Improving Spatial Data Processing by Clipping Minimum Bounding Boxes
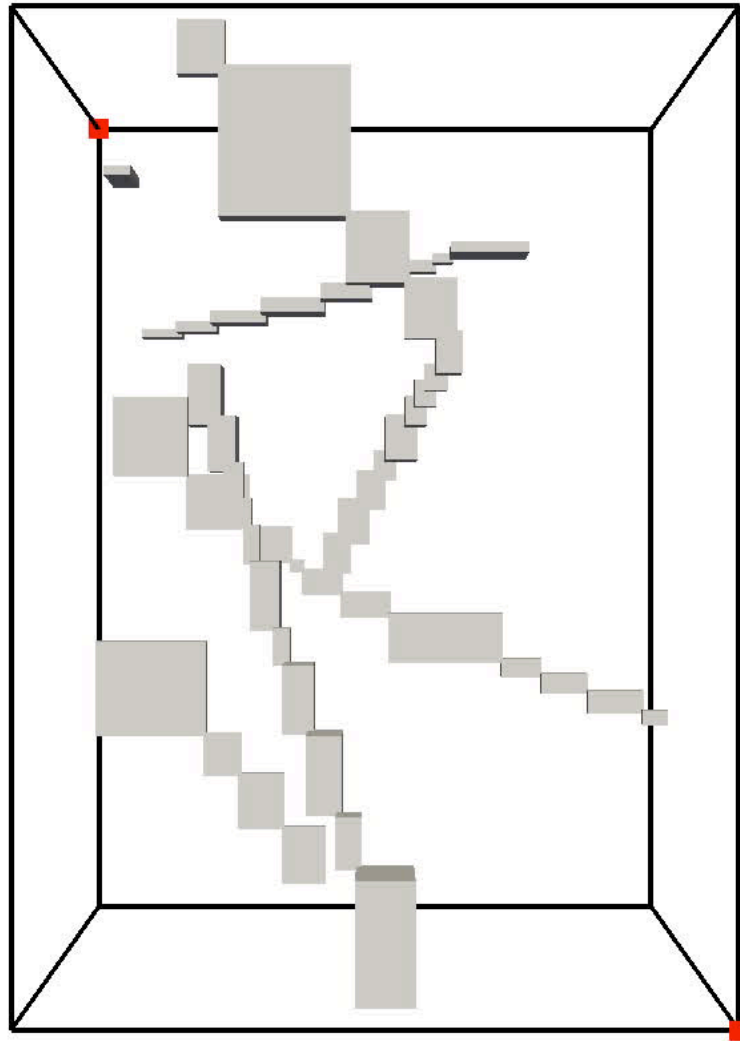
Darius Sidlauskas
EPFL

Sean Chester
NTNU

Eleni Tzirita Zacharatou
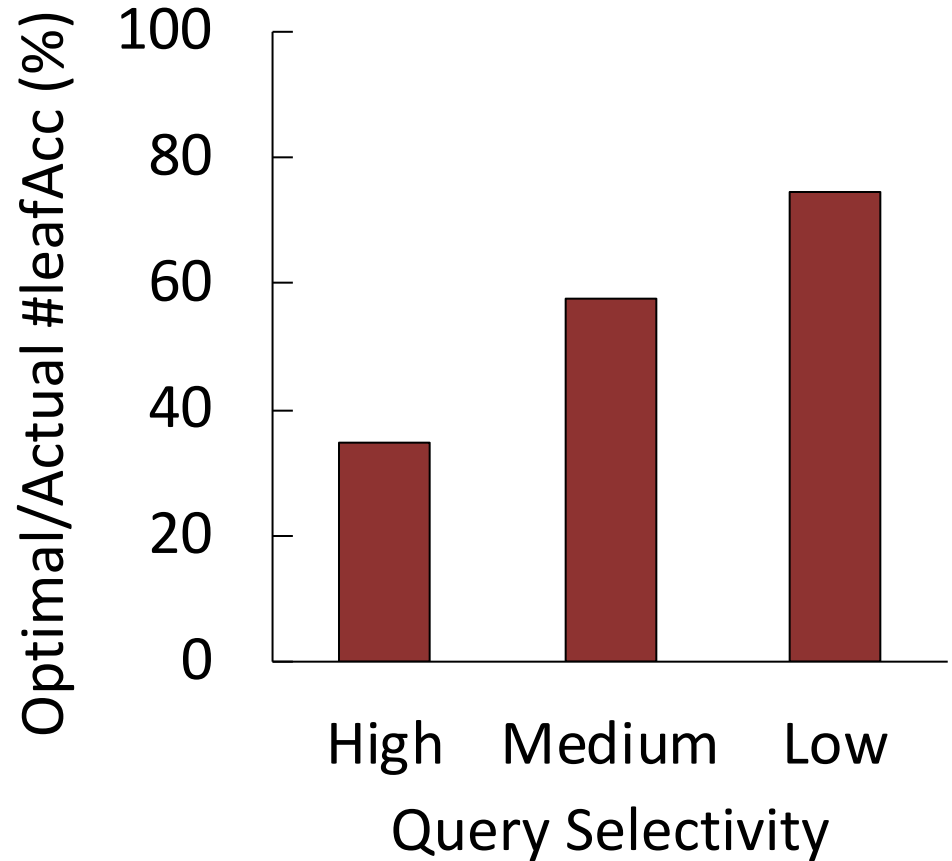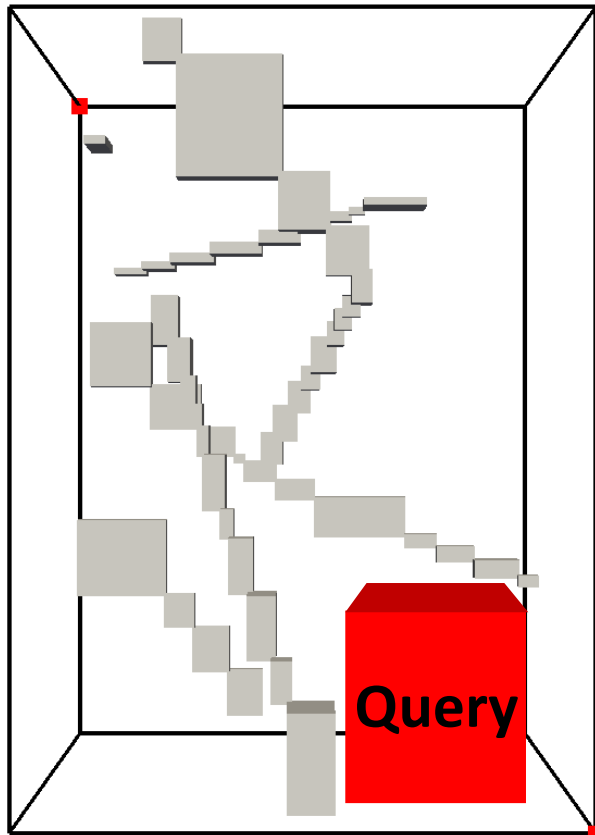EPFL

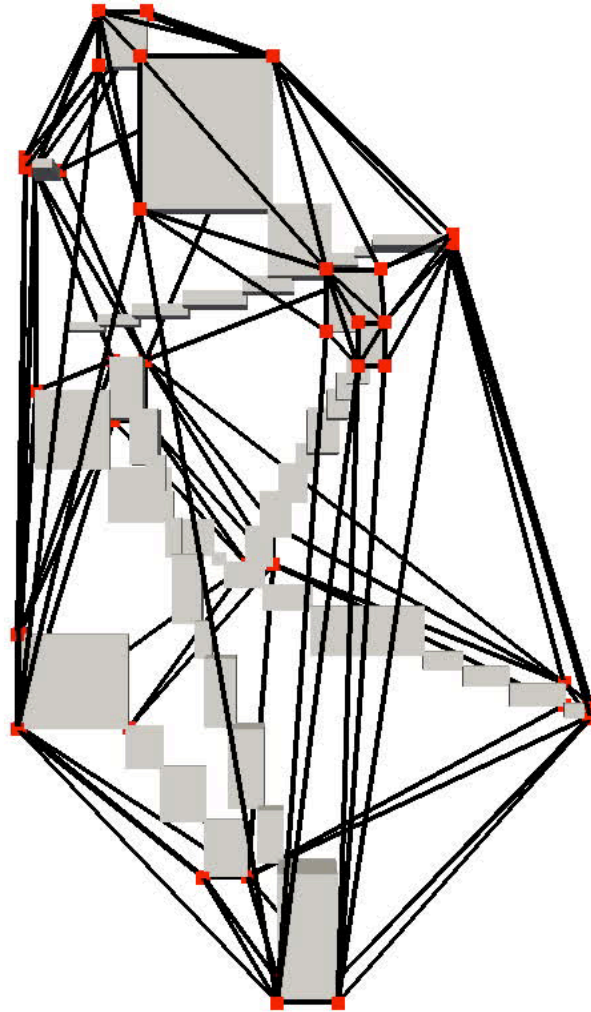Anastasia Ailamaki
EPFL

Brain model (axons)

97% of the Minimum Bounding Box is empty
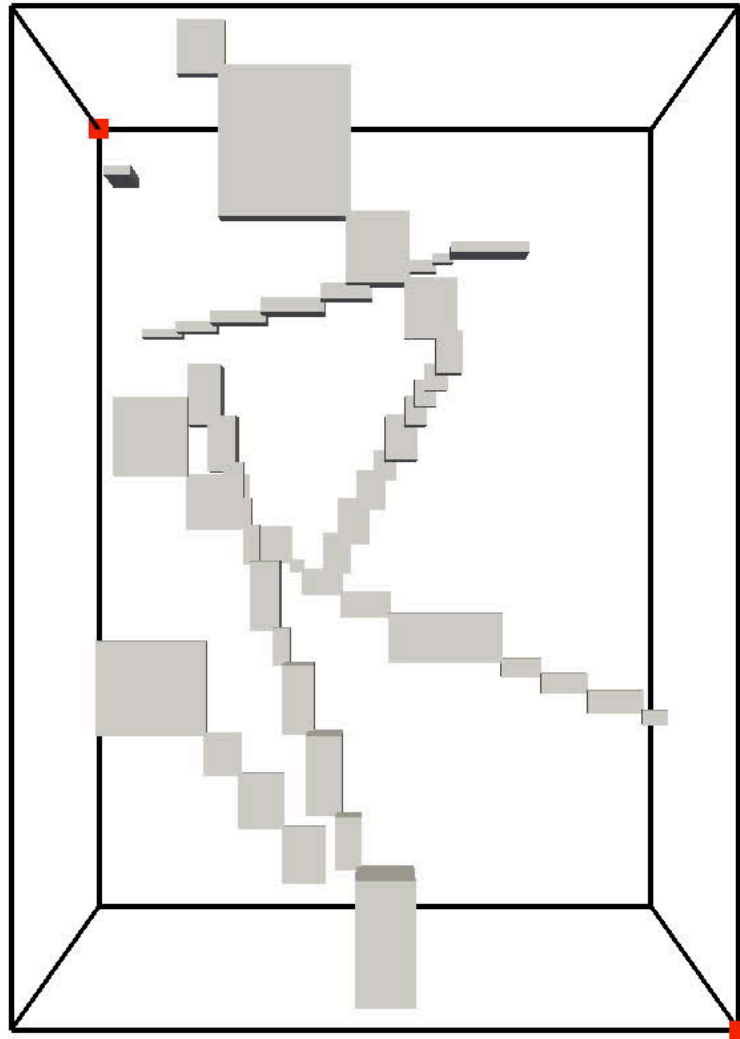
Empty space ➔ unnecessary I/Os

Up to 64% of the accessed leaf nodes are false hits
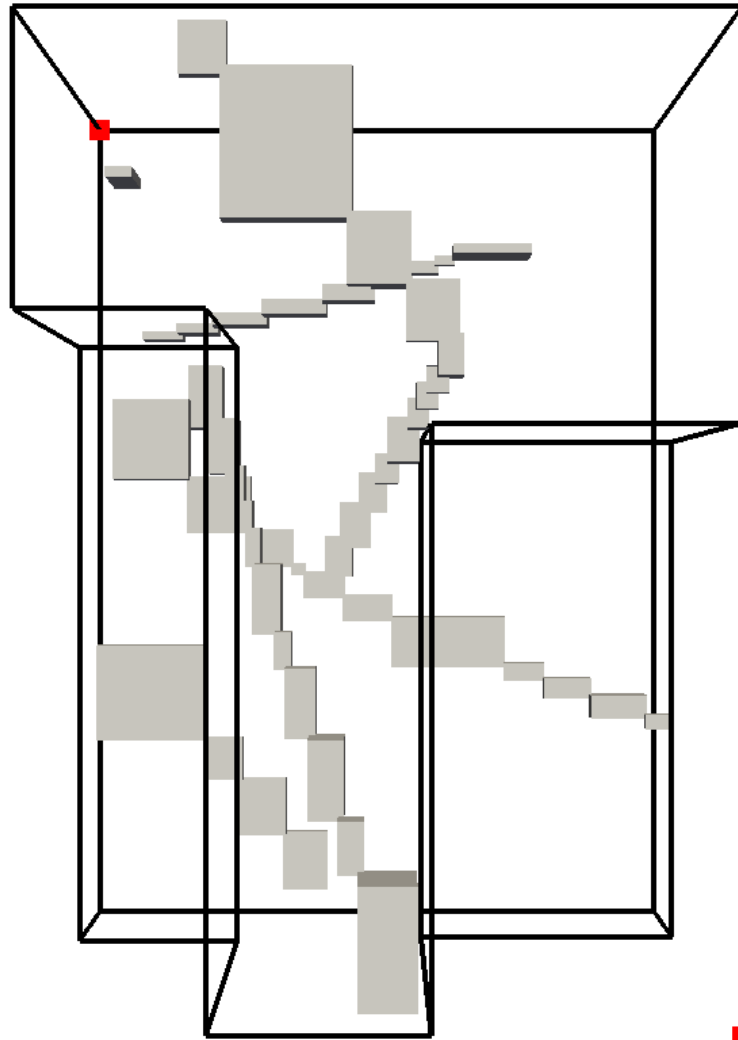
Tighter structure (convex hull)

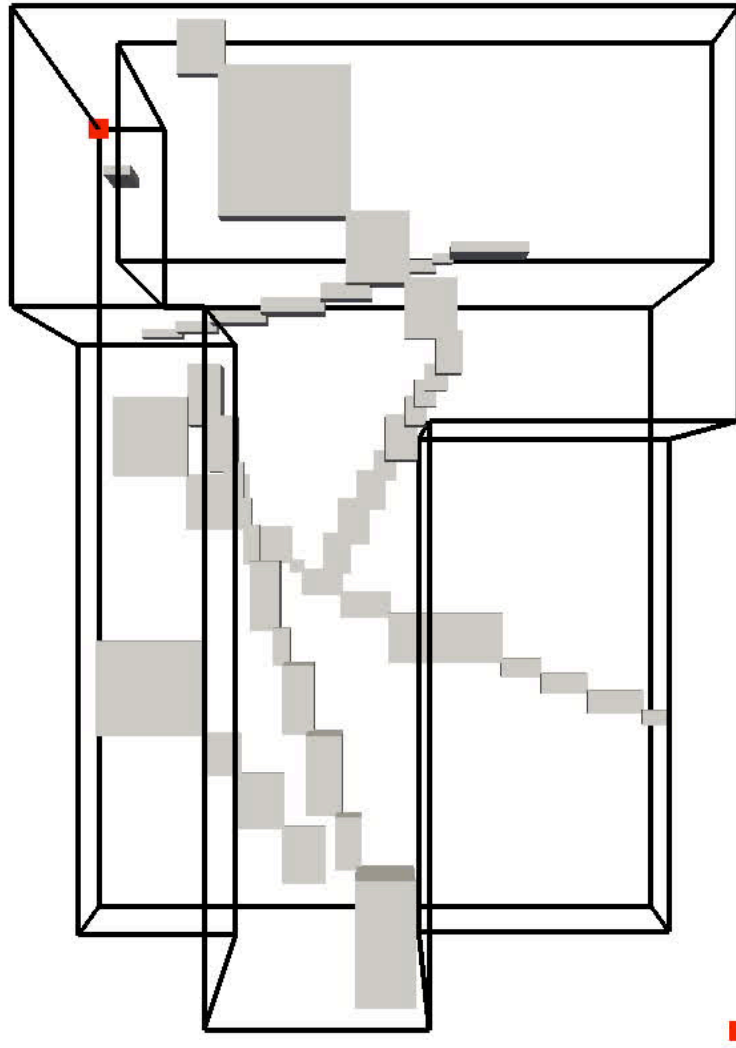

Empty space from 97% to 37%, but requires 49+ points

How to reduce dead space with only few extra points
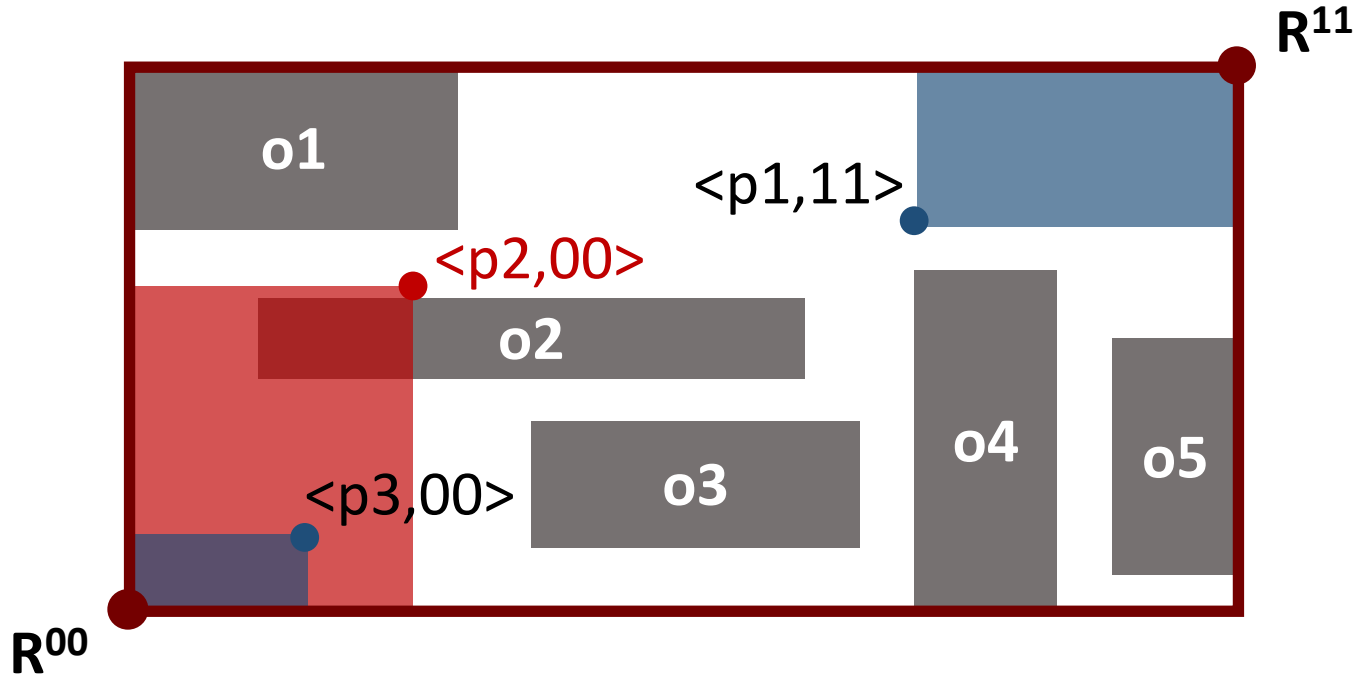
"Light cuts" using only few extra points

"Light cuts" using only few extra points

45% reduction in empty space with just 3 extra points

Clip point

- Relevant to a corner of the Minimum Bounding Box.
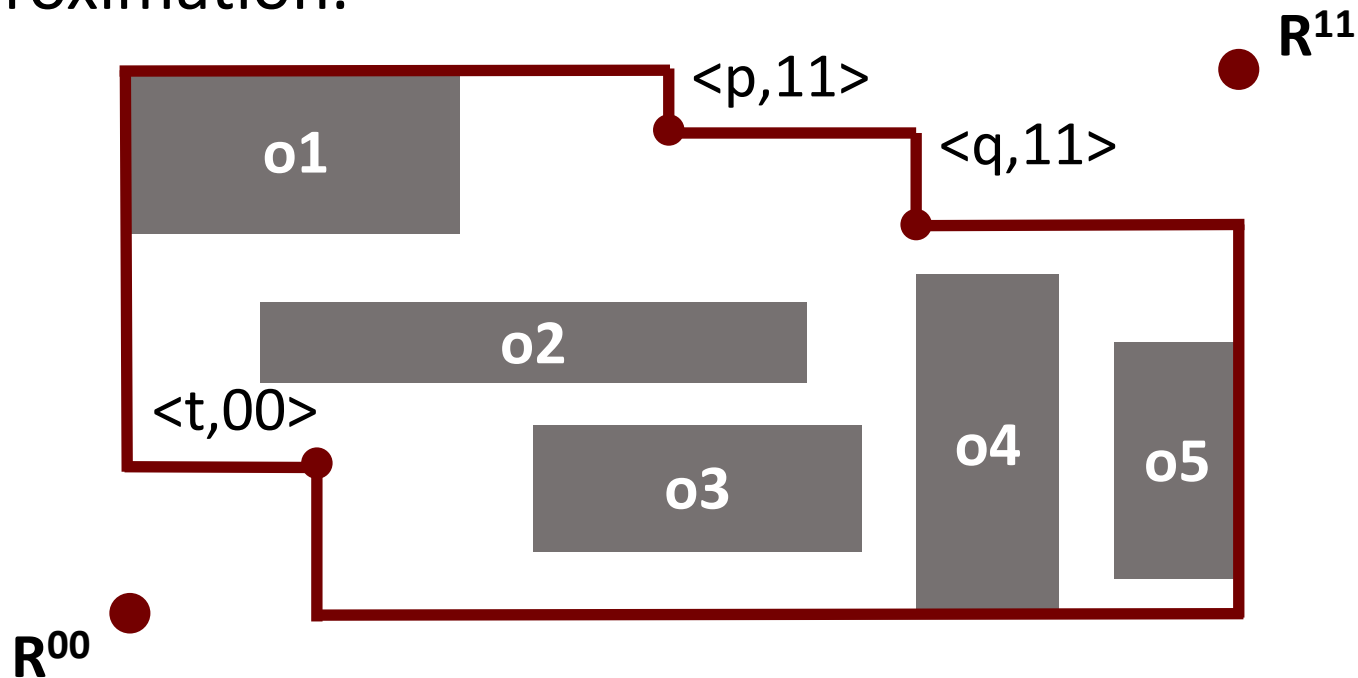- The rectangular area between the clip point and the corner is dead.

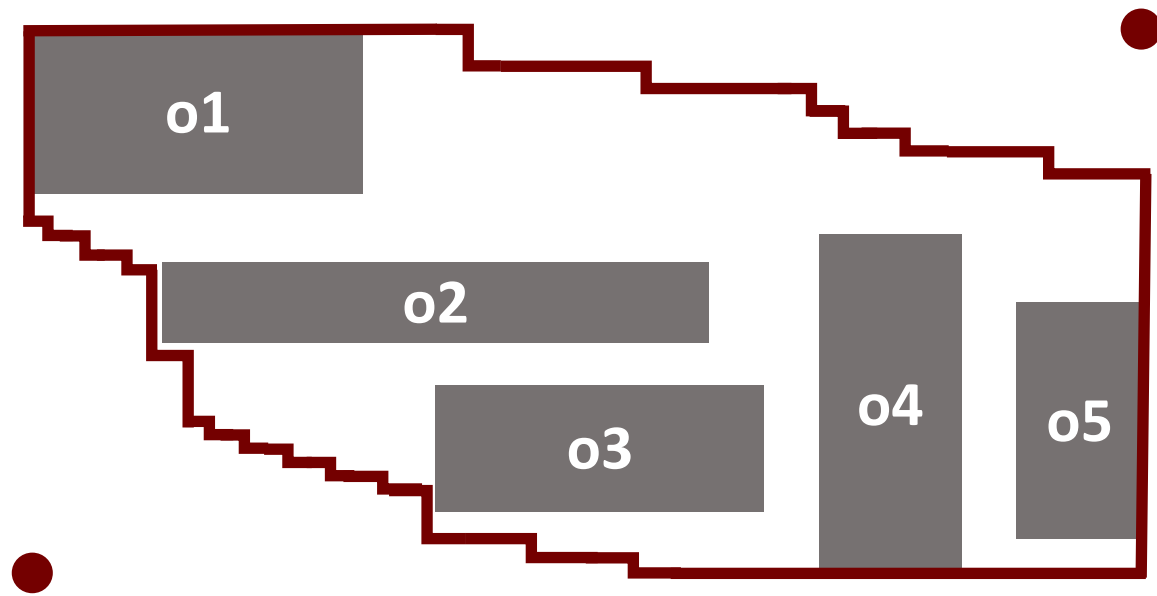**Low representation overhead for clipped areas**

Clipped Bounding Box (CBB)

- Augments the Minimum Bounding Box with a set of clip points.
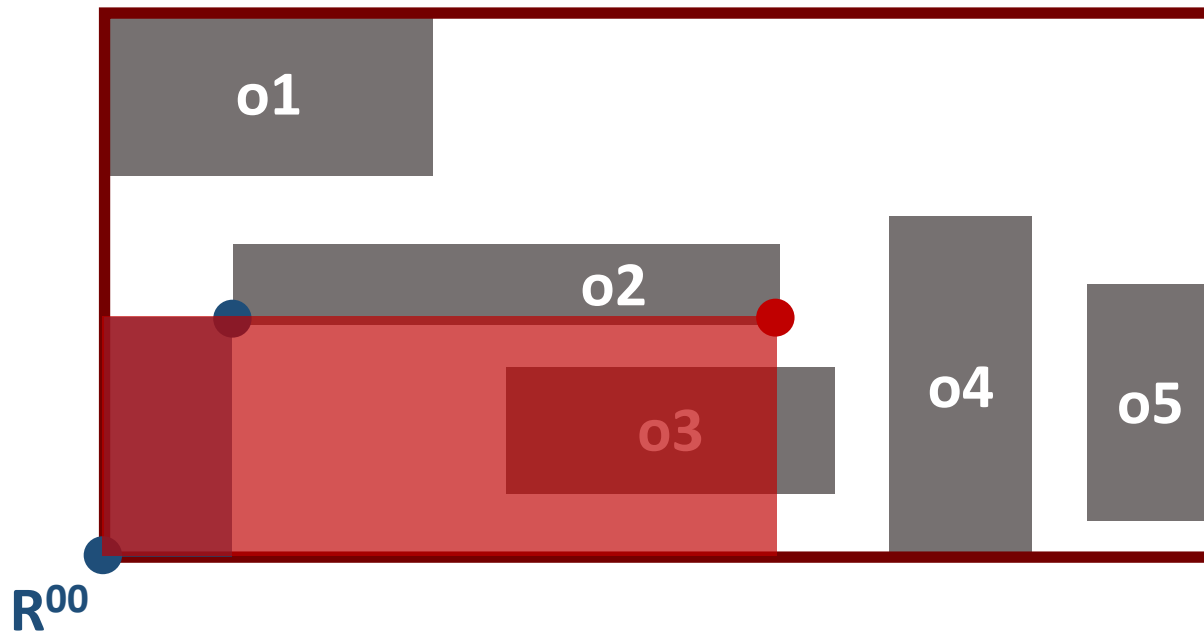- The lesser the retained volume, the better the approximation.

Challenge: Choice of clip points

Choose ≤ k clip points that maximize the eliminated volume

Candidate clip points

- For given corner R^b:
  - Consider only points in the outer surface of the objects o_i.
  - Consider only the closest corner o_i^b.
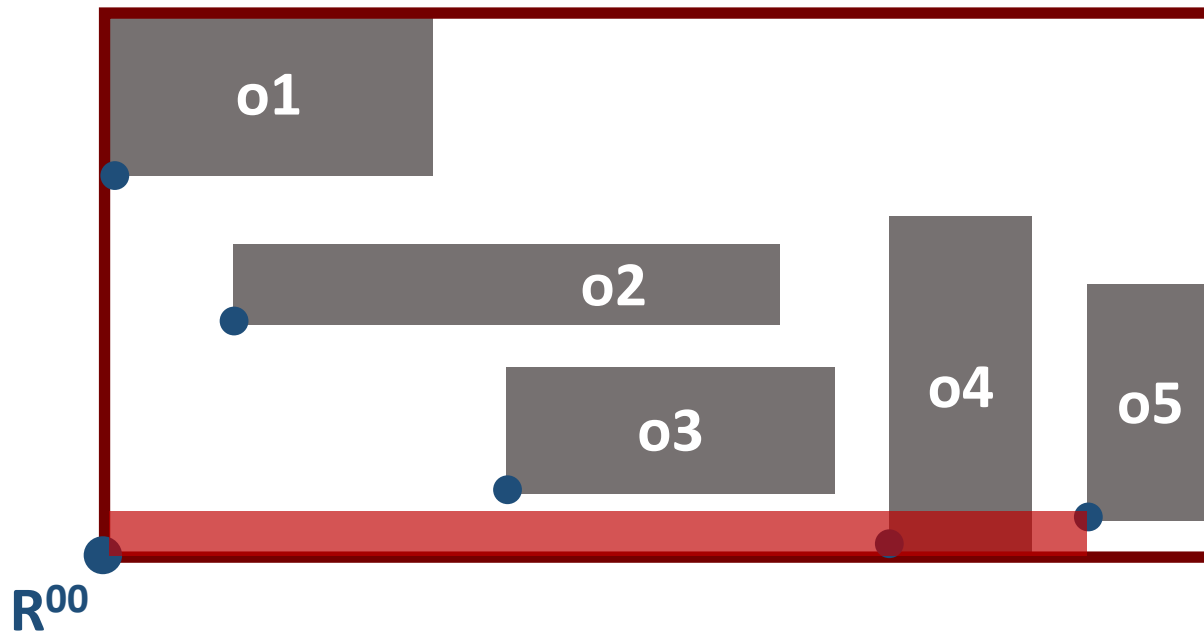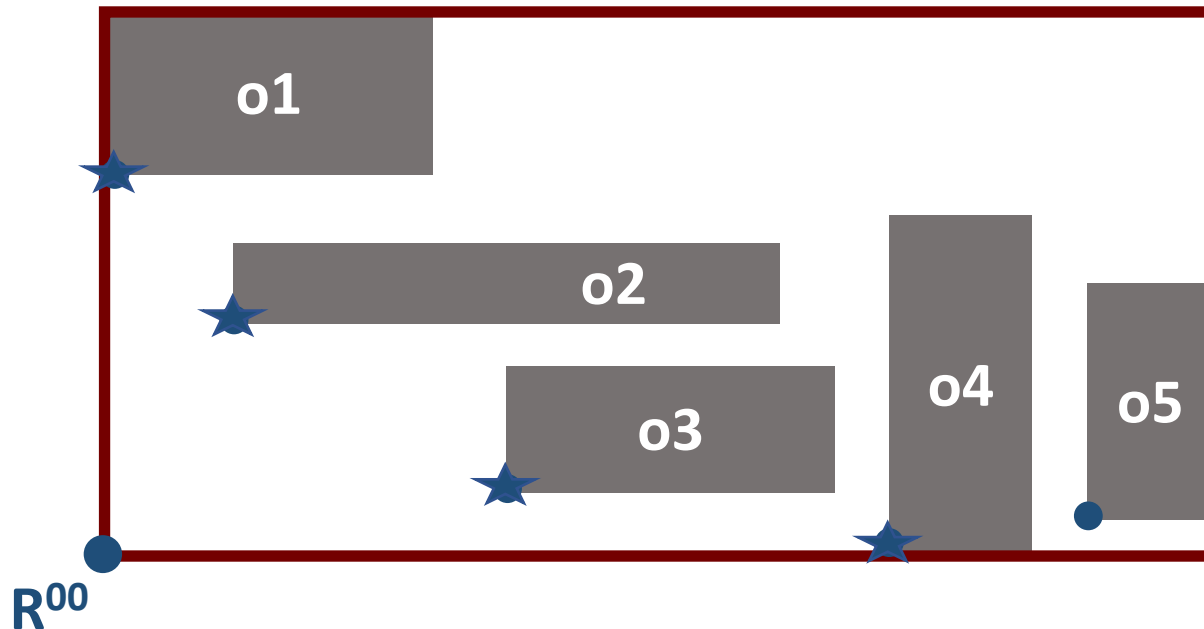
Candidate clip points

- For given corner R^b:
    - Consider only points in the outer surface of the objects o_i.
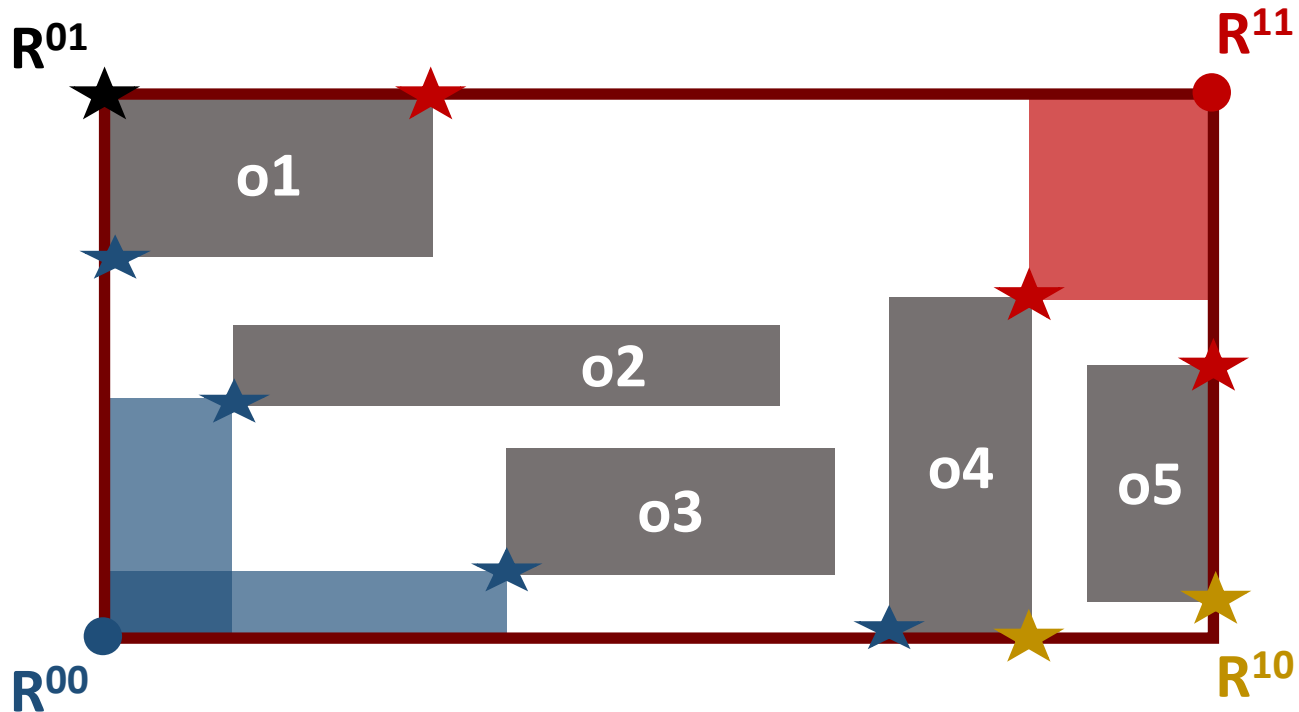   - Consider only the closest corner o_i^b.

Skyline clip points

- For given corner R^b:
  - Consider only points in the outer surface of the objects o_i.
  - Consider only the closest corner o_i^b.
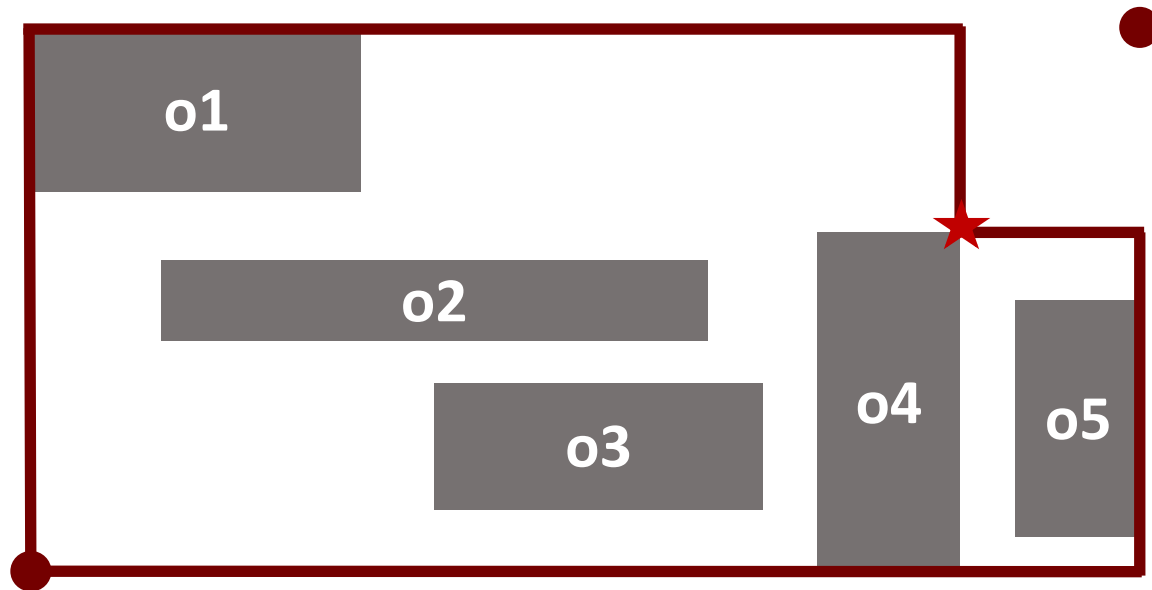 - Only the clip points in the Skyline of $\{o_i^b\}$ are valid clip points!
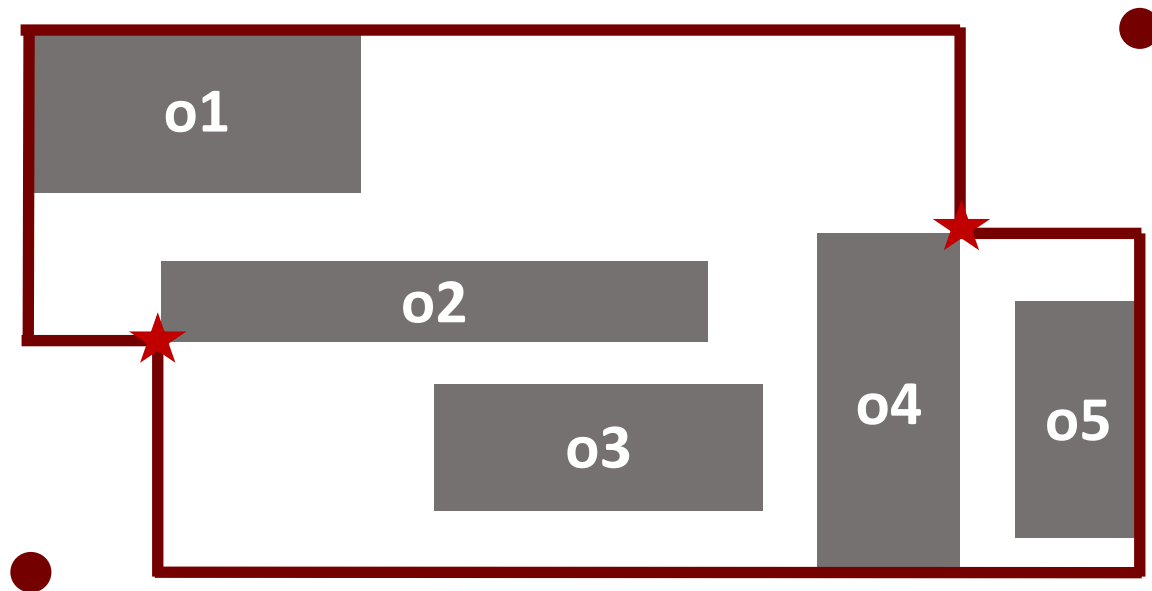
Skyline-based CBB

- Get skyline points with respect to each corner R^b.
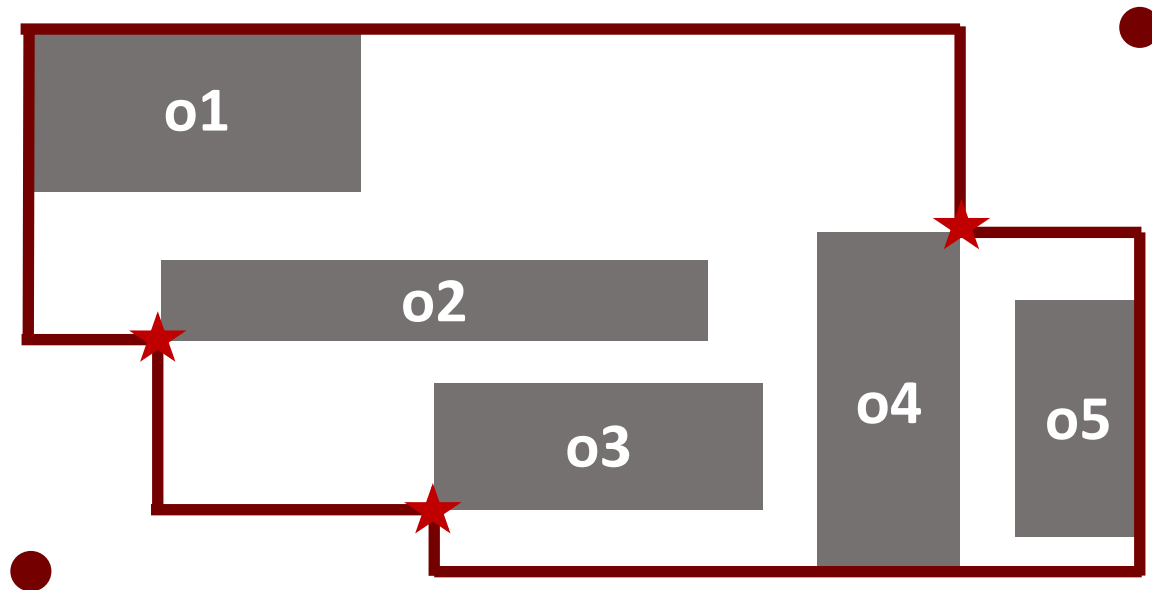- Choose up to k points.

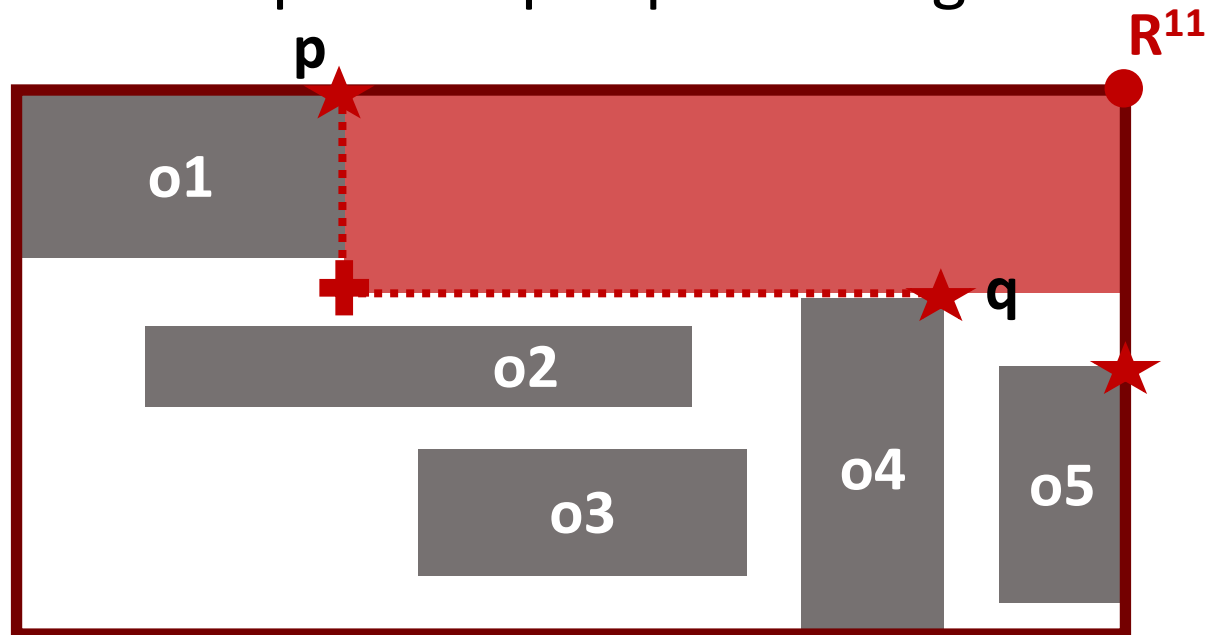# *Skyline*-based CBB (*k* = 1)

Skyline-based CBB (*k* = 2)
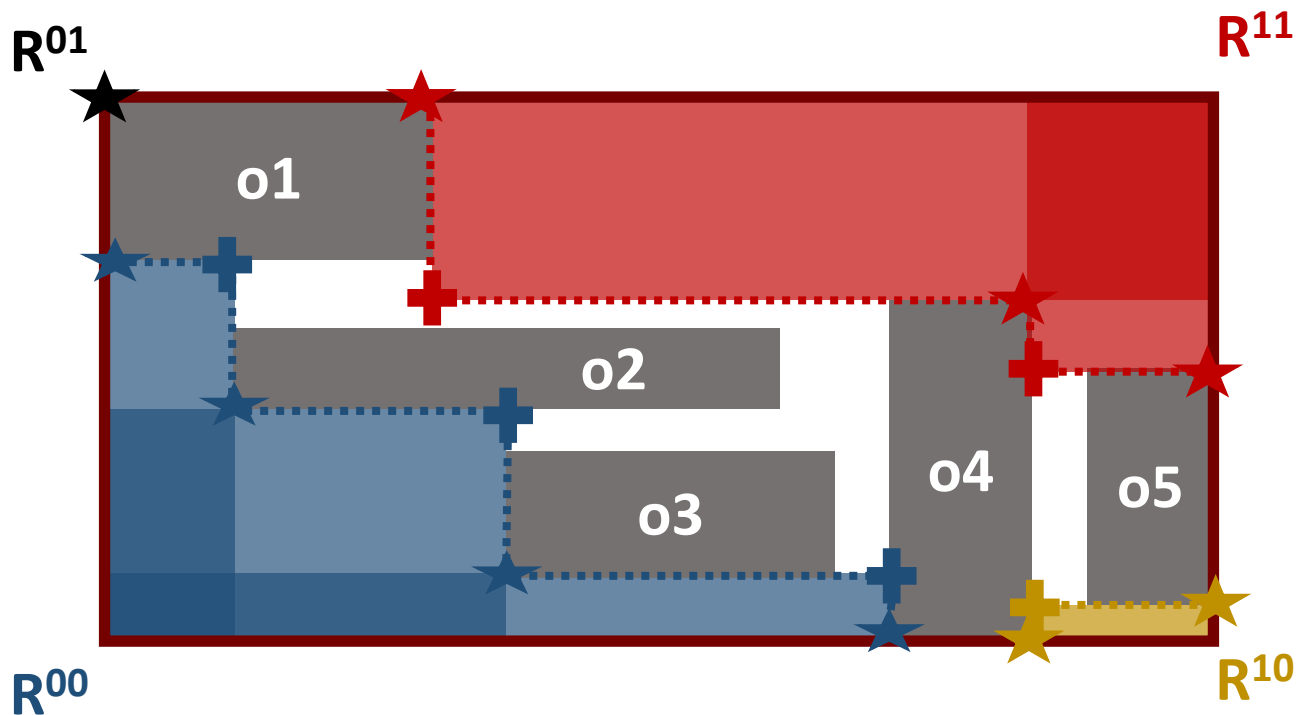
Skyline-based CBB (*k* = 3)

Stairline clip points

- "Between" two skyline points.
- Retain the "best" value in each dimension.
- Clip away significantly more dead space.
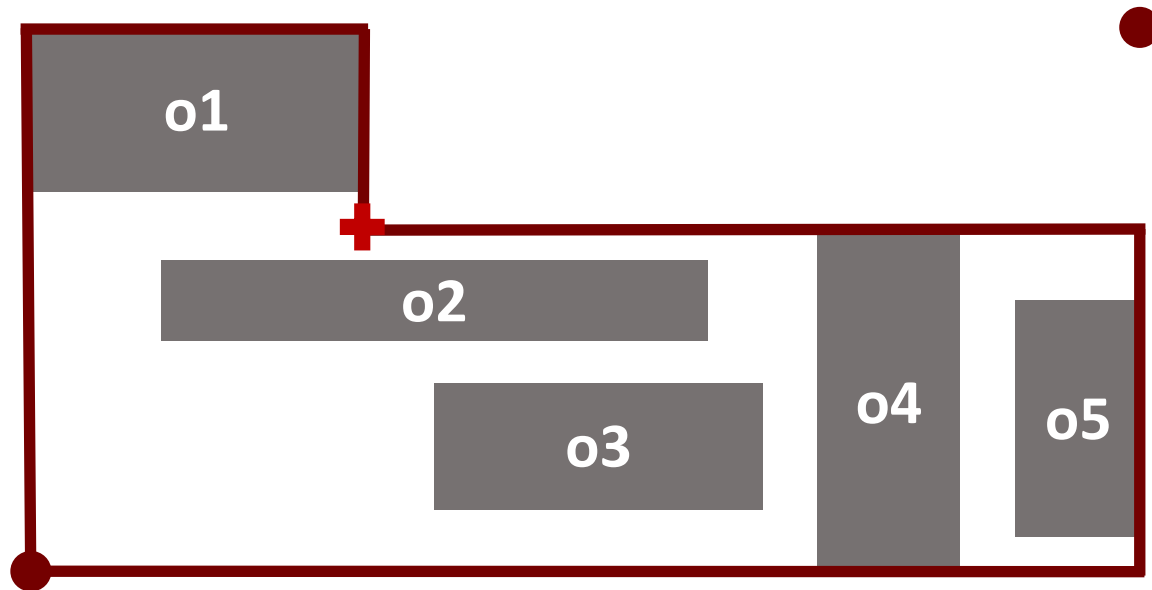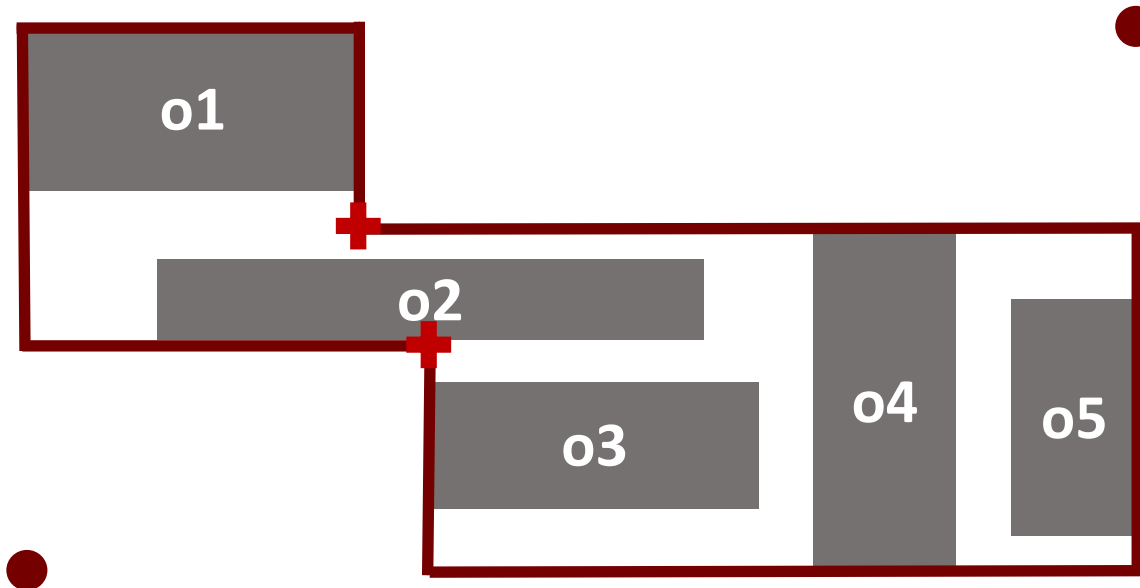- Require more expensive pre-processing.

Stairline-based CBB

- Get stairline points that are valid clip points with respect to each corner R^b.
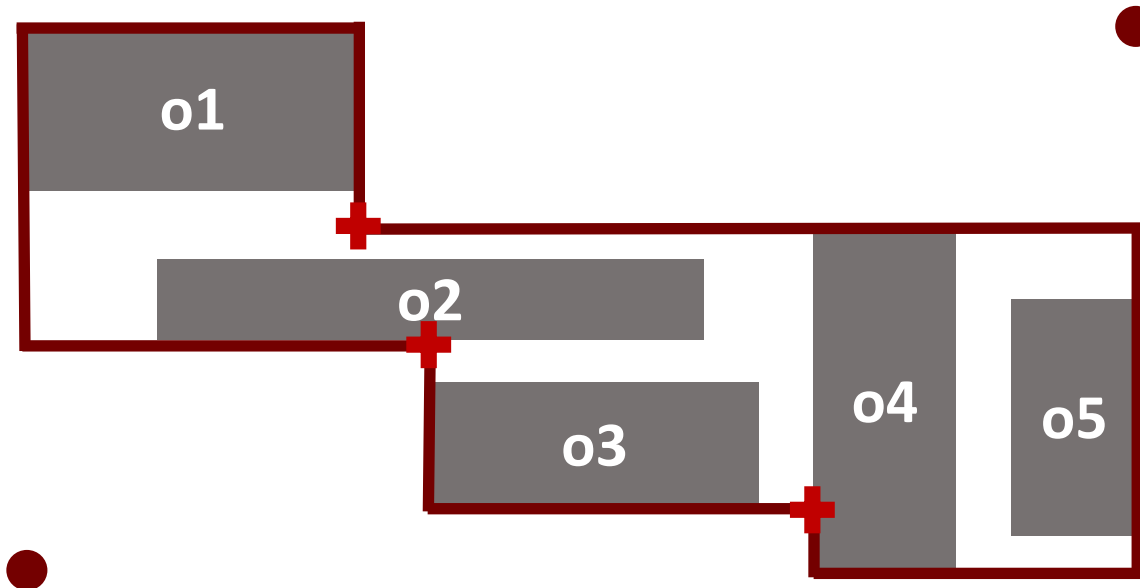
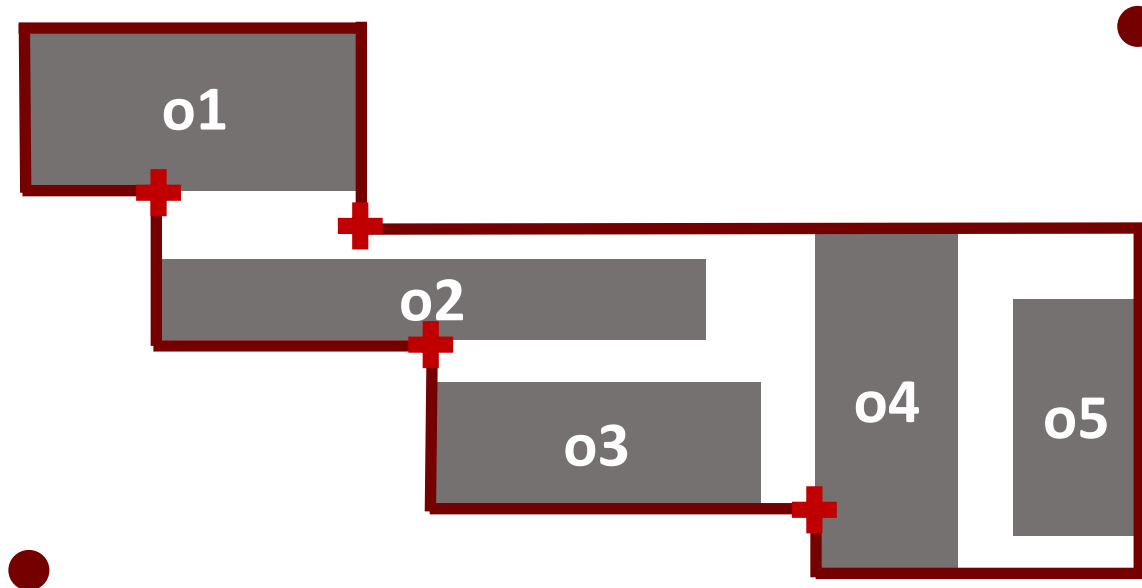- Choose up to k points.

Stairline-based CBB (*k* = 1)

Stairline-based CBB (*k* = 2)

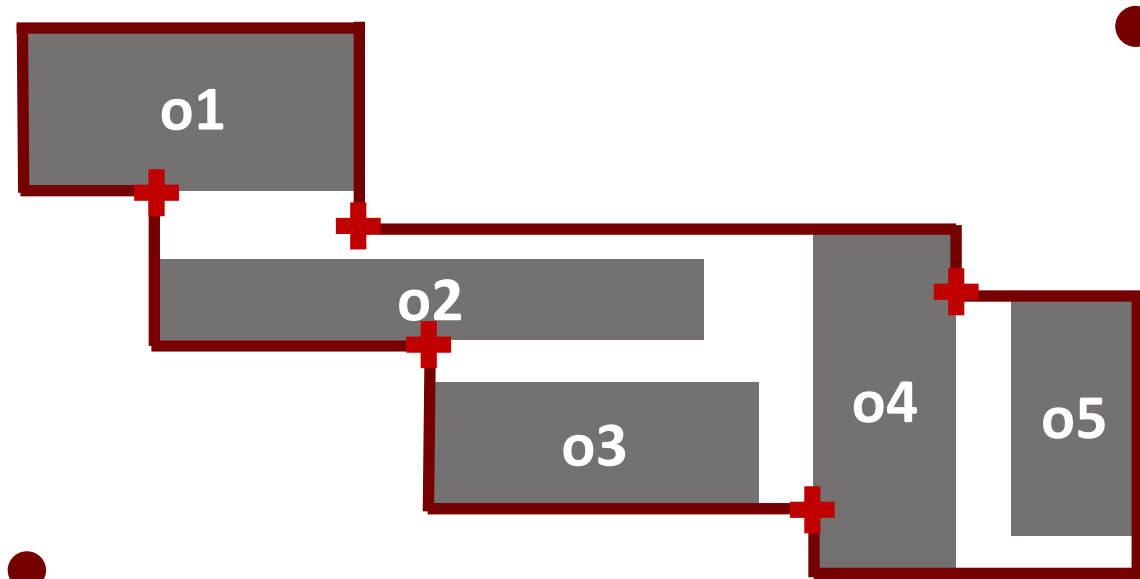# *Stairline*-based CBB (*k* = 3)

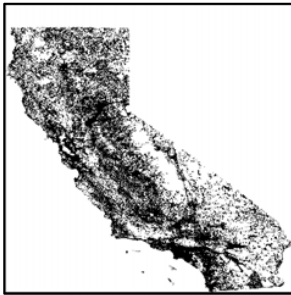# *Stairline*-based CBB (k = 4)

Stairline-based CBB (*k* = 5)

Experimental Setup

- **R-tree variants** Quadratic [QR-tree], Hilbert [HR-tree], R*-tree, Revised R*-tree [RR*-tree]

- **Range queries**
  - High: ≈ 1 object per query
  - Medium: ≈ 10 objects per query
  - Low: ≈ 100 objects per query

- **Spatial Join**

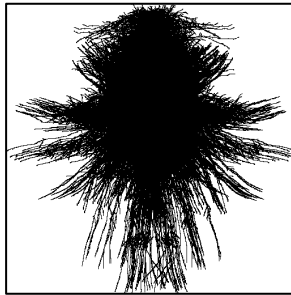- **Hardware** Quad-core Intel Core i7-3770 CPU @ 3.4GHz, 16GB RAM, 500GB HDD - 7200RPM
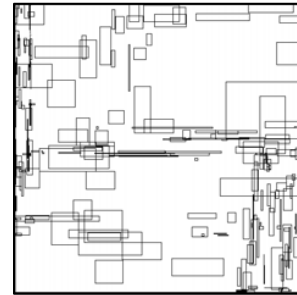
rea02
~2M elements

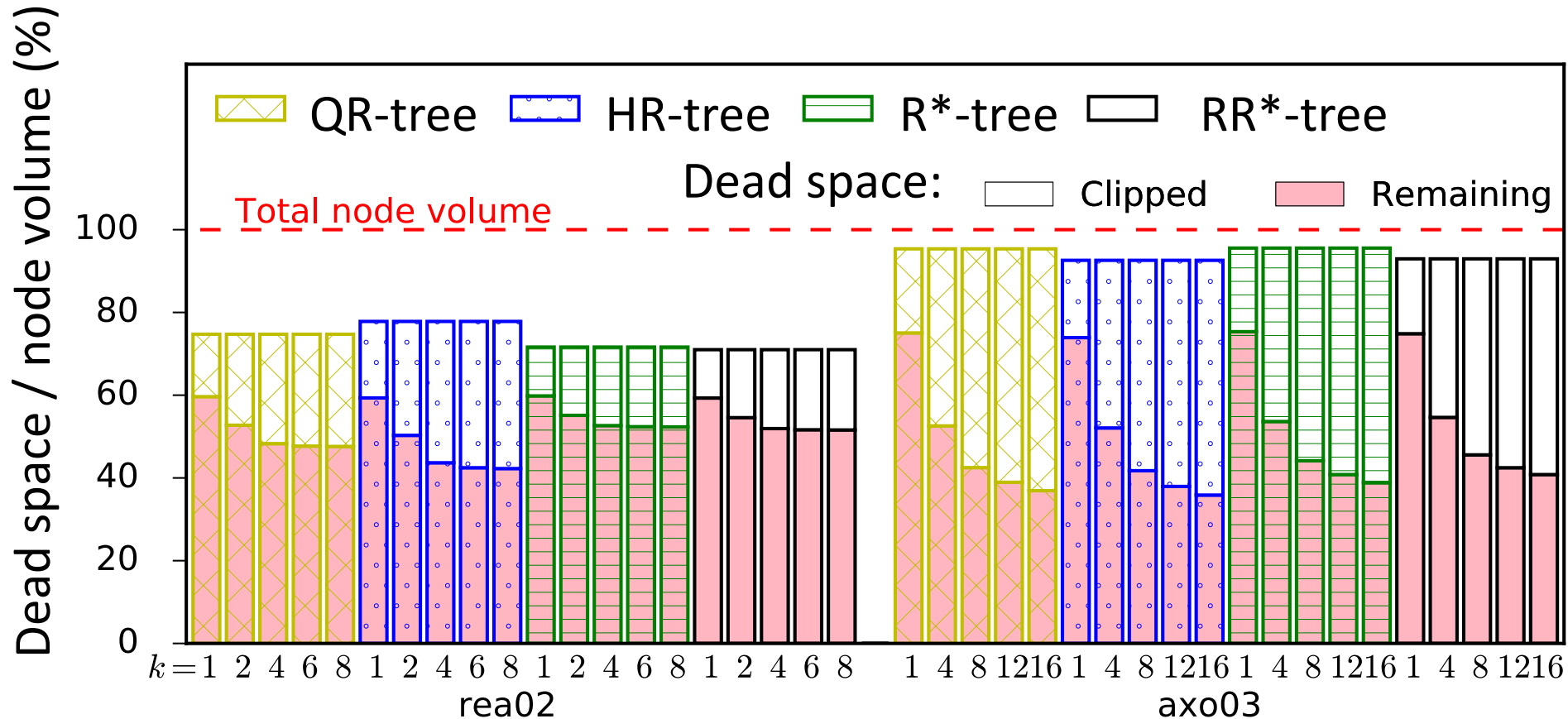**axo03**
~2.5 M elements

par02/par03
2^{30} elements
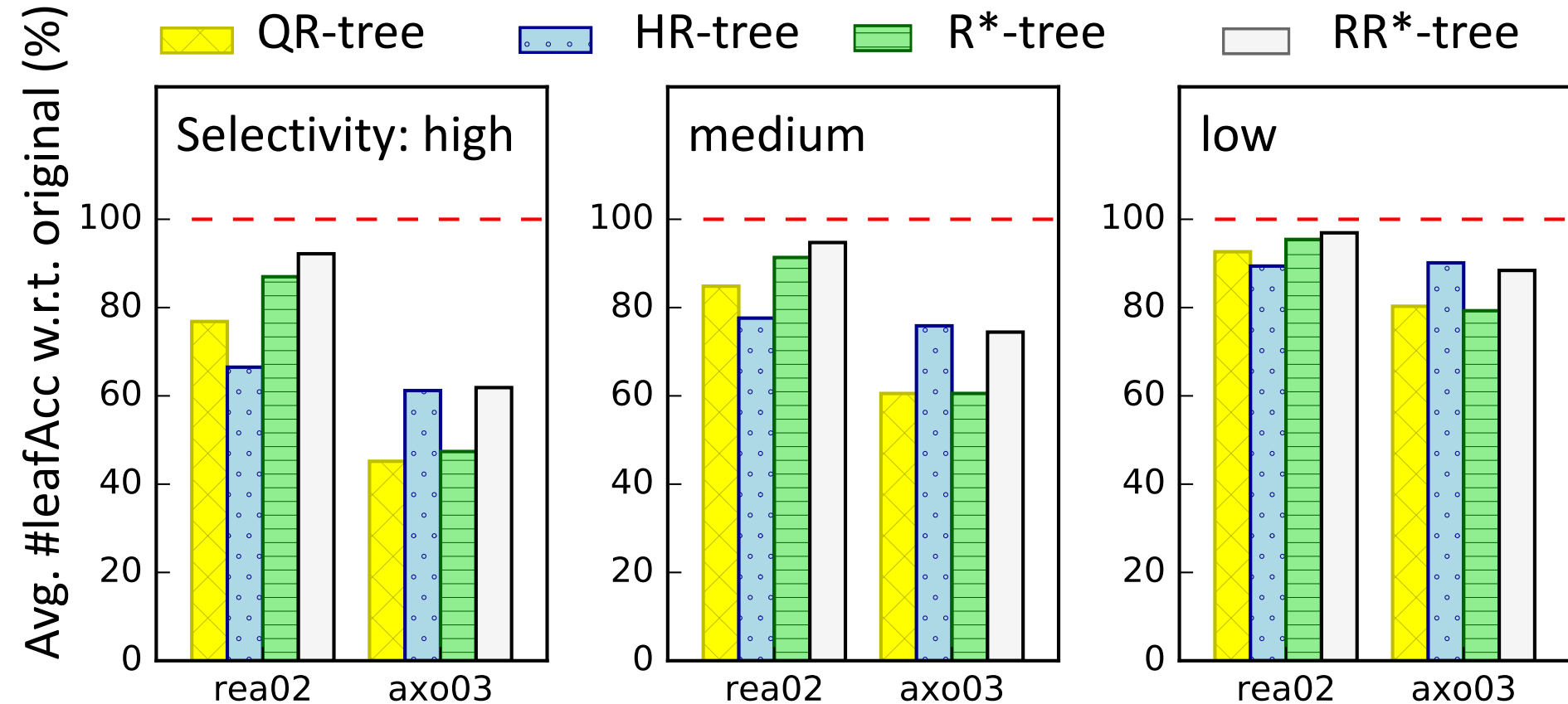
Dead space elimination

Stairline clipping

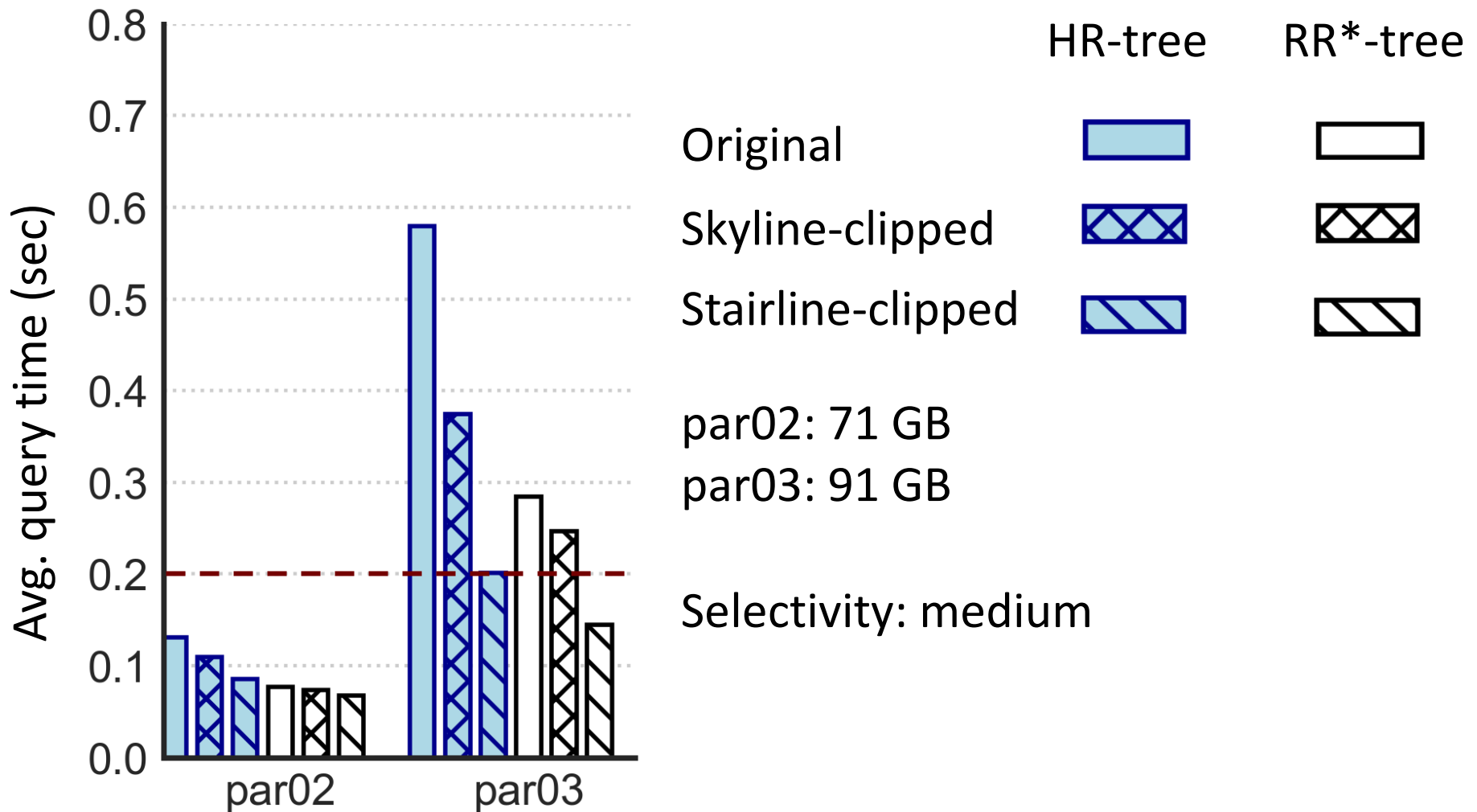**CBBs remove 27% - 60% of dead space**

Range query performance

Stairline clipping

≈26% I/O reduction across all R-trees/workloads

Querying 1B spatial objects

Enabling interactive times for 1B objects

Take home message

- **The Minimum Bounding Box (MBB) is ubiquitous**
  - Compact
  - Cheap intersection tests
  - Poor approximation of real data: can be > 90 % empty
    ➔ up to 64% unnecessary I/Os!

- **The Clipped Bounding Box**
  - Augments the MBB with few additional clip points
  - Retains the simplicity of the MBB
  - Eliminates up to 60 % of dead space
  - Enables interactive exploration of 1B objects

Thank you!