Thermal Aware Multi-Core Scheduler

Christiana Ioannou *, [1], Yiannakis Sazeides *, [1],
Pierre Michaud †, [2],
Martha Vasiliadou *, [1]

** University of Cyprus, 75 Kallipoleos Street, CY-1678 Nicosia, Cyprus*
† Campus de Beaulieu, 35042 Rennes Cedex, France

ABSTRACT

Multi-core chips allow thread and program level parallelism thus increasing performance. However, this comes with the cost of temperature problem. Multi-core chips require more power, creating non uniform power map and hotspots. Activity migration is one of the solutions to distribute power in a more uniform manner over the chip. Thermal Aware Multi Core Scheduler combines Dynamic Thermal Management (DTM) techniques, thread priorities and activity migration algorithms to achieve a uniform temperature map with the best performance.

KEYWORDS: multi-core processors, activity migration, temperature

1 Introduction

There is an increase in demand for high performance, low cost and low power processors to satisfy the consumer's needs. One of the industry's responses was to place in a single chip more than one core. With multi-core chips, thread and program level parallelism can increase performance with low cost. However, multi-cores increase power consumption which in turn creates temperature problems and non uniform power density map.

Temperature is now a critical concern not only for packaging, but it is also important for circuit, micro architecture, and OS design. Improving packaging and advancing new cooling methods can be expensive and not effective enough. Dynamic Thermal Management (DTM) mechanisms showed that they can help temperature problems with low performance degradation [DM06]. Migrating threads among cores also helps to distribute power uniformly in the chip at a minimal performance loss [HBA03].

The aim of this project is to develop a thermal aware multi-core scheduler which uses migration of threads and DTM mechanisms to reduce temperature and maintain high performance.

[1] E-mail: {cioannou,yanos,cs04mv1}@cs.ucy.ac.cy
[2] E-mail: pmichaud@irisa.fr

2 Conventional Thread Scheduling on Multi-Cores

Linux scheduler in one core chip schedules threads to run based on thread priorities. There exists a queue that contains all threads that are ready to be assigned for processing. The thread with the highest priority is first scheduled to run first. For a multi-core system each core has its own queue. The scheduler allows migration between cores only when necessary in order to minimize cache misses [BC00].

Our thermal aware scheduler has only one queue for all cores and it does schedule threads according to priority. Migration occurs when considered necessary according to temperature thresholds and if a certain period of time has passed since the last migration.

3 Thermal Aware Techniques

High power consumption results in high temperature and vice versa [DM06]. Power consists of two parts: dynamic power and static power. Dynamic power is consumed during processing activity, whereas static power is consumed without any activity. Dynamic power depends on activity, capacitance, frequency, and voltage. Activity depends on the program behavior. DTM methods aim at decreasing dynamic power consumption by modifying voltage and/or frequency [DM06]. Static Power is a function of voltage and leakage. Leakage depends on temperature, thus by decreasing temperature we decrease static power.

3.1 Dynamic Thermal Management Mechanisms

Most of the recent processors have adopted clock gating and Dynamic Voltage and Frequency Scaling as two reactive DTM mechanisms. Clock gating refers to switching off the core clock for a certain period of time and then switching it on. Switching off the clock eliminates the dynamic power leaving only static power. During this time the core or chip (depending on the number of voltage domains) is slowing down the total processing time and in return the temperature is dropping.

Dynamic Voltage and Frequency Scaling does not stop the processing of data as clock gating. Instead it dynamically adapts the processing time of the data. DVFS dynamically adapts voltage and frequency taking into consideration temperature. If the core or chip is thermally saturated, voltage and frequency will decrease to achieve lower temperatures and relatively high performance at the same time.

In our thermal aware scheduler we are simulating both mechanisms. At small time intervals we evaluate if clock gating has to be engaged and if the chip is still thermally saturated then we engage DVFS. Every 1 ms we check if we have exceed the temperature threshold and if we did the voltage is dynamically reduced. During this 1 ms, at every 1 μs the temperature of each core is evaluated and if the core is thermally saturated its clock is switched off for 2 μs.

3.2 Migration

Migration can help even more the temperature problem [HBA03][MSSCF]. By migrating threads in a chip can help distribute power around the chip and create a more uniformly distributed power map. Migration period and temperature threshold are two conditions that have to first be satisfied to allow migration. Deciding where to assign a thread is based on the migrations policies and thread characteristics. There are two categories of migration policies, the thermal aware migration policies and non thermal aware migration policies. The thermal aware policies depend on the cores' thermal state and may or may not take into consideration thread priorities. The non thermal aware policies take into consideration only thread priorities.

Every thread has a priority and thermal activity characteristic. Thread priorities in our simulation are based on the processing time of each thread. Also, every thread in our simulation has thermal characteristics depending on thermal activity. We created three main categories: hot, warm, and cold threads. The difference between the categories is the dynamic power density on the hot spot of each core. Threads under the category hot threads use 4W, warm threads use 2W, and the cold threads use 1W power density.

4 Simulator

ATMI [MS07] is used as the temperature model of our simulator. It takes as input the total power of a unit and returns the temperature. ATMI is used at the beginning of the simulation to compute the steady state of each core. During simulation time, every 1 µs ATMI takes as input the new power densities for each unit in the chip and it returns the new temperature.

Configuration

Chip Size

- 0.02 x 0.02 m

Cores

- 4 Cores
- Single and Multiple Voltage
- Each core is composed of 9 units
- Every Unit has a sensor

Thread

- 4 Threads
- Thermal Behaviour and Priorities
    - 1 Thread with Priority 10 Thermal Behaviour Cold
    - 1 Thread with Priority 5 Thermal Behaviour Cold
    - 2 Threads with Priority 1 Thermal Behaviour Hot and Warm

Thermal Management Techniques

- Clock Gating
- DVFS

Temperature Model

- ATMI

Migration

- Every timeslice

Duration of the simulation : 2 s

Timeslice: 1 ms
Timestep: 1 µs

Temperature

- Relative Temperature 45 C
- Ambient Temperature 40 C

Clock Gating off period : 2 µs

Figure 1: Example of possible experiment

Figure 1 shows an example of a possible experiment. The only option omitted is the mapping policies.

5 Ongoing Work

The scheduler is still under development but so far we have observed that indeed migration creates a uniform power map. Depending on the migration policy as well as the configuration, temperature is distributed more uniformly than not having any migration.

Temperature increase has let to new ways of scheduling threads on a multi-core system. Our experiments showed that some migration policies can give good performance and good thermal behavior. One promising approach is using single voltage domain for all cores and mapping threads only by their thermal activity characteristics.

References

[BC00] Daniel P. Bovet and Marco Cesati. Understanding the Linux Kernel. Chapter 10, Process Scheduling. October 2000.

[DM06] James Donald and Margaret Martonosi. Techniques for Multicore Thermal Management: Classification and New Exploration. In *Proceedings of* ISCA'06, 2006

[HBA03] Seongmoo Heo, Kenneth Barr, and Krste Asanovic. Reducing Power Density through Activity Migration. In *Proceedings of* ISPLED'03, 2003.

[PGV04] Michael D. Power, Mohamed Gomaa, and T. N. Vijaykumar. Heat-and-Run: Leveraging SMT and CMP to Manage Power Density Through the Operating System. In *Proceedings of* ASPLOS'04, 2004

[MS07]Pierre Michaud and Yiannakis Sazeides.ATMI: Analytical Model of Temperature in Microprocessors. MoBS'07, 2007

[MSSCF] Pierre Michaud, Yiannakis Sazeides, Andre Seznec, Theofanis Constantinou, Damien Fetis. A Study of Thread Migration in Temperature-Constrained Multi-Cores. To appear in ACM Transactions of Architecture and Code Optimization