Temperature Sensitive Microarchitecture Design Circuit Design

Tamanna Afroze
MSc. Student
Computer Science and Engineering, BUET

S. M. Farhad
Associate Professor
Computer Science and Engineering, BUET

ABSTRACT

Microprocessors are designed with very tiny microchips and heat induced due to operation makes the chip deteriorate their performance in many extents. Heat causes a portion of chip-area to get beyond tolerable temperature range which can degrade performance of many applications in chip-level. This work addresses many issues in this range. The main contribution of the work lies in reconsidering the heat transition among chips, or inside a chip in order to decrease heat inside a microprocessor. With this end in view, a renewed design architecture in circuit-level has been considered. To design the total work inside microprocessor in response to dynamic temperature change a different level of operation-mechanism has been proposed. To watch the applications running in pipeline, and then by utilizing slack time in hardware level this work wants to improve performance of the processor.

In brief, total work proposes two new heat-control mechanisms, one is at operation-level and the other is at architectural-level. At operation-level, this work proposes a prediction mechanism to predict the useful operations inside the microprocessor that performs as a sink for heat dissipation. At architectural-level, this work proposes a drain system. This work has simulated the proposed system using Matlab and observed that the system works perfectly well. A comparison with existing mechanism has been devised which shows the proposed work increases performance of running application.

General Terms

Computer Architecture Design.

Keywords

Heat detection, watcher, application, circuit, logical operation, logic-gates

1. INTRODUCTION

Microprocessors are the base part of a computer of our day-to-day computing system. Heat generation limits processor performance and limits the number of transistors incorporated in a single chip. Research on how to reduce the generated heat, is gaining popularity due to its many aspects of improvement [1, 2, 3, 4, 5, 6, 7].Different layer of microprocessor including clocked pulse is also playing significant role in performance improvement. As the performance improvement continues so does the heat generation both in chip and architectural level [6]. Each of these works has addressed several aspects of improving performance that have induced heat. Bit- partitioning mechanism improves performance, but the proposed chip-level architecture increases fabrication area [5]. Moreover, the super-scaling operation like decoder spacing expansion in this

system requires logic states to remain active simultaneously that in turn increases heat at chip-level.

Very little work addressed the issue of operational techniques for heat control. Architecture-level issues for controlling heat by register-transistor logic is discussed in [15]. We addressed the logical gates operation for performance improvement and thermal awareness. Some operations can be predicted early which can increase performance and/or propagation delay induced by many gates causes transition of heat to increase.

Microchips architecture is designed using universal gates, that is, NAND and NOR gates. We are designing dynamically controlled heat detector by detecting power. In that way, we can make our processor chips cool and, simultaneously longevity may increase. Concurrently, it will give runtime improvement or working time improvement. Runtime is the time which is the active time of a particular operation or a particular mechanism. It indicates the operations improvement or how much time a circuit remains active which in turn makes the circuit heated. This work has shown how runtime affects chip's output in operation and heat control.

NAND and NOR gates are designed in the chip level by the use of inverter logic gate. If we can add some logic for inverters, so that inverter gates will be aware of temperature and will take necessary operation for temperature minimization, then we can achieve some good logic for temperature minimization with low cost. A new method of heat detection with RTL logic in the chip-level layer absorbing the generated heat is presented in [3]. Changing clock frequency and checking delay or overhead of operation is essential in performance improvement [5]. But extra logic for this operation may increase time. We want to predict some operation early to check the operation of some gates and then make the performance improved, heat reduction simultaneously.

Use of different metal instead of silicon is also a think for temperature management. We can use Murcury in doping to fabricate the semicondector. Not always we can use other metal with optimum cost. Some research is necessary for gaining some good effect. Isotopes of Fe (Iron) can be used in a small portion to increase conductivity. This thought is motivated by the general observation of our daily activities. A normal day-to-day observation can motivate the mechanism: If we store water in the bucket, the water remains cold depending on the surrounding environment. So, if we can flow some thermally cold metal beside the chips or tiny gates it can also significantly reduce temperature with improving performance of running application though. \\

In this paper, we propose two mechanisms that reduce the heat inside processor. For circuit level optimization, we need to check the level of minimum voltage, which can keep the logic states without attaining the highest voltage levels. We

can add capacitor for this purpose. In the second methodology, if we can add different metal instead of the silicon chip for storing the voltage levels we can get an optimum result for temperature management.

The main message of this paper we want to deliver is that smart viewing of applications and runtime improvement of those running application can significantly improve temperature reduction process and simultaneously improve performance. In this work we have also considered smart error and fault detection with tolerable fault tolerance. Watcher needs to take care of performance of the running application, and tries to find out the slack time and stall times to inject other necessary instruction checking to improve performance.

2. RELATED WORKS

With the advent of newer techniques of multiprocessor design, temperature and power has become a crucial issue in chip designing. Every design focus on power consumption, power dissipation and heat tolerance to improve the microchip designing keeping the performance of program execution better and reliable. Power and heat management is not only related to microchip designing mechanisms, but also essential for network-chip designing process.

Most of the work addressed the internal parameters of microchips. Simultaneously, register file, memory store/load operation, reading registers, and cache architecture also came into the heat control and detection mechanism.

Heat control by using separate hardware architecture and bit-partitioning method is illustrated in [1]. The author considered an extra efficient mechanism by inducing newer memory chips for accessing register files. Bit-partitioned Register File (BPRF) considered their designing mechanism from basic cache organization mechanism. It is designed for, in fact, designed based on a conventional dynamically scheduled superscalar processor. They showed how much energy is consumed in the separated bank of register files, and bit partitioned method, while preserving early de-allocation of registers usage for processors performance. Energy, that is, otherwise related to power consumption greatly improves performance of microprocessor in this methodology.\\

In [3], the authors considered heat detection and control mechanism for Architectures. They designed the hot-leakage mechanism for the micro-architecture going inside into the main digital logic designing gates, like NAND or NORs CMOS fabrication level. They showed how temperature leak can be controlled in caches by using several techniques, like, lowering the Quiescent Vdd, multiple threshold CMOS, Drowsy caches, hot-leakage parameters. Dynamic thermal management by monitoring chip-wide temperature at run-time and dynamically inducing power reduction schemes is discussed in [2].Reducing register ports for memory read/write operation in explained in [4]. Thermal relationship and thermal management for subarrayed data cache has been discussed in [2, 13, 16].

Skadron proposed most of the temperature control mechanisms in literature [3, 11, 15]. His several publications show many different improvements for heat control. In [15], the authors showed a micro-architecture, which is temperature aware. In brief, they showed chip-level hardware techniques for good illustration of both the benefits and challenges of runtime thermal management. They named the architecture as hot-spot. In this paper, they showed how to find out the het-induced chip and then to reduce the heat of the detected chip by using RTL circuits. In [8, 9], they showed die area for

fabrication of microchips and several parameters that are necessary for die area fabrication.\\

Heat has a direct relation to time, which has been already said in introduction of this paper. [10] Takes attention of time-variant design issues for micro-architecture.\\

The deep level of micro-architecture is illustrated in [5, 6, 7, 8, 12]. The logic gates have induced delay. Any state transition causes delay for propagation from input to output [5]. Totem pole for inverter circuit has been induced in the CMOS logic design in order to decrease propagation delay. Now-a-days, we do not have a single processor in our computer. This work has considered simultaneous multithreaded processors, chip multiprocessors, and many cores. Induced heat and their checking and minimization are discussed in [6, 7, 8].\\

In [14], the author showed power, thermal view for multicore. They described the temperature issue as spatial distribution. They showed how heat is detected for the floorplan of the system. It is important to care about the geometric characteristics of the floorplan. They summarize all those impacting on chip heating as follows:

1. Proximity of hot units. If two or more hotspots come close,this will produce thermal coupling and therefore raise the temperature locally.
2. Relative positions of hot and cold units. A floorplan interleaving hot and cold units will result in lower global power density (therefore lower temperature).
3. Available spreading silicon. Units placed in such a position, that limits its spreading perimeter, will result in higher temperature, e.g. the units placed in a corner of the die.

Finally, we want to conclude this section by saying something about fabrication of silicon chips. Now-a-days, different fabrication methods are used for chip design. In [15], they said some very important step of the fabrication process. They said, chip today are typically packaged with the die placed against a spreaded plate, often made of Aluminum, copper, or some other highly conductive metal, which is in turn placed against a heat sink of Aluminum or copper that is cooled by fan. In introduction section, we said about using different metal for designing the fabrication of chips or for designing the drain system for heat control. In strategy section, we will say in detail what things are important for our proposed mechanism.

3. TEMPERATURE SENSITIVE DESIGN

Our principal goal is the make the processor design using some separator logic gate. Most of the part of microprocessor are designed by universal gates NAND and NOR. These gates are designed by basic gates and including an inverter circuit in front of the gates. We can add heat dissipation mechanism in the inverter circuit. We think of making the gates design with a different type of metal, and the chip can sustain 3 logic states instead of 2. It will improve voltage switching easier and fast, and we can drain the tri-state easily.

How the temperature is related to these mathematical equations we will describe in later of this section. Briefly we can say, the less time the chip will remain on (bit state 1), the less will be the voltage and current flow. That way power will reduce. When power reduces it will decrease heat. Moreover, we have said already, the chip will remain on for shorter period. So, time will be small. We want to make the change in

bit level in less than nano-scale level. We want to optimize our performance of the running application, that is, we want to reduce runtime. In this way, we want to contribute in supercomputing. Our programming simulation will tell details about it and will show light inside the nano architecture.

Main design logic is tested to keep in mind about following improvement of microchips.

- o Register induced die expansion controlling.
- o Heat reduction by using different metal.
- o Performance improvement.
- o Reduction of gate delay.
- o Checking heat control inducing sampling switch
- o Modifying heated area by detecting dynamic temperature of the semiconductor floors.
- o Predicting, checking and controlling thread operation for Simultaneous Multithreaded Processors.
- o Core switching heat transfer for many core, multi core and chip multiprocessors.
- o Prediction about load/store operation to minimize heat transfer

Proposed system comprises of the buffer and the sampling switch. An existing high-performing bit-interleaved system is implemented using same technique to check the performance of the proposed system. Simulation identifies the proposed system wont degrade performance of any exixting work. So, this proposal want to prose the system shown in figure 1. The bit-interleaved system is shown in figure 2 using the proposed methodology. The proposed circuit diagram is shown in figure 3 and the corresponding floorplan design in chip level is shown in figure 4. buffer in the circuit (AD22050N) is used for retrieval of output of the prediction/actual output. The switch can be an XOR gate. The operating circuit can be anything like output of ALU, memory calculation output for load/store operation, branch prediction logic, task switching operation, control logic for program execution, etc. Buffer and the xor gate works as a sink which absorbs heat and reduces heat generation by storing for a short interval of time.
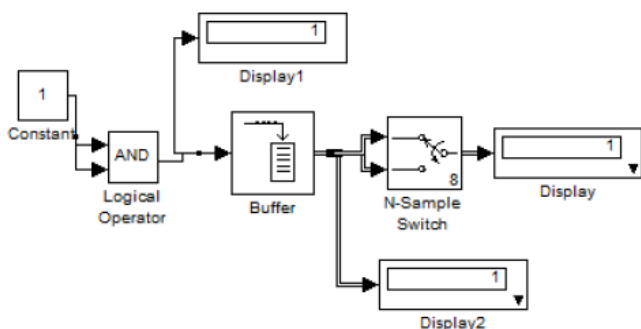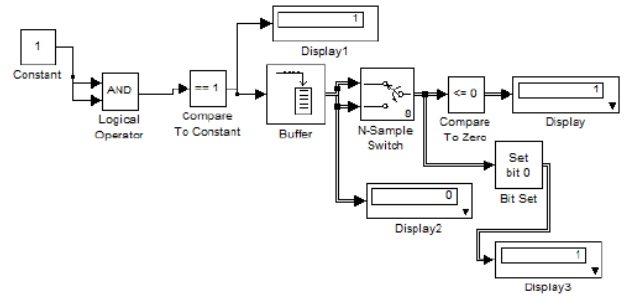

Figure 1: Proposed Simplistic Design

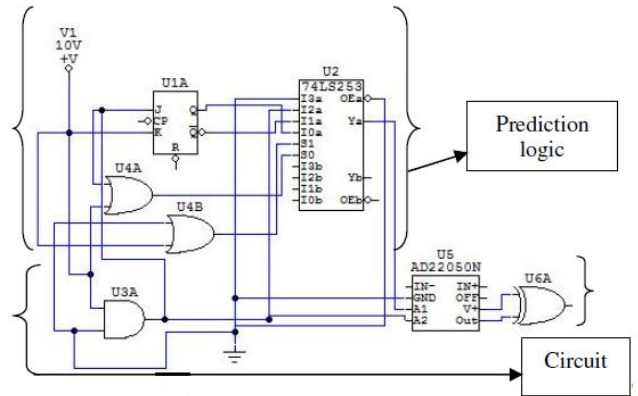**Figure 2: Bit Inter-leaved Design**

Figure 3: Proposed Circuit Design

Figure 1 includes some "Display" block which shows output at each step after every clock tick. For a certain number of clock tick proposed system woks fine as like as bit-interleaved system.This type of mechanism is not only necessary for control-logic design of microprocessors, but also is an essential subsystem for network-on-chip. We can leave some work for Operating System. We can design a patch for thread's activity to control heat in case of Simultaneous Multithreaded Multiprocessors.
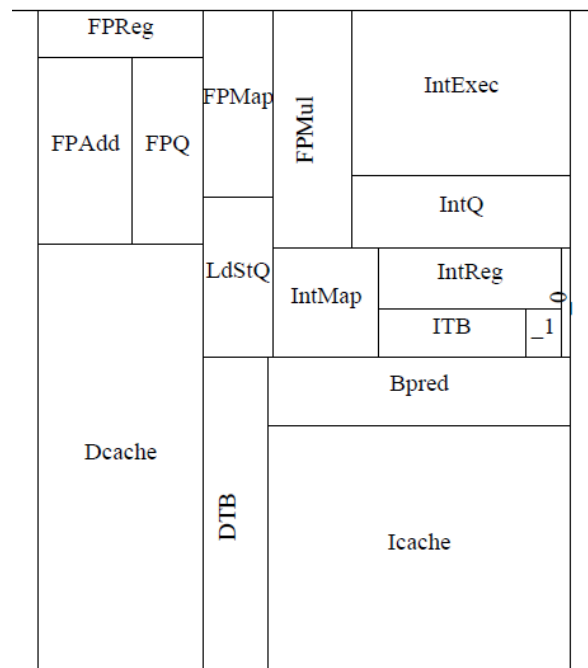

\begin{figure}[htbp!]

Figure 4: Underlying Floorplan Design in Chip

4. CONCLUSION

Proposed techniques reduce the generated heat in the microprocessor and thus improve performance. We will further extend our work for development of an integrated system. Though the work considers different architectures of microprocessor, it will be extended specific further for Simultaneous Multithreaded Multiprocessor and Chip Multiprocessors.

5. REFERENCES

[1] M. Kondo and H. Nakamura. A Small, Fast and Low-Power Register File by Bit-Partitioning/Proceedings of the 11th Int'l Symposium on High-Performance Computer Architectur e,2005,pp. 1-10.

[2] J. Hu, K. John, and S. Wang, "Thermal-Aware Subarrayed Data Cache Microarchitectures," International Journal of Intelligent Control and Systems, Vol.13, No. 4, December 2008, pp. 251-263.

[3] Y. Zhang, D. Parikh, K. Sankaranaraya nan, K. Skadron, and M. Stan, "HotLeakage: A Temperature-Aware Model of Subthreshold and Gate Leakage for Architects," University of Virginia Department of Computer Science Tech. Report CS-2003-05.

[4] I. Park, M. D. Powell, and T. N. Vijayku mar, "Reducing Register Ports for Higher Speed and Lower Energy", In Proceedings of MICRO, 2002.

[5] M.S. Hrishikesh, N. P. Jouppi, K. I. Farkas, D. Burger, S. W. Keckler, and P. Shivakumar, "The Optimal Logic Depth Per Pipeline Stage is 6 to 8 FO4 Inverter Delays", In the proceedings of the 29th International Symposium on Computer Architecture.

[6] J. Donald and M. Martonosi, "Tempera ture Aware Design Issues for SMT and CMP Architectures", Work shop on Complexity-Effective Design, 2004.

[7] J. Donald and M. Martonosi, "Techn iques for Multicore Thermal Management: Classification and New E xploration", ACM SIGARCH Computer Architecture News, 2006.

[8] Sheng-Chih Lin, N. Srivastava and K. Banerjee, "A Thermally −Aware Methodology for Design −Specific Optimization of Supply and Threshold Voltages in Nanometer Scale ICs", International Conference of Computer Design, 2005.

[9] G. H. Loh, Y. Xie, and B. Black, "Proce ssor Design in 3D Die-Stacking Technologies," IEEE Computer Society, 2007. Pp. 31-48.

[10] H. Yu, Yu Hu, C. Liu, and Lei He, "Mi nimal Skew Clock Embedding Considering Time Variant Temperature G radient," In the Proceedings of ISPLD, 2007, Austin, Texas, USA.

[11] Z. Qi, B. H. Meyer, W. Huang, R. J. Ri bando, K. Skadron, M. R. Stan, "Temperature-to-Power Mapping," In the Proceedings of ICCD, 2010.

[12] S. Borkar, "Thousand Core Chips-A T echnology Perspective," In the Proceedings of DAC, 2007, San Diego, California, USA.

[13] J. K. John, J.S. Hu , and S. G. Ziavras, "Optimizing the Thermal Behavior of Subarrayed Data Caches, " International Conference of Computer Design, 2005.pp. 625-630.

[14] M. Monchiero, R. Canal, and A. Gonzalez , "Design Space Exploration for Multicore Architectures: A Power/Performance/T hermal View," In the Proceedings of ICS, 2006, Queensland, Australia.

[15] K. Skadron,M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-Aware M icroarchitecture," International Symposium on Computer Architecture, 2003.

[16] Dan Ernst, Nam Sung Kim, Shidhartha Das, Sanjay Pant, Rajeev Rao, Toan Pham,Conrad Ziesler, David Blaauw, Todd Austin, Krisztian Flautner, and Trevor Mudge, "Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation" in MICRO 36 (2003).

[17] Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems

[18] Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[19] Sannella, M. J. 1994 Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[20] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[21] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE.

[22] Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[23] Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender