Research and Application of SIMD-Based Online Data Processing Acceleration Technology

Pengfei Xiao[1,3], Xiaolu Ji[2,3], Ping Cao[1,2]
[1]University of Science and Technology of China, Hefei
[2]State Key Laboratory of Particle Detection and Electronics, Institute of High Energy Physics, CAS, Beijing
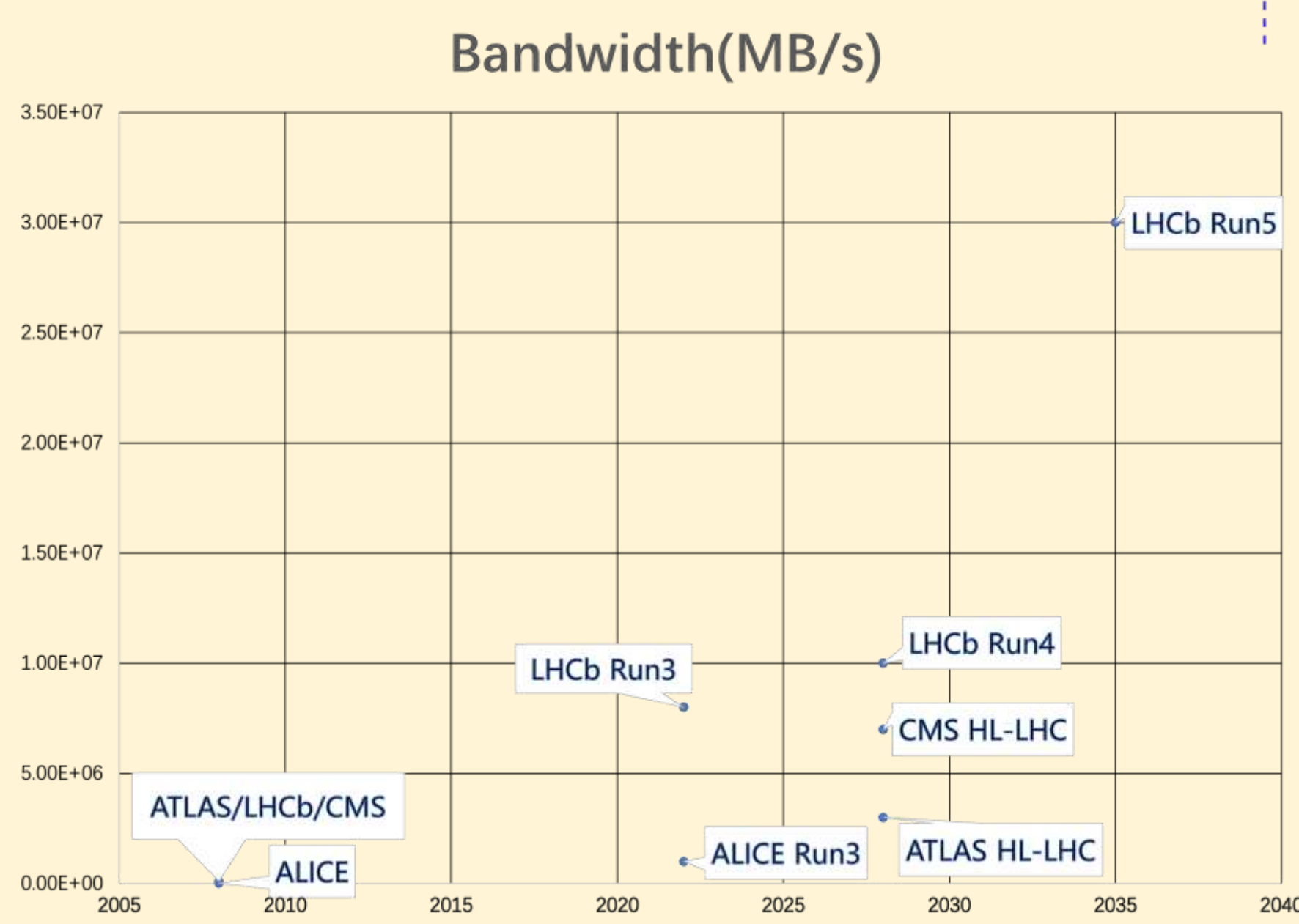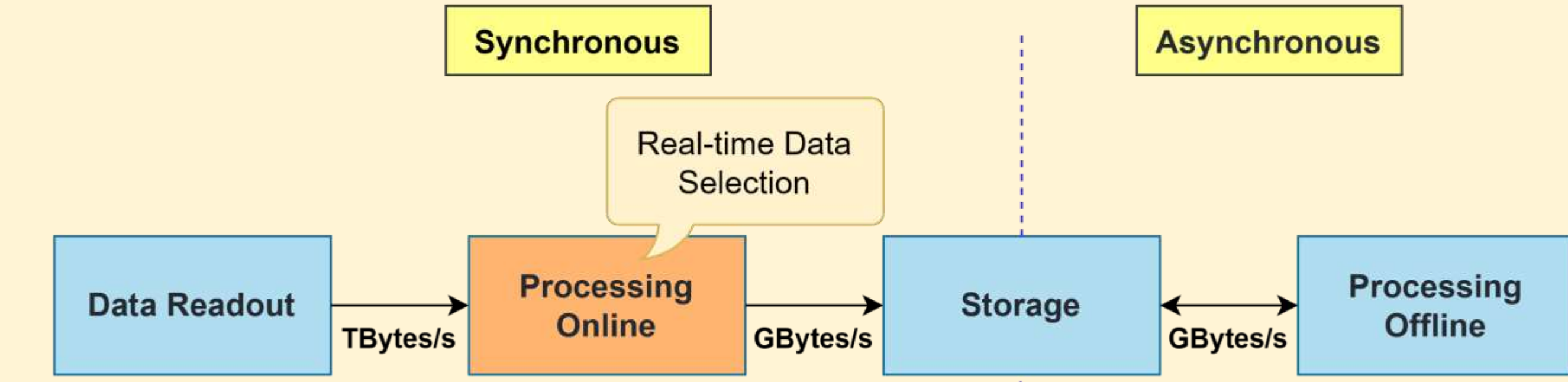[3]Institute of High Energy Physics, CAS, Beijing

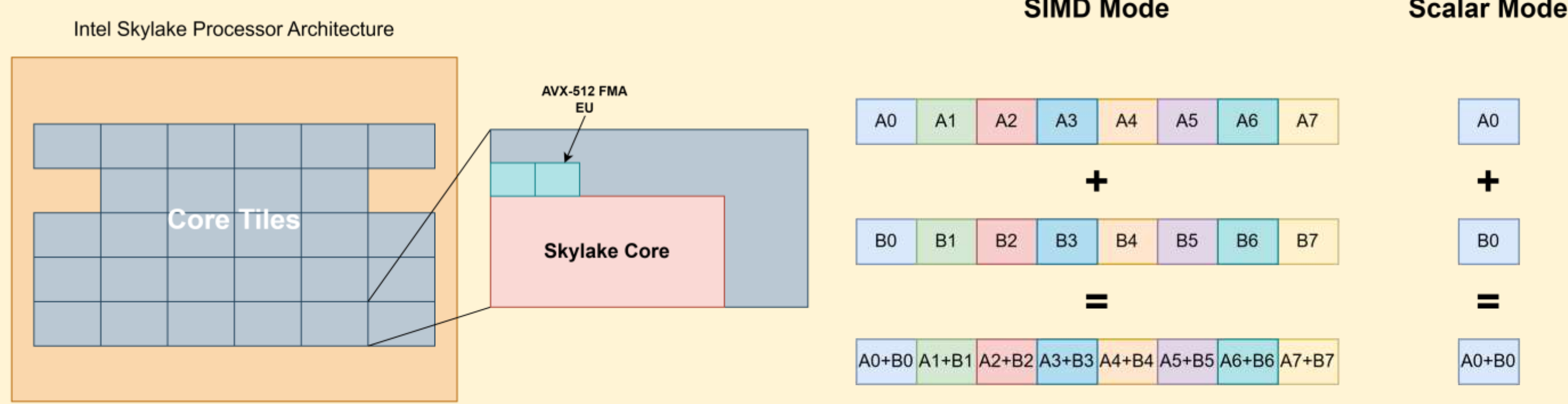Email: pengfeixiao@mail.ustc.edu.cn

MOTIVATION

Online Data Processing Challenge

- The rapid growth of the data bandwidth has brought unprecedented challenges to online data processing.
- The large volume of data necessitates a rapid method for handling them.

INTRODUCTION

CPU Single Instruction Multiple Data

- The selected Intel CPU core is equipped with a dedicated register set for parallel data processing.
- Each register for the Intel AVX-512 instruction set can simultaneously process 8 double-precision or 16 single-precision numbers.

**GPU Single Instruction Multiple Data**
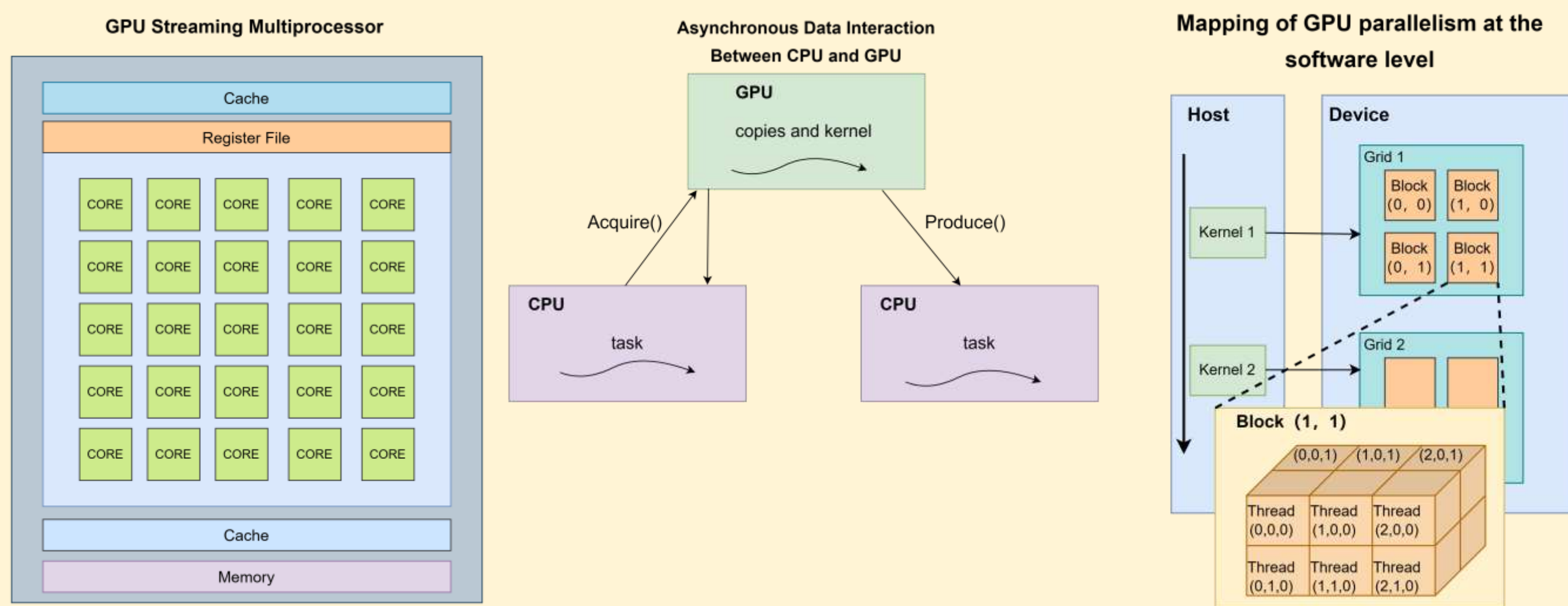
- GPU computing is a variant of SIMD, with more computing cores, achieving high-performance parallel data processing through grid and thread block invocation.
- Typically, heterogeneous computing can be achieved as the CPU asynchronously transfers data with the GPU.

RESEARCH

(OS: Linux CentOS 8 **CPU:** Intel Xeon Platinum 8462Y+ **GPU:** Nvidia A30)

Research On CPU SIMD Applied In Image Coefficient Correction
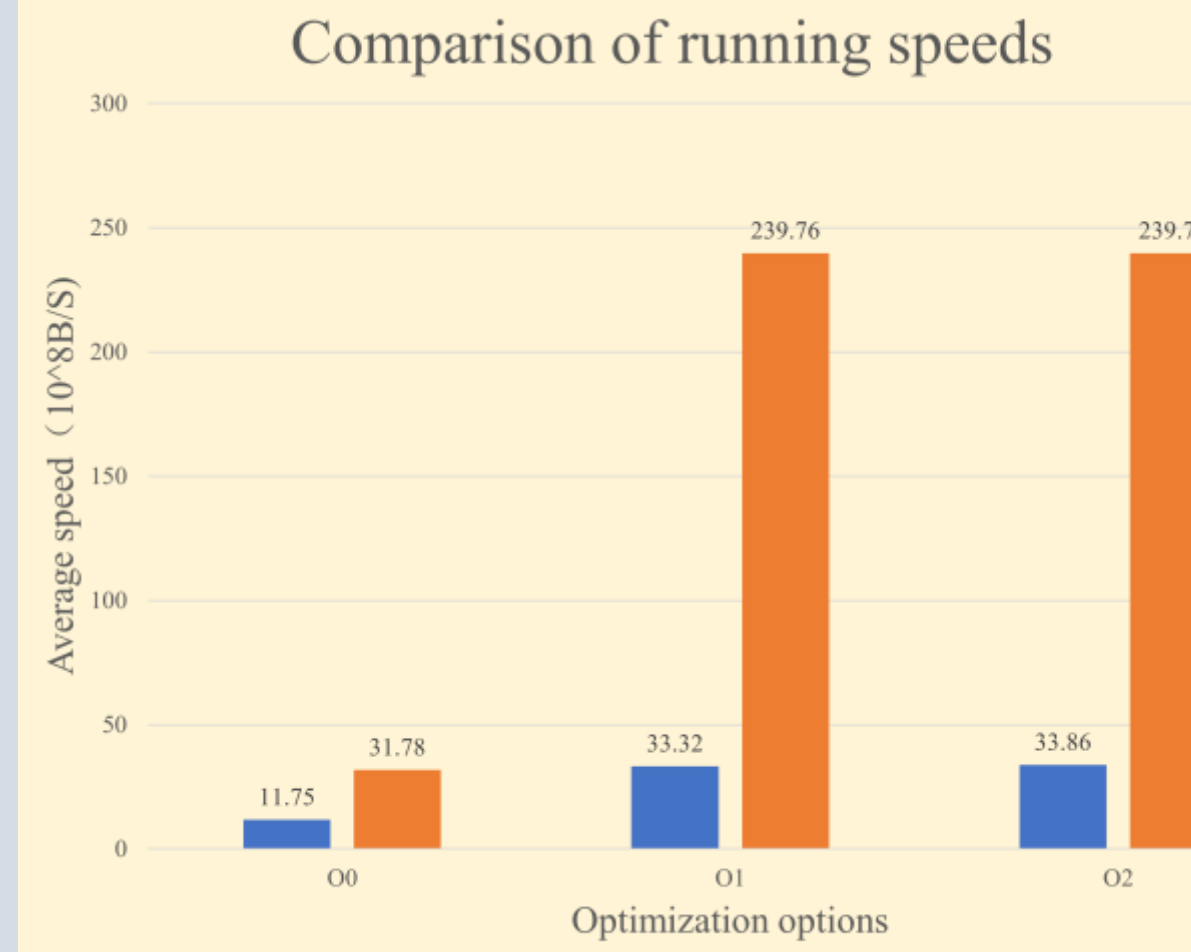
Running With Different Compilation Optimization Options

- Image coefficient correction involves dense independent multiplication calculations, suitable for SIMD processing.
- The tested pixel array size is 256×576.
- In a single-threaded scenario, the performance improvement reaches a maximum of 719.57%.
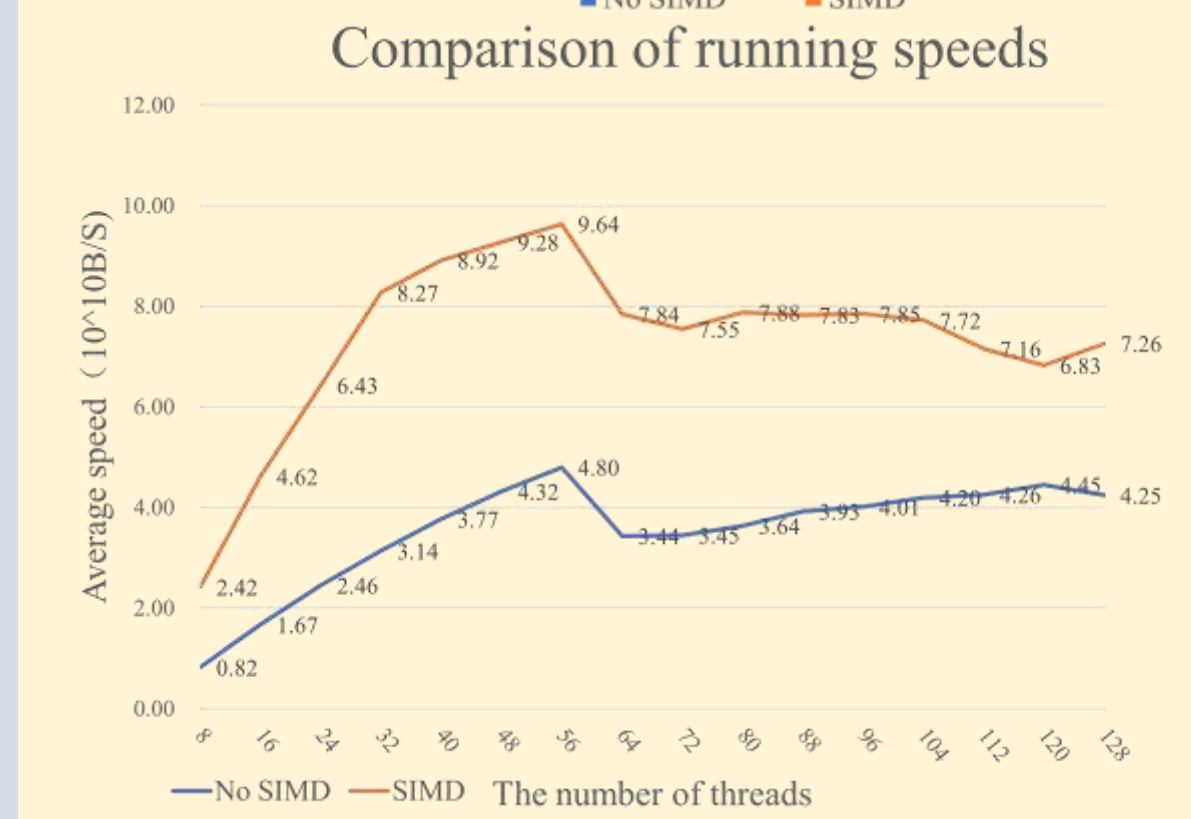
Running on different numbers of threads

- The tested pixel array size is 1000×1000, tested with -O0 compilation optimization option.
- The performance improvement is affected by the number of threads, with the highest improvement reaching 293.68%

**Research On GPU SIMD Applied In Waveform Point Processing**

Waveform Point Processing parallelized on GPU

Inspired by JUNO OEC, the implementation of parallelized waveform point processing on GPU is conducted .
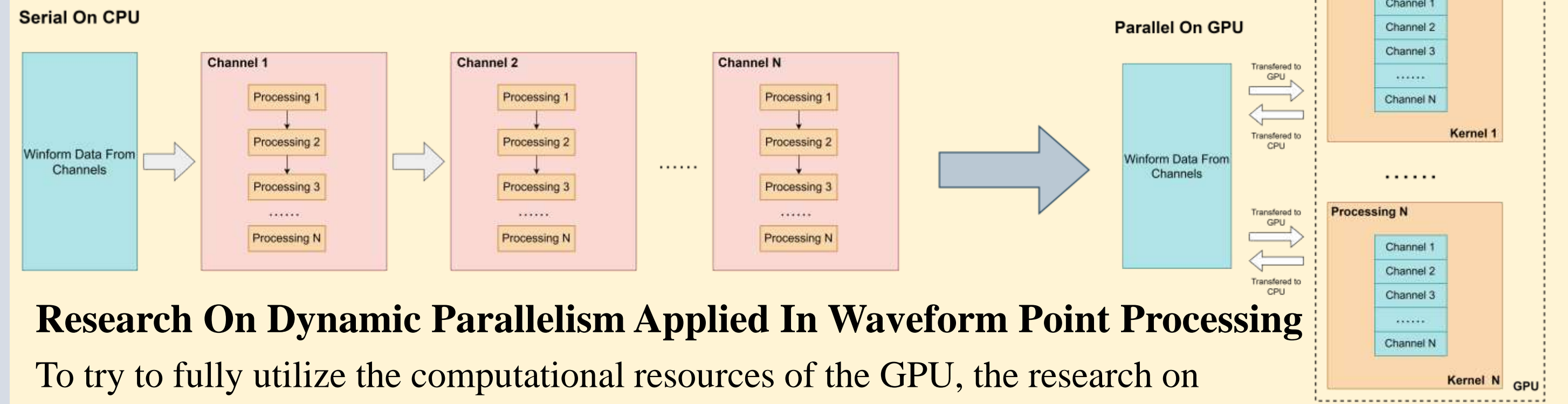
**Research On Dynamic Parallelism Applied In Waveform Point Processing**

To try to fully utilize the computational resources of the GPU, the research on dynamic parallelism is conducted.
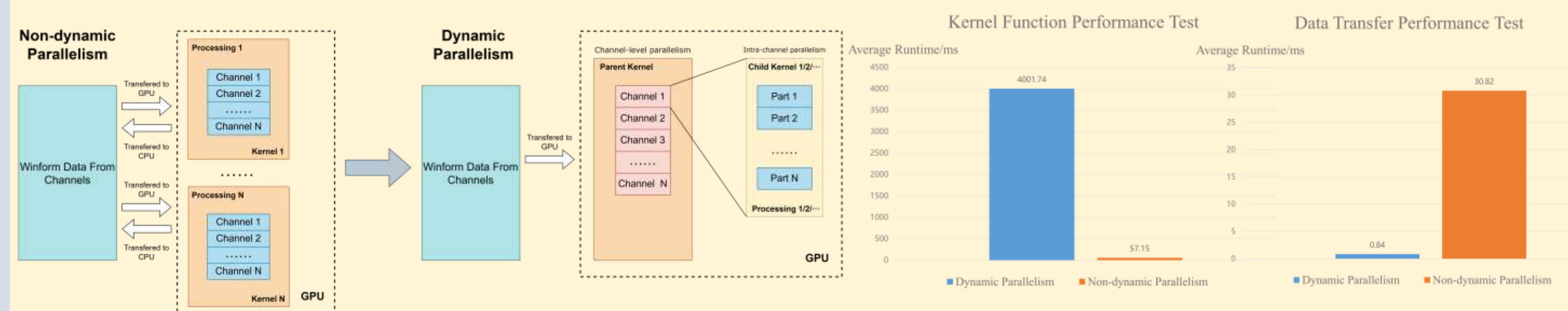
- According to the test result of processing 1800 channels × 1000 points, this method is suitable for scenarios where data transfer time accounts for a relatively large proportion.

IMPLEMENTATION

SIMD Technology Implemented In Silicon Pixel Detector DAQ
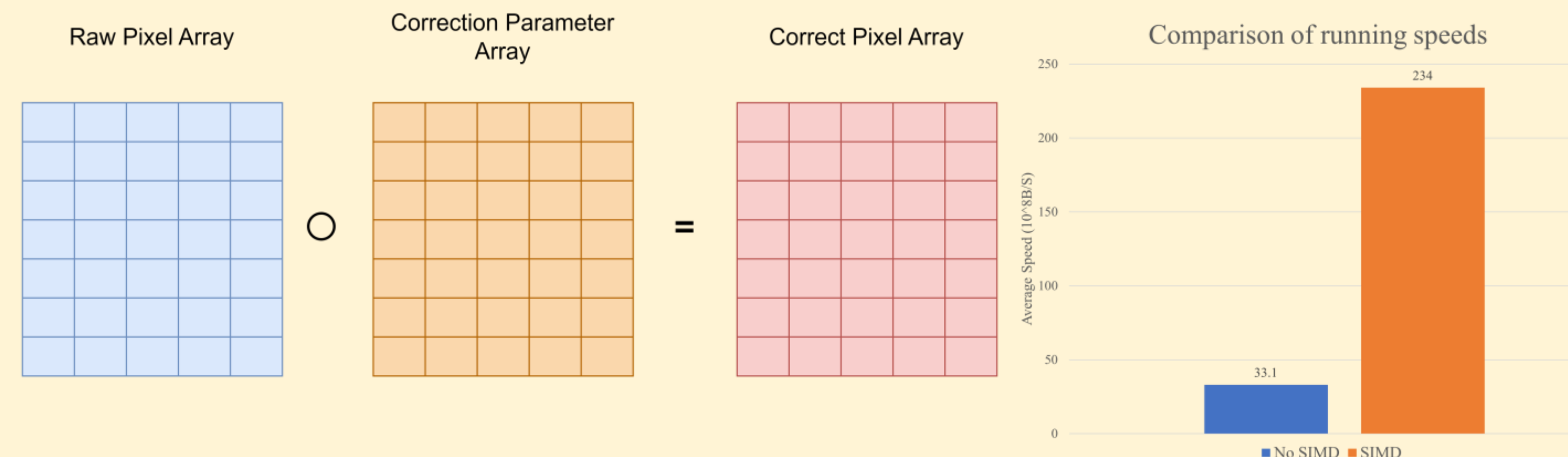
- Combined with the O2 compilation optimization option, the performance of flat field correction processing for 256×576 silicon pixels is improved to 708.09%.

SUMMARY

- **The SIMD (Single Instruction Multiple Data) acceleration technology based on CPU and GPU can effectively accelerate various online data processing scenarios in high-energy physics experiments, and has great potential for practical application.**
- **The subsequent plan is to further explore the potential application of SIMD acceleration in online data processing, in order to improve the real-time performance and throughput of online data processing, and provide effective solutions for future large-scale high-energy physics experiments.**