Towards Reliable and Ultra Low Energy Digital Circuits with Sub/Near Threshold Supply Voltage

Yu Pu[1,2,3] José Pineda de Gyvez [1,2] Henk Corporaal[1] Yajun Ha[3]

Technische Universiteit Eindhoven[1] NXP Research Eindhoven[2] National University of Singapore[3]

{y.pu, h.corporaal}@tue.nl jose.pineda.de.gyvez@nxp.com elehy@nus.edu.sg

Abstract – A sub/near threshold supply voltage can be used for digital applications to achieve ultra-low energy consumption, but it brings along big challenges to reach the required throughput and good tolerance of process variations. In this paper a series of techniques are proposed to handle the above two issues. Parallelism is used to overcome throughput degradation. A configurable V_T balancer is introduced to balance V_T mismatch in the sub-threshold at the process corners. We also show that utilizing V_T mismatch between parallelized transistors can improve transistor drivability in the sub-threshold and help reducing layout area. An autonomous local V_{DD} adjustment method based on the crystal ball flip-flop characterization method [2][3] and digital V_{DD} actuators is proposed to reduce design margin. Finally, a novel multi-standard JPEG encoder is fully implemented in RTL to demonstrate these ideas. Our simulation results show that, the energy of the DCT and Quantization Engine can be reduced by 9× at a 0.4V supply compared to the nominal 1.2V supply in a CMOS 65nm Standard V_T process.

Keywords: Sub-threshold circuits, variability, reliablity

I. Introduction

While the capacity density of battery doubles every ten years, the number of transistors integrated in a circuit doubles approximately every two years. The ever-increasing transistor density not only greatly challenges the limited battery life, but also the thermal issue. Exploring design methodology for low energy, "green" digital circuits is thus of very great importance. An effective means to achieve these goals is to scale the supply voltage V_{DD} along with the operating frequency. As V_{DD} scales, not only does the dynamic energy reduce quadratically, the leakage current also reduces super-linearly. As a result, the total energy consumption can be considerably reduced. In addition, V_{DD} scaling mitigates the transient current hence lowering the notorious ground bounce noise. This also helps to improve the performance of the sensitive analog circuit on the chip, such as the delay-lock loop (DLL), which is crucial for the functioning of large digital circuits.

Although the CMOS digital gate can work seamlessly from full V_{DD} to well below threshold voltage V_T, the design rules provided by semiconductor vendors normally set 2/3 of the full V_{DD} as the lower bound for V_{DD} scaling. Taking the Samsung's DVFS Design Technology [1] and the TSMC design rule as examples, the constraint of V_{DD} for digital circuits designed in 65nm Standard V_T Process is between 0.8V~1.2V. The reasoning behind this constraint is essentially twofold. First, as V_{DD} scales, the driving capability of transistors accordingly reduces. Because most consumer electronic applications need operating frequencies in the range of tens of MHz, which might not be fulfilled with aggressive V_{DD} scaling, 2/3 full V_{DD} is tested to be a safe lower bound. Second, the digital circuits become particularly sensitive to process variations when V_{DD} scales below 2/3 full V_{DD}. Process variations are likely to cause malfunctioning, and both the timing yield and functional yield tremendously decrease. As a result, 2/3 full V_{DD} is chosen to maintain adequate margin to prevent high yield loss and to keep quality to the industry standard.

However, it is an undeniable fact that setting the 2/3 full V_{DD} as the lower bound has limited any further energy reduction we could gain from utilizing voltage scaling. In our research, we are particularly interested in exploring a design methodology that allows V_{DD} to go below the 2/3 full V_{DD} and even to the sub/near threshold region without causing the aforementioned performance and yield problems. The features of this work include:

1) A sub/near threshold V_{DD} has been used to achieve a near maximum energy per operation reduction.
2) Parallelism has been used to overcome throughput degradation.
3) For each parallel unit, an online autonomous supply voltage adjustment methodology which uses its own digital voltage actuators with the crystal ball flip-flop characterization method [2][3] has been proposed. A large safety margin reservation is avoided.
4) A configurable threshold voltage balancer which helps mitigating the V_T mismatch between pMOS and nMOS transistors at process corners in the sub/near threshold supply mode, has been proposed to increase yield.
5) An interesting approach to improve sub-threshold driving capability by exploiting V_T mismatch between parallelized transistors has been proposed.
6) A baseline JPEG image encoder has been fully implemented in RTL to demonstrate these ideas. The JPEG encoder architecture is parallelized in an efficient manner to minimize the associated area overhead.

The remainder of this paper is organized as follows. Section II briefly presents the speed, energy models used for the first-order analysis in this paper. In Section III, we discuss the possibility to use massive parallelism to compensate the throughput degradation due to aggressive

V_{DD} scaling. Then, Section IV, V, VI introduce the approach of utilization transistor mismatch for an increased drivability, the configurable V_T balancer and autonomous local V_{DD} adjustment scheme respectively. Section VII shows a baseline multi-standard JPEG encoder as case study. Finally, section VIII draws the conclusion of this work.

II. Speed, Energy and Optimal EOP Point

Dynamic energy and leakage energy are the two major sources of energy dissipation in CMOS circuits. The dynamic energy per one operation cycle is

$$E_{dynamic} = \alpha C V_{DD}^2 \quad (1)$$

where α is the average switching activity factor of all the output nodes, C is the total capacitance of all the output nodes, V_{DD} is the supply voltage.

In nanometer technology the off-state leakage current I_l of a digital block consists of the sub-threshold leakage I_{sub} and the gate-oxide tunneling leakage I_{gt},

$$I_l = I_{sub} + I_{gt} \quad (2)$$

The sub-threshold leakage current is modeled as

$$I_{sub} = I_s e^{\frac{(-V_T + \eta V_{DD} - \gamma V_{SB})}{nU}} (1 - e^{-\frac{V_{DD}}{U}}) \quad (3)$$

where n is the sub-threshold swing factor, η is the DIBL coefficient, γ the linearized body effect factor. U is the so-called thermal voltage kT/q which is around 26mV at room temperature. I_s is the zero-threshold leakage current. The gate-oxide tunneling leakage is modeled as [4]:

$$I_{gt} = A(V_{DD}/t_{ox})^2 \cdot e^{(\frac{-B(1-(1-V_{DD}/\phi_{ox})^{3/2})}{V_{DD}/t_{ox}})} \quad (4)$$

where A and B are physical parameters that are intrinsic to the process technology, Φ_{ox} is the barrier height for the tunneling particle (electron or hole), and t_{ox} is the oxide thickness. Equations (2), (3) and (4) clearly indicate a super-linear decrease of leakage current due to V_{DD} scaling.

Therefore, the total energy per operation EOP of a circuit can be obtained as

$$EOP = \alpha C V_{DD}^2 + I_l V_{DD} T_c \quad (5)$$

where T_c is the operation cycle time.

We assume C_{load} the load capacitance of a FO4 inverter and I_d the average drive current of a FO4 inverter. L_d is the logic depth which represents how many inverters are chained to mimic the critical path delay. The inverter delay T_g can be derived as

$$T_g = C_{load} V_{DD} / I_d \quad (6)$$

thus the critical path delay is

$$T_{cp} = L_d T_g \quad (7)$$

V_{DD} scaling lowers power dissipation, but increases the operation cycle time T_c, there exists an optimal V_{DD} point where the Energy per Operation (EOP) is minimal. Taking a standard-cell based CMOS circuit in a 65nm Low Power Standard V_T (SV_T) technology with an area of 6mm^2 as a reference, the values of typical circuit parameters are derived. The average switching activity factor α is 0.12, the effective switching capacitance for the entire block is 4.9nF, the nominal V_{DD} is 1.2V, average V_T of pMOS and nMOS is 0.41V, and off-state leakage I_l 648μA. I_d in equation (6) is obtained from transistor-level simulation. This baseline processor has a L_d =24 and we assume that it is running at its maximum speed, i.e., T_c=T_{cp}. Fig.1(a) shows how the dynamic, leakage and total energy of the baseline processor vary when V_{DD} scales. The simulated optimal V_{DD} point V_{opt} is annotated. Since nowadays high V_T process is a popular option for low power digital design, a simulation has also been carried for the same block implemented with a High V_T (HV_T) process. Fig.1(b) compares the total energy per operation for SV_T and HV_T processes. The behavior of these curves is close to one another.

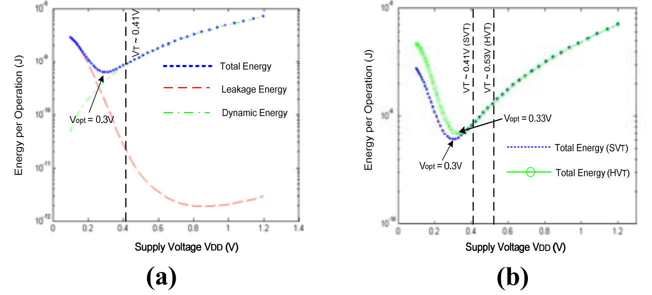

Fig. 1. (a) Dynamic/Leakage/Total energy per operation and the optimal V_{DD} (SV_TProcess) (b) Total energy per operation and the optimal V_{DD} points for SV_T and HV_T Process

As indicated by Fig.1, the optimal operating supply voltage V_{opt} is in the sub-threshold region, as has been observed in [5]-[7]. Lowering further V_{DD} below V_{opt} does not render any additional energy benefits. This V_{opt} is larger than 0.25V, where the static gates in the existing super-threshold SV_T digital library operate without problems according to our transistor-level Monte-Carlo simulations. Therefore, these gates can be reused safely in the sub-threshold design. In addition, we observe that using HV_T process consumes 13% higher EOP compared to SV_T process. In the following analysis, SV_T process is used for investigation and simulations.

III. Parallelism for Fixed Throughput

As V_{DD} scales, the circuit throughput degrades. To maintain a fixed throughput, we use parallel processing units. We assume that the tasks of individual units are independent, meaning that no performance penalty due to data or control dependencies is incurred from parallelism. This is a realistic assumption that suits well for applications such as sound/graphic, streaming processing, etc. Ideally, for a fixed V_{DD}, the degree of parallelism does not affect the EOP whereas a larger throughput can be obtained simply by using more parallelized units. However, in reality the muxing/demuxing circuit also contributes to the EOP. To better take into account the overhead brought by the muxing/demuxing circuits, the area and timing overhead are approximated in equations (8) and (9), respectively,

$$Area = Area_{baseline} \times M^{\rho} \quad (8)$$

$$T_{muxing\text{-}demuxing} = log_2 M \times FO4 \quad (9)$$

where M is the associated degree of parallelism. ρ is the area growth factor, which indicates that the circuit area grows super-linearly with M. In our simulation, we choose ρ=1.1.

Refering to equation (5), the area overhead affects C and I_l, the timing overhead affects T_C.

Fig. 3 shows the normalized EOP for different V_{DD}, with the same throughput as that of the baseline processor operating at the nominal 1.2V supply voltage. The needed degrees of parallelism for a few V_{DD} points are annotated in the plot. As can be seen, we could obtained 5×,4×,3× EOP reduction when V_{DD} is at 0.4V, 0.5V, 0.6V, respectively. At first glance it is horrible to see the associated 245, 31and 12 parallel widths, which implies a huge silicon area. Additionally, the larger the circuit size, the more likely it will contain defects and it will fail to function correctly. As a result, the fabrication line limits the size of the largest chip producible with commercially viable yields. However, it should be noted that in the analysis we assume the baseline processor is operating at its maximum speed, which means close to 300MHz frequency. For some consumer electronic applications which only need a few or tens of MHz, the associated parallel width can become much smaller and thus more affordable. For applications that only require KHz range frequencies, such as sensor networks, biomedical instrumentations and audio processors, operating at the V_{opt} is possible and there is even no need to use parallel paths.

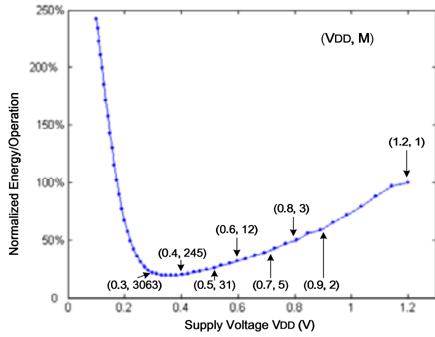

Fig. 3. Normalized EOP for different V_{DD}

IV. Improve Drivability by Exploiting V_T Mismatch between Parallelized Transistors

The intra-die V_T variation of a single transistor has been modeled in [13]

$$\sigma V_T = \frac{A_{\Delta V_T}}{\sqrt{WL}} \tag{10}$$

where $A_{\Delta V_T}$ is a technology conversion constant (in mVμm), WL is the transistor's active area. $A_{\Delta V_T}$ is about 10mVμm for nMOS and pMOS transistors in the sub-threshold region in a 65nm CMOS process. Common sense tells that V_T mismatch is always catastrophic to circuit design. In the later part of this section, we will propose an interesting approach to improve sub/near threshold current drivability by exploiting the V_T mismatch between parallel transistors. Our approach is based on the theoretical analysis and simulation results that V_T mismatch between parallelized transistors always results in an increased driving current in the sub-threshold.

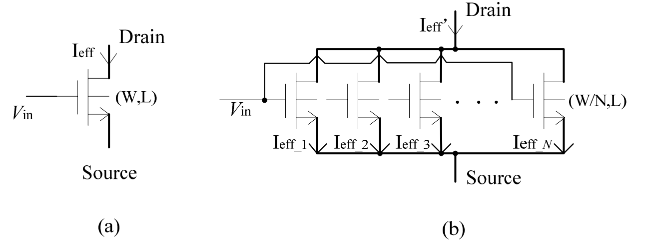

Fig. 4. (a) nMOS transistor with aspect ratio (W, L) (b) N-parallelized nMOS transistors with aspect ratio (W/N, L)

A. Theory

Suppose μV_T, σV_T are the mean and standard deviation of V_T for an nMOS transistor in Fig. 4.(a). According to statistic theory and considering V_T variation, the mean of I_{eff} (μ_{Ieff}) is:

$$\mu_{Ieff} = I_{0n} e^{\frac{(V_{GS}-\mu V_T+\eta V_{DS}-\gamma V_{SB})}{nU}+\frac{\sigma V_T^2}{2}} (1-e^{-\frac{V_{DS}}{U}}) \tag{11}$$

Suppose the transistor is equally divided as N-parallelized nMOS transistors (see Fig. 4.(b)). For every individual small transistor we have

$$\mu V_{T1} = \mu V_{T2} = \cdots = \mu V_{Tn} = \mu V_T \tag{12}$$

$$\sigma V_{T1} = \sigma V_{T2} = \cdots = \sigma V_{Tn} = \sigma V_T \tag{13}$$

The mean value of the total sub-threshold current ($\mu_{Ieff'}$) in Fig. 4(b) is calculated in equation (14),

$$\mu_{Ieff}' = \sum_{i=1}^{N} I_{eff_i}$$

$$= I_{0n} e^{\frac{(V_{GS}-\mu V_T+\eta V_{DS}-\gamma V_{SB})}{nU}+\frac{\sigma V_T'^2}{2}} (1-e^{-\frac{V_{DS}}{U}}) \tag{14}$$

Based on equation (10), we have

$$\sigma V_{T1} > \sigma V_T \tag{15}$$

By comparing equations (11) and (14), we can obtain

$$\mu_{Ieff}' > \mu_{Ieff} \tag{16}$$

As can be seen, dividing a large transistor into smaller parallelized transistors helps to increase the sub-threshold current due to larger V_T mismatch. Therefore, statistically a larger drivability can be accomplished without incurring any additional overhead such as increasing V_{DD} or transistor size.

B. Simulation Results

The Monte-Carlo simulation results have confirmed the effectiveness of this approach. Assume that a Standard V_T (SV_T) nMOS transistor with aspect ratio W/L = 0.96μm/0.065μm is divided as N-transistors (N=1,2,3,4,6,8). Its gate voltage V_{in} and drain-to-source voltage V_{DS} are set as 200mV. The simulated mean and standard deviation values of the driving current I_{eff} (μI_{eff}, σI_{eff}) are listed in Table I. As seen, the larger the N, the larger the V_T mismatch, consequently the larger sub-threshold driving current.

C. Possible Applications

This interesting attribute can be applied to the pass-transistor based logics, such as transmission gate, multiplexer and power-switch. Because the sub-threshold drivability is increased, the necessary transistor size can be reduced. In addition, since a wide transistor is thus divided into small ones, multiple-finger structured layout is therefore

allowed. As a result, the layout of very huge transistors, such as power switch, becomes much more compact which may reduce silicon area considerably.

Table I
MEAN AND STANDARD DEVIATION OF DRIVING CURRENT

N	$\mu I_{eff}(nA)$	$\sigma I_{eff}(nA)$
1	*5.3904*	*2.4949*
2	*5.9910*	*3.0229*
3	*7.6670*	*4.2326*
4	*9.3239*	*4.8853*
6	*12.9342*	*6.2544*
8	*13.3155*	*7.3793*

V. CONFIGURABLE V_T BALANCER

Our previous work [8] has already shown that the V_T variation is the dominant component for sub-threshold current variation due to its exponential correlation to the current. A novel V_T balancer which uses only one bulk line to balance the V_T of pMOS and nMOS transistors has also been introduced. In this paper, a configurable V_T balancer has been extended from that work (see Fig. 5).

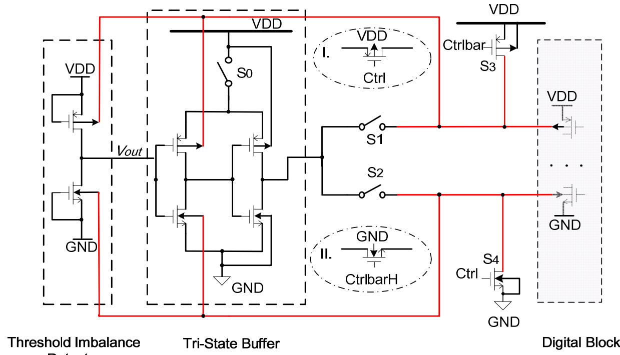

Fig. 5. Proposed configurable V_T Balancer

When the processor is set in super-threshold mode, the power switch S_0 is configured as *off*, the power switch transistor S_3,S_4 are *on*, and S_1,S_2 are *off*, so the bulk of pMOS transistors is connected to V_{DD}, and the bulk of nMOS transistors is connected to G_{ND}. When the processor is configured to be in the sub/near threshold mode, the V_T balancer starts to function. In this mode, the power switch S_0 is set as *on*, S_1, S_2 are *on*, and S_3, S_4 are *off*. Therefore, the buffer's output voltage passes through S_1, S_2 to supply the bulk of the logic gates. The fluctuation of the signal V_{out}, which is generated from a process-corner V_T imbalance detector, is thus detected and amplified by the tri-state buffer. Therefore, the buffer's output voltage passes through S_1, S_2 to supply the bulk of the logic gates. This output is also fed back to the bulk of the V_T balancing detector to force pMOS/nMOS V_T balancing.

The design of the power switch transistor S_0, $S_1$and S_2 should be careful as their equivalent on-resistance R_{on} must be small enough to avoid large voltage drop across the transistors. Small R_{on} also improves the configuration setup time. If pMOS transistors are used as the power switches (see scheme I in Fig. 5), note that the $|V_{GS}|$ for switches S_1 and S_2 is only around $V_{DD}/2$ ($V_{DD} < V_T$). As a result, the R_{on} of a unit pMOS transistor becomes many orders bigger than in the super-threshold mode, so they must be upsized largely, which would introduce a huge amount of area as well as configuration energy overhead. Instead, we use much smaller nMOS transistors with their gate voltage boosted (see scheme II in Fig. 5). As nMOS transistor has better on-current characteristic than pMOS transistor and the boosted gate voltage over-drives the transistor, the R_{on} and transistor area can be greatly reduced. The potential drop across a transistor is also avoided. The boosted gate voltage can be obtained either from other high voltage domain in the SoC or from the periphery I/O power rails. Fig. 6 shows the layout of the proposed configurable V_T balancer in a 65nm CMOS process. The total layout area is $25\times30\mu m^2$. As shown, all the power switches are divided as small transistors by using multiple-finger structured layout style, which not only reduces silicon area considerably, but also enables a largely increased sub-threshold drivability from the final stage of the tri-state buffer.

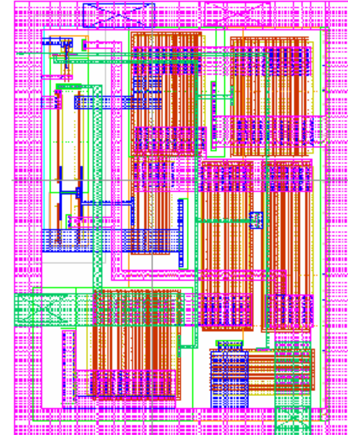

Fig. 6. Layout of configurable V_T Balancer with Multiple-Finger Structured Power Switch in a 65nm CMOS

Fig. 7 shows the Monte-Carlo simulated transition time for an inverter with aspect ratio of Wp/Wn=1.1µm/0.4µm to drive a capacitive load of 5fF at V_{DD}=400mV in a CMOS 65nm SV_T process technology. Compared to the conventional design, the standard deviation σ is reduced by 4.7× and the σ/µ is reduced by 3.6× when the proposed configurable V_T balancer is used.

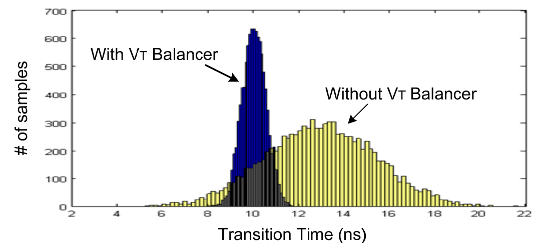

Fig. 7. Transition time for an inverter from Monte-Carlo simulation (Wp/Wn=1.1µm/0.4µm, C_{load}=5fF)

VI. Online V_{DD} Adjustment for High Parametric Yield

The V_T balancer has shown to be effective to reduce the impact of D2D (Die-to-Die) and WID (Within-Die) variations, while the effect of Ran (Random) process variations, which also present comparable importance as feature size shrinks, still needs to be considered. In addition, integration issues such as IR-Drop, cross-talk, may also be the malefactors for yield loss. To counter with these variations, a safety margin δ must be added to the global supply voltage V_{DD}. When V_{DD} scales, δ grows larger since the circuits becomes more sensitive to process variations. According to our simulations, targeting for the worst case, δ could be as high as 100mV at V_{DD}<400mV, such large safety margin limits the amount of energy reduction.

Characterizing online critical paths and adjusting global V_{DD} for each individual die with a close-loop V_{DD} regulation help to relieve the energy overhead introduced by the high safety margin reservation. One method is duplicating the critical paths to mimic the timing behavior of the slowest path. However, in the sub/near threshold region, the critical path can exhibit very large deviation from its duplicated counterpart. A more robust in-situ delay characterization method has been introduced in [3]. This is realized through the “crystal ball flipflop” [2], which is a flipflop (FF) with an increased setup time placed in parallel to the regular FF (see Fig. 7(a)). The longer setup time can be achieved by using an internal signal of the regular FF as input signal. Monte-Carlo simulation is carried out to insure that the crystal ball flip-flop is slower than the regular FF under any process variations. The crystal ball FFs are only placed at the end of critical paths. When the circuitry powers on, the start-up V_{DD} is set to a value with the adequate safety margin δ, which ensures that no timing error can happen. The V_{DD} control loop monitors the circuit periodically. As V_{DD} is slowly scaled down, a timing error occurs first at the crystal ball FF. This timing error can be detected by a XOR-gate comparing the outputs of the regular FF and the crystal ball FF. At this time the regular FF still functions correctly. The V_{DD} scaling is then stopped to prevent any timing error occurring on the regular FF. Once the V_{DD} is fixed at the minimum error-free value, the V_{DD} characterization circuits can be disabled. In this way, this scheme avoids a large safety margin reservation.

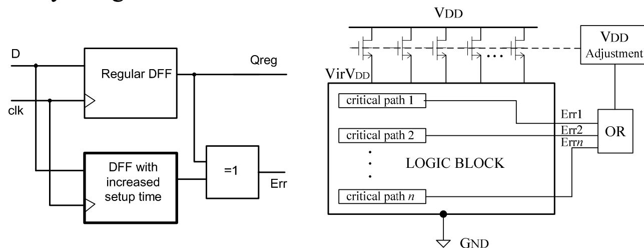

Fig. 7. (a) crystal ball flipflop (b) V_{DD} digital actuator

Since parallel circuit blocks of circuits will be used as discussed in Section III, it is preferable to characterize and to determine the error-free V_{DD} for each block individually. In [3], the V_{DD} close-loop control is implemented by an off-chip DC-DC converter. It is unwise to have a DC-DC converter for every individual block, because the overhead such as external analog components, static biasing currents, pins, additional global power routing and distribution is significant. In our design, we use only 1 DC-DC converter to supply a global voltage and several digital power supply actuators (also shown in Fig. 7(b)) to adjust V_{DD} locally. The V_{DD} adjustment scheme includes 2 steps: first, the global V_{DD} from the DC-DC converter is adjusted to reduce δ_{global}, until a crystal ball FF from one block reports error. Second, the local digital actuators adjust the V_{DD} for each individual blocks to reduce the δ_{local}.

In the actuator, the combination of parallelized transistors generates different header resistor values, so that a virtual V_{DD} can be obtained. The number of pMOS transistors and their sizing determine the resolution and V_{DD} control range, and need to be properly designed. The voltage actuator takes less than a microsecond to reach a steady-state for each tuning step, compared to an off-chip DC-DC converter which needs a response time in the order of tens of microseconds.

VII. Case Study: A JPEG Encoder Processor

JPEG compression hardware accelerator is widely used in the applications such as digital still cameras, wireless still image, medical imaging, etc. The proposed JPEG encoder architecture is shown in Fig. 8.

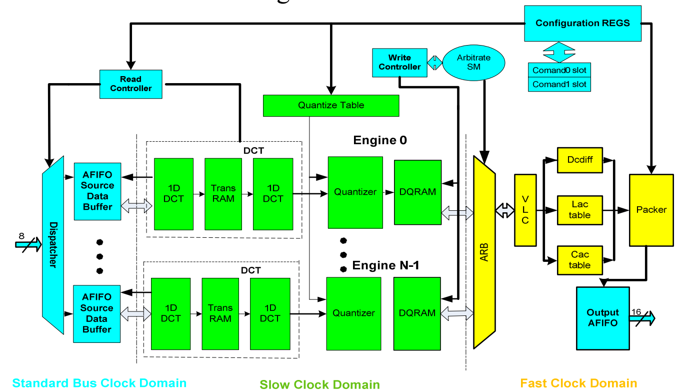

Fig. 8. Proposed JPEG encoder with Parallelized engines

The design is compliant with the JPEG encoder baseline standard [12]. The synthesizable scripts include 17,000 Verilog and 2,000 Perl lines. Asynchronous FIFOs (AFIFOs) are located at the front-end of the data-path to enable an interface flexible to commercial standard bus interface such as PCI/PCI-X/PCI-Express. For each frame, the external main CPU issues a command to the configuration register file of the JPEG processor. The command includes information such as the source data start address/length, destination data start address, YUV sampling ratio, programmable quantization table coefficients, etc. In our architecture, 2 command slots are accommodated in the configuration register file, so the main CPU can issue a command for the next frame while the processor is still processing the current frame. Otherwise the processor should have been stalled for tens of clock cycles in between

of two frames and it could be re-started only when the reconfiguration for the next frame is completed. The JPEG data-path has three main stages: (1) 2D-DCT transformation, (2) Quantization, (3) Huffman encoding. We denote a pair of DCT and Quantization modules as an "engine". Instead of parallelizing the entire data-path, we parallelize only the engine for two reasons. First, the 2D-DCT computation and Quantization modules occupy 50% of the total chip area but they consume 75% of the total power. Second, the Huffman encoding for the DC value of an 8×8 block depends on the DC value of the previous block. If the Huffman encoder is also parallelized, additional effort must be drawn to handle this data dependency. Since it would be difficult to align the output streams from each encoder which have unpredictable lengths, a memory shuffler and many memory operations would become unavoidable.

The final JPEG encoder processor exploits 2 supply voltage domains, 3 frequency domains, and 8 engines. The configuration and interface operate with bus clock and V_{DDH}, the Huffman encoder functions with fast clock and V_{DDH}, while the engines function with slow clock and V_{DDL}. Signals across different clock domains are hand-shaked to increase robustness. The engines have 2 operation modes: the sub/near threshold mode and super-threshold mode. When in the sub/near threshold mode, the V_T balancer works. Online V_{DD} adjustment is applied to every individual engine. Fig. 9 shows that its energy/operation can be reduced by 6× and 9× when at 0.5V and 0.4V, respectively.

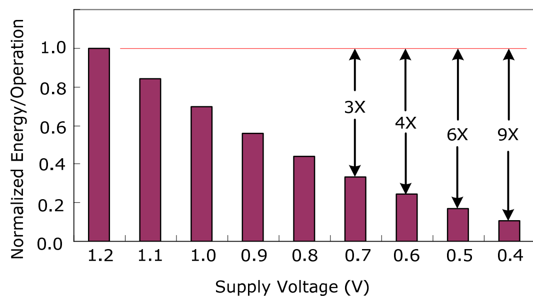

Fig. 9. Normalized Energy/Operation for the Engine

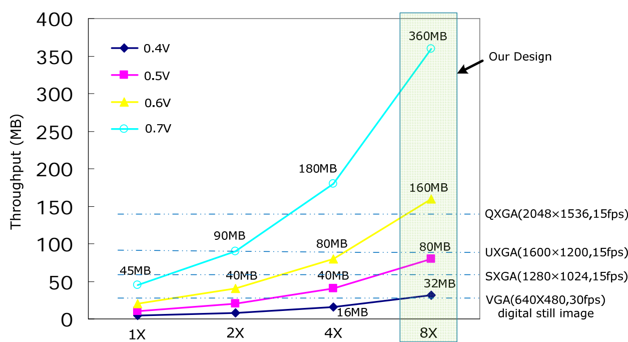

Fig. 10. Area (# of Engines) vs. Throughput for the Engines and possible real-time image applications

If the application has no hard real-time constraint, such as for a digital still image camera, the V_{DD} of the engines can be scaled to a value very close to the V_{opt} which leads to the optimal energy/operation. This processor also handles multi-standard applications which require real-time processing. For instance, it can provide solutions for some wireless image applications. Fig.10 indicates the throughput vs. area (# of engines) tradeoff for the engines. Some achievable applications are annotated.

VIII. Summary and Conclusions

In this paper we have illustrated the possibility to achieve an ultra-low energy/operation for digital circuit by taking advantage of a sub/near threshold supply voltage. We use parallelism to compensate for the throughput degradation. A configurable V_T balancer has been introduced to balance the V_T mismatch in the sub-threshold at process corners. An approach which improves transistor driving capability by exploiting V_T mismatch between parallelized transistors has also been introduced. An autonomous local V_{DD} adjustment method, which uses crystal ball flipflop characterization and digital voltage actuators, has been proposed to reduce design margin individually. Finally, a multi-standard JPEG encoder is proposed to demonstrate these ideas. Our simulation results show that, the energy of the DCT and Quantization Engine can be reduced by 9× at a 0.4V supply compared to the value at the nominal 1.2V supply.

References

[1] http://www.samsung.com/global/business/semiconductor/products/asic/ Products_DesignTechnology.html

[2] L. Angehel and M. Nicolaidis, "Cost reduction and evaluation of a temporary fault detecting technique," *Design, Automation and Test in Europe*, 2000

[3] Matthias Eireiner, Stephan Henzler, Georg Georgakos, Joerg Berthold, Doris Schmitt-Landsiedel, "Local Supply Voltage Adjustment for Low Power Parameteric Yield Increase," in Proceedings of *ESSCIRC*, 2006

[4] S.Mukhopadhyay, "Gate Leakage Reduction for Scaled Devices Using Transistor Stacking," *IEEE TVLSI*, Vol.11, No.4, Aug 2003

[5] Benton H.Calhoun, et al. , "Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuits," *JSSC*, Vol.40, No.9, September 2005

[6] Alice Wang, et al. , "Optimal Supply and Threshold Scaling for Subthreshold CMOS Circuits," *IEEE Computer Society Annual Symposium on VLSI*, April 2002

[7] Benton H. Calhoun, Anantha P. Chandrakasan, "Characterizing and Modeling Minimum Energy Operation for Subthreshold Circuits," *ISLPED*, August 2004

[8] ~~Deleted for blind review~~

[9] Masakatsu Nakai, et al., "Dynamic Voltage and Frequency Management for a Low-Power Embedded Microprocessor," *IEEE JSSC*, Vol.40, No.1, January 2005

[10] Bo Zhai, et al., "A 2.60pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency," *IEEE Symposium on VLSI Circuits*, June 2006

[11] Myeong-Eun Hwang, et al., "A 85mV 40nW Process-Tolerant Subthreshold 8×8 FIR Filter in 130nm Technology," *IEEE Symposium on VLSI Circuits*, June 2007

[12] Gregory K. Wallace, "The JPEG Still Picture Compression Standard," *IEEE Trans. Consumer Electronics*, 1991

[13] M.J.M. Pelgrom, *et al.* , "Matching properties of MOS transistors," *JSSC*, vol.24, October 1989