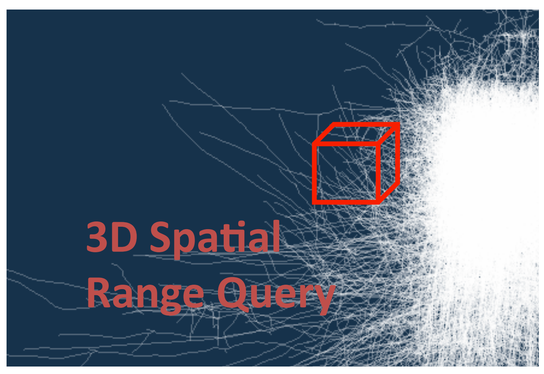
Data-driven Neuroscience: Enabling Breakthroughs Via Innovative Data Management

Alexandros Stougiannis, Mirjana Pavlovic, Farhan Tauheed, Thomas Heinis, Anastasia Ailamaki

Indexing Dense Spatial Datasets

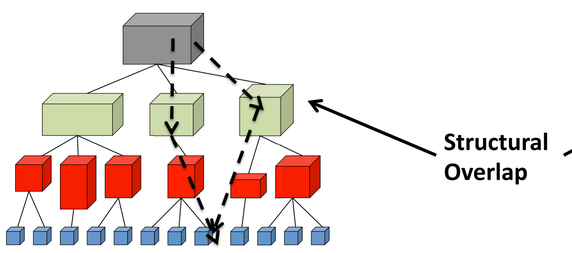


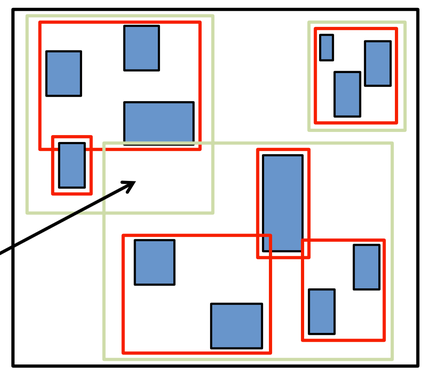
3D Neuron Model

Dataset: Rat Neocortex
Model Size: 1692 Neurons
Dataset Size: 12.5 GB

GOAL: Simulate Human Brain
Model Size: 86 Billion Neurons
Expected Dataset Size: **606 PB**

Spatial Index: R-Tree: Hierarchy of Minimum Bounding Rectangles **(MBR)**

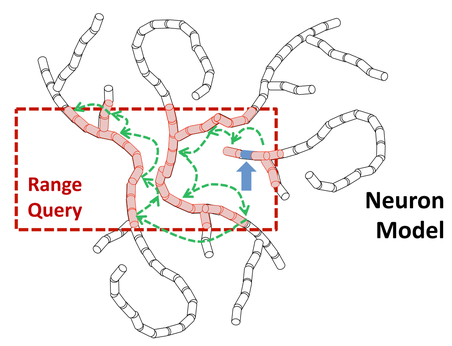




Increase in Spatial Data Density => More Overlap

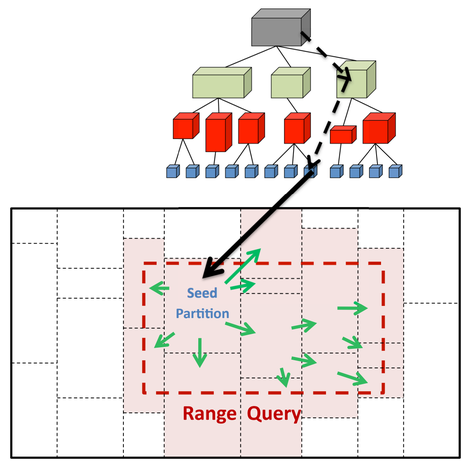
FLAT: Use Connectivity to Avoid Overlap

SEEDING: Find any one object
CRAWLING: Traverse neighbors

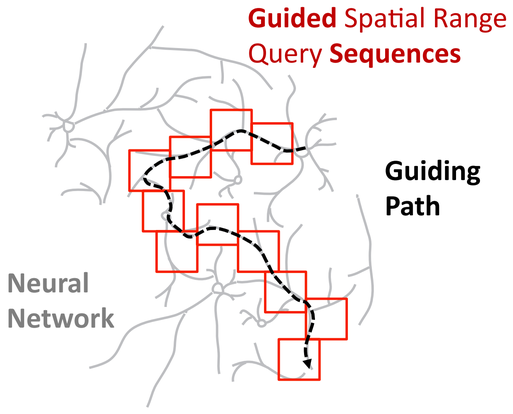


SEEDING PHASE:
Uses R-Tree
Picks any child node in case of overlap

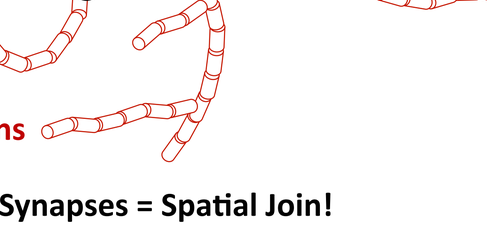
CRAWLING PHASE:
Uses Graph Traversal
Based on Linked Partitions of Data



Prefetching for Spatial Sequences



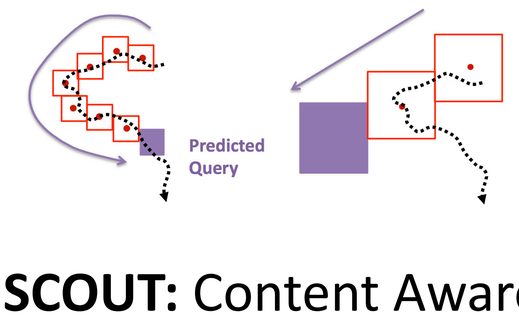
Interactive Analysis



Effective Prefetching Requires Accurate Prediction

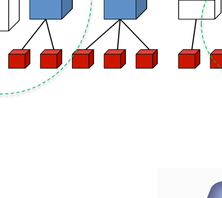
Existing Approaches:

Trajectory Extrapolation Does Not Work Well With Irregular Paths in Neural Networks



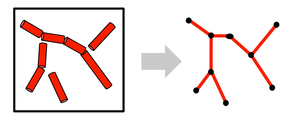
SCOUT: Content Aware Prefetching Approach

Examine: Query Results
Identify: Guiding Path
Predict: Using Guiding Path



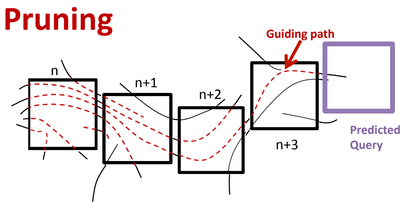
Graph Summarization:

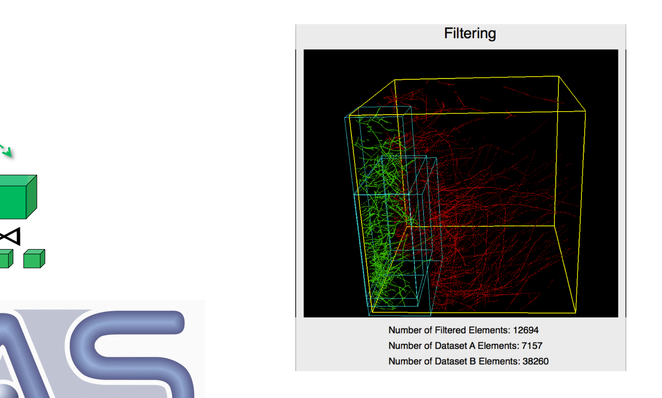
G(V,E) : **V** = spatial objects,
Edge between close objects.



Iterative Candidate Pruning

Neural Structures
Candidate Set





In-Memory Spatial Join

Locating Synapses = Spatial Join!

TOUCH: Hierarchical Data Oriented Partitioning

1) **Avoid replication through data-oriented partitioning**
2) **Avoid overhead of overlap through direct assignment and batched join**
3) **Use filtering**

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Blue Brain Project

DATA-INTENSIVE APPLICATIONS AND SYSTEMS