ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME

ERCIM
European Research Consortium
for Informatics and Mathematics

Scientific Report

First name / Family name	Kameswar Rao Vaddina
Nationality	Indian
Name of the *Host Organisation*	Norwegian University of Science and Technology (NTNU)
First Name / family name of the *Scientific Coordinator*	Lasse Natvig
Period of the fellowship	01/07/2014 to 30/06/2015

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Processor power density and heat dissipation are becoming major limiting factors for their performance evolution. Considering that leakage power depends exponentially on temperature and the future trend towards massively parallel systems which are striving for energy efficiency, on-chip temperature is fast becoming a major hindrance which needs to be managed by balancing chip thermal profile. A network of thermal sensors are being used to monitor temperature on the chip for this purpose. But as they age they become inaccurate and require constant recalibration. This work tries to address that issue.

As technology scales down and power density increases, the temperature sensor characteristics will drift, leading to temperature errors which increase over time. Transistor aging is one of the leading contributors to temperature sensing inaccuracies. The prominent aging failure mechanisms like Negative Bias Temperature Instability (NBTI), Hot Carrier Injection (HCI) and electromigration have emerged as the main sources of system unreliability which manifest as an increase in the propagation delay over time. On-chip thermal sensors are not immune to this phenomenon and get affected by these aging mechanisms. Thermal sensor aging exacerbated by increased temperatures leads to temperature sensing inaccuracies requiring repeated sensor calibration. During

this fellowship, we proposed a novel approach of using performance metrics to predict the transient temperature profile of an application as seen by the aging thermal sensor. Firstly, as proposed in our FP research training programme we make offline profiling of applications and then cluster them into groups using k-means clustering mechanism. Then we use a neural network model to predict the thermal profile of a new application given its performance metrics. The forecasting ability of our model is accessed using MSE and RMSE.

We present an implementation flow for our prediction model. Based on the implementation flow, we first gather offline profiling information for ROI bechmark run for all the benchmarks of the PARSEC suite. In order to cluster the applications we use steady-state temperature (Tss) as a classification metric in k-means clustering method. At run-time, when a new application is introduced to the system, we classify it to one of the groups and select an appropriately trained neural network model which takes in performance metrics to predict the transient temperature profile of the application. As a case study we trained our neural network model with data samples from Facesim benchmark run and predicted transient temperature values for Fluidanimate and Ferret benchmarks. The MSE and RMSE of our neural network model for the prediction of Fluidanimate benchmark run is $2.76(°C)^2$ and 1.66°C respectively. For the Ferret benchmark run they are $2.77(°C)^2$ and 1.66°C respectively.

This approach is highly scalable and can be used to predict future temperatures which can then be used for run-time dynamic thermal management of multi-core systems as discussed below.

Future work: As technology scales down and power density increases, dynamic thermal management of multi-core systems is becoming more challenging. It is shown that a prediction based proactive thermal management techniques outperforms reactive based ones. Our future research work will focus in that direction. Firstly, we make further investigations into improving the accuracy of temperature prediction. Later, the work in this paper will be extended to predict future temperature values which will then be used to develop an efficient thermal-aware scheduling schemes for prediction based proactive thermal management. We expect that this work would improve system reliability and increase its lifetime.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Submitted a 5-page paper titled “Transient Temperature Prediction for Multi-Core Systems Using Artificial Neural Network” to ReCoSoC 2015 which was not accepted. This work has been extended and a paper titled “Transient Temperature Prediction for Aging Thermal Sensors Using Artificial Neural Network” has been submitted to

Kameswar Rao Vaddina, Juan M. Cebrian and Lasse Natvig, “Transient Temperature Prediction for Aging Thermal Sensors Using Artificial Neural Network”, 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP 2016), Pending.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

During my fellowship at NTNU, I have attended several presentations of members of the CARD group during informal lunch sessions. I also gave an informal talk on "Performance metrics to evaluate multiprogram workloads – current state of the art" during one such session.

The computer architecture and design research group (CARD) at NTNU (19th May 2015, at 13:15h – 16:00h) together with EECS has hosted a mini-seminar covering topics within Task Based Programming, TBB (Thread Building Blocks), multithreading, scheduling and challenges for Real Time systems. It consisted of four independent presentations from faculty members, students and industry. Prof. Theo Ungerer gave an introductory lecture on "Multi-core Research Topics Relevant to Automotive and Avionics Safety-critical Control Systems". Later, I also gave a short presentation to Prof. Ungerer on my PhD thesis and the current ongoing work at NTNU. He suggested some ideas to fine tuning my work which I later implemented. More information on the mini-seminar can be found in the following link.

https://research.idi.ntnu.no/multicore/seminar-2015

Also attended EECS guest lecture on Mill Computing and the Mill Architecture at NTNU (21st October 2014, at 15:15h – 17:00h). More information on the EECS seminar can be found in the following link.

https://research.idi.ntnu.no/multicore/mill-lecture-2014

Prof. Mike O'Boyle from the university of Edinburgh and part of HiPEAC steering committee has visited CARD group and presented his work on "Machine Learning based Compiler Optimization for Parallel Systems". I attended his talk and also presented my work to him on "Thermal management framework for heterogeneous architectures".

Attended a guest lecture by Kenneth Østby, ARM-Trondheim. The ARM Mali GPU architecture and its memory hierarchy (caches etc.)

IV – RESEARCH EXCHANGE PROGRAMME (REP)

The research exchange programme (REP) was done at IRISA/INRIA, Rennes, France. The host is Pierre Michaud (Chargé de Recherche) from the ALF group. The research visit was for two weeks (from April 27 to May 7, 2015). During the research visit, I presented my research work at INRIA to 30 people (researchers and students) from two different campuses. I also discussed my ongoing work with Pierre and discussed various methods I could use to extend it. Following that discussion I have extended my work which is ready for submission. I have also networked with other members of the ALF team and proposed to write joint funding applications with INRIA as the host organization.