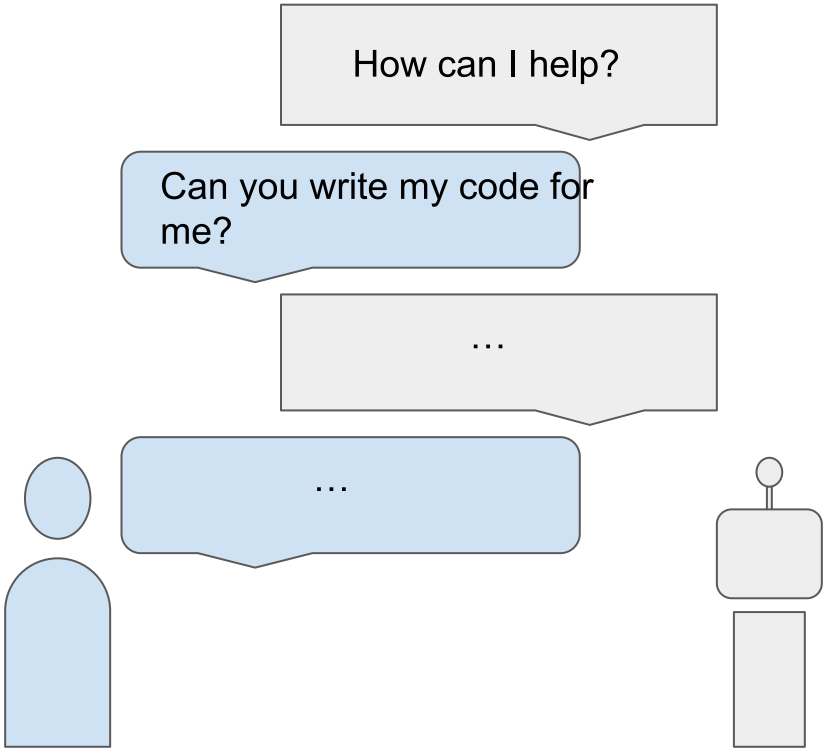
Bamboo

Making Preemptible Instances Resilient for Affordable Training of Large DNNs

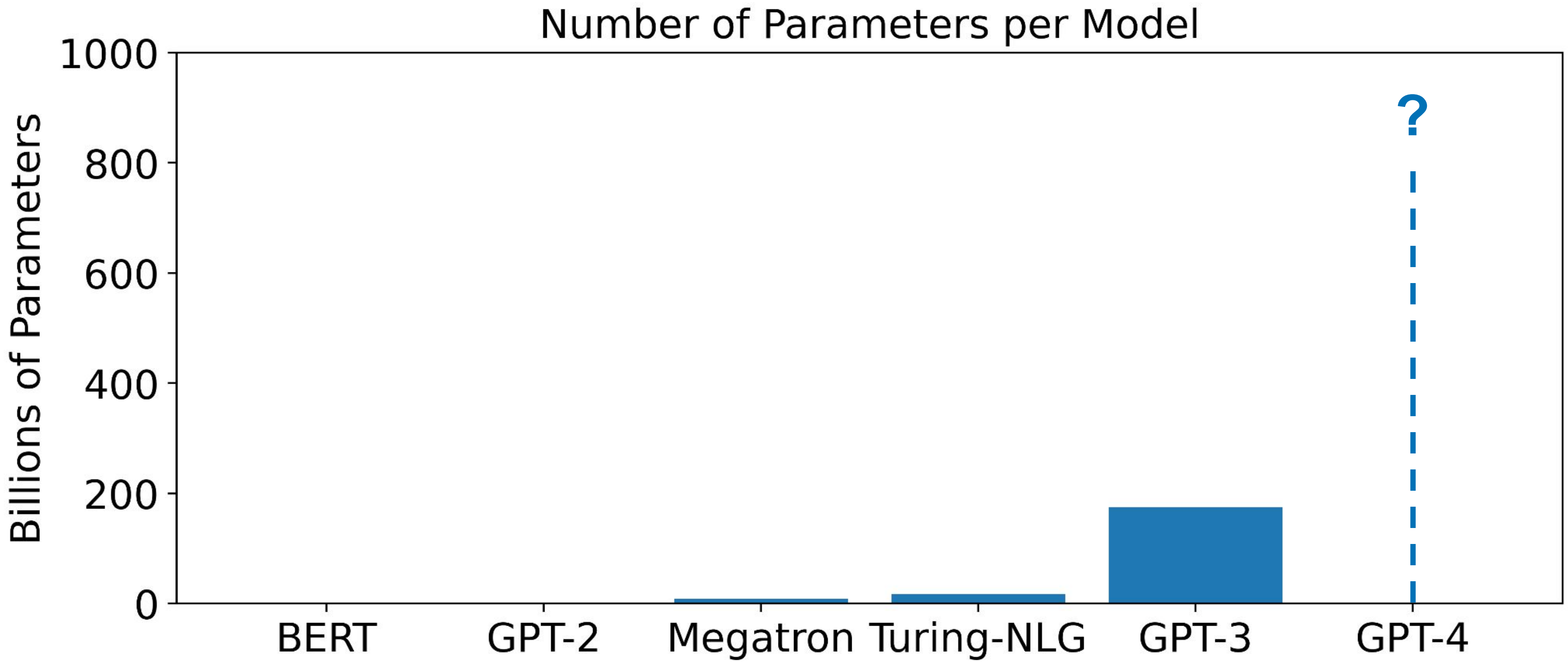
John Thorpe*, Pengzhan Zhao*, Jonathan Eyolfson, Yifan Qiao,
Zhihao Jia, Minjia Zhang, Ravi Netravali, Harry Xu

UCLA CMU Microsoft Research Princeton University

Generative AI Is Changing the World



Model Sizes are Increasing



Prohibitive Costs for Most Organizations

32GB GPU cannot scale past 1.4B parameters

- Many accelerators needed to scale to today's 100B+ parameter models
- $4.6 million to train GPT-3

Model Compression

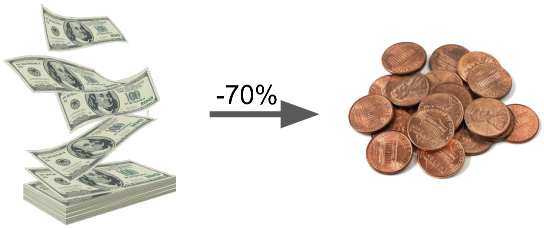
- Accuracy tradeoffs

Can we take advantage of particular resources in the cloud to train large models with much lower costs?

Spot Instances Can Lower Costs

Instances can be acquired cheaply

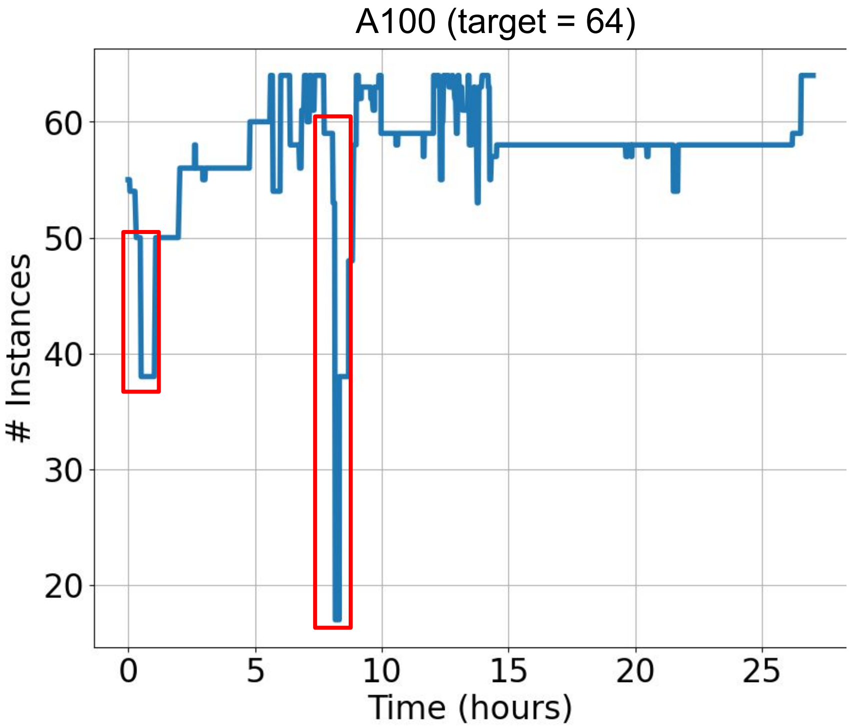
- Up to 70% lower costs

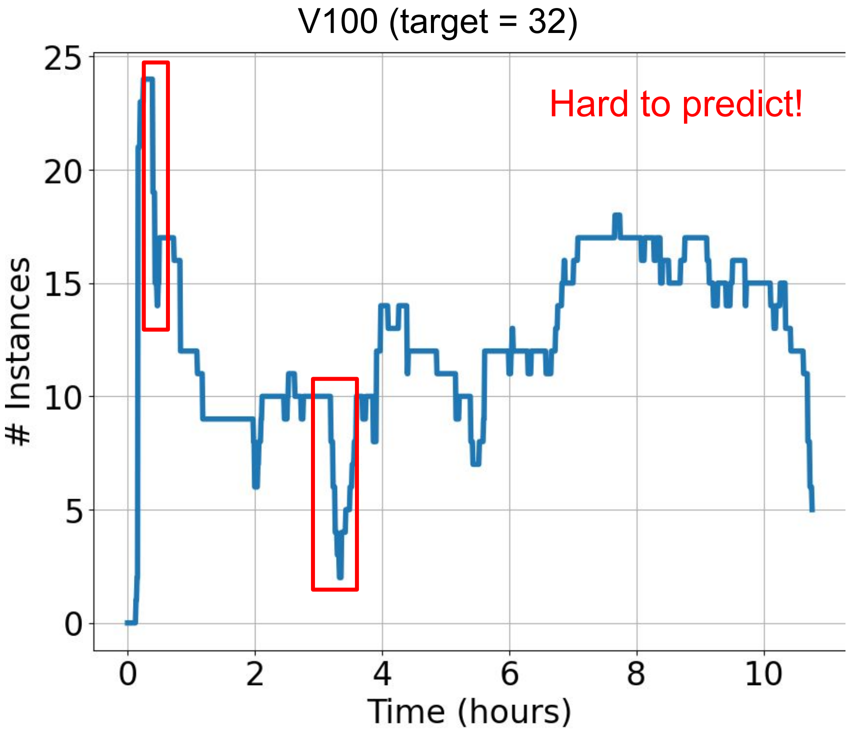


Preemptions can be unavoidable

- Price based preemptions: Just raise bid
- Capacity preemptions: No excess capacity left
  - Unavoidable!

Spot Instances Can Have High Failure Rates





UCLA

How To Deal With Preemptions?

Approximation

Sample dropping assumes that losing some samples is acceptable

- Remains true with smaller failure rates
- Loss severely impacted with higher rates

Checkpointing

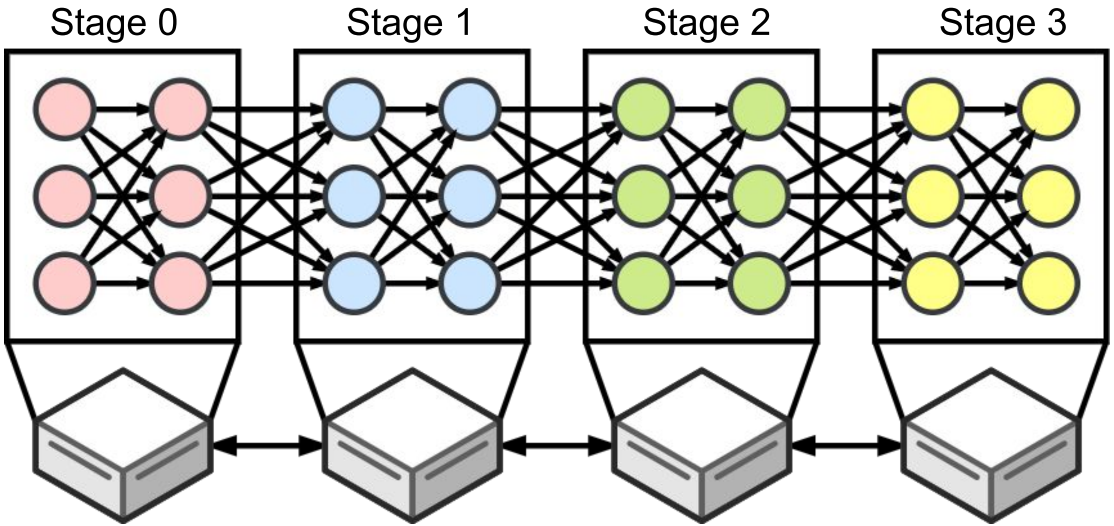
Roll back to stable state upon failures

- Maintains accuracy
- Acceptable performance for low fail rates
- Frequent restart when failures frequent

UCLA

Pipeline Parallelism Enables Large Model Training

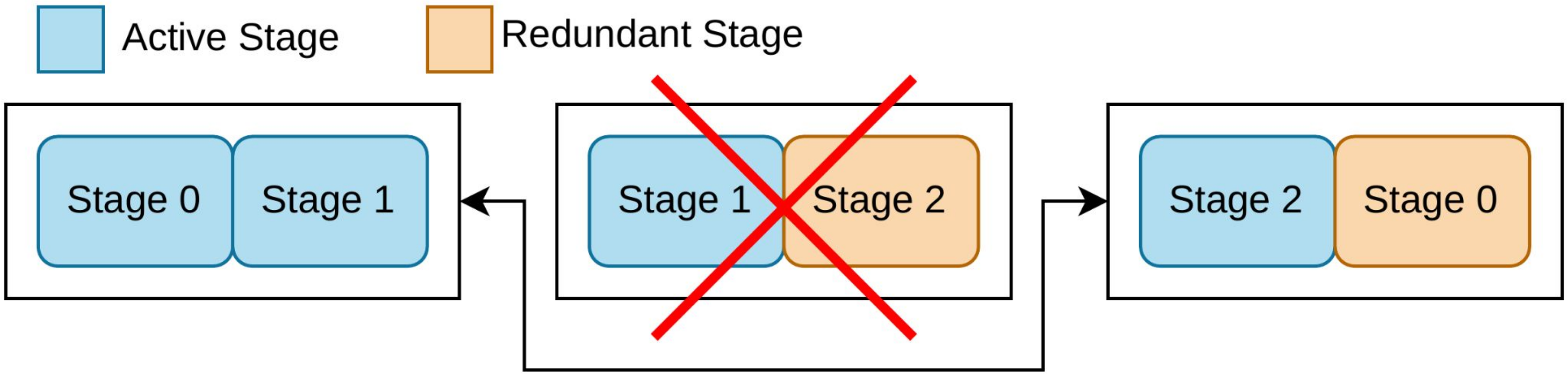
Pipeline Parallelism partitions the model among workers



Redundancy Provides Resilience

Redundant stages provide redundancy more quickly ✅

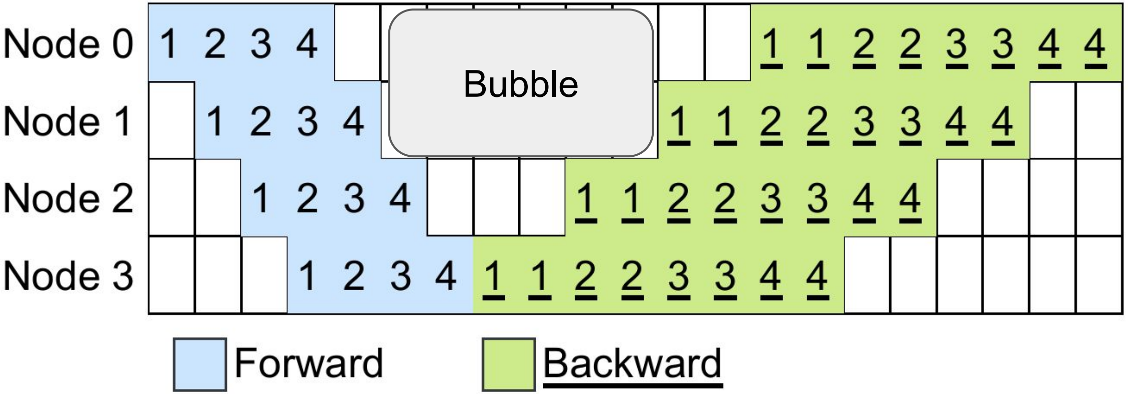
High overheads if done naively! ❌



Pipeline Parallelism Has Bubbles

Each mini-batch split up into micro-batches

Accumulate micro-batch gradients to get full batch gradients

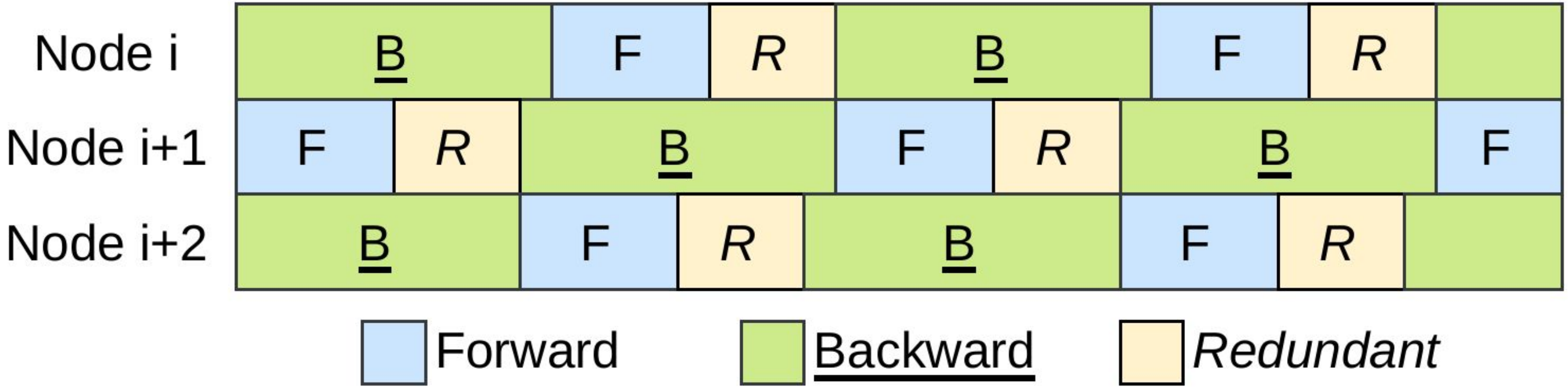


UCLA

Pipeline Bubbles Still Exist

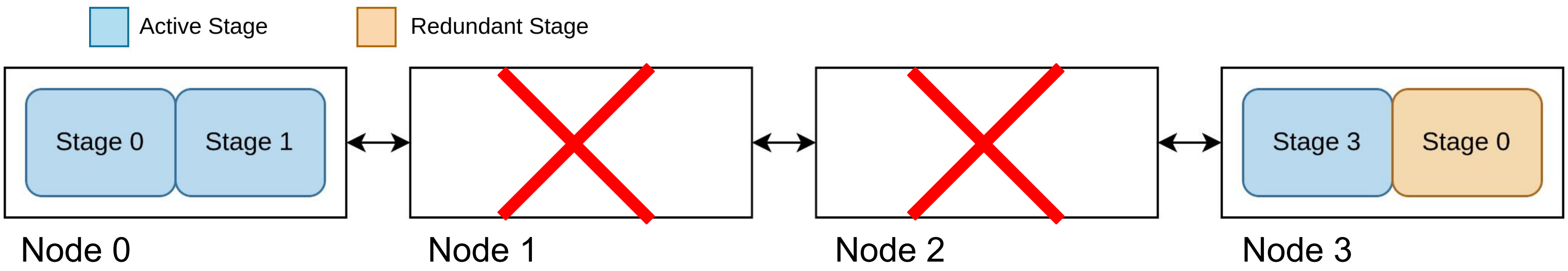
Bamboo

Can we use this idle time to minimize redundancy overhead?



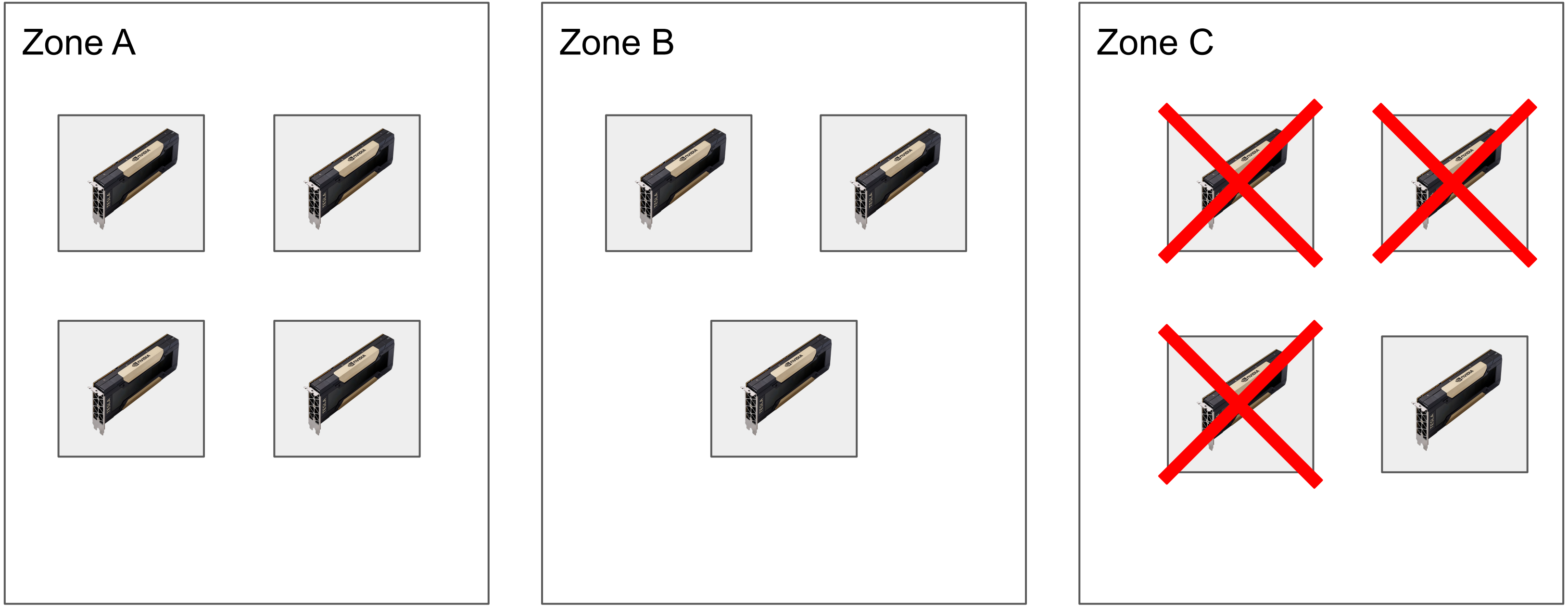
UCLA

Consecutive Failures Are Fatal

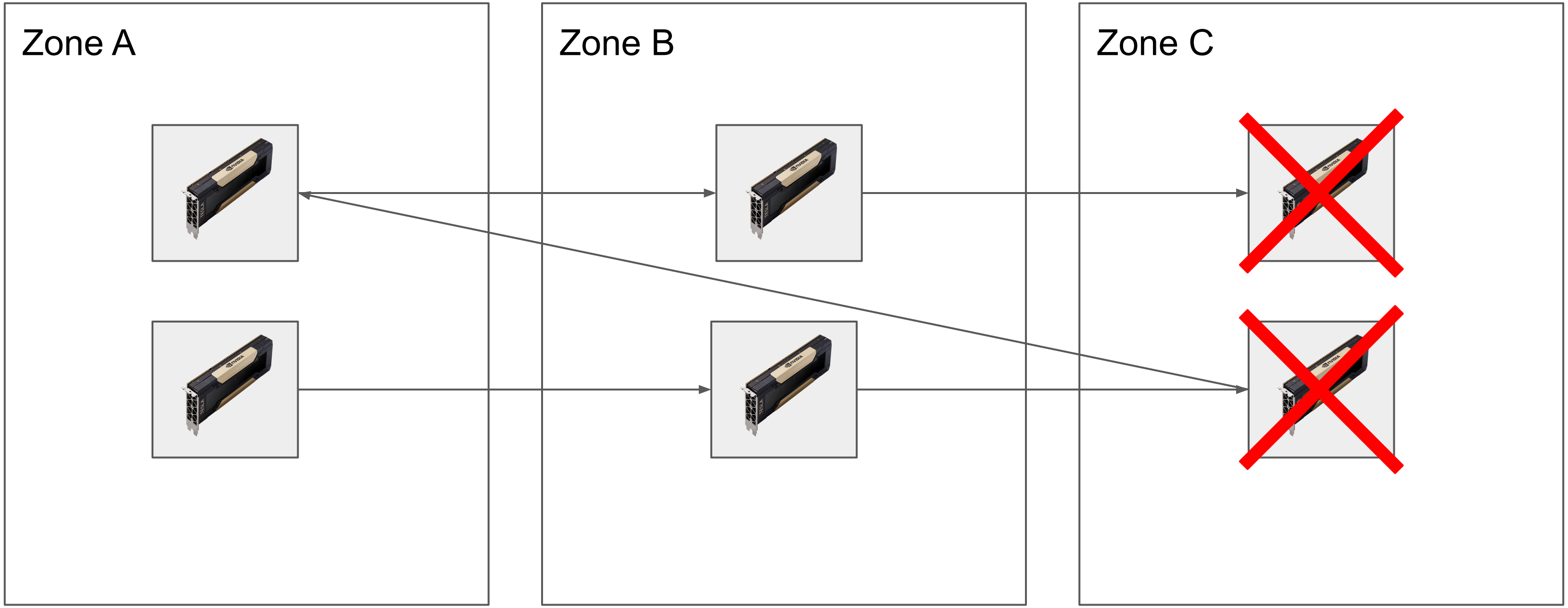


How To Avoid Consecutive Failures?

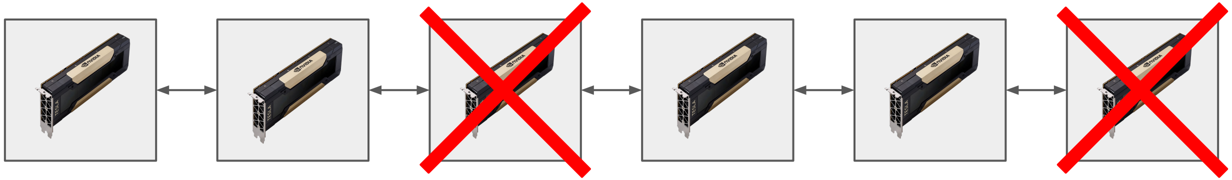
Bulk failures tend to happen in the same zone



Careful Placement Avoids Consecutive Preemptions



Careful Placement Avoids Consecutive Preemptions



Evaluated Bamboo On Many Datasets

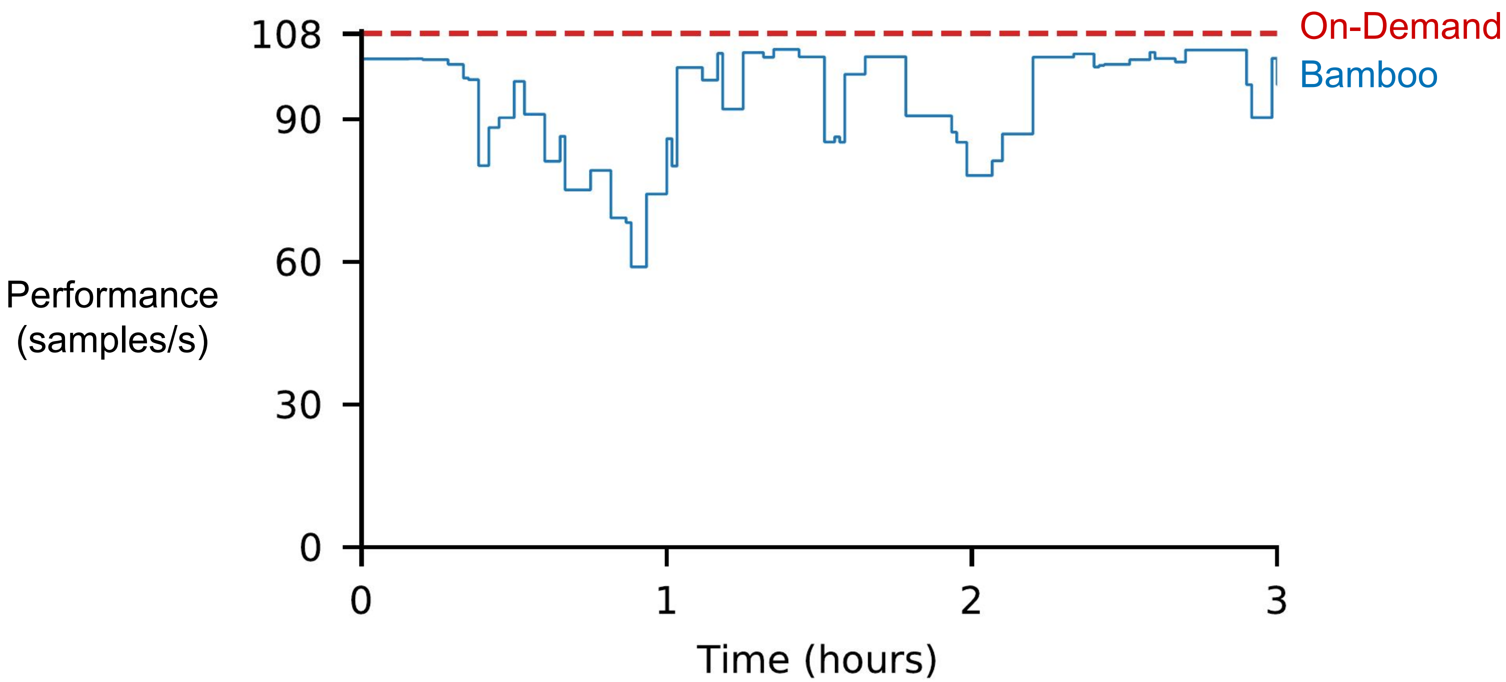
Model	DataSet	Data Parallel Size	Pipeline Size
ResNet-152	ImageNet	4	8x1.5 (12)
VGG-19	ImageNet	4	4x1.5 (6)
AlexNet	Synthetic Data	4	4x1.5 (6)
GNMT-16	WMT16EN-De	4	4x1.5 (6)
BERT-Large	Wikicorpus En	4	8x1.5 (12)
GPT-2	Wikicorpus En	4	8x1.5 (12)

Experiments

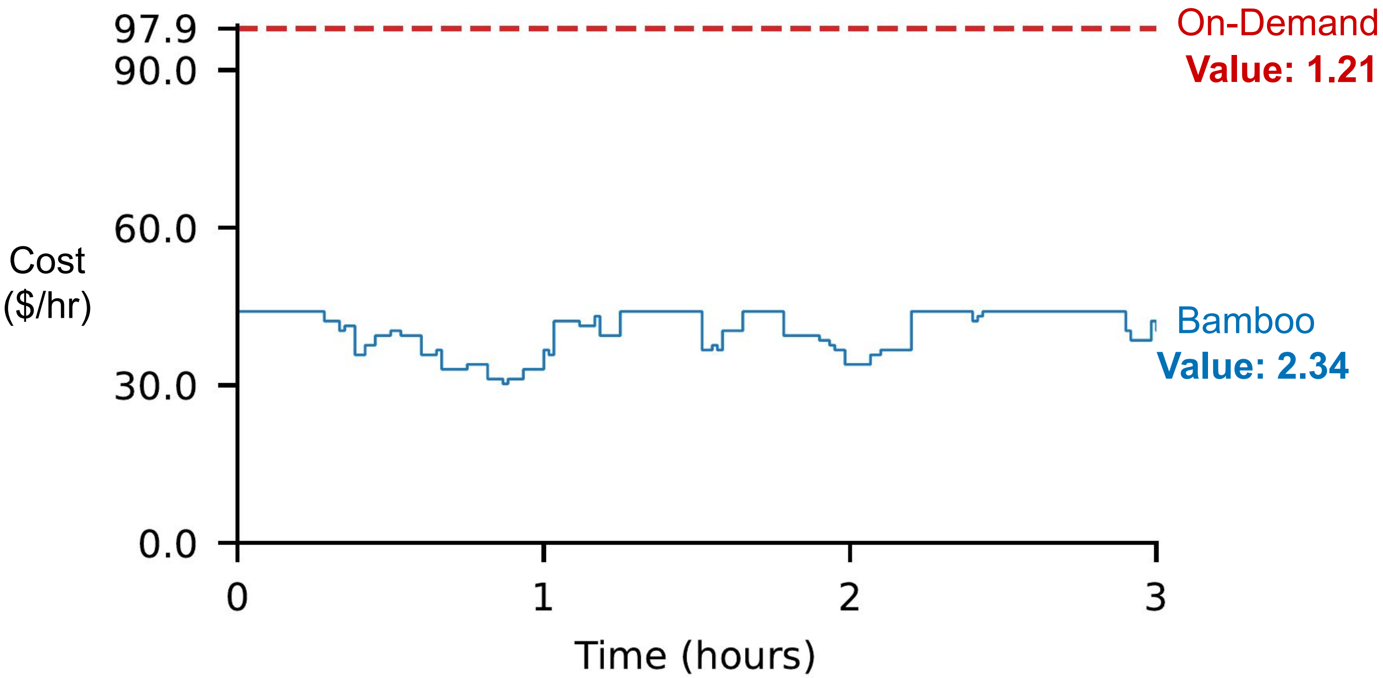
- Overall performance with respect to training costs
- Full simulations at different preemption rates
- Comparison against existing systems

Value: Performance-per-dollar

We Provide Comparable Performance to On-Demand



Bamboo Significantly Reduces Cost



Simulation of BERT to Completion

On-Demand Value: 1.1

Simulated BERT at different levels of preemption

Even at high preemptions maintain high value

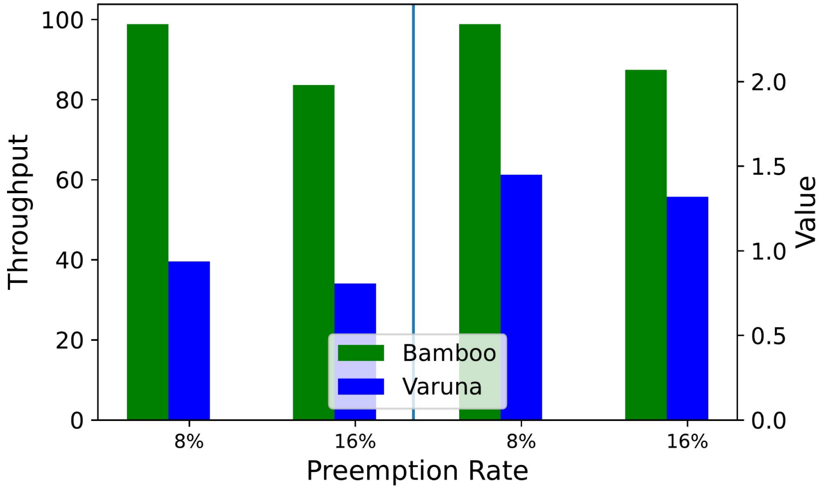
Probability	Throughput	Cost ($/hr)	Value
0.01	87.99	41.11	2.10
0.05	76.35	39.73	1.90
0.10	72.12	37.94	1.88
0.25	60.12	32.58	1.82
0.50	40.37	24.53	1.59

UCLA

Bamboo Provides More Value Than Similar Systems

Bamboo provides more value at different levels of preemption than Varuna

Frequent restarts and checkpoints slow Varuna



Bamboo Provides Resilience on Preemptible GPUs

Redundancy allows quick recovery from preemptions

Training efficiently on a changing set of resources

Provides 1.9x more value than On-Demand and 1.5x more than Varuna