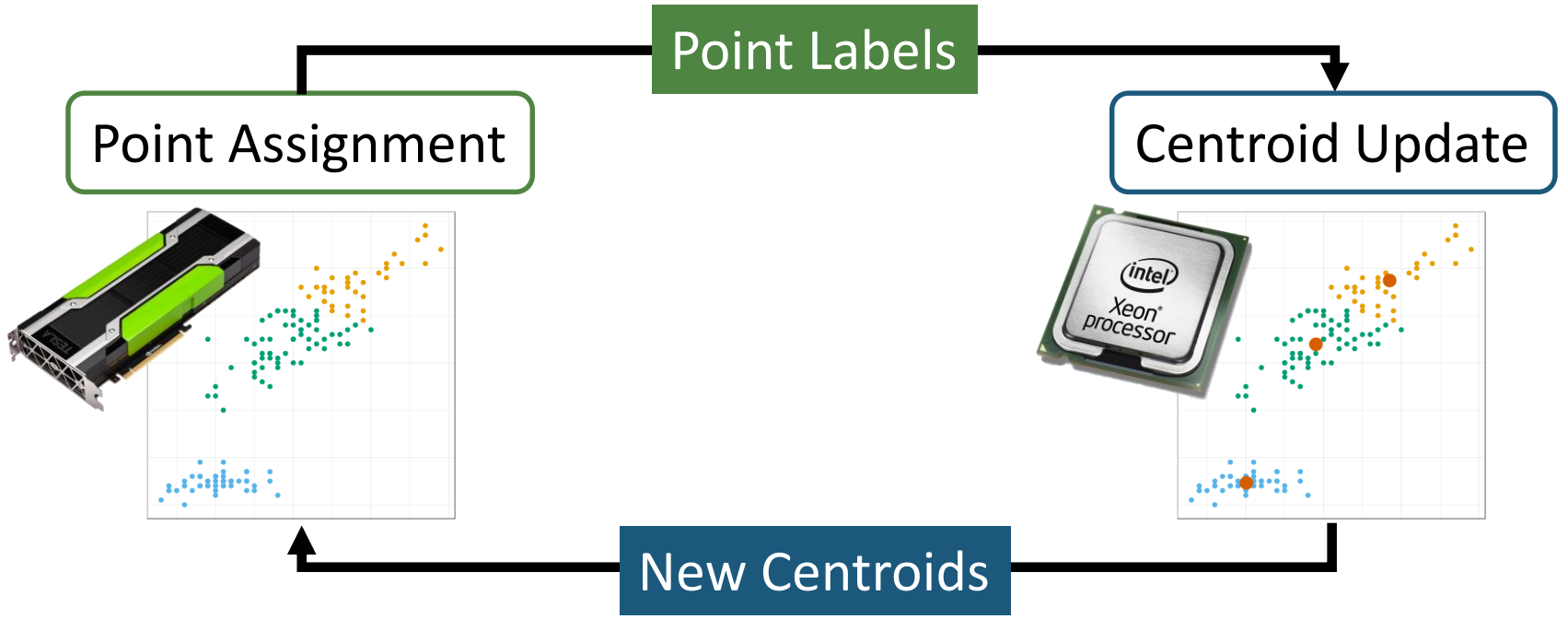
Efficient k-Means on GPUs

Clemens Lutz, Sebastian Breß, Tilmann Rabl

Steffen Zeuch, Volker Markl

k-Means State-of-the-Art

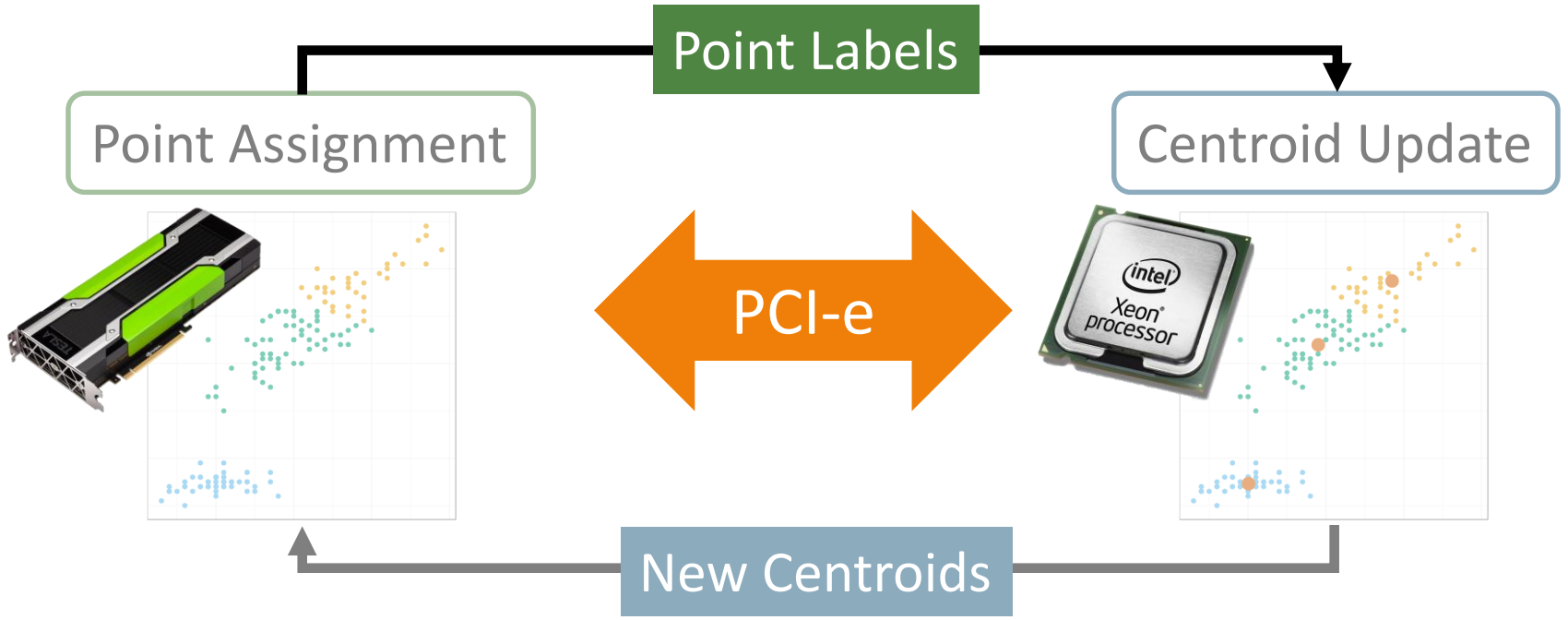
- Accelerate point assignment on GPU
  - High parallelization power



Problem 1: Cross-Processing

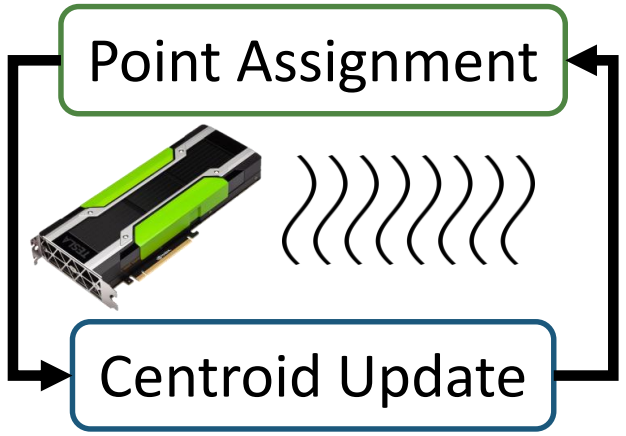
Incurs high PCI-e transfer costs

- Processing split between two processors
  - PCI-e bus transfer for labels and centroids



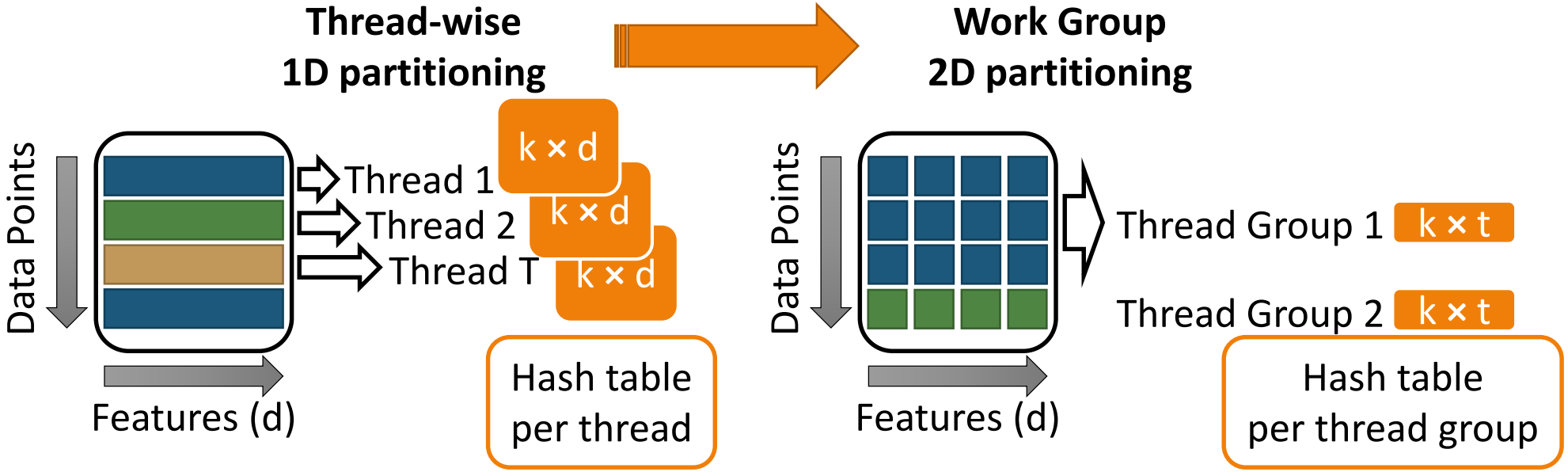
Solution: Update Centroids on GPU

- New algorithm for centroid update
  - Reduce cache footprint



Untie cache footprint from data features

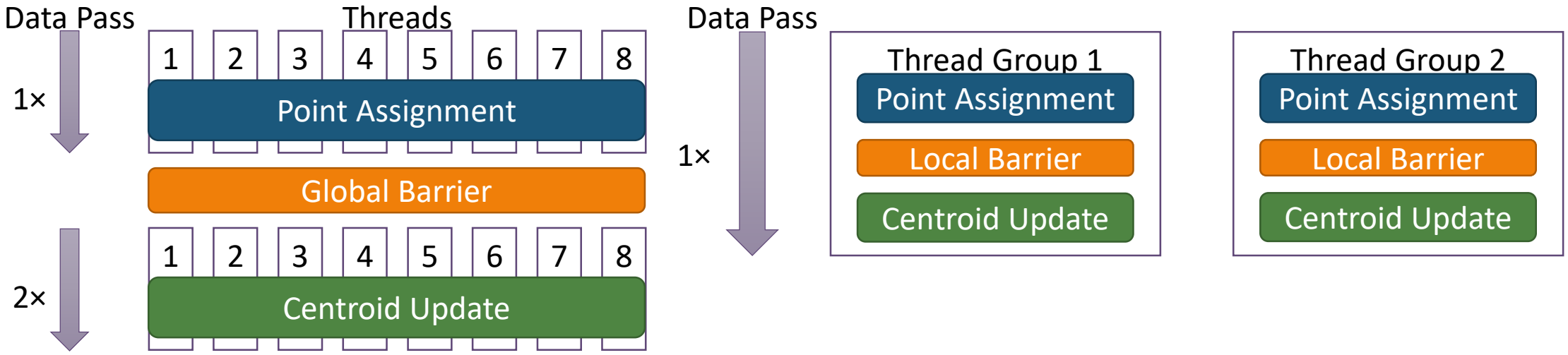
Eliminate PCI-e transfer costs



Problem 2: Multi-Pass

Solution: Single pass per iteration

- Point assignment and centroid update each make a data pass
  - Global barrier necessary due to transposed data access
  - Transpose on-the-fly and use local barrier inside thread group instead



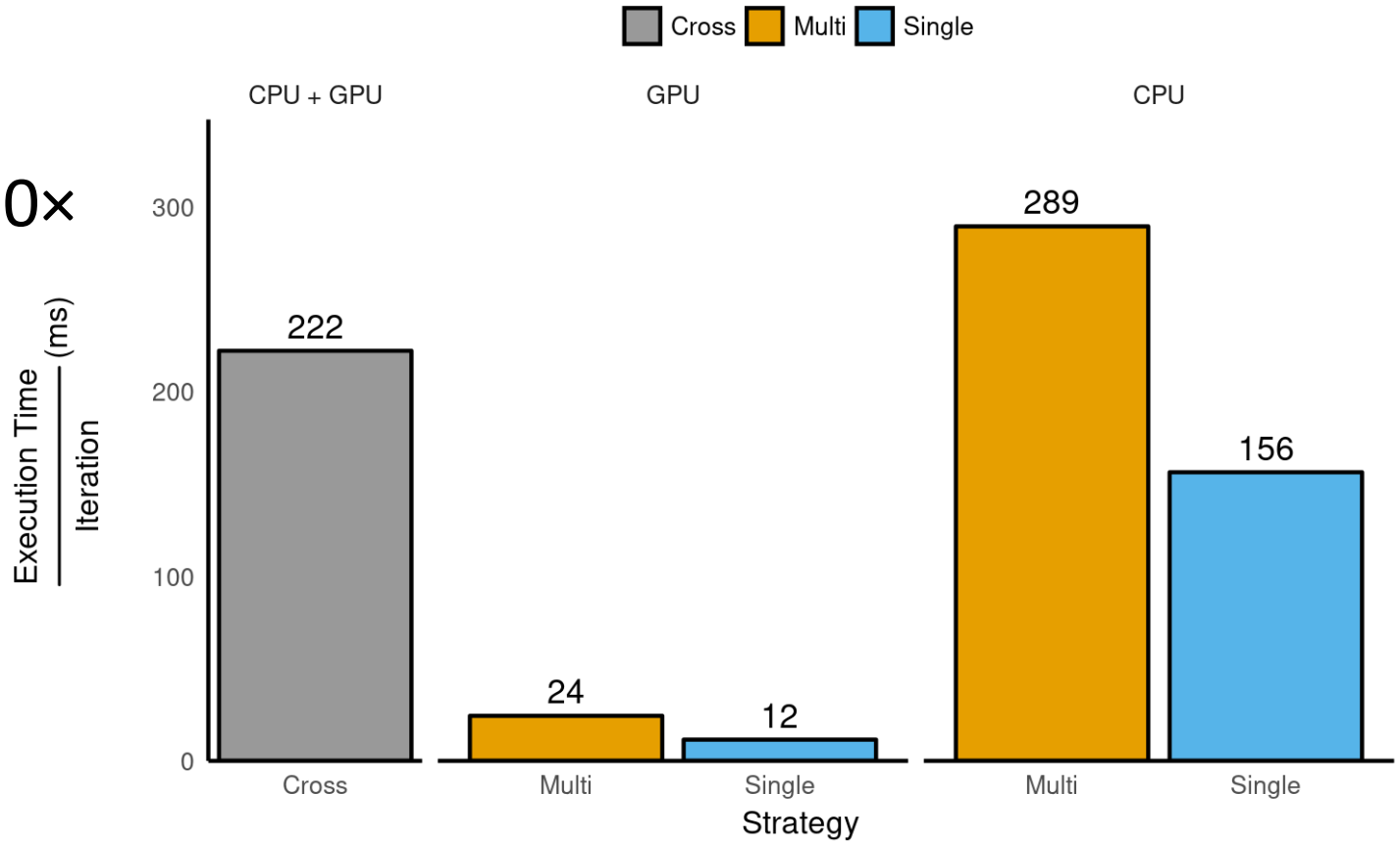
Benefits of Centroid Update on the GPU and a Single Data Pass

GPU:

- Cross-processing problem: 10×
- Multi-pass problem: 2×
- Overall: 20×

CPU:

- Multi-pass problem: 1.8×



Contributions – “Efficient k-Means on GPUs”

Cross-Processing Problem

→ Efficient Centroid Update for GPUs

Multi-Pass Problem

→ Single-Pass k-Means

https://www.dfki.de/web/forschung/publikationen?pubid=9767