交通部中央氣象署
Central Weather Administration

Comparative Performance Analysis of Global Ensemble model Verification on CPU and GPU Platforms (Team: CWA GVER)

Team Members: 王志嘉(WANG,CHIH-CHIA)、蘇胤瑞(SU,YIN-RUEI)

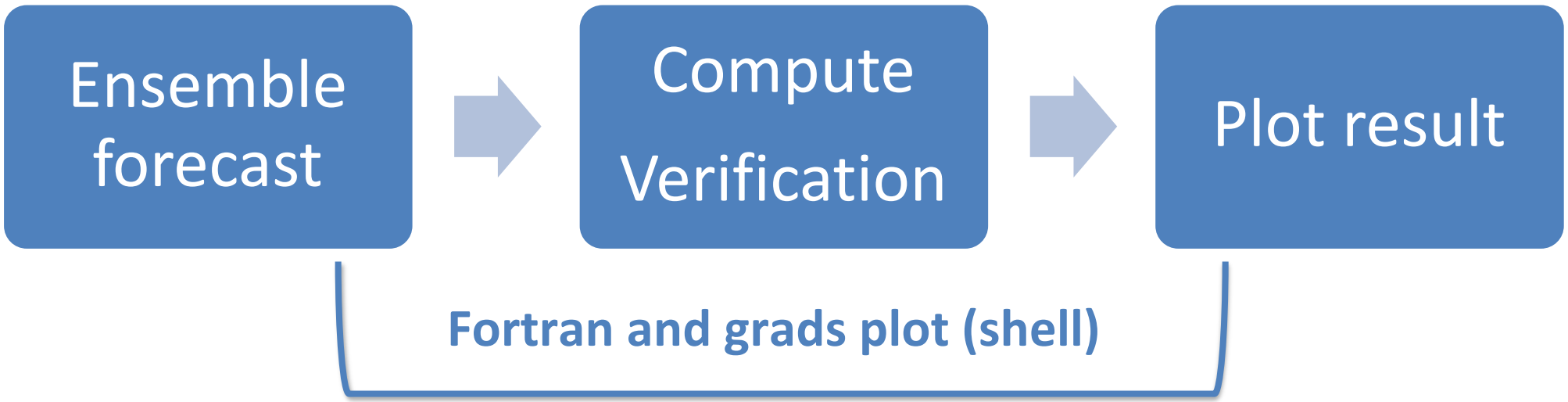
Mentors: Ming Huang、Jay Chen、Kuan-Ting Yeh

NARLabs 國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing

OPEN HACKATHONS

Introduction

Data dimension: (30, 10, 71, 73,144) single model



Total cost time: Exceed12 hours

Introducing GPU acceleration

Evolution and Strategy

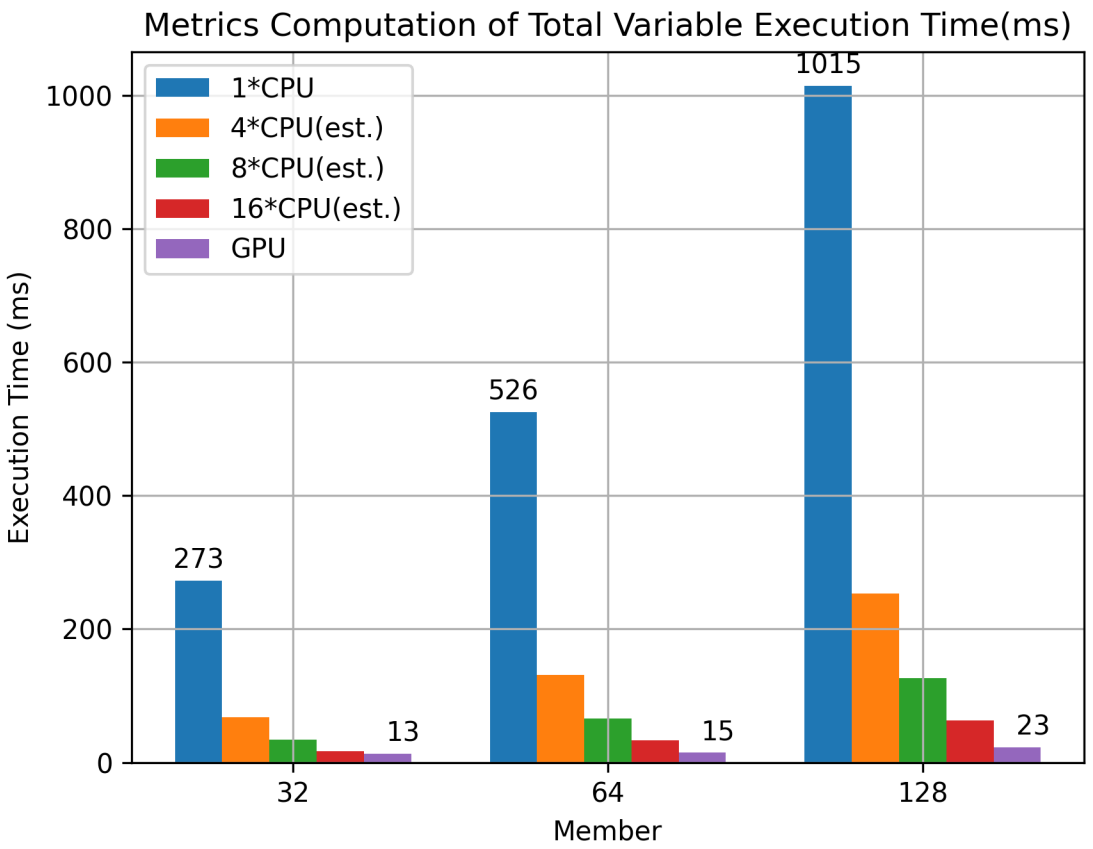
- Fortran code compute → python code compute
- Python Jax JIT Compilation improve performance
- Compare CPU and GPU runtime

Experiment design

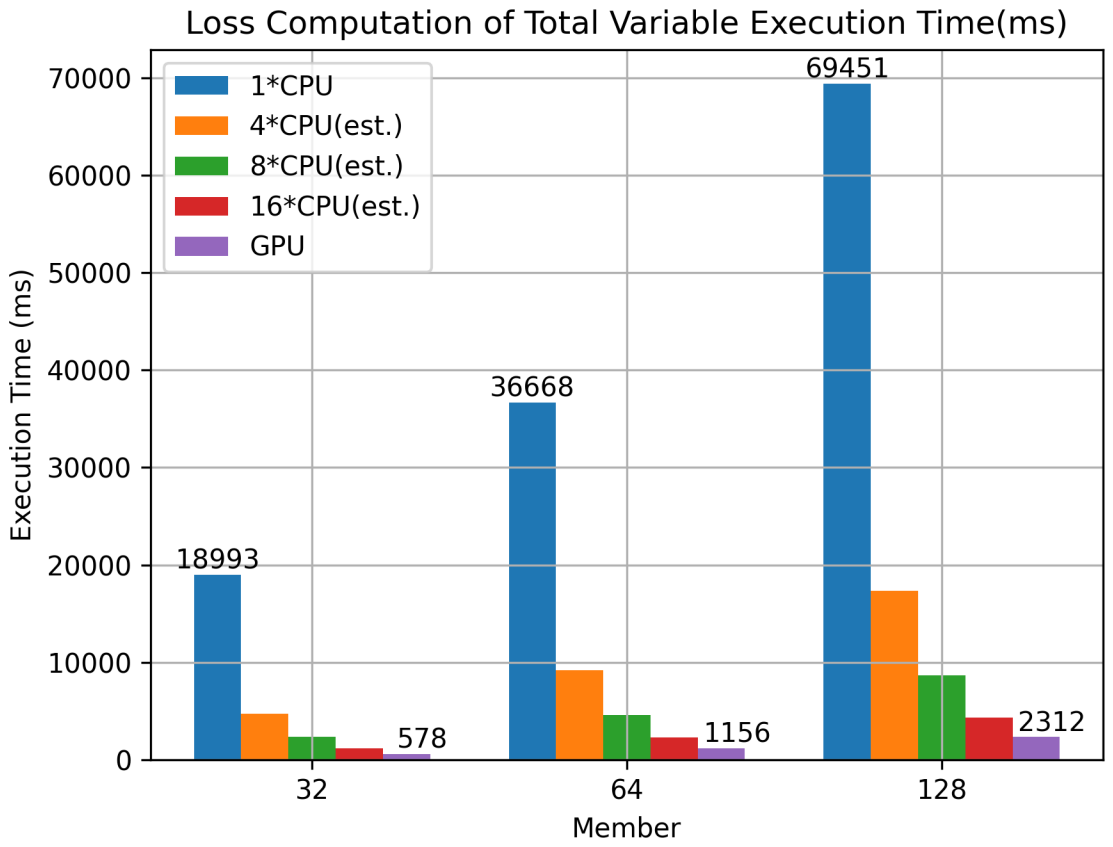
Compute Metrics	Compute Loss
RMSE	RPSS
MAE	CRPS
SPREAD	CRPSS
BIAS	BS
	BSS
	HISTOGRAM

Experiment Efficiency

Metrics

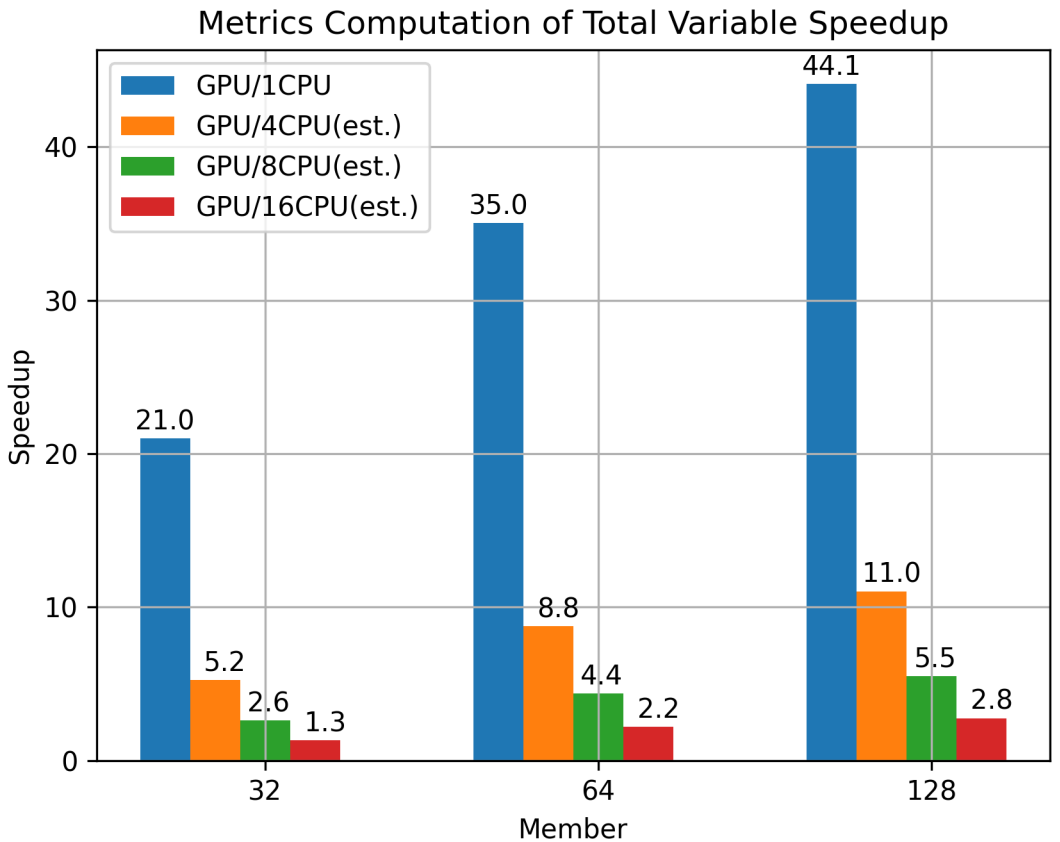


Loss

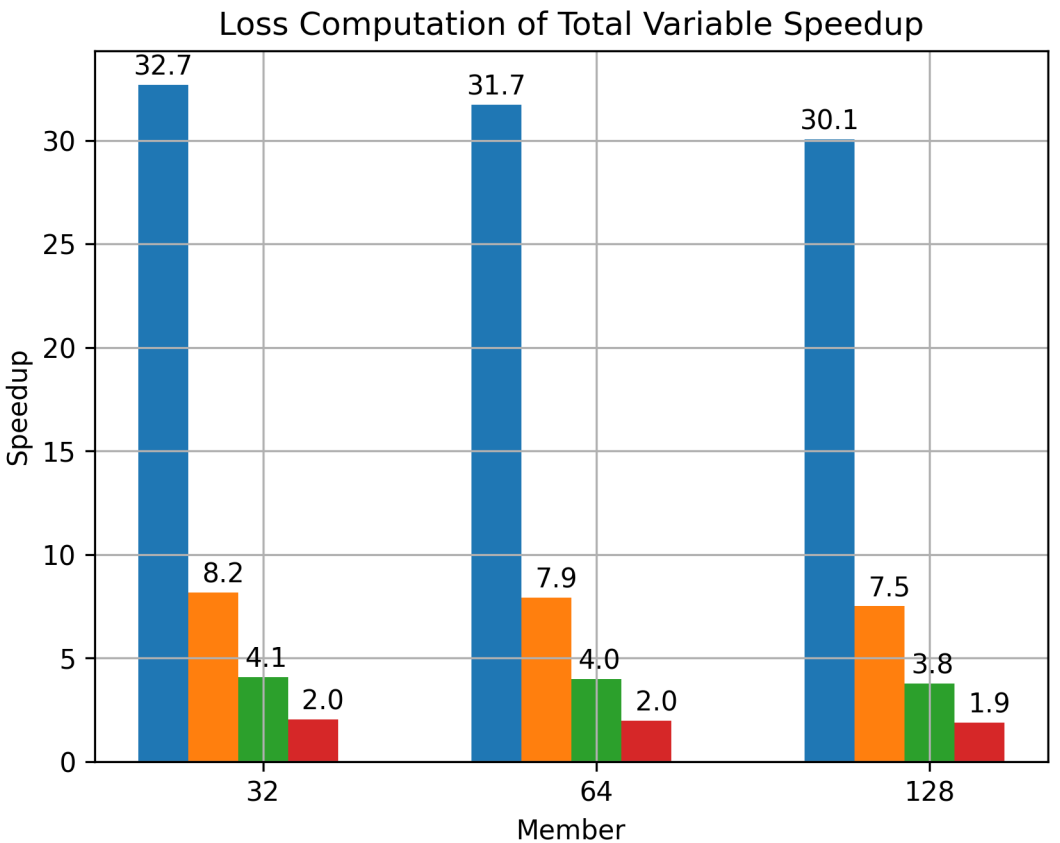


Experiment Efficiency(Speedup)

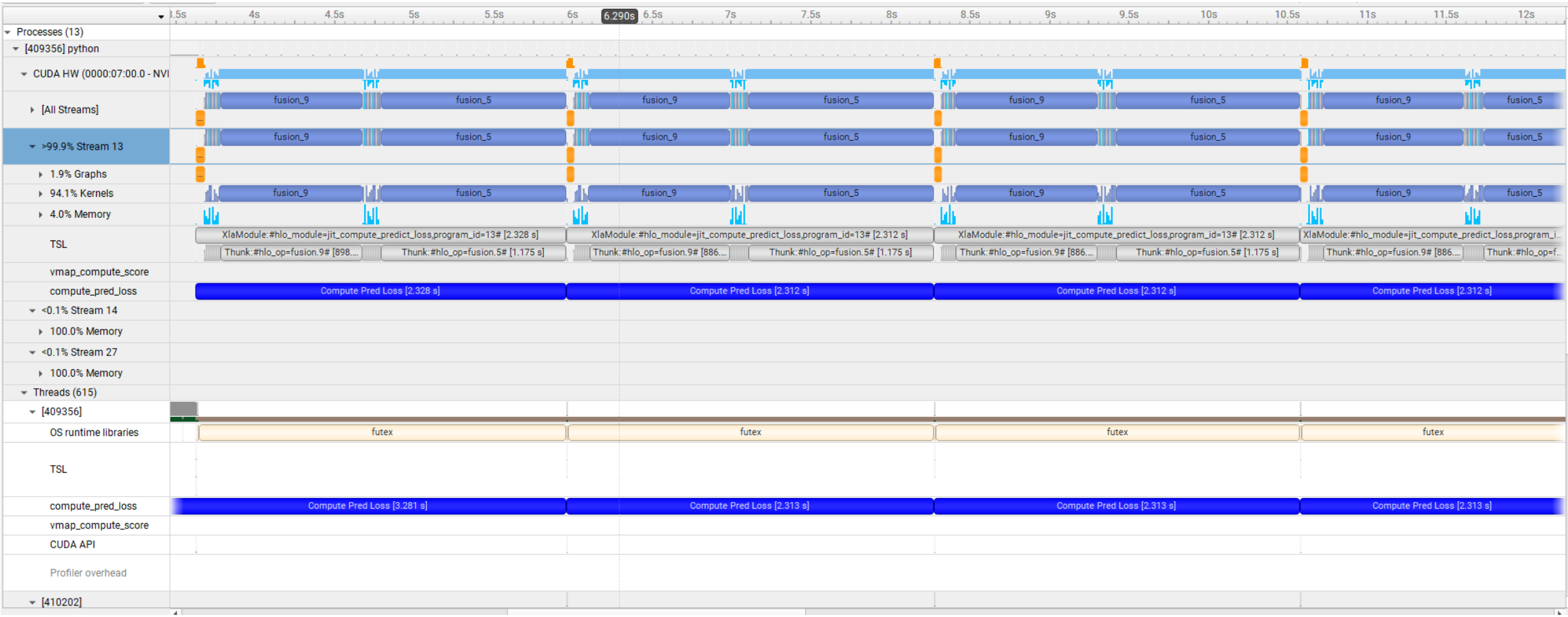
Metrics



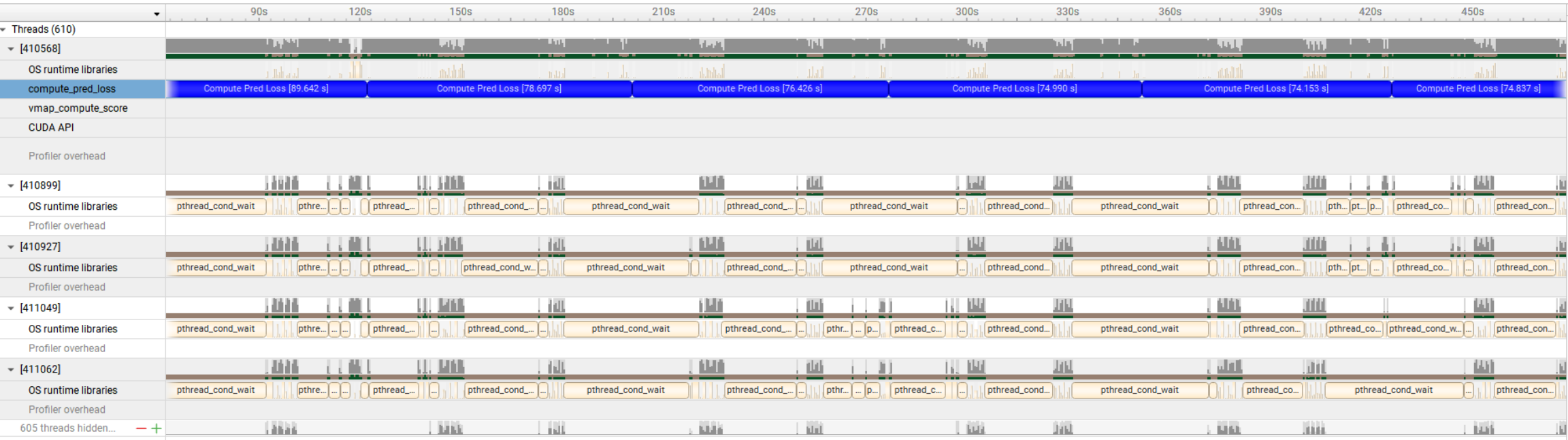
Loss



Nsight 128 member GPU run



Nsight 128 member CPU run



Summary

Experiment	before/after
compute metrics	21x (32mem), 35x (64mem), 44x (128mem)
compute loss	32.7x (32mem), 31.7x (64mem), 30.1x (128mem)

on 32 members, total times on CPU(19 seconds) to GPU(0.5 seconds)
on 64 members, total times on CPU(37 seconds) to GPU(1.1 seconds)
on 32 members, total times on CPU(70 seconds) to GPU(2.3 seconds)

Future work

- Implementing Additional Metrics
  - Expand the scope of metric calculations to provide a more comprehensive analysis.
  - Develop and integrate additional metrics to enhance the evaluation process.
- Introducing I/O Operations
  - The current status of IO acceleration is still under investigation
- Introducing CWA Ensemble Verification

Summarize achievements during this Hackathon

1. Successfully applied GPU acceleration for computing scores, improving overall computational efficiency.
2. Acquired proficiency in utilizing Nsight for performance evaluation.
3. Applying NVTX allows for detailed performance analysis by marking specific sections in CUDA code with NVTX ranges.
4. Mastered running GPU-accelerated code with PyTorch and leveraged JAX for efficient GPU computation, contributing to optimized code.
5. Finally, thank you to Ming, Jay, and Kuan-Ting for their guidance.

Github: https://github.com/yinruei/cwb_gver/tree/master
Contact: yinruei@cwa.gov.tw