

The background features a dark blue gradient with numerous thin, glowing lines in shades of cyan and magenta. These lines are interspersed with small, bright dots, creating a sense of motion and digital connectivity.

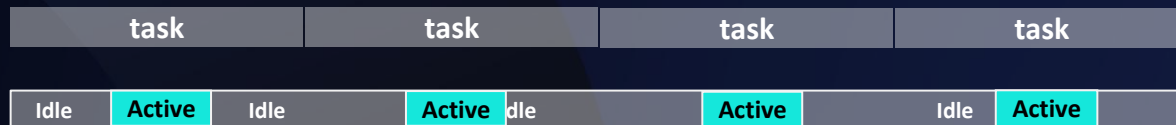
**Make your GPUs work harder:
AI computing broker**

Powerful GPUs, Underutilized

Most apps have GPU idle time

Conventional

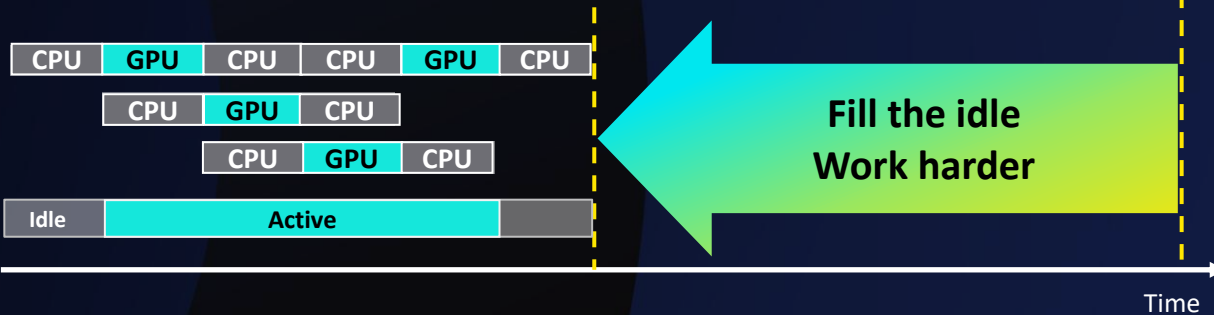
GPU util.



AI Computing Broker fills the idle gaps



GPU util.



ACB Demo ①

AlphaFold2 : AI application for predicts the 3D structure of a protein

Conventional

AF2 running on 2 GPUs



AI Computing Broker

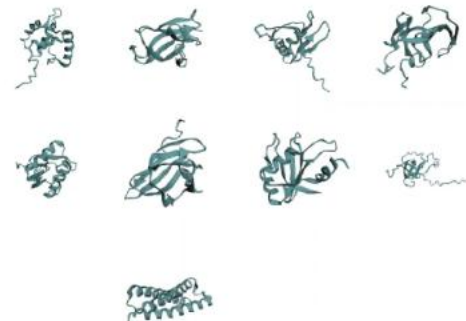
AF2 running on 1 GPU — same performance



2 GPUs



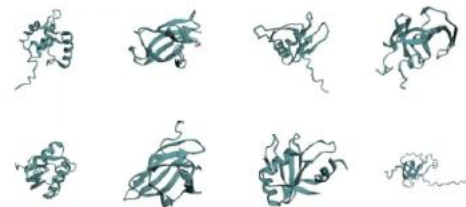
Throughput [proteins/min]: 9.3



1 GPU + AI Computing Broker



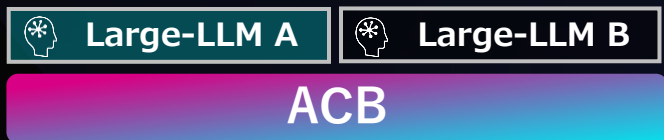
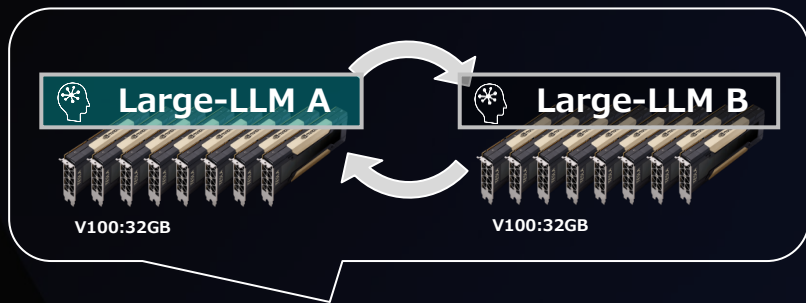
Throughput [proteins/min]: 9.1



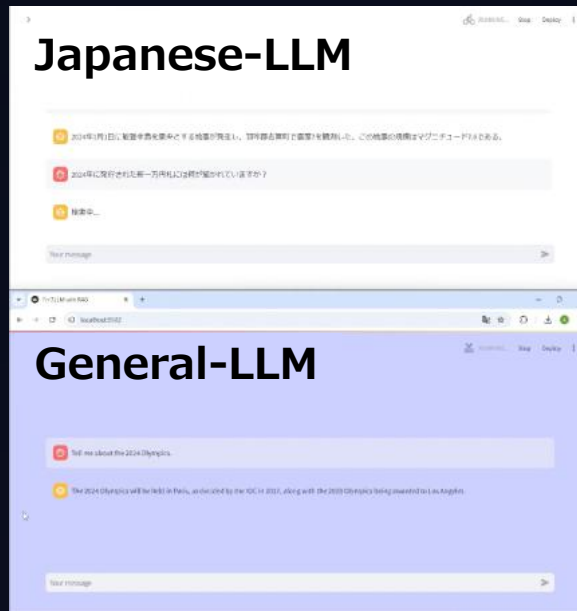
ACB Demo ②

LLM : Large Language Model inference across multiple GPUs

1 LLM on a server with 8 GPUs



Multiple LLM can run on a server



The background features a dark blue gradient with numerous thin, wavy lines in shades of cyan and magenta. These lines are punctuated with small, glowing dots of the same colors, creating a sense of dynamic energy and data flow.

ACB — Real-world Impact

Improving GPU AI workload efficiency and reducing computational costs

- Now running on Raplase RTX PRO 6000 Blackwell Server Edition instances -

Xtreme-D's supercomputing-based system tuning + ACB:
more concurrent jobs, higher throughput,
no loss in training performance

Business Issues

Low GPU utilization & memory limits bottleneck workloads

Solution

ACB optimizes GPU allocation, complementing the existing scheduler

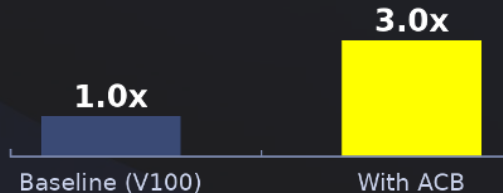
Impact

Over 3x throughput on the same GPU (based on Xtreme-D's benchmark)

GPU Jobs Throughput per unit time

Raplase

3x or more



About Xtreme-D (Xtreme-D inc. <https://xtreme-d.net/>)

Xtreme-D offers a multi-cloud compatible and high-speed AI platform service Raplase(Ra+)

Your GPUs work hard.
Could they work smarter?

Ask us for a demo or PoC with ACB