CO-DESIGN OPPORTUNITIES WITH ARM-BASED ARCHITECTURES

Ananta Tiwari, PhD
Director of Research Operations
EP Analytics, Inc.

Collaboration with Ames Lab (Dr. Mark Gordon)

EPAnalytics

EP Analytics Overview

- We are performance and power modeling and analysis experts. We utilize in-house specialized tools to develop power & performance characterizations of the software/applications and the hardware.
  - Optimized tools enable working efficiently on large scale systems and production HPC applications
  - Methodologies to determine computations within applications that are affected by (or are sensitive to) the limitations of the hardware
    - ✓ ***Identify sections of source code that map well to future hardware & why***
    - ✓ ***Identify what architectural components affect performance***
    - ✓ ***Develop efficient porting & optimization strategies for new hardware***

- Expertise analyzing HPC workloads from DoE, NNSA, NSF, & DoD
- Research collaborations with DOE: Ames, LLNL, LBL, NERSC, ORNL, ANL, Sandia, PNNL, Los Alamos; DOD: HPCMP, CREATE, PETTT, MHPCC; & NSF: TACC, PSC, NCSA

What is co-design?

"Co-design refers to a computer system design process where scientific problem requirements influence architecture design and technology and constraints inform formulation and design of algorithms and software."

[From: http://science.energy.gov/ascr/research/scidac/co-design/]

Co-design challenges

- Investment in software is large – decades of application development in “legacy” applications (same for ISVs)
  - Heterogeneous architectures with accelerators require even larger investment in development
- Performance constraints
  - Solutions need to exploit different types of parallelism with minimal investment in development
- Power constraints
  - Operational costs dominating overall cost of total purchase.
  - DOE recommends a 20MW power budget for a >100X performance improvement

Addressing co-design challenges

- Power and performance → co-design solutions that enhance energy efficiency
- Software development investment → solutions that are easy to deploy and use
- Potential solutions:
  - ***Reactive***: wait for traditional processors to improve in energy efficiency
    - Single-path solution: you get what hardware designers offer you; slow process
  - ***Proactive***: leverage ARM's extensible ISA model and low-power design to influence the design of future supercomputers
    - Many-path solution: work with multiple vendors within the ARM ecosystem to design HPC processor architectures
    - Ever maturing system software stack

(Credit: Jim Ang's presentation titled "ARM as an Enabler for HPC co-design" at ISC 2014)

Enabling components -- co-design w ARM64

- Performance analysis framework that allows co-designers to develop clear understanding of:
  - Application requirements (i.e., code characterizations) in the form of memory requirements, data parallelism opportunities, communication requirements, etc.
  - Properties and capabilities of hardware (i.e., machine characterizations) in the form of memory bandwidth/latency, size of vector units, communication bandwidth/latency, etc.
- Combine the code and machine characterizations to find:
  - What architectural elements might be bottlenecking application performance
  - What changes can be made on application side to make better use of the hardware

Our ARM-based co-design efforts

- DOE funded SBIR grants: developing binary instrumentation toolkit (EPAX) for ARM architectures
  - Full-featured static analysis toolkit (EPAX) complete
  - Released as open-source to foster tool ecosystem for ARM
  - Performance efficiency of instrumented binaries at scale is key

- AFOSR CODAASH (A co-design approach for advances in software and hardware, PI: Mark Gordon, IA State)
  - Analyze multiple ARM-based systems for performance and power specifically for GAMESS, which is a quantum chemistry code, and other computational chemistry codes

Example co-design efforts

- Energy efficiency and performance implications of many-core ARM64 designs (Cavium ThunderX) (paper under preparation)
- Analysis of architectural bottlenecks for X-Gene1 [1, 2] and ThunderX processors
- Impact of reduced per-core memory bandwidth to application performance [3]

[1] Tiwari, A., Keipert, K., Jundt, A., Peraza, J., Leang, S., Laurenzano, M., Gordon, M., Carrington, L. (2015): **Performance and Energy Efficiency Analysis of 64-bit ARM using GAMESS**. *Second International Workshop on Hardware-Software Co-Design for High Performance Computing (Co-HPC)*.

[2] Laurenzano, M., Tiwari, A., Cauble-Chantrenne, A., Jundt, A., Peraza, J., Ward, W., Campbell, R., and Carrington, L. (2016): **Characterization and Bottleneck Analysis of a 64-bit ARMv8 Platform.** *In International Symposium on Performance Analysis of Systems and Software (ISPASS)*

[3] Tiwari, A., Gamst, A., Laurenzano, M., Schulz, M., Carrington, L. (2014): **Modeling the impact of Reduced Memory Bandwidth on HPC Applications**. *In 20th International Conference in Parallel Processing (Euro-Par)*.

Conclusions

- ARM enables transformational research and development in co-design with its extensible ISA model
- Community needs performance analysis framework to be able to understand the requirements of applications and map that to proposed new innovations in hardware
- We look forward to working with the community to advance HPC performance on ARM

Come talk to us to learn more about EPAX and our co-design efforts!

Acknowledgements

Some of the work presented in the talk was supported by the U.S. Air Force Office of Scientific Research under AFOSR Award No. FA9550-12-1-0476 and DOE SBIR funding [DE-SC0009497 and DE-SC0013164].

Contributors:
Laura Carrington
Mark Gordon
Kris Keipert
Adam Jundt
Allyson Cauble-Chantrenne

QUESTIONS:

ananta.tiwari@epanalytics.com
laura.carrington@epanalytics.com

http://www.epanalytics.com/