ERCIM "ALAIN BENSOUSSAN" FELLOWSHIP PROGRAMME

Scientific Report

First name / Family name	Shounak Chakraborty
Nationality	Indian
Name of the *Host Organisation*	Norwegian University of Science and Technology (NTNU)
First Name / family name of the *Scientific Coordinator*	Magnus Själander
Period of the fellowship	01/01/2019 to 31/12/2020

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

The following scientific activities have been taken place during the two years of fellowship:

A. Developed a novel technique called *WaFFLe* that improves thermal and power efficiencies of heterogeneous chip multi-processor with one paper under review.

B. Explored thermal and power characteristics of FinFET based 3D chip multi-processor and developed novel technique that improves the thermal and power efficiency of the chip. This work is in collaboration with the Cyprus University of Technology, Cyprus, and the respective paper is under preparation.

C. Established a research collaboration with University of Essex, UK and published primary research ideas and outcomes (*RePAiR*) in the ESWEEK (CODES+ISSS) 2020 conference. The objective of this work is to improve thermal and power efficiency in approximated real-time computing. Currently, one paper is under review.

D. Collaborating with my previous university on improving runtime thermal and power efficiency of heterogeneous multi-core real-time systems. One paper is still under review and one (SEAMERS) has been published in Journal of Systems Architecture [JSA], Elsevier**.**

E. Participated in a two-day workshop (in grant writing) arranged by the Faculty of Information Technology and Electrical Engineering, NTNU, during the summer of 2019. This workshop helped me in writing and submitting a research proposal for Marie Skłodowska-Curie Individual Fellowship (MSCA-IF). The proposal (TECTONIC) was accepted in February 2020 for funding by MSCA-IF-2019.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Journals:

1. **S. Moulik, Z. Das, Rajesh D., and *S. Chakraborty*, "SEAMERS: A Semi-partitioned Energy Aware scheduler for heterogeneous MulticorE Real-time Systems." Journal of Systems Architecture [JSA], Elsevier (in press).**
   Abstract: Over the years, the nature of processing platforms is witnessing a significant shift in most of the battery supported real-time systems, which now support a combination of specialized multicores to meet the demands of modern applications. Devising energy-efficient schedulers has become a critical issue for such kinds of devices. Hence, this research presents a low-overhead heuristic strategy named SEAMERS, for DVFS based energy-aware scheduling for a set of real-time periodic tasks on a heterogeneous multicore platform. The presented strategy operates in four phases, namely Deadline Partitioning, Core Clustering, Task Allocation and Energy-Aware Scheduling. Our experimental analysis shows that the presented strategy improves upon the state-of-the-art in terms of energy savings (16% to 47% on average) and enables significant improvement in resource utilization.
2. ***S. Chakraborty*, and H. K. Kapoor, "Exploring the Role of Large Centralised Caches in Thermal Efficient Chip Design." ACM Transactions on Design Automation of Electronic Systems [TODAES], Vol. 24, Issue 5, Article 52, June 2019.**
   Abstract: In the era of short channel length, Dynamic Thermal Management (DTM) has become a challenging task for the architects and designers while engineering modern Chip Multi-Processors (CMPs). Ever increasing demand of processing power along with the developed integration technology produces CMPs with high power density, which in turn increases effective chip temperature. This increased temperature leads to increase in the reliability issues for the chip-circuitry with significant increment in leakage power consumption. Recent DTM techniques apply DVFS or Task Migration to reduce temperature at the cores, the hottest on-chip components, but often ignore the on-chip hot caches. To commensurate the high data demand of these cores, most of the modern CMPs are equipped with large multi-level on-chip caches, out of which on-chip Last Level Caches (LLCs)

occupy the largest on-chip area. These LLCs are accounted for their significantly high leakage power consumption which can also potentially generate on-chip hotspots at the LLCs similar to the cores. As power consumption constructs the backbone of heat dissipation, hence, this work dynamically shrinks cache size while maintaining performance constraint to reduce LLC leakage, primarily. These turned off cache portions further work as on-chip thermal buffers for reducing average and peak temperature of the CMP without affecting the computation. Simulation results claim that, at a minimal penalty on the performance, proposed cache based thermal management having 8MB centralised multi-banked shared LLC gives around 5°C reduction in peak and average chip temperature, which are comparable with a Greedy DVFS policy.

Conferences/Seminars:

1. ***S. Chakraborty*, S. Saha, M. Själander, and K. D. McDonald-Maier, "RePAiR: A Strategy for Reducing Peak Temperature while Maximising Accuracy of Approximate Real-Time Computing: Work-in-Progress"-In Proceedings of IEEE/ACM CODES + ISSS, 2020.**
   Abstract: Improving accuracy in approximate real-time computing without violating thermal-energy constraints of the underlying hardware is a challenging problem. The execution of approximate real-time tasks can individually be bifurcated into two components: (I) execution of the mandatory part of the task to obtain a result of acceptable quality, followed by (ii) partial/complete execution of the optional part, which refines the initially obtained result, to increase the accuracy without violating the temporal-deadline. This paper introduces RePAiR, a novel task-allocation strategy for approximate real-time applications, combined with fine-grained DVFS and on-line task migration of the cores and power-gating of the last level cache, to reduce chip-temperature while respecting both deadline and thermal constraints. Furthermore, gained thermal benefits can be traded against system-level accuracy by extending the execution-time of the optional part.
2. ***S. Chakraborty* and M. Själander, "TECTONIC: Towards Employing Compilers for Thermal Management and Optimal Data Placement in Hybrid Cache" at ARM Research Summit, Texas, USA, 2019.**
   Abstract: The goal of the TECTONIC project is to alleviate the challenging problem of hot-spots in 3D stacked chip-multiprocessors by employing a software-hardware based combined approach. With the stagnation in process technology scaling, new emerging memory technologies are investigated. Promise of better scalability with reduced power consumption makes Non-volatile memories (NVMs) as the potential

candidates to replace conventional SRAM based caches. However, many of the proposed NVMs are sensitive to heat, which raises the issue of reliability. Considering heat dissipation as an exclusive issue of hardware design will not be an appropriate approach towards finding potential solutions, as runtime resource utilisation profile of the application has direct impact on on-chip thermal imbalance. Hence, TECTONIC will manage on-chip temperature and eliminate hot-spots by leveraging application specific knowledge extracted during compilation in combination with new hardware mechanisms for distributing computational work and memory accesses for even heat distribution while maintaining high performance.

Under Preparation/Review:

1. *S. Chakraborty* and M. Själander,
   "WaFFLe: Gated Cache-Ways with Per-core Fine-grained DVFS for Reduced On-chip Temperature and Leakage Consumption".
2. *S. Chakraborty*, S. Saha, M. Själander, and K. D. McDonald-Maier,
   "PrinCess: Power-Aware Scheduling of Approximate Computing Based Real-time Tasks".
3. *S. Chakraborty,* M. Själander, and V. Soteriou,
   "Thermally Safe Temperature Effect Inversion-Aware FinFET based Hybrid 3D Multi-core (STIFF)".
4. Z. Das, Y. Sharma, *S. Chakraborty,* and S. Moulik,
   "ETA-HP: An Energy and Temperature-Aware Real-time Scheduler for Heterogeneous Platforms".

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

1. International Symposium on Microarchitecture (MICRO), 2020.
2. Embedded Systems Week (ESWEEK), 2020.
3. IEEE International Conference on Computer Design 2019.
4. ARM Research Summit 2019.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

This programme could not be completed primarily due to the outbreak of COVID-19. There was a plan for a physical visit at the INRIA Lab in France during December 2020. Later we had to cancel the plan because of the second wave of the COVID-19 in Europe. We attempted to arrange it virtually but with the additional burden of COVID-19 lock downs and the obstacle of initiating new collaborations based on only virtual meetings this was also not a viable option. Hence, I have unfortunately not been able to do the REP.