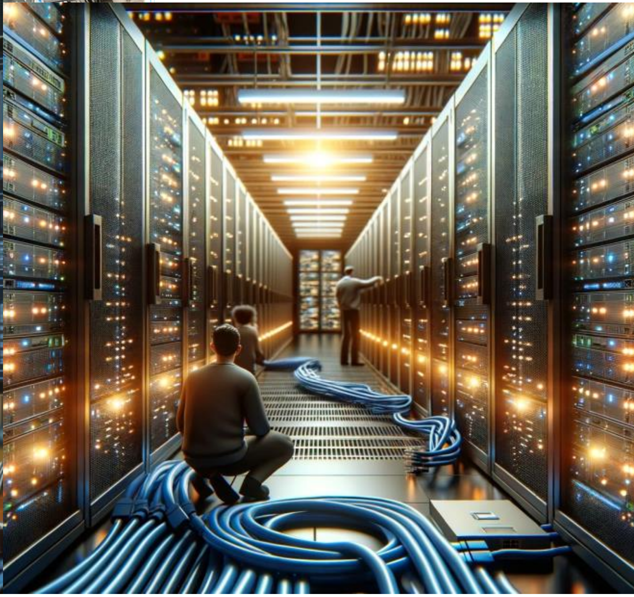


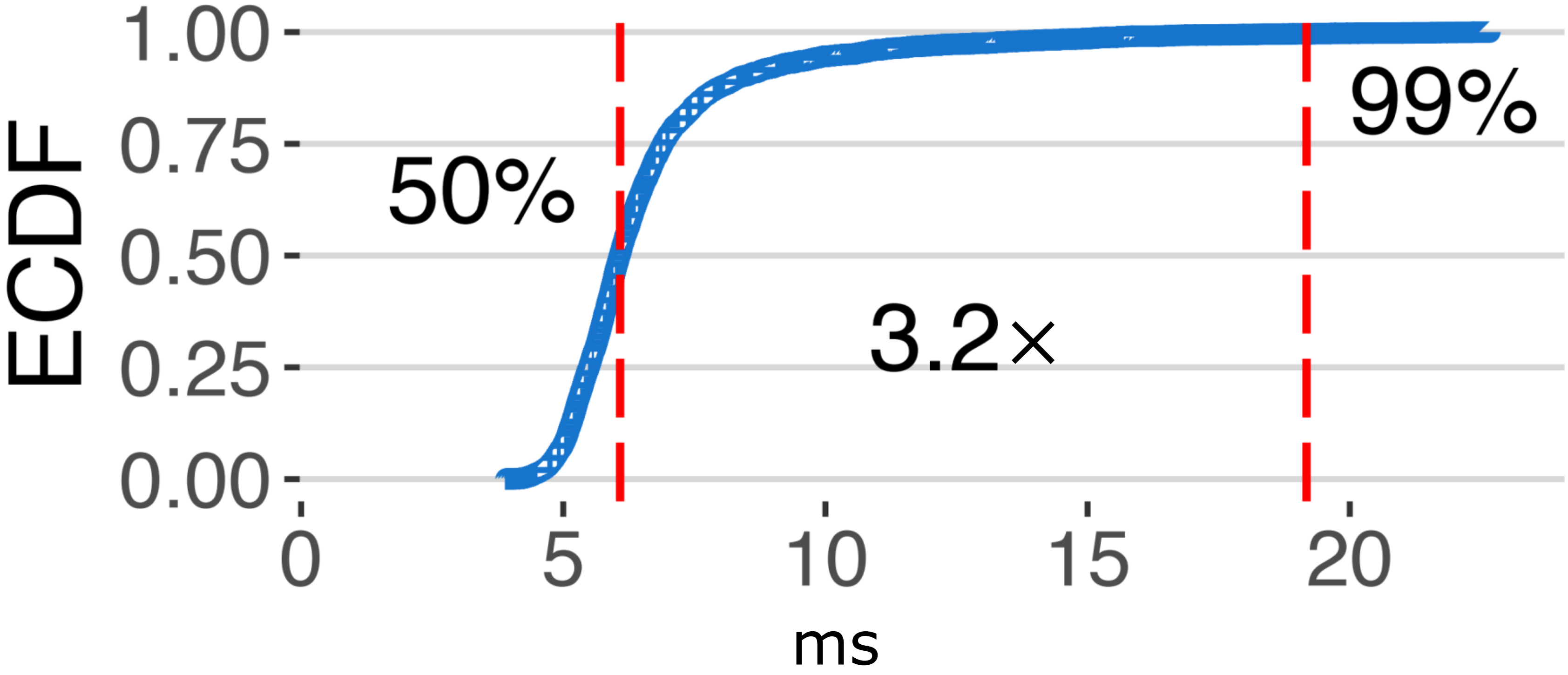
Lossy Network Transport for Large-Scale AI Insights and Future Directions

The current lossless abstraction

Everything sent shalt be received.

- If we lose something, e.g., a packet, we’ll retransmit it.

The tail latency problem



(OPTIREDUCE, NSDI 2025)

The tail latency problem

- Happens due to various reasons:
  - Packet drops due to congestion/corruption.
  - Queue buildups.
  - Accelerator failures.
  - Switch failures.
  - Link failures.
  - Straggler nodes.
  - …
- Exacerbated when training cross data centers.
- Mainly happens due to system events!

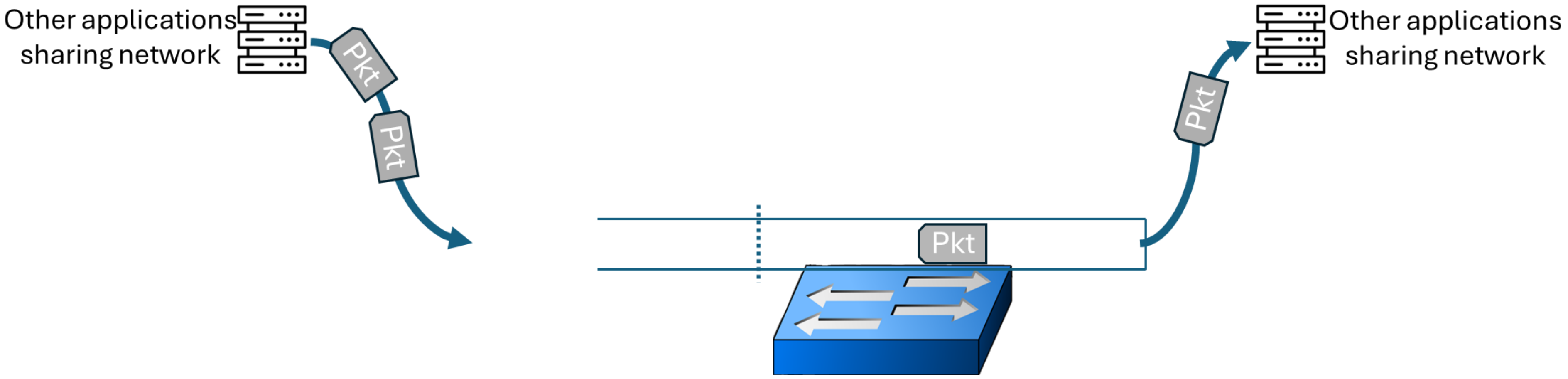
Embracing loss, a new paradigm

- We propose allowing the system flexibility in the gradient synchronization to mitigate the causes of the tail latency.
- To control the added loss, we need to formalize the allowed errors.

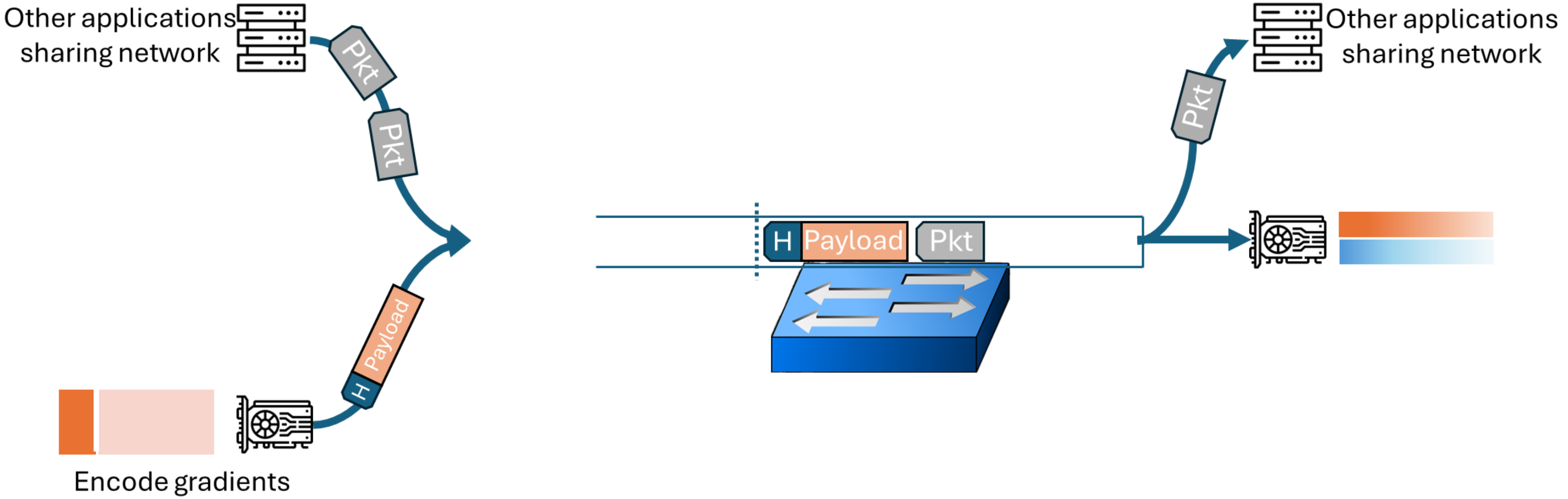
Example: Mitigating congestion with just-in-time compression (HotNets 2024)

- Congestion can increase latency of gradient synchronization and is not always avoidable.
- What if switches compress the data whenever needed?
- Can we do it with existing hardware switches?

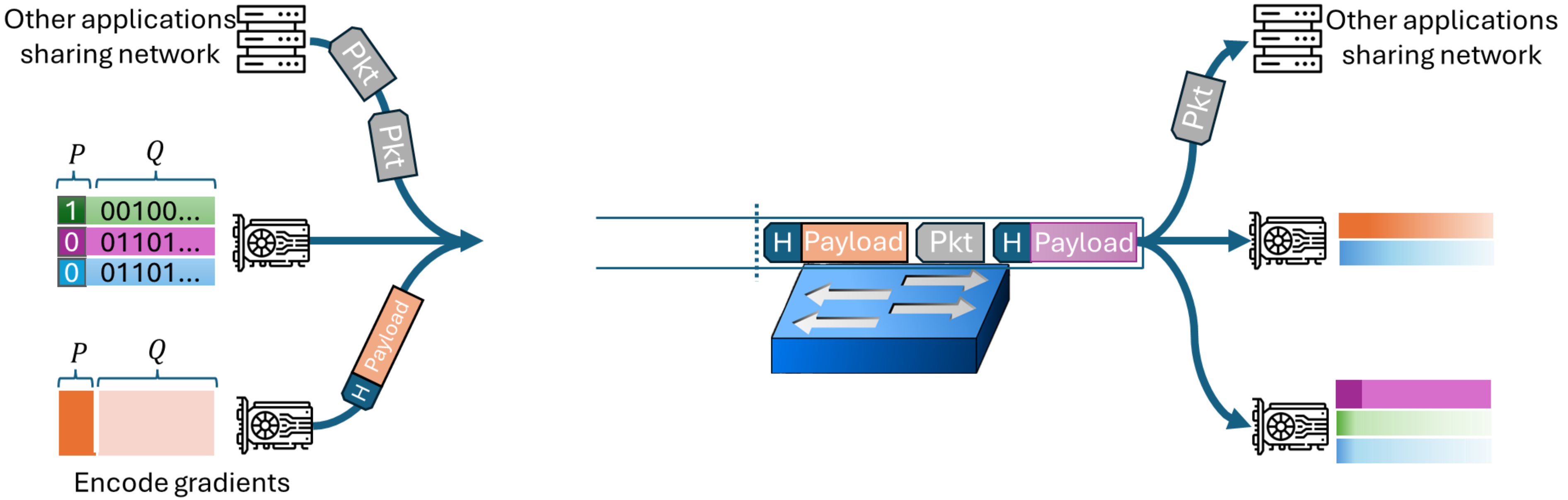
Example: Mitigating compression with just-in-time compression (HotNets 2024)


Other applications sharing network
Pkt
Pkt
Pkt
Pkt
Other applications sharing network

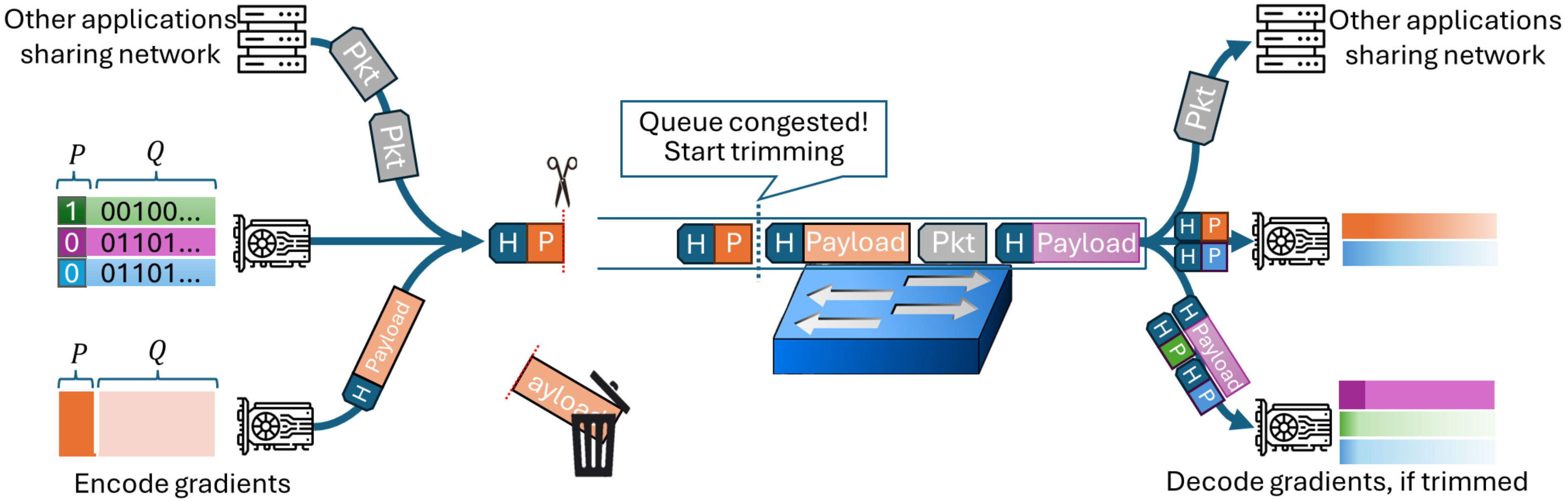
Example: Mitigating compression with just-in-time compression (HotNets 2024)


Other applications sharing network
Pkt
Pkt
Payload
H
Encode gradients
H
Payload
Pkt
Pkt
Other applications sharing network

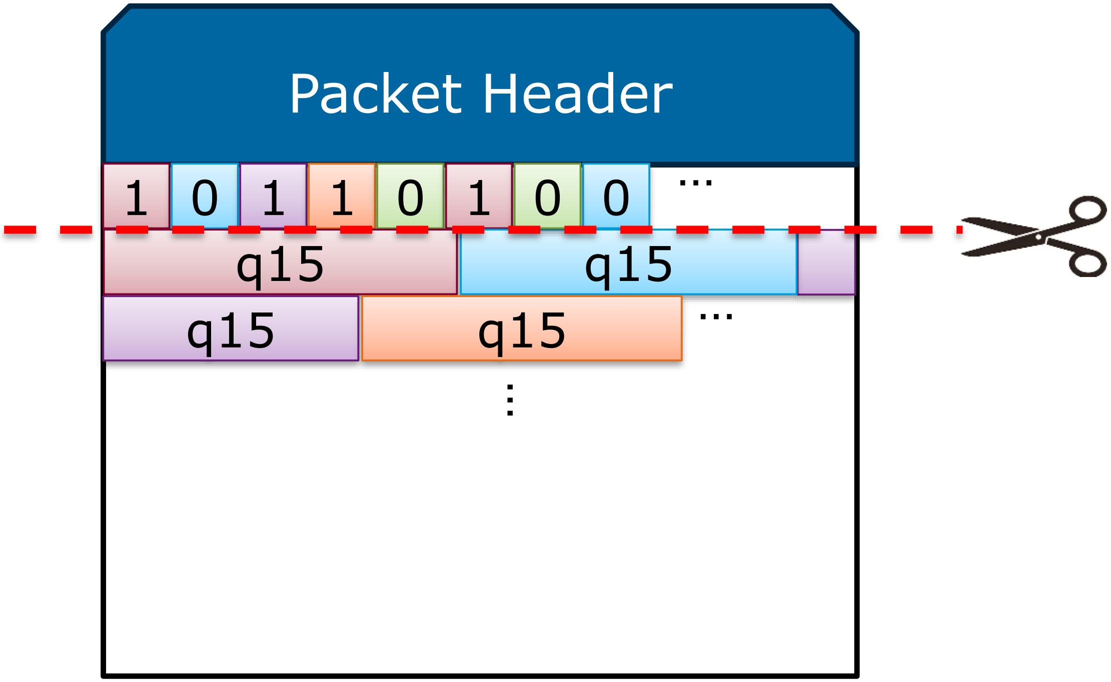
Example: Mitigating compression with just-in-time compression (HotNets 2024)



Example: Mitigating compression with just-in-time compression (HotNets 2024)


Other applications sharing network
Pkt
Pkt
P
Q
1 00100...
0 01101...
0 01101...
P
Q
H
Payload
Encode gradients
H P
ayload
Queue congested! Start trimming
H P
H Payload
Pkt
H Payload
Pkt
Other applications sharing network
H P
H P
H
Payload
H
P
H
P
Decode gradients, if trimmed

Example: Mitigating compression with just-in-time compression (HotNets 2024)



- The one-bit that makes it serves as state-of-the-art compression algorithm! (DRIVE, NeurIPS 2021).

Reproducibility

- Lossy and stochastic synchronization does **not** mean non-reproducible results!

- By logging the sources of loss, we can replay the execution of the process. For example, we log:
  - Which stragglers were dropped.
  - Which packets were trimmed.
  - …

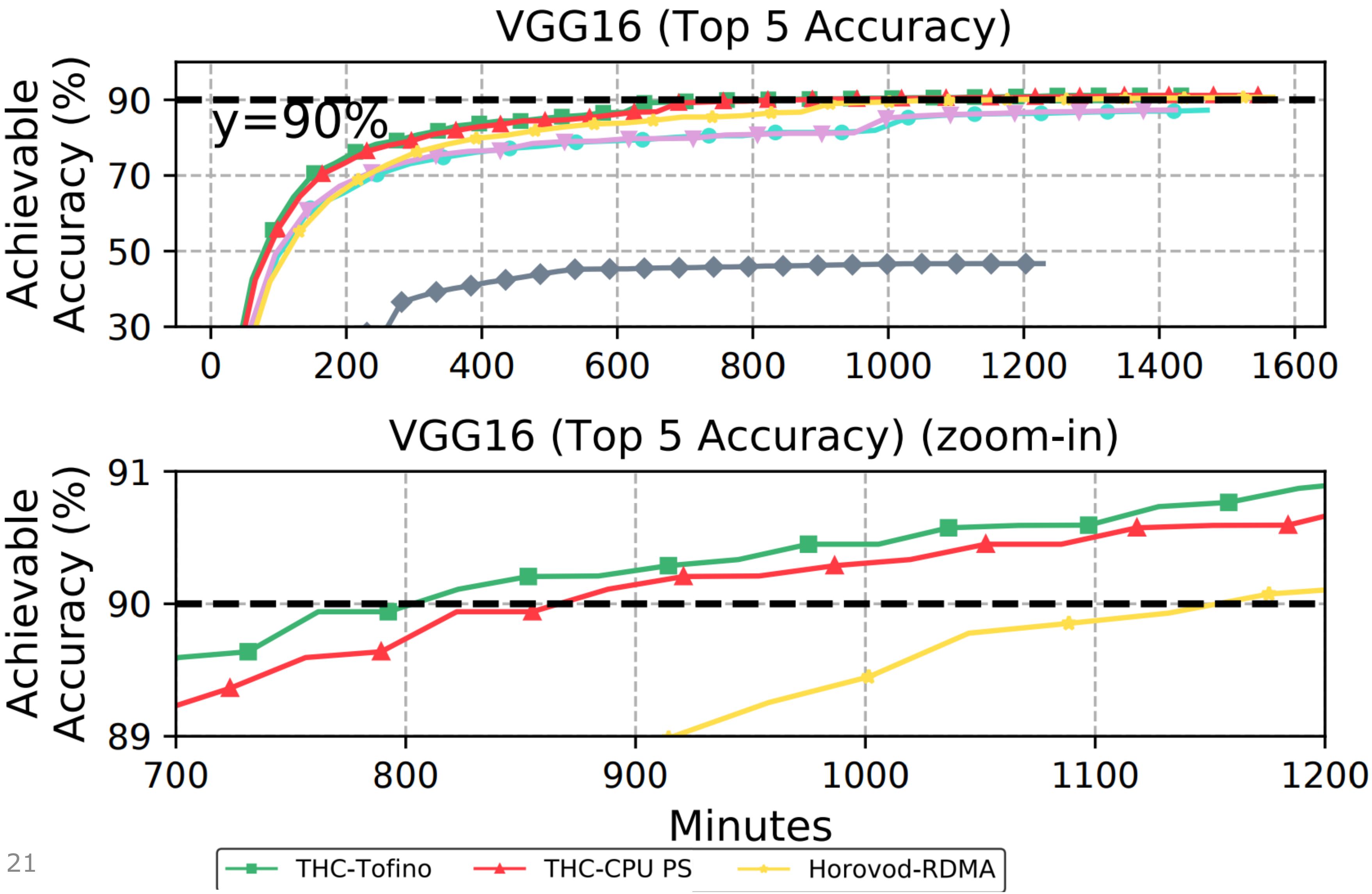
Vision

- We develop an understanding of how system optimizations affect gradient variance and thus the convergence rate.

- We can trade a small (e.g., 5%) increase in either the #rounds or batch size into a “budget” that the system optimizations can leverage.

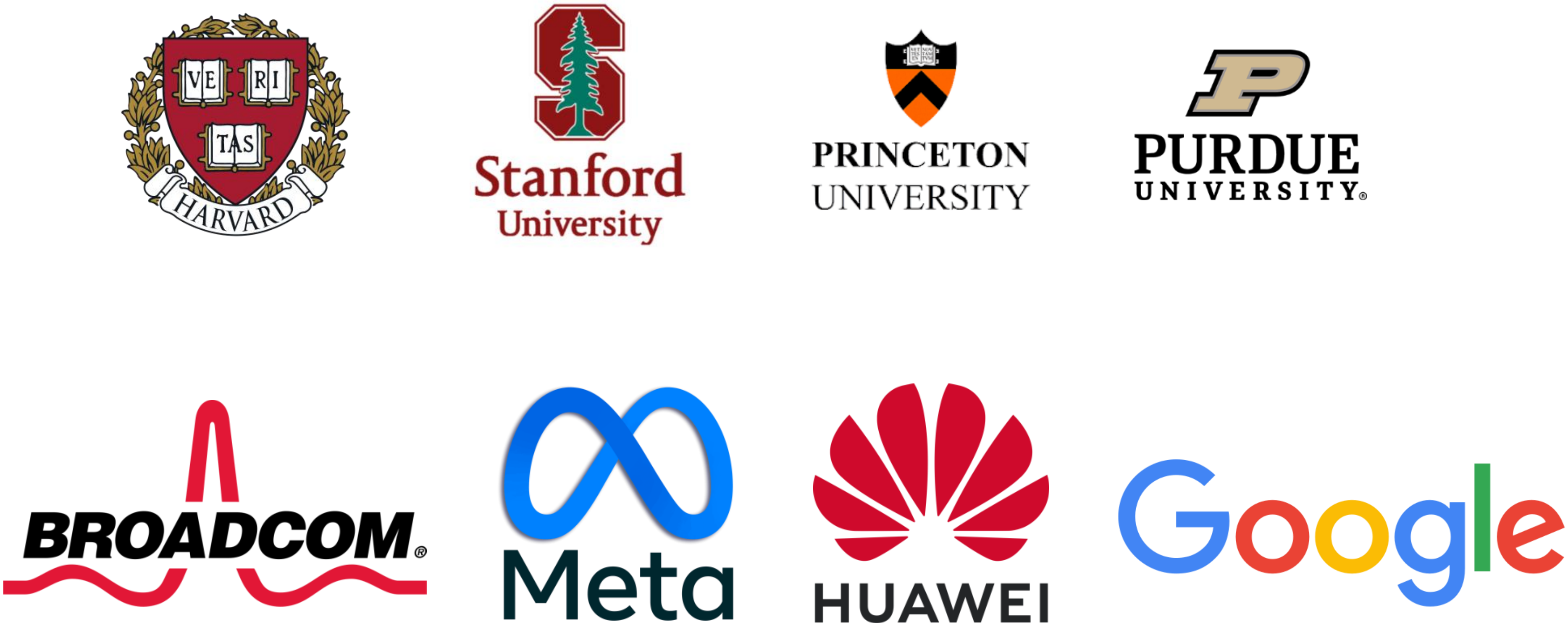
- We measure the benefit in **time** taken to train a model.

Example (THC, NSDI 2024)



Summary

Collaborators and funders


VE RI TAS
HARVARD
Stanford University
PRINCETON UNIVERSITY
PURDUE UNIVERSITY
BROADCOM
Meta
HUAWEI
Google