IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 12 February 2015; revised 3 July 2015; accepted 4 July 2015. Date of publication 11 February 2015;
date of current version 9 December 2015.

Digital Object Identifier 10.1109/TETC.2015.2454854

Powering Up Dark Silicon: Mitigating the Limitation of Power Delivery via Dynamic Pin Switching

Shaoming Chen[1], Lu Peng[1], Yue Hu[1], Zhou Zhao[1], Ashok Srivastava[1], Ying Zhang[1], Jin-Woo Choi[1], Bin Li[2], and Edward Song[1]

[1]Division of Electrical and Computer Engineering, School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, LA 70803 USA

[2]Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803 USA

CORRESPONDING AUTHOR: L. PENG (lpeng@lsu.edu)

This work was supported in part by the Division of Computing and Communication Foundations through the National Science Foundation under Grant CCF-1017961 and Grant CCF-1422408.

ABSTRACT The end of Dennard scaling has led to a large amount of inactive or significantly underclocked transistors on modern chip multiprocessors in order to comply with the power budget and prevent the processors from overheating. This so-called dark silicon is one of the most critical constraints that will hinder the scaling with Moore's Law in future. While advanced cooling techniques, such as liquid cooling, can effectively decrease the chip temperature and alleviate the power constraints, the peak performance, determined by the maximum number of transistors, which are allowed to switch simultaneously, is still confined by the amount of power pins on the chip package. In this paper, we propose a novel mechanism to power up the dark silicon by dynamically switching a portion of I/O pins to power pins when off-chip communications are less frequent. By enabling extra cores or increasing processor frequency, the proposed strategy can significantly boost the performance compared with the traditional designs.

INDEX TERMS Multiprocessor systems, impact of VLSI on system design, design studies.

I. INTRODUCTION

The continuous shrinking of modern semiconductors has stalled Dennard scaling, which has hitherto sustained the achievement of Moore's Law over the past few decades. That is, the supply voltage of a transistor – hence the per-transistor switch energy – is no longer scaled with its geometric dimensions [32]. Since the on chip transistor density doubles every 18 months, the chip power is increasing faster than the power delivery system can handle. Consequently, a large portion of transistors need to be significantly under-clocked or even completely turned off to enclose the processor power consumption within a reasonable envelope. This phenomenon, which is termed ''utilization wall'' or ''dark silicon'' [20], is predicted to be one of the most critical constraints preventing us from obtaining commensurate performance benefits by adding transistors in the future.

In the current industry, there are two commonly accepted reasons for power constraints that cause dark silicon: thermal constraints and power delivery [22]. The slow improvement of per-transistor switch energy along with the fast growing transistor density has led to a considerable rise in the power consumption per unit area (i.e., power density). Provided that inexpensive cooling techniques such as air cooling are still the mainstream solution to heat dissipation for desktop and mobile platforms, such increasing of the power density tends to generate substantial heat that outstrips the chip's heat spreading capability. In this situation, the maximum power consumption of the chip cannot go beyond a threshold in order to maintain a safe working temperature for the entire processor. This power limit is usually referred to as the thermal design power (TDP). Some high-end processors with a higher TDP use backplate liquid cooling [16] to avoid thermal issues.

The underlying power delivery system, on the other hand, constrains the amount or the frequency of simultaneously active transistors as it determines the maximum power that is

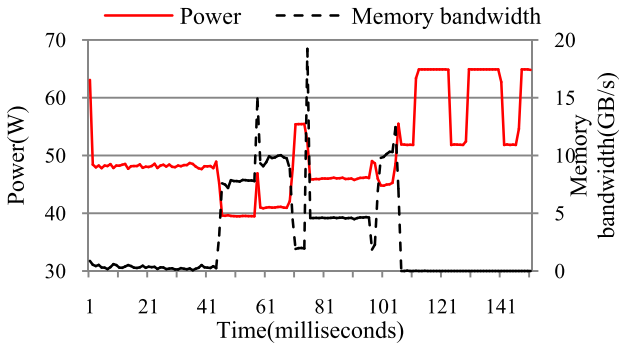

FIGURE 1. Power and memory bandwidth (8 copies of *DEALII* from SPEC2006).

able to be provided to the chip irrespective of the thermal concern. To alleviate this constraint, we consider increasing the power envelope with minimum circuit change to the existing computer systems, in order to enable more transistors or raise the operating frequency in the power-hungry phases during program execution. Figure 1 plots a snapshot of the execution of 8 copies of DEALII from SPEC2006 on an 8-core processor, visualizing a representative scenario that motivates our work. The off-chip memory traffic and processor power consumption both vary in different execution phases. More interestingly, the two traces generally show an opposite trend during the execution; when the memory traffic is relatively light, the total power consumption is quite considerable (e.g., time interval 106 – 151ms). On the other hand, a duration of memory-intensive execution will correspond to a low-power period (e.g., the plateau from 46ms to 70ms). The underlying reason for this phenomenon is that frequent misses in the last-level cache and the resultant off-chip memory accesses will largely slow down the overall execution rate, leading to a decrease in the processor's power consumption.

This intuitive observation implies an important opportunity for performance improvement and dark silicon mitigation by appropriately balancing the power delivery and off-chip traffic. To exploit this potential benefit, we propose a novel mechanism to dynamically switch a portion of I/O pins for power delivery when off-chip memory accesses are infrequent in order to mitigate the limitation of power delivery due to a lack of power pins, thus powering up the dark silicon for performance boost. During a phase when off-chip activities are relatively high, we switch back the pins for signal transmission.

Specifically, we make the following main contributions in this work:

- We propose a novel switchable pin design in which switchable pins can dynamically switch between power pins and I/O pins depending on a control voltage.
- We propose to replace a fraction of the existing processor I/O pins with switchable pins to optimize performance by dynamically switching pins to "I/O mode" for signal transmission during memory intensive phases, or to "power mode" to deliver extra power during computation intensive phases.
- We give a circuit implementation for the proposed dynamic pin switching design, using minor changes to existing processor and motherboard circuitry.
- We further design a rigorous statistical model that correlates the historical execution behaviors and off-chip access intensities in upcoming intervals. The established model can be employed by the operating system or equivalent supervisor to guide pin switching at runtime.
- We conduct a series of simulations to evaluate the performance, energy efficiency, and thermal impact of the proposed design on a chip multiprocessor (CMP) in both the *dim silicon* [34] and *dark silicon* modes.

The remainder of this paper is organized as follows. We review related works and introduce the background of C4 pads in Section 2. Section 3 elaborates the dynamic pin switching design and its impact on the power delivery network, signal transmission, and thermal issues. We then describe the prediction model and experimental setup in Section 4 and Section 5, respectively. The evaluation results will be demonstrated and analyzed in Section 6. We finally conclude our work in Section 7.

II. RELATED WORK AND BACKGROUND

A. DARK SILICON & DIM SILICON

Dark silicon has emerged as an increasingly important issue that will menace the scaling of Moore's Law in the deep submicron era and beyond. Esmaeilzadeh et al. [18] use an analytical model to predict processor scaling for the next few generations. They demonstrate that dark silicon will be heavily exacerbated by the continued shrinking of manufactured technology. Esmaeilzadeh *et al.* [18], Goulding-Hotta *et al.* [20], and Hardavellas *et al.* [21], [22] commonly attribute the cause of dark silicon to physical power and off-chip bandwidth constraints. Kim et al. [25] proposes to integrate memory with processors as a 3D chip. This integration can mitigate off-chip bandwidth constraints but it brings more challenges for power delivery and cooling since extra power will be consumed by the integrated memory. Hardavellas et al. [22] investigate this problem and believe even if an advanced liquid cooling technique was applied the power delivery would still result in dark silicon. On the other hand, instead of powering on partial transistors in dark silicon, researchers propose *dim silicon* in which all transistors are powered on but run at a lower frequency [34].

B. C4 PADS

Integrated circuit (IC) packaging is the final process of the IC fabrication in which the silicon die (i.e., the core of the device) is encased in a support and connected to the chip package for power delivery and off-chip communication. There are two main technologies for connecting the silicon die with the chip package: wire bonding and flip chip. Wire bonding uses bonding wires to connect the pads located on the perimeter of the silicon die to the package.

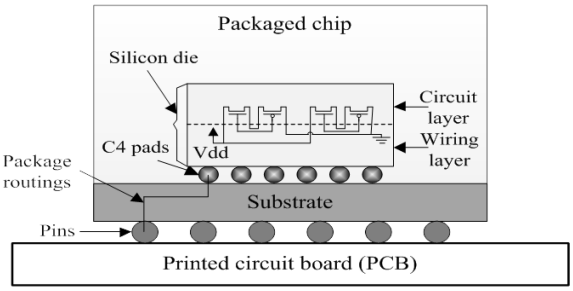

FIGURE 2. Structure of a packaged chip.

The flip chip technology, also called Controlled Collapse Chip Connection (C4) technology, is shown in Figure 2. The silicon die faces downwards, and is connected to the substrate directly with C4 pads. C4 technology greatly increases pad density, compared with wire bonding, by allowing C4 pads to be placed over the entire chip area. This eases wiring requirements by allowing shorter wire lengths and fewer global wires, and provides better power distribution as circuits in the middle of the die can access VDD/GND directly. The size of a silicon die is smaller than that of the chip package. This means the cross-sectional area of a C4 pad is smaller than that of a pin as shown in Figure 2. According to a recent study [36], it is concluded that I/O pad shortage will limit power delivery in future sub-16nm technology. In addition, increasing the number of C4 pads will linearly increase chip packaging costs, which have already started to exceed the silicon fabrication costs [23].

The ITRS [6] predicts that C4 pad density will increase 7.7% annually, and fail to ever meet demand, which is increasing at 15.7% annually. Zhang et al. [36] evaluate the usage of C4 pads in a multicore processor and conclude that we will see a C4 pad shortage starting from 16nm technology node. The shortage comes from an increasing demand in power delivery and off-chip bandwidth but a slow improvement in C4 pad technology. Previous works [21], [33] observe that the required number of C4 pads increases exponentially with the number of processor cores. Therefore, an exponentially larger number of C4 pads are needed to increase off-chip communication. Moreover, more power pads are needed for current delivery as each new technology increases the power density. Barowski *et al.* [11] from IBM also observe the C4 pad shortage and propose to utilize the heat sink to deliver power. A recent work demonstrates the impact of the pad shortage on power delivery quality [37]. In another recent work, Chen et. al [14] propose to use switchable pins to increase memory bandwidth. Instead, we propose to increase power delivery using pin switching.

C. ON-CHIP VOLTAGE REGULATOR (VR)

Theoretically, an on-chip VR can be used to deliver more power by supplying a chip with a relatively high voltage and convert the voltage to a normal value inside the chip. However, on-chip VR has large area [27]. Our proposal presents another alternative approach to the power delivery problem.

III. DYNAMIC PIN SWITCHING MECHANISM

Here we elaborate the dynamic pin switching mechanism in detail. In Section 3.1, we show the overall design on the motherboard to illustrate how the computer system functions while utilizing switchable pins to deliver power. Then we analyze the pin allocations of an Intel Xeon processor to see how many switchable pins we can design in a processor in Section 3.2. In Section 3.3 and Section 3.4, we study the impact of the switchable pins on the power delivery network and the signal transmission respectively. Section 3.5 studies area overhead and Section 3.6 discusses thermal issues. Finally, we describe the workflow of dynamic pin switching in Section 3.7.

A. OVERVIEW DESIGN

We now discuss how the computer system functions while utilizing switchable pins to deliver power. Figure 3 shows an overview of the dynamic pin switching design illustrating the layout of the microprocessor and SDRAM on the motherboard. The 64-bit data path of the integrated memory controller in the microprocessor connects to the SDRAM via 64 pins, specifically 16 conventional pins and 48 switchable pins. The 16 conventional pins are always used as I/O pins, while the switchable pins can switch between power pins and I/O pins dynamically. Our COMSOL-based [2] simulation which models the electromigration phenomenon on the traces/interconnects shows that using wires connecting to I/O pins to deliver the current studied in this work will not result in reliability issues. When the control voltage is low, the computer system works in the default I/O mode since the switchable pins are used as I/O pins. In this mode, signals circumvent the shift register components on the microprocessor and motherboard, which causes the 64-bit data path of the integrated memory controller in the microprocessor to connect to the SDRAM via 64 I/O pins directly. On the other hand, when the control voltage is pulled up the switchable pins are used as power pins; thus the computer system works in the power mode. In this mode, all shift register components are enabled to implement the signal transmission via the limited 16 I/O pins. The shift registers are bi-directional, one is parallel-in serial-out while the other is serial-in parallel-out, and have a negligible area overhead [3], [4]. When switchable pins are used for signal transmission, shift registers steer the signal from input to output without buffering them. Otherwise, they are used to send signals over a single line instead of 4 lines. The shift registers can be integrated into the microprocessor and motherboard and synchronized by the clock signal of SDRAM interface. We also add a delay circuit to balance the delay between lines with and without signal buffers.

The shift registers work at the same frequency as the SDRAM and integrated memory controller.

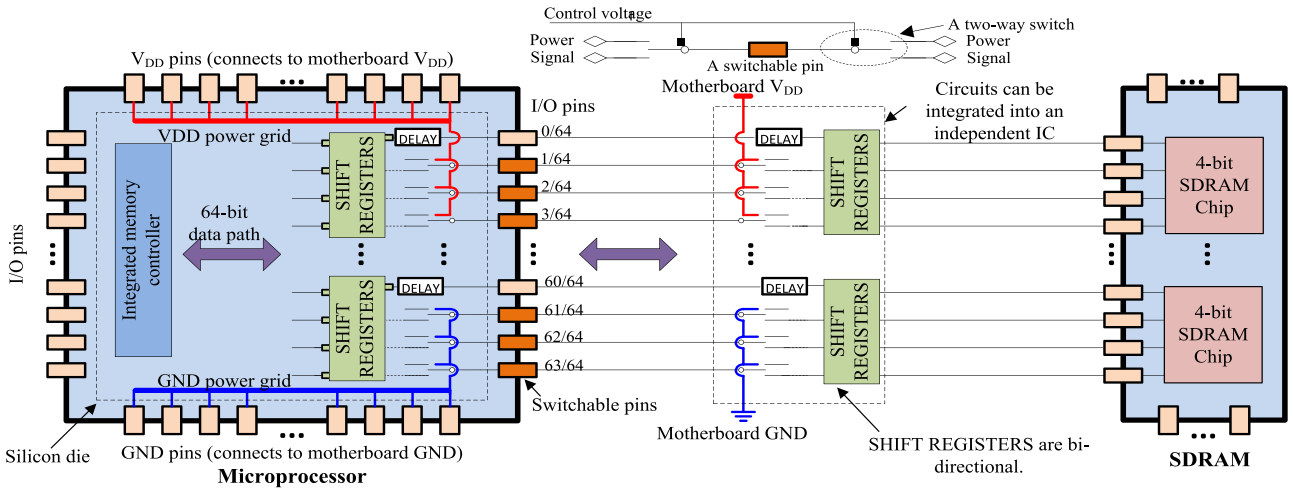

FIGURE 3. **Design overview on the proposed scheme.**

TABLE 1. **Pin allocation of the Intel Xeon Processor E5-2450L.**

V_{DD}	GND	DDR3	PCIE	QPI	DMI2	Others	Total
151	353	483	102	45	16	206	1356

Therefore, in power mode it takes four times as many cycles to transfer data over the bus via 16 I/O pins as it does via 64 I/O pins. The equivalent bus frequency is decreased to 25% of its default value when the switchable pins are used for power delivery although only data I/O pins are influenced (i.e., the number of effective I/O pins is decreased from 64 to 16), which can reduce the bus power [15]. Although the design increases the time required for transferring data over the bus, it will not affect bank access time or the queuing delay.

To minimize the change to the computer system, we only consider one-way pin switching, i.e., dynamically allocating a portion of I/O pins to power pins. In fact, it is feasible to switch from power pins to I/O pins by designing extra I/O units (e.g. memory controllers) and related control logics. Switching from power pins to signal pins will increase the off-chip communication bandwidth, which boosts the performance of memory intensive workloads significantly. This work focuses on switching from signal pins to power pins since the major purpose is to find an approach to power up dark silicon.

B. PIN ALLOCATION

To see how many switchable pins can be designed in a processor, we study the pin allocations of an Intel Xeon Processor E5-2450L [5] as listed in Table 1. We assume an equal number of C4 pads are designed on the chip with a pad density of 1356 pads/cm2 approximates to about 1200 pads/cm2 in the typical pad design [36]. Although it is feasible to design denser pads, the current that each pad can deliver will be smaller. The Xeon is an 8-core processor with a 20MB last-level cache and three memory channels. As can be seen, most pins are used for power delivery and off-chip communication. Among the off-chip communication pins, three 64-bit DDR3 memory channels occupy 483 C4 pins. Out of the pins on a 64-bit data path, 48 pins can be designed as switchable pins as discussed in Section 3.2. Correspondingly, three memory channels have 144 switchable pins which can increase the number of power pins by 28.6% (i.e., 144/(151+353)). On the other hand, among power delivery pins the number of the GND pins is more than that of the VDD pins. This helps lower the ground voltage in the silicon die, increasing circuit reliability, since the ground voltage is also used as a reference voltage for signal transmission. Conservatively following the same VDD/GND ratio, we allocate 144 switchable pins to 45 VDD pins and 99 GND pins in the pin switching mode. More switchable pins can be designed from other pins in DDR3 and pins in PCIE, QPI, DMI2 and etc. As an initial study, we only consider the 144 switchable pins from a portion of the data I/O pins in the three memory channels (DDR3).

C. POWER DELIVERY NETWORK

Here we study the impact on the power delivery network when the switchable pins switch from I/O mode to power mode. In the power delivery network (PDN) shown in Figure 4 we assume the voltage regulator module is a fixed voltage source since its feedback control mechanism can maintain a steady output voltage regardless of current magnitude. The power delivery path across the printed circuit board (PCB), the package, and the silicon die are modeled as the RL (i.e., resistor and inductor) components connected in series. Decoupling capacitances are introduced between each sub power network to reduce the voltage bounce. Power grids and processor circuits of the silicon die are modeled separately as RL components and an ideal current source.

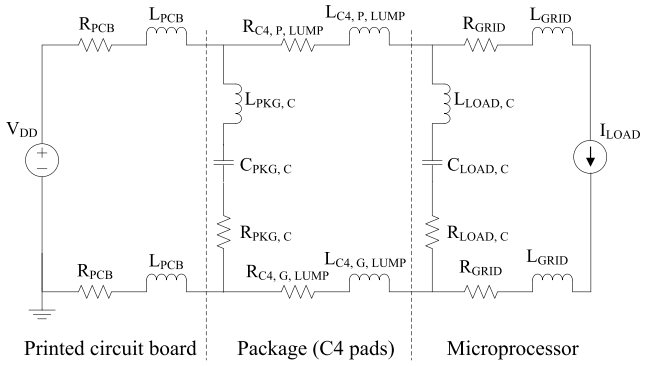

FIGURE 4. **RLC power delivery model.**

TABLE 2. **Power network model parameters.**

Resistance	Value	Inductance	Value
R_{PCB}	0.015 mΩ	L_{PCB}	0.1 nH
$R_{PKG,\,C}$	0.2 mΩ	$L_{PKG,C}$	1 pH
$R_{LOAD,C}$	0.4 mΩ	$L_{LOAD,C}$	1 fH
R_{GRID}	0.01 mΩ	L_{GRID}	0.8 fH
$R_{C4,\,SINGLE}$	40 mΩ	$L_{C4,\,SINGLE}$	72 pH
$R_{PowerSwitch,ON}$	1.8 mΩ	$C_{PowerSwitch,Parasitic}$	0.232pf
Default I/O mode			
$R_{C4,\,P,\,LUMP}$	0.265 mΩ	$L_{C4,\,P,\,LUMP}$	0.48 pH
$R_{C4,\,G,\,LUMP}$	0.113 mΩ	$L_{C4,\,G,\,LUMP}$	0.20 pH
Pin switching mode			
$R_{C4,\,P,\,LUMP}$	0.206 mΩ	$L_{C4,\,P,\,LUMP}$	0.37 pH
$R_{C4,\,G,\,LUMP}$	0.089 mΩ	$L_{C4,\,G,\,LUMP}$	0.16 pH
Capacitance			
$C_{PKG,C}$	250 µF	$C_{LOAD,C}$	500 nF

Table 2 gives the parameter values obtained from prior works [24], [26].

We perform static PDN simulations using SPICE 0. There is an IR drop between the supply voltage and the load voltage as current flows through the PDN. As the total current increases, the IR drop increases due to the resistance on the power delivery path. We assume the normalized IR drop should be limited to be less than 5% as a design convention used by previous work [28], [36] to ensure signal integrity and energy efficient power delivery. Thus, the maximum allowable currents are respectively 116A and 144A for the baseline and the pin switching design. In other words, the pin switching design can supply an extra 24.1% (i.e., (144−116)/116) current with 28.6% more power pins as discussed in Section 3.2. The pin switching design can supply a larger current since it provides more power pins

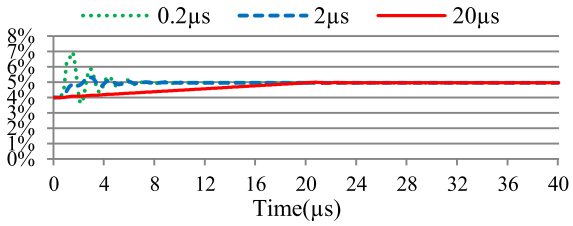

FIGURE 5. **Dynamic simulation.**

that reduce the package resistance. The percentage of current increase is less than that of power pin increase because the IR drop also depends on the resistance on the PCB and power grids.

To demonstrate performance benefits of the extra current, we employ two scenarios: dim silicon mode and dark silicon mode. As defined in Section 2, we refer dim silicon to the scenarios where all 8 cores are kept active but running at a lower frequency to comply with the power constraints. In dark silicon mode, we always boost the processor to its highest frequency while changing the number of active cores to satisfy the power limitations.

In addition, our processor power model shows that the extra current can boost the frequency of an 8-core processor from 2.0GHz to 3.0GHz in dim silicon mode. As listed in Table 3, the delivered power increases from 75.4W (0.65V×116A) to 111.6W (0.775V×144A) by 48.0% (i.e., (111.6−75.4)/75.4)). Note that the supply voltage is different for different processor frequency as shown in Table 4. In dark silicon mode, the extra current can enable 6 cores instead of 4 cores running at the same time at frequency 3.8GHz. In this mode, the delivered power increases from 101.5W (0.875V×116A) to 126.0W (0.875V×144A) by 24.1%. Figure 5 presents the dynamic IR drop while switching from I/O mode to power mode within 0.2μs, 2μs and 20μs. The IR drop fluctuation exceeds 5% for switching time 0.2μs and 2μs, while it is within 5% for 20μs case. Therefore, we use 20μs as the overhead for each pin switching operation.

Figure 6 plots the impedance for the default I/O mode and the pin switching mode. The impedance does not change much when switchable pins are used for power delivery.

D. POWER SWITCH

In our design, we use a large power transistor switch with ultra-low on-resistance and low parasitic capacitance. The switch is of comparably large size (like multiple NMOS or PMOS transistors connected in parallel). Figure 7 shows

TABLE 3. **Processor configurations under different cooling techniques.**

Configuration	Dim silicon mode		Dark silicon mode	
	Frequency (GHz)	Limitation	Frequency (GHz)	Limitation
Air cooling	8×1.6	Temperature < 85 ºC	4×1.8	Temperature < 85 ºC
Liquid cooling	8×2.0	Power<75.4W(0.65V×116A)	4×3.8	Power<101.5W(0.875V×116A)
Liquid cooling + Static pin switching	8×3.0	Power<111.6W(0.775V×144A)	6×3.8	Power<126.0W(0.875V×144A)
Liquid cooling + Dynamic pin switching	8×2.0 or 8×3.0	Power < 75.4W or 111.6W	4×3.8 or 6×3.8	Power < 101.5W or 126.0W

TABLE 4. Parameters of the performance and power models.

Parameters	Values
Technology	16 nm
Die area	10mm × 10mm
Voltage(V)	0.6, 0.625, …, 0.875, 0.9
Frequency (GHz)	1.6, 1.8, …, 3.8, 4.0
Fetch / Issue/ Commit Width	4/ 4/ 5
INT/ FP Window Size	96/ 64
LoadStore/ INT/ FP Units	2/ 2/ 3
Load/ Store Queue Size	80/ 80
Latency of INT ALU/ Mult/ Div	1/ 4/ 12 cycles
Latency of FP ALU/ Mult/ Div	1/ 2/ 10 cycles
L1 Instruction/ Data Cache Size	64/ 64 KB
L1 ICache/DCache Associativity	8/ 8
L1 Instruction/ Data Block Size	64/ 64 B
L2 Cache Size	8 MB
L2 Cache Associativity	8
L2 Cache Block Size	64 B
Memory parameters	
Number of channels	3
Frequency	800MHz
Data bus width	64
Peak memory bandwidth in I/O mode: 38.4GB/s = 3(channels)×2(double data rate)×0.8(GHz)×64(bus width)÷8(bits per byte)	
Peak memory bandwidth in power mode: 9.6GB/s = 3(channels)×2(double data rate)×0.2(GHz)×64(bus width)÷8(bits per byte)	

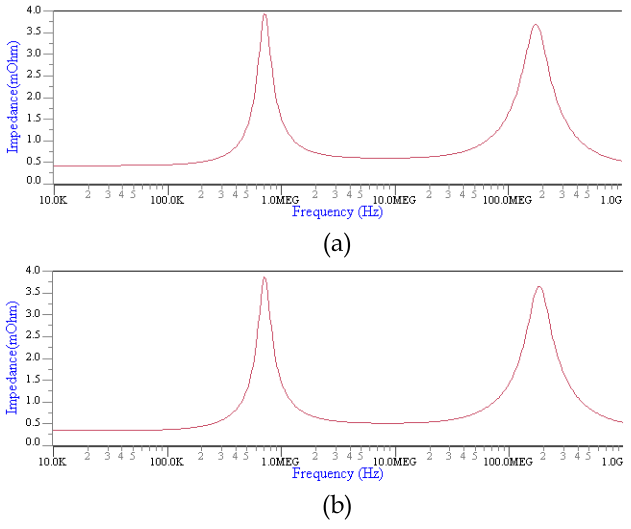

FIGURE 6. Impedance plots. (a) Default I/O mode. (b) Pin switching mode.

a layout design of such a large PMOS transistor switch of W/L = 80 based on 16nm technology [9]. Since the estimated resistance of the single switch is nearly 0.47Ω, we connected 262 switches in parallel to achieve the desired 1.8mΩ on-resistance [31] with a 0.232pf parasitic capacitance using 2601μm^2 of area overhead. Similar calculations for the large NMOS power switch show lower on-resistance and the same parasitic capacitance. The large power switches incur the main processor die area overhead of our design. For 144 switchable pins, they consume 0.00374544cm^2 of area

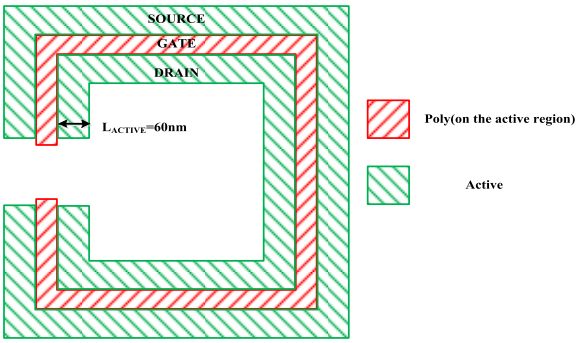

FIGURE 7. Layout of wrapped around large transistor.

on the processor die, incurring less than a 0.4% area overhead if the total die area is 1 cm^2.

E. SIGNAL TRANSMISSION

Figure 8 shows the circuits of switchable pin design. The switchable pin can either be used for power delivery or the signal transmission. To compensate the parasitic capacitance of the power switches, we add four tri-state buffers for a switchable pin since it can increase signal drive capability. We investigate the impact of adding power switches on the signal transmission path by observing the received eyes for memory writing and reading. We place both switches and buffers close to the processor, which can minimize the trace shared by power and signal lines. Since buffers on the signal line may cause an impedance mismatch, we have added 50Ω termination impedances on the side of memory devices to match the 50Ω transmission line; these minimize the signal reflections due to impedance mismatching. As shown in Figure 9, both eye diagrams show open eyes.

The pin switching design will cause delay on the signal transmission path since extra circuits are introduced as shown in Figure 8. For each I/O pin, simulation shows the extra circuits, two tri-state buffers and the two shift registers' components, cause 2.6ns delay in total.

F. THERMAL ISSUES

The switchable pins deliver more power in both the dim silicon and dark silicon modes as listed in Table 3. Our simulation shows air cooling is an unfeasible solution since the worst case processor temperature will be more than 100 °C in both the dim silicon and dark silicon modes which will result in serious reliability and lifetime issues. Therefore, we use traditional backplate liquid cooling [16] to increase heat dissipation while delivering more power via dynamic pin switching mechanism.

G. DYNAMIC PIN SWITCHING BASED ON PROGRAM PHASES

Programs tend to show phase behaviors, which can be classified as memory-intensive or computation-intensive. In our design, we will use the switchable pins for off-chip communication to achieve higher communication bandwidth during memory intensive phases. On the other hand, during computation intensive phases where the memory access

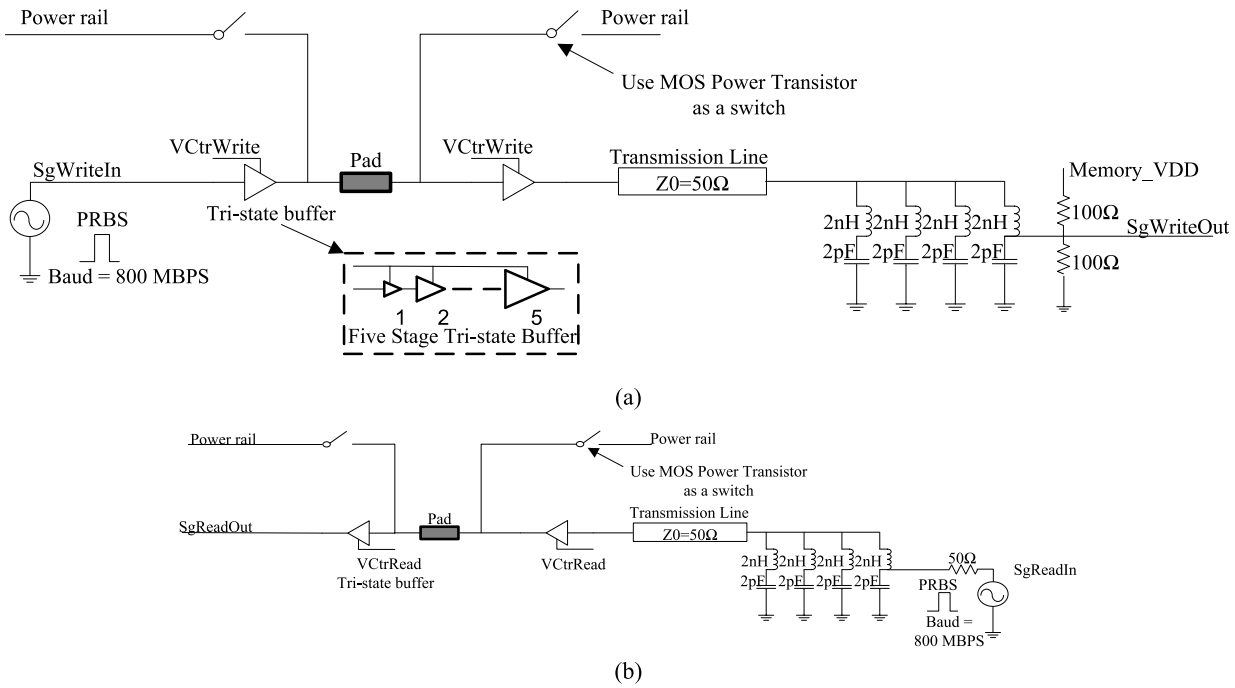

FIGURE 8. Circuits when a switchable pin is used for signal transmission. (a) Writing to memory (VCtrWrite=1, The tri-state buffers are enabled while the power switches are off). (b) Reading from memory (VCtrRead=1. The tri-state buffers are enabled while the power switches are off).

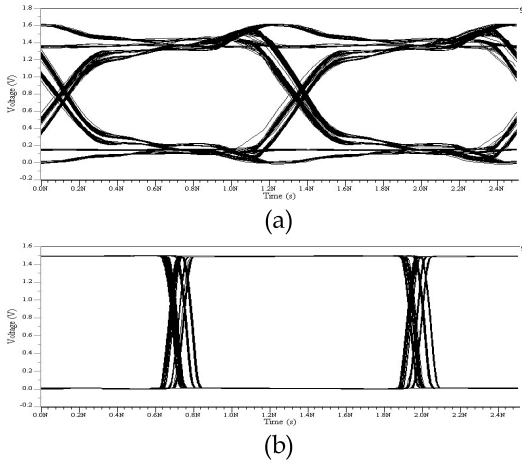

FIGURE 9. Received eye diagram. (a) Write to memory. (b) Read from memory.

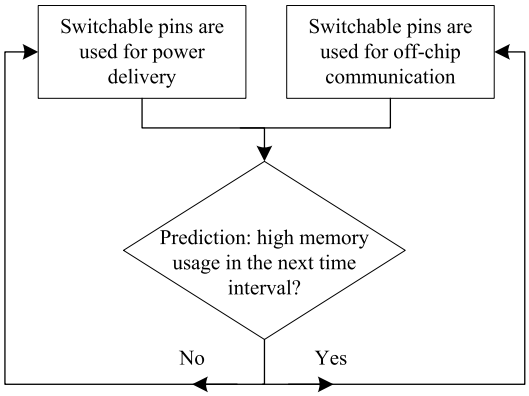

FIGURE 10. Workflow of dynamic switching.

frequency is low, the switchable pins can be utilized to deliver extra power to mitigate dark silicon. This extra power can either be used to activate dark cores or to increase the frequency of the running processors. Figure 10 illustrates the workflow of the pin switching mechanism which favors both the memory intensive and the computation intensive phases dynamically. A predictor, using the program's history (i.e., patterns of performance counters), is employed to predict the memory usage in the next time interval. When a memory-intensive phase is predicted, the switchable pins will be used for off-chip communication; otherwise, the switchable pins will be utilized to deliver power. Predictions are made in real-time, meaning incorrect predictions can be corrected in the next time interval.

IV. PREDICTION MODEL

In this section, we describe the prediction model training procedure employed by the dynamic pin switching scheme. In general, the goal of the predictor is to determine whether the bandwidth requirement of the upcoming intervals is high (or low) enough to require a pin switching for optimal performance. The prediction model is trained as follows.

First, we run workloads on the processor and collect common performance metrics including branch

mispredictions and cache misses from all cores and shared components at a preset frequency. By doing this, we obtain the following tuple from each time interval.

$$\langle X_1^1, X_1^2, ..X_1^p, X_2^1 ... X_2^p ... X_q^p, X_S^1, ..X_S^r, \text{MB}\rangle$$

In the above expression, each variable X_a^b represents a performance metric of a specific component. The subscript a is the component identity (e.g., core ID) and the superscript b corresponds to the index of the metric. For example, X_1^2 denotes the second performance metric observed on the first core. We assume that the number of cores on chip is q and we monitor p performance metrics for each of them. This results in a total of $p \times q$ metrics from the integrated cores. The r variables with the subscript S (i.e., X_S^1 through X_S^r) indicate the performance metrics from shared components such as the last-level cache. In this work, we collect 180 counters from each core and 20 counters from the shared components for each time interval. The notation MB represents the average memory bandwidth of this interval.

Second, we reorganize the collected data and train a statistical model to correlate the historical execution behaviors and the memory bandwidth in future intervals. To form a training instance, we combine input variables (i.e., all X_a^b) from M consecutive intervals and use all of them as the input for this sample. The response value (i.e., output) of this training instance is a Boolean flag which is defined as follows. We calculate the average bandwidth of intervals M+1 to M+N; if the average value is greater than a preset threshold, we set the flag to 1, indicating the following N intervals require high memory bandwidth. In contrast, if the average bandwidth is less than the threshold, the flag will be set to 0. By doing this, we are essentially building a rigorous relationship between past execution behaviors (i.e., interval 1 to M) and the future bandwidth requirement (interval M+1 to M+N). After obtaining these training instances, we employ a regression tree model [17] to select 10 input factors that most significantly impact the output value (i.e., the Boolean flag). We then feed the chosen 10 variables, along with their corresponding responses, to a model implementing a bump-hunting algorithm [19] in order to generate a set of rules to guide the pin switching. The rules are interpreted in a group of "IF-ELSE" conditions and are able to identify the regions with the maximum output values. We keep comparing the collected performance metrics to the generated rules at runtime. When the conditions are satisfied, a pin switching will be triggered to deliver more power or bandwidth to the processor to improve performance. Note that we randomly sample 80% of all the instances for training and use the remaining 20% for validation as the conventional statistical model training does.

V. EXPERIMENTAL SETUP

We simulate an 8-core chip multiprocessor (CMP) and set the maximum allowable temperature to be 85 °C. Power constraints lead to numerous execution modes in terms of different core frequencies and the number of active cores. For example, decreasing the frequency is effective for reducing per-core power consumption, thus enabling more cores to run simultaneously without exceeding the power limits. For simplicity, we conduct two groups of studies to make our observations and conclusions more comprehensive. The first category of the study is mainly concentrated on the dim silicon mode, while the second category of the study is focused on the dark silicon mode. Note that this work focuses on demonstrating the effectiveness of the pin switching design instead of comparing these two modes. We explore 13 frequency levels from 1.6GHz to 4.0GHz with a step frequency of 200MHz on the target CMP. Thermal and power constraints cause the core frequency and number of active cores to be different depending on the execution mode. The specific configurations of each execution mode are listed in Table 3.

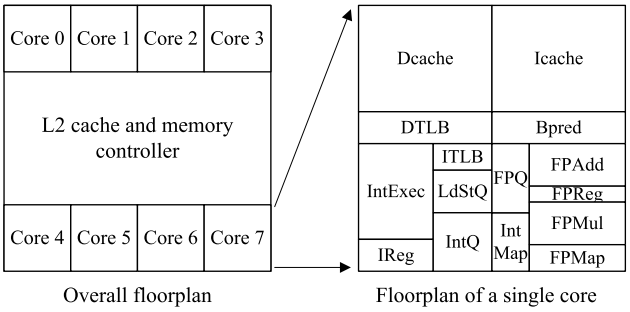

FIGURE 11. Floorplan of the chip multiprocessor.

We use Gem5 [13] to conduct our performance evaluation, and use McPAT [29] for processor power modeling. The corresponding parameters are listed in Table 4. We modify HotSpot [7] to simulate the floorplan shown in Figure 11 using air cooling and backplate liquid cooling. Both multi-threaded benchmarks (i.e., SPLASH-2 [8], PARSEC [12], ALPBench [30]) and the SPEC2006 [10] multi-program benchmarks are used in the evaluation of dim silicon mode. All multi-threaded applications are running with 8 threads and are executed until completion to guarantee that the task performed is identical across different simulations. Eight copies of the representative instructions are used to create a multi-program workload. The multi-program workloads can be categorized into two types: the first mixes eight copies of the identical SPEC2006 programs, while the second type mixes eight copies of different SPEC2006 programs. As listed in Table 5, P8MIX1 and P6MIX1 are 8-copy and 6-copy multi-program workloads used for dim silicon and dark silicon evaluation separately. We use multi-program workloads (each includes 6 copies of SPEC programs) for the study in the dark silicon mode. We implement fine-grained multithreading [35] so that all threads take turns being active and thus will be finished at roughly the same time even when the number of active cores is less than the number of threads. This can achieve the best performance by fully utilizing all active cores.

TABLE 5. Simulated multi-program workloads.

Name	Combinations
P8MIX1	4×NAMD + 4×MCF
P8MIX2	4×NAMD + 4×BZIP2
P8MIX3	4×BZIP2 + 4×SJENG
P8MIX4	2×BZIP2 + 2×DEALII + 1×HMMER + 1×GOBMK + 1×H264REF + 1×SJENG
P6MIX1	3×NAMD + 3×MCF
P6MIX2	3×NAMD + 3×BZIP2
P6MIX3	3×BZIP2 + 3×SJENG
P6MIX4	1×BZIP2 + 1×DEALII + 1×HMMER + 1×GOBMK + 1×H264REF + 1×SJENG

As for the prediction model and online pin switching, we use the execution behaviors in the previous three time intervals to predict the bandwidth in the next interval, with each interval lasting for one millisecond. In this case, the 20μs overhead discussed in Section 3.4 is 2% of a time interval. Another important parameter is the memory bandwidth threshold, which is used to evaluate if a pin switching is needed. The threshold should be less than 9.6GB/s, which is the peak memory bandwidth in power mode as listed in Table 4. We set the bandwidth threshold to be 1.6GB/s in this work to achieve optimal overall performance. Note that these empirically selected parameters do not impact the effectiveness of our proposed scheme and can be changed to other values in a practical system.

VI. RESULT ANALYSIS

In this section, we demonstrate the effectiveness of the pin switching mechanism by comparing the performance between traditional designs and our proposed scheme.

A. RULES EXPLANATION

We start by analyzing the generated rule-set used to guide pin switching at runtime. In the dim silicon execution mode, all 8 cores are kept busy and the processor frequency can switch between 2.0GHz and 3.0GHz. Since the frequency is changing, two individual prediction models are necessary to guide the power-to-I/O (i.e., 3.0GHz to 2.0GHz) and the I/O-to-power (i.e., 2.0GHz to 3.0GHz) pin switching. Recall that our pin switching technique is, in essence, a one-way conversion. Therefore, the switch from power to I/O mode means the procedure of returning to the default I/O pin configuration. Assuming that the switchable pins are currently on the power path and the processor is running at 3.0GHz, the following rules indicate that the upcoming interval is very likely to be memory-intensive where off-chip memory access is frequent, and therefore the switchable pins should be switched to the I/O path:

$$\begin{aligned} &Int3\_L2\_totalmiss > 36929 \;\&\& \\ &Int2\_L2\_totalmiss > 32100 \;\&\& \\ &Int1\_L2\_totalmiss > 3630 \end{aligned}$$

The conditions are expressed in a format of IntID_component_metric > X, meaning that the performance counter metric of component in interval IntID (one of the M intervals used as input) should be larger than a certain value X. Given this notion, the first condition in the rule-set listed above indicates that the total misses (read misses and write misses) in the L2 cache in the immediately preceding interval should be larger than 36929; recall that we use 3 intervals to predict the ensuing interval. Similarly, the second and third conditions set the lower bound for the L2 total misses in the second (Int2_L2_totalmiss) and the first previous intervals (Int1_L2_totalmiss) respectively. The rules identify memory-intensive execution periods by testing the conditions in each interval, and guide switchable pins to switch to the I/O path or stay in power delivery.

When the switchable pins have been set for signal transmission and the processor is running at lower frequency, we also need a rule-set to govern when to switch to the power path. The corresponding rules are listed as follows.

$$\begin{aligned} &Int3\_L2\_readmiss < 1848 \;\&\& \\ &Int1\_L2\_readmiss < 9192 \;\&\& \\ &Int1\_L2\_totalmiss < 6171 \end{aligned}$$

The rules can be explained similarly and we thereby omit the analysis. We also train prediction models for the dark silicon execution for which the rules are similar to those of dim silicon mode and therefore are not listed due to the space limitation.

B. DIM SILICON RESULT

Recall that in the dim silicon mode all 8 cores are enabled while running at a low frequency determined by the power delivery and cooling configurations listed in Table 3. Figure 12 shows the normalized performance for multi-program and multi-threaded workloads under four evaluated configurations. In the air cooling mode, the 8 cores are running at 1.6GHz because the TDP, restricted by thermal constraints, is relatively small. Using liquid cooling, we are able to raise the frequency to 2.0GHz. The remaining two configurations both implement the pin switching mechanism using liquid cooling; therefore there is extra power allowing the core frequency to go up to 3.0GHz. These two final configurations use different pin switching schemes. The first uses a static scheme in which the switchable pins are always set to the power delivery path throughout the entire execution. The second configuration uses a dynamic pin switching scheme guided by the prediction model. Note that all results are normalized against the baseline configuration which uses air cooling.

As shown in Figure 12, the scaling trends for most benchmarks are reasonable because higher frequencies lead to faster execution. However, the relative performance among the four configurations is different for different benchmarks. For example, while running 8 copies of MCF and DEALII, the static switch scheme (3.0GHz) has longer execution time compared to the runs with a lower frequency (2.0GHz).

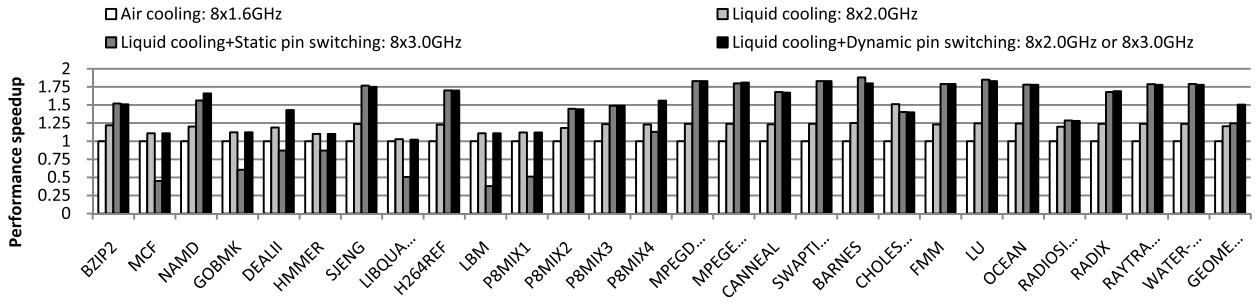

FIGURE 12. **Performance speedup when the processor is in dim silicon mode.**

TABLE 6. **Number of L2 cache misses per 1K instructions on a processor configured to 8×2.0GHz (liquid coiling).**

Workloads	BZIP2	MCF	NAMD	GOBMK	DEALII	HMMER	SJENG	LIBQUANTUM	H264REF
L2 cache misses per 1K inst.	3.421	61.727	0.059	8.173	8.350	2.784	0.438	49.560	0.436
Workloads	LBM	P8MIX1	P8MIX2	P8MIX3	P8MIX4	MPEGDEC	MPEGENC	CANNEAL	SWAPTIONS
L2 cache misses per 1K inst.	30.216	32.460	0.414	0.817	1.450	0.020	0.060	1.583	0.010
Workloads	BARNES	CHOLESKY	FMM	LU	OCEAN	RADIOSITY	RADIX	RAYTRACE	WATER-SPATIAL
L2 cache misses per 1K inst.	0.012	2.955	0.260	0.016	0.363	0.052	4.450	0.047	0.030

TABLE 7. **Prediction accuracy on a processor in dim silicon mode.**

Workloads	BZIP2	MCF	NAMD	GOBMK	DEALII	HMMER	SJENG	LIBQUANTUM	H264REF
Prediction accuracy	0.86	1	0.96	0.8	0.98	0.98	0.76	1	0.9
Workloads	LBM	P8MIX1	P8MIX2	P8MIX3	P8MIX4	MPEGDEC	MPEGENC	CANNEAL	SWAPTIONS
Prediction accuracy	0.94	1	0.85	0.82	0.8	1	1	1	1
Workloads	BARNES	CHOLESKY	FMM	LU	OCEAN	RADIOSITY	RADIX	RAYTRACE	WATER-SPATIAL
Prediction accuracy	1	0.9	0.95	0.95	0.9	1	0.95	0.96	0.94

Similar trends can also be observed from the execution of P8MIX1, which includes the memory-intensive program MCF. The main reason for the longer execution time here is the substantial penalty from lower memory bandwidth in 3.0GHz compared with 2.0GHz case. More details will be given shortly to expound upon this observation. On the other hand, for applications that are intrinsically computation-intensive, executions using the pin switching technique will significantly outperform those with traditional configurations. Typical examples include H264REF from SPEC2006 and MPEGDEC from ALPBench. For these benchmarks, even the static pin switching leads to an impressive speedup because memory-bound intervals are fairly rare during the execution. Therefore, maintaining a higher core frequency is more beneficial.

Furthermore, in benchmark DEALII when the pin switching is guided by the prediction model, we notice further performance enhancement compared with the static switching. This is because with the dynamic approach, the predictor will estimate how much off-chip traffic will be generated during upcoming execution period, thus determining the most appropriate path for the switchable pins. Compared with the static scheme which blindly sets the switchable pins to the power delivery path, the dynamic switching strategy can more effectively balance the requirement of power delivery and off-chip bandwidth. In general, the geometric mean of the performance speedup delivered by our optimal scheme (liquid cooling + dynamic pin switching) is 1.50X compared with the baseline (air cooling), while the static pin switching and liquid cooling accelerate the execution by 1.24X and 1.20X respectively.

To further understand the scaling trend of each workload, we plot the number of L2 cache misses per 1K instructions in Table 6. The figure shows whether a workload is computation-intensive or memory-intensive. In addition, a high-accuracy predictor stands as one of the most important factors in determining the effectiveness of dynamic switching; therefore it is necessary to evaluate the accuracy of our prediction model. Recall that, in our model, the response of each training instance is set as a Boolean flag. Consequently, by counting the occurrences of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), we calculate the prediction accuracy as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

As shown in Table 7, the prediction accuracy is fairly high for most benchmarks. For applications

TABLE 8. Number of L2 cache misses per 1K instructions on a processor configured to 4×3.8 GHz.

Workloads	BZIP2	MCF	NAMD	GOBMK	DEALII	HMMER	SJENG
L2 cache misses per 1K instructions	3.411643	56.41	0.058589	4.095244	3.1	2.20258	0.24
Workloads	LIBQUANTUM	H264REF	LBM	P6MIX1	P6MIX2	P6MIX3	P6MIX4
L2 cache misses per 1K instructions	40.03515	0.303765	30.21711	26.51531	0.290195	0.581876	0.48

TABLE 9. Prediction accuracy on a processor in dim silicon mode.

Workloads	BZIP2	MCF	NAMD	GOBMK	DEALII	HMMER	SJENG
Prediction accuracy	0.99	0.99	0.99	0.83	0.9	0.98	1
Workloads	LIBQUANTUM	H264REF	LBM	P6MIX1	P6MIX2	P6MIX3	P6MIX4
Prediction accuracy	1	1	0.9	0.82	0.85	0.83	0.92

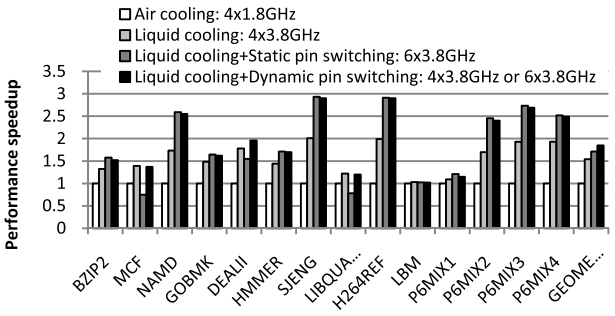

FIGURE 13. Performance speedup when the processor is in dark silicon mode.

(e.g. BZIP2 and SJENG) where the accuracies are slightly lower, the predictor still results in impressive performance improvements over the static switching scheme.

C. DARK SILICON RESULT

We now shift focus to the dark silicon mode. Figure 13 compares the performance of 14 multi-program workloads, each of which contains 6 programs from SPEC2006. The figure shows that using the pin switching technique always yields significant speedup for computation-intensive benchmarks. For example, for benchmark H264REF, the geometric means of the speedups under dynamic switching, static switching and liquid cooling are 2.9X, 2.91X, and 1.99X, respectively. The slightly lower speedup in dynamic switching compared with static switching is because the processor is conservatively set to low power mode (i.e., 4×3.8GHz) to avoid memory penalties when a program starts to run.

The performance scaling for memory-bound workloads is obviously different compared with the computation-intensive counterparts. More specifically, the static switching approach lags behind all other strategies while dynamic switching delivers comparable performance improvement as the liquid cooling (4×3.8GHz). This is because this set of benchmarks is more sensitive to the change in memory performance. Nevertheless, our predictor can effectively identify the execution behaviors of these workloads and force the switchable pins to mainly stay on the I/O path, thus minimizing the performance degradation.

In general, the dynamic switch leads to 1.85X speedup while the static switch and liquid cooling respectively accelerates the executions by 1.71X and 1.54X. Table 8 shows the number of L2 misses per 1K instructions for the evaluated workloads. Obviously, workloads such as MCF (i.e., memory-intensive) result in much larger L2 misses than H264REF (i.e., computation-intensive), thus leading to the scaling behavior shown in Figure 13. The prediction accuracies are plotted in Table 9. As can be seen, our model is sufficiently accurate to capture the memory-bound and compute-bound phases at runtime and facilitate the system to optimize the overall performance by connecting the switchable pins to the most suitable path.

D. ENERGY EFFICIENCY EVALUATION

We evaluate the energy efficiency for all configurations in both the dim silicon and dark silicon modes and plot their normalized geometric means in Figure 14. The liquid cooling configurations cost more energy since the processor works at a higher frequency compared to the corresponding air cooling configurations. Interestingly, the dynamic pin switching configuration is the best in metrics Energy-Delay-Squared Product (ED2) since remarkable performance speedups are achieved compared to the baseline air cooling design.

E. THERMAL ANALYSIS

To investigate the corresponding thermal issues, we conduct a series of simulations and show the results in Figure 15. The baseline configurations in the dim silicon and dark silicon modes are individually 8×1.6GHz and 4×1.8GHz. With air cooling, their maximum temperatures, individually 84.9 °C and 83.7 °C, are very close to the safe temperature threshold of 85 °C. Therefore, we utilize liquid cooling to lower the thermal constraints and allow the processor to run at a higher frequency with more enabled cores. In the dim silicon mode, the temperature of a processor using dynamic pin switching (8×2.0GHz or 8×3.0GHz) is

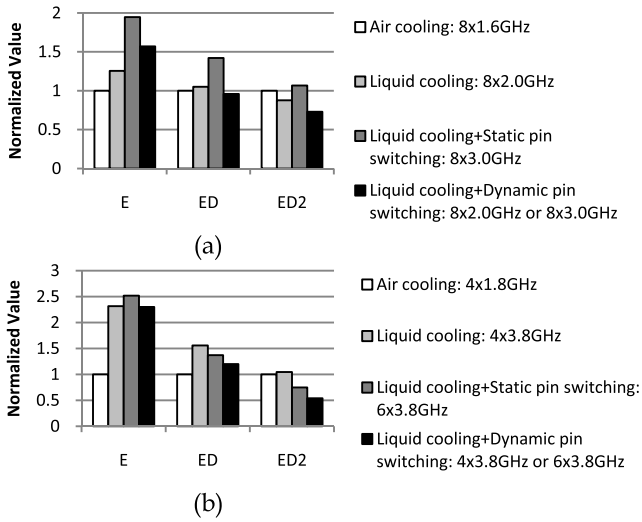

FIGURE 14. Energy and performance evaluation (E: Energy, ED: Energy-Delay Product, ED^2: Energy-Delay-Squared Product). (a) Dim silicon mode. (b) Dark silicon mode.

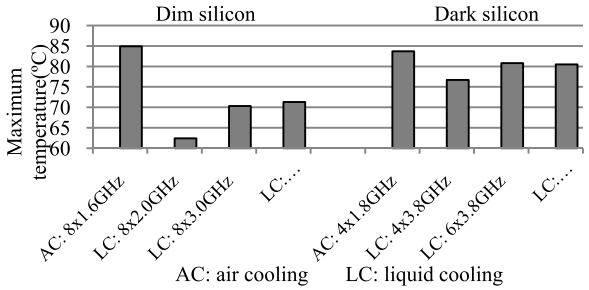

FIGURE 15. Thermal study on the dim and dark silicon cases.

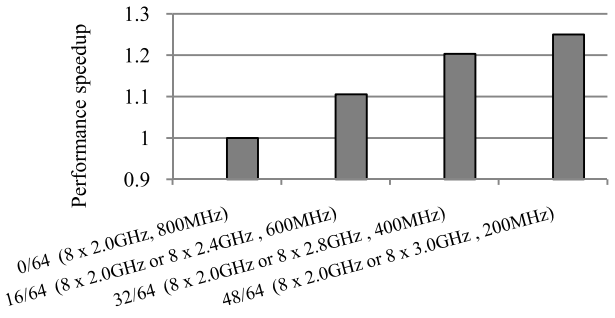

FIGURE 16. Sensitivity study on designing different ratio of switchable pins in a data path. (The numbers in brackets mean the processor frequency and memory frequency individually.)

only 1.0°C higher than that of a processor using static pin switching (8×3.0GHz) while achieving much better performance. Similarly, this can also be observed in the dark silicon mode.

F. SENSITIVITY STUDY

Alternatively, a different portion of pins in a data path can be designed as switchable pins. When the number of switchable pins per data path is increased, cores can be delivered with more power and run at a higher frequency at cost of lower off-chip bandwidth. We conduct this study to find the optimal portion for the performance of the workloads by comparing performance of the dynamic switching strategy when there are 16, 32, and 48 switchable pins per data path in the dim silicon mode respectively. This will create multi-level (16/64, 32/64, 48/64) pin switching, facilitating design space exploration. Figure 16 shows the performance speedup in geometric mean when the processor is configured with different pin switching levels in the dim silicon mode. We see a higher ratio of switchable pins achieves better performance. This work uses the pin switching level 48/64 to explore the power delivery capacity of the pin switching mechanism.

VII. CONCLUSION

With the stall of Dennard scaling, dark silicon is gradually becoming a daunting conundrum that threatens the scaling of Moore's Law in the future. While thermal constraint are widely believed to be the main cause of this phenomenon, the limited number of pins on the chip package also confines the maximum number of simultaneously active transistors, thus preventing us from obtaining a sufficient performance improvement by increasing transistor density. To mitigate this limitation, we propose a novel mechanism to dynamically switch a portion of I/O pins to power pins in order to light up dark silicon by delivering extra power. We also employ an advanced statistical model to train a prediction model that can be employed by the OS to govern the pin switching. Our evaluation results demonstrate that the proposed pin switching mechanism can remarkably enhance the overall performance compared with conventional designs.

ACKNOWLEDGEMENT

The authors acknowledge the computing resources provided by the Louisiana Optical Network Initiative (LONI) HPC team. Finally, they appreciate invaluable comments from anonymous reviewers.

REFERENCES

[1] *Mentor Graphics SPICE Simulator ELDO*. [Online]. Available: http://www.mentor.com/products/ic_nanometer_design/analog-mixed-signal-verification/eldo/, accessed Jul. 1, 2015.
[2] *COMSOL Multiphysics*. [Online]. Available: http://www.comsol.com/, accessed Jul. 1, 2015.
[3] *4-Bit Parallel-to-Serial Converter*. [Online]. Available: http://www.micrel.com/_PDF/HBW/sy10-100e446.pdf, accessed Jul. 1, 2015.
[4] *4-Bit Serial-to-Parallel Converter*. [Online]. Available: http://www.micrel.com/_PDF/HBW/sy10-100e445.pdf, accessed Jul. 1, 2015.
[5] *Intel Xeon Processor E5-2450L*. [Online]. Available: http://ark.intel.com/products/64610/Intel-Xeon-Processor-E5-2450L-20M-Cache-1_80-GHz-8_00-GTs-Intel-QPI, accessed Jul. 1, 2015.
[6] *ITRS Assembly and Packaging Technical Working Group*. [Online]. Available: http://www.itrs.net/Links/2012ITRS/2012Tables/AssemblyPkg_2012Tables.xlsx, accessed Jul. 1, 2015.
[7] *HotSpot*. [Online]. Available: http://lava.cs.virginia.edu/HotSpot/, accessed Jul. 1, 2015.
[8] *Modified SPLASH-2 Benchmarks*. [Online]. Available: http://www.capsl.udel.edu/splash/, accessed Jul. 1, 2015.
[9] *Model for a 16 nm, 0.9 V Process*. [Online]. Available: http://ptm.asu.edu/modelcard/LP/16nm_LP.pm, accessed Jul. 1, 2015.

[10] (2006). *SPEC CPU 2006*. [Online]. Available: http://spec.org/cpu2006/, accessed Jul. 1, 2015.
[11] H. Barowski *et al.*, "Heat sink integrated power delivery and distribution for integrated circuits," U.S. Patent 2012 0 106 074 A1, May 3, 2012.
[12] C. Bienia, S. Kumar, J. P. Singh, and K. Li, "The PARSEC benchmark suite: Characterization and architectural implications," in *Proc. 17th Int. Conf. Parallel Archit. Compilation (PACT)*, 2008, pp. 72–81.
[13] N. Binkert *et al.*, "The gem5 simulator," *ACM SIGARCH Comput. Archit.*, vol. 39, no. 2, pp. 1–7, 2011.
[14] S. Chen *et al.*, "Increasing off-chip bandwidth in multi-core processors with switchable pins," in *Proc. 41st Annu. Int. Symp. Comput. Archit. (ISCA)*, 2014, pp. 385–396.
[15] H. David, C. Fallin, E. Gorbatov, U. R. Hanebutte, and O. Mutlu, "Memory power management via dynamic voltage/frequency scaling," in *Proc. 8th ACM Int. Conf. Auto. Comput. (ICAC)*, 2011, pp. 31–40.
[16] Y. Deng and J. Liu, "Optimization and evaluation of a high-performance liquid metal CPU cooling product," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 3, no. 7, pp. 1171–1177, Jul. 2013.
[17] L. Duan, B. Li, and L. Peng, "Versatile prediction and fast estimation of architectural vulnerability factor from processor performance metrics," in *Proc. IEEE 15th Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2009, pp. 129–140.
[18] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *Proc. 38th Annu. Int. Symp. Comput. Archit. (ISCA)*, 2011, pp. 365–376.
[19] J. H. Friedman and N. I. Fisher, "Bump hunting in high-dimensional data," *Statist. Comput.*, vol. 9, no. 2, pp. 123–143, 1999.
[20] N. Goulding-Hotta *et al.*, "The GreenDroid mobile application processor: An architecture for silicon's dark future," *IEEE Micro*, vol. 31, no. 2, pp. 86–95, Mar./Apr. 2011.
[21] N. Hardavellas, M. Ferdman, A. Ailamaki, and B. Falsafi, "Power scaling: The ultimate obstacle to 1 K-core chips," Dept. Elect. Eng. Comput. Sci., Northwestern Univ., Evanston, IL, USA, Tech. Rep. NWU-EECS-10-05, 2010.
[22] N. Hardavellas, M. Ferdman, B. Falsafi, and A. Ailamaki, "Toward dark silicon in servers," *IEEE Micro*, vol. 31, no. 4, pp. 6–15, Jul./Aug. 2011.
[23] B. Jacob, S. W. Ng, and D. T. Wang, *Memory Systems: Cache, DRAM, Disk*. Amsterdam, The Netherlands: Elsevier, 2008.
[24] R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Köse, and E. G. Friedman, *Power Distribution Networks With On-Chip Decoupling Capacitors*. New York, NY, USA: Springer-Verlag, 2011.
[25] D. H. Kim *et al.*, "3D-MAPS: 3D Massively parallel processor with stacked memory," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, Feb. 2012, pp. 188–190.
[26] J. Kim, J. Shim, J. S. Pak, and J. Kim, "Modeling of chip-package-PCB hierarchical power distribution network based on segmentation method," in *Proc. Elect. Design Adv. Packag. Syst. Symp.*, Dec. 2008, pp. 85–88.
[27] W. Kim, D. M. Brooks, and G.-Y. Wei, "A fully-integrated 3-level DC/DC converter for nanosecond-scale DVS with fast shunt regulation," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, Feb. 2011, pp. 268–270.
[28] K. L. Kishore and V. S. V. Prabhakar, *VLSI Design*. New Delhi, India: I.K. International Publishing House, 2009.
[29] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proc. 42nd Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Dec. 2009, pp. 469–480.
[30] M.-L. Li, R. Sasanka, S. V. Adve, Y.-K. Chen, and E. Debes, "The ALPBench benchmark suite for multimedia applications," Dept. Comput. Sci., UIUC, Champaign, IL, USA, Tech. Rep. UIUCDCS-R-2005-2603, Jul. 2005.
[31] T. N. Miller, X. Pan, R. Thomas, N. Sedaghati, and R. Teodorescu, "Booster: Reactive core acceleration for mitigating the effects of process variation and application imbalance in low-voltage chips," in *Proc. IEEE 18th Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2012, pp. 1–12.
[32] A. Raghavan *et al.*, "Computational sprinting," in *Proc. IEEE 18th Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2012, pp. 1–12.
[33] B. M. Rogers, A. Krishna, G. B. Bell, K. Vu, X. Jiang, and Y. Solihin, "Scaling the bandwidth wall: Challenges in and avenues for CMP scaling," in *Proc. 36th Annu. Int. Symp. Comput. Archit. (ISCA)*, 2009, pp. 371–382.
[34] M. B. Taylor, "Is dark silicon useful? Harnessing the four horsemen of the coming dark silicon apocalypse," in *Proc. 49th ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Jun. 2012, pp. 1131–1136.
[35] D. M. Tullsen, S. J. Eggers, and H. M. Levy, "Simultaneous multithreading: Maximizing on-chip parallelism," in *Proc. 22nd Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 1995, pp. 392–403.
[36] R. Zhang, B. H. Meyer, W. Huang, K. Skadron, and M. R. Stan, "Some limits of power delivery in the multicore era," in *Proc. Workshop Energy Efficient Design (WEED)*, 2012, pp. 1–7.
[37] R. Zhang, K. Wang, B. H. Meyer, M. R. Stan, and K. Skadron, "Architecture implications of pads as a scarce resource," in *Proc. ACM/IEEE 41st Int. Symp. Comput. Archit. (ISCA)*, Jun. 2014, pp. 373–384.

Shaoming Chen, photograph and biography not available at the time of publication.

Lu Peng, photograph and biography not available at the time of publication.

Yue Hu, photograph and biography not available at the time of publication.

Zhou Zhao, photograph and biography not available at the time of publication.

Ashok Srivastava, photograph and biography not available at the time of publication.

Ying Zhang, photograph and biography not available at the time of publication.

Jin-Woo Choi, photograph and biography not available at the time of publication.

Bin Li, photograph and biography not available at the time of publication.

Edward Song, photograph and biography not available at the time of publication.