Big Data and Extreme Computing: Software Ecosystem for Sunway Processors

Jingheng Xu
National Supercomputing Center in Wuxi
Department of Computer Science, Tsinghua University
May 15th, 2019

国家超级计算无锡中心
National Supercomputing Center in Wux

Outlines

- Overview of the Exa-scale Supercomputers in China
- Development of Performance Tools and Libraries on Sunway Processors
- HPC & Big Data Applications on Sunway-TaihuLight

Outlines

- **Overview of the Exa-scale Supercomputers in China**
- Development of Performance Tools and Libraries on Sunway Processors
- HPC & Big Data Applications on Sunway-TaihuLight

Exascale Preparation Status – Sunway Plan

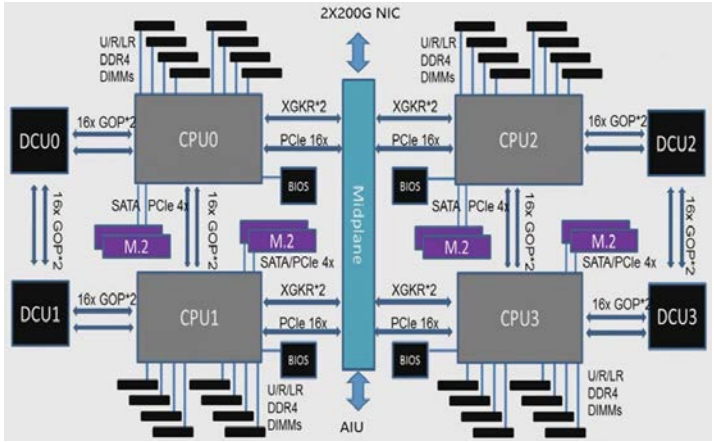
- Prototype of Sunway exascale supercomputer:
  - Chip: SW26010 whose peak performance is 3.06TFlops
  - Node number / Peak Perf: 512 / 3.13PFlops
  - Linpack efficiency: 81.5%
- Exa-scale Plan:
  - New generation of CPU to improve single-chip performance
  - Faster memory systems to improve bandwidth
  - Similar master/slave core model will be adopted to facilitate program deployment
- Located in Tsingtao
  - National Laboratory for Marine Science and Technology
  - Marine applications are the main focus

Exascale Preparation Status – Tianhe Plan

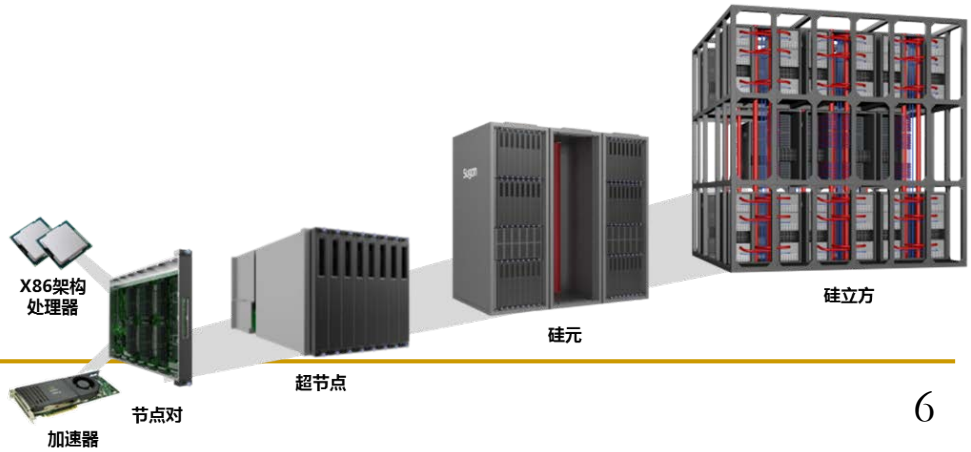
- Prototype of Tianhe exascale supercomputer
  - Chip: Matrix 2000+ whose peak performance is 2.08TFlops
  - Node number / Peak Perf: 512 / 3.14PFlops
  - Linpack efficiency: 78.5%
- Exa-scale Plan:
  - New generation of CPU (Matrix 3000)
  - Faster memory systems
  - High scalable 3D butterfly network enables the good scalability of the supercomputer system (enabling connection of more than 100,000 nodes)
- Located in TianJin

Exascale Preparation Status – Sugon Plan

- Prototype of Sugon exascale supercomputer
  - Chip: Hygon x86 processors + DCU accelerators
  - Node number / Peak Perf: : 512 / 3.18PFlops
  - Linpack efficiency: 71.5%
  - Efficient cooling design



- Exa-scale Plan:
  - 100P machine will be deployed in Zhengzhou by 2020
  - Exa-scale supercomputer is expected to be deployed by around 2022



Outlines

- Overview of the Exa-scale Supercomputers in China
- **Development of Performance Tools and Libraries on Sunway Processors**
- HPC & Big Data Applications on Sunway-TaihuLight

Performance Tools and Libraries on Sunway

- General Performance Tools have been well discussed in previous talks and paper

Figure from the presentation of Prof. Depei Qian in the SC2018;

The software-stacks could be divided into four levels;

All of these software-stacks are ready since the deployment of Sunway-TaihuLight supercomputer.

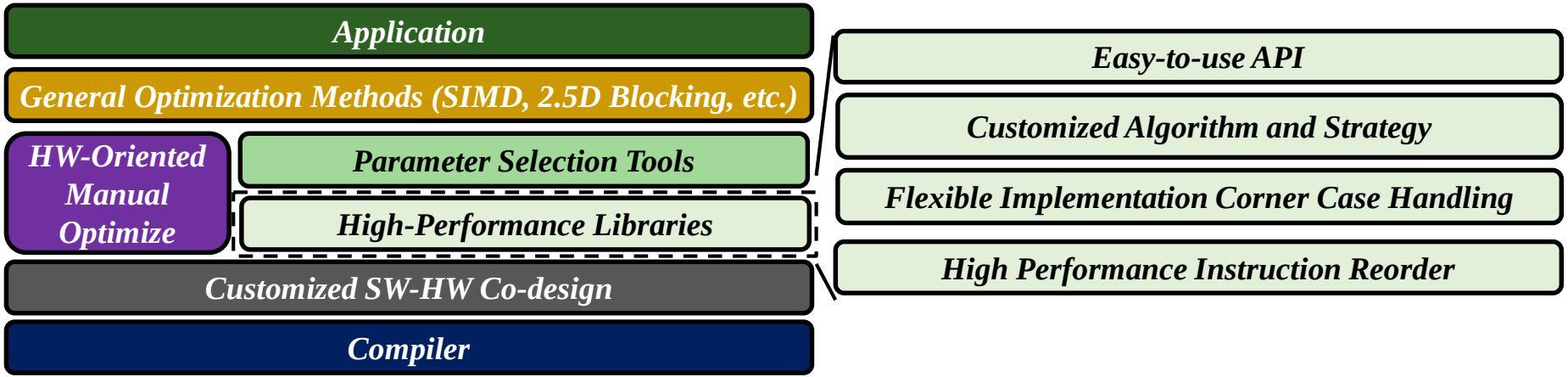
Performance Tools and Libraries on Sunway

- General Performance Tools have been well discussed in previous talks and paper
- Performance-Oriented Libraries are still needs to be fully developed on Sunway

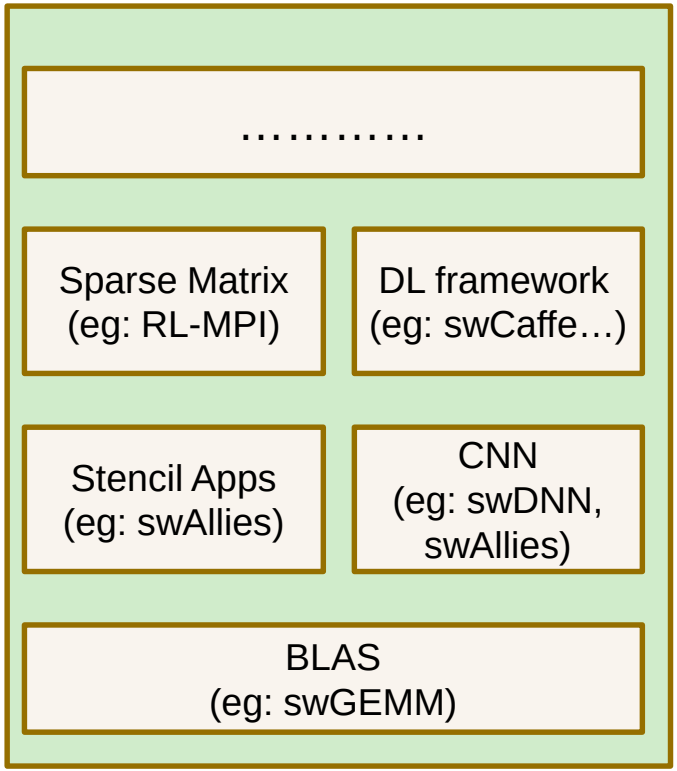
	Intel CPU	POWER CPU	Nvidia GPU	SW26010
Compute Unit	CPU core	CPU core	CUDA core	CPE
Exclusive Cache Space	L1-Dcache	L1-Dcache L2-cache	**None**	Local Data Memory (LDM)
Multi-core Shared Cache	L2-cache	L3-cache	L1-cache / SMEM	**None**
Chip Shared Cache	L3-cache	L4-cache	L2-cache	**None**
Main Memory	Main mem	Main mem	Global mem	Main mem

Performance Tools and Libraries on Sunway

- General Performance Tools have been well discussed in previous talks and paper
- Performance-Oriented Libraries are still needs to be fully developed on Sunway
- Application optimization stack to guide the library design



Libraries for HPC & Deep Learning



General-Used DMA_Read API

```
1: for (i=0; i<NY+ynh+yph, i++)
2:         dma_read_base(dma_get, src, dst, length);
3: dma_wait();
```

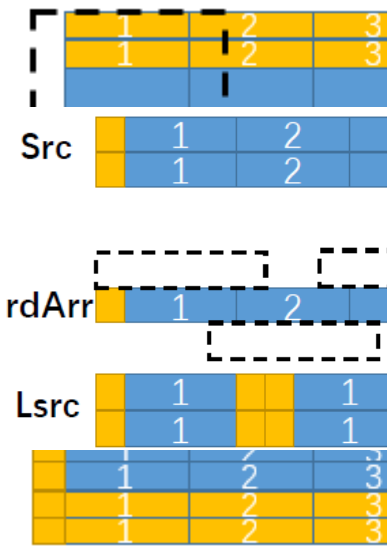
DMA_Read API of swAllies Strategy 1 (Detailed in Section 5)

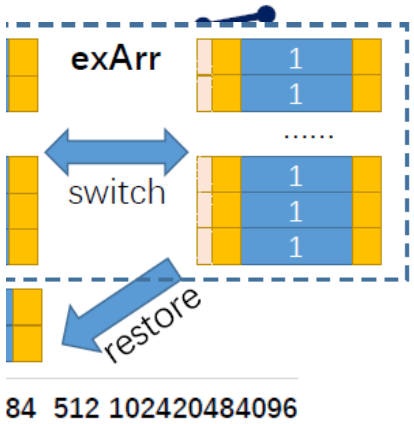
```
1: for (i=ynh; i<NY+ynh, i++)
2:         dma_read_base(type, src, dst, length, xnh, xph);
3: for (i=0; i<ynh, i++)
4:         dma_border_up(type, src, dst, upbdAdd, length, xnh, xph);
5: for (i=NY+ynh; i<NY+ynh+yph, i++)
6:         dma_border_down(type, src, dst, downbdAdd, length, xnh, xph);
7: dma_wait();
```

DMA_Read API of swAllies Strategy 2 (Detailed in Section 6)

```
1: for (i=0; i<NY+ynh+yph, i+=8)
2:         int rmd = NY+ynh+yph-i < 8 ? NY+ynh+yph-i : 8;
3:         dma_read_alg2(type, src, dst, rmd, length, xnh, xph);
4: dma_wait();
```

Libraries f



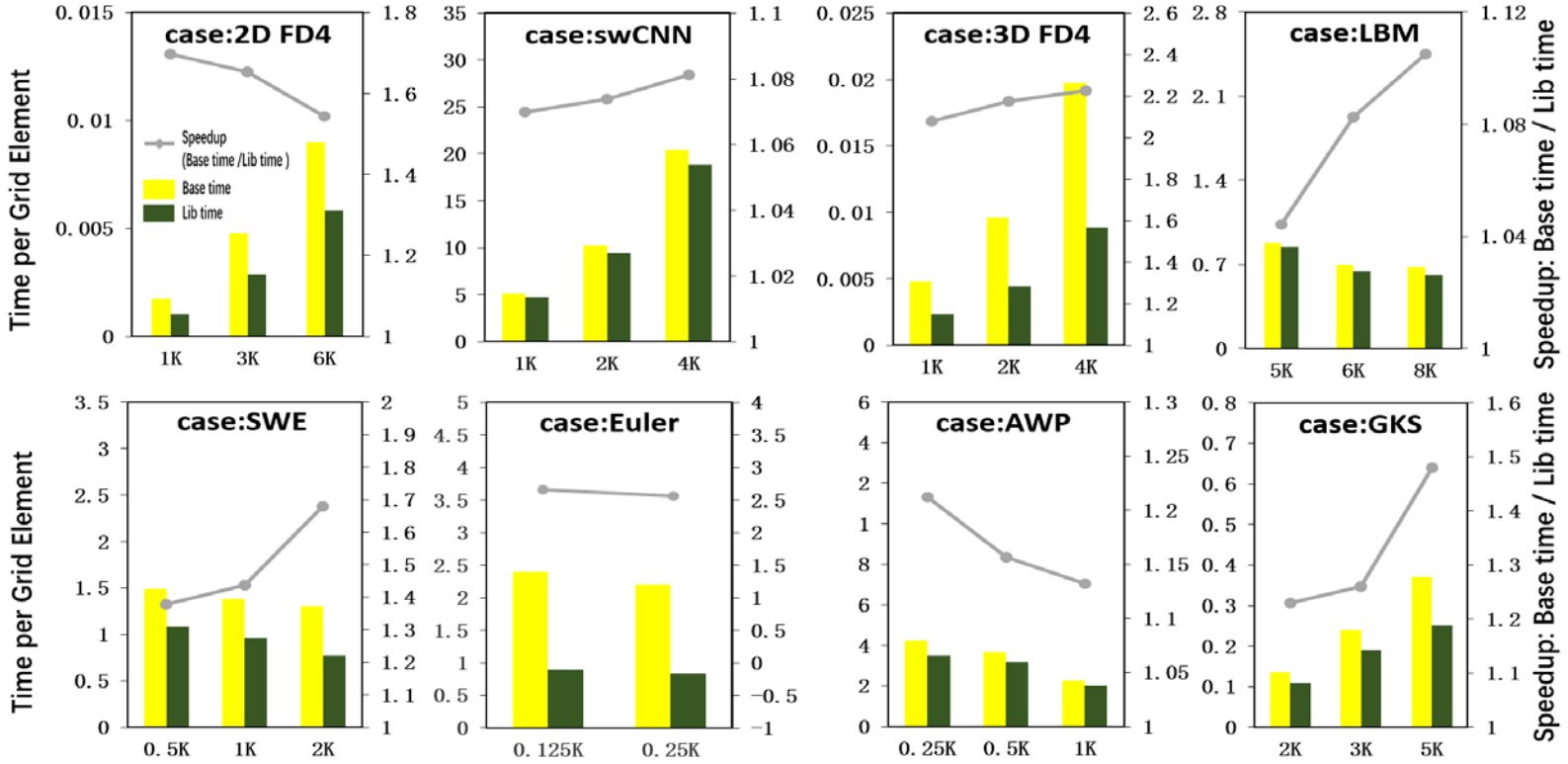


Two challenges t

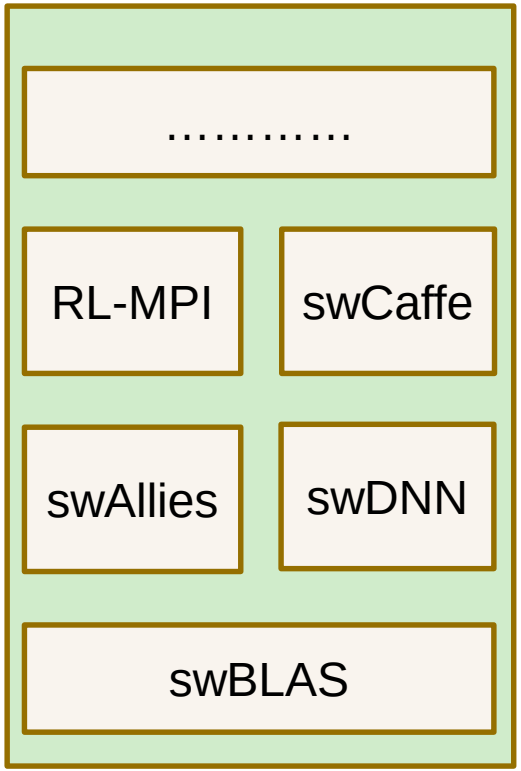
- Redundant DMA
- Insufficient use

ave been reduced;
eviated.

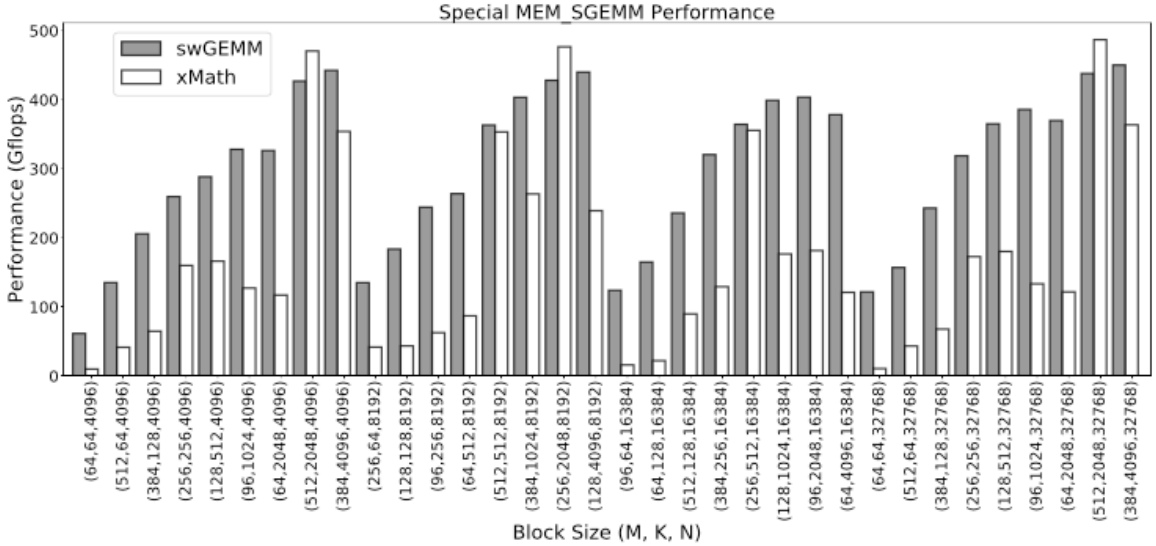
Libraries for HPC & Deep Learning



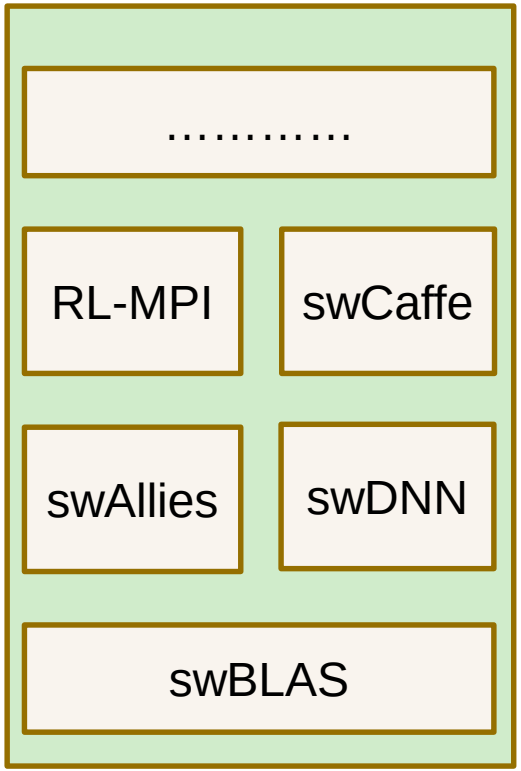
Second Cases: swBLAS



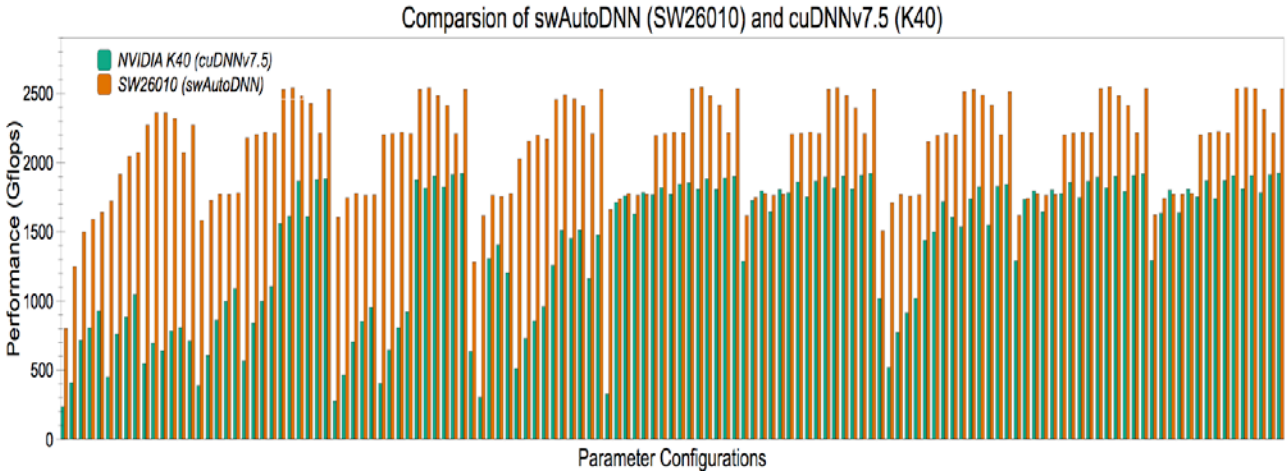
- Customized changes to swBLAS functions for deep learning tasks: **SWGEMM**
- Over 3x speedup compared with the **basic version** on 167 cases among 200 test cases



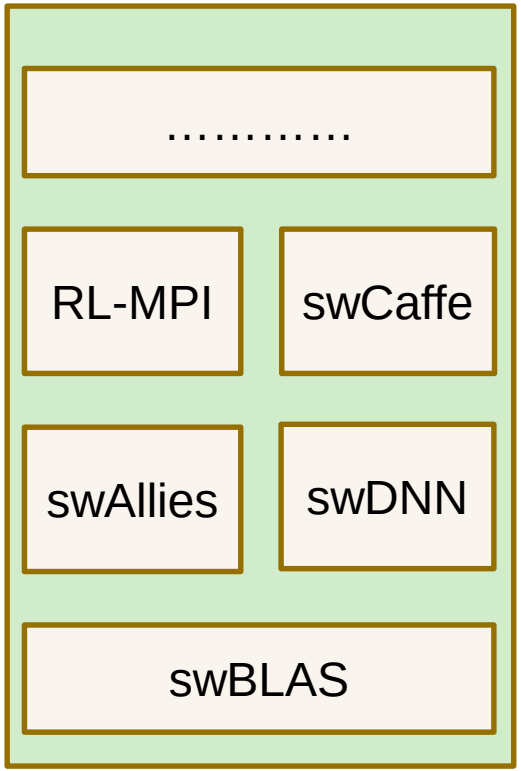
Third Case: swDNN



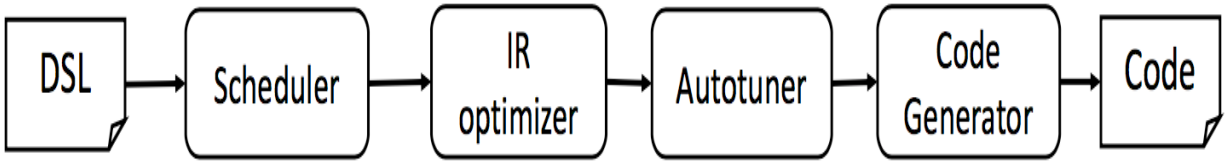
- Conv Layer is averaged **2.26x** faster than cuDNNv7.5 on K40 (159 test cases)



Libraries for HPC & Deep Learning



- Conv Layer is **2.26x** faster than cuDNNv7.5 on K40 (159 test cases)
- swAutoDNN is coming soon



- **Scheduler:** Loop, layout, vectorization transformation
- **IR optimizer:** DMA inference, Hiding Memory Access Latency, Boundary Processing.
- **Autotuner:** Selecting the best candidate with Performance Model
- **Code Generator:** Generate CPE code

Outlines

- Overview of the Exa-scale Supercomputers in China
- Development of Performance Tools and Libraries on Sunway Processors
- **HPC & Big Data Applications on Sunway-TaihuLight**

HPC & Big Data Applications

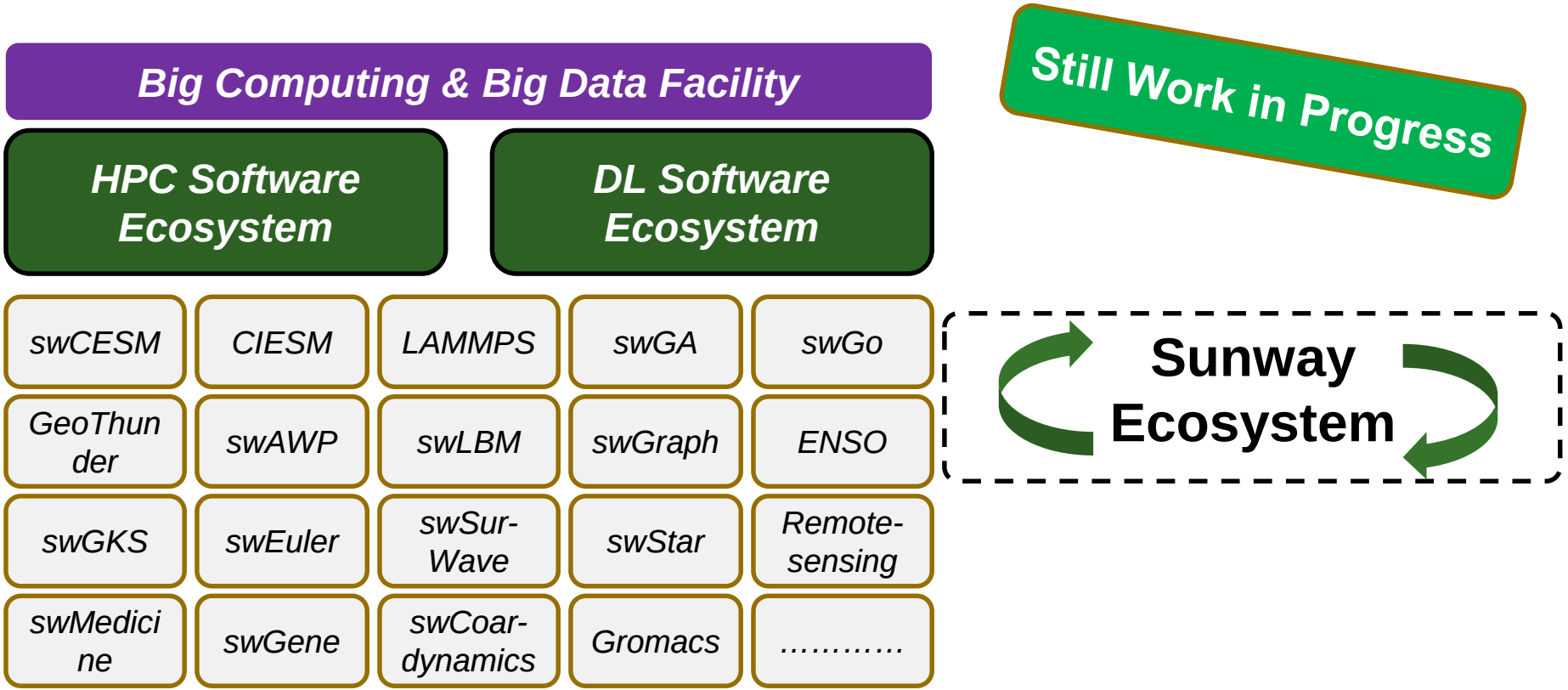
Overview:

- ***In three years, over 200 large-scale applications from over 300 research institutes covering 19 application domains are deployed on Sunway-TaihuLight Supercomputer;***

- ***Among which there are 22 full-scale, 40 half-scale, around 100 million-core-scale applications;***

- ***5 Gordon Bell Finalists, and 2 Gordon Bell Prizes.***

HPC & Big Data Applications



Thanks for your listening!

Jingheng Xu
National Supercomputing Center in Wuxi
Department of Computer Science, Tsinghua University
May 15th, 2019

国家超级计算无锡中心
National Supercomputing Center in Wux