Journal of Computational Science 19 (2017) 112–120

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Thermal-aware power-efficient deadline based task allocation in multi-core processor

Sumarga Kumar Sah Tyagi [a,*], Deepak Kumar Jain [b], Steven Lawrence Fernandes [d], Pranab K. Muhuri [c]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing, China
[c] Department of Computer Science, South Asian University, New Delhi, India
[d] Dept. of Electronics and Communication Engineering, Sahayadri College of Engineering & Management, Mangalore, Karnataka, India

ARTICLE INFO

Article history:
Received 22 June 2016
Received in revised form 25 October 2016
Accepted 21 November 2016
Available online 15 December 2016

Keywords:
Multicore systems
Thermal management
Task allocation
Dynamic voltage and frequency scaling
Real-time constraints

ABSTRACT

This paper address the power-efficient task allocation problem in a heterogeneous multi-core processor for thermally challenged tasks with real-time constraints. A thermally challenged task is one which, if executed on a core at the maximum possible speed, resulting the core temperature exceeds a given safe temperature limit. Our problem formulation is based on the well-praised thermal models HotSpot-4, leakage power and delay models with DVFS (dynamic voltage and frequency scaling). The solution to this problem is not so straight forward due to the fact that there is a nonlinear-circular dependency between the overall power dissipation and the temperature. It becomes further complicated with the consideration of a crucial parameter, task's deadline. We have solved this problem with a novel idea of *deadline based task allocation scheme* (DBTAS) that takes care of the permissible thermal dissipation, power consumption and speed of the processor. We have included suitable numerical test cases to demonstrate the efficacy of the proposed scheme for task allocation.

1. Introduction

During the last decade, numerous thermal management schemes have been proposed by the researches to address prevalent thermal issues in the single as well as multi-core processors. In the multi-core processor, thermal hot spots and large temperature gradients produced due to increased no. of threads and intra and inter-core process variations. Which drastically affects performance, leakage power and lifetime of the system. Leakage power dissipation in a processor scales exponentially with the voltage overdrive [1–3]. According to [4] for every 9 °C temperature reduction, there is 50% reduction in the leakage power. This reduction is particularly paramount important for such processor that has real-time tasks to be finished by it is deadlines. It is very challenging to design a comprehensive technique that can address such issues related to heat-dissipation, power consumption and task deadline. Such technique should be able to empower the processor to autonomously modify the task execution and power dissipation characteristics along with safe temperature regulation.

Most of the existing techniques, such as Thread Migration [5,6], Core Hopping [7], Predictive Thermal Aware Task Migration Techniques [8–12] and policies [13,14,8,15–20] presented in the literature has centralized approach. They require a central co-coordinator that examines the temperature and load distribution of individual core on the entire chip and take global decisions of resource allocation. The major drawbacks of such centralized approaches are poor scalability and huge communication overhead among subunits of large no of cores [21]. Similarly, the papers [22–24] discuss about power-performance characteristics in multi-thread environment for mobile gaming platform under thermal constraints where meeting deadline is not crucial. Albeit in [25–29], the authors proposed mathematical models for resource distribution (e.g. DsREM) and dynamic thermal management (DTM) techniques considering thermal dynamic power (TDP) and power budgets in dark silicon multi core chips. However, they do not consider real-time constraint especially task-deadline. Therefore, they are not suitable for real-time applications where the system response should be fast enough and meeting task deadline is must.

* Corresponding author.
E-mail address: sumarga@ict.ac.cn (S.K. Sah Tyagi).

http://dx.doi.org/10.1016/j.jocs.2016.11.012
1877-7503/

1.1. Paper contribution

In this paper, we present a thorough study of a closed form power-efficient task allocation problem in a heterogeneous multi-core processor for thermally challenged tasks with real-time constraints. A thermally challenged task is one which, if executed on a core at the maximum possible speed, resulting the core temperature exceeds a given safe temperature limit. Our problem formulation is based on the well praised thermal models HotSpot-4, leakage power and delay models with dynamic voltage and frequency scaling (DVFS) method. It accounts for the differences in power consumption of tasks and cores as well as the differences in the thermal characteristics of the functional blocks in the die. The model also considers various components of the package and chip temperature dependent leakage power. The solution to this problem is not so straightforward because the temperature and leakage power are nonlinear-circularly dependent to each other. It becomes further complicated with the consideration of real-time constraints of the task i.e. deadlines. To address this problem we propose our novel idea of *deadline based task allocation scheme (DBTAS)* using key parameter, *lateness*, of the task. Because Lateness assures meeting task's deadline in early [30]. Simulating suitable test cases has assessed the proposed scheme.

The rest of the paper is organized as follows: Section 2 provides descriptions of *Thermal*, *Delay* and *Leakage Power* models, throwing light on the analytical basis of the Thermal model's parametric choices and proposes task-to-core allocation scheme. In Section 3 problem has been formulated as an optimization problem which further analyzed by our proposed scheme. Section 4 invalidate proposed scheme through experiments.

2. Thermal and power models of processor

We are using well-praised HotSpot-4 based thermal model as a tool to analyze thermal characteristics of a multi-core processor in this paper. Basically HotSpot-4 described as analogy of electrical network where thermal interaction between several blocks are regarded as RC network while current source represented as power input whereas heat spreading and storing phenomenon is represented as resistance and capacitance respectively. Fig. 1 describes a typical 4-core processor. Each core is partitioned into m blocks on both the layers of die and the thermal interface material (TIM). The layers of heat sink and spreader below the TIM have 9 and 5 blocks, respectively. Altogether the total number of thermal blocks is $N = 2nm + 14$. For a four-core processor with 20 blocks per core, the total number of blocks is 174.

Therefore with the help of state-space model the thermal model can be expressed as per [31]:

$$\frac{dT(t)}{dt} = \alpha T(t) + \beta Pw(f, V, T, t) \tag{1}$$

where Pw and T are power and temperature vectors respectively. Similarly n is the number of cores. α and β are constants. The total power Pw is nothing but sum of dynamic (Pw_d) and leakage (Pw_l) powers. The Pw_d is expressed as relation (2) whereas Pw_l is expressed using empirical model from [1] as relation (3).

$$Pw_{d,c,b}(t) = Pw^{\max}_{d,c,b}(t) f_c(t) V_c^2(t) \tag{2}$$

$$Pw_{l,c,b}(t) = g_1 V_c(t) T^2_{c,b}(t) e^{\frac{\omega V_c(t)+\theta}{T_{c,b}(t)}} + g_2 e^{\gamma V_c(t)+\delta} \tag{3}$$

where $Pw^{\max}_{d,c,b}$ is the dynamic power dissipated by block b of core c when the core is at the maximum *speed* and *voltage* and g_1, g_2, ω, θ, γ and δ are parameters that depend on circuit topology, size, technology, and design.

The non-linear *leakage power dependent on temperature and voltage (LDTV)* and the cyclic dependency of the leakage power over the temperature in relation (3) has been simplified using piece-wise linear, PWL, approximation. Thus the relation (3) is rewritten as:

$$Pw_{l,c,b}(t) = Pw_{l0,c,b} + g_{T,c,b} T_{c,b}(t) + P_{v,c} V_c(t) \quad \forall c, b, t \tag{4}$$

where, $g_{T,c,b}$ and $P_{v,c}$ represent the temperature and the voltage coefficients.

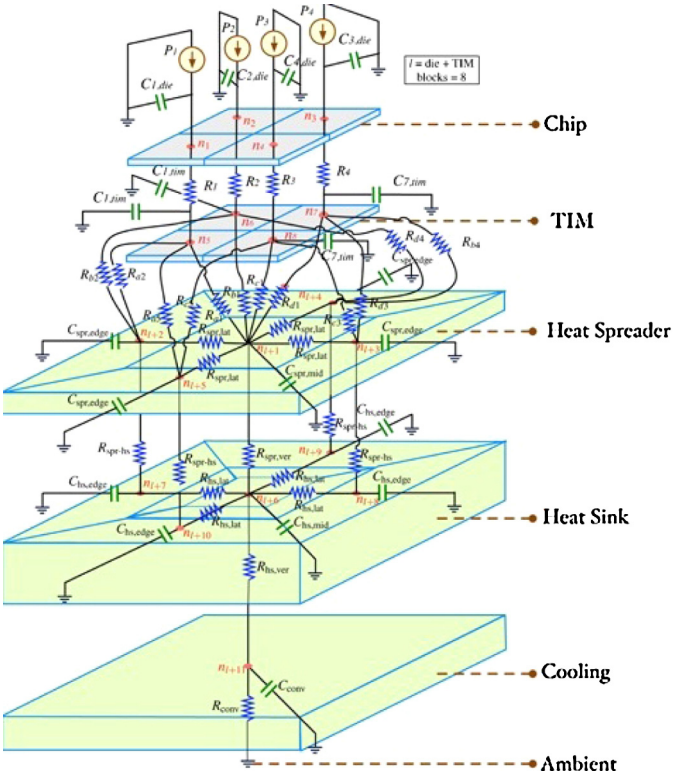

Fig. 1. HotSpot-4 thermal model for a four-core processor.

2.1. Thermal-delay model

It is known that delay in circuits is directly proportional to its temperatures. This can be nullified with adjustment of supply voltage. The relationship among maximum *Frequency, Voltage* and *Temperature* of a circuit is represented empirically for 65 nm technology by [1] as below: (core index c is omitted for simplicity)

$$f^{\max} = \beta \frac{(V - V_{th})^{1.2}}{V(T_{\max})^{1.19}} \tag{5}$$

where f^{max} is the maximum operational frequency of the core c. T, V and V_{th} are temperature, supply and threshold voltages of core respectively. β is proportionality constant and 1.19 and 1.2 are estimated values for 65-nm technology.

3. Problem formulation and DBTAS

The system power basically consists of two components, dynamic and leakage power. Due to the homogeneity, different task mappings have little impact on the dynamic power. However, they change the temperature distribution across the system and can potentially affect the leakage power. The leakage power is determined by the temperature and core-voltage, while dynamic power is dictated by the speed and core-voltage and task allocation to it. Hence they require some optimization heuristics to optimize power and performance. The author in [32] proposed makespan minimization technique, however, it could not guarantee meeting the task's deadlines. Our analysis shows that the overall power can