SPEEDUP AND POWER SCALING MODELS FOR HETEROGENEOUS MANY-CORE SYSTEMS

Parallelization has been used to maintain a reasonable balance between energy consumption and performance in computing systems, especially modern multi- and many-core platforms. This University Booth presents the expansion of speedup and power scaling models to heterogeneous many-core systems.

In this study, traditional speedup and power scaling models are expanded to cover a ***normal form*** of core heterogeneity, no longer restricting to existing homogeneous and the limited 'asymmetric' assumption of heterogeneity. The new models are validated through extensive experimentation with two systems covering two different types of core heterogeneity:

1. ARM big.LITTLE architecture with iso-ISA core performance and power heterogeneity. Functional execution equivalence is established at the instruction level.

2. Windows laptop with CPU+GPU+GPU architecture representing ISA-level heterogeneity. Functional execution equivalence is established at the workload item level.

We will demonstrate the validity of the method by comparing our experimental results with model-calculated results, and with a software tool http://async.org.uk/prime/PER/.

[A. Rafiev, M. Al-Hayanni, F. Xia, R. Shafik, A. Romanovsky and A. Yakovlev, "Speedup and Power Scaling Models for Heterogeneous Many-Core Systems," in *IEEE Trans. Multi-Scale Comp. Syst.*]

ARCHON – AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK

The analysis for extra-functional properties such as power and performance takes a critical role in the system design workflow. Hardware-software co-simulation is one of the commonly used ways to perform this type of analysis. However, with the modern development of many-core systems the problem of scalability is becoming a bottleneck for all analysis techniques including simulation, especially when a simple extrapolation from the single core results is unacceptable. For instance, complications with memory access affect the speedup and power of many-core systems as much as core heterogeneity does.

This University Booth demonstrates ArchOn, a tool based on representing systems as resource graphs, which is layer-agnostic and facilitates selective abstraction, helping to minimize the complexity of simulations.

A set of Networks-on-Chip and bus-based topologies will be presented as use case examples to demonstrate the usefulness of ArchOn in analyzing speedup and power for systems with both core heterogeneity and memory architecture heterogeneity.
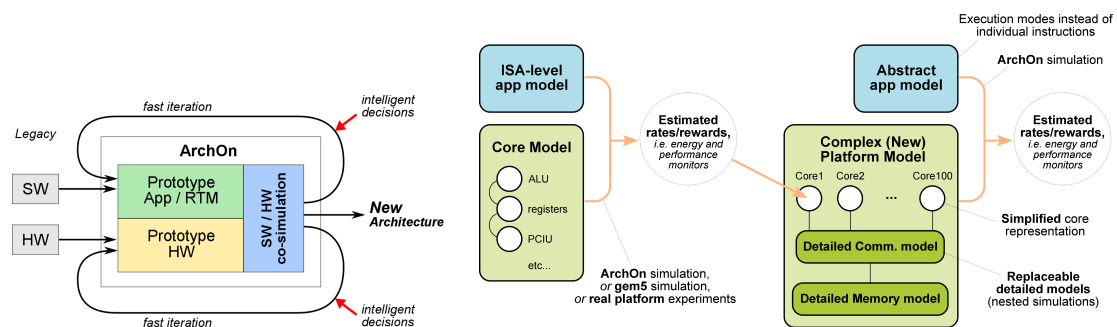

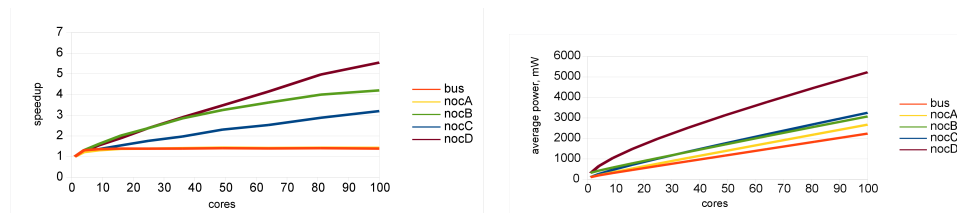Figure 1 ArchOn in the system design process (left) and selective abstraction using ArchOn.

Figure 2 Speedup and power comparisons of different memory architectures obtained using ArchOn.

[A. Rafiev, F. Xia, A. Iliasov, A. Romanovsky, A. Yakovlev. "Selective Abstraction for Estimating Extra-Functional Properties in Networks-on-Chips Using ArchOn," In ACSD2017, Zaragoza, Spain, 2017.]

http://async.org.uk/