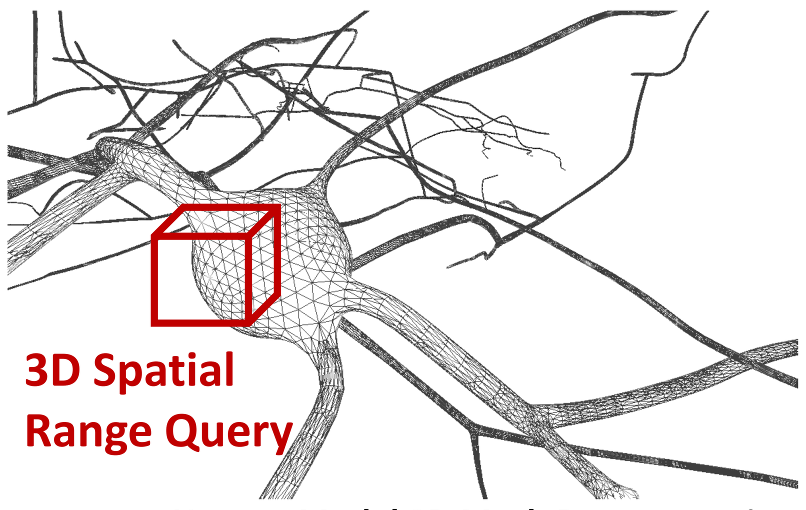
# Accelerating Spatial Range Queries

Alexandros Stougiannis, Farhan Tauheed,
Thomas Heinis, Anastasia Ailamaki

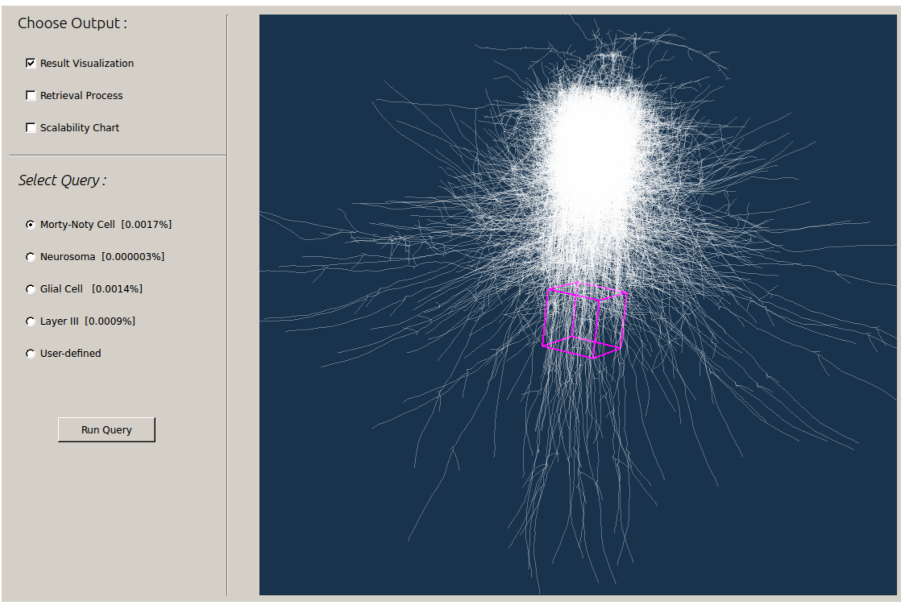
## Spatial Analysis

Blue Brain Project: simulates brain tissue by building massive neural **spatial models**.

Analyzing Neuron Models require Efficient **Spatial Range Query** Execution




Neuron Model 3D Mesh Representation

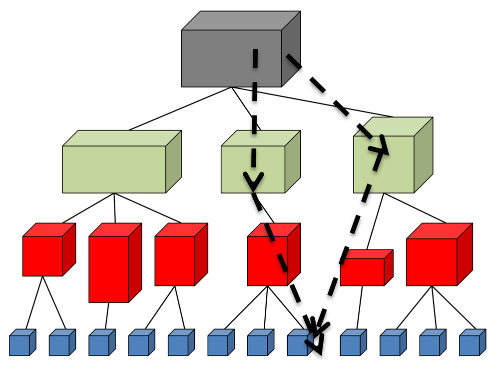


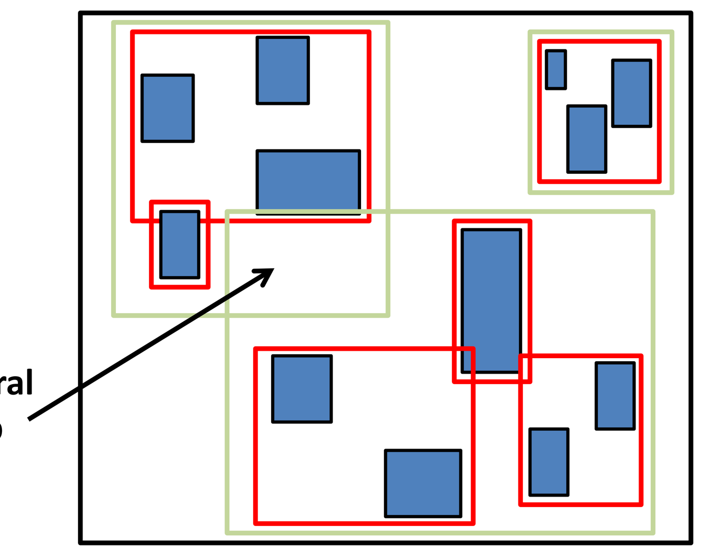
**Dataset: Rat Neocortex**
Model Size: 1692 Neurons
Dataset Size: 12.5 GB

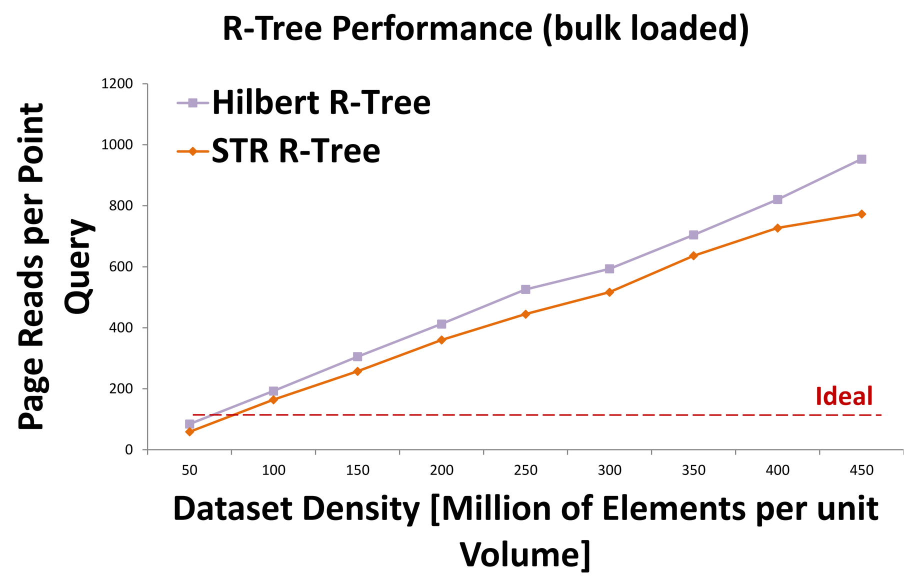
**GOAL: Simulate Human Brain**
Model Size: 86 Billion Neurons
Expected Dataset Size: **606 PB**

## State of the Art

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





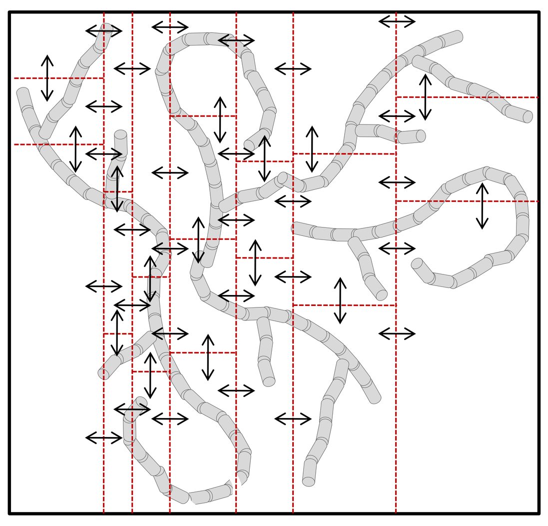




**Increase in Spatial Data Density => More Overlap**

**Point Query Analysis:**
R-Tree variants do not scale with data density

## FLAT Algorithm

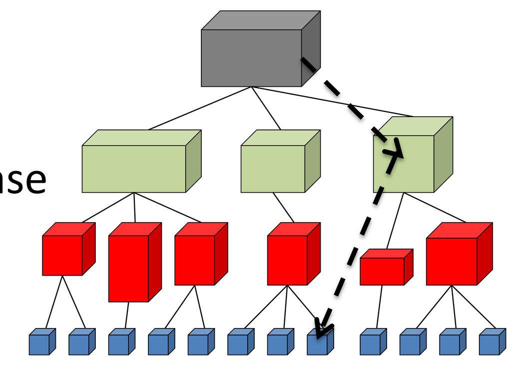
**Two Phase Query Execution**
1) **SEEDING:** Find any one object arbitrarily inside the query region.
2) **CRAWLING:** Retrieve remaining results by traversing the neighbors.



**SEEDING PHASE:** Use R-Tree

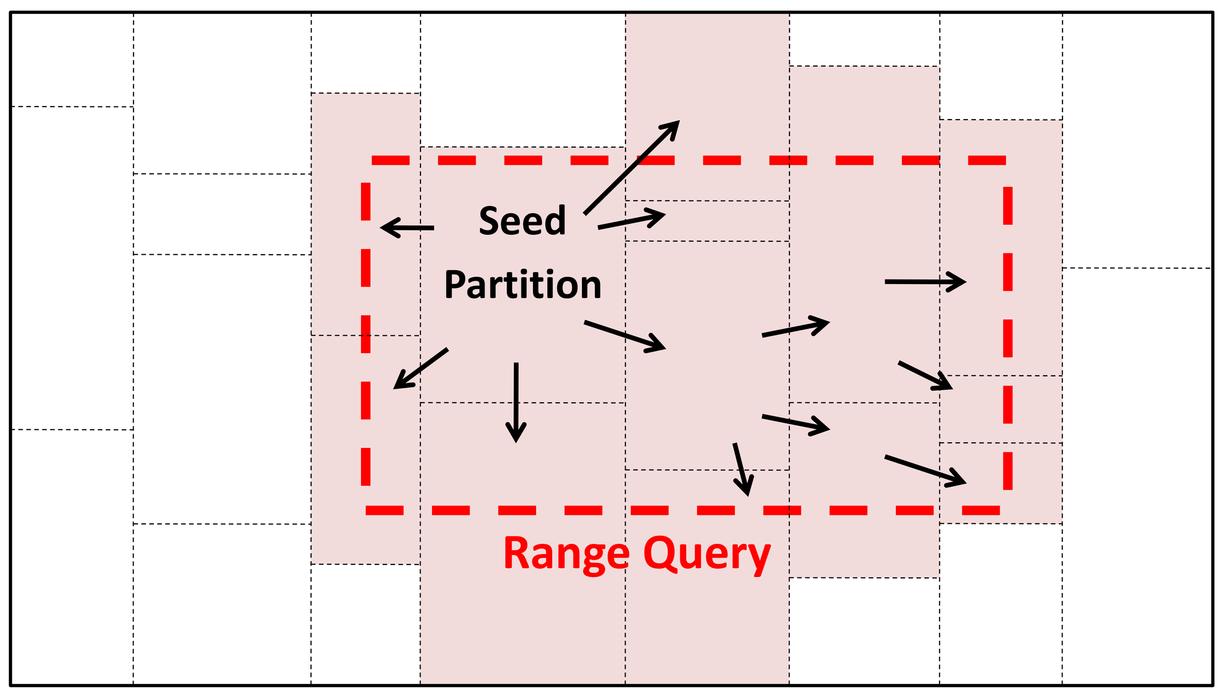
**Range Query**: Find **ALL** objects inside query
**Seed Query**: Find **ANY ONE** objects inside query

picks one child arbitrarily, in case of structural overlap



**Seeding** requires I/O equal to height of tree.

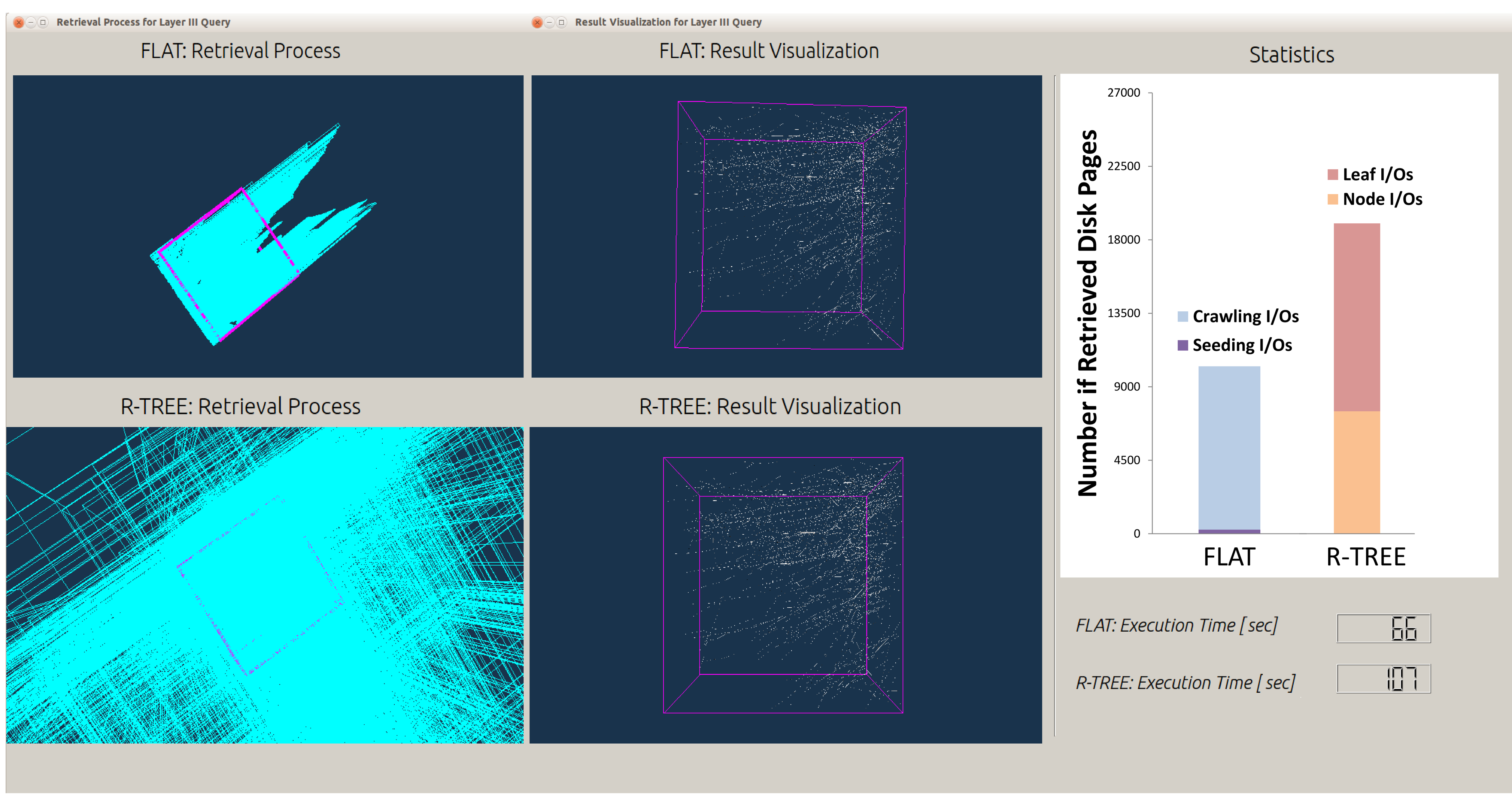
**CRAWLING PHASE:** recursive graph traversal starting from the seed partition




**Index Construction**
1) **Partitioning**: Recursive tiling to group spatial close objects together.
2) **Linking:** Connect neighboring partitions together.

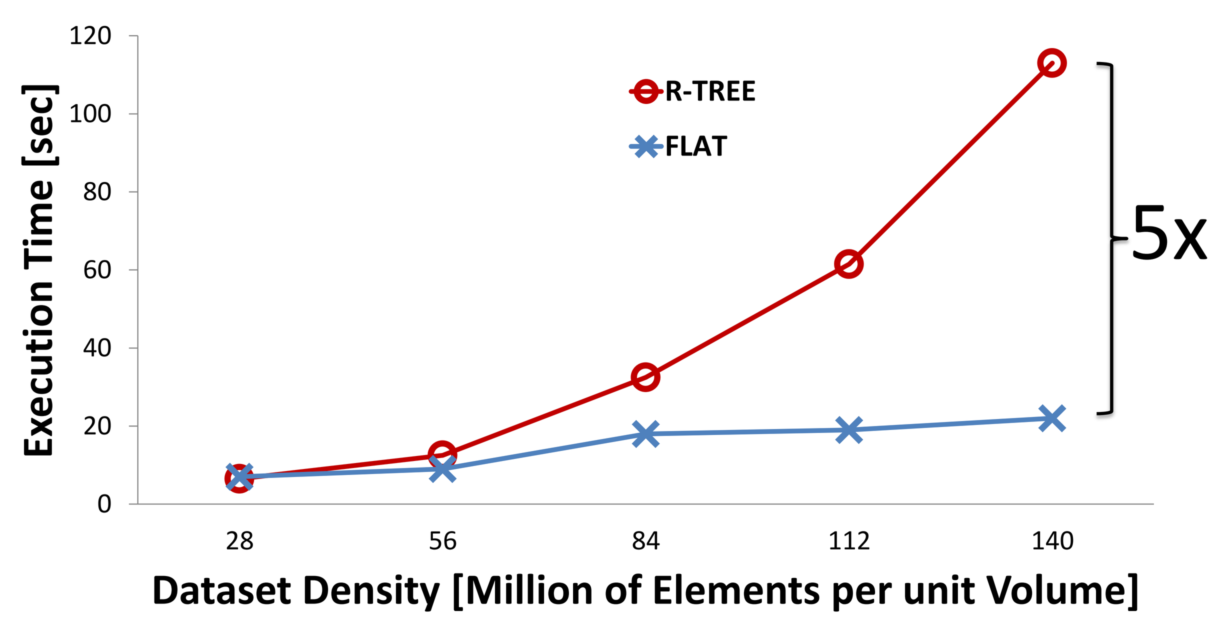
→ Linear complexity in terms of graph edges and vertices
→ Performance depends on the selectivity rather than density

## FLAT Performance




**Scalability:**

**Range Queries:** Morty-Noty Cell Query
**Measure:** Query Execution Time as a function of dataset density




ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Blue Brain Project

DATA-INTENSIVE APPLICATIONS AND SYSTEMS