



# Global-QSGD

Allreduce-Compatible Quantization for  
Distributed Learning with Theoretical Guarantees

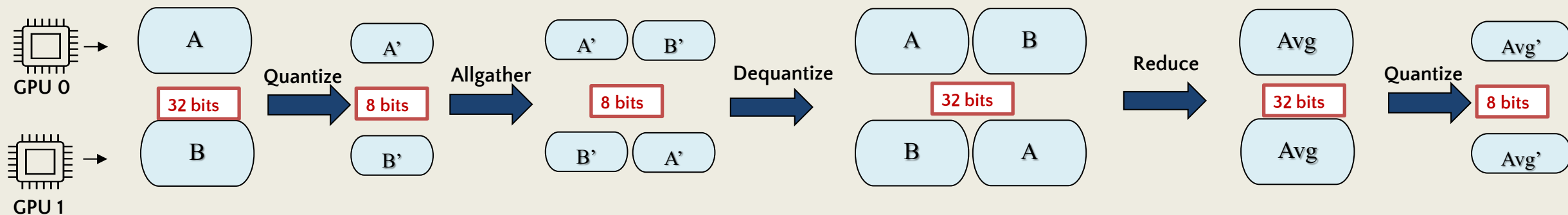
**Jihao Xin**  
KAUST

**Marco Canini**  
KAUST

**Peter Richtárik**  
KAUST

**Samuel Horváth**  
MBZUAI

## Problem: Quantization is popular in distributed learning but does not support Allreduce



## Solution: Design an Allreduce-compatible Quantization using Global Norm

- Quantization requires a “Scaling Factor”
- “Scaling Factor” are calculated locally
- Scaled by different factor, quantized gradients cannot be aggregated directly

## Global-QSGD: 2 round Allreduce protocol

- Each GPU finds local norm
- 1<sup>st</sup>-round Allreduce to exchange the Global Norm
- Each GPU Quantize by Global Norm
- 2<sup>nd</sup>-round Allreduce to exchange quantized data

