INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	Shounak Chakraborty
Roll Number	:	11610111
Programme of Study	:	Ph.D.

Thesis Title: Energy and Thermal Management of CMPs by Dynamic Cache Reconfiguration

Name of Thesis Supervisor(s)	:	Prof. Hemangee Kalpesh Kapoor
Thesis Submitted to the Department	:	Computer Science & Engineering
Date of completion of Thesis Viva-Voce Exam	:	22nd February, 2018
Key words for description of Thesis Work	:	Chip Multi-Processors, Thermal Efficiency, Energy Efficiency, Cache Memory, LLC, Dynamic Cache Reconfiguration,

SHORT ABSTRACT

Ever increasing demand of processing speed and parallelism, along with the modern shrunk transistors, motivates the architects to increase the number of cores on a single chip leading to Chip Multi-Processors (CMPs). To commensurate the data demand of these high number of cores, large on-chip Last Level Caches (LLCs) are integrated. After studying a plethora of prior works, it has been concluded that, LLCs play a vital role in maintaining system performance by accumulating more data on-chip. But large sized LLCs are accounted for their significant leakage energy consumption, which has a circular dependency on the effective temperature of the chip-circuitry. In addition to curtailing the circuit's reliability, this increased chip temperature (caused due to heavy power consumption) has enough potential to damage the on-chip circuitry permanently, and to exacerbate the battery life in embedded systems.

Towards leakage minimised on-chip cache design, with due consideration to the Locality of Reference, we propose a set of performance constrained Dynamic Cache Resizing techniques to reduce leakage in LLCs. The resizing is done by turning off/on some cache banks which can be implemented by power gating at circuit level. The cache resizing decision is triggered based upon the dynamic (cache) usage and change in system performance. We get 65% savings in leakage energyconsumption. In the next contribution, apart from bank level granularity, we proposed cache resizing at way-level granularity, where performance degradation is handled by incorporating DAM (Dynamic Associativity Management) techniques. In this policy,

we obtain 70% improvement in the leakage energy. Both of these techniques outperform a prior state preserving leakage reduction technique called, Drowsy cache; both in terms of leakage savings and EDP gains. All of these techniques are evaluated for tiled CMP having a multi-banked shared L2 cache as its on-chip LLC.

From the thermal efficiency perspective, it has been noticed that, larger caches fabricated in smaller technology nodes are the potential candidates for generating hotspots similar to the CPU cores. The increased power density at heavily used cache zones can heated them up faster, whereas the lightly used cache portions unnecessarily consume high leakage, which has a circular dependency with chip temperature. To mitigate these issues, our proposal selectively turns off LLC banks to reduce temperature by reducing their power consumption. Moreover, these turned off cache portions are utilised as on-chip thermal buffers that reduce effective chip temperature. The proposal is evaluated for tiled CMP architecture as well as CMPs having large multi-banked centralised LLCs. These approaches reduce average chip temperature by around 4 ◦ C and 6 ◦ C, respectively, for both architectures. The results are also compared with DVFS methods. The thesis has thus demonstrated that, large LLCs on-chip need resizing for optimal power management and they can also assist in controlling chip temperature.