Machine Learning Systems

LLM Post Training

Siyuan Feng

Shanghai Innovation Institute

01

RL Post Training for LLMs

PPO and GRPO Workflow

1. Rollout
2. Inference
3. Calc Advantage
4. Train Models
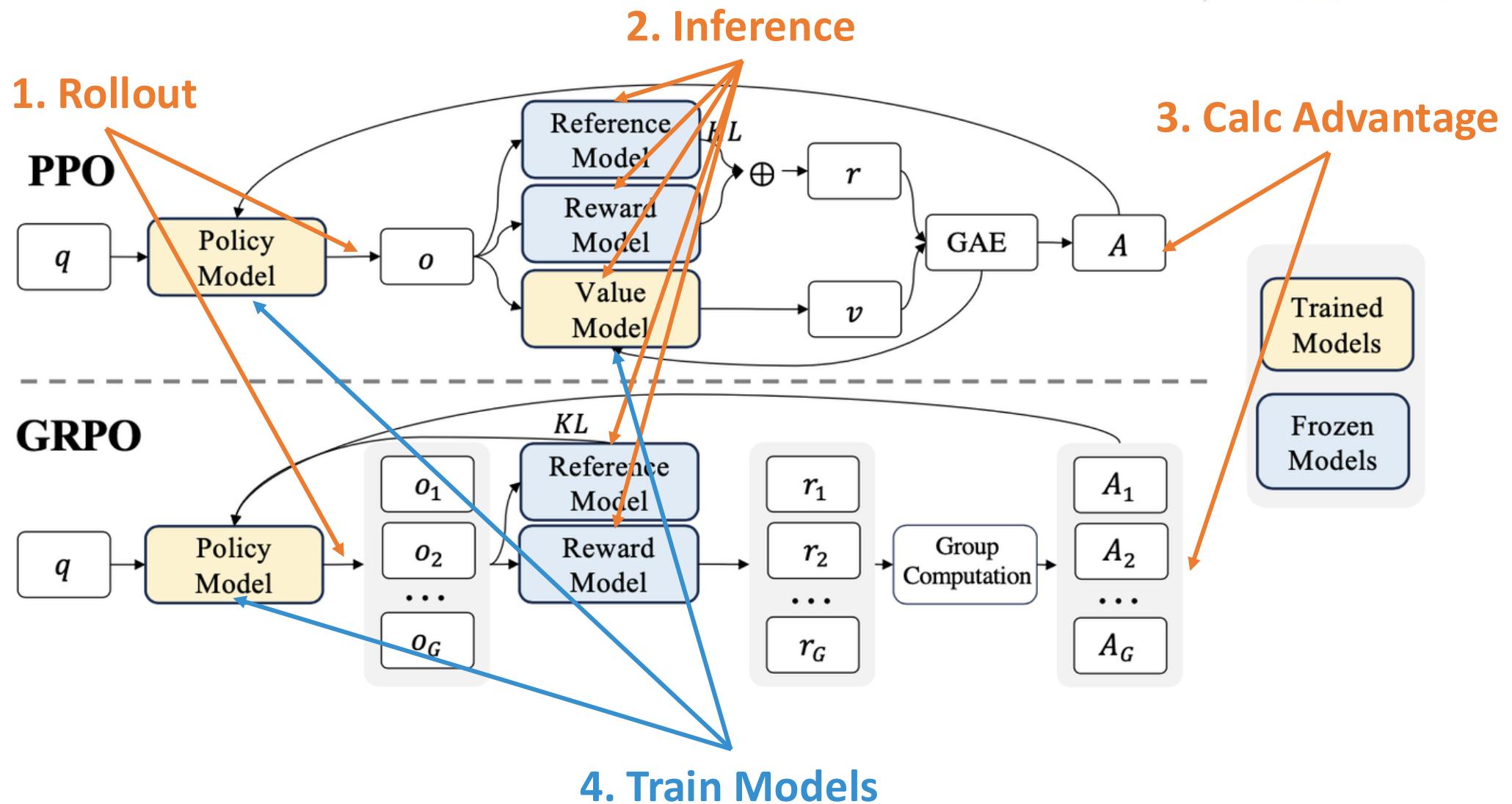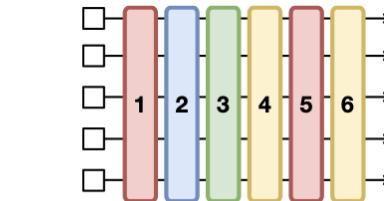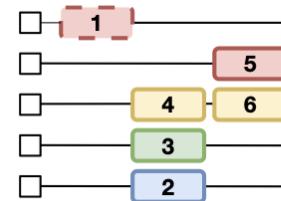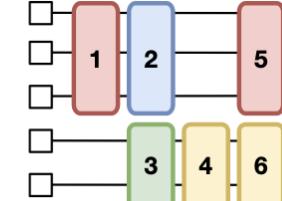
Dataflow of Different RL Algorithm

(a) PPO

(b) Safe-RLHF

(c) ReMax

Sheng, Guangming, et al. "Hybridflow: A flexible and efficient rlhf framework." *Proceedings of the Twentieth European Conference on Computer Systems*. 2025.
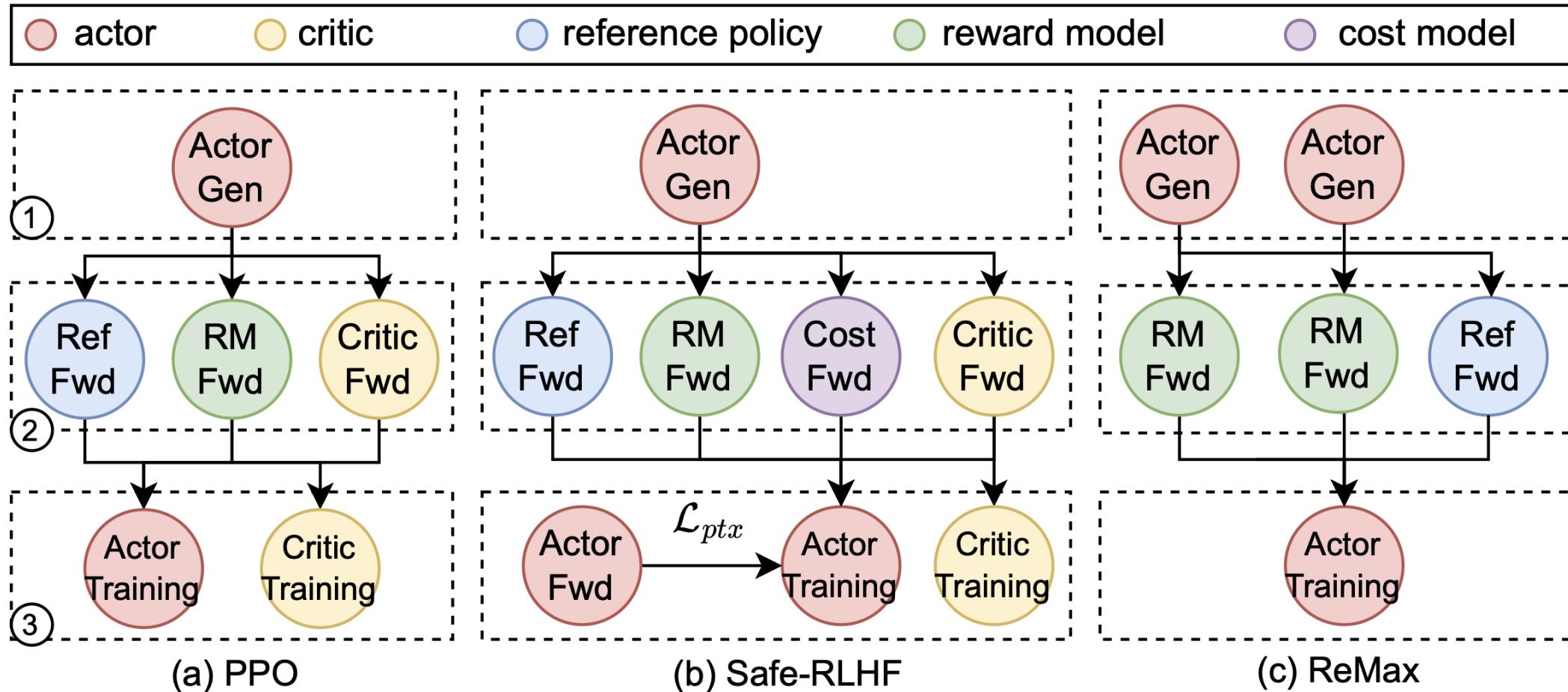
Naïve Model Placement Plan

Sheng, Guangming, et al. "Hybridflow: A flexible and efficient rlhf framework." *Proceedings of the Twentieth European Conference on Computer Systems*. 2025.
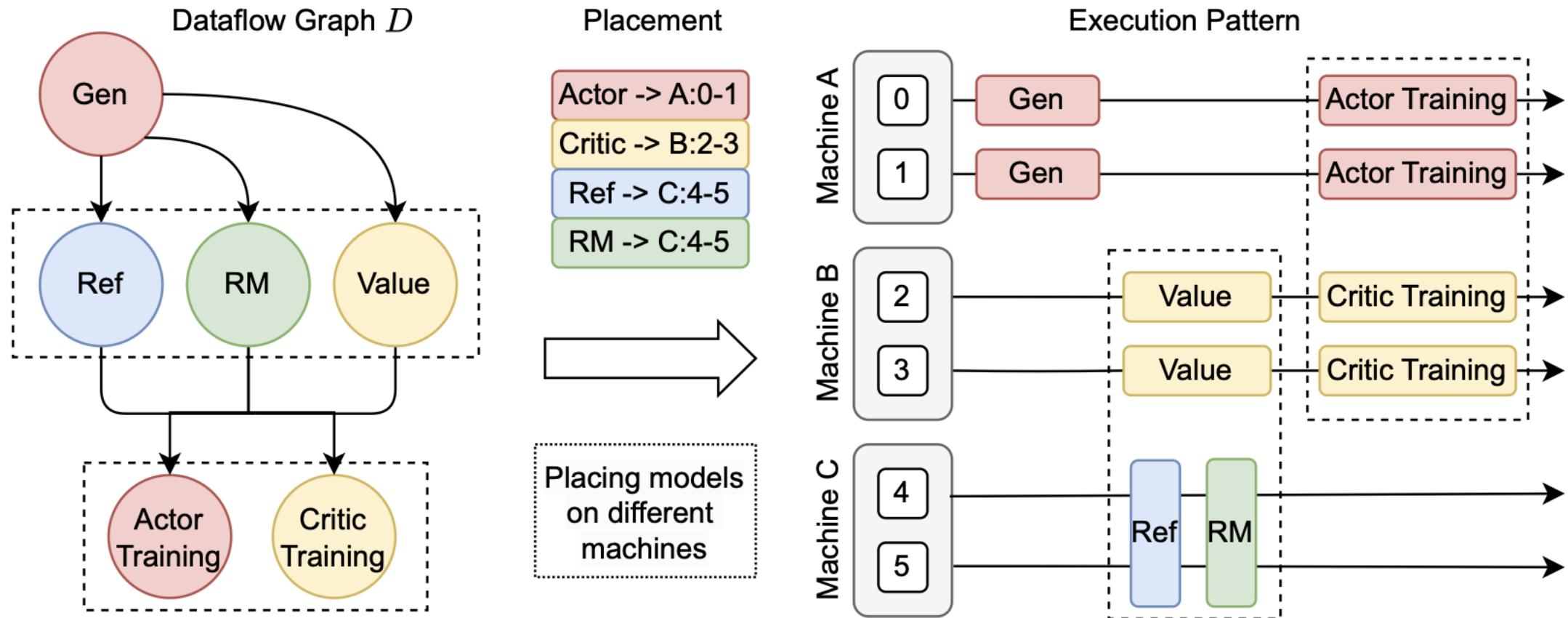
Different Co-located Model Placement

RLHF system	DeepSpeed-Chat	OpenRLHF	NeMo-Aligner
Parallelism	Training: ZeRO Generation:TP	Training: ZeRO Generation:TP	3D Parallelism for both training and generation
Actor weights in training & generation	Model resharding from ZeRO to TP	Using two copies of actor weights for the two stages	Using identical model partition in two stages (shared weights)
Model Placement	Colocate all models on the same set of devices	Each model placed on separate devices	Actor/Ref colocated on some GPUs Critic/RM colocated on other GPUs
Execution Pattern			

Sheng, Guangming, et al. "Hybridflow: A flexible and efficient rlhf framework." *Proceedings of the Twentieth European Conference on Computer Systems*. 2025.

3D-HybridEngine Workflow in veRL
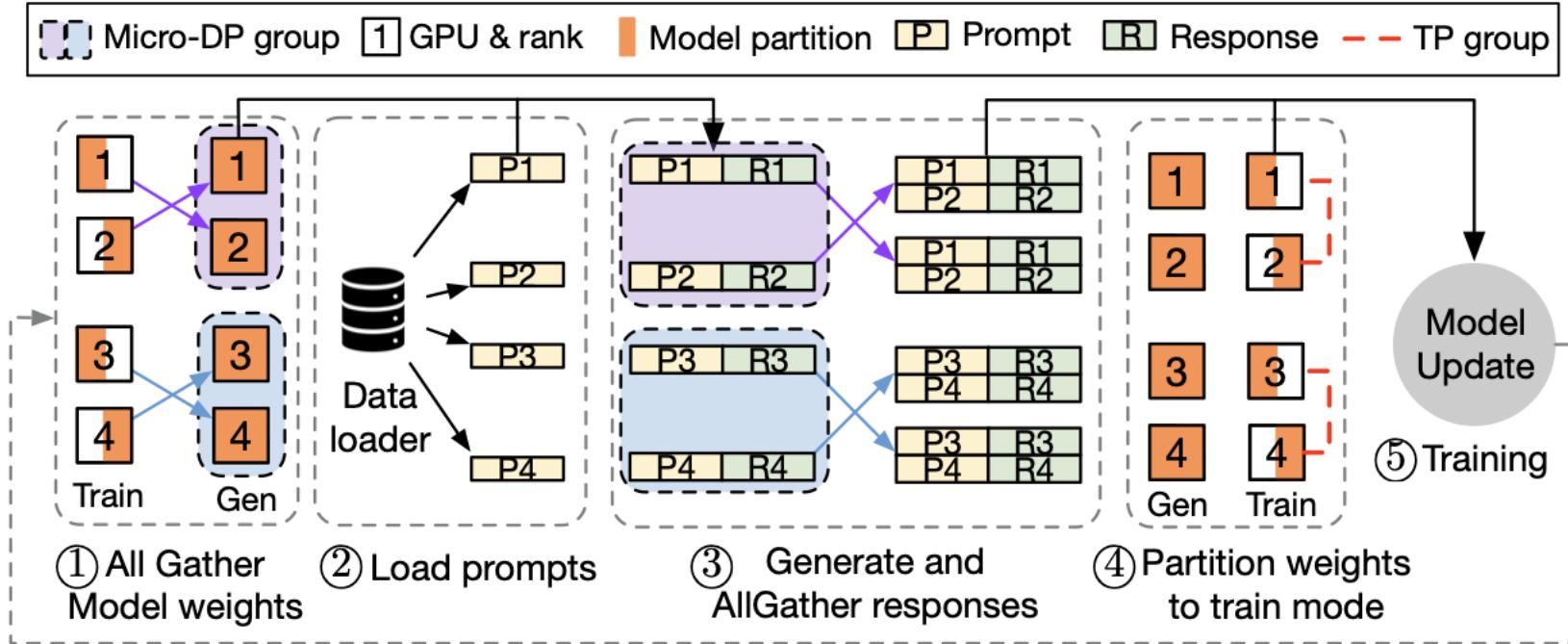
1. Utilize GPU in every timestamp
2. Enable flexible parallelism between training and rollout

Sheng, Guangming, et al. "Hybridflow: A flexible and efficient rlhf framework." *Proceedings of the Twentieth European Conference on Computer Systems*. 2025.

Model Weights Sharding

Optimized for communication

(a) Same grouping methods between training and generation (HybridFlow-V)

(b) Optimized parallel grouping methods (HybridFlow)

Sheng, Guangming, et al. "Hybridflow: A flexible and efficient rlhf framework." *Proceedings of the Twentieth European Conference on Computer Systems*. 2025.

9
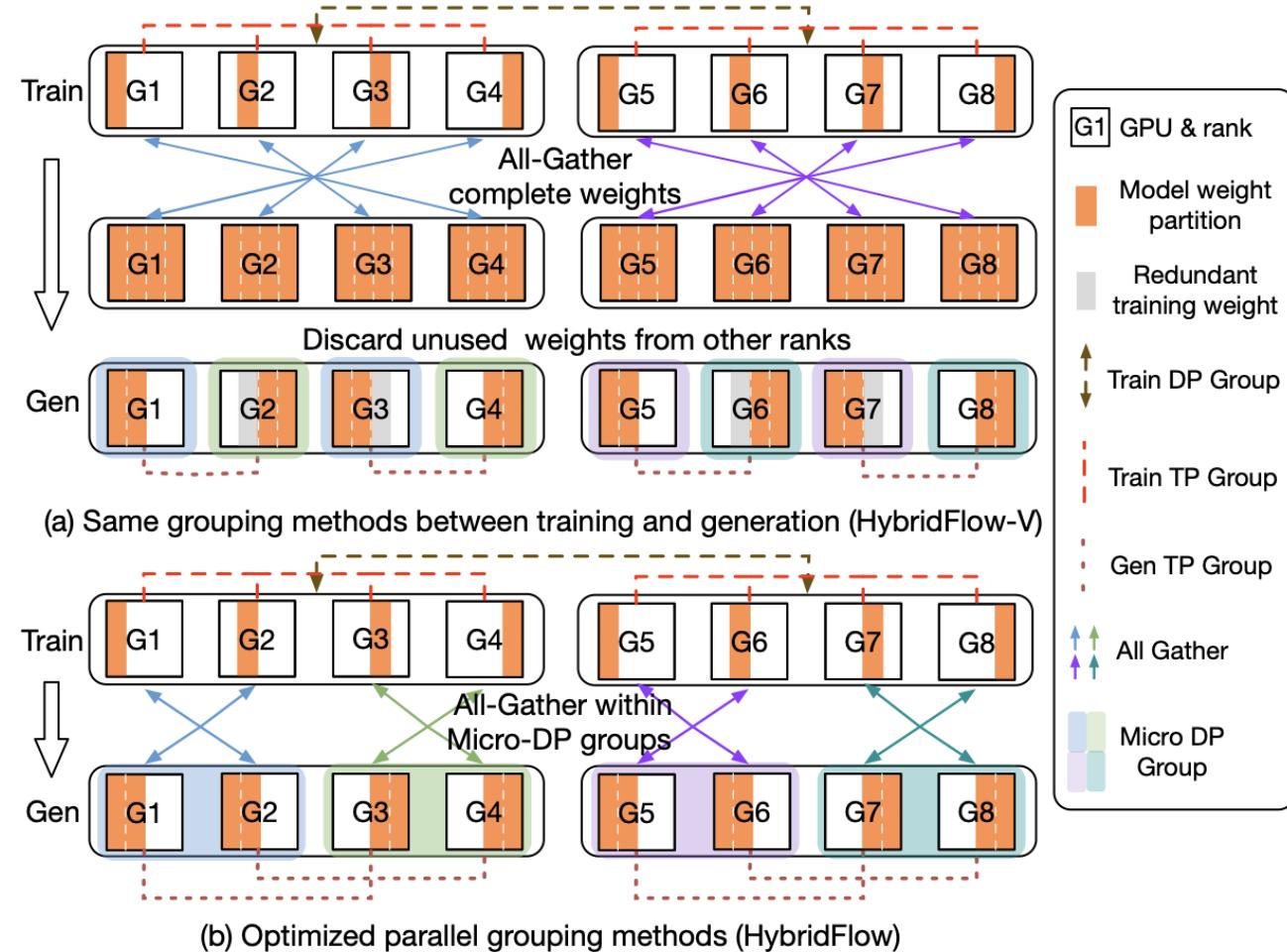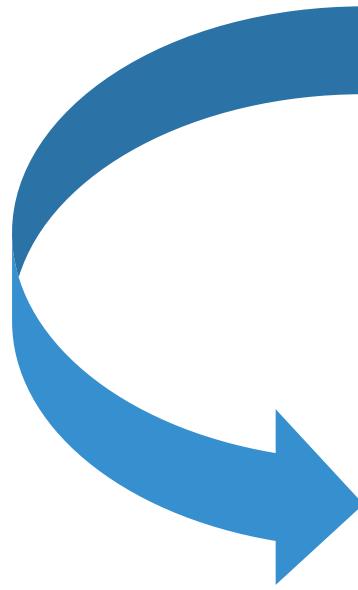
02

Asynchronous RL for LLMs

Synchronous vs. One-Step Asynchronous

Fu, Wei, et al. "AReaL: A Large-Scale Asynchronous Reinforcement Learning System for Language Reasoning." arXiv preprint arXiv:2505.24298 (2025).

Synchronous vs. One-Step Asynchronous

Fu, Wei, et al. "AReaL: A Large-Scale Asynchronous Reinforcement Learning System for Language Reasoning." arXiv preprint arXiv:2505.24298 (2025).
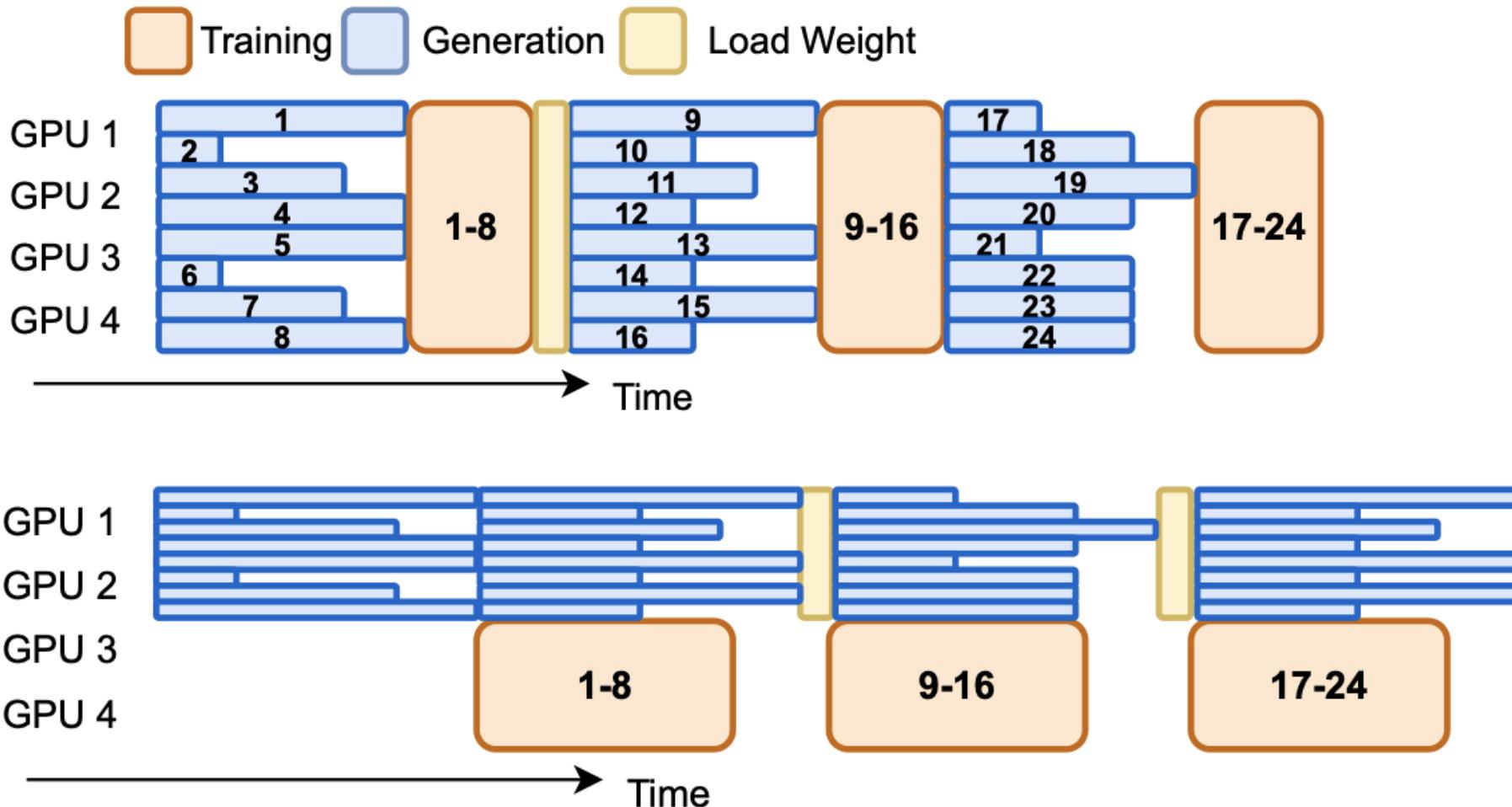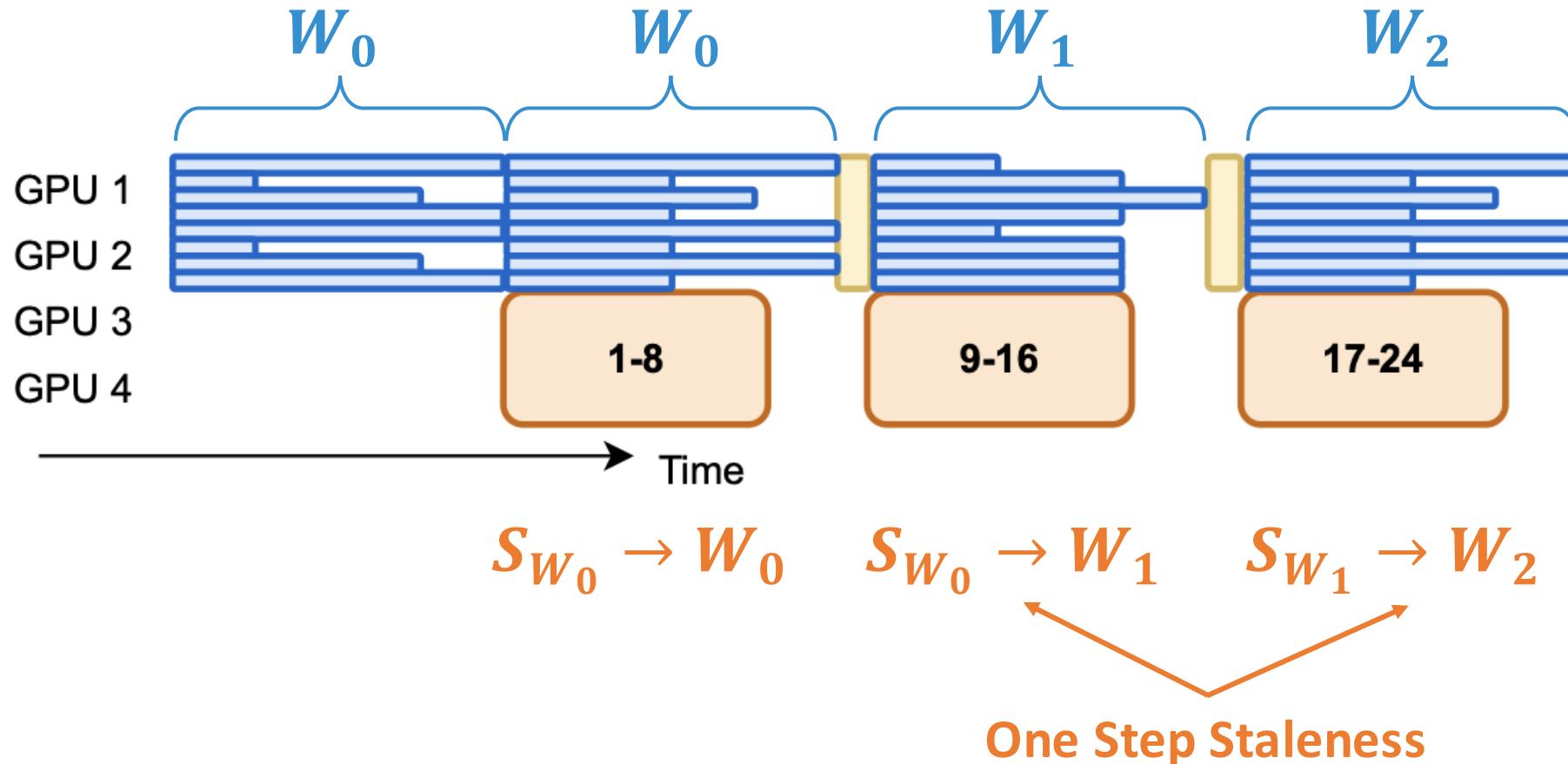
Why Asynchronous?

- Different parallelism world size for training and rollout

- Heterogeneous compute

- **Long, variable-length rollout trajectories**

Partial Rollout

Fu, Wei, et al. "AReaL: A Large-Scale Asynchronous Reinforcement Learning System for Language Reasoning." arXiv preprint arXiv:2505.24298 (2025).

14
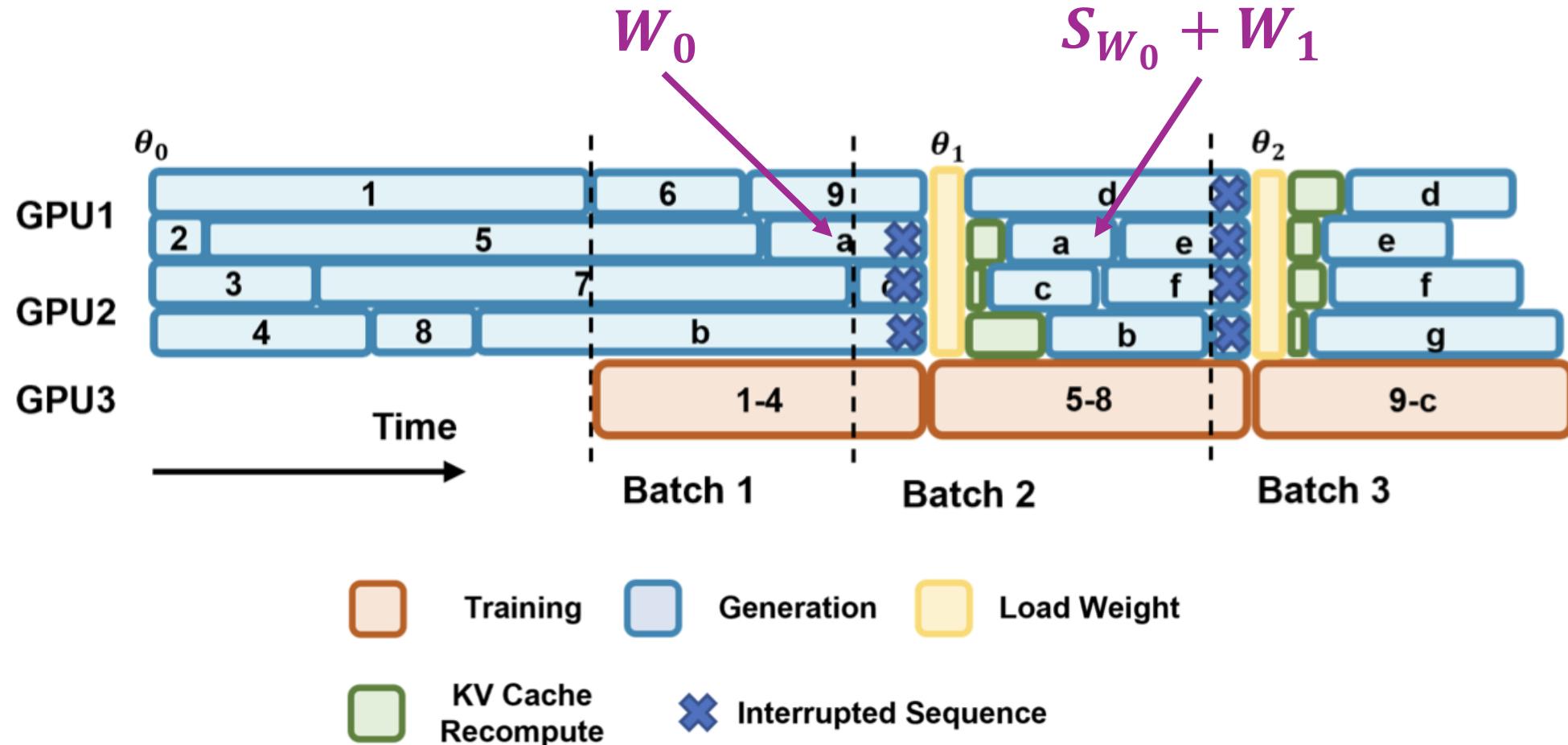
Fully Asynchronous RL
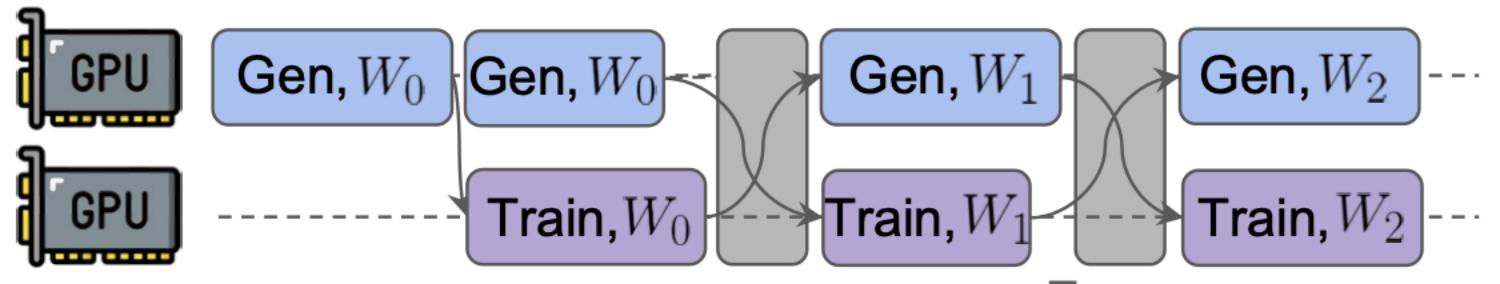
(c) One-step Asynchronous Pipelining

With Streaming
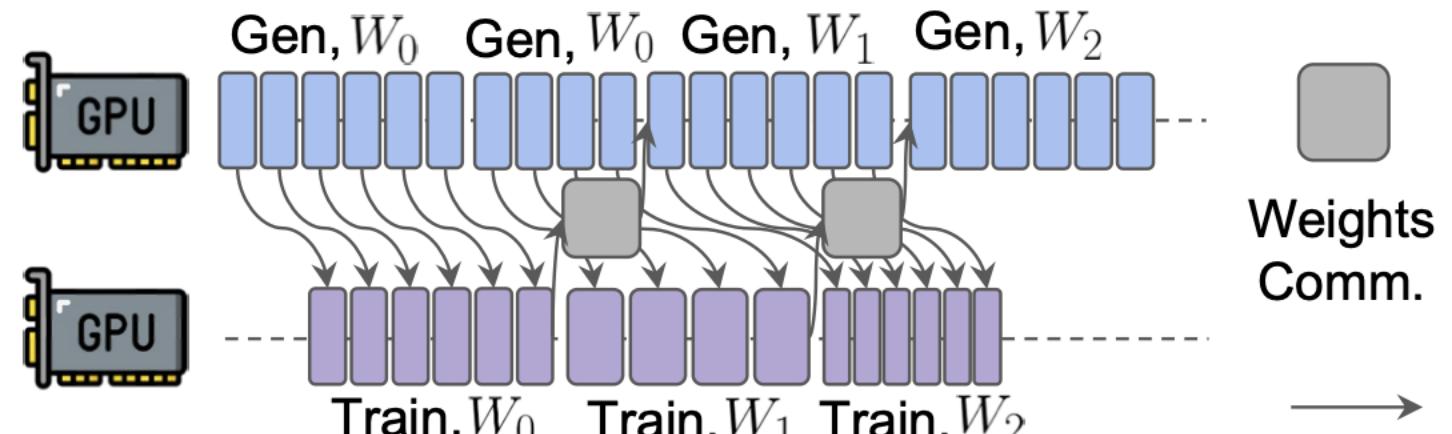
(d) Fully Asynchronous Pipelining

Weights Comm.

Dependency

Zhong, Yinmin, et al. "StreamRL: Scalable, Heterogeneous, and Elastic RL for LLMs with Disaggregated Stream Generation." arXiv preprint arXiv:2504.15930 (2025).

15

Acknowledgement

The development of this course, including its structure, content, and accompanying presentation slides, has been significantly influenced and inspired by the excellent work of instructors and institutions who have shared their materials openly. We wish to extend our sincere acknowledgement and gratitude to the following courses, which served as invaluable references and a source of pedagogical inspiration:

- Machine Learning Systems[15-442/15-642], by **Tianqi Chen** and **Zhihao Jia** at **CMU**.

- Advanced Topics in Machine Learning (Systems)[CS6216], by **Yao Lu** at **NUS**

While these materials provided a foundational blueprint and a wealth of insightful examples, all content herein has been adapted, modified, and curated to meet the specific learning objectives of our curriculum. Any errors, omissions, or shortcomings found in these course materials are entirely our own responsibility. We are profoundly grateful for the contributions of the educators listed above, whose dedication to teaching and knowledge-sharing has made the creation of this course possible.

System for Artificial Intelligence

Thanks

Siyuan Feng

Shanghai Innovation Institute