EU HiPEAC thematic session: Collective characterization, optimization and design of computer systems

Grigori Fursin

Thematic Session Proposal
Collective characterization, optimization and design of computer systems

Grigori Fursin

**** Motivation ****

Continuing innovation in science and technology is vital for our society and requires ever increasing computational resources. However, delivering such resources is becoming intolerably complex, costly and error prone due to the following fundamental reasons:

* ever growing number of possible design and optimization choices for computer systems and their components including many core heterogeneous architectures with varying topology, frequency and memory hierarchy, multiple programming models, compiler optimization, run-time resource contentions among many others;

* multiple strict end-user requirements placed at the same time on system performance, power consumption, thermal budget, resiliency, quality of service, size, response, bandwidth, portability, design time among other important characteristics;

* complex interaction between all components of a computer system;

* growing number of incompatible tools and techniques with ad-hoc, intuition based heuristics.

As a result, understanding and modeling of the overall relationship between end-user algorithms, applications, compiler optimizations, hardware designs, data sets and run-time behavior, essential to provide better solutions and computational resources, became simply infeasible as confirmed by many recent long-term international research visions about future computer systems.

**** State of the art ****

During the last 2 decades, researchers attempted to tackle the above complexity and improve performance and power consumption of computer systems through empirical automatic exploration of large design and optimization spaces, probabilistic focused search, automatic and speculative parallelization, polyhedral modeling, machine learning techniques and run-time adaptation. Although these techniques managed to considerably outperform the state-of-the-art approaches of that time, they did not become a mainstream in production environments. Besides the usual inertia for adopting novel approaches, the main problem with such auto-tuning techniques is an excessively large number of training runs during exploration needed to find good solutions or model behavior of various components of computer systems.

On the other hand, the research and development methodology for computer systems has hardly changed in the past decades. Users and developers have to resort to non-scientific, non-systematic, non-rigorous, intuitive and error-prone methods combined with multiple ad-hoc tools and benchmarks, complex interfaces and data formats to select the most appropriate solution that satisfies all their needs. Such an outdated technology results in an enormous waste of expensive computing resources and energy, and considerably increases development costs and time-to-market for the new systems. Peak performance of the new systems is often achieved only for a few previously optimized benchmarks, forcing scientists and engineers to resort to a tedious and often manual process of understanding, modeling and improving the behavior of their applications on numerous, rapidly evolving devices that in turn slows down research and innovation. At the end, researchers and engineers find themselves in a state of chaotic and desperate transitions between various forgotten or recent techniques and tools, trying to fix or combine them but finding no real breakthrough solutions.

**** Research direction ****

We propose to fundamentally rethink, systematize and automate the optimization and design of computer systems using collective methodology based on interdisciplinary techniques originating from physics, biology and artificial intelligence combined with collective participation of multiple users.

We propose to build an extensible public repository and infrastructure to enable continuous sharing (collection) of various data (applications, data sets, tools) and
statistics about behavior of computer systems from multiple users. Researchers and engineers will be providing collaborative data mining (machine learning) modules to classify this behavior, correlate various characteristics and optimizations,
and build predictive models. Such methodology enables gradual and continuous learning and
improvement of a complex behavior of existing computer systems in realistic environments in terms of performance, power consumption and other important metrics, while boosting innovation and dramatically reducing time to market for new systems.

With continuously increasing and systematized collective knowledge about computer
systems, we should be able to provide a quick and scientifically motivated advice to end-users about optimizing their programs or choosing architectures to maximize performance and minimize power or balance other important characteristics depending on user’s needs. Furthermore, we should be able to suggest the researchers and engineers where to focus their effort and creativity when designing or optimizing computer systems, thus boosting innovation and considerably reducing development and optimization costs.

Eventually, collective methodology should fundamentally change the research, development and educational process for the future computer systems, making it systematic, scientific, statistically rigorous, reproducible and truly collaborative, while opening many ground-breaking research opportunities.

This research direction is strongly backed up by recent theoretical and practical findings from HiPEAC members and EU FP6 projects such as MILEPOST. We released the first version of collaborative repository and infrastructure at cTuning.org in 2009 and it has been successfully used to collaboratively train predictive models of the machine learning enabled compiler (MILEPOST GCC).

This technology considerably improved performance and power consumption of the existing computer
systems from IBM, Intel, AMD, and Synopsys (ARC) over the state-of-the-art technology
and helped to reduce time to market for new software designs by several orders of
magnitude.

cTuning now has more than 150 officially registered users and has been extended in several international projects together with Google, IBM, CAPS, NCAR, ICT (China)
and others. Our workshop related to this topic (EXADAPT) have been accepted at FCRC/PLDI 2011 and ASPLOS 2012. Therefore, we believe that collective methodology
is ready to become the HiPEAC3 Theme while consolidating various research and engineering
efforts.

**** References ****

* Grigori Fursin, Yuriy Kashnikov, Abdul Wahid Memon, Zbigniew Chamski, Olivier Temam, Mircea Namolaru, Elad Yom-Tov, Bilha Mendelson, Ayal Zaks, Eric Courtois, Francois Bodin, Phil Barnard, Elton Ashton, Edwin Bonilla, John Thomson, Chris Williams, Michael O'Boyle. MILEPOST GCC: machine learning enabled self-tuning compiler. International Journal of Parallel Programming (IJPP), June 2011, Volume 39, Issue 3, pages 296-327

* Grigori Fursin and Olivier Temam. Collective Optimization: A Practical Collaborative Approach. ACM Transactions on Architecture and Code Optimization (TACO), December 2010, Volume 7, Number 4, pages 20-49

* Grigori Fursin. Collective Tuning Initiative: automating and accelerating development and optimization of computing systems. Proceedings of the GCC Summit'09, Montreal, Canada, June 2009

* ACM SIGPLAN International Workshop on Adaptive Self-Tuning Computing Systems for the Exaflop Era (EXADAPT 2011) at PLDI 2011/FCRC 2011, http://exadapt.org

* Public repository and infrastructure for collaborative program and architecture characterization and optimization, http://cTuning.org