An ultra-low-energy multi-standard JPEG co-processor in 65 nm CMOS with sub/near threshold supply voltage

Citation for published version (APA):
Pu, Y., Pineda de Gyvez, J., Corporaal, H., & Ha, Y. (2010). An ultra-low-energy multi-standard JPEG co-processor in 65 nm CMOS with sub/near threshold supply voltage. *IEEE Journal of Solid-State Circuits*, *45*(3), 668-680. https://doi.org/10.1109/JSSC.2009.2039684

DOI:
10.1109/JSSC.2009.2039684

Document status and date:
Published: 01/01/2010

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

An Ultra-Low-Energy Multi-Standard JPEG Co-Processor in 65 nm CMOS With Sub/Near Threshold Supply Voltage

Yu Pu, *Member, IEEE*, Jose Pineda de Gyvez, *Fellow, IEEE*, Henk Corporaal, *Member, IEEE*, and Yajun Ha, *Senior Member, IEEE*

***Abstract*—We present a design technique for (near) subthreshold operation that achieves ultra low energy dissipation at throughputs of up to 100 MB/s suitable for digital consumer electronic applications. Our approach employs i) architecture-level parallelism to compensate throughput degradation, ii) a configurable $V_{\rm T}$ balancer to mitigate the $V_{\rm T}$ mismatch of nMOS and pMOS transistors operating in sub/near threshold, and iii) a fingered-structured parallel transistor that exploits $V_{\rm T}$ mismatch to improve current drivability. Additionally, we describe the selection procedure of the standard cells and how they were modified for higher reliability in the subthreshold regime. All these concepts are demonstrated using *SubJPEG*, a 1.4×1.4 mm^2 65 nm CMOS standard-$V_{\rm T}$ multi-standard JPEG co-processor. Measurement results of the discrete cosine transform (DCT) and quantization processing engines, operating in the subthreshold regime, show an energy dissipation of only 0.75 pJ per cycle with a supply voltage of 0.4 V at 2.5 MHz. This leads to $8.3\times$ energy reduction when compared to using a 1.2 V nominal supply. In the near-threshold regime the energy dissipation is 1.0 pJ per cycle with a 0.45 V supply voltage at 4.5 MHz. The system throughput can meet 15 fps 640×480 pixel VGA standard. Our methodology is largely applicable to designing other sound/graphic and streaming processors.**

***Index Terms*—JPEG, parallel architecture, sub-threshold, ultra low energy.**

I. Introduction

With the ever-shrinking feature size, the number of transistors integrated in one digital core doubles approximately every two years. The increasing transistor density greatly challenges the limited battery life and thermal properties of the IC. Exploring a design methodology for ultra low-energy, "green" digital circuits is thus very important. One of the most effective means to achieve these goals is to scale the supply voltage $V_{\rm DD}$ along with the operating frequency. As $V_{\rm DD}$ scales, not only does the dynamic energy reduce quadratically, but also the leakage current does reduce super-linearly due to the drain-induced barrier-lowering (DIBL) effect. Therefore, the total energy dissipation of a circuit can considerably be reduced. In addition, $V_{\rm DD}$ scaling reduces transient current spikes, hence lowering the notorious ground bounce noise. This also helps to improve the performance of sensitive analog circuits on the chip, such as delay-lock loops (DLL), which are crucial for the functioning of large digital circuits.

In contrast to analog circuit design where lowering the $V_{\rm DD}$ to the subthreshold region is generally avoided because of the small values of the driving currents and the exceedingly large noise, CMOS digital logic gates can work seamlessly from full $V_{\rm DD}$ to well below threshold voltage $V_{\rm T}$. Theoretically, operating digital circuits in the near/sub-threshold region ($V_{\rm GS} < V_{\rm T}$) can help obtain huge energy savings. However, the design rules provided by foundries normally set 2/3 of the full $V_{\rm DD}$ as the practical limitation for $V_{\rm DD}$ scaling. Taking Samsung's DVFS Design Technology [1] and TSMC's design rules as examples, the constraint of $V_{\rm DD}$ for digital circuits designed in CMOS 65 nm Standard $V_{\rm T}$ Process is in the 0.8 V $\sim$ 1.2 V range. The reasoning behind the limitation is twofold. First, as $V_{\rm DD}$ scales, the driving capability of transistors reduces accordingly. Most consumer electronic applications need operating frequencies in the range of tens of MHz to reach certain throughput, which might not be fulfilled with aggressive $V_{\rm DD}$ scaling. Second, digital circuits become particularly sensitive to process variations when $V_{\rm DD}$ scales below 2/3 full $V_{\rm DD}$. Process variations are likely to cause malfunctioning, and both the timing yield and functional yield may tremendously decrease. As a result, $0.67V_{\rm DD}$ is generally chosen to maintain an adequate margin to prevent high yield loss and to keep quality according to industrial standards. The goal of our work is to safely evade this limitation so as to enable wide range voltage scaling, from nominal supply to near/sub threshold.

Sub/near threshold techniques have been explored in recent years. Fig. 1 shows a comparison of the computation efficiency (GOPS/W) and throughput (MOPS) of our *SubJPEG* co-processor and other existing subthreshold processors. Likewise, Table I summarizes the most relevant work in the field. In contrast to the work presented in those publications, our work has some unique features. Firstly, we explore the use of architecture-level parallelism to compensate throughput degradation at ultra-low supply values. Parallelism along with sub/near threshold techniques is best suited for low-energy and medium frequency applications, such as mobile image processing. Secondly, this work proposes a configurable $V_{\rm T}$ balancer to lessen the $V_{\rm T}$ mismatch between nMOS and pMOS transistors, such that both the functional and the timing yield

Manuscript received June 24, 2009; revised September 09, 2009. Current version published February 24, 2010. This paper was approved by Associate Editor Bevan Baas.

Y. Pu was with the Ultra Low Power DSP Processor Group, IMEC-NL, 5656 AE Eindhoven, The Netherlands, and is now with the Sakurai Lab, University of Tokyo, Tokyo 153-8505, Japan (e-mail: y.pu@tue.nl).

J. Pineda de Gyvez is with NXP Semiconductors, 5656 AE Eindhoven, The Netherlands.

H. Corporaal is with the Faculty of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands.

Y. Ha is with the Department of Electrical and Computer Engineering, National University of Singapore, 117576 Singapore.

Digital Object Identifier 10.1109/JSSC.2009.2039684

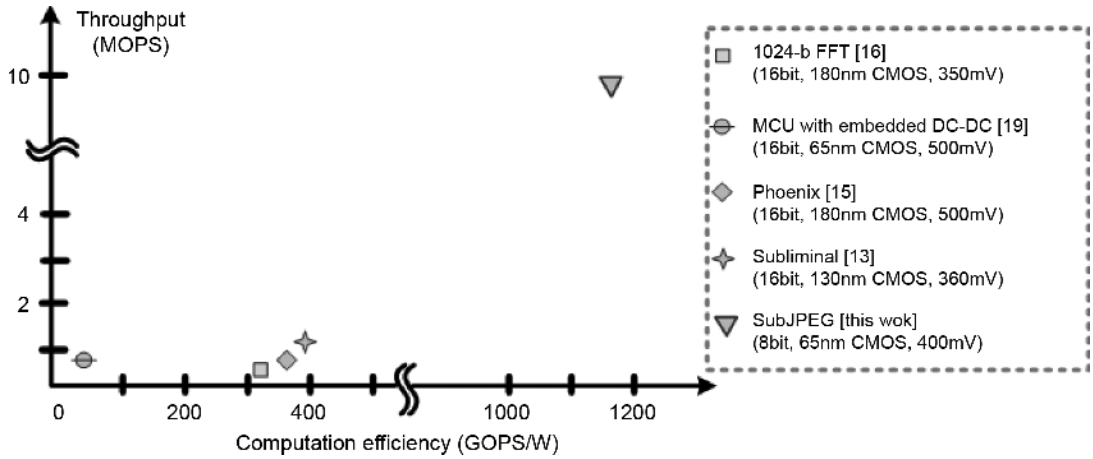

Fig. 1. Computation efficiency and throughput of this work and other works.

TABLE I
SUMMARY OF EXISTING SUB-THRESHOLD WORK

Category	Existing sub-threshold work
Sub-threshold modeling	[2][3]: built up the analytical models for sub-threshold current, delay, energy and variations
Sub-threshold logic design	[4][5][6][7]: explored sub-threshold logic cells
Sub-threshold memory	[8]: 256kb 10-T dual-port SRAM in 65nm CMOS
	[9]: 512×13bit dual-port SRAM in 180nm CMOS
	[10]: 480kb 6-T dual-port SRAM in 130nm CMOS
	[11]: 2kb 6-T single-port SRAM in 130nm CMOS
	[12]: 256 kb 8-T dual-port SRAM in 65nm CMOS
Sub-threshold processors	[13][14][15]: sensor node processors in 130nm and 180nm CMOS
	[16]: 180mV FFT processor in 180nm CMOS
	[17]: 0.4V UWB baseband processor in 65nm CMOS
	[18]: 85mV 40nW 8×8 FIR filter in 130nm CMOS
	[19]: 2-stage pipelined micro-controller with embedded SRAM and DC-DC converter in 65nm CMOS

are increased. Thirdly, we make use of design approaches that exploit parallel-transistor V_{T} mismatch to improve drivability in power switches, and of design strategies that select a reliable cell library for logic synthesis, and that turn ratioed logic into non-ratioed logic to improve the robustness of our design in the subthreshold regime. To demonstrate these ideas, we have designed and implemented a 65 nm SV_{T} CMOS ultra-low energy multi-standard JPEG co-processor.

The remainder of this paper is organized as follows. Section II presents the physical-level effort we have made for an enhanced circuit yield. In Section III, the architecture of *SubJPEG* is introduced in detail. Section IV presents key design issues and the evaluation results of the prototype chip. Finally, Section V draws conclusions of this work.

II. PHYSICAL LEVEL EFFORT FOR AN ENHANCED YIELD

A. Configurable V_{T} Balancer

V_{T} mismatch dominates the subthreshold current variation due to its exponential correlation to the current. Since transistor V_{T} is controlled by an independent doping process, pMOS/nMOS V_{T} can vary significantly with respect to each other. Consequently, this variability can result in lower circuit yield. For example, at the fast nMOS slow pMOS corner (FNSP) where the nMOS network is much leakier than the pMOS network, a sufficiently high output voltage V_{OH} may not be reached. Similarly, an insufficiently low voltage V_{OL} can happen when at the fast pMOS slow nMOS corner (SNFP). Even if the noise margin can be met, either the rising or falling time becomes exceedingly long at process corners, which also dramatically deteriorates the timing yield. Therefore, it is very important to balance the V_{T} of pMOS and nMOS transistors. We propose a configurable V_{T} balancing scheme (Fig. 2), which enables ultra wide range V_{DD} scaling from the nominal supply voltage to sub-threshold. This configurable V_{T} balancer is an extension of our previous work [20]. Our V_{T} balancer is also different from the regulator presented in [21] since it uses an imbalance detector which has a better sensitivity. Also, it uses an amplifier in the feedback loop to enhance the sensitivity, and, it is configurable to support wide V_{DD} tuning. Let us address now the operation of our V_{T} balancer. When the processor works in the super-threshold mode, S_0 is off such that the tri-state buffer is configured to be in a high impedance state. Since the power switch transistors S_3 and S_4 are on, and S_1, S_2

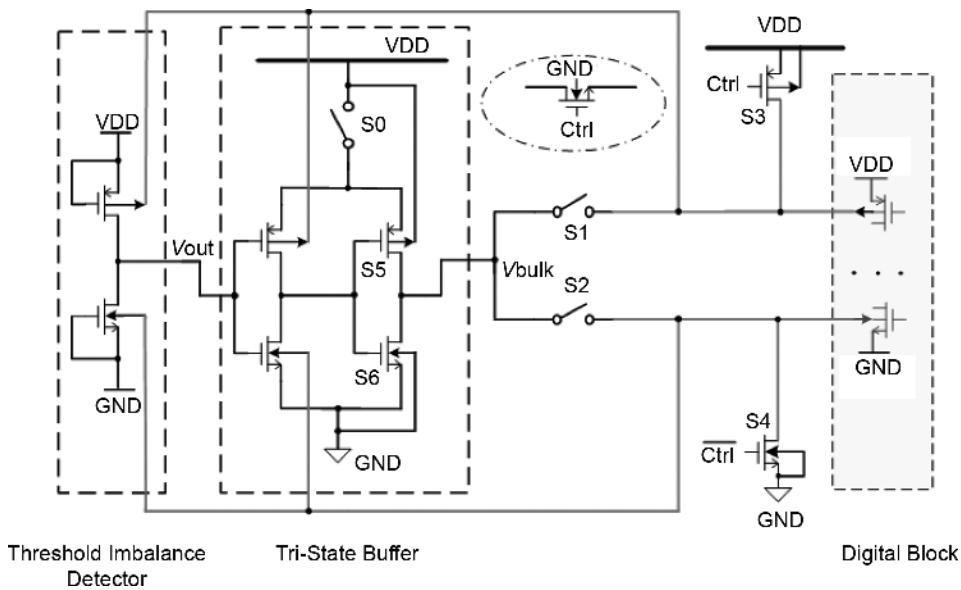

Fig. 2. Proposed configurable V_{T} balancer.

are off, the bulk of the pMOS transistors is connected to V_{DD}, and the bulk of the nMOS transistors is connected to $\mathrm{G_{ND}}$. When the processor is configured to work in the subthreshold mode, S_0 is on, and thus the tri-state buffer is functional. In this mode, S_1, S_2 are on, and S_3, S_4 are off. Therefore, the buffer's output voltage passes through S_1, and S_2 to feed the bulk of the logic gates. A CMOS inverter, whose pMOS and nMOS transistors are off, functions as a process-corner V_{T} imbalance detector. Observe that V_{bulk} is never higher than $|V_{\mathrm{T}}|$ preventing in this way the junction diodes from turning on in the P-well and N-well under control. V_{out} and V_{bulk} are designed in advance to be at $V_{\mathrm{DD}}/2$ in the typical process corner (TT). V_{out} fluctuates with the variations of process and temperature. The buffer detects and amplifies the swing of V_{out}. The buffer's output V_{bulk}, which feeds the bulk voltage for the logic gates, is fed back to the bulk of the threshold balancing detector to force the pMOS/nMOS V_{T} balancing. For instance, if the nMOS is leakier than the pMOS, V_{out} will decrease, triggering a much larger drop on V_{bulk}. This drop will make the nMOS increase its V_{T} and the pMOS decrease its V_{T}, such that the process-corner V_{T} imbalance is mitigated. In our design, the power switch transistors S_0, S_1 and S_2 are nMOS transistors overdriven by a boosted gate voltage. Hence, their R_{on} is small enough to avoid the potential drop across a transistor. The boosted gate voltage can be obtained either from other high voltage domains or from the periphery I/O power rails.

We use a metric $\zeta = (V_{\mathrm{out}} - V_{\mathrm{DD}}/2)/V_{\mathrm{DD}}$ to represent the V_{T} imbalance. In fact, ζ depicts how far V_{out} deviates from $V_{\mathrm{DD}}/2$ due to unbalanced V_{T} devices. The larger ζ is, the larger the V_{T} imbalance is. Fig. 3(a) shows the simulated 3σ range of ζ, with and without our V_{T} balancing scheme. As can be seen, the imbalance between V_{T} of pMOS and nMOS transistors is confined to a much tighter range after V_{T} balancing. Fig. 3(b) shows the Monte Carlo simulated propagation delay for an inverter with aspect ratio of $Wp/Wn = 1.1\,\mu\mathrm{m}/0.40\,\mu\mathrm{m}$ to drive a capacitive load of 5 fF at $V_{\mathrm{DD}} = 400$ mV in the CMOS 65 nm $\mathrm{S}V_{\mathrm{T}}$ process. After V_{T} balancing, the average propagation delay of the inverter is reduced from 14 ns to 10 ns. This speed improvement is because both the p/nMOS transistors are forward-biased when the balancer is turned on. Most importantly, the standard deviation σ is reduced by $4.7\times$ and the σ/μ is reduced by $3.6\times$ when the proposed configurable V_{T} balancer is used, as an exceedingly long rising/falling time is avoided.

B. Improving Driving Capability by Exploiting Parallel V_{T} Mismatch in Power Switches

Even though V_{T} mismatch is known to be catastrophic for circuit functionality, we have developed 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 proof and simulation results that show that in the subthreshold regime the V_{T} mismatch between parallelized transistors always results in an increased mean driving current. This interesting property has been applied to the power-switches of the V_{T} balancer circuit.

Suppose $\mu(V_{\mathrm{T}})$, $\sigma(V_{\mathrm{T}})$ are the mean and standard deviation of V_{T} of an nMOS transistor as shown in Fig. 4(a). Considering

$$\mu(I_{\mathrm{eff}}) = I_{\mathrm{On}} e^{[V_{\mathrm{GS}} - \mu(V_{\mathrm{T}}) + \eta V_{\mathrm{DS}} - \gamma V_{\mathrm{SB}}]/nU + [\sigma(V_{\mathrm{T}})/nU]^2/2} (1 - e^{-V_{\mathrm{DS}}/U}) \tag{2}$$

$$\sigma(I_{\mathrm{eff}}) = I_{\mathrm{On}} e^{[V_{\mathrm{GS}} - \mu(V_{\mathrm{T}}) + \eta V_{\mathrm{DS}} - \gamma V_{\mathrm{SB}}]/nU + [\sigma(V_{\mathrm{T}})/nU]^2/2} (1 - e^{-V_{\mathrm{DS}}/U}) \sqrt{e^{[\sigma(V_{\mathrm{T}})/nU]^2} - 1} \tag{3}$$

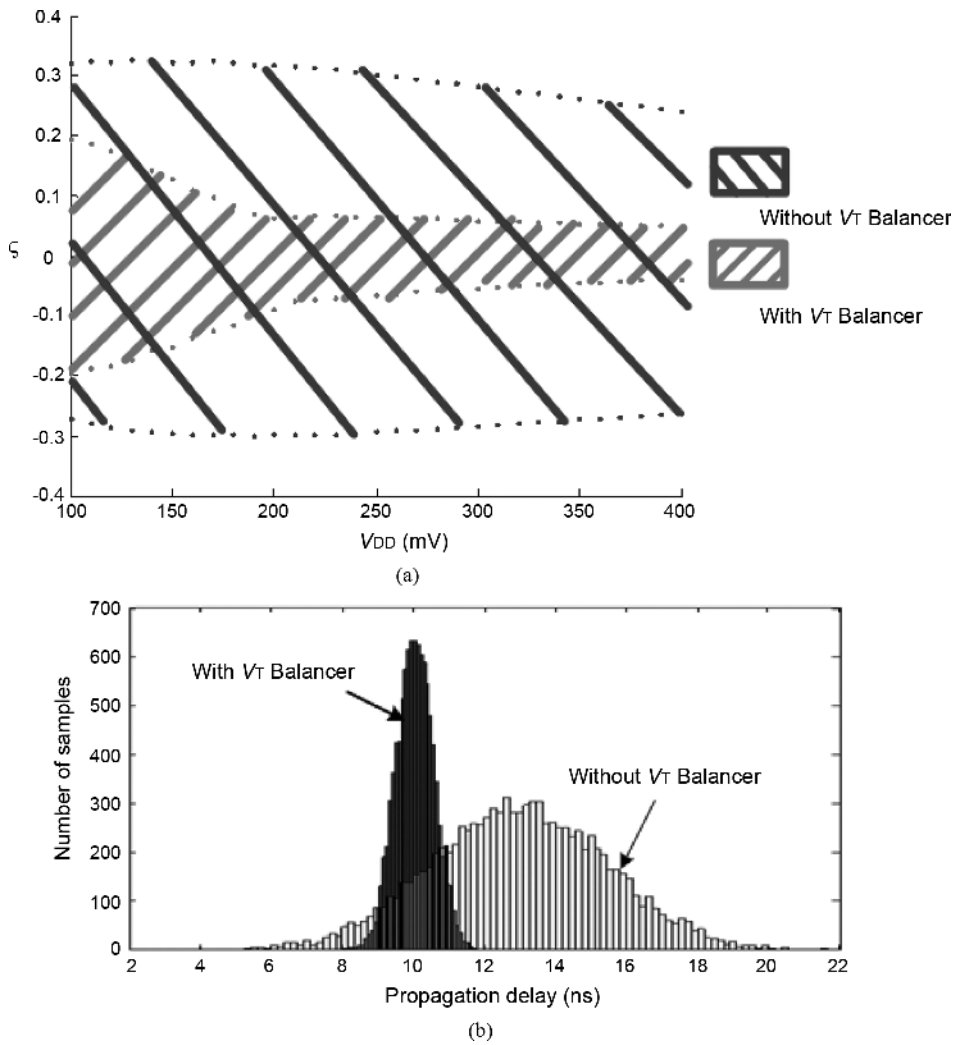

Fig. 3. (a) Simulated 3σ range of ζ. (b) Propagation delay for an inverter in 65 nm CMOS from Monte Carlo simulation ($W_p/W_n = 1.1\,\mu\text{m}/0.40\,\mu\text{m}$, $\text{C}_{\text{load}} = 5$ fF).

the intra-die V_{T} variation of a single transistor modeled as in [22], we have

$$\sigma(V_{\text{T}}) = \frac{A\Delta V_{\text{T}}}{\sqrt{WL}} \tag{1}$$

where $A_{\Delta V_{\text{T}}}$ is a technology conversion constant (in mVμm), and *WL* is the transistor's active area. Since V_{T} follows a normal distribution, the transistor's on-current I_{eff} follows a log-normal distribution in sub-threshold. Using the properties of a log-normal distribution, the mean value and standard deviation of I_{eff} are as shown in (2) and (3) at the bottom of the previous page, where V_{GS} is the gate source voltage, U the intrinsic thermal voltage, and n the junction gradient coefficient. Suppose the transistor is equally divided in N-parallel nMOS transistors, $T_1 \ldots T_{\text{N}}$ [see Fig. 4(b)]. Without loss of generality, let us denote the mean and standard deviation of the threshold voltage of any of these parallel transistors T_x as

$$\mu(V_{\text{T},1}) = \mu(V_{\text{T},2}) = \cdots = \mu(V_{\text{T},N}) = \mu(V_{\text{T}x}) \tag{4}$$

$$\sigma(V_{\text{T},1}) = \sigma(V_{\text{T},2}) = \cdots = \sigma(V_{\text{T},N}) = \sigma(V_{\text{T}x}) \tag{5}$$

$$\begin{aligned}\mu(I'_{\text{eff}}) &= \sum_{\text{i}=1}^{N} \mu(I_{\text{eff,i}}) \\ &= I_{\text{On}} e^{[V_{\text{GS}} - \mu(V_{\text{Tx}}) + \eta V_{\text{DS}} - \gamma V_{\text{SB}}]/nU + [\sigma(V_{\text{Tx}})/nU]^2/2} (1 - e^{-V_{\text{DS}}/U})\end{aligned} \tag{7}$$

$$\frac{\sigma(I'_{\text{eff}})}{\mu(I'_{\text{eff}})} = \sqrt{e^{[\sigma(V_{\text{Tx}})/nU]^2} - 1}. \tag{8}$$

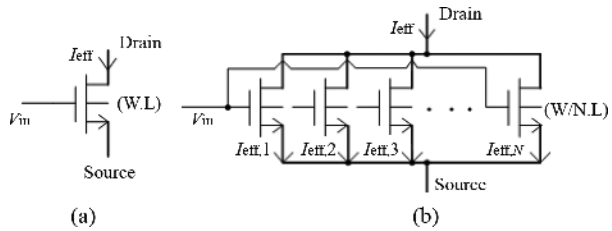

Fig. 4. (a) nMOS transistor with aspect ratio (W, L); (b) N-parallelized nMOS transistors with aspect ratio (W/N, L).

where

$$\sigma(V_{\mathrm{T}x}) = \frac{A_{\Delta V_{\mathrm{T}}}}{\sqrt{\frac{WL}{N}}} = \frac{\sqrt{N}A_{\Delta V_{\mathrm{T}}}}{\sqrt{WL}} \tag{6}$$

Then, the mean value of the total subthreshold current $\mu(I'_{\mathrm{eff}})$ in Fig. 4(b) is obtained as shown in (7) and (8) at the bottom of the previous page. Comparing (1) and (6), and since $N > 1$, we have that

$$\sigma(V_{\mathrm{T}x}) > \sigma(V_{\mathrm{T}}). \tag{9}$$

Then, by comparing (2) and (7), we obtain

$$\mu(I'_{\mathrm{eff}}) > \mu(I_{\mathrm{eff}}). \tag{10}$$

As can be seen, dividing a large transistor into smaller parallelized transistors helps to increase the subthreshold current due to larger V_{T} mismatch. We also did Monte Carlo simulations to confirm the effectiveness of this approach. As way of reference assume an $\mathrm{S}V_{\mathrm{T}}$ nMOS transistor with aspect ratio $\mathrm{W/L} = 0.96\ \mu\mathrm{m}/0.065\ \mu\mathrm{m}$, divided in N-transistors $(N = 1, 2, 3, 4, 6, 8)$, with its gate voltage V_{in} and drain-to-source voltage V_{DS} set at 200 mV. The reason why 200 mV $|V\mathrm{ds}|$ and $|V\mathrm{gs}|$ is chosen, is because in the V_{T} balancer the $|V\mathrm{gs}|$ and $|V\mathrm{ds}|$ of the power switches operating in the subthreshold regime is approximately 200 mV (half of 400 mV V_{DD}). Since the power switches' output will forward bias the bulk of p/n transistors in digital blocks, a close to $|200\ \mathrm{mV}|$ output voltage is the right magnitude which can bring V_{T} unbalance from 3σ; deviation to typical value without incurring too much excessive leakage current. The simulated mean and standard deviation values of the effective driving current I_{eff} are listed in Table II. As seen, the larger the number of segments N, the larger the V_{T} mismatch, consequently the larger the mean subthreshold driving current. However, Table II also shows an increasing driving current variability and larger $\sigma(I_{\mathrm{eff}})/\mu(I_{\mathrm{eff}})$ as the transistor becomes narrower. According to (8), this is due to an increased V_{T} shift caused by narrow width effects. To mitigate such effect, instead of dividing all transistors into minimal width transistors, our design constrained the transistor width to be not smaller than a certain limit. By constraining a maximum $\sigma(I_{\mathrm{eff}})/\mu(I_{\mathrm{eff}}) = 20\%$, a same driving current can be achieved with approximately 10% transistor area reduction. In addition, the multi-finger layout can avoid a very strange aspect-ratio and easily fit into the layout of the other devices hence making the entire layout more compact.

TABLE II
MEAN AND STANDARD DEVIATION OF DRIVING CURRENT

N	$\mu(I_{\mathrm{eff}})$ (nA)	$\sigma(I_{\mathrm{eff}})$ (nA)
1	5.390	2.495
2	5.991	3.023
3	7.667	4.233
4	9.324	4.885
6	12.934	6.255
8	13.316	7.379

C. Sub-Threshold Library Selection

The standard library cells optimized for super-threshold design must be revised for reliable logic synthesis. The cells having a large effective driving current variability will have a remarkably low yield. We identified these cells through Monte Carlo simulations and filtered them out before logic synthesis. The metric we used is that, after applying V_{T} balancing, the cells that have $\sigma(I_{\mathrm{eff}} - I_{\mathrm{idle}})/\mu(I_{\mathrm{eff}} - I_{\mathrm{idle}}) > 20\%$ at $V_{\mathrm{DD}} = 400\ \mathrm{mV}$, are eliminated, where I_{idle} is the leakage current for off-transistors. These cells have some typical structures:

1) More Than Four Parallel Transistors and More Than Four Stacked Transistors: The standard cells are composed of narrow transistors to increase area efficiency. As the number of parallel transistors and the number of stacked-transistors increases, the leakage current variability increases dramatically, as shown in Section II-B. We simply discarded logic gates with more than four parallel transistors or more than four stacked transistors, such as 4-input NAND and NOR gates.

2) Ratioed Logic: Ratioed logic can reduce the number of transistors required to implement a given logic function, but it must be sized carefully to guarantee that the active current is stronger than the static current. Therefore, the correct functioning of ratioed logic cells depends largely on the sizing. In the subthreshold region, the largest current variability is due to V_{T} variation. Even a small variation on V_{T} has a heavy impact on the active or static current. Therefore, logic cells totally relying on transistor sizing are dangerous and should be avoided.

3) Feedback Logic: Feedback logic is a special type of ratioed logic which uses positive feedback loops to help change the logic values. Due to V_{T} variation, the output of the logic can have stuck-high or stuck-low failures and thus never flip.

D. Turning Ratioed Logic Into Non-Ratioed Logic

Latches and registers are the feedback logic that must be used in sequential circuits. To reduce loading on clock net and ease ultrahigh speed designs, some latches/registers use weak but always-on feedback inverters. Fig. 5 shows how to turn them into non-ratioed logic. By using the *clk* and $\overline{clk}$ signals, we prevent the slave inverters (I_2, I_4) from directly cross-coupling with the master inverters (I_1, I_3). As a result, when writing into the latch, the slave inverter is always disabled, so the writing to the master inverter is facilitated. After the writing is done, the slave inverter is enabled to help maintain the logic value. Therefore, the race

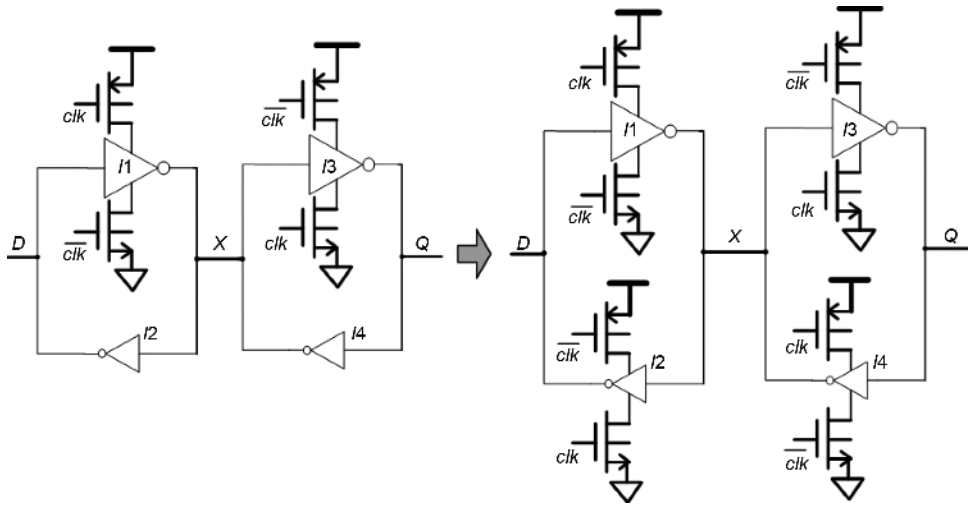

Fig. 5. Turning ratioed logic into non-ratioed logic.

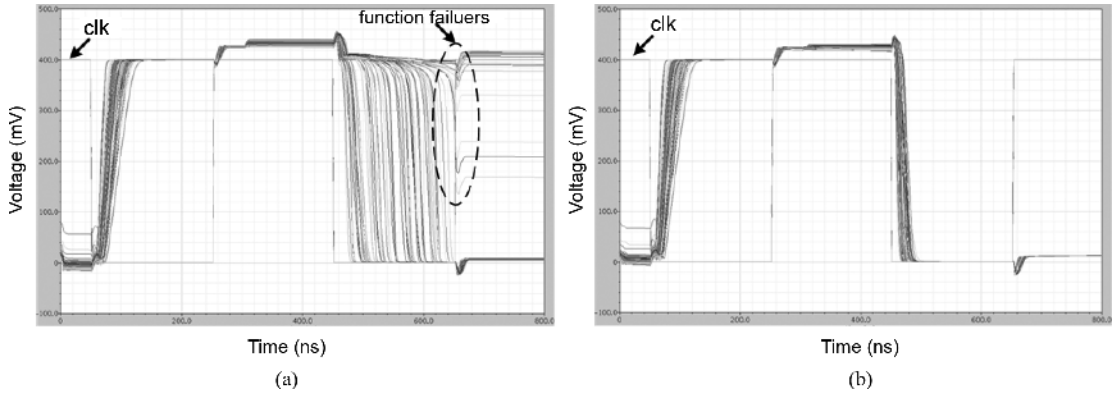

Fig. 6. Monte Carlo simulation results at node X at $V_{\mathrm{DD}} = 400$ mV: (a) before turning ratioed logic into non-ratioed logic; (b) after turning ratioed logic into non-ratioed logic.

between the slave and master inverters is avoided. Fig. 6 compares the Monte Carlo simulation results at node X (the output from the negative latch) at $V_{\mathrm{DD}} = 400$ mV before and after turning ratioed logic into non-ratioed logic. With this modification, the stuck high and stuck low failures are avoided. In addition, the propagation delay becomes more than an order tighter.

III. SubJPEG Architecture

JPEG is an international compression standard for continuous-tone still images, both grayscale and color [23], [24]. As a generic image compression standard, JPEG supports a wide variety of image applications. The baseline JPEG encoding processing has three primary steps: 8×8 discrete cosine transformation (DCT), quantization, entropy encoding. Our goal is to design a JPEG compression co-processor that consumes extremely low energy and thus can be used in application fields such as image sensoring, digital still cameras, mobile image, etc. The design challenge is to explore an architecture with efficient parallelism to trade-off area, throughput and energy.

Our baseline design was built from scratch to accommodate architectural changes required for subthreshold operation in a 65 nm CMOS SV_{T} process. Its area and energy breakdown are shown in Fig. 7. The term "engine" denotes a combined 2D-DCT and Quantization module. As seen, the engine dominates both the energy and area. At the nominal supply voltage the engine occupies less than 50% of the total silicon area but consumes around 70% of the total energy. The rest of the components, such as the Huffman encoder and the configuration logic, are of less importance. Thus, minimizing the energy consumption of the engine becomes our primary target when designing the new architecture. Therefore, instead of parallelizing the entire data-path, we decided to parallelize only the engine. Another reason for making this decision is because of the difficulty in parallelizing the Huffman encoder. 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. Also, it would be difficult to align the output streams from each Huffman encoder which have unpredictable lengths, a memory shuffler and many memory operations would become unavoidable. Fig. 8 indicates the estimated throughput versus area tradeoff for the engines with annotated application standards. Four parallel engines were chosen in our design because from simulations we observed that the encoder was already capable of meeting 15 fps VGA standard at 0.4 V with $9\times$ energy reduction (in subthreshold mode), 30 fps VGA standard at 0.5 V with $6\times$ energy reduction (in near-threshold mode), 15 fps

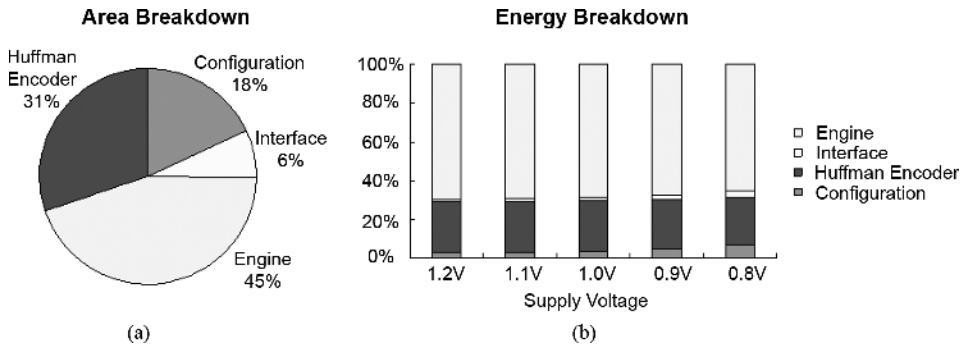

Fig. 7. (a) Area and (b) energy breakdown for baseline JPEG encoder.

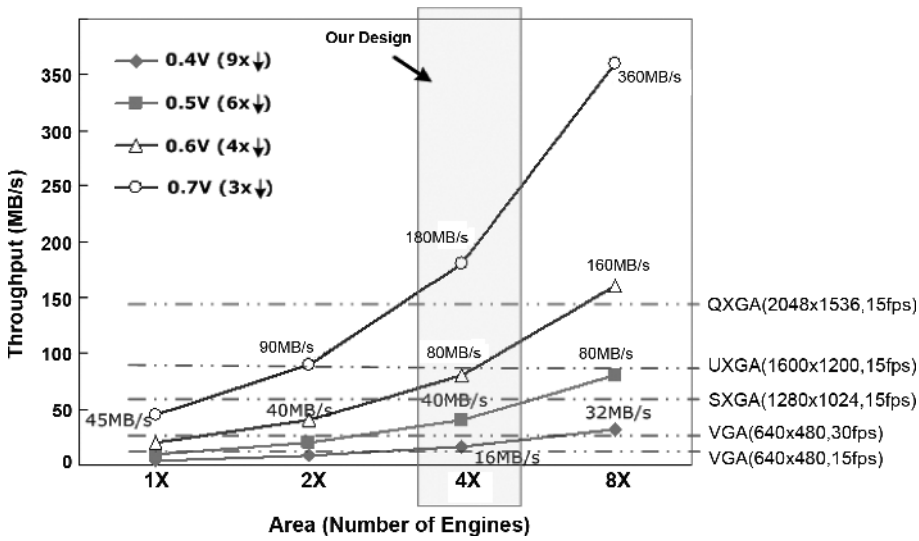

Fig. 8. Estimated throughput and area tradeoff.

QXGA standard at 0.7 V with 3× energy reduction (super-threshold mode). If the application has no hard real-time constraints, such as for a still image of a digital camera, then, ideally, the V_{DD} of the engines can be scaled to a value very close to V_{opt} which leads to the optimal energy per engine operation.

SubJPEG is a co-processor hosted by a main CPU. The main CPU can communicate with *SubJPEG*, issue commands and access the status registers in *SubJPEG* through the control lines. *SubJPEG* interfaces directly with a commercial standard bus, such as PCI/PCI-X/PCI-Express. It has direct-memory-access (DMA) which supports fetching the image data stored in an external memory without going through the main CPU. Fig. 9 shows the *SubJPEG* processor diagram. The final JPEG encoder processor exploits two supply voltage domains $(V_{\mathrm{DDH}}, V_{\mathrm{DDL}})$, three frequency domains (bus_clk, engine_clk, Huffman_clk). The control path and data path are described below.

A. Data Path Design

Before going into the details of the data path design, let us first address how we handled internal storage banks. We compared all memory banks synthesized as register files (RF) using standard cells (mainly DFFs) with fast dual-port SRAMs generated from a commercial memory generator. At 1.2 V nominal supply, the standard cell based RF is not only faster but also more energy efficient than the dual-port SRAM. This is because the energy overhead from the SRAM's peripheral read-out circuitry, such as the sense-amplifiers, dominates the energy when the memory's width and depth are too small. Since SRAMs have worse energy and frequency scaling factors when compared to those of standard cells under voltage scaling, using SRAMs in our design would result in more energy consumption. Also, considering that the reliability of the standard cell based RF is superior to that of the SRAM-based RF at low voltage, we decided to use the synthesized RF with the dedicated subthreshold library throughout our design. We did not adopt the existing subthreshold memory solutions [8]–[12] because all these solutions severely degrade speed and energy efficiency when compared to conventional SRAMs in the super-threshold mode.

Asynchronous FIFOs are located at the front and back of the data-path to enable a flexible interface to a commercial standard bus interface. The AFIFOs are connected with bus_clk, engine_clk and operated with V_{DDH}. The intermediate results being produced from the first 1D-DCT are stored in the Transposed Memory (TransRAM) which is actually a flip-flop based RF. The Transposed Memory behaves as a dual port RAM. While the Transposed Memory is written in row-major order, the second stage of processing reads data from the Transposed Memory in a column-major order, effectively performing a transposition of the intermediate results. The TransRAM contains two block RAM entries, which enable a macro-level

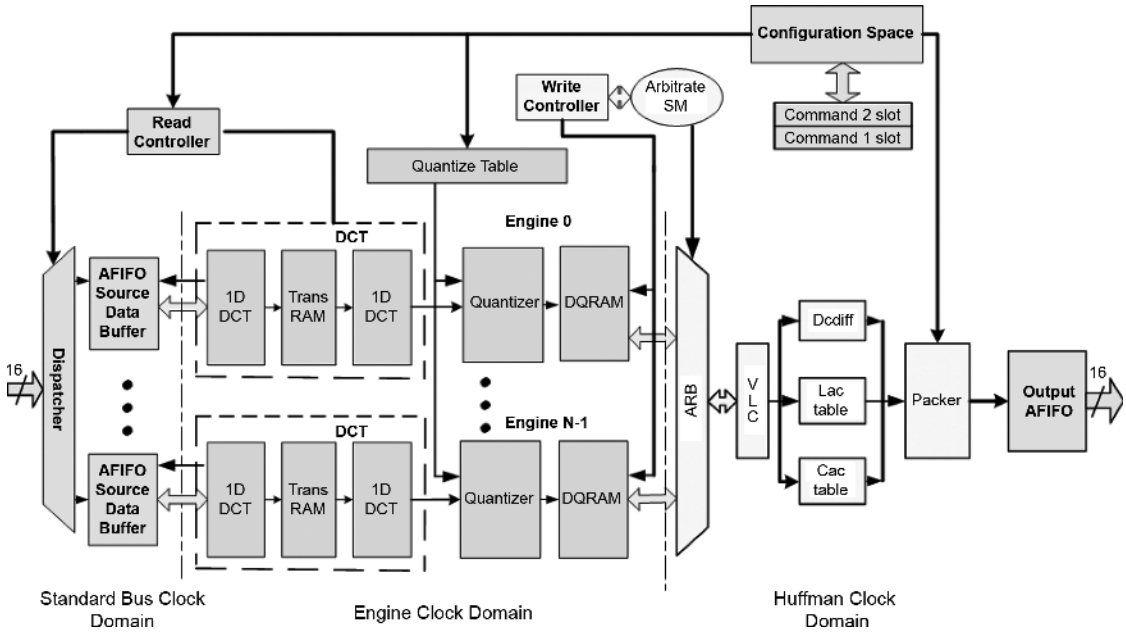

Fig. 9. *SubJPEG* diagram.

TABLE III
REGISTER FILES USED IN *SUBJPEG* DATA PATH

Register Files	V_{DD}, clk	Description
AFIFO Source Data	V_{DDH}, bus_clk, engine_clk	Input buffer, 8×64 bits for each engine
TransRAM	V_{DDL}, engine_clk	12×64 bits per entry, 2 entries per engine
DQRAM	V_{DDL}, engine_clk	8×64 bits per entry, 2 entries per engine
Output AFIFO	V_{DDH}, bus_clk, engine_clk	Output buffer, 16x4 bits

pipelined processing to enhance throughput. That is, the first 1D-DCT can start processing and writing intermediate output into one entry while the second 1D-DCT is still reading data from the other entry. The pipeline latency for 1D-DCT is 80 engine_clk cycles. The output from the second 1D-DCT goes to the quantizer. After the quantization process, the data is stored in a "DQRAM" (also a RF). For the same reason as the TransRAM, the DQRAM contains also two block RAM entries. The engines work with engine_clk and V_{DDL}. Finally, the arbitrator selects data from each entry, and sends the data to the Huffman coder for entropy coding. The Huffman encoder works with its own clock (Huffman_clk) and powered from V_{DDH}. The Huffman encoder takes 80 Huffman_clk cycles to finish processing data from one DQRAM entry. Therefore, the Huffman_clk should be at least $4\times$ faster than the engine_clk since four engines are used, otherwise the Huffman encoder becomes the system's throughput bottleneck. The RFs used for data storage on the data path are summarized in Table III.

B. Control Path Design

The configuration space, read controller (RDC), and write controller (WRC) are the three main modules of the control path. The configuration space is used for the external main CPU to configure *SubJPEG* and to request its computation status. It is operated with bus_clk and V_{DDH}. For each frame, the external main CPU issues a command to the configuration space of the JPEG co-processor. The configuration commands include information such as the source data start address/length, destination data start address, YUV sampling ratio, programmable quantization table coefficients, etc. In our architecture, two command slots are accommodated in the configuration space, so the main CPU can issue a command for the next frame while the co-processor is still processing the current frame. Otherwise the processor must be stalled for hundreds of clock cycles between of two frames and be re-started only when the reconfiguration for the next frame is completed.

The read controller (RDC) works with bus_clk and V_{DDH}. Its main function is to read blocks of source data from standard bus according to the configuration information. A status table is maintained to record the status of the AFIFOs and information of the last block. Once new data coming from the bus has been fed into the AFIFOs, the source data counter will count the incoming data length and will update the AFIFOs' status in the table and also move the head pointer. The RDC issues a data request periodically according to the configured interval time T_0. The requested data length is based on the minimal of the remaining data length (this is initialized as the source data length at start run), maximum bus payload size and AFIFOs' empty size (how many AFIFOs are empty). As soon as the requested data length is calculated, the tail pointer will jump to AFIFO

```
I.  Access configuration space;
II. Initialization:
      1. Request data addr == Source data addr;
      2. Remaining data length == source data length;
III. Start:
    1. Count the coming data length;
    2. Update AFIFO status;
    3. Move head_ptr;
                                          Thread 1: Receive
                                          source data from bus

    1. Request length =
       min(Max bus payload, Remaining data length, empty AFIFOs'size);
    2. Send request (Request data addr, Request length);
    3. Move tail_ptr;
    4. Request data addr = Request data addr + Request length;
    5. Remaining data length = Remaining data length - Request data addr;
    6. if (Remaining data length != 0 && timer ==T0)   go to step III;
    7. else
          { update Last block info status; go to step IV; }
                                          Thread 2: Send data
                                          request to bus
IV. Idle:
      1. if (configuration for next frame is finished)  go to step I;
      2. else go to step IV;
```

Fig. 10. Pseudo code algorithm for RDC.

where the latest requested source data block will be stored. The new requested data address and the remaining data length are also updated. If the remaining data length is zero, meaning that the last requested data block is the ending block of the current frame, the column logging the information of the last block in the status table will be updated. Fig. 10 shows the pseudo code algorithm of the RDC.

The write controller (WRC) works with the Huffman_clk and uses V_{DDH} as power supply. It checks the status of the DCT-Quantization RAM (DQRAM), from each engine, and controls writing data from DQRAMs to the arbitrator. Similar to the RDC, the WRC also maintains a status table to log the DQRAMs' status and the last block information. Once a DQRAM entry of an engine is full, the header pointer will move to the next engine's DQRAM entry and the DQRAMs' status will update. If the entropy encoder is idle, the WRC will indicate the arbitrator to push the data out of an engine's DQRAM. Once the data is completely pushed out, the DQRAM status will be updated and the tail pointer will jump to the next engine's DQRAM entry. In this way the engines' DQRAMs are circulated for writing and reading. Fig. 11 shows the pseudo code algorithm of the WRC.

IV. Implementation and Evaluation

The implemented core is fully compliant with the JPEG encoder baseline standard. Signals across different clock domains are hand-shacked to increase communication robustness. We used a hierarchical logic synthesis approach: the engines are synthesized with a dedicated subthreshold library, as mentioned in Section II. The other blocks are synthesized with a conventional CMOS65 standard cell library. According to synthesis results, the engines and the Huffman encoder can operate easily beyond 250 MHz with a 65 nm $\mathrm{S}V_{\mathrm{T}}$ CMOS process at 1.2 V nominal supply voltage. Some signals in the design have to cross the V_{DDL} and V_{DDH} domains. Therefore, a level shifting scheme is needed. In addition, the digital I/O pads in 65 nm CMOS must use a reference voltage of 1.2 V, so we also need a level shifting scheme to convert the signal level from the *SubJPEG* core to the I/O pads. Shown in Fig. 12 is the 2-stage level shift scheme used in *SubJPEG*. The first stage level shifting is performed through simple buffers which are capable enough of pulling up signals from subthreshold V_{DDL} to V_{DDH}. The difference between V_{DDL} and V_{DDH} is less than 300 mV. The second stage level shifting is performed through positive feedback structured level-shifters from V_{DDH} to 1.2 V I/O pads.

Each engine has its own deep n-well to separate its bulk from the rest of the chip and also has a V_{T} balancer located at one of its corners. Each V_{T} balancer is $25 \times 30\ \mu\mathrm{m}^2$ and the core size is 1.4×1.4 mm^2. The testchip was fabricated using TSMC's 65 nm seven-layer low-power standard V_{T} CMOS process. The core layout and the microphotograph of the prototype chip are shown in Fig. 13. Compared to the baseline processor, the area of *SubJPEG* is about $2.5\times$ larger, including overhead from implementing parallel engines and bulk biasing, etc. The area and simulated energy breakdown in the digital still image mode are shown in Fig. 14. The circuits that are required to parallelize the engines, i.e., dispatcher, RDC, WRC, arbiter and interface AFIFOs, occupy 8% area of the core. For digital still image processing ($V_{\mathrm{DDL}} = 0.4$ V and $V_{\mathrm{DDH}} = 0.5$ V in simulation) and $f_{\mathrm{Huffman\_clk}} = 2f_{\mathrm{bus\_clk}} = 4f_{\mathrm{engine\_clk}}$, these circuits would dissipate approximately 12% of the total energy.

To test the functionality of the chip, a 9-layer PCB was designed. On the board a Xilinx Spartan-3 FPGA chip functions as the main CPU and *SubJPEG* functions as its co-processor. The 1.2 V V_{ref} and 2.5 V I/O voltages are generated with on-board DC-DC converters. The other supply voltages are supplied from external voltage generators.

The measured behavior of the configurable V_{T} balancer at $V_{\mathrm{DD}} = 400$ mV is shown in Fig. 15. An off-chip capacitor is needed to mitigate the ripple. As it can be seen, before the V_{T} balancer is activated, the n-well is connected to V_{DD} and the p-well is connected to GND. Then, within 1 ms after the V_{T} balancer is turned on, the supply voltages of both n-well and p-well converge at near $V_{\mathrm{DD}}/2$. At $V_{\mathrm{DD}} = 400$ mV, the tested samples could not function correctly with a 2 MHz engine_clk frequency without V_{T} balancing. With the help

```
x: Engine index [0,1,2,3]
y: DQRAM entry [0,1]

for ( 0 ≤ y ≤ 1 )
  {
    for (0 ≤ x ≤ 3)
     {

       if (Engine_x.DQRAM_y is full)
         {
           update DQRAM status column in table;
           move head_ptr;
         }
      x++;
     }
    y++;
  }
```

Thread 1: log the status of writing from quantizers to engines' DQRAMs

```
for ( 0 ≤ y ≤ 1 )
  {
    for (0 ≤ x ≤ 3)
     {

       if (Engine_x.DQRAM_y is full && entropy encoder is idle)
         {
           push data from Engine_x.DQRAM_y to arbitrator;
           update DQRAM status column in table;
           move tail_ptr;
         }
      x++;
     }
    y++;
  }
```

Thread 2: control reading data from engines' DQRAMs to arbitrator

Fig. 11. Pseudo code algorithm for WRC.

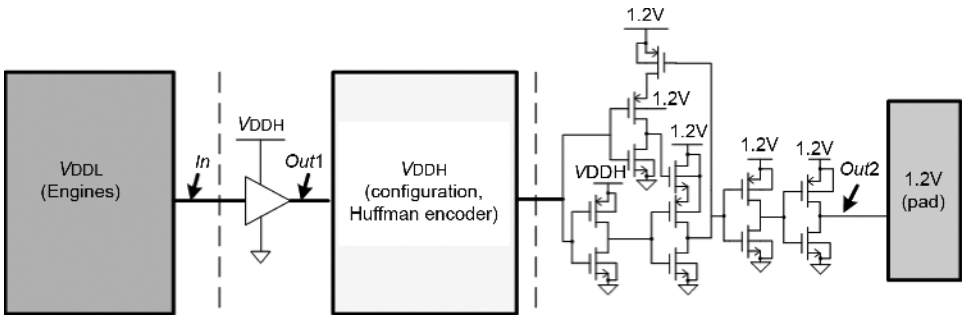

Fig. 12. Two-stage level-shifting scheme in *SubJPEG*.

of V_{T} balancing, the samples could run at 2.5 MHz. In this case, the average leakage current is increased by $2\times$. At this time, the ratio between the leakage and the dynamic energy is about 1/30, meaning that the V_{DD} can still be further reduced to reach V_{opt} which leads to a 1/1 ratio. Unfortunately, we cannot operate the engines with V_{DDL} lower than 0.4 V. This testing limitation is from the lowest V_{DDH} that the second stage level shifters can tolerate. The second stage level shifters function erroneously when V_{DDH} is lower than 0.6 V. This lowest V_{DDH} limitation affects directly the lowest V_{DDL} that the first stage level shifters can handle, in spite of the fact that the engines are likely to function correctly below 0.4 V with a lower frequency. The estimated V_{opt} is around 0.35 V. Fig. 16 shows the transient current at $V_{\mathrm{DDL}} = 0.4$ V, 0.8 V, 1.2 V at an engine_clk of 2.5 MHz, 5 MHz, 10 MHz respectively. Note that 2.5 MHz is the maximum operating frequency at $V_{\mathrm{DDL}} = 0.4$ V supply, but 5 MHz and 10 MHz are not the maximum operating frequencies at $V_{\mathrm{DDL}} = 0.8$ V and $V_{\mathrm{DDL}} = 1.2$ V. Fig. 17 shows the $\mathrm{energy/(engine \cdot cycle)}$ savings. The term $\mathrm{energy/(engine \cdot cycle)}$ denotes the energy consumed per cycle by a single engine. More measurements of system energy and speed performance are summarized in Table IV. In the subthreshold mode the engines can operate with 2.5 MHz frequency at 0.4 V, with 0.75 pJ $\mathrm{energy/(engine \cdot cycle)}$. This leads to $8.3\times$ $\mathrm{energy/(engine \cdot cycle)}$ reduction as compared to operating at 1.2 V nominal supply. Correspondingly, the Huffman coder should be operated at 10 MHz at 0.5 V, with 1.2 pJ per entropy encoding cycle. In the near-threshold mode the engines can operate with 4.5 MHz frequency at 0.45 V, and consume about 1.0 pJ $\mathrm{energy/(engine \cdot cycle)}$. The Huffman

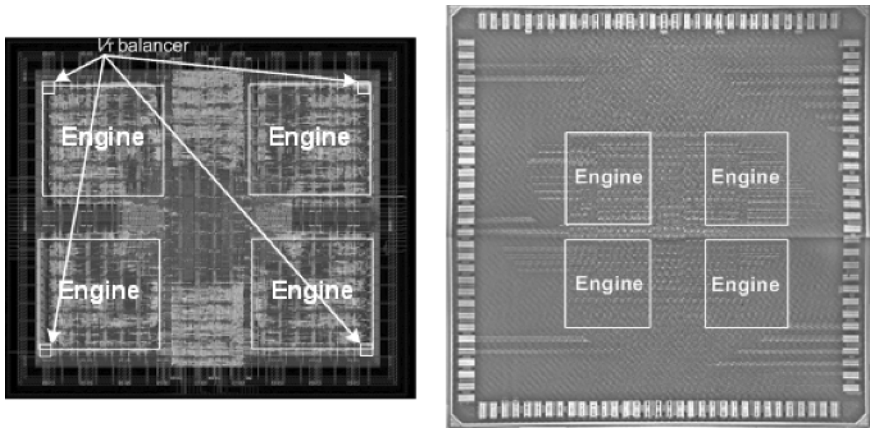

Fig. 13. Core layout and prototype chip microphotograph.

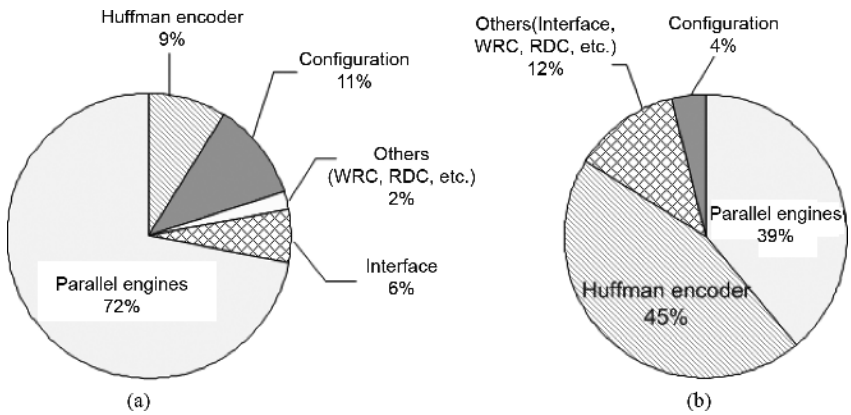

Fig. 14. *SubJPEG* (a) area (b) energy breakdown in digital still image mode.

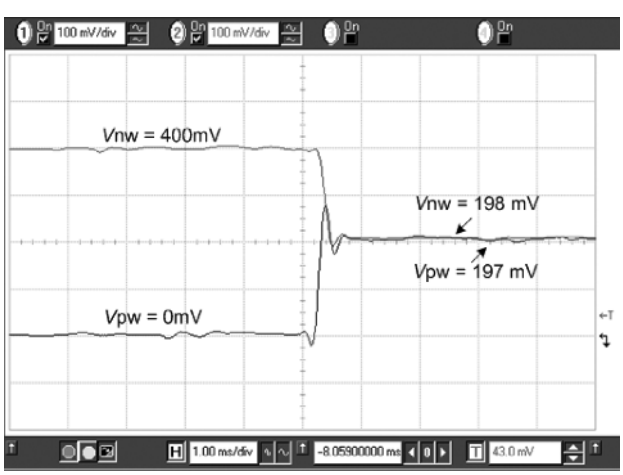

Fig. 15. Measurement results of switching on the V_{T} balancer.

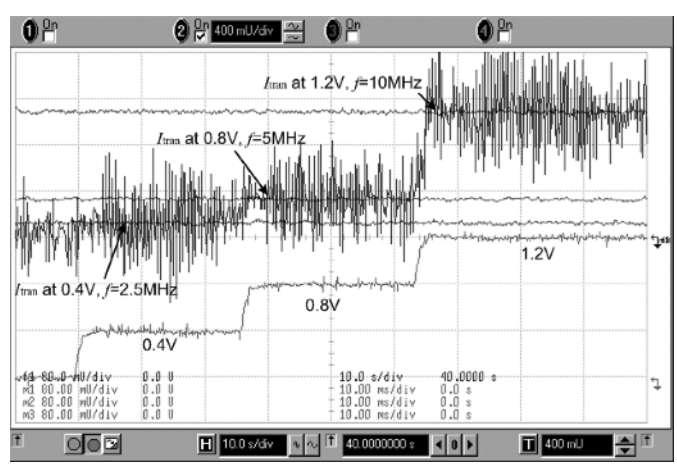

Fig. 16. Transient and average current with $1000\times$ amplified magnitude at (0.4 V, 2.5 MHz), (0.8 V, 5 MHz) and (1.2 V, 10 MHz).

coder operates at 18 MHz frequency with a less than 0.7 V supply, and dissipates around 2.0 pJ per entropy encoding cycle. The overall system throughput meets the 15 fps VGA compression requirement. By further increasing both V_{DDH} and V_{DDL}, and exploring distinct $(V_{\mathrm{DDH}}, V_{\mathrm{DDL}})$ combinations, the prototype chip can achieve multi-standard image encoding.

V. Conclusion

This paper presents our work on exploiting a sub/near threshold supply voltage in the design of ultra low energy and medium throughput (up to 100 MB/s) consumer digital electronic applications. We utilize architecture-level parallelism to compensate for throughput degradation at very low voltage. Several physical-level design techniques were developed to improve circuit robustness. Among them is a configurable V_{T}

TABLE IV
SYSTEM THROUGHPUT AND POSSIBLE IMAGE APPLICATIONS

Engine Mode	V_{DDL}, energy/engine_clk	V_{DDH}, energy/Huffman_clk	Throughput (MB/s), f_{engine_clk}	Possible Applications
Sub-threshold	0.4V, 0.75pJ	0.5V,1.2pJ(estimated) 0.6V, 1.8pJ	10, 2.5MHz	digital still image
Near-threshold	0.45V, 1.0 pJ	<0.7V, 2.0pJ	18, 4.5MHz	VGA (640×480,15fps)
	0.5V, 1.33pJ	0.8V, 3.1pJ	40, 10MHz	VGA (640×480, 30fps)
Super-threshold	0.6V, 1.7pJ	0.9V, 4.3pJ	100, 25MHz	SXGA(1280×1024,15fps)
	0.7V. 2.2pJ	1.0V, 5.8pJ	160, 40MHz	UXGA(1600×1200, 15fps) QXGA(2048×1536, 15fps)

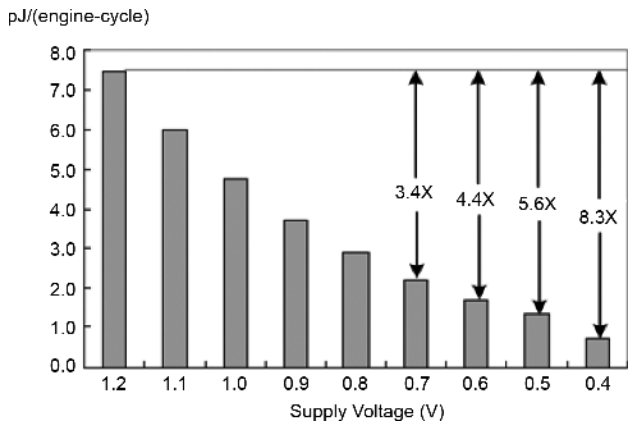

Fig. 17. Energy per operation cycle for each engine $[\mathrm{pJ}/(\mathrm{engine}\cdot\mathrm{cycle})]$.

balancer which is used to mitigate the V_{T} mismatch of nMOS and pMOS transistors in the sub/near threshold at all process corners. Another design technique to improve transistor driving capability in subthreshold was presented as well. This technique exploits V_{T} mismatch between parallelized transistors in the implementation of power switches. In addition, we describe how the "common" standard cells are selected and modified for robust operation. All these ideas are demonstrated using *SubJPEG*, a 1.4×1.4 mm^2 CMOS 65 nm standard V_{T} multi-standard DMA based JPEG co-processor. For DCT and Quantization processing, a single engine in subthreshold mode dissipates only 0.75 pJ of energy with a 0.4 V supply voltage at 2.5 MHz frequency, which leads to $8.3\times$ energy reduction compared to using a 1.2 V nominal supply. In the near-threshold mode it dissipates 1.0 pJ with a supply voltage of 0.45 V at 4.5 MHz frequency, and the system throughput meets 15 fps (640×480 pixel VGA standard). In general, our methodology is largely applicable to designing other sound/graphic and streaming processors.

ACKNOWLEDGMENT

The authors thank Leo Sevat, Maurice Meijer, Cas Groot and Agnese Bargagli-Stoffi, all from NXP Research Eindhoven, for their support during backend and testing of the chip. The authors also thank Leo Warmerdam, also from NXP Research Eindhoven, for funding the project.

REFERENCES

[1] DVFS Design Technology. Samsung. [Online]. Available: http://www.samsung.com/global/business/semiconductor/products/asic/Products_DesignTechnology.html
[2] B. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1778–1786, Sep. 2005.
[3] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "Analysis and mitigation of variability in subthreshold design," in *Proc. IEEE Int. Symp. Low Power Electronics and Design (ISLPED)*, Aug. 2005, pp. 20–25.
[4] J. Keane, H. Eom, T. Kim, S. Sapatnekar, and C. Kim, "Subthreshold logical effort: A systematic framework for optimal subthreshold device sizing," in *Proc. Design Automation Conf. (DAC'06)*, Jul. 2006, pp. 425–428.
[5] B. Calhoun, A. Wang, and A. Chandrakasan, "Device sizing for minimum energy operation in subthreshold circuits," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC'04)*, Oct. 2004, pp. 95–98.
[6] J. Kwong and A. Chandrakasan, "Variation-driven device sizing for minimum energy subthreshold circuits," in *Proc. IEEE Int. Symp. Low Power Electronics and Design (ISLPED)*, Oct. 2006, pp. 8–13.
[7] H. Soeleman and K. Roy, "Ultra-low power digital subthreshold logic circuits," in *Proc. IEEE Int. Symp. Low Power Electronics and Design (ISLPED)*, Aug. 1999, pp. 94–96.
[8] B. Calhoun and A. Chandrakasan, "A 256 kb subthreshold SRAM in 65 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2006, pp. 2592–2601.
[9] J. Chen, L. Clark, and T. Chen, "An ultra-low-power memory with a subthreshold power supply voltage," *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2344–2353, Oct. 2006.
[10] T. Kim, J. Liu, J. Keane, and C. Kim, "A high-density subthreshold SRAM with data-independent bitline leakage and virtual ground replica scheme," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2007, pp. 330–606.
[11] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "A variation-tolerant sub-200 mV 6T subthreshold SRAM," *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2338–2348, Oct. 2008.
[12] N. Verma and A. Chandrakasan, "A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 141–149, Jan. 2008.
[13] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, D. Blaauw, and T. Austinand, "A 2.60 pJ/inst subthreshold sensor processor for optimal energy efficiency," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2006, pp. 154–155.
[14] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, "Exploring variability and performance in a sub-200 mV processor," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 881–891, Apr. 2008.
[15] M. Seok, S. Hanson, Y. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "The phoenix processor: A 30 pW platform for sensor applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2008, pp. 188–189.
[16] A. Wang and A. Chandrakasan, "A 180 mV subthreshold FFT processor using a minimum energy design methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.
[17] V. Sze and A. Chandrakasan, "A 0.4-V UWB baseband processor," in *Proc. IEEE Int. Symp. Low Power Electronics and Design (ISLPED)*, Aug. 2007, pp. 262–267.
[18] M. Hwang, A. Raychowdhury, K. Kim, and K. Roy, "A 85 mV 40 nW process-tolerant subthreshold 8×8 FIR filter in 130 nm technology," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2007, pp. 154–155.

[19] J. Kwong, Y. Ramadass, N. Verma, and A. Chandrakasan, "A 65 nm sub-Vt microcontroller with integrated SRAM and switched capacitor DC-DC converter," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 115–126, Jan. 2009.
[20] Y. Pu, J. P. de Gyvez, H. Corporaal, and Y. Ha, "V_{T} balancing and device sizing towards high yield of subthreshold static logic gates," in *Proc. IEEE Int. Symp. Low Power Electronics and Design (ISLPED)*, Aug. 2007, pp. 355–358.
[21] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. Ellis-Monaghan, and E. J. Nowak, "Low-power CMOS at $\mathrm{Vdd} = 4\,\mathrm{kT/q}$," in *Proc. Device Research Conf.*, Jun. 2001, pp. 22–23.
[22] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.
[23] G. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electron.*, vol. 38, no. 1, pp. XVIII–XXXIV, Feb. 1992.
[24] *Digital Compression and Coding of Continuous Tone Still Images, Part 1, Requirements and Guidelines*, ISO/IEC JTC1 Draft International Standard 10918-1, Nov. 1991.
[25] Y. Pu, J. P. de Gyvez, H. Corporaal, and Y. Ha, "An ultra low-energy/frame multi-standard JPEG co-processor in 65 nm CMOS with sub/near threshold power supply," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2009, pp. 146–147.

Yu Pu (M'09) received the B.S. degree (*cum laude*) in electrical engineering from Zhejiang University, Hangzhou, China, in 2004. In 2009, he received the Ph.D. degree in electrical engineering from the Eindhoven University of Technology, The Netherlands, in association with the National University of Singapore.

From November 2006 to February 2009, he was with the Mixed-Signal Circuit and System Group in NXP Research Eindhoven. From March 2009 to September 2009 he was a research scientist in the Ultra Low-Power DSP Processor Group of IMEC, The Netherlands. He is now with the Sakurai Lab, University of Tokyo, Japan. His research interests focus on ultra low-energy digital circuit design and EDA methodologies.

Jose Pineda de Gyvez (F'09) received the Ph.D. degree from the Eindhoven University of Technology, The Netherlands, in 1991.

From 1991 until 1999 he was a Faculty member in the Department of Electrical Engineering at Texas A&M University. He is currently a Senior Principal at NXP Semiconductors in The Netherlands. Since 2006 he also holds the professorship "Deep Submicron Integration" in the Department of Electrical Engineering at the Eindhoven University of Technology.

Dr. Pineda de Gyvez has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART I and PART II, and also Associate Editor for Technology of the IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. He is also a member of the editorial board of the *Journal of Low Power Electronics*. He has co-authored more than 100 combined publications in the fields of testing, nonlinear circuits, and low power design. He is author or co-author of three books, and holds several granted patents. His work has been acknowledged in academic environments as well as in patent portfolios of many companies. His research has been funded by the Dutch Ministry of Science, U.S. Office of Naval Research, and U.S. National Science Foundation, among others.

Henk Corporaal (M'09) received the M.Sc. degree in theoretical physics from the University of Groningen, and the Ph.D. degree in electrical engineering in the area of computer architecture from Delft University of Technology, The Netherlands.

He has been teaching at several schools for higher education, has been Associate Professor at the Delft University of Technology in the field of computer architecture and code generation, had a joint professor appointment at the National University of Singapore, and has been scientific director of the joint NUS-TUE Design Technology Institute. He also has been department head and chief scientist within the DESICS (Design Technology for Integrated Information and Communication Systems) division at IMEC, Leuven (Belgium). Currently he is a Professor in Embedded System Architectures at the Eindhoven University of Technology (TU/e), The Netherlands. He has co-authored over 250 journal and conference papers in the (multi-)processor architecture and embedded system design area. Furthermore, he invented a new class of VLIW architectures, the Transport Triggered Architectures, which is used in several commercial products, and by many research groups. His current research projects are on multiprocessor architectures and the predictable design of soft and hard real-time embedded systems.

Yajun Ha (SM'09) received the B.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 1996, the M.Eng. degree in electrical engineering from the National University of Singapore (NUS), Singapore, in 1999, and the Ph.D. degree in electrical engineering from Katholieke Universiteit Leuven, Leuven, Belgium, in 2004. Between 1999 and 2004, he did his Ph.D. research project at IMEC, Leuven.

He has been an Assistant Professor in the Department of Electrical and Computer Engineering, NUS, since 2004. His research interests lie in the embedded system architecture and design methodologies, particularly in the area of reconfigurable computing. He holds one U.S. patent and has published more than 50 internationally refereed technical papers in his areas of interest.