# Design and Analysis of Ultra Low Power Processors Using Sub/Near-Threshold 3D Stacked ICs

Sandeep Kumar Samal, Yarui Peng, Yang Zhang, and Sung Kyu Lim
School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
Email: {sandeep.samal, yarui.peng}@gatech.edu, limsk@ece.gatech.edu

*Abstract*—In this paper, we study a 3D IC micro-controller implemented with sub-threshold supply for ultra-low power applications. Our study is based on GDSII layouts of a sub-threshold 8052 micro-controller that consumes $3.6\mu W$ power running at 20 KHz clock frequency and 0.4V logic supply. Our study confirms that sub-threshold circuits indeed offer a few orders of magnitude power vs performance tradeoff. In addition, our 3D sub-threshold design reduces the footprint area by 78% and wirelength by 33% compared with the 2D counterpart. Our studies also show that thermal and IR drop issues are negligible in this sub-threshold 3D implementation due to its extreme low power operation. Lastly, we demonstrate the low power and high memory bandwidth advantages of many-core 3D sub-threshold circuits.

## I. INTRODUCTION

One of the most effective ways to reduce the total power consumption of VLSI circuits is by reducing 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 [1]. This supply voltage usually falls below the threshold voltage of the transistors. The digital gates can still function in subthreshold regions by utilizing sub-threshold current, which has exponential dependence on gate voltage. The sub-threshold operating conditions require the individual gates to be designed in a robust manner, which requires larger size of transistors. For large designs, these result in area overhead. The circuit may also need to interface with on-chip or off-chip memory, which further complicates a subthreshold design. Though, the operating frequency reduces heavily, the major concern is power and longevity of battery.

Three dimensional ICs (3D ICs) is one of the most promising technologies that enables higher integration and further miniaturization with increase in memory bandwidth for off chip memories while reducing power dissipation and improving performance. Memory stacked over logic integrates off chip memories to on-chip. Studies have shown that significant increase in memory bandwidth is obtained by moving to 3D [2]. Much shorter interconnects due to die stacking results in lower power consumption compared to long interconnect designs. The overall footprint of the design also reduces significantly further satisfying the needs of miniaturized designs. However, 3D ICs suffer from the thermal and power integrity issues, which has to be taken care of during design and fabrication.

Through the work in this paper, we propose a new way to meet most of the requirements of low power miniaturized designs by using sub-threshold circuits stacked with memory. This results in low power and much smaller die footprint with an increased memory bandwidth. Our study is based on GDSII layouts and sign-off analysis of a sub-threshold 8052 micro-controller with $3.6\mu W$ power consumption at 20 KHz clock frequency operating at 0.4V logic supply. A similar near-threshold circuit but in 2D IC that runs at 1MHz frequency, 0.4V supply voltage and consumes $79\mu W$ has been recently presented [3].

In another work, 0.65V for logic and around 0.87V for SRAM is used in a near-threshold 3D stacked system designed for high energy efficiency [4]. The major contributions of this paper are as follows: (1) To the best of our knowledge, this is the first work on 3D design using subthreshold circuits. (2) A full comparison of performance metrics of the 2D and 3D nominal and sub-threshold circuits has been done showing the 3D benefits. (3) The thermal and IR drop issues that are generally the major drawbacks of going to 3D have been shown to have no effect in this design due to extreme low power further stabilizing 3D performance. (4) We have further demonstrated 3D advantage with many-core design.

## II. DESIGN METHODOLOGY

### A. Issues Related to Sub-threshold Designs

At very low supply voltage, the logic gates become extremely sensitive to noise and there may be logic failure. The standard cells may not function at sub-threshold region for several reasons. First, the optimum ratio between NMOS and PMOS is mainly determined by the difference between mobility of electrons and holes for super-threshold design. But in sub-threshold operation, the difference of threshold voltage determines the current variation between NMOS and PMOS. Using typical NMOS to PMOS ratio used in super-threshold operation may result in an unsymmetrical Voltage Transfer Characteristic in sub-threshold supplies. Second, the technology variation increases the difficulty for sub-threshold design. For super-threshold design, the on/off current ratio is typically more than $10^3$. Even though the technology causes some transistors to be stronger or weaker, the on-transistor still overwhelms the off-transistor. In sub-threshold region, only sub-threshold leakage current can be used to switch the logic gates. The current relation becomes exponential with respect to threshold voltage. Therefore any variation in technology may cause current to vary exponentially. Third, many standard memory components such as standard 6T SRAM cannot work in sub-threshold region since reduced noise margin (SNM) in sub-threshold design causes signal integrity issues. The memory cells need to be modified to reach the energy minimum voltage [5] [6].

### B. Sizing Techniques

To reduce process variation effects in our design, we used the GlobalFoundries 130nm technology that has a nominal supply voltage of 1.5V. We studied the transistor characteristics around the threshold voltage values (NMOS-0.53V and PMOS-0.56V). The spice simulations with supply varying from 0-0.8V show that during super-threshold operation, NMOS is faster than PMOS as expected but under subthreshold conditions, PMOS is stronger.

To determine our supply voltage we study the energy per cycle of a 20 inverter chain and find the optimum supply voltage at 0.3V
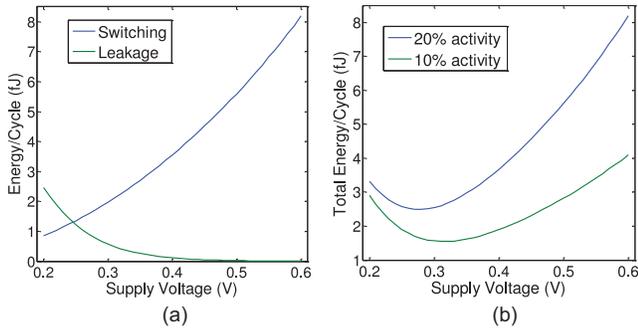
Fig. 1. Energy consumption per cycle for 20 inverter chain. (a) Switching and Leakage Energy, (b) Total Energy.

| | Power ($\mu W$) | | Avg $t_{delay}$ (ns) | | Power-delay product ($fJ$) | |
|------|-----------|---------|-----------|---------|-----------|---------|
| Cell | Super-vth | Sub-vth | Super-vth | Sub-vth | Super-vth | Sub-vth |
| BUF | 18.64 | 0.00081 | 0.164 | 323 | 3.06 | 0.262 |
| DFF | 29.66 | 0.00343 | 0.432 | 660 | 12.81 | 2.264 |
| INV | 14.97 | 0.00066 | 0.103 | 213 | 1.54 | 0.141 |
| MUX | 22.62 | 0.00155 | 0.207 | 595 | 4.68 | 0.922 |
| NAND2 | 21.92 | 0.00139 | 0.122 | 290 | 2.67 | 0.403 |
| NOR2 | 24.77 | 0.00160 | 0.132 | 293 | 3.27 | 0.469 |

(Fig. 1). However, for reliable operation of larger cells like D-flipflop, we set our supply voltage at 0.4V after careful spice simulations.

The logic gates are sized accordingly equalizing the strength of NMOS and PMOS to reduce propagation delay mismatch at 0.4V [7]. To save time and keep our focus on the objective targeted, we confine our standard cell library to the basic cells only viz. Inverter, NAND, NOR, Buffer, D-flipflop and MUX and avoid sizing of all gates for subthreshold operations. We used the same set of cells in superthreshold designs with nominal sizing to have a fair comparison. While sizing of standard cells, we use the load capacitance value of 16fF at a clock frequency of 100 KHz and input rise/fall time of $1\mu s$. The wider standard cells are used as per the need of subthreshold operation.

### C. Library Characterization

After sizing of the cells at the above mentioned settings and completing the new layout, we characterize our subthreshold standard cell library using the post extraction netlist with Cadence Encounter Library Characterizer. This new sub threshold library is used in the digital circuit design. The cell power consumption and delay comparison for 1.5V nominal voltage and 0.4V sub-threshold voltage for the respectively sized cells is shown in Table I. There is significant power reduction but due to increase in propagation delays, the power-delay product reduction is not so significant. Since, we set our supply to be 0.4V, the minimum energy per cycle supply voltage is not reached for all cells.

Since subthreshold operation is highly sensitive to variation in supply and temperature, we study the standard cell library behavior with changes in these parameters. The performance changes are shown in Table II.

### D. Memory Designs

For practical applications, the requirement of memory is unavoidable. There exists a number of works on subthreshold memory
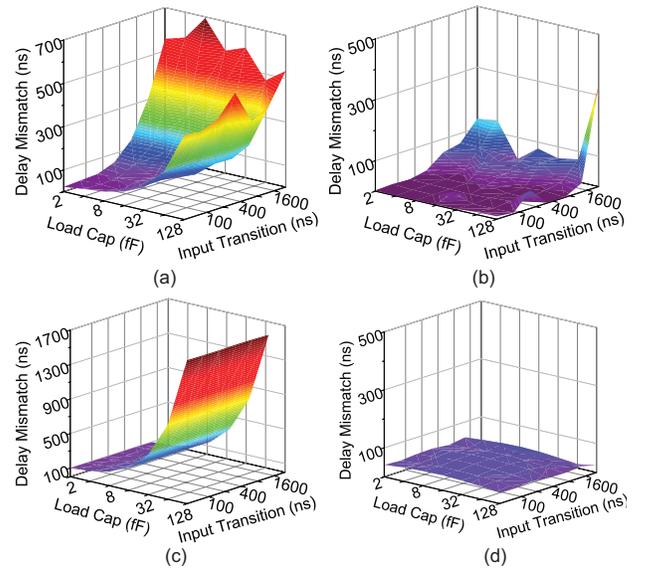


Fig. 2. Delay mismatch before and after sizing at 0.4V supply. (a) INV before sizing, (b) INV after sizing, (c) DFF before sizing, (d) DFF after sizing.

| Process Corner | FF | TT | SS |
|----------------|----|----|----|
| Rise Delay ($ns$) | 239 (0.599x) | 420 (1.0x) | 1080 (2.57x) |
| Fall delay ($ns$) | 251 (0.615x) | 408 (1.0x) | 1370 (3.35x) |
| Temperature ($K$) | 323 | 298 | 273 |
| Rise delay ($ns$) | 309 (0.736x) | 420 (1.0x) | 503 (1.20x) |
| Fall delay ($ns$) | 322 (0.789x) | 408 (1.0x) | 580 (1.42x) |
| Supply Voltage ($V$) | 0.44 | 0.40 | 0.36 |
| Rise delay ($ns$) | 336 (0.800x) | 420 (1.0x) | 542 (1.29x) |
| Fall delay ($ns$) | 287 (0.703x) | 408 (1.0x) | 575 (1.41x) |

design [5] [6]. However, we used commercial memory compilers to generate large memory macros at nominal supply voltages. For smaller memory sizes less than 1KB, we design register files using our subthreshold cells to have very low power consumption. The larger memory macros generated by commercial memory compilers uses 6T SRAM cells. We reduce the supply voltage of memory to 0.8V, which is determined from spice simulations as the minimum voltage for reliable 6T cell SRAM operation for the same technology. The libraries are also scaled as per the scaling factors obtained from spice simulations at reduced voltage supply of 0.8V. However there is a requirement of multiple voltage supplies of 0.4V and 0.8V for logic and memory respectively for a whole system design. Once again 3D 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. Signal interfacing between logic and memory requires the use of level shifters. As the voltage has to be raised from a sub-threshold voltage to a high voltage, we use a modified level-shifter circuit as shown in Fig. 3. Since the input is sub-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. These level-shifters are used at the address and data pin outputs of logic in our full-chip design.

### III. FULL-CHIP ANALYSIS

For our study, we use the 8052 micro-controller as our digital circuit and design it in 3D using subthreshold logic circuit and near-threshold memory. The 8052 micro-controller uses internal RAM of
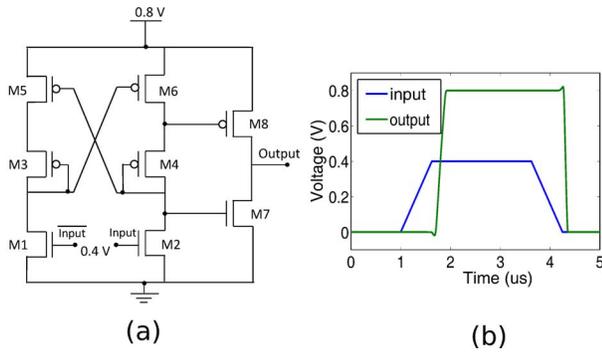
Fig. 3. Level Shifter Circuit used for 0.4V to 0.8V shift (a) Schematic (b) Transient Waveform



2D sub-Vth



5-tier 3D sub-Vth (die 0 and die1 shown only)



TSVs and their connections
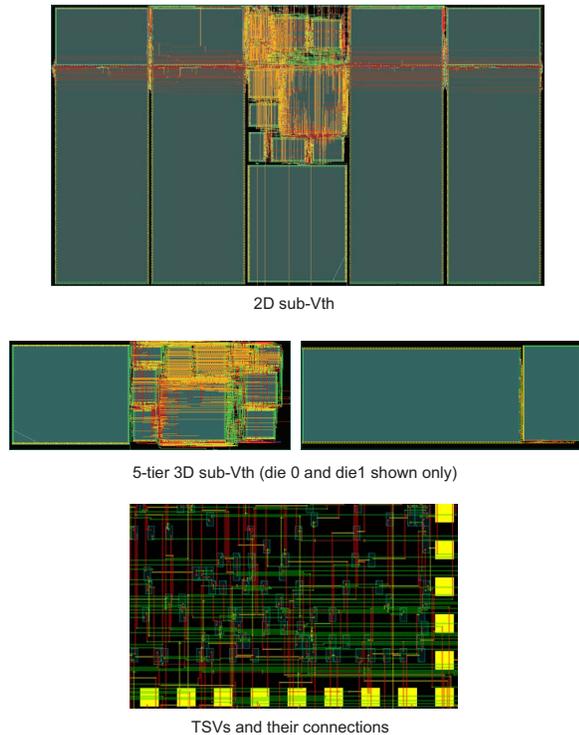
Fig. 4. Sub-threshold layout with 64KB external memory. The design for die 1 to die 4 (= memory dies) in 3D design are almost identical.

256 bytes and external RAM up to a maximum of 64KB. It also has a ROM with size 64KB. 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 3D design. The full-chip layouts are shown Fig. 4.

### A. Area and Wirelength Comparisons

We build 4 designs to compare the sub-threshold and 3D impact on the performance, the footprint area and wirelength. The design specifications are listed in Table III. The super-threshold designs operate at supply voltage of 1.5V, while the sub-threshold designs have a 0.4V supply to drive the logic and internal RAM and a 0.8V supply to drive the external memory. 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 reducing the total wirelength in the 3D

design. Each 16KB external RAM along with 16KB external ROM makes one die. We use 60 TSVs of $5\mu m$ in diameter to connect between two dies, and we set the TSV pitch to be $10\mu m$.

We observe that by implementing the designs in 3D, we obtain a 78% and 47% reduction in footprint area for 64KB and 16KB external memory size respectively. Because of the footprint reduction, the wires that are needed to connect the blocks are shortened, and this results is reduction of the top-level interconnect wirelength by 33% for 64 KB sub-threshold design. The wirelength saving is small in 16KB design due to very few 3D connections.

### B. Timing and Power Comparisons

Table IV and Table V shows the power and timing results of the system with 16KB and 64KB memory respectably. For super-threshold designs, the clock frequency can reach 66.7MHz. 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 power in super-threshold designs. By reducing the supply voltage and going to sub-threshold computing, the design with 16KB external memory shows 9099 times reduction in total power at the cost of 3333 times lower clock frequency. We use the symbol 'X' from here for comparison. The total power reported is sum of the internal, switching and leakage powers. The memory power has been reported independently. For certain low power applications like sensors, the work load for each nodes 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 sub-threshold designs are working at 0.8V, the larger the size of memory, the more the leakage 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 64KB external memory design, the internal and switching power is reduced by almost 35000X by changing from super-threshold design to sub-threshold logic. But the leakage power is not directly related to clock frequency change, so we achieve 6.49X times saving in leakage power saving. The larger the memory size, smaller are the power savings. The 64KB external memory design shows 3008X overall power reduction in contrast with the 9099X power saving for 16KB external memory design.

By introducing 3D IC, we obtain a significant saving in footprint area. As a result, the wirelength needed to connect the blocks is reduced and we observe a performance improvement and some power reduction due to smaller interconnect wire load. We are using the same blocks for both 2D and 3D design, and we assume each TSV has 100fF load capacitance and $0.5\Omega$ wire resistance in our 3D design analysis. Moving from 2D sub-threshold to 3D sub-threshold, we observe 4.85% timing improvement after timing analysis including TSV parasitic information. The internal power and switching power is reduced by 1.5% and 2.5% respectively in the sub-threshold 3D design with 64KB memory compared to sub-threshold 2D design, and the total power is only reduced a little because the leakage power remains almost the same and it is the dominating part in total power.

### C. Variation Study

As discussed in the previous section, subthreshold operation is highly sensitive to process, voltage and temperature variations. We

TABLE III.     FOOTPRINT AREA AND WIRELENGTH COMPARISON FOR 8052 MICROPROCESSOR WITH DIFFERENT SIZES OF EXTERNAL MEMORY

| | super-threshold | | | | sub-threshold | | | | comparison | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2D | | 3D | | 2D | | 3D | | 3D/2D (sub-Vth) | |
| Memory size ($KB$) | 16 | 64 | 16 | 64 | 16 | 64 | 16 | 64 | 16 | 64 |
| No. of dies ($KB$) | 1 | | 2 | 5 | 1 | | 2 | 5 | 2 | 5 |
| Supply voltage ($V$) | 1.5 | | | | 0.4 (logic and int RAM) + 0.8 (ext memory) | | | | | |
| Area ($\mu m \times \mu m$) | $940 \times 1300$ | $2300 \times 1300$ | $500 \times 1300$ | $500 \times 1300$ | $940 \times 1300$ | $2300 \times 1300$ | $500 \times 1300$ | $500 \times 1300$ | 0.53 | 0.22 |
| Top wirelength ($mm$) | 249.8 | 371.9 | 204.9 (bottom die) 3.80 (each top die) | | 244.2 | 376.7 | 237.8 (bottom die) 4.02 (each top die) | | 0.99 | 0.67 |

TABLE IV.     TIMING AND POWER RESULTS FOR 8052 MICROPROCESSOR WITH 16KB EXTERNAL MEMORY. TIMING VALUES ARE IN $ns$, AND POWER IN $mW$.

| | super-Vth | sub-Vth | | comparison | |
|---|---|---|---|---|---|
| | 2D | 2D | 3D | $\frac{2Dsub}{2Dsuper}$ | $\frac{3Dsub}{2Dsub}$ |
| Target clock | 15 | 50000 | | 3333.33 | 1 |
| Timing slack | 1.626 | 9331 | 9938 | | |
| Internal power | 4.713 | 0.0001714 | 0.0001713 | 1/27497 | 1 |
| Switching power | 5.893 | 0.0001322 | 0.0001303 | 1/44576 | 0.985 |
| Leakage power | 0.00595 | 0.0008611 | 0.0008609 | 1/6.9 | 1 |
| Memory power | 0.7145 | 0.0008861 | 0.0008861 | 1/806 | 1 |
| Total power | 10.6 | 0.001165 | 0.001163 | 1/9099 | 1 |

TABLE V.     TIMING AND POWER RESULTS FOR 8052 MICROPROCESSOR WITH 64KB EXTERNAL MEMORY. TIMING VALUES ARE IN $ns$, AND POWER IN $mW$.

| | super-Vth | sub-Vth | | comparison | |
|---|---|---|---|---|---|
| | 2D | 2D | 3D | $\frac{2Dsub}{2Dsuper}$ | $\frac{3Dsub}{2Dsub}$ |
| Target clock | 15 | 50000 | | 3333.33 | 1 |
| Timing slack | 1.399 | 7895 | 9938 | | |
| Internal power | 6.214 | 0.0001800 | 0.0001773 | 1/35047 | 0.985 |
| Switching power | 4.799 | 0.0001424 | 0.0001389 | 1/34550 | 0.975 |
| Leakage power | 0.02165 | 0.003334 | 0.003333 | 1/6.5 | 1 |
| Memory power | 0.7865 | 0.003363 | 0.003361 | 1/234 | 1 |
| Total power | 11 | 0.003657 | 0.003649 | 1/3015 | 0.998 |

TABLE VI.     EFFECT OF PVT VARIATIONS ON SUBTHRESHOLD LOGIC ONLY : POWER NUMBERS BASED ON 20KHz FREQUENCY OPERATION

| Process Corner | FF | TT | SS |
|---|---|---|---|
| Longest path delay ($ns$) | 9801 | 40062 | 327905 |
| Core Leakage Power ($\mu W$) | 0.185 | 0.037 | 0.021 |
| Total Core Power ($\mu W$) | 0.420 | 0.278 | 0.264 |
| Temperature ($^oC$) | 50 | 25 | 0 |
| Longest path delay ($ns$) | 23316 | 40062 | 104595 |
| Core Leakage Power ($\mu W$) | 0.121 | 0.037 | 0.024 |
| Total Core Power ($\mu W$) | 0.364 | 0.278 | 0.266 |
| Supply Voltage ($V$) | 0.44 | 0.4 | 0.36 |
| Longest path delay ($ns$) | 22424 | 40062 | 111500 |
| Core Leakage Power ($\mu W$) | 0.050 | 0.037 | 0.033 |
| Total Core Power ($\mu W$) | 0.347 | 0.278 | 0.230 |

TABLE VII.     EFFECT OF PROCESS VARIATIONS ON MEMORY ONLY (OPERATING CORNERS OBTAINED FROM MEMORY COMPILER AND SCALED DOWN)

| Corner | FF/0.88V/-40$^oC$ | TT/0.8V/25$^oC$ | SS/0.72/125$^oC$ |
|---|---|---|---|
| Leakage Power ($\mu W$) | 3.12 | 3.3 | 6.52 |
| Total Power ($\mu W$) | 3.20 | 3.37 | 6.60 |

therefore, analyze our design at different process corners and with temperature and voltage variations. The results for only logic variations is shown in Table VI. Only-memory process variation effects are shown in Table VII. We observe that the design becomes faster at higher temperature unlike standard circuits and consumes higher power. This is because sub-threshold current increases exponentially 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*

Since the 3D IC is stacking several dies into a single package, therefore heat dissipation is always a major concern. Also, the performance of sub-threshold cells are very sensitive to temperature, therefore, we need to carefully simulate the thermal effects on our 3D designs. Since current tools cannot handle 3D designs properly, we use our in house tools to build a thermal model for 3D IC and perform thermal simulation using ANSYS Fluent. First we build a mesh for our chip, and compute the thermal conductivity for each grid using layout information. Then we export power information from Primetime and build a power density map. Finally we use Fluent to solve the thermal differential equations using the power density map and thermal conductivity information and obtain the temperature map of our design. In this simulation, we assume adiabatic boundary conditions on all 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 $25^oC$.

The temperature map for 2 die design is shown in Fig. 5 respectively. Since in super-threshold design, the memory power is only 7.3% of the total power, therefore, 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. Also, since TSVs are made of copper which has the highest thermal conductivity among all the materials on the chip, it can transfer heat quickly from the bottom die to the top die. Therefore, around the TSV arrays, the temperature is relatively low and that area becomes the coolest part of the full-chip. By lowering the voltage supply and performing sub-threshold 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 $72.96^oC$ in super-threshold design to $0.00852^oC$ in sub-threshold design. Also, since the memory has much larger power than the logic portion, the temperature on the bottom die is heavily affected by the top die. In this case, the ROM has the largest power density within the chip, so the maximum temperature appears on the part where ROM is placed. From the results, we can conclude that by stacking sub-threshold circuits in 3D, 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 supply voltage variation. This makes the power distribution within the chip a very critical design step. Even though the external supply may not vary, the internal IR drop may result in reduced supplies to certain logic gates within the chip. To study this effect, we use a very simple Power Distribution Network (PDN) for our chip and analyze the static IR drop for sub-threshold operation. The IR drop issues will mainly
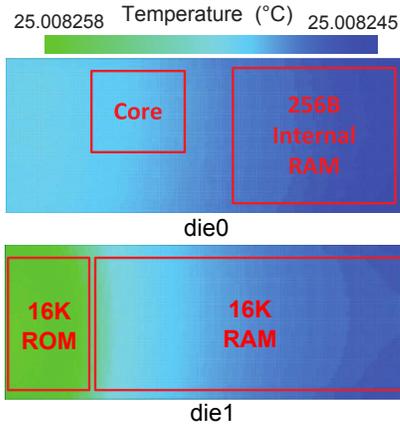
Fig. 5. Temperature Map for Sub-threshold 3D design

TABLE VIII. 3D FULL-CHIP TEMPERATURE AND IR DROP (LOGIC DIE) ANALYSIS

| | Power density $(mW/cm^2)$ | | Max Temp $(^oC)$ | | Max Static IR Drop |
|---|---|---|---|---|---|
| | Die0 | Die1 | Die0 | Die1 | |
| 3D Super-Vth | 10410 | 750 | 97.964 | 97.821 | 26mV |
| 3D Sub-Vth | 0.269 | 0.381 | 25.008 | 25.008 | $0.34\mu V$ |

affect the logic portion operating at sub-threshold 0.4V because the memory contains dies are operating at 0.8V and have a dedicated PDN.

The blocks used in the top level logic design have power rings on their boundaries. We use simple minimum PDN for top level with only rings at the die boundary in the top metal layers and use the Metal1 VDD and VSS rails to connect to the individual block rings as well as the top level buffers and inverters used for timing closure. 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. 6 shows the IR drop map. 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 $0.34\mu V$ in sub-threshold design. The values are so small because the current drawn by each cell from the supply is sub-threshold and hence very small. Therefore, we can conclude that IR drop is not an issue for our sub-threshold 3D design.

Another major advantage of going 3D is the integration of sub-threshold and near-threshold (as in our work) or super-threshold dies without any additional circuits for separate PDNs . Since each die operates at its own supply, we can have dedicated power ground TSVs supplying the top tiers. The same design in 2D may require isolation and decoupling of the different supply networks.

## IV. 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
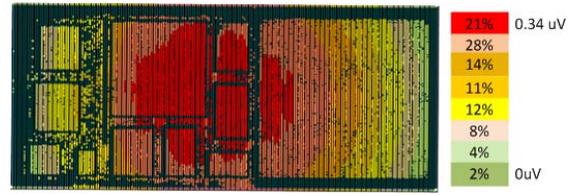


Fig. 6. IR drop map for sub-threshold logic tier
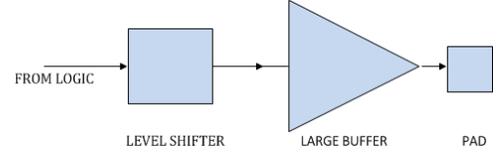


Fig. 7. I/O circuit for logic to off-chip memory connections in many core 2D sub-threshold design

huge amount of power, their count usually has an upper bound. However, when we use 3D 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. Since the memory to processor connections are TSV based in 3D design, they consume much less power than I/O pads. We study this feature quantitatively by implementing a many core sub-threshold design with off chip memory and comparing it with the equivalent 3D implementation.

### A. I/O Driver Design

The I/O pads provided in the standard library are large with complicated circuits in them and therefore consume large amount of power and cannot be used for sub-threshold circuits. They are meant for high performance in standard 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 ESD and other circuits from our simple design. The representative diagram for the output pad is shown in Fig. 7. We use a capacitive load of 5pf 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 off chip design. Level shifters are required as the processor output is 0.4V while the memory operates at 0.8V. We set the I/O supply voltage as 0.8V. The total power dissipated by a single I/O pad is calculated from spice simulations to be $1.066\mu W$ at 20KHz with 5pF load and $1\mu s$ input slew. This is inclusive of switching power, cell power and leakage power.

### B. Power Saving in Many-core Sub-Vth 3D Designs

Using 3D implementation for the micro-controller design helps us put together many processors as per requirement in reduced area with reduced interconnect and hence lesser power. Since each of the cores are smaller in size in 3D, the inter-core connections become shorter and there is less loss of power in wires. As we move to lower technology processes, the wirelength parasitic become more critical and contribute significantly to power. Also, we can achieve significant memory bandwidth increase by going to 3D.

For initial study, we used 128 cores in 2D and 3D designs and analyzed the area, wirelength, processor to memory I/O power and

2D 128-core design     5-tier 3D 128-core design
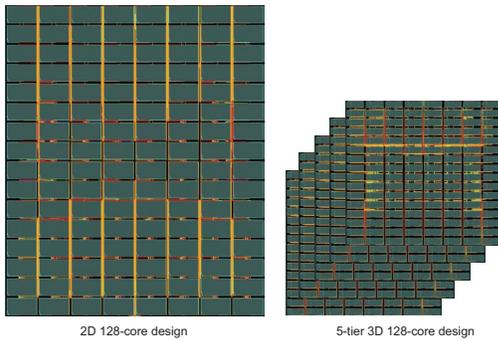
Fig. 8. Many core sub-threshold layouts.

TABLE IX.     DESIGN COMPARISON FOR SUB-THRESHOLD MANY CORE
IMPLEMENTATION IN 2D AND 3D

| Property | 2D | 3D (5-tier) |
|---|---|---|
| Area ($cm \times cm$) | $2 \times 2.4$ (100%) | $1.12 \times 1.2$ (28%) |
| Wirelength ($m$) | 54.70 (100%) | 23.94 (44%) |
| Processor to Memory I/O Power ($\mu W$) | 104.5 (100%) | 0.768 (0.007%) |
| Total Power ($\mu W$) | 564.5 (100%) | 460.8 (81.5%) |
| Data Connections | 48 (I/O) | 3072 (TSV) |
| Total Connections | 98 (I/O) | 3108 (TSV) |
| Memory Bandwidth ($bits/cycle$) | 16 (100%) | 1024 (6400%) |

the memory bandwidth. For the 2D design we use a single large off chip memory while we use 5 tier stacking for 3D design. The many-core layouts are shown in Fig. 8. The comparison results are shown in Table IX. We use a 2-channel off chip 2D memory as our baseline for comparison and then analyze the benefits of 3D design. We observe that the reduction in wire power is proportional to the reduction in wirelength. Also the processor to memory I/O power, which is 18.5% of the total power in 2D design is completely removed in the 3D implementation. The theoretical maximum bandwidth also increases from 16bits/cycle for 2D many core to 1024 bits/cycle for 3D many core as compared to 8bits/cycle for single core. Therefore, going for 3D designs will contribute significantly to power savings above the sub-threshold savings with increase in memory bandwidth for specific applications.

## V.     DESIGN LESSONS AND GUIDELINES

The design of sub-threshold gates involves a number of issues which need to be taken care of. Wider cells (2X or 4X) need to be used for sub-threshold operation for better performance in driving same load as super-threshold circuits. Also the narrow-width effect is much more significant in sub-threshold operation of gates [8] and proper cell sizing is dependent on it. While determining the supply voltage for the design, we need to take care of proper functionality of all the gates because the energy-minimum supply voltage occurs at deep sub-threshold region where cells may fail to operate reliably. The D-flipflop is the critical cell in our case whose correct operation determines our supply voltage of 0.4V. The cell characterization has to be done carefully based on sub-threshold operation based input slew values. For 3D sub-threshold design, we need to be very careful about the thermal and IR drop variations that the chip may encounter during operation. Sub-threshold cells are highly sensitive to temperature and voltage variations. Since 3D stacking causes thermal and power integrity issues, this is a very critical part of the design. The TSV parasitic also need to be taken into account during full chip timing and power analysis as they may be a non-negligible portion of total power dissipation. To reduce power and improve performance,

circuit techniques like power gating, adaptive body biasing or design of sub-threshold memory cells is a very good option. We need to consider these in our future designs. In 3D sub-threshold design, it is preferable to use different supply cells in different dies in case we have a multiple supply design. The reason is that we can have dedicated power supply to the dies without major design issues. However, we need to design sub-threshold level shifter circuits for proper interfacing of signals with different voltage values. Another important factor is that the I/O pads used for sub-threshold circuits need to be modified. Otherwise, the objective of low power operation will be nullified by the power hungry I/O pads.

## VI.     CONCLUSIONS

In this paper, for the first time we explore the 3DIC benefits in ultra low power designs using subthreshold circuits. While logic circuits show an excellent reduction in power consumption, memory contributes to maximum power in our designs because of its near threshold region of operation and high leakage. Larger the memory in a design, lesser are the power savings. However, the footprint area reduction of around 78% is quite significant when we move to 3D and this plays a key factor in miniaturization of digital systems. We showed that 3DIC implementation of subthreshold circuits is free from internal thermal and IR drop related issues. The problem of dominating memory power can be improved significantly by using special low voltage memory cells [6]. This is one of the very important steps to be considered in our future work. We have also demonstrated the idea of many-core design with increase in memory bandwidth and further reduction in power consumption due to the removal of processor to memory I/O pads. Therefore, 3D stacked subthreshold circuits with proper memory design approach present a major improvement for both ultra-low power and miniaturization in processors.

## VII.     ACKNOWLEDGMENT

## REFERENCES

[1]  A. Wang and A. Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, 2005.

[2]  D. H. Kim, *et al*, "3D-MAPS: 3D Massively Parallel Processor with Stacked Memory," in *ISSCC Dig. Tech. Papers*, 2012, pp. 188–189.

[3]  M. Konijnenburg, *et al*, "Reliable and energy-efficient 1MHz 0.4V dynamically reconfigurable SoC for ExG applications in 40nm LP CMOS," in *ISSCC Dig. Tech. Papers*, 2013, pp. 430–431.

[4]  D. Fick, *et al*, "Centip3De : A 3930 DMIPS/W Configurable Near-Threshold 3D Stacked System with 64 ARM Cortex-M3 Cores," in *ISSCC Dig. Tech. Papers*, 2012, pp. 190–191.

[5]  S. Hanson, *et al*, "A Low-Voltage Processor for Sensing Applications With Picowatt Standby Mode," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1145–1155, 2009.

[6]  S. Hanson, *et al*, "Ultralow-voltage, minimum-energy CMOS," *IBM Journal of Research and Development*, vol. 50, no. 4/5, pp. 469–490, 2006.

[7]  B. H. Calhoun, *et al*, "Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1778–1786, 2005.

[8]  J. Zhou, *et al*, "A 40 nm inverse-narrow-width-effect-aware sub-threshold standard cell library," in *Proc. ACM Design Automation Conf.*, 2011, pp. 441–446.