Dr. Haohuan Fu is an associate professor in the Ministry of Education Key Laboratory for Earth System Modeling, and Center of Earth System Science in Tsinghua University. He is also the associate director of the National Supercomputing Center in Wuxi. His research interests include design methodologies for highly efficient and highly scalable simulation applications that can take advantage of emerging multi-core, many-core, and reconfigurable architectures, and make full utilization of current Peta-Flops and future Exa-Flops supercomputers; and intelligent data management, analysis, and data mining platforms that combine the statistical methods and machine learning technologies.

Sunway TaihuLight Supercomputer: System and Application Updates
The Sunway TaihuLight Supercomputer: Updates on System and Application

Haohuan Fu

Tsinghua University and
National Supercomputing Center in Wuxi

Nov, 16, 2016 @ HPC Connection
Sunway TaihuLight: An Overview

The SW26010 CPU

Application Updates
Outline

- Sunway TaihuLight: An Overview
- The SW26010 CPU
- Application Updates
The Sunway Series Supercomputers

Sunway BlueLight
2012
• National Supercomputing Center in Jinan
• 1 PFlops
• multi-core processor (16-core)

Sunway TaihuLight
2016
• National Supercomputing Center in Wuxi
• 125 PFlops
• many-core processor (260-core)
Sunway TaihuLight: an Overview

Homegrown many-core processor: SW26010

- 260 cores per chip
- 3 Tflops

The first system in the world that provides over 100 Pflops performance with over 10 million cores

- theoretical peak 125 Pflops, 2.5 times improvement over before
- LINPACK performance 93 Pflops, 3 times improvement over before

High efficiency of the overall system

- 6.05 Gflops/Watt, 3 to 6 times improvement over Tianhe-2, Titan, and K

Three full-scale applications elected as 2016 Gordon Bell finalists
Outline

- Sunway TaihuLight: An Overview
- The SW26010 CPU
- Application Updates
SW26010 Processor

- China’s first homegrown many-core processor
  - Vendor: Shanghai High Performance IC Design Center
  - supported by National Science and Technology Major Project (NMP): Core Electronic Devices, High-end Generic Chips, and Basic Software
SW26010: General Architecture

- 260 cores per processor
- 4 Core Groups (CGs), each of which has:
  - 1 Management Processing Element (MPE)
  - 64 (8x8) Computing Processing Elements (CPEs)
SW26010: MPE and CPE

- **Management Processing Element (MPE)**
  - 64-bit RISC core
  - support both user and system modes
  - 256-bit vector instructions
  - 32 KB L1 instruction cache, and 32 KB L1 data cache
  - 256 KB L2 cache

- **Computing Processing Element (CPE)**
  - 8x8 CPE mesh
  - 64-bit RISC core
  - support only user mode
  - 256-bit vector instructions
  - 16 KB L1 instruction cache, and a Scratch Pad Memory (SPM)
SW26010: General Architecture

- Memory Level
- LDM Level
- Register Level
- Computing Level

- LDM = SPM
SW26010: General Architecture
SW26010: General Architecture
DMA Transfer

- CPE initializes the contiguous or strided data transfers between the main memory and the SPM
  - support for specific DMA and barrier instructions
- Support for asynchronous DMA

![Diagram of DMA Transfer]

**Synchronous DMA**

- Load A
- Store A
- Load B
- Store B
- Compute A
- Compute B

**Asynchronous DMA**

- Load A
- Load B
- Store A
- Store B
- Compute A
- Compute B
DMA Transfer: Multicast

- Multicast DMA for computations that require data sharing among CPEs
  - requested by one CPE, data sent to multiple CPEs
- improve data reuse and reduce the memory bandwidth requirement
On-chip register communication

- A tight coupling between the computing pipeline and the on-chip network, to enable direct data exchange between register files of different CPEs
  - register communication instructions
  - fine-grained, low-latency
  - support for both point-to-point and multicast modes

- Compared with the traditional on-chip inter-core communication, register communication can bypass the cache hierarchy, and provides a better efficiency and lower power consumption.
Hardware Synchronization Support

- CPEs compute in a collaborative way, which requires frequency synchronizations.

- Traditional software synchronization uses atomic operations in the main memory space, leading to long latency and memory overhead.

- Hardware synchronization to improve efficiency:
  - the synchronization controller collects the request signal from all the involved CPEs, and returns the confirmation signal to each of them.
In fact, the sum of the performance for computers in position 2 through 7 is just barely greater than the performance of the Sunway system.

The Sunway TaihuLight system, based on a homegrown processor, demonstrates the significant progress that China has made in the domain of designing and manufacturing large-scale computation systems.

The fact that there are sizeable applications and Gordon Bell contender applications running on the system is impressive and shows that the system is capable of running real applications and not just a stunt machine.

It is clear that they are on a path which will take them to an exascale computer by 2020, well ahead of the US plans for reaching exascale by 2023.
Comments from Satoshi Matsuoka

Satoshi Matsuoka
@ProfMatsuoka

I was quite different in Japan in

下午4:40

Satoshi Matsuoka
@ProfMatsuoka

TaihuLight physical design is excellent with low num. of chips, dual-sided surface mounting of all components for dense cold plate cooling.

下午5:57 - 2016年11月3日
Comments from Satoshi Matsuoka

Satoshi Matsuoka
@ProfMatsuoka

I was quite different in Japan.

Satoshi Matsuoka
@ProfMatsuoka

TaihuLight's dual-sided plate cooler.

Also impressive was their software and application efforts. Contrary to my speculations OpenACC does work, used in many of their real apps.
Finally their design was cost&utility conscious. No expensive parts, quack architecture, etc. Sunway apparently plans to sell the machine.

下午6:08 - 2016年11月3日
Outline

- Sunway TaihuLight: An Overview
- The SW26010 CPU
- Application Updates
Application Summary

- Over 100 research institutes, over 60 large-scale applications
- 19 application domains (weather and climate, aeronautic and astronautic, marine environment, life science, new energy, ship engineering, new material)
- 6 full-scale applications, and 3 Gordon Bell finalists, 18 half-scale applications, and 22 million-core-scale applications
Sunway TaihuLight: Applications

Key application domains:

- Earth system modeling / weather forecasting
- Advanced manufacturing (CFD/CAE)
- Life science
- Big data analytics
Sunway TaihuLight: Applications

Key application domains:

- Earth system modeling / weather forecasting
- Advanced manufacturing (CFD/CAE)
- Life science
- Big data analytics
Earth System Modeling

- Refactoring and Optimizing the Community Atmospheric Model (CAM)

- A High-Resolution Experimental Atmospheric Model for Sunway TaihuLight
Porting of CESM onto Sunway TaihuLight

- Porting of 4 component models (over 1 millions lines of code)
- Accuracy verification by 3-5 year simulation of different compsets

- Large-scale run on Sunway TaihuLight
  - 24,000 CGs, million cores
  - high resolution ATM (25-km) and OCEAM (10-km) modeling

- Refactoring of CAM
  - 20-30x speedup for compute kernels
  - 2x speedup for the entire model

Multi-disciplinary team (Tsinghua, BNU)
CAM kernels: MPE+64CPE vs MPE

Dynamic core

Physics schemes
CAM model: scalability and speedup

- million core scale, 2.81 SYPD
- many-core refactoring for the entire model
- competitive simulation speed to the same model on NCAR Yellowstone
Earth System Modeling

- Refactoring and Optimizing the Community Atmospheric Model (CAM)

- A High-Resolution Experimental Atmospheric Model for Sunway TaihuLight
A High-Resolution Experimental Atmospheric Model for Sunway TaihuLight

- A experimental project started by National Supercomputing Center in Wuxi

- Hardware-Software Co-Design
  - structural change of the model components
  - many-core accelerations of all compute-intensive parts
  - a loosely coupling scheme between the dynamic core and the physics scheme to further improve the scalability

- Targeting at high-resolution simulation scenarios
  - 10 KM to 3 KM global resolution
  - scaling to the entire system of TaihuLight
10 KM Results

Precipitation rate (mm/6hour)

Upwelling longwave flux at top of model (W/m²)
Collaboration between NSCC-Wuxi and National Climate Center on Season-to-Season Climate Prediction System Development and Operation

High-resolution short-term climate prediction operational system based on CWRF (regional)

Multi-model ensemble short-term prediction operational system, collaborated with other climate research institutes (global)
CIESM (Community Integrated Earth System Model) for CMIP6 (Climate Model Inter-comparison Project 6)
Sunway TaihuLight: Applications

Key application domains:

- Earth system modeling / weather forecasting
- Advanced manufacturing (CFD/CAE)
- Life science
- Big data analytics
Existing Projects on Sunway TaihuLight
Supercomputing for Advanced Manufacturing

World-leading Computing Facility → Industrial Design and Manufacturing Simulation Platform → Industrial Upgrading

World-famous Manufacturing Area → Industrial Design and Manufacturing Simulation Platform → Advanced Center for Manufacturing Innovation
Tsinghua-ANSYS Collaborative Plan (for the x86 System)

12 Components of ANSYS

7,012-core computing capability (2,048-core Fluent, 2,048-core CFX)

The largest industrial simulation center in China
Sunway TaihuLight: Applications

Key application domains:

- Earth system modeling / weather forecasting
- Advanced manufacturing (CFD/CAE)
- Life science
- Big data analytics
cryo-EM Imaging Processing

- We have employed a new empirical Bayesian approach to refinement of 3D reconstructions of proteins in electron cryo-microscopy (cryo-EM).
  - Much faster: weeks to hours
  - Much more precise: resolutions improve to 3 Angstrom below

100000+ projection images
Sunway TaihuLight: Applications

Key application domains:

- Earth system modeling / weather forecasting
- Advanced manufacturing (CFD/CAE)
- Life science
- Big data analytics
swDNN: Deep Learning Framework for Sunway

Over 2Tflops learning performance per CPU
Sunway TaihuLight

International Workshop on HPC Architecture, Software, and Application at an Extreme Scale
Sep 17, Wuxi, China
THANK YOU