshold ($V_{\rm DD} < V_{\rm TH}$), the delay can be approximated as exponential functions of $V_{\rm DT}$ [9] as shown in (4). The parameters k_3 and k_4 are fitting coefficients.

$$D = \frac{k_3 V_{\rm DD}}{{V_{\rm DT}}^{\alpha}}. \qquad (V_{\rm DD} \gg V_{\rm TH}) \qquad (3)$$

$$D = k_4 V_{\rm DD} e^{-\frac{V_{\rm DT}}{N_{\rm s}}}. \qquad (V_{\rm DD} \leq V_{\rm TH}) \qquad (4)$$

Since the MEP is found at a point where the delay and energy contours are tangent, the following equation holds at the MEP:

$$\frac{-\dfrac{\partial E_{\rm d}}{\partial V_{\rm TH}} - \dfrac{\partial E_{\rm s}}{\partial V_{\rm TH}}}{\dfrac{\partial E_{\rm d}}{\partial V_{\rm DD}} + \dfrac{\partial E_{\rm s}}{\partial V_{\rm DD}}} = \frac{\dfrac{\partial D}{\partial V_{\rm TH}}}{-\dfrac{\partial D}{\partial V_{\rm DD}}}. \qquad (5)$$

The left and the right of (5) represent gradients of the energy contour and the delay contour at the MEP, respectively. If we denote the slope of a constant delay contour as $s_{\rm d}$, which corresponds to the right of (5), it can be converted into (6) by partially differentiating the circuit delay represented by (3) with respect to $V_{\rm DD}$ and $V_{\rm TH}$, respectively. Note that $s_{\rm d} \approx 1$ when $V_{\rm DD} \leq V_{\rm TH}$ [5]. By partially differentiating (1) and (2) with respect to $V_{\rm DD}$ and $V_{\rm TH}$, respectively, the slope of a constant energy contour $s_{\rm e}$ can be expressed as (7).

$$s_{\rm d} = \frac{\dfrac{\partial D}{\partial V_{\rm TH}}}{-\dfrac{\partial D}{\partial V_{\rm DD}}} = \frac{\alpha V_{\rm DD}}{\alpha V_{\rm DD} - (V_{\rm DD} - V_{\rm TH})} \qquad (6)$$

$$s_{\rm e} = \frac{-\dfrac{\partial E_{\rm d}}{\partial V_{\rm TH}} - \dfrac{\partial E_{\rm s}}{\partial V_{\rm TH}}}{\dfrac{\partial E_{\rm d}}{\partial V_{\rm DD}} + \dfrac{\partial E_{\rm s}}{\partial V_{\rm DD}}} = \frac{E_{\rm s} V_{\rm DD}}{(2E_{\rm d} + E_{\rm s}) N_{\rm s}} \qquad (7)$$

The value of $s_{\rm d}$ can be easily obtained if we know $V_{\rm DD}$ and $V_{\rm TH}$ assigned to the processor. The values of α in (6) and $N_{\rm s}$ in (7) can be obtained through model fitting for the target process technology at a specific temperature. Note that the value of $N_{\rm s}$ is proportional to the absolute temperature. The value of $s_{\rm e}$ can be estimated at runtime by measuring the temperature, dynamic and static energy values. Once $E_{\rm d}$, $E_{\rm s}$ and temperature are measured, the MEP can be easily identified at runtime by a very simple algorithm as shown in Fig. 3. Suppose a pair of $V_{\rm DD}$ and $V_{\rm TH}$ is on the delay contour which satisfies a performance requirement. At the MEP, $s_{\rm d} = s_{\rm e}$ since (5) holds. If $s_{\rm d} > s_{\rm e}$, both of $V_{\rm DD}$ and $V_{\rm TH}$ should be stepped down to get close to the MEP along the delay contour. Contrarily, if $s_{\rm d} < s_{\rm e}$, both of $V_{\rm DD}$ and $V_{\rm TH}$ should be stepped up to the MEP along the delay contour.

B. Related Work and Our Contributions

The algorithm presented in the previous subsection can be easily implemented with a dynamic power sensor, a static power sensor, a temperature sensor, and a critical-path monitor. An MEP tracking processor chip integrating all-digital power sensors, a temperature sensor and a critical-path monitor is presented in [3]. It employs a critical-path monitor to replicate

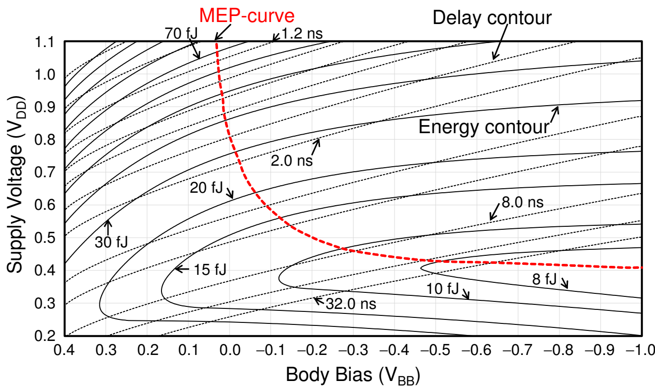

Fig. 2. Energy and delay contours.

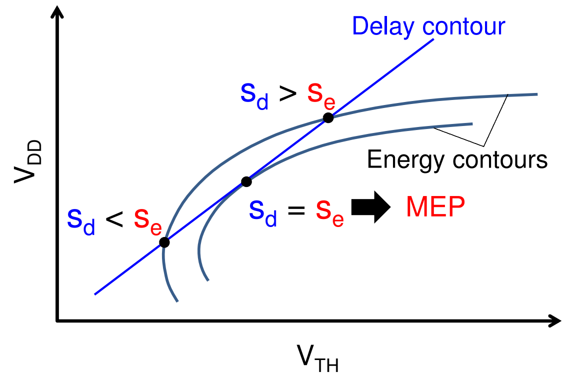

Fig. 3. A concept of minimum energy point tracking algorithm.

the critical-path delay of the processor. With the delay tracking techniques such as presented in [2], [10], [11], the pair of V_{DD} and V_{BB} can be dynamically tuned so that the critical-path delay of the processor is closely tracking the target clock cycle time. A set of performance counters is used to estimate the dynamic power consumption at runtime. In [12], Krishnaswamy et al. reported that the dynamic power consumption can be accurately estimated using performance counters which represent the activity factor of the processor. The maximum error of their approach is less than 5%. A leakage sensor which estimates the static power consumption is proposed in [13]. The leakage sensor is based on a ring oscillator which is driven by sub-threshold leakage current. Since the oscillation frequency is proportional to the leakage current, the leakage current can be estimated by counting the oscillation frequency. V_{DD} and V_{BB} supplied to the processor are also given to the leakage monitor so that the leakage current of the processor can be accurately represented by the leakage current of the monitor. If we supply fixed constant V_{DD} and V_{BB} to the leakage monitor, it can work as a temperature sensor because the leakage current in this case depends on the temperature only [13]. This algorithm used for the MEP tracking is very accurate. The energy loss introduced by this MEP tracking algorithm reported in [3] is less than 3%. However, the fundamental disadvantage of the algorithm is a large latency involved for finding the MEP since it is based on time-consuming power measurement techniques.

More simplified MEP tracking approaches are proposed in [1], [2], [4]. Those approaches are based on the empirical observation that the energy consumption under a delay constraint is minimized when leakage is about half of dynamic power [1], [14], [15]. The algorithm can be quite similar to what presented in Fig. 3. When the leakage is half of the dynamic power, this means that the current operating point is MEP. If the leakage is less than half of the dynamic power, both of V_{DD} and V_{TH} should be stepped down along the delay contour to get close to the MEP. Contrarily, if the leakage is more than half of the dynamic power, both of V_{DD} and V_{TH} should be stepped up along the delay contour toward the MEP. The approaches presented above are largely simplified compared with the method presented in [3]. However, those approaches are still very slow since they still rely on the time-consuming power measurement for finding the MEP. The power measurement technique proposed in [4] is based on the fact that DC-DC converter frequency is proportional to its load current. Therefore, by counting the DC-DC converter frequency, the power consumption of the target processor can be measured. The DC-DC converter frequency is generally limited to a few hundred kHz and if the processor is running at a low power mode, the frequency is lowered to several kHz. If we average out the DC-DC converter frequency to measure the average power consumption of the processor, it needs several tens of cycles in the DC-DC converter frequency, which corresponds to a few hundreds of microseconds.

Unlike the approaches presented above, our method can find the MEP in a real-time manner by skipping the time-consuming power measurement when finding the MEP. This can be done using a predetermined voltage setting strategy which is optimized at a boot phase for a given processor chip.

TABLE I
SITUATIONS FOR CHANGING THE OPERATING POINT (V_{DD} AND V_{TH})

Change of Situation	Responsiveness	Energy Savings
Speed requirement	Order of μs	Largest
Chip temperature	Order of ms	Medium
Process variation	Order of seconds	Large
Aging status	Order of hours	Large
Activity factor	Order of μs	Small

III. Real-time MEP Tracking Algorithm

A. Situations to Update the Voltage Setting

If we look at situations for changing the voltage setting, we can categorize them into two categories in terms of response time needed and impact on the energy reduction as shown in Table I. As presented in Section I, the MEP is heavily dependent on the operating condition such as chip temperature, activity factor, process variation, aging status, and speed requirement for the processor. When the speed requirement for the processor has changed, the operating point should be immediately shifted to the MEP. Otherwise, we may lose the chance for reducing the energy consumption. Since the largest energy saving can be obtained in this situation, we should fully exploit this short chance by quickly finding the MEP and shifting V_{DD} and V_{TH} toward the MEP. Contrarily, since the chip temperature changes slowly, we can use several milliseconds for finding the MEP. The MEPs of different chips may be different from each other due to the chip-to-chip process variation. The operating point (i.e. V_{DD} and V_{TH}) of a chip can be accurately shifted to its own MEP by running the traditional MEP tracking algorithm [3] only once at the booting phase. The aging status of transistors changes very slowly and the time to find the MEP is negligible compared to the aging speed. When a task running on the processor has changed, the activity factor of the processor may change, where the activity factor is the average probability that the internal gate transitions from 0 to 1 in a clock cycle [7]. This change is, however, not very large typically and the MEP also does not change widely. Therefore, we assume that the activity factor of the processor is always at a typical value of 10% [7] in this paper. As a conclusion, only the change of the speed requirement needs the real-time response and the responses to the other factors such as changes of the temperature and aging status can spend longer time on the order of milliseconds.

B. Outline of Real-Time MEP Tracking Algorithm

Based on the situations summarized in Table I, this paper proposes a real-time MEP tracking method using a predetermined MEP-curve which is characterized at a boot phase for a given processor chip. The key idea of the method is first separating the response to the change of the speed requirement and that to the others such as chip-to-chip process variation and temperature change. Then, upon the change of the speed requirement, which needs the real-time response, this method

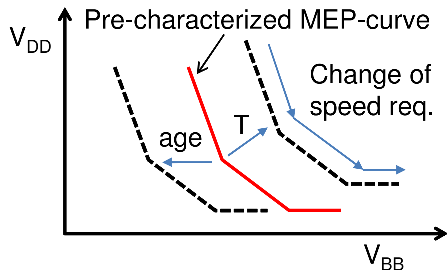

Fig. 4. A concept of real-time minimum energy point tracking.

skips the time-consuming power measurement and just moves the operating point along the pre-characterized MEP-curve when finding the MEP as shown in Fig. 4. This can be done using the MEP-curve which is characterized at the boot phase for a given processor chip. For the response to the temperature change, where a slow response speed on the order of milliseconds is allowed, we shift the entire MEP-curve along the delay contour. The shift amount is calculated statically based on the process parameter of the target technology. If the chip aging status changes, both the delay contours and the entire MEP-curve are shifted horizontally as shown in Fig. 4 since the device aging corresponds to the V_{TH} shift.

C. Pre-Characterization of MEP-Curve

Although we call it as a curve, the MEP-curve we use in the real-time MEP tracking algorithm is the concatenation of three straight lines as shown in Fig. 5. It shows the actual MEP-curves and corresponding linear approximation for three different process conditions for pMOS and nMOS, slow-slow (SS), typical-typical (TT), and fast-fast (FF). The idea of the linear approximation for the MEP-curve arises from the observation that MEP-curve in the super-threshold region is almost vertical and that in the sub-threshold region is horizontal in a V_{DD}-V_{BB} two-dimensional space [5]. Based on the observation, our method first characterizes the MEP-curve as the concatenated linear model for a given processor chip. For adapting to the chip-to-chip process variation, the traditional MEP tracking algorithm is executed at the booting phase to find the linear approximation model of the MEP-curve for each chip. The linear approximation is based on four-point characterization. The characterization algorithm starts from identifying the outer two MEPs. In this work, we use $V_{\mathrm{DD}} = 1.0$ and $V_{\mathrm{BB}} = -0.9$ as the two outer MEPs as shown in Fig. 5. We can find the MEP at $V_{\mathrm{DD}} = 1.0$ using the traditional MEP tracking algorithm by sweeping V_{BB} with the fixed V_{DD} at 1.0 V. Every time the V_{BB} is shifted, we measure the dynamic energy (E_{d}) and the static energy (E_{s}) separately to calculate the s_{e} value in (7) at runtime using power sensors. This power measurement and calculation of s_{e} and s_{d} are repeated until the MEP is found. At the MEP, $s_{\mathrm{d}} = s_{\mathrm{e}}$ as presented in Section II. The MEP for $V_{\mathrm{BB}} = -0.9$ is similarly found by sweeping V_{DD} with the fixed V_{BB}. Once the two outer MEPs are found, the next step is identifying the two inner MEPs. The inner two MEPs are specified at a 2-to-1 distance from the outer two MEPs along the x and y axes, respectively as shown in Fig. 5. Finally, we calculate the slopes of the three straight lines based on the four points, which are used when the performance requirement changes. With this linear approximation, once the request of the performance change has been sent from a real-time OS, for example, we can quickly track the MEP just by shifting the operating point along the straight lines without any power measurement and calculations for finding the MEP.

D. Strategy for Updating the Coordinate of MEP-Curve

For responding to the chip aging, the method periodically calibrates the coordinate of the linearized MEP-curve at runtime in a specific frequency. The frequency of calibrating the coordinate can be very low for example at 1Hz or lower since the time constants for the aging speed is on the order of minutes to hours. At each update, the coordinate of the linearized MEP model is shifted horizontally in the V_{DD}-V_{BB} two-dimensional space since the chip aging corresponds to a horizontal shift as shown in Fig. 6. The shift amount of the coordinate can be identified by the amount of the V_{BB} shift when achieving a specific clock frequency, which corresponds to the shift amount of the delay contours. The change of the chip temperature is also corresponding to the V_{TH} shift since the magnitude of the V_{TH} decreases nearly linearly with temperature and approximated as $V_{\mathrm{TH}} = V_{\mathrm{TH0}} - k_{vt}T$, where k_{vt} is typically about 1 to 2 mV/K [7]. The value of

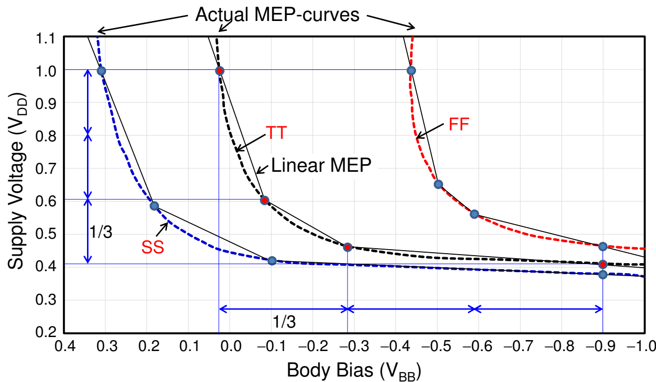

Fig. 5. Linear approximation of MEP-curve.

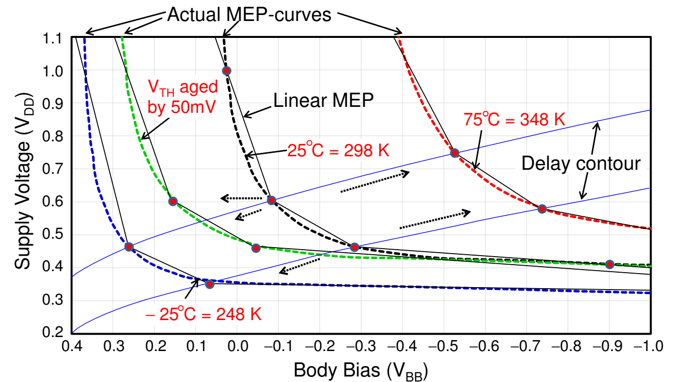

Fig. 6. Linear approximation of MEP-curve for different temperature.

N_s is also proportional to the absolute temperature. However, unlike the chip aging, the delay of gates manufactured with the advanced process technology is almost unchanged against the chip temperature change. This is because the change of the electron mobility in MOSFETs and the V_{TH} shift due to the temperature change are canceled out to each other. In this case, the coordinate of the MEP-curve shifts along the constant delay contour in the near-threshold region. The shift amount of the MEP-curve along the delay contour can be estimated by the process parameter of the target technology.

IV. Experimental Results

To evaluate the accuracy of the linearized MEP-curve model, we design a 50-stage fanout-4 inverter chain to reflect the behavior of a microprocessor pipeline using a commercial 55 nm deeply depleted channel (DDC) CMOS process technology. The threshold voltage (V_{TH}) in this process can be almost linearly controlled by tuning the body bias (V_{BB}). The dynamic and static power consumption values are obtained using gate-level circuit simulation. The activity factor is set to a 10% throughout this simulation.

We assume that the MEP-curve is characterized on a chip-by-chip basis at a boot phase using the traditional MEP tracking algorithm presented in Section II. The linearized MEP model is generated by the four-point characterization method presented in the previous section. Figure 7 shows the results of the linearized MEP-curve model for typical-typical (TT) process condition. The four-point MEPs found at the beginning already include the error since the slope models of s_e and s_d do not accurately express the actual slopes of energy and delay contours. If they are accurate, the coordinate of the topmost MEP of the four-point MEPs is (1.0, 0.03) while the estimated coordinate is (1.0, 0.07). Therefore, the linearized MEP-curve is slightly shifted left as shown in Fig. 7 compared to the ideal position shown in Fig. 5. Although the linearized model is shifted left, the worst-case energy loss introduced by this linear approximation is only 0.3%, which occurs at the highest performance point in this result. The reason why the energy loss introduced by the displacement of the operating point is very small is because the energy consumption curve is fairly flat around the MEP.

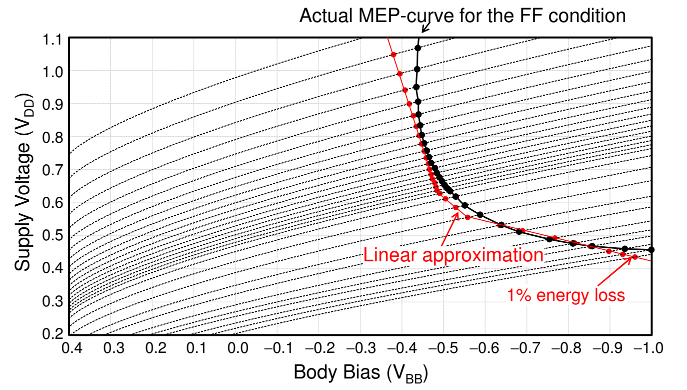

Fig. 8. Results of linearized MEP model for a chip of an FF condition.

Figure 8 shows the results for fast-fast (FF) process condition. When we execute the four-point characterization for a given chip, regardless of the manufacturing process condition of the chip, we use the TT values for the parameters such as α used in (6) and N_s in (7) since the parameter values for the chip are generally unknown in advance. This may cause an error of the linear approximation. However, the results in Fig. 8 show very good accuracy of the linear approximation model. The worst case energy loss introduced by the linearized MEP-curve model is only 1%. This is because the location of the MEP is mainly determined by the E_s to E_d ratio which is measured at runtime and values of the α and N_s do not have a big impact on the location of the MEP. Figure 9 shows the results for slow-slow (SS) process condition. This linearized MEP-curve is also generated with the TT values for α and N_s when finding the four initial points of the MEPs. The worst case energy loss introduced in this case is 0.6%.

For responding to the temperature change, we extract the V_{TH} shift amount based on the process parameter of the target technology. As presented in the previous section, the V_{TH} shift due to the temperature change is $k_{vt}T$, where T is absolute temperature and k_{vt} is typically about 1 to 2 mV/K. For example, when the absolute temperature increases from 298K to 348K, the V_{TH} value is reduced by a factor of 0.88 in the target process technology. Contrarily, the N_s value increases by a factor of 1.17 ($= 348/298$). In this case, the value of

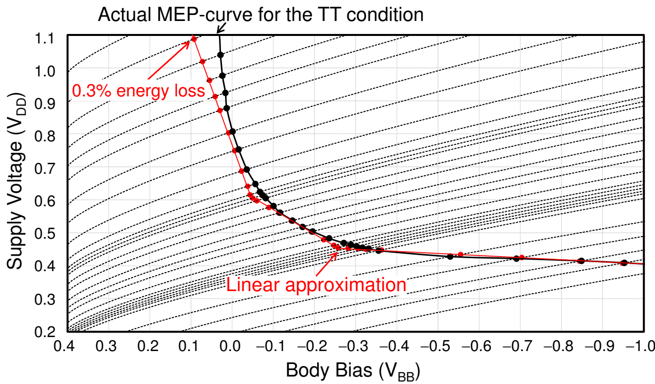

Fig. 7. Results of linearized MEP model for a chip of a TT condition.

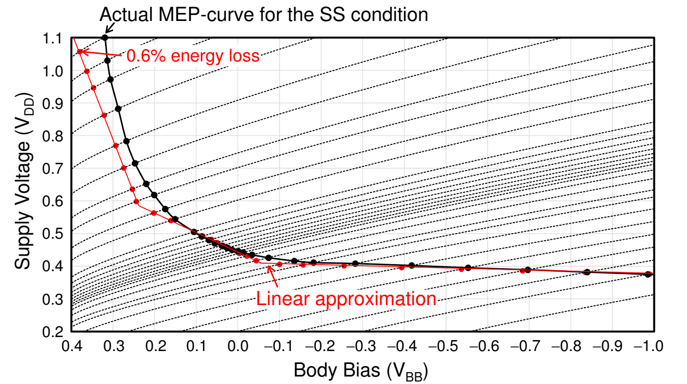

Fig. 9. Results of linearized MEP model for a chip of a SS condition.

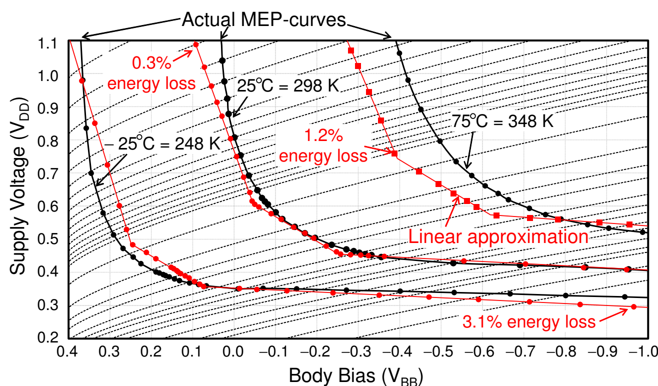

Fig. 10. Results of linearized MEP model for 75 and −25 degree C.

$V_{\mathrm{TH}}/N_{\mathrm{s}}$ which exponentially affects the leakage current as shown in (2) is reduced by a factor of 0.75, which is equivalent to 82 mV when converted to a change in V_{TH}. The slopes of the delay contours passing through the two inner MEPs are 1.66 and 1.4, respectively. Since the two inner MEPs on the linearized MEP-curve move along the two delay contours when the temperature changes, the increase of the temperature by 50 degree C corresponds to the vertical shift amounts of 136 mV and 115 mV, respectively. Note that the vertical shift means the V_{DD} shift of the MEP-curve when it moves along the delay contours. Similarly, the change of the absolute temperature from 298K to 248K is corresponding to the V_{TH} shift of 82 mV. In this way, the shift amount of the MEP-curve along the delay contour can be calculated analytically based on the process parameters of the target technology. Figure 10 shows the results for 75 and −25 degree C. As can be seen from the figure, the linearized model for 75 degree C is largely displaced from the actual MEP-curve. However, surprisingly, the worst case energy loss introduced by the linearized model for the 75 degree C is only 1.2%. This is because the energy and delay contours are almost in parallel around the MEP. Therefore, even if the distance between the MEPs on the linear model and the actual MEP-curve is considerably large, the corresponding energy difference is limited.

V. Conclusion

This paper proposes a real-time MEP tracking method using a pre-characterized linear MEP model which is generated for a given chip at a boot phase. The key of the method is first separating the response to the change of the required performance and that to the others such as process variation and temperature change. Then, for the change of the required performance, which needs the real-time response, the method skips the time-consuming power measurement and any calculations to identify the MEP when finding the MEP. The experimental results obtained using a 50-stage fanout-4 inverter chain designed with a commercial 55 nm DDC process demonstrate that the method using the linear MEP model accurately tracks the MEP even if the operating conditions including process variation, chip aging status, temperature, and performance requirement widely vary. The worst case energy loss introduced by the linearized MEP model is only 3.1%. More importantly, this method can quickly track the MEP when the performance requirement changes as it does not need any time-consuming power measurement or calculations for identifying the MEP. Our future work will focus on improving the model accuracy and evaluating the method using actual microprocessor chips. Implementing more specific software programs which characterize the MEP-curve as a linear model at boot time and shift the operating point appropriately according to the required performance is also our future work.

References

[1] V. von Kaenel, M. Pardoen, E. Dijkstra, and E. Vittoz, “Automatic Adjustment of Threshold and Supply Voltages for Minimum Power Consumption in CMOS Digital Circuits,” in *Proceedings of IEEE Symposium on Low Power Electronics*, October 1994, pp. 78–79.

[2] M. Nomura, Y. Ikenaga, K. Takeda, Y. Nakazawa, Y. Aimoto, and Y. Hagihara, “Delay and Power Monitoring Schemes for Minimizing Power Consumption by Means of Supply and Threshold Voltage Control in Active and Standby Modes,” *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 805–814, April 2006.

[3] S. Hokimoto, J. Shiomi, T. Ishihara, and H. Onodera, “All-Digital On-Chip Heterogeneous Sensors for Tracking the Minimum Energy Point of Processors,” in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, 3 2018, pp. 128–133.

[4] J. Jeongsup Lee, Y. Zhang, Q. Dong, W. Lim, M. Saligane, Y. Kim, S. Jeong, J. Lim, M. Yasuda, S. Miyoshi, M. Kawaminami, D. Blaauw, and D. Sylvester, “A 6.4pJ/Cycle Self-Tuning Cortex-M0 IoT Processor Based on Leakage-Ratio Measurement for Energy-Optimal Operation Across Wide-Range PVT Variation,” in *IEEE International of Solid-State Circuits Conference*, Feb 2019, pp. 314–316.

[5] T. Takeshita, T. Ishihara, and H. Onodera, “Guidelines for Effective and Simplified Dynamic Supply and Threshold Voltage Scaling,” in *Proceedings of International Symposium on VLSI Design, Automation and Test*, 4 2016, pp. 1–4.

[6] A. K. Coskun, T. S. Rosing, and K. C. Gross, “Proactive Temperature Management in MPSoCs,” in *Proceedings of International Symposium on Low-Power Electronics and Design*, August 2008, pp. 165–170.

[7] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*. Addison Wesley, 4th edition, 2010.

[8] T. Sakurai and R. Newton, “Alpha-Power Law MOSFET Model and Its Applications to CMOS Inverter Delay and Other Formulas,” *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, 4 1990.

[9] S. Keller, D. Harris, and A. Martin, “A Compact Transregional Model for Digital CMOS Circuits Operating Near Threshold,” *IEEE Transactions on Very Large Scale Integration Systems*, vol. 22, no. 10, pp. 2041–2053, 10 2014.

[10] J. Park and J. Abraham, “A Fast, Accurate and Simple Critical Path Monitor for Improving Energy-delay Product in DVS Systems,” in *Proceedings of IEEE/ACM International Symposium on Low-Power Electronics and Design*, 8 2011, pp. 391–396.

[11] T. Kuroda, K. Suzuki, S. Mita, T. Fujita, F. Yamane, F. Sano, A. Chiba, Y. Watanabe, K. Matsuda, T. Maeda, T. Sakurai, and T. Furuyama, “Variable Supply-Voltage Scheme for Low-Power High-Speed CMOS Digital Design,” *IEEE Journal of Solid-State Circuits*, vol. 33, no. 3, pp. 454–462, 3 1998.

[12] V. Krishnaswamy, J. Brooks, G. Konstadinidis, C. McAllister, H. Pham, S. Turullols, J. Shin, Y. Yifan, and Z. Haowei, “Fine-grained Adaptive Power Management of the SPARC M7 Processor,” in *IEEE International of Solid- State Circuits Conference*, Feb 2015, pp. 1–3.

[13] M. Islam, J. Shiomi, T. Ishihara, and H. Onodera, “Wide-Supply-Range All-Digital Leakage Variation Sensor for On-Chip Process and Temperature Monitoring,” *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2475–2490, Nov 2015.

[14] K. Nose and T. Sakurai, “Optimization of V_{DD} and V_{TH} for Low-Power and High Speed Applications,” in *Proceedings of Asia and South Pacific Design Automation Conference*, 2000, pp. 469–474.

[15] D. Markovic, M. Horowitz, and R. Brodersen, “Methods for True Energy-Performance Optimization,” *IEEE Journal of Solid-State Circuits*, vol. 39, no. 8, pp. 1282–1293, August 2004.