Ultralow Power Circuit Design With Subthreshold/Near-Threshold 3-D IC Technologies

Sandeep Kumar Samal, Student Member, IEEE, Yarui Peng, Student Member, IEEE, Mohit Pathak, Student Member, IEEE, and Sung Kyu Lim, Senior Member, IEEE

Abstract—The requirement of ultralow power and energy efficient systems is becoming more and more important with the increase in the use of miniaturized portable devices and unsupervised remote sensor systems. 3-D integration is an emerging technology that helps in reducing footprint as well as power. In this paper, we study in detail the combined benefits of 3-D ICs and low-voltage supply designs to obtain maximum energy efficiency. We implement different types of circuits in conventional 2-D and through-silicon-via-based 3-D designs at different supply voltages varying from nominal to subthreshold voltages. The impact of 3-D integration on these different types of circuits is analyzed. Our study is based on power and energy comparison of full GDSII layouts. Our study confirms that subthreshold/near-threshold circuits indeed offer a few orders of magnitude power versus performance tradeoff with further improvement due to 3-D implementation. In addition, 3-D designs reduce the footprint area up to 78% and wirelength up to 33% compared with the 2-D counterpart for individual design benchmarks. Our studies also show that thermal and IR drop issues are negligible in subthreshold 3-D implementation due to its extreme low-power operation. Finally, we demonstrate the low-power and high-memory bandwidth advantages of many-core 3-D subthreshold circuits.

Index Terms—3-D IC, subthreshold/near-threshold operation, through-silicon via (TSV), ultralow power.

I. INTRODUCTION

One of the most effective ways of reducing the total power consumption of very-large-scale integration circuits is scaling down the supply voltage. Previous works have shown that under optimum power supply voltage, a circuit can attain minimum energy consumption per operation, which is the primary goal for applications that require long battery life [2]. This supply voltage usually falls near the threshold voltage of the transistors. Though the operating frequency reduces significantly, the energy efficiency improves, reducing energy consumption and hence increasing longevity of battery, which is the major requirement for remote and portable systems like sensor networks. Low-voltage and subthreshold operating conditions require the individual gate to be designed in a robust manner, which requires a larger size of the transistors. For large designs, these result in area overhead. Since area is directly proportional to cost, it is desirable to have the same silicon usage but to get low power and high energy efficiency.

3-D ICs is one of the most promising technologies that enables higher integration and further miniaturization with an increase in memory bandwidth for off-chip memories while reducing power dissipation and improving performance. Much shorter interconnects due to die stacking results in lower switching power dissipation compared with long interconnect designs. Memory stacked over logic integrates off-chip memories to on-chip. There is a significant increase in memory bandwidth by moving to 3-D [3]. The overall footprint of the design also significantly reduces further satisfying the needs of miniaturized designs. A very important advantage of 3-D integration is that it helps us in integrating heterogeneous dies that use different technologies or have different supply voltages. The potential benefits of 3-D stacking, if used properly, can help improve circuits and systems significantly.

Through this paper, we propose a new way to meet most of the requirements of low-power miniaturized designs by using subthreshold/near-threshold circuits stacked with memory as well as mixed with memory in the same die. This results in low power and much smaller die footprint with an increased memory bandwidth. The goal is to investigate and quantify the benefits and issues of using 3-D integration of subthreshold/near-threshold circuits using through-silicon vias (TSVs) and identify what kind of ultralow-power circuits can be improved by 3-D design. Our studies are based on full GDSII layouts and signoff analysis of various types of designs which vary from memory dominated to gate and interconnect dominated. The contributions of this paper are as follows.

1) We design and investigate full-chip subthreshold/near-threshold 3-D circuits for various types of large circuits with and without memory. We study the impact of random variation effects for different supply voltages (Section III-C). We also implement special level shifters...
for interfacing logic with memory when they operate at different supply voltages (Section III-D).

2) A full comparison of performance metrics of the 2-D and 3-D designs for different supply voltages varying from nominal to subthreshold voltages has been done analyzing the low supply voltage and 3-D benefits (Section IVA-C).

3) The thermal and \( IR \) drop issues that are generally the major drawbacks of 3-D integration have been shown to have almost no adverse impact on low-voltage designs due to extreme low power. This reduces the intrinsic voltage and temperature changes which are critical at low voltages (Sections IV-D and IV-E).

4) We discuss the impact of circuit type and characteristics on the design approach to get maximum benefits in Section V.

5) We have further demonstrated 3-D IC advantage with a many-core design to show massive saving on input/output power and the increase in memory bandwidth (Section VI).

Section VII summarizes the design lessons and outlines for subthreshold/near-threshold 3-D IC design. Finally, we conclude our work in Section VIII.

II. REVIEW OF EXISTING WORKS

One of the very first works on low-voltage design was published two decades ago. Burr and Shott [4] demonstrate a 625X power-delay product (PDP) saving by using 200-mV supply compared with the then standard 5 V supply. PDP is a key metric to fairly compare the quality of designs at different supply voltages. With advancement of technology, research related to subthreshold/near-threshold design has gained much more momentum. Wang and Chandrakasan [2] designed a fast Fourier transform processor operating at 180 mV. They used a 180-nm process with a transistor threshold voltage of around 450 mV. As a follow-up of this paper, Calhoun et al. [5] studied the techniques for optimum sizing in the subthreshold regime for minimum-energy operation. They solved equations of total energy to provide analytical solutions for minimum-energy operating conditions and also studied gate-sizing impact on energy at subthreshold supplies. For our initial gate design and sizing work, we considered these techniques to obtain the best cell designs with the given process design kit (PDK), new supply voltage, and size constraints.

Hanson et al. [6] published their work on a full low-voltage processor with special techniques in memory design to reduce standby power to picowatt range. In our paper, we use level shifters with a standard 6T static RAM (SRAM) cell to enable memory operation at higher supply voltage while the logic operates at lower voltages. The industry has also shown keen interest in near-threshold designs for obtaining best energy efficiency [7]. Kaul et al. [7] discuss in detail the opportunities and challenges of near-threshold voltage designs that operate at maximum energy efficiency. They give various real design examples and measured values to validate the great potential of near-threshold designs.

Another 2-D IC work on near-threshold operation is a processor that runs at a 1-MHz frequency and a 0.4 V supply voltage and consumes 79 \( \mu \)W [8]. On similar lines, in our work, we broaden our analysis space compared with that in [1] by adding near-threshold voltages (0.6 and 0.8 V) and different types of circuits.

With the advent of 3-D IC technology with TSVs, there has been some important research works on using 3-D implementation to save power and increase memory bandwidth. A memory bandwidth of 63.8GB/s is measured in a two-tier 3-D IC in 130-nm process [3]. Jung et al. [9] demonstrated a 21.2% power reduction in TSV-based 3-D over 2-D by implementing various techniques, such as 3-D floorplanning, block folding, metal layer usage control, and multi-\( V_{th} \) designs. Though their work is focused on supersupertreshold voltage operation with nominal supplies, the techniques developed are important and can be used in low-voltage designs as well for getting further benefits from 3-D IC implementation. The first near-threshold 3-D IC system was designed with 0.65 V for logic and around 0.87 V for SRAM [10]. Fick et al. [10] highlight the feasibility of thermal-constrained 3-D IC designs by combining it with near-threshold architecture that also results in high energy efficiency. We also carry out thermal and \( IR \) drop analysis to demonstrate that, internally, 3-D IC with subthreshold/near-threshold supply voltages has almost zero-temperature increase and a negligible 34-\( \mu \)V max \( IR \) drop from a thermal and power delivery point of view, respectively.

III. LIBRARY DESIGN AND VARIATION STUDY

A. Issues Related to Subthreshold/Near-Threshold Designs

At a very low supply voltage, the logic gates become extremely sensitive to any noise and there may be logic failure. The original cells in a given PDK are designed for nominal operation and may not be well optimized for direct use in a wider range of voltage supplies. These standard cells may not function properly at very low subthreshold/near-threshold supply voltage. The relative strengths of nMOS and pMOS significantly change with change in supply voltage and hence require resizing. The process, voltage, and temperature (PVT) variations increase their relative impact for subthreshold/near-threshold circuits. For a superthreshold design, the \( ON/OFF \) current ratio is typically more than 10\(^3\). Even though variations cause some transistors to be stronger or weaker, the \( ON \)-transistor still overwhelms the \( OFF \)-transistor. However, in the subthreshold region, only the subthreshold leakage current can be used to switch the logic gates. The current relation becomes exponential with respect to threshold voltage, gate voltage, and temperature change. Therefore, any PVT variation may cause current to vary exponentially.

Other problems include memory design and operation. Many standard memory components such as standard 6T SRAM cell cannot work in the subthreshold region since reduced noise margin in subthreshold design causes signal integrity issues. Either the memory cells need to be modified for proper functioning at the energy minimum voltage or other system design techniques have to be used [6], [11].
B. Supply Voltage Selection and Gate Sizing

To minimize process variation effects in our design, we used the GlobalFoundries 130-nm process that has a nominal voltage of 1.5 V. We studied the transistor characteristics around the threshold voltage values ($V_{thN} = 0.53$ V and $V_{thP} = 0.56$ V). From the energy per cycle curve of a 20-inverter chain with supply varying from 0 to 0.6 V, we observed the optimum supply voltage to be around 0.3 V [Fig. 1(b)]. Fig. 1(a) shows the contribution of switching and leakage energies at a 20% switching activity. Fig. 1(b) compares the effect of switching activity on the total energy per cycle. As switching activity increases, the optimum supply voltage goes into a deeper subthreshold region. This is because the switching energy starts dominating an earlier voltage compared with leakage energy. However, for reliable operation of larger cells like Dflip-flop, we set our minimum supply voltage at 0.4 V after careful SPICE simulations.

It is to be noted that the 0.3 V optimum supply is for an inverter chain, which is a logic-only design. For actual designs with memory, the leakage component of energy increases significantly, and therefore, at lower voltages, a slower clock results in the leakage power dissipating for a longer period and dominating the overall energy dissipation. For designs with a high-leakage component, the optimum voltage for maximum energy efficiency increases. This is one of the major components of our study with a comparison between 2-D and 3-D ICs at different supply voltages and different types of designs. The two other supply voltages different from nominal supply of 1.5 V are chosen to be 0.6 and 0.8 V. The supply voltage of 0.6 V is chosen for a detailed study at near-threshold voltage operation. It has been shown that near-threshold voltage operation is more energy efficient than subthreshold operation [7]. 0.6 V is around 10% greater than the transistor threshold voltages in this 130-nm process. Memory in all designs except the nominal (1.5 V) design is operated at 0.8 V to ensure reliable operation. Therefore, we choose 0.8 V as another supply voltage for logic. In this case, no level shifters and multiple power domains will be required for logic and memory. Overall, we have four different supply voltages varying from nominal (1.5 V) to subthreshold (0.4 V).

The logic gates are sized for each supply voltage, respectively, to equalize the strength of nMOS and pMOS and reduce propagation delay mismatch [5]. While sizing standard cells, we use a load capacitance value of 16 fF and input rise/fall time of 2% of the input switching period. Default layouts are used for 1.5 V operation. Fig. 2 elaborates the necessity of resizing of the gates and the impact of resizing on propagation delay mismatch. Fig. 2(a) and (c) shows the rise and fall delay mismatch when operating default designs of inverter and Dflip-flop, respectively, at 0.4 V for different load and input transition times. On proper gate sizing, the mismatch reduces by a large amount across the entire loads and transition times [Fig. 2(b) and (d)]. Therefore, each gate has four implementations with different transistor sizes for each supply voltage. The boundary areas of the cells are kept the same as in a nominal cell to have a fair total chip area comparison at different supplies. The wider standard cells are used as per the need of subthreshold operation.

As discussed above, a good design and sizing of cells for proper operation at low voltages is in itself a huge task. The purpose of our study here is to study the impact of 3-D IC implementation of low-voltage designs. Therefore, we confine our standard cell library to the basic cells only, viz., inverter, NAND, NOR, buffer, Dflip-flop, and multiplexer, and avoid sizing of all gates for subthreshold/near-threshold operations. We use the same set of cells resized for propagation delay matching and characterized at respective supply voltages. The cell power consumption and delay comparison corresponding to intermediate load and respective input transition for a 1.5 V nominal voltage and a 0.4 V subthreshold voltage for the respective sized cells is shown in Table I.

C. Variation Study

Since low-voltage operation is highly sensitive to variation in supply and temperature, we study the standard cell library behavior with changes in these parameters and carry out Monte Carlo simulations to get a better picture of the impact
of variation. We choose the inverter as our target cell and carry out the relevant simulations. We carry out Monte Carlo simulation with 10000 runs. A Gaussian distribution for threshold voltage and transistor dimensions is used with a standard deviation of 50 mV for threshold voltage variation and 10 nm for a dimensional variation and all the variations are random. Fig. 3 shows the inverter layouts at different voltages with major differences highlighted. While there is a small change in 0.6 and 0.8 V inverter nMOS sizes (highlighted by double-headed arrows in Fig. 3) compared with that in the default 1.5 V layout provided in the original library, the modification in the 0.4 V design is quite large. As supply goes below threshold voltage of transistors, major resizing is necessary to match the rise and fall delay. The top right of Fig. 3 shows the inverter layout at 0.4 V operation and one pMOS transistor in the pull-up network is removed (highlighted by black circle) compared with the inverter layout at an original 1.5 V cell operation in order to reduce the delay mismatch [Fig. 2(a) and (b)].

The histograms in Fig. 3 show the normalized average delay distribution of the inverters at the respective supply voltages. The mean values are also specified to have an absolute comparison of the speed at different voltages. As we move to low voltages near and below the threshold voltage of transistors, the current varies exponentially with the change in threshold voltage. At a subthreshold voltage of 0.4 V, the OFF current is responsible for switching of transistors and the variation is high. There is a longer tail in the distribution of propagation delay at a 0.4 V supply. The 0.4 V inverter has only one pMOS transistor in its pull-up network and two nMOS transistors in the pull-down network, while the cell designs at other voltages have two transistors each in both pull-up and pull-down networks. Therefore, the average of random variation effects is lesser for the 0.4 V cell compared with cells operating at other voltages.

The relative variations at 0.8 V are just slightly more than 1.5 V though the supply voltage is different. The reason is that we use a 130-nm process, and hence there is almost no impact of supply voltage on threshold voltage change, i.e., negligible drain-induced barrier lowering, which is a dominant short-channel effect.

The performance changes of the subthreshold inverter with PVT variations are shown in Table II. These results are obtained from SPICE simulation using 10-µs input and 16-fF load at the output.

**D. Memory Designs**

For practical applications, the requirement of memory is unavoidable. There exist a number of works on subthreshold memory design [6], [11]. However, we used commercial memory compilers to generate large-memory macros at nominal supply voltages. For smaller memory sizes less than 1 kB, we designed register files using our standard cells at the respective operating voltage to have very low power consumption. We reduce the supply voltage of memory to 0.8 V, which is determined from SPICE simulations as the minimum voltage for reliable 6T cell SRAM operation for the same 130-nm process. The memory libraries are also scaled as per the scaling factors obtained from SPICE simulations at a reduced voltage supply of 0.8 V. However, there is a requirement of multiple voltage supplies of 0.4/0.6 and 0.8 V for logic and memory, respectively, for the whole system design at those logic supplies. 3-D helps us in this respect as we use memory macros in dies separate from logic die, and hence, they can have dedicated power supply without additional circuits.

Another critical issue is signal interfacing between logic and memory that require the use of level shifters. As the voltage has to be raised from a subthreshold/near-threshold voltage to a high voltage, we use a modified level shifter circuit as shown in Fig. 4(a). Since the input is subthreshold/near-threshold, the transistors are not strongly driven and the output may not change. Therefore, we add transistors in diode connection to control the pull-up current. The diode-connected transistors help to drive a larger current to switch the output successfully.
The sizing of the transistors in this circuit has been done to satisfy the requirements and verified with SPICE simulations shown in Fig. 4(b).

**IV. FULL-CHIP DESIGN AND ANALYSIS**

**A. Design Flow**

We study the impact of 3-D designs and low voltages for various types of circuits. The different test-cases we use are the 8052 microcontroller, leon3 processor, and low-density parity check (LDPC) circuit. The rationale behind choosing these circuits is to have different kinds of test cases. While LDPC is pure logic design with a very high degree of interconnection, a four-core leon3 is heavily dominated by memory. The 8052 microcontroller uses an internal RAM of 256 bytes and an external RAM up to a maximum of 64 kB. It also has a ROM with size up to 64 kB. We implement two versions of 8052 with 16- and 64-kB external memory, respectively.

All these designs are implemented in 2-D and 3-D. Memory is always operated at 0.8 V except for the nominal 1.5 V design when it is operated at 1.5 V itself.

We use Synopsys Design Compiler to obtain the netlist from the RTL and then use Cadence SoC Encounter to do the full-chip layout. Synopsys Primetime is used for timing and power analysis for 3-D design. The different test cases are designed in different ways. For 8052 designs, based on the area of logic and memory, we completely separate both into different tiers for the 3-D design. Since the system has input and output pins that are connected to the logic portion only, we put the logic and internal RAM on the bottom die (die0) and the external memory are stacked on top. Each 16-kB external RAM along with 16-kB external ROM makes one die.
We use 60 TSVs of 5-µm diameter to connect between two dies and we set the TSV pitch to be 10 µm. The TSV parasitics are calculated accordingly for 3-D TSV first approach [12]. For the leon3 and LDPC designs, we first carry out multiway partitioning of the netlist using flare [13]. Then TSV insertion with true 3-D placement is carried out using a 3-D placer [14]. In four-core leon3 designs, the memory modules are spread in all the four tiers at the same 2-D locations and they do not have any dedicated tier. For the LDPC test case, a two-way partitioning followed by the 3-D cell and TSV placement is done. The full-chip layouts for 64-kB 8052 designs are shown in Fig. 5.

We build 2-D and 3-D designs for all these test cases at four different supply voltages to compare the subthreshold/near-threshold and 3-D impact on the power, energy, and footprint area. The design specifications along with the number of tiers and total TSV count in 3-D implementation are listed in Table III. The wirelength for 0.4 V designs are also reported. We observe that by implementing the designs in 3-D, we obtain up to a 78% reduction in footprint area (for 64 kB 8052). The greater the number of tiers in 3-D, the greater the footprint area improvement. Because of the footprint reduction, the wires that are needed to connect the blocks in 8052 designs are shortened, and this results in reduction in the top-level interconnect wirelength by 33% in the 64-kB design. The wirelength saving is small in the 16-kB design due to very few 3-D connections. The wirelength increase gets reduced in the 3-D four-core leon3 design because memory dominates the design and 3-D folding with a balanced area gives lesser benefits. However, in LDPC design, which is heavily interconnect dominated, we see up to a 12% wirelength reduction by going to 3-D.

### B. Timing and Power Comparisons

Table IV shows the detailed results of the different implementations at different supply voltages. We analyze each of the test cases individually and then try to summarize how the extent of 3-D advantages depends on the type of design. The clock periods are kept the same for 2-D and 3-D designs in each case to have an isoperformance comparison. The total power reported is the sum of the internal, switching, and leakage powers. The memory power has been reported independently.

In 8052 designs, the clock frequency is up to 66.7 MHz for superthreshold design. Therefore, the internal power and the switching power are the main part of the total power consumption. Since both the logic and memory are under the same supply voltage and the switching activity for memory is very low, the memory power is not a big portion of total power. By reducing the supply voltage and going to subthreshold computing, the design with a 16-kB external memory shows a 9099 times reduction in total power at a cost of 3333 times lower clock frequency. For certain low-power applications like sensors, the workload for each node is not heavy, and the performance of each computing node is not the major concern in most cases. Therefore, by reducing the supply voltage we can reduce the power and energy per cycle and ensure a longer battery life.

Since the external memory for subthreshold designs is working at 0.8 V, the larger the size of memory, the more the leakage is, as is clear from the results. By reducing the clock frequency and supply voltage, we can achieve a significant reduction in internal power and switching power. In the 64-kB external memory design, the internal and switching powers are reduced by almost 35 000 times by changing from superthreshold to subthreshold logic. But the leakage power is not directly related to clock frequency change, so we achieve a 6.49-fold saving in leakage power. The larger the memory size, the smaller the total power saving. The 64-kB external memory design shows a 3008 times overall power reduction in contrast
TABLE IV

<table>
<thead>
<tr>
<th></th>
<th>1.5 V</th>
<th>0.8 V</th>
<th>0.6 V</th>
<th>0.4 V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2D</td>
<td>3D</td>
<td>2D</td>
<td>3D</td>
</tr>
<tr>
<td><strong>8052 with 16kB memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target clock (ns)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal power (mW)</td>
<td>4.713</td>
<td>4.589</td>
<td>0.1843</td>
<td>0.1775</td>
</tr>
<tr>
<td>Leakage power (mW)</td>
<td>0.00959</td>
<td>0.00959</td>
<td>0.00098</td>
<td>0.00098</td>
</tr>
<tr>
<td>Total power (mW)</td>
<td>10.6</td>
<td>10.3</td>
<td>0.3841</td>
<td>0.3616</td>
</tr>
<tr>
<td>Power Delay Product (ps)</td>
<td>159</td>
<td>154.5</td>
<td>57.62</td>
<td>54.24</td>
</tr>
<tr>
<td>Memory power (mW)</td>
<td>0.7145</td>
<td>0.7141</td>
<td>0.0225</td>
<td>0.0216</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>8052 with 64kB memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target clock (ns)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal power (mW)</td>
<td>6.214</td>
<td>5.30</td>
<td>0.249</td>
<td>0.2398</td>
</tr>
<tr>
<td>Leakage power (mW)</td>
<td>4.799</td>
<td>5.09</td>
<td>0.199</td>
<td>0.186</td>
</tr>
<tr>
<td>Total power (mW)</td>
<td>0.02165</td>
<td>0.021</td>
<td>0.00346</td>
<td>0.00346</td>
</tr>
<tr>
<td>Power Delay Product (ps)</td>
<td>11.03</td>
<td>11.01</td>
<td>0.4515</td>
<td>0.4293</td>
</tr>
<tr>
<td>Memory power (mW)</td>
<td>0.7865</td>
<td>0.7841</td>
<td>0.08995</td>
<td>0.08635</td>
</tr>
</tbody>
</table>

Leon3

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>100</th>
<th>1000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target clock (ns)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal power (mW)</td>
<td>237</td>
<td>239.8</td>
<td>7.265</td>
<td>7.269</td>
</tr>
<tr>
<td>Leakage power (mW)</td>
<td>54.4</td>
<td>55.1</td>
<td>1.317</td>
<td>1.317</td>
</tr>
<tr>
<td>Total power (mW)</td>
<td>0.009</td>
<td>0.009</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>Power Delay Product (ps)</td>
<td>294.7</td>
<td>292.97</td>
<td>8.592</td>
<td>8.596</td>
</tr>
<tr>
<td>Memory power (mW)</td>
<td>0.7865</td>
<td>0.7841</td>
<td>0.08995</td>
<td>0.08635</td>
</tr>
</tbody>
</table>

LDPC

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>400</th>
<th>4000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target clock (ns)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal power (mW)</td>
<td>37.33</td>
<td>34.43</td>
<td>0.413</td>
<td>0.474</td>
</tr>
<tr>
<td>Leakage power (mW)</td>
<td>216.9</td>
<td>199.1</td>
<td>1.463</td>
<td>1.247</td>
</tr>
<tr>
<td>Total power (mW)</td>
<td>0.0034</td>
<td>0.0029</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>Power Delay Product (ps)</td>
<td>254.5</td>
<td>233.5</td>
<td>1.876</td>
<td>1.721</td>
</tr>
</tbody>
</table>

with the 9099 times power saving for a 16-kB external memory design. The internal power and switching power is reduced by 1.5% and 2.5%, respectively, in the subthreshold 3-D design with a 64-kB memory compared with the subthreshold 2-D design, and the total power is reduced only a little because the leakage power remains almost the same and it is the dominating part in total power. The best PDP is obtained at 0.6 V because memory leakage starts to dominate at 0.4 V. We can see from Table IV that degradation in PDP is more severe for 64-kB designs compared with that for 16-kB designs as there is more memory contributing more leakage.

The four-core Leon3 has huge memory content. Memory power is more than 60% of the total power for the superthreshold 1.5 V design. Therefore, even though there is some minor improvement in switching power in 3-D implementation, the overall benefit is almost negligible. We also know that leakage power does not scale proportional to switching power. Therefore, as we reduce the supply voltage, dominance of memory increases further. Once again best PDP is obtained at 0.6 V, which agrees with earlier studies of full processors [7].

For the LDPC design, we do not have any memory but only logic. In addition, the design is highly interconnected and hence we see a significant reduction in switching power in the 3-D designs (Table IV). Since switching power is a major portion (85%) of the total power, we observe a 9% reduction in total power, which is a direct consequence of 10% switching power saving. It is interesting to observe that the best PDP for this design is obtained at 0.4 V unlike the other test cases where it was at 0.6 V. This follows directly from the fact that leakage is negligible in LDPC due to absence of memory, and hence, we have a similar minimum energy point as in the 20-inverter chain evaluated in Section III. More discussion is presented in Section V.

C. Full-Chip Variation Study

Since subthreshold operation is highly sensitive to PVT variations, we analyze the 8052 subthreshold (0.4 V) design at different process corners and with temperature and voltage variations. The results for only-logic variations are shown in Table V. Only-memory process variation effects are shown in Table VI. We observe that the design becomes faster at higher temperatures unlike standard superthreshold circuits and consumes higher power. This is because subthreshold current exponentially increases with increase in temperature. The other variations affect the design performance and power consumption as expected. Since the critical path in our analysis is only through logic, variations in memory do not affect the timing performance.

D. Thermal Analysis

Heat dissipation is always a major concern for stacked dies. Also, the performances of low-voltage cells especially the subthreshold ones are very sensitive to temperature. Therefore,
we need to carefully simulate the thermal effects on our 3-D designs. Since current tools cannot handle 3-D designs properly, we use our in-house tools to build a thermal model for 3-D IC and perform thermal simulation using ANSYS Fluent [15]. In this simulation, we assume adiabatic boundary conditions on all the four sides and the bottom side of the package, and the top side of the package is directly in contact with static air without any heat sink. The ambient temperature is 25 °C.

The temperature map for logic die for the 8052 design with a 16-kB memory is shown in Fig. 6. Since the memory power is only 7.3% of the total power in superthreshold design, the temperature of the chip is mainly determined by the blocks on the bottom die. Therefore, the center of the blocks will usually have the highest temperature within that block. In addition, TSVs are made of copper, which has the highest thermal conductivity among all the materials on the chip and therefore, it provides an easy heat transfer path from the bottom die to the top die. As a result, the temperature is relatively low around the TSV arrays, and that area becomes the coolest part of the full-chip. By lowering the voltage supply and performing subthreshold computing, the power density on each die is significantly reduced. As a result, the maximum temperature increase from ambient temperature within the chip is reduced from 72.964 °C in the superthreshold design to negligible 0.008 °C in the subthreshold design (Table VII).

We carried out static thermal analysis without considering leakage positive feedback mechanism where an increase in temperature increases leakage power, which in turn further increases temperature. As a result, the temperature rise after the first iteration is negligible in the subthreshold design and therefore will not have a huge impact on leakage. The feedback needs to be considered in superthreshold designs for accurate temperature estimation, but since leakage contribution is extremely low in total power at 1.5 V supply (Table IV), the feedback impact can be ignored. From the results, we can conclude that by stacking subthreshold circuits in 3-D, we do not encounter serious thermal problems from within the chip. However, there will be external temperature effects on performance.

**E. IR-Drop Analysis**

We have shown that the performance of standard cells, and therefore, the full design is highly affected by the supply voltage variation. Even though the external supply may not vary, the internal IR drop may result in reduced supplies to certain logic gates. To study this effect, we use a very simple power distribution network (PDN) for the 8052 design and analyze the static IR drop for subthreshold operation. The IR drop issues will mainly affect the logic portion operating at a subthreshold of 0.4 V because the memory-containing dies are operating at 0.8 V and have a dedicated PDN. We use simple minimum PDN for top level with rings only at the die boundary in the top metal layers and use the Metal1 VDD and VSS rails to connect to the individual cells. The power supply bump locations for IR drop analysis are set at the four corners of the dies at the power ring intersections. We use Cadence VoltageStorm for static IR drop analysis.

Detailed placement and layout information is used for IR drop analysis to ensure exact calculations even within the hard blocks. Fig. 7 shows the logic tier IR drop maps for both superthreshold and subthreshold designs. The individual cell power consumption is obtained from Primetime simulations. We scale up the initial power consumption of each cell by
a factor of 10000 to obtain accurate results of every minor IR drop and then scale down the voltage drop values back to original after obtaining the results from analysis. We observe that even for a minimal PDN design, the maximum static IR drop is only 0.34 µV in subthreshold design (Table VII). The values are so small because the current drawn by each cell from the supply is subthreshold current and hence very small. Therefore, we can conclude that IR drop is not an issue for our subthreshold 3-D design.

V. IMPACT OF CIRCUIT CHARACTERISTICS ON DESIGN APPROACH TO MAXIMIZE BENEFITS

In this section, we discuss the dependency of maximum energy efficiency on the choice of supply voltage on the 3-D IC versus 2-D IC implementation using our results and observations.

A. Choice of Operating Voltage

As observed from the results in Table IV, the four-coreleon3 and 8052 designs show minimum PDP at 0.6 V, which is above the threshold voltage of transistors. The high contribution of memory leakage power to total power results in the voltage supply for minimum energy increasing from deep subthreshold voltages. Memory leakage power is lowered at low voltages compared with higher voltages. But it heavily dominates dynamic power and flows for a longer clock period and therefore, leakage energy dissipation exceeds the dynamic energy saving. On the other hand, LDPC benchmark has no memory and maximum energy efficiency is observed at 0.4 V. The efficiency will actually increase by further lowering voltage as demonstrated in Fig. 1 but it is difficult to ensure reliable operation at such low voltages.

Therefore, for the best low-power, high energy efficiency design, it is necessary to identify and sort the blocks based on the percentage of memory in them and then use multivoltage design approach to have separate voltage supplies for memory-dominated designs and logic-dominated designs, respectively.

This will require extra physical design effort with multiple voltage islands and use of level-shifters. In the following section, we discuss how 3-D IC implementation can help reduce some physical design effort in addition to offering other benefits.

B. 3-D IC Over 2-D IC

3-D IC implementation provides many benefits over 2-D IC in terms of footprint, wirelength, power, and physical design effort for multivoltage design. The use of TSVs helps reduce long 2-D interconnects to short 3-D connections resulting in good savings on wirelength and hence switching power. The LDPC benchmark which is interconnect dominated shows up to a 10% improvement in switching power, which is a direct consequence of 12% reduced wirelength. In addition to that, it has been demonstrated that buffer usage is reduced to a good extent in 3-D IC implementation [9]. This helps in saving more dynamic power and hence energy. Therefore, 3-D IC helps in the overall objective of low-voltage designs, which is reducing energy dissipation. However, 3-D implementation has much less power improvement for memory-dominated designs. We do not fold individual memory modules but just partition the different memory macros to be placed as a single block in a die. Memory leakage power dominates total power, and switching power savings do not have much impact on total power.

The use of separate tiers for separate voltage domains simplifies the PDN design where we will not require different voltage islands on the same die (8052 design case). Each die can have its own dedicated supply connection with the far tiers getting power through power/ground TSVs. While this is a good benefit of using 3-D IC, it is not always feasible to have exclusively all memory on single dies if the design is heavily dominated by memory (leon3), as there will be wastage of silicon area in the logic tier and unnecessary increase in the number of TSVs. This necessitates the requirement of good mixed 3-D placement that can simultaneously handle logic cells and memory macros and place them in a 3-D fashion. Bad placement can result in the degradation of power with unnecessary long wires and hence poor design quality. Though, we still need to have voltage islands like 2-D IC, we can use 3-D IC to save power with good partitioning and placement.

Another advantage that 3-D IC provides is footprint area reduction. Subthreshold/near-threshold applications with high energy efficiency requirement includes ultralow power remote applications which also need to be made with minimum footprint area, and therefore, 3-D IC helps in that respect. By stacking multiple tiers over one another, we were able to reduce the footprint by 78% for the used benchmarks. Not only that, 3-D IC will also enable stacking of dies from different technology nodes provided the 3-D connections are properly placed and designed.

VI. POWER BENEFITS IN MANY-CORE DESIGNS

The power consumption discussed so far does not include the I/O power. If we have an OFF-chip memory, the number of I/O pads will limit the bandwidth of memory access. As the
I/O pads consume a huge amount of power, their count usually has an upper bound. However, when we use 3-D integration, we can integrate off-chip memory to on-chip and get rid of the processor to memory I/O pads. We not only reduce the power consumption but also increase the memory bandwidth close to theoretical maximum.

A. I/O Driver Design

The I/O pads provided in the standard library are large with complicated circuits in them and therefore consume a large amount of power and cannot be used for subthreshold circuits. They are meant for high performance in standard superthreshold circuits. Therefore, to have a reasonable quantitative power analysis of many-core implementation, we design our own I/O pads using level shifter and buffers to drive a large load. We exclude electrostatic discharge (ESD) and other related circuits from our simple design. The purpose of our study here is to show the power benefits obtained by 3-D IC implementation by removing logic to memory I/O pads. The addition of ESD circuit in such I/O pads properly designed for low-voltage operation will only add further to 2-D IC power and therefore reinforce the 3-D benefits.

The representative diagram for the output pad without any ESD is shown in Fig. 8. We use a capacitive load of 5 pf to size the large buffer with SPICE simulations. The large load is representative of the pin capacitance and interconnect capacitance between the processor output pad to memory input pad for the off-chip design. Level shifters are required as the processor output is 0.4 V while the memory operates at 0.8 V. We set the I/O supply voltage as 0.8 V. The total power dissipated by a single I/O pad is calculated from SPICE simulations to be 1.066 µW at 20 kHz with a 5-pF load and a 1-µs input slew. This is inclusive of switching power, cell power, and leakage power.

B. Power Saving in Many-Core Sub-\(V_{th}\) 3-D Designs

Using 3-D implementation for the microcontroller design helps us put together many processors as per requirement in reduced area with reduced interconnect and hence lesser power. For our study, we used 128 cores in 2-D and 3-D designs and analyzed the area, wirelength, processor to memory I/O power, and the memory bandwidth. We use five-tier stacking for the 3-D design. The comparison results are shown in Table VIII. We use a two-channel off-chip 2-D memory as our baseline for comparison and then analyze the benefits of the 3-D design.

We observe that the reduction in wire power is proportional to the reduction in wirelength. As we move to advanced process nodes, the wirelength parasitic will become more critical and contribute significantly to power, and therefore, wire reduction is an important benefit offered by 3-D ICs. In addition, the processor to memory I/O power, which is 18.5% of the total power in the 2-D design, is completely removed in the 3-D implementation. 3-D ICs use TSVs and μ-bumps to connect logic and memory that consume negligible power compared with the power hungry I/O pads. In addition, the theoretical maximum bandwidth increases from 16 bits/cycle for 2-D many-core to 1024 bits/cycle for 3-D many-core compared with 8 bits/cycle for single core.

VII. DESIGN LESSONS AND GUIDELINES

The design of low-power subthreshold/near-threshold gates involves a number of issues which need to be taken care of. Wider cells need to be used for low-voltage operation for better performance. We need to take care of proper functionality of all the gates because the energy-minimum supply voltage for gates may occur at the deep subthreshold region. The D-flip-flop is the critical cell in our case whose correct operation determines the minimum supply voltage of 0.4 V.

In general, for the 3-D IC design, we need to be very careful about the thermal and IR drop variations that the chip may encounter during operation. However, the internal thermal and static IR drop is negligible for 3-D subthreshold designs due to extreme low power, and therefore, preplanning of thermal budget and power delivery resources is not necessary in the subthreshold 3-D IC physical design.

In the 3-D subthreshold/near-threshold design, it is preferable to use different supply voltages in different dies in case we have a multiple-supply design. This facilitates dedicated power supply to the dies without major design overhead and issues. However, we need to design good level shifter circuits for proper interfacing of signals with different voltage values and without any large delays. Signals in different voltage domains communicate TSVs. All TSV parasitics need to be taken into account during full-chip timing and power analysis as they are a nonnegligible portion of total power dissipation, especially when the number of 3-D connections is very high.

The impact of circuit type on design approach for maximum benefits has already been discussed in detail in Section V. In our many-core design demonstration, we ignored ESD circuits for simplification and estimation of power savings from logic to memory I/Os, but I/Os are still necessary for external package connections. But it is important to modify these I/O pads used for low-voltage circuits to consume less power. In addition, circuit techniques like power

---

**Table VIII**

<table>
<thead>
<tr>
<th>Property</th>
<th>2D</th>
<th>3D (5-tier)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footprint Area (cm x cm)</td>
<td>2 x 2.4 (100%)</td>
<td>1.12 x 1.2 (28%)</td>
</tr>
<tr>
<td>Wirelength (m)</td>
<td>54.70 (100%)</td>
<td>23.94 (44%)</td>
</tr>
<tr>
<td>Memory I/O Power (µW)</td>
<td>104.5 (100%)</td>
<td>0.768 (0.007%)</td>
</tr>
<tr>
<td>Total Power (µW)</td>
<td>564.5 (100%)</td>
<td>460.8 (81.5%)</td>
</tr>
<tr>
<td>Data Connections</td>
<td>48 (I/O)</td>
<td>3072 (TSV)</td>
</tr>
<tr>
<td>Total Connections</td>
<td>98 (I/O)</td>
<td>3108 (TSV)</td>
</tr>
<tr>
<td>Memory Bandwidth (bits/cycle)</td>
<td>16 (100%)</td>
<td>1024 (6400%)</td>
</tr>
</tbody>
</table>

---

Fig. 8. I/O circuit for logic to off-chip memory connections in a many core 2-D subthreshold design.
gating, adaptive body biasing, or specific custom design of subthreshold/near-threshold memory cells provide very good options to reduce power and improve performance.

VIII. CONCLUSION

In this paper, we explore and quantify the 3-D IC benefits in ultralow-power designs using subthreshold/near-threshold circuits of different kinds. 3-D IC implementation helps in further saving switching power by reducing wirelength. While logic circuits show an excellent reduction in power consumption with good PDP improvement, memory contributes to maximum power in designs with memory because of its near-threshold region of operation and high leakage. The larger the memory in a design, the lesser the power savings. We also carried out detailed variation studies at cell level as well as full-chip level. We showed that 3-D IC implementation of subthreshold/near-threshold circuits has negligible internal thermal and IR-drop-related issues. In Section VI, we also demonstrated the idea of many-core designs with increase in memory bandwidth and further reduction in power consumption due to the removal of processor to memory I/O pads. Therefore, 3-D stacked subthreshold/near-threshold circuits with proper memory design approach enable a major improvement for both ultralow-power operation and miniaturization in processors.

REFERENCES


Sanddeep Kumar Samal (S’12) received the B.Tech. degree in electronics and electrical communication engineering from IIT Kharagpur, Kharagpur, India, in 2012, and the M.S. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2013, where he is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering.

His current research interests include 3-D integrated circuit technology.

Sung Kkyu Lim (S’04–M’00–SM’09–F’14) received the B.S., M.S., and Ph.D. degrees from the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA, in 1994, 1997, and 2000, respectively.

He joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2001, where he is currently a Professor. He led the Cross-Center Theme on 3-D Integration for the Focus Center Research Program with Semiconductor Research Corporation, Durham, NC, USA, from 2010 to 2012. He has authored the books entitled Practical Problems in Very Large Scale Integration Systems Physical Design Automation (Springer, 2008) and Design for High Performance, Low Power, and Reliable 3-D Integrated Circuits (Springer, 2013). His current research interests include architecture, design, test, and EDA solutions for 3-D integrated circuits.

Dr. Lim was a recipient of the National Science Foundation Faculty Early Career Development Award in 2006, and the ACM Special Interest Group on Design Automation (SIGDA) Distinguished Service Award in 2008. He received the best paper award from the Asian Test Symposium in 2012. His work was nominated for the best paper award at the International Symposium on Physical Design in 2006 and 2014, the International Conference on Computer-Aided Design in 2009, the ACM Custom Integrated Circuits Conference in 2010, the Design Automation Conference in 2011, 2012, and 2014, and the International Symposium on Low Power Electronics and Design in 2012. He was on the Advisory Board of the ACM SIGDA from 2003 to 2008. His research was featured as the Research Highlight in the Communication of the ACM in 2014. He was an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS from 2007 to 2009. Since 2013.

Mohit Pathak (S’05) received the B.Tech. degree in computer science and engineering from IIT Kharagpur, Kharagpur, India, in 2004. He is currently pursuing the Ph.D. degree with the Georgia Institute of Technology, Atlanta, GA, USA.

He was with Magma Design Automation, Noida, India, for a year, as a Technical Staff Member. He is a Graduate Research Assistant with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He is also with Cadence Design Systems, San Jose, CA, USA. His current research interests include 3-D integrated circuits (ICs) physical and CAD design, including the extraction of parasitics and optimization for signal integrity, thermal, and power delivery issues in 3-D ICs.

Mr. Peng was a recipient of the Best-in-Session Award at SRC TECHCON in 2014.