Efficient Bundled Spatial Range Queries

Eleni Tzirita Zacharatou‡ Darius Sidlauskas ‡

Thomas Heinis* Farhan Tauheed‡ Anastasia Ailamaki‡

DIAS Data-Intensive Applications and Systems

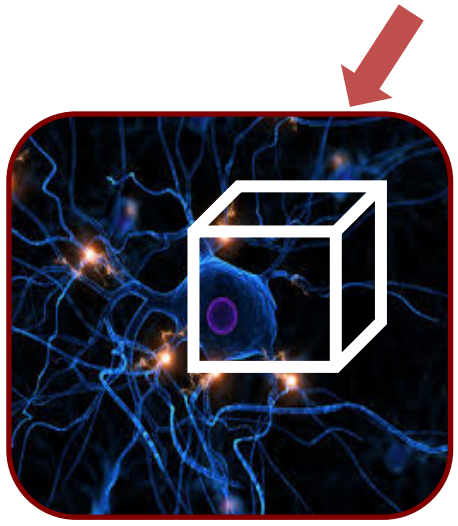
* Imperial College London

‡ EPFL

Exploring Multiple Spatial Datasets

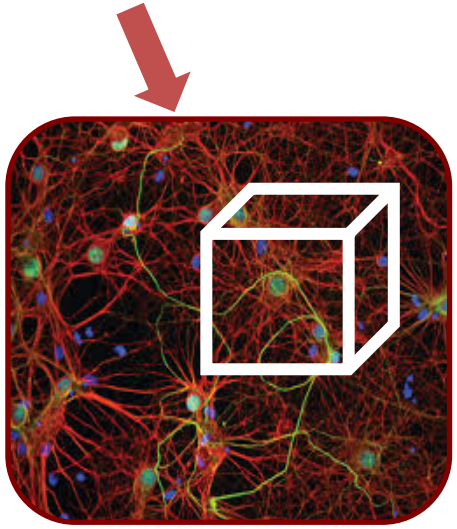
Problem: Spatial range query on multiple categories of spatial objects within the same spatial universe

Query: { 3D Spatial Range; Category preference }



Category 1

Category 2

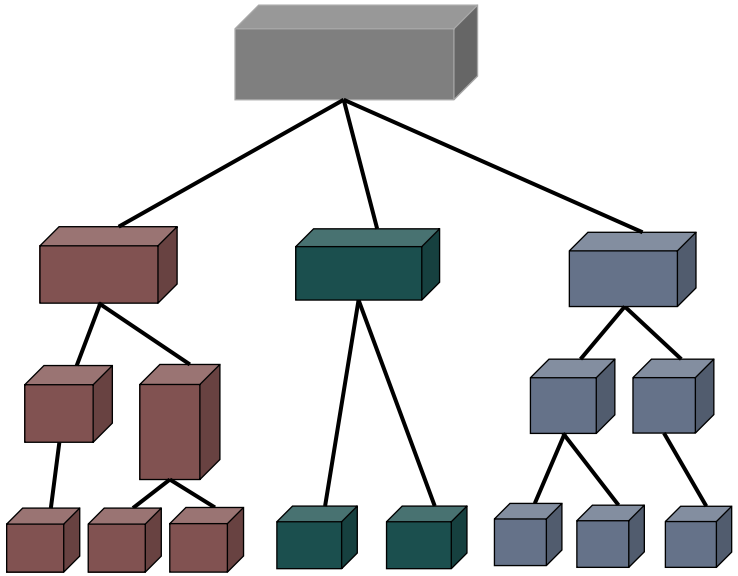


Category 3

Challenge: Scale with an increasing number of categories

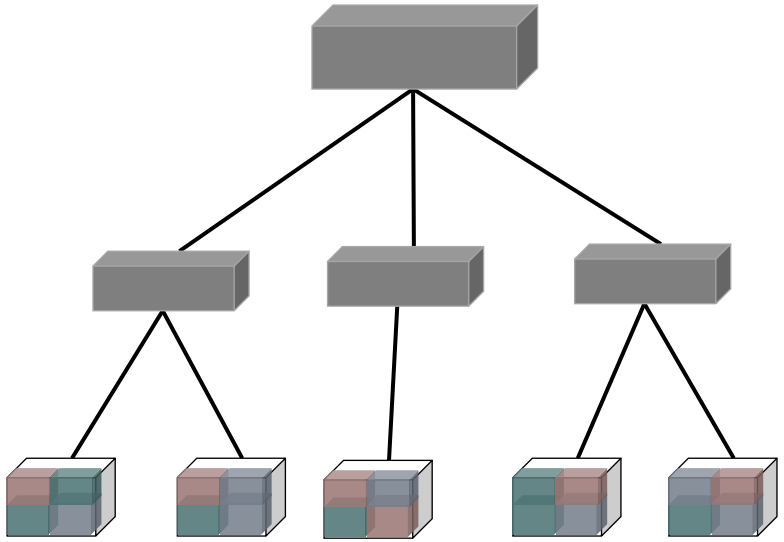
Shortcomings of Existing Techniques

One index per category



Independent search space

Index over union

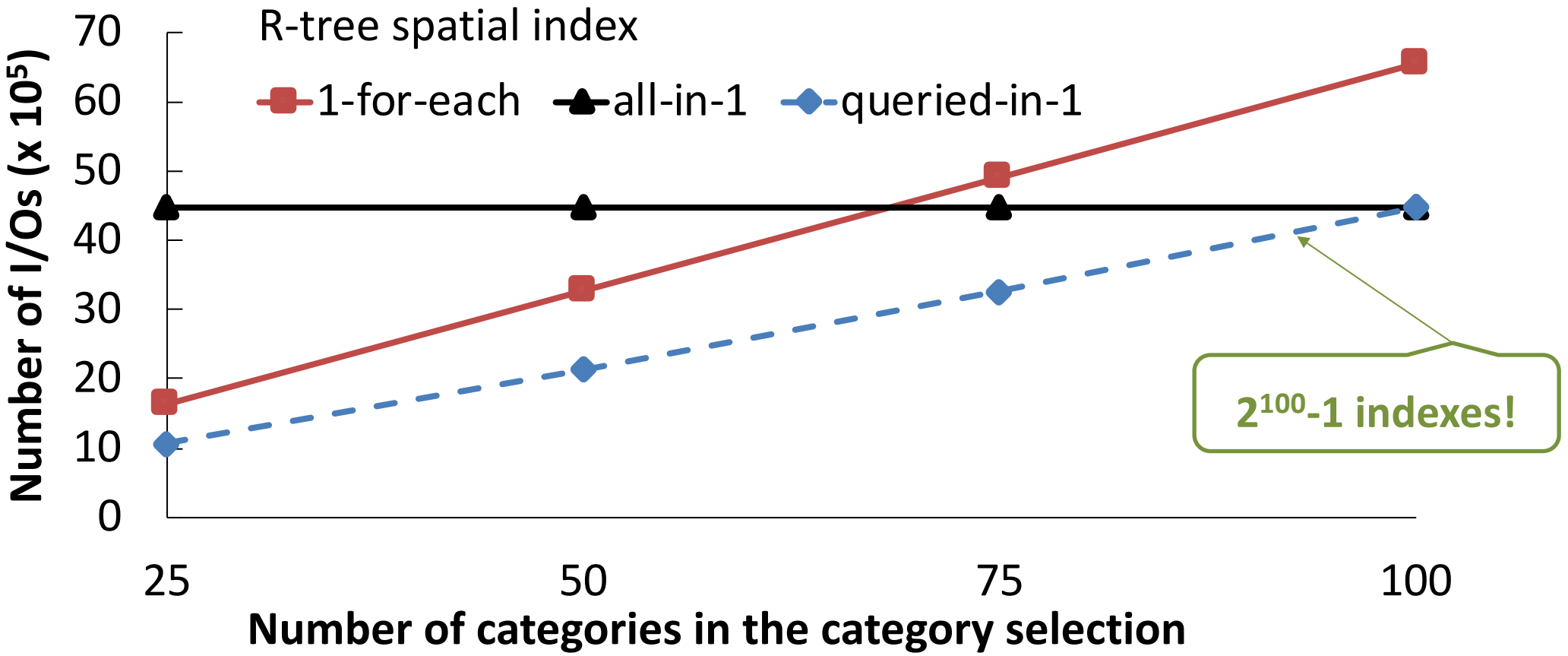


Common search space

Unnecessary data retrieval

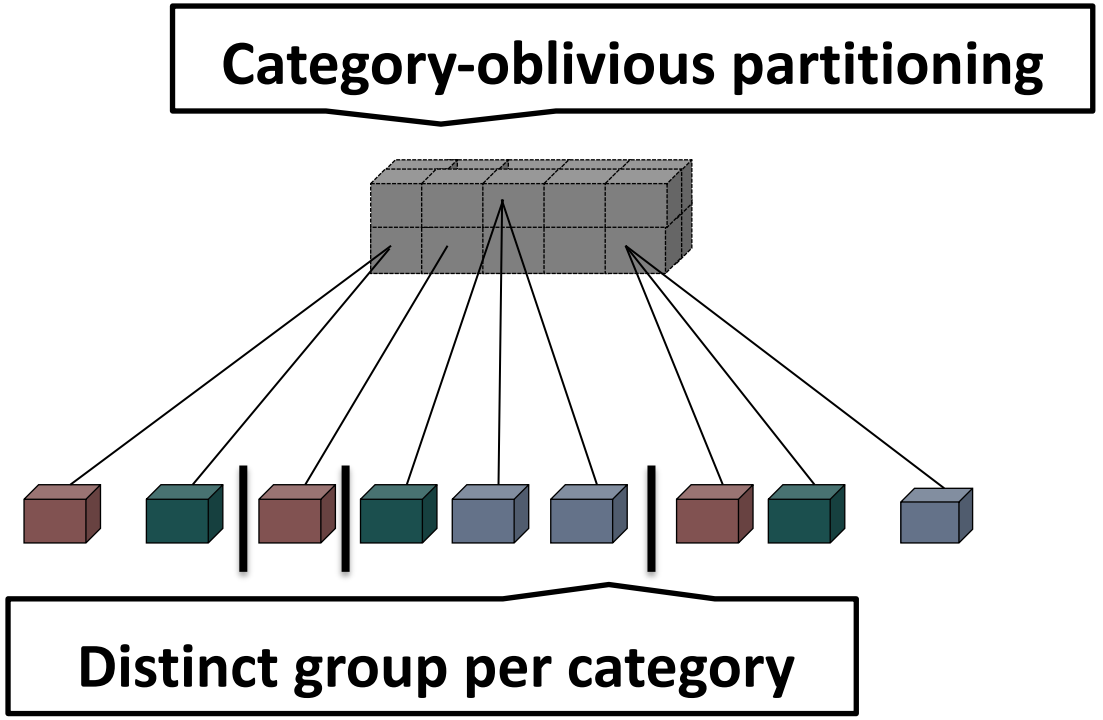
Existing Techniques in Action

Category size: ~1GB, 100 categories, 200 range queries, Query size: 10^{-3} % of total volume



Goal: queried-in-1 performance for ad-hoc category selection

Category-Aware Spatial Data Organization

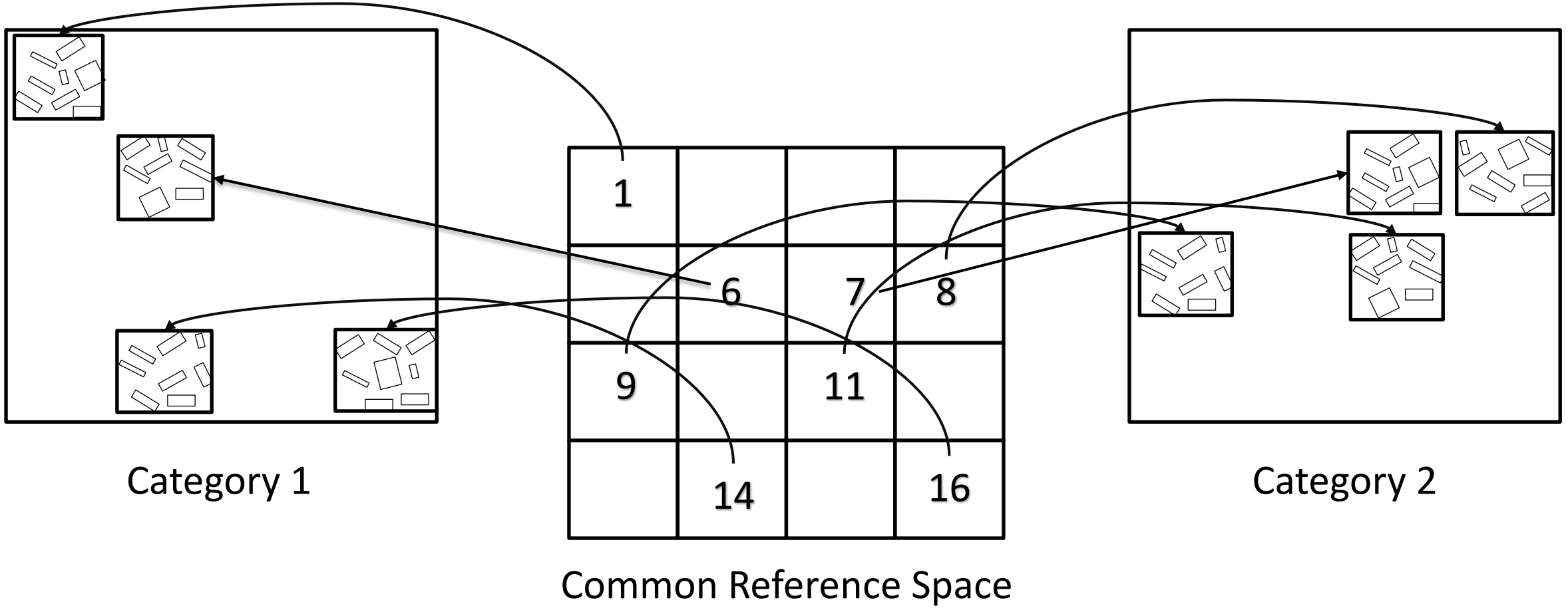


Common search space

Access to specific categories

STITCH Bundled Index

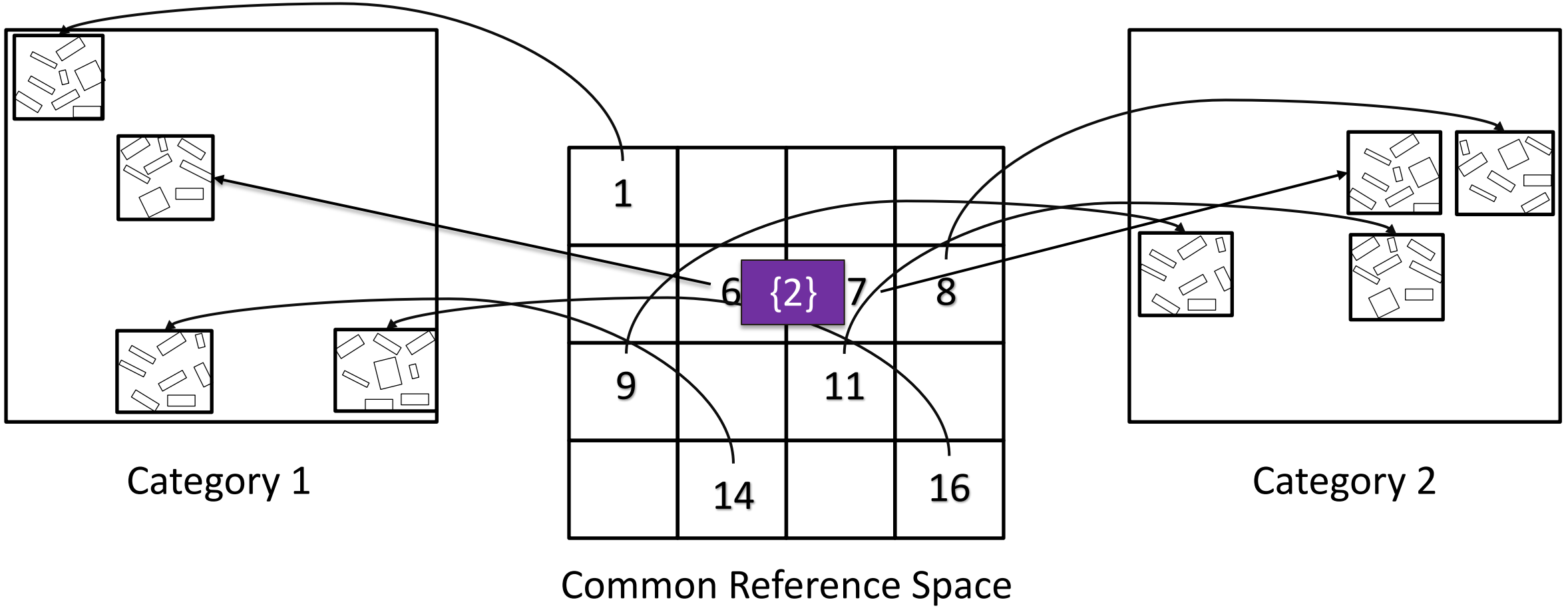
Link intersecting partitions



Combination of data-oriented partitioning with space-oriented indexing

STITCH Query Execution

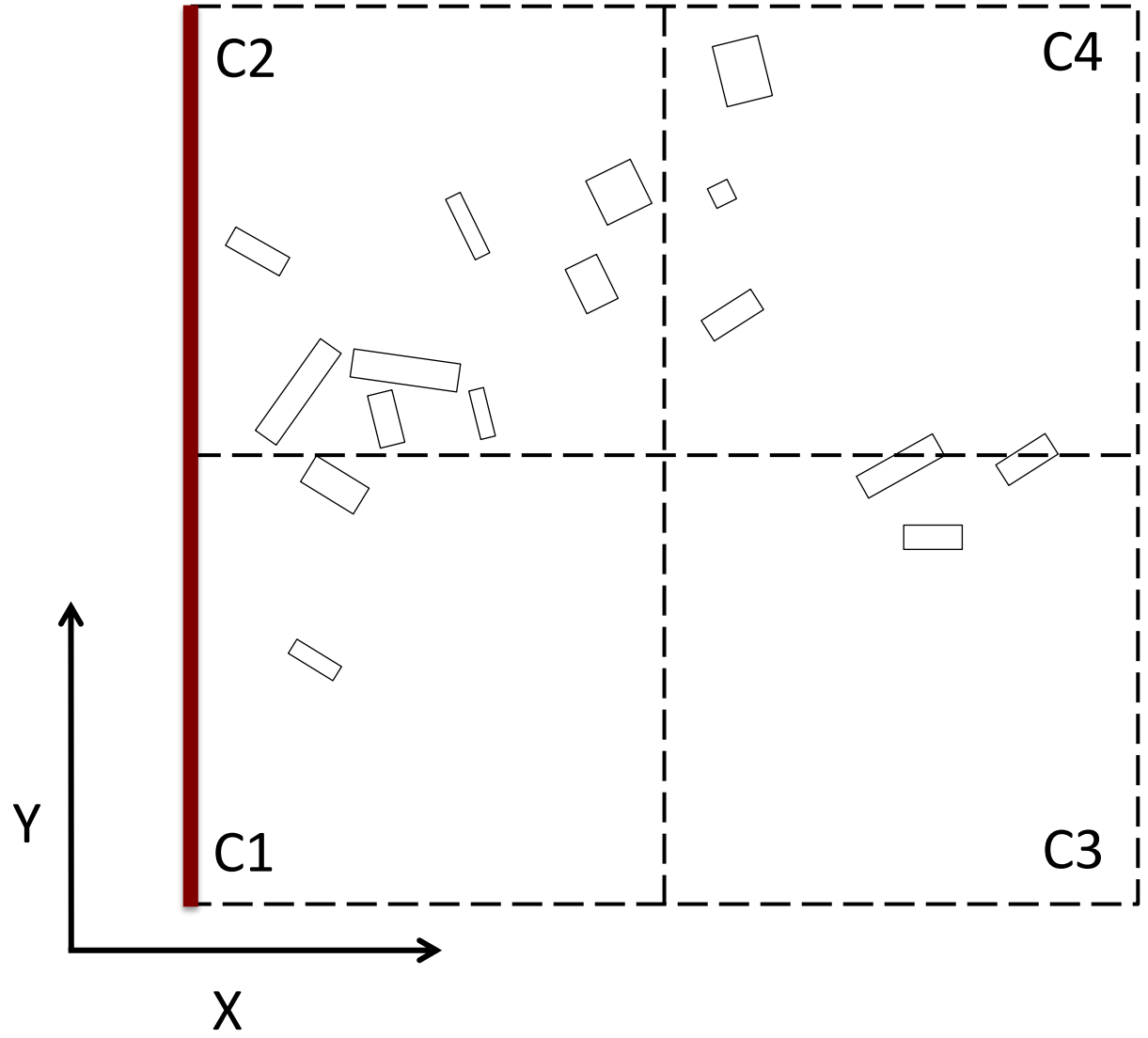
Link intersecting partitions



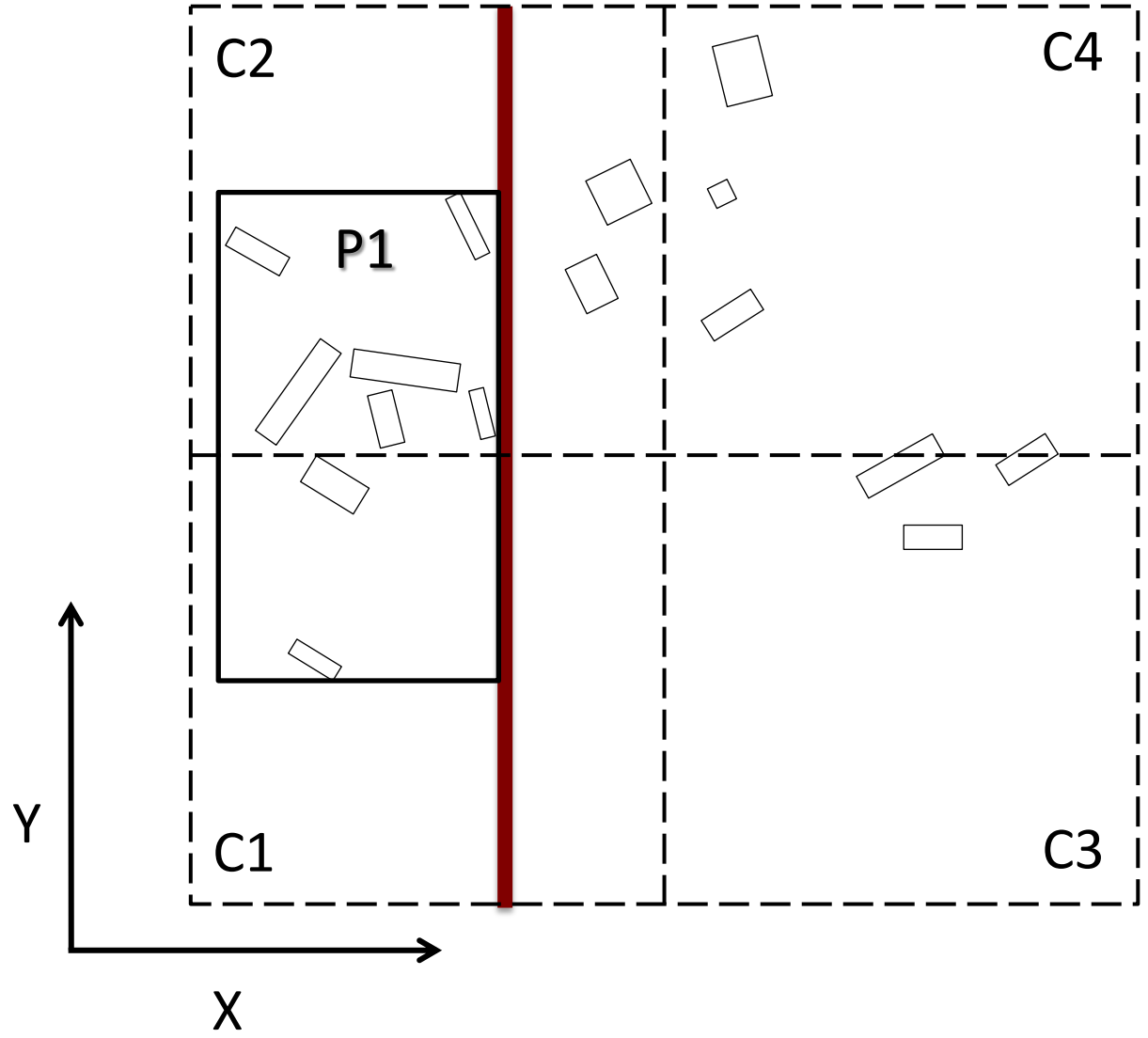
Grid: Locate the spatial region once

Links: Prune irrelevant categories

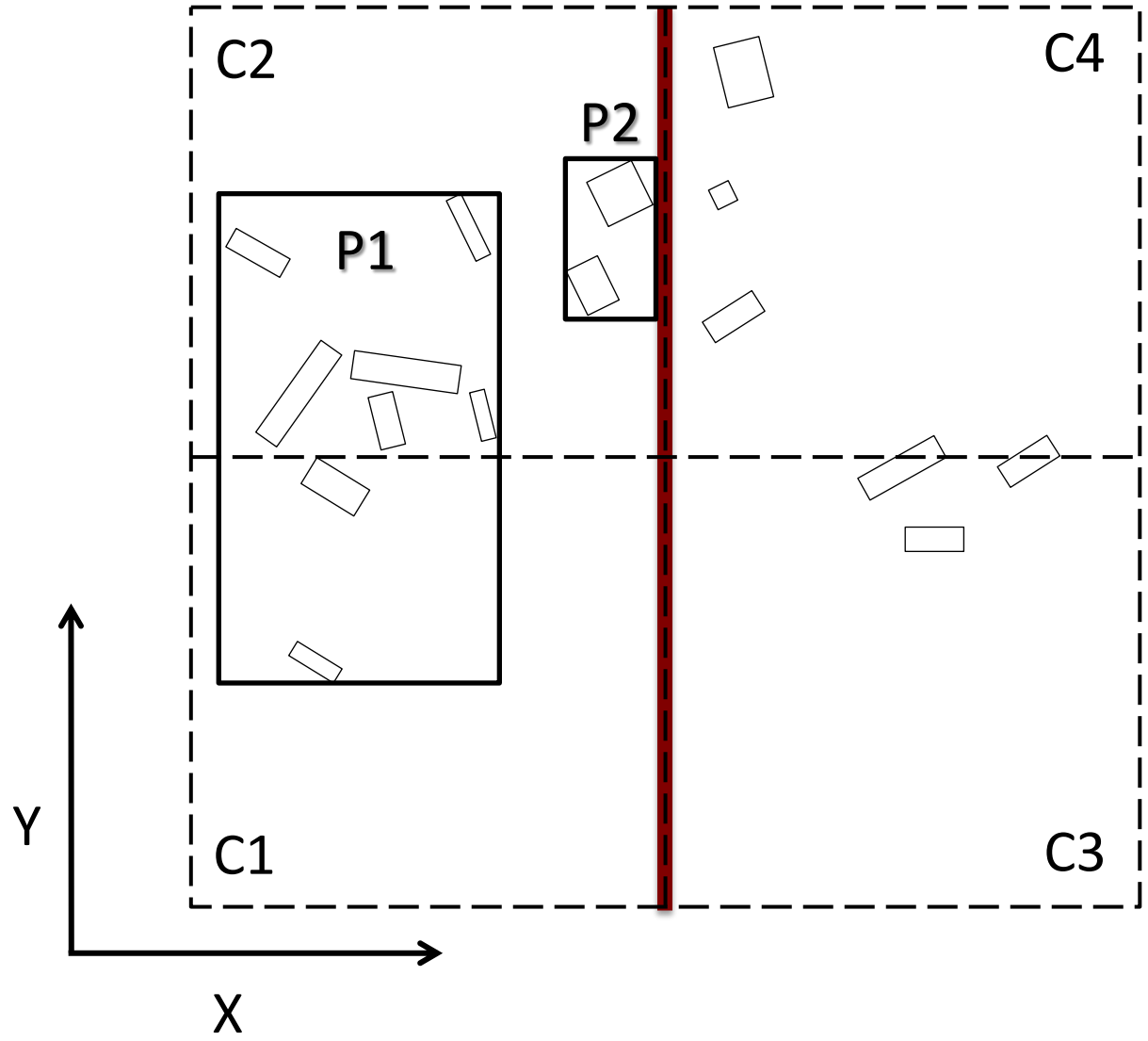
Sliced Data-Oriented Partitioning: I. X Dimension



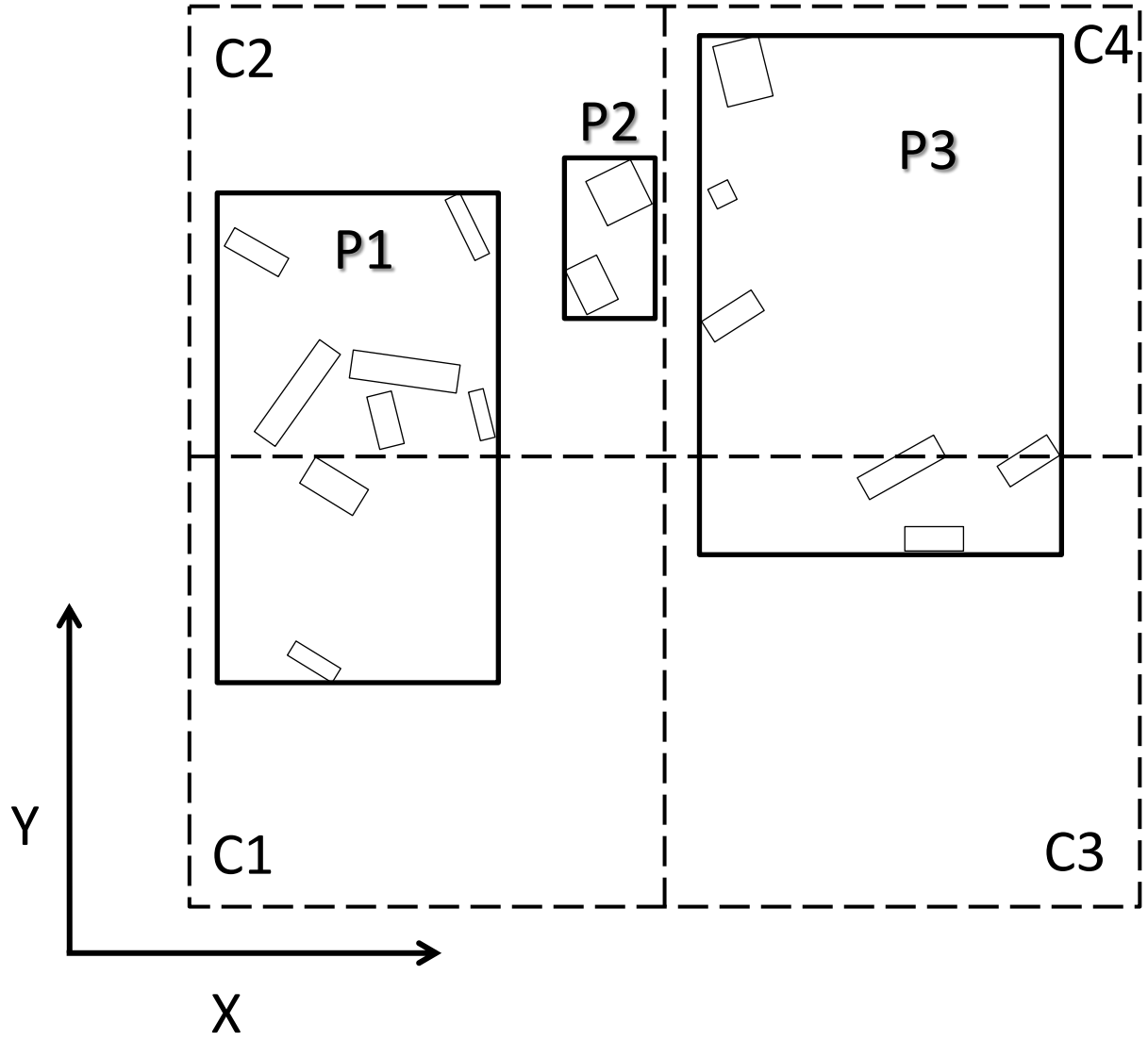
Sliced Data-Oriented Partitioning:
I. X Dimension



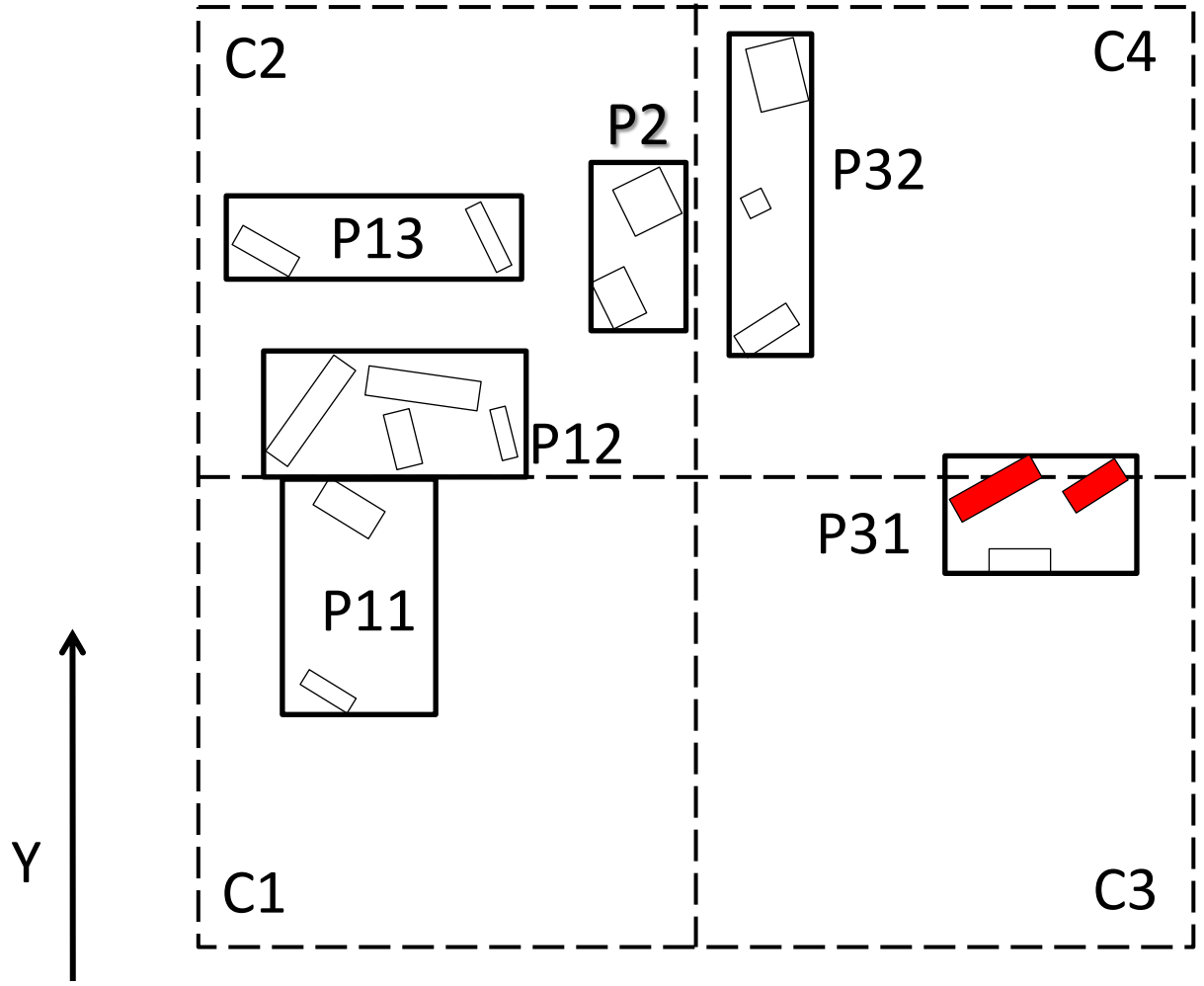
Sliced Data-Oriented Partitioning:
I. X Dimension



Sliced Data-Oriented Partitioning: I. X Dimension



Sliced Data-Oriented Partitioning: II. Y Dimension

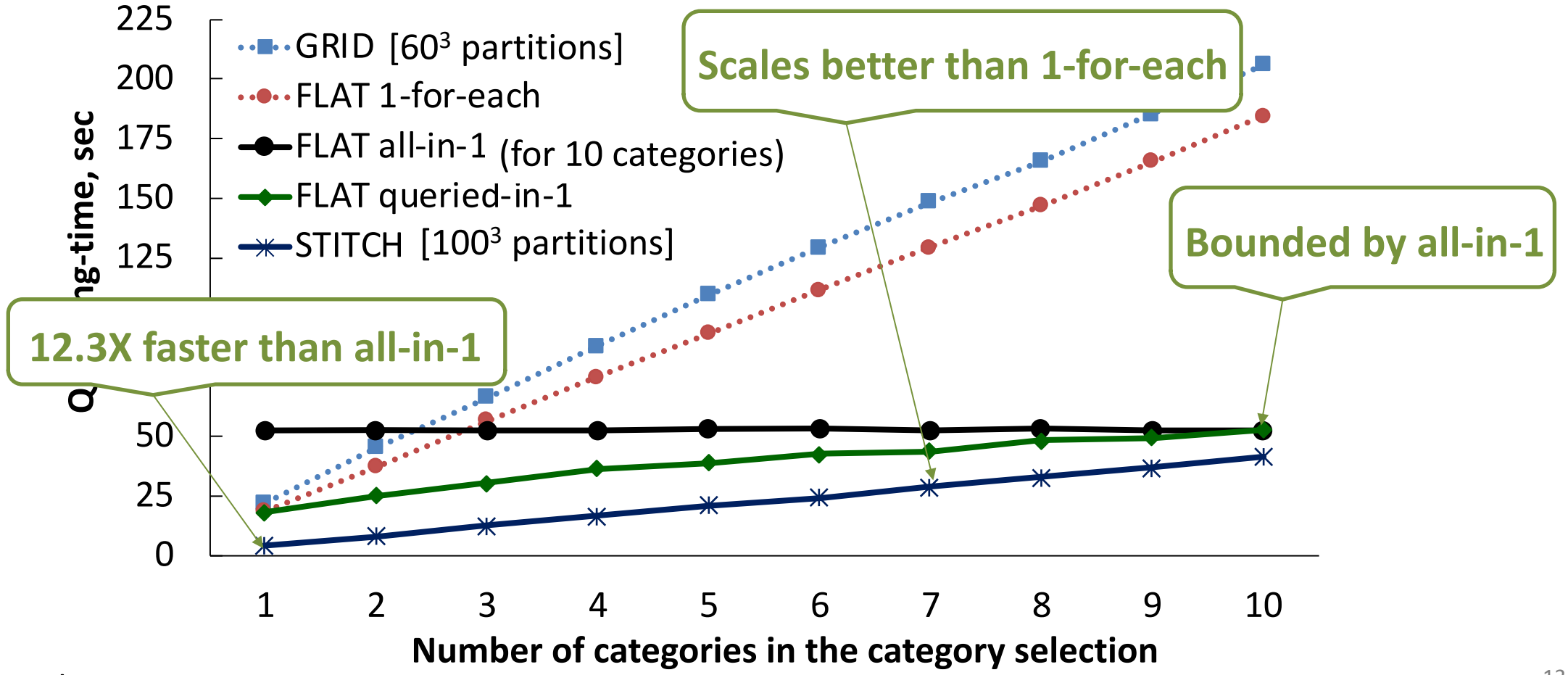


Grid-guided data-driven partitioning

Scaling with the Number of Queried Categories

Datasets: 3D triangular mesh, 10 neuron categories, ≤ 5GB each, 45GB in total.

Benchmark: 200 spatial range queries. Random aspect ratio, location and size (avg. 10^{-6}%).

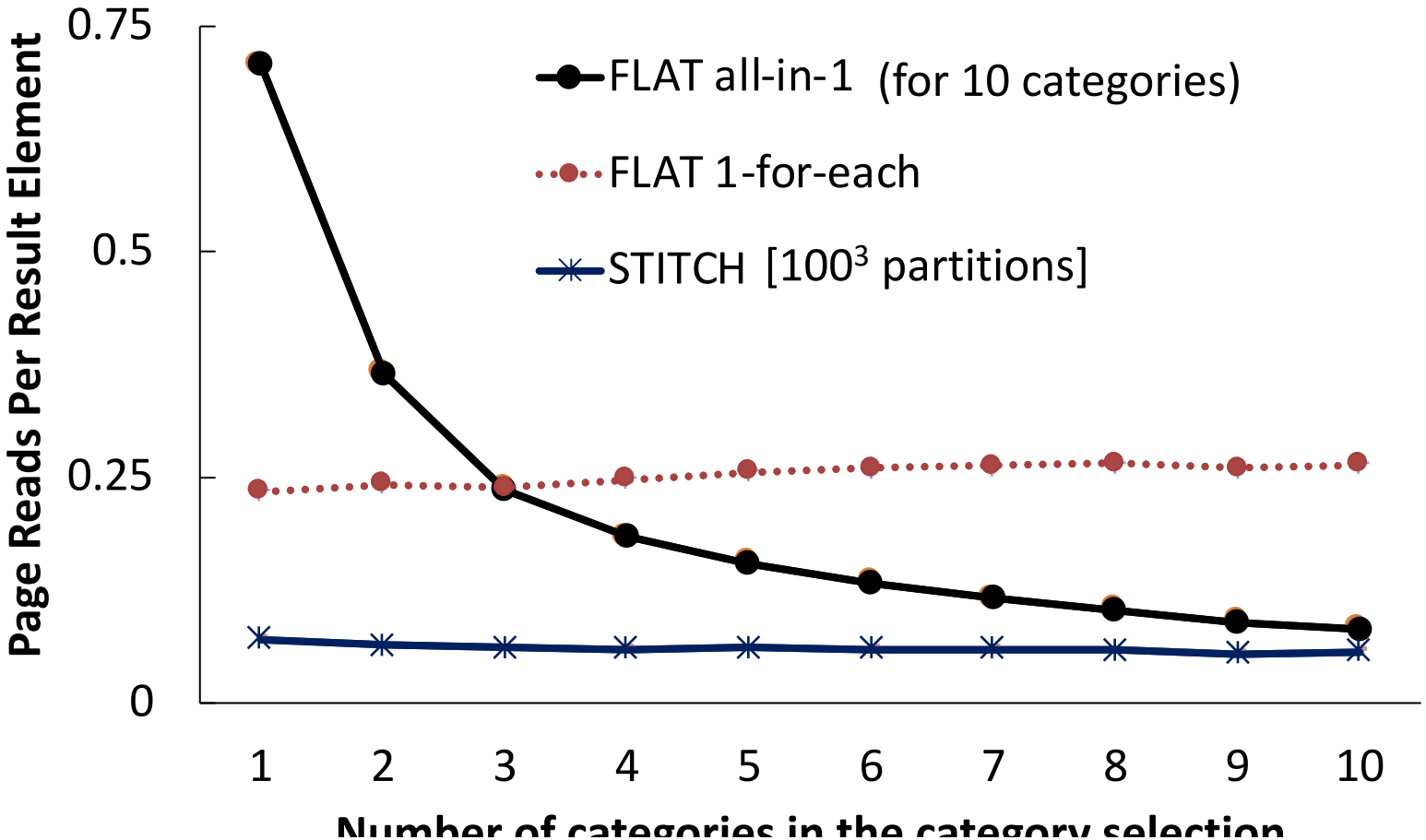


Hardware: Intel Xeon @2.8GHz, 48GB RAM

Overhead Analysis

Datasets: 3D triangular mesh, 10 neuron categories, ≤ 5GB each, 45GB in total.

Benchmark: 200 spatial range queries. Random aspect ratio, location and size (avg. 10^{-6}%).



STITCH reduces the amount of unnecessary data retrieved

Scalable Exploration of Multiple Spatial Datasets

- Single physical index → Common search space
- Hybrid partitioning scheme → Access to selected datasets
- Scalable performance for ad-hoc dataset selection

Thank You!