High Performance Computing Architecture and System Laboratory (HPCAS)

Yonghong Yan, Kewei Yan, Anjia Wang, Yaying Shi, Xinyao Yi
Department of Computer Science | UNC Charlotte | https://passlab.github.io

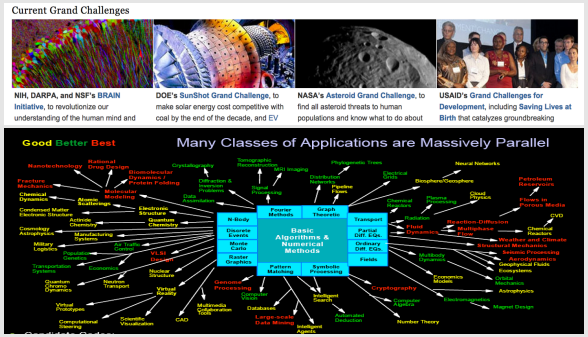
Introduction

The **High Performance Computing Architecture and System Laboratory** (HPCAS) at **University of North Carolina at Charlotte** does computer system research for parallel and high performance computing. We develop parallel programming models, compiler and runtime systems to accelerate performance improvement, application development and energy efficiency improvement for HPC applications. We develop, improve and use software and hardware tools to deepen our understanding of parallel execution and resource utilizations. We explore new hardware architecture and techniques to achieve these goals and we collaborate with scientists to apply our solutions to real world problems. We interact, educate and mentor students and peers to promote the area of high performance computing and computer systems.

HPC Applications, Systems and Architectures

HPC solve large and complex computational science problems!
- Need large amount of computational resources
- Process large amount of data

HPC systems and architectures
- Supercomputers, e.g. the 500 fastest, http://top500.org
- Parallel computing is the solution so far.
- Performance, programmability and power are the challenges.

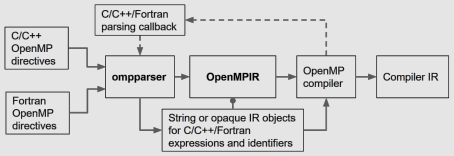


Parallel Programming Models, Compiler and Tools (System Software)
Performance, Programmability and Power

Parallel Programming Models and Compilers

OpenMP and compiler support
- ROSE compiler and LLVM OpenMP runtime systems
- Loop transformation and code generation for CPU, GPU and FPGA

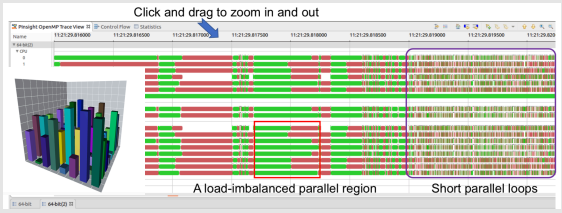
ompparser - standalone unified OpenMP parser for C/C++ /Fortran.
- Can be used standalone for static source code analysis
- Integrating ompparser into an OpenMP compiler
- A complete OpenMP grammar in the Backus-Naur Form



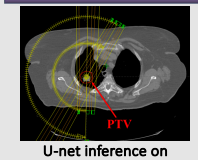
HPC Performance Tools

Trace, analyze and visualize parallel execution
- Support OpenMP, MPI, CUDA, AMD GPUs
- 2D and 3D visualization, online and offline

Pinpoint performance problem and corelate with sources
- Load imbalance, bottleneck, excessive synchronization
- Data movement and cache behavior, etc

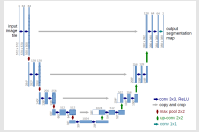


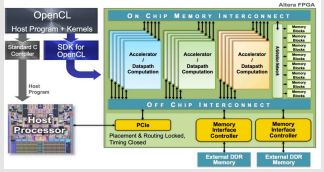
Efficient Contouring of Medical Images



U-net inference on embedded system and FPGA

- Implement a U-net model on embedded system
- Utilized the advantage of FPGA to increase the inference speed on embedded system
- Design the U-net architecture on FPGA
- Highly optimized convolution layer with max-pooling
- Increase inference speed by mix-precision, special design workflow and reusable data in convolution layer.

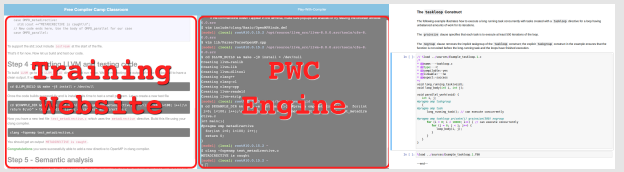




CloudRDL (Research, Development and Learning from Cloud)

FreeCompilerCamp.org: A free and open online training platform to quickly extend compilers and help developers learn the skills of compiler development.
- Online interactive, self-paced learning platform using Play-With-Docker Engine: free, open, pre-configured, accessible via a browser.
- Crowd-source: Users can contribute and use their own tutorials.
- Cloud-based: self-deployable.
- Instructors or students can easily make customization and deploy it on any local server or even their own laptop.

Jupyter-notebook-based online tutorial: The OpenMP 5.0 examples are translated into notebooks to enable interactions between tutorial and users.
- Online, jupyter-notebook-based self-learning materials : free, open, browser accessible.
- Multiple kernels available: python/python3 kernels for default, C based Native kernel applied.
- Flexible: users may translate other tutorials written by LaTeX to notebooks via our python script.



Publications

[1] FreeCompilerCamp.org: Free and Open Online Training for Developing OpenMP Extensions Anjia Wang, Alok Mishra, Chunhua Liao, Yonghong Yan and Barbara Chapman OpenMPCon 2019

[2] Cloud-based Collaborative Development Environments for Research Software Tools and Applications, Yonghong Yan, and Chunhua Liao Workshop on Best Practices and Tools for Computational and Data-Intensive Research at ICS 2019, June 26, 2019

[3] ompparser: A Standalone and Unified OpenMP Parser Anjia Wang, Yaying Shi, Xinyao Yi, Yonghong Yan, Chunhua Liao, and Bronis R. de Supinski IWOMP'19: International Workshop on OpenMP (IWOMP) 2019

[4] Extending OpenMPMetadirectiveSemantics forRuntime Adaptation Yonghong Yan, Anjia Wang, Chunhua Liao, Thomas R.W. Scogland, and Bronis R. de Supinski IWOMP'19: International Workshop on OpenMP (IWOMP) 2019

Sponsors

NSF · Lawrence Livermore National Laboratory · XILINX · MAXELER Technologies Maximum Performance Computing