

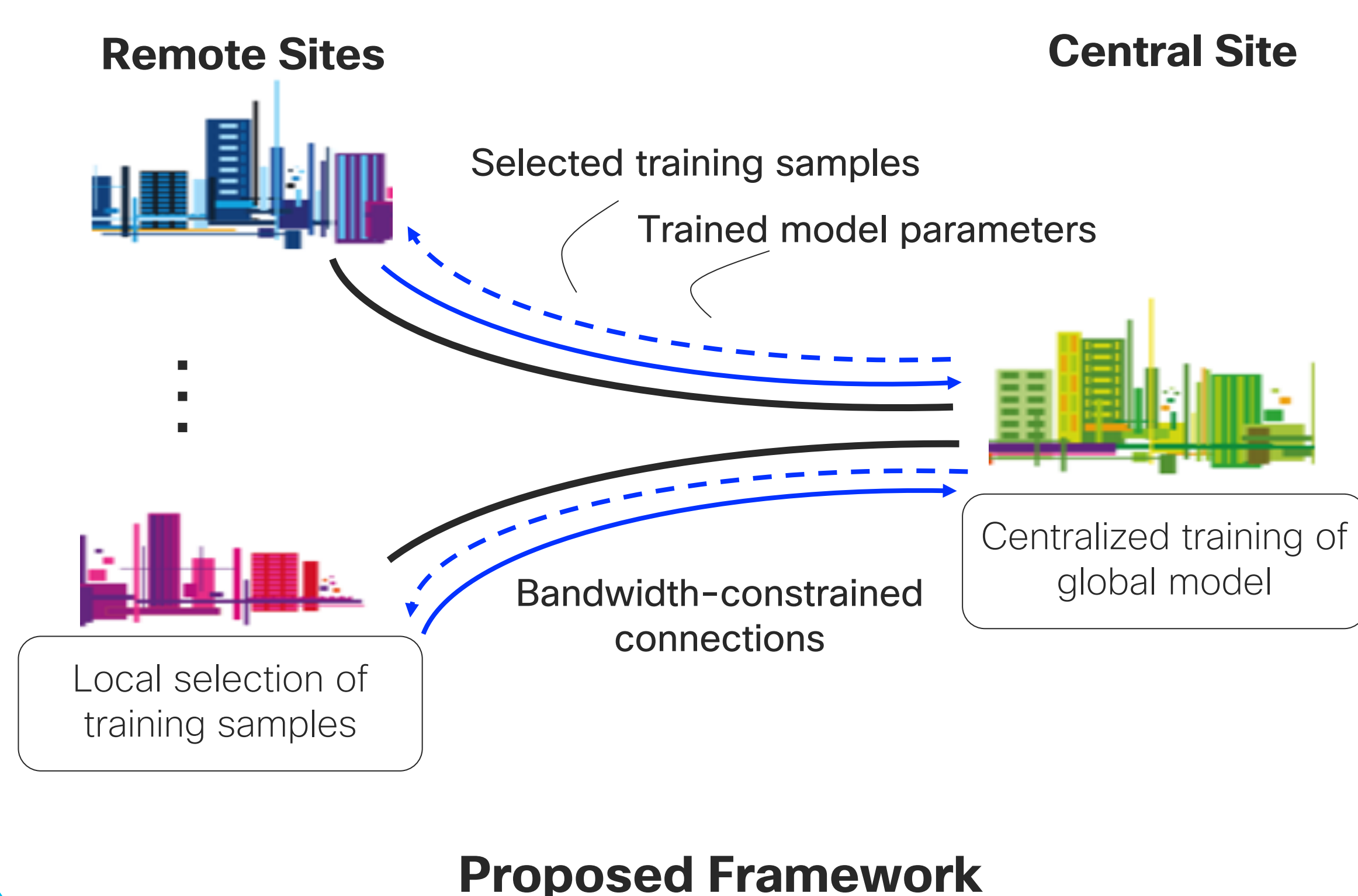
Training Sample Selection for Deep Learning of Distributed Data

Zheng Jiang, Xiaoqing Zhu, Wai-tian Tan, and Rob Liston

Chief Technology and Architecture Office (CTAO), Cisco Systems Inc.



A Framework for Learning Distributed Data



Application Scenarios

- Multi-site video surveillance
- Wireless network telemetry
- ...

Key Challenges

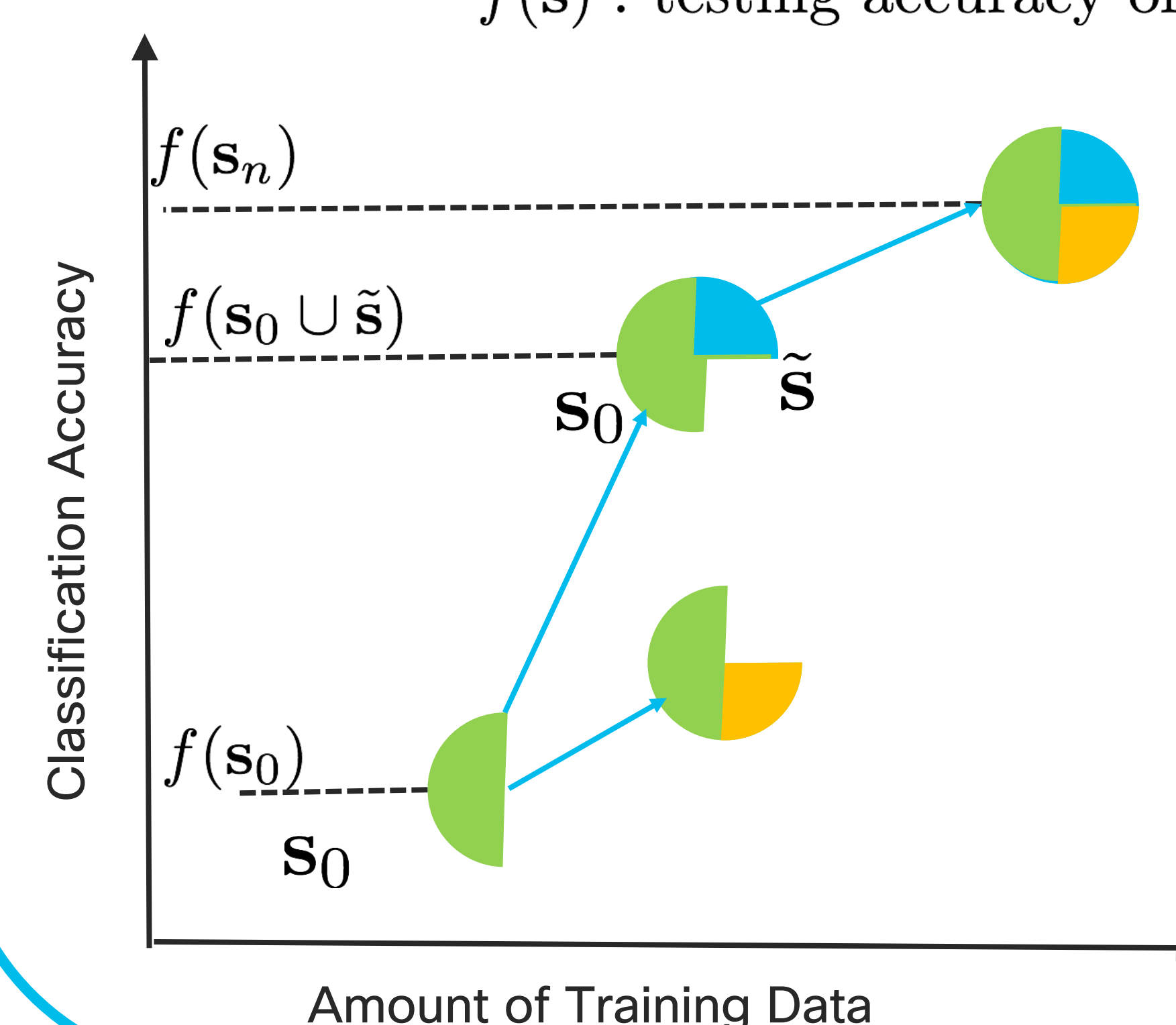
- Massive locally observed data
- Bandwidth-constrained connections
- Need for aggregating learning across multiple sites
- Need for continuous learning

Training Sample Selection

s_0 : subset of existing training samples at the central site

\tilde{s} : subset of selected training samples from the remote site

$f(s)$: testing accuracy of training using dataset s



Optimization Objective

Improvement in testing accuracy

$$\Delta_{\tilde{s}} = f(\tilde{s} \cup s_0) - f(s_0).$$

$$\max_{\tilde{s}} \Delta_{\tilde{s}}$$

$$\text{s.t. } \tilde{s} = \{s_i | s_i \in s, z_i = 1\}$$

$$z_i \in \{0, 1\} \quad \forall i \in \{1, \dots, n\}$$

$$\sum_{i=1}^n z_i < \rho |s| = \rho n.$$

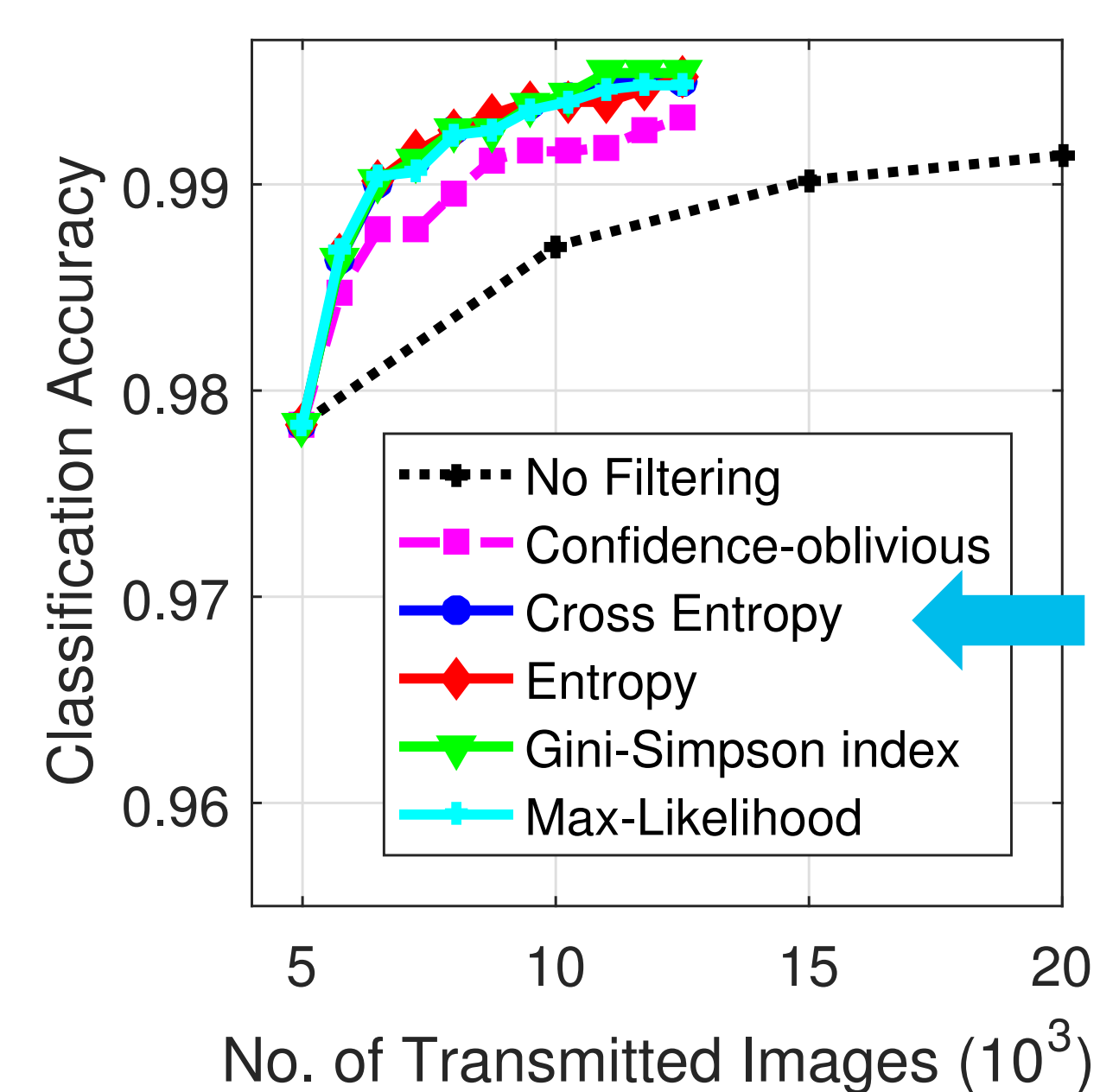
Binary selection vector

Constraint on sample count

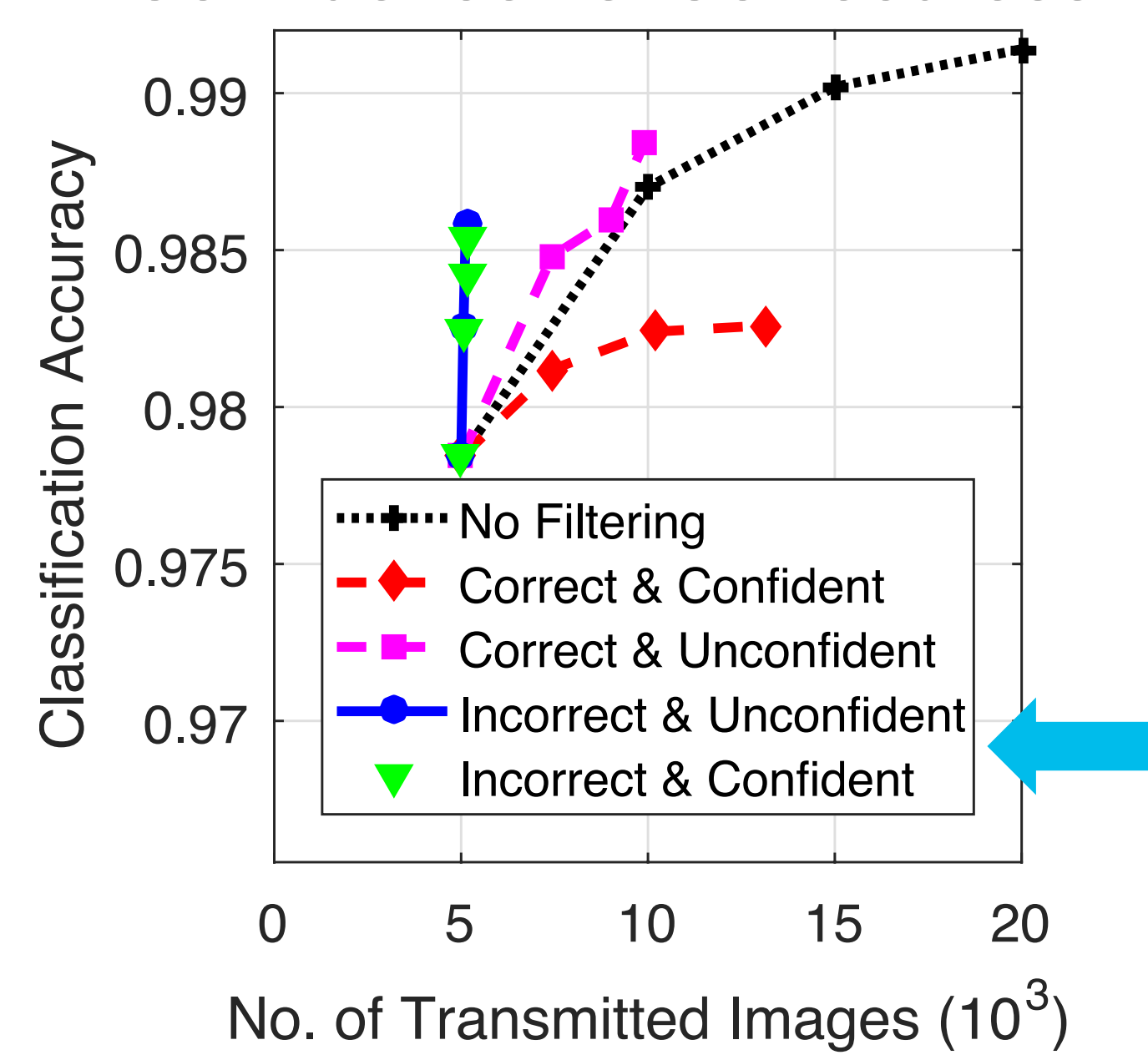
Proxy Function Design

- Greedy approximation: $\Delta_{\tilde{s}} \approx \sum_{s_i \in \tilde{s}} \Delta_{s_i}$,
- Experimental evaluation of the per-sample proxy function Δ_{s_i}

Choice of Confidence Metric



Impact of Confidence vs. Correctness



Evaluations and Summary

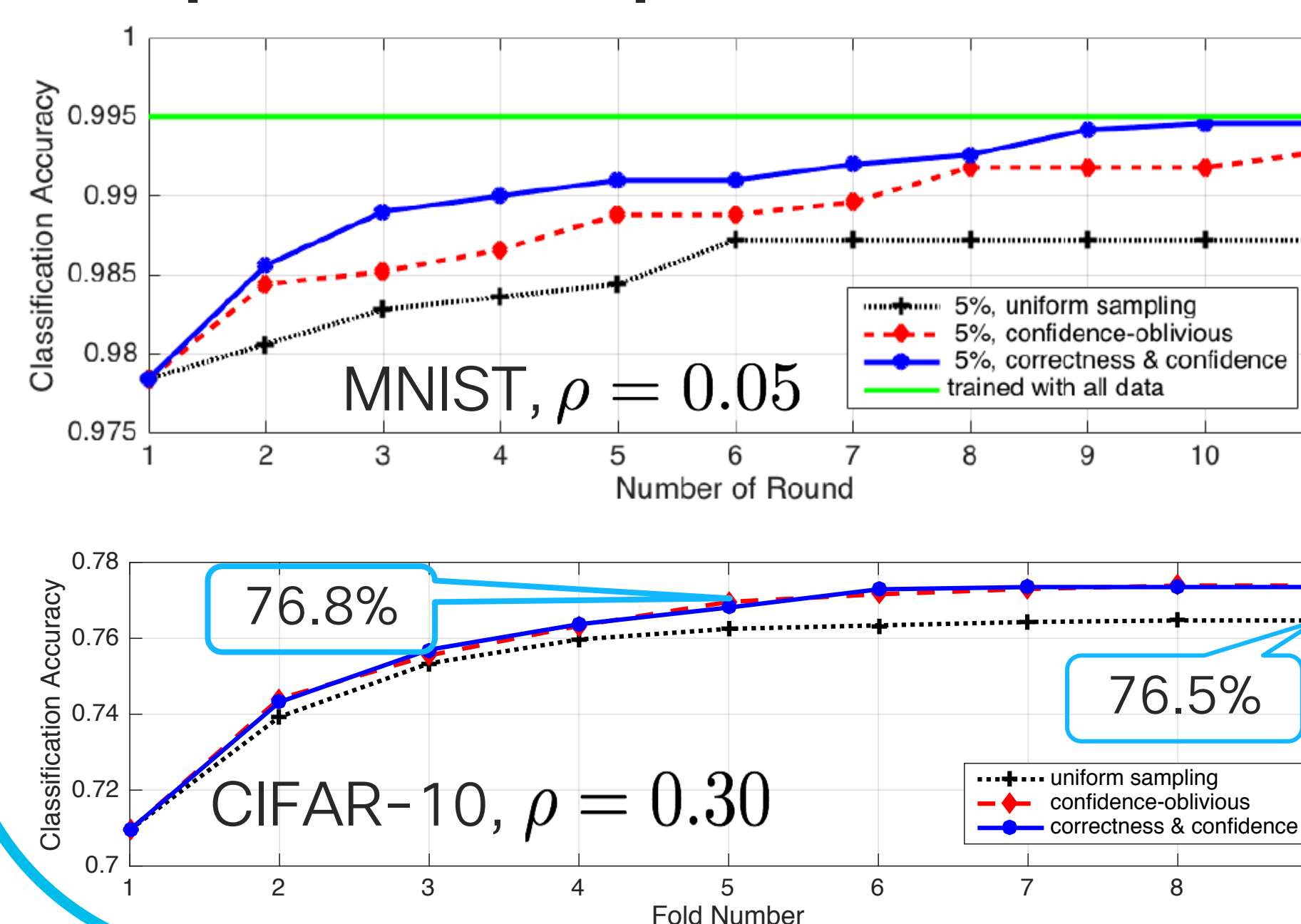
Datasets

- Handwritten digits: MNIST
- Tiny images: CIFAR-10

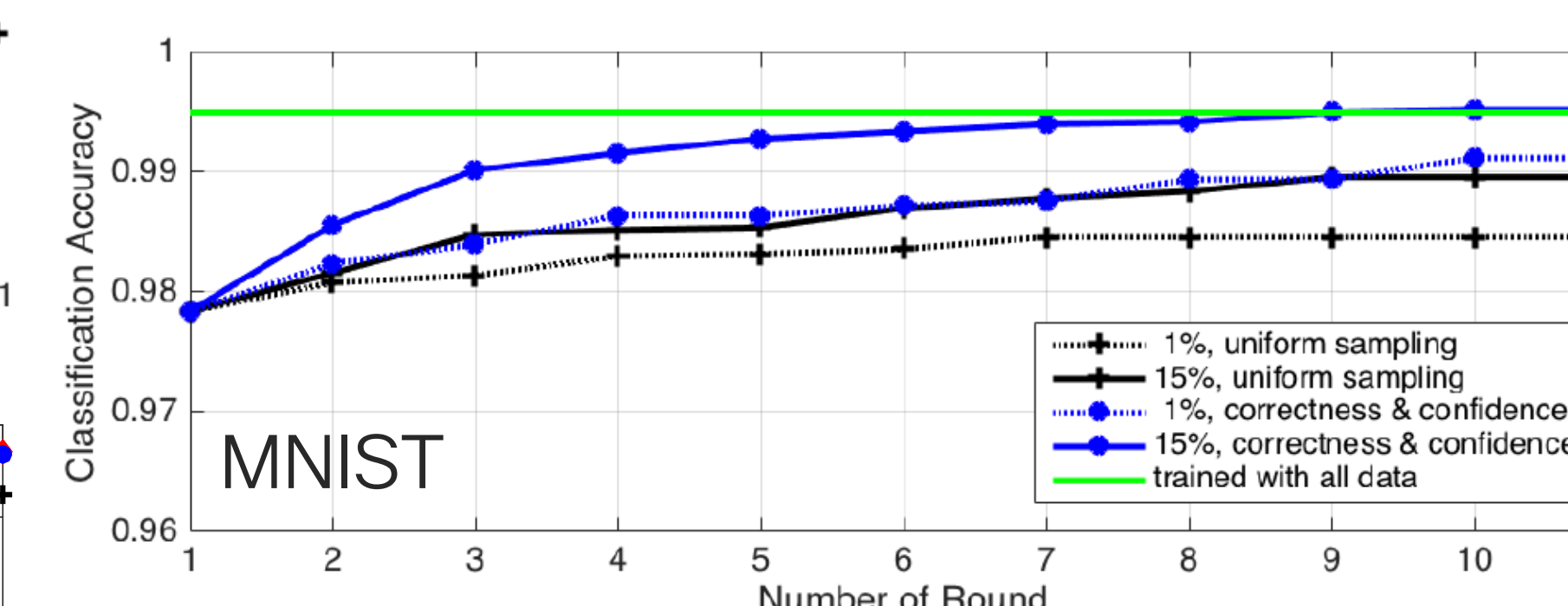
Procedure

- Initialize central classifier with a subset of training data; divide the rest into multiple rounds
- Remote site uses local copy of the classifier to select & forward samples from the next round
- Central site re-train classifier and feedback updated copy to local site

Comparison of Sample Selection Schemes



Impact of bandwidth constraints



Summary

- Not all training samples are equal: some more important than others
- Correctness & confidence from local test trials can be used as an effective proxy for their importance
- Benefits: bandwidth saving by 15x for MNIST; training time reduction by 44% for CIFAR-10

