TEAM POWERful Peking University, China

Yun Liang (advisor), Xiaolin Wang (advisor), Yilong Li, Lei Yang, Wenbin Hou, Zhenxin Fu, Haoze Wu, Zhuohan Li
{ericlyun,wxl,yilong,yangl1996,catchyrime,fuzhenxin,functioner,zhuohan123}@pku.edu.cn

About Peking University

Peking University (PKU) is a comprehensive national key university in China. Founded in 1898, Peking University was the first national university covering comprehensive disciplines in China, and has been a leading institution of higher education in China since its establishment. Here, top Chinese students seek academic excellence and dedication to the society.

Peking University ranks as one of the best universities in China and a top university in Asia and worldwide.

HPC in Peking

High-performance computing education and infrastructure are being developed rapidly in Peking University. For example, the PKUHPC Platform built in 2015 is one of the supercomputers owned and operated by the university that ranks top 100 in China.

Peking University has strong background in HPC related research, our advisors are all experts in high performance computing. Peking University also includes the HPC-related courses such as "Introduction to Parallel and Distributed Computing" , "Introduction to Computer Systems", "Heterogeneous High Performance Computing", etc. in the curriculum.

About Our Team

This is the first time that we participate in the Student Cluster Competition, and most of us are junior and sophomore students majoring in Computer Science.

LI Yilong Senior
Team leader and coordinator
Expert in Linux Maintenance
Focus: Benchmark, system administration

YANG Lei Junior
Team leader
Expert in networking and FPGA
Focus: Password recovery

HOU Wenbin Junior
Expert in CUDA and Xeon Phi
Focus: ParaView

FU Zhenxin Junior
Research interests in NLP
Focus: Reproduction Task (ParConnect)

WU Haoze Sophomore
Expert in Algorithm
Won medals in ACM/ICPC
Focus: Power management and Mystery

LI Zhuohan Sophomore
Expert in Algorithm
Won medals in ACM/ICPC
Focus: System resource visualization and Mystery

Prof. LIANG Yun
Advisor, School of EECS
Research Interest: High performance computing, compilation and architecture.
Publication in Top-tier conferences: HPCA, ISCA, MICRO, FPGA, DAC, etc.

Prof. WANG Xiaolin
Advisor, School of EECS
Research interests : High performance computing, algorithm design.
Publication in Top-tier conferences: TACO, USENIX ATC, ICS, etc.

Our Preparation

Self-study

We started preparing the SC16 SCC as soon as the applications were released. All of us have taken training courses for parallel programming, like OpenMP, MPI and CUDA programming, and got to be more familiar with Linux systems. Lei has studied FPGA programming using HLS and SDAccel, and he succeeded in cross-compiling the SDAccel binaries and runtime libraries in POWER systems, which makes SDAccel programming in POWER possible.

We have delved deeply into algorithms behind the ParConnect and ParaView task, compared among common methods in the password cracking task. We have even simulated the ParaView task using bioinformatics and astro-nomics data in order to have a better view about the running time of the tasks.

Teamwork

Everyone was expected to develop some general knowledge in all areas, and specialize in one area. We have weekly discussions with our advisors since February. Everyone could report the knowledge they learnt, the progress they made and difficulties they encountered.

Support from our vendors

We have benefitted from the collaboration with OMNISKY, IBM, NVIDIA and Xilinx. OMNISKY provided us with high-end clusters and devices, and hosted our machines. IBM and NVIDIA have provided vendor-optimized benchmark applications, toolchains and math libraries so that we can fully optimize our applications. We have also learnt a lot of skills about optimizing scientific applications and HPC system maintenance from the collaboration.

Special Thanks

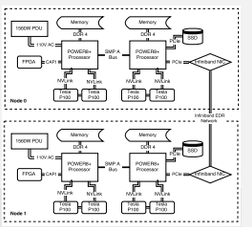
OMNISKY 睿天 IBM NVIDIA XILINX

Hardware and Software Config

Hardware Configuration

OMNISKY & IBM "Minsky" S822LC Server × 2

CPU	4×IBM POWER8+	8 core 64 threads, 4GHz
MEM	2×256GB	DDR4 2133
GPU	8×NVIDIA P100	with NVlink
HDD	2×Intel 750 Pro	400GB PCI-e SSD
NIC	2×Mellanox ConnectX-4 100Gb Infiniband EDR	
FPGA	2×Semptian NSA-120 Xilinx Ultrascale board	



The IBM "**Minsky**" server is using Nvlink ports for **P100** accelerators, providing 80GB/s bandwidth between GPUs. **FPGA** accelerators are also used for tasks like password recovery, and the **SDAccel** toolchains provided by Xilinx has made programming on FPGA boards much easier.

Software Configuration

This year we are going to use **Ubuntu Linux 16.04**, which supports POWER well and toolchains provided by **IBM** and **NVIDIA**, in order to get best performance on Minsky.

Math Libraries ESSL
XL C/Fortran Compiler
Spectrum MPI

IBM ubuntu CUDA C/C++

Runtime Libraries
CUDA Profilers

allinea PERFORMANCE REPORTS

Optimizations and Strategies

Benchmark

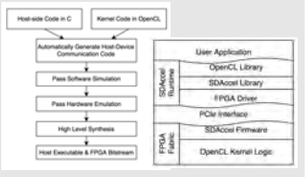
We are using HPL and HPCG binaries provided by our sponsor NVIDIA and IBM which are specially optimized for MINSKY. Within the 3kW power limitation, HPL can score over **25 TFlops** and HPCG benchmark can score over **660 GFlops**.

ParaView

We **rewrote some filters** so that much more filters can utilize our GPU accelerators using NVIDIA CUDA.
Checkpointing are specifically implemented in ParaView by our members, making it prepared for sudden power cuts.

Password Recovery

For **MD5crypt** passwords, we are using open-source application hashcat supporting CUDA and OpenCL.
For **bcrypt** consisting of many random memory accesses and logic arithmetic, we use **FPGA accelerators** to customize data path and memory configuration. The **Xilinx SDAccel** HLS toolchain facilitates the development of FPGA based applications.



Development and runtime flow of SDAccel programs

Reproduction Task

We have run different datasets on our 2-node clusters using the SMT hardware threading technologies and Allinea profilers to reproduce charts in the original paper.

Mystery Application

From the experience of application optimization before, differenet levels of optimization will be considered.

Running Strategies

We have profiled all the applications to characterize them. We have deep understanding of their computation and memory details. When we get the size and complexity of the application during the competition, we will estimate the runtime for each application based on this. Then, we schedule the applications to maximize the resource utilization.

Power Management

- During the test runs we **record per-node power consumption** and break it down
- **Frequency** scaling and power capping techniques are used to limit power consumption
- We use **visualization tools like Grafana** to get real-time power consumption curves
- The **sudden power cut** are considered in all applications and **checkpointing** tools are used

We Will Win!

- We have **passionate members** who worked together for the competition!
- Each application is well studied and optimized for months, so we have a **good understanding of different applications** for the competition.
- Members are experienced in areas like **System Administration, Power Management** and **Computer Networking**
- Well-designed **hardware and software configurations** are prepared, and we have fully exploited the performance of **P100 and FPGAs**.
- Support from **OMNISKY, IBM, NVIDIA and Xilinx**.