Ph.D. in Information Technology
Thesis Defense

April 23rd, 2025
at 11:00
Room BIO1 – building 21

Lev DENISOV– XXXVII Cycle

Construction of Precision Tuning Tools

Supervisor: Prof. Giovanni Agosta

Abstract:

Modern computing applications increasingly face performance and energy constraints, especially in fields such as machine learning, scientific computing, and embedded systems. Many of these applications do not require full numerical precision for every computation—introducing opportunities to optimize efficiency by reducing precision while maintaining acceptable accuracy. Precision tuning is a technique that adjusts the numerical representation of variables and operations to improve execution speed, reduce power consumption, and lower memory requirements. However, existing precision tuning tools often require manual expertise, making them impractical for large-scale adoption.
This work explores two key challenges in tool design for precision tuning: (1) the trade-offs between static and profile-guided analysis and (2) the integration of precision tuning in a hardware-software co-design framework.
Static analysis estimates precision requirements based on mathematical reasoning alone, often leading to overly conservative precision choices. In contrast, profile-guided tuning gathers real execution data to dynamically determine the precision needs of different computations, achieving a better balance between accuracy and efficiency. Our evaluation on the PolyBench/C benchmark suite shows that profile-guided analysis improves numerical accuracy by an order of magnitude in over 80% of benchmarks, while maintaining comparable execution speed. In specific cases, such as heat-3d, profile guided tuning achieves a 10x speedup, whereas static analysis leads to a slowdown due to conservative estimates. Similarly, in deriche, profile-guided tuning delivers a 3x speedup by overcoming inefficiencies in the static approach.
Most precision tuning tools optimize software without considering the underlying hardware. We introduce a co-design approach, where precision tuning is applied simultaneously at the software and hardware levels to maximize efficiency. By tuning software computations to match hardware capabilities, we achieve better performance and energy savings. When applied to an FPGA-based floating-point unit, our design approach reduces energy consumption of the program by up to 55% compared to the 32-bit

floating-point baseline, while maintaining the specified error threshold, and reduces the design time by 2700x compared to the traditional gate-level simulation-based design approach.
Additionally, we apply precision tuning to real-world applications, including Field- Oriented Control (FOC) for motor control and bicubic image scaling. In FOC, profile guided tuning achieves a 594% speedup while maintaining a mean relative error below 10^−7, compared to 293% speedup for static tuning. In bicubic scaling, static tuning results in a 15% slowdown, while profile-guided tuning delivers a 795% speedup with no accuracy loss, demonstrating its advantage in computationally intensive scenarios.
By combining automation, adaptability, and co-design principles, this research makes precision tuning more practical and scalable, paving the way for more energy-efficient and high-performance

PhD Committee

Prof. Alessandro Barenghi, Politecnico di Milano

Prof. Biagio Cosenza, Università degli Studi di Salerno

Prof. Jari Nurmi, Università di Tampére