

The Case for Distance-Bounded Spatial Approximations

Eleni Tzirita Zacharatou

TU Berlin

Ibrahim Sabek

MIT

Harish Doraiswamy

NYU

Andreas Kipf

MIT

Varun Pandey

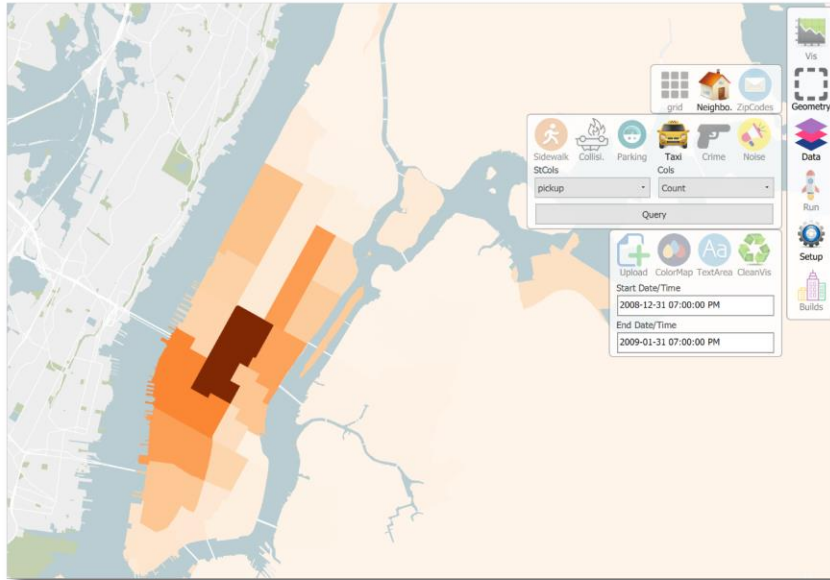
TU Munich

Volker Markl

TU Berlin



Visual Exploration of Mobility Data

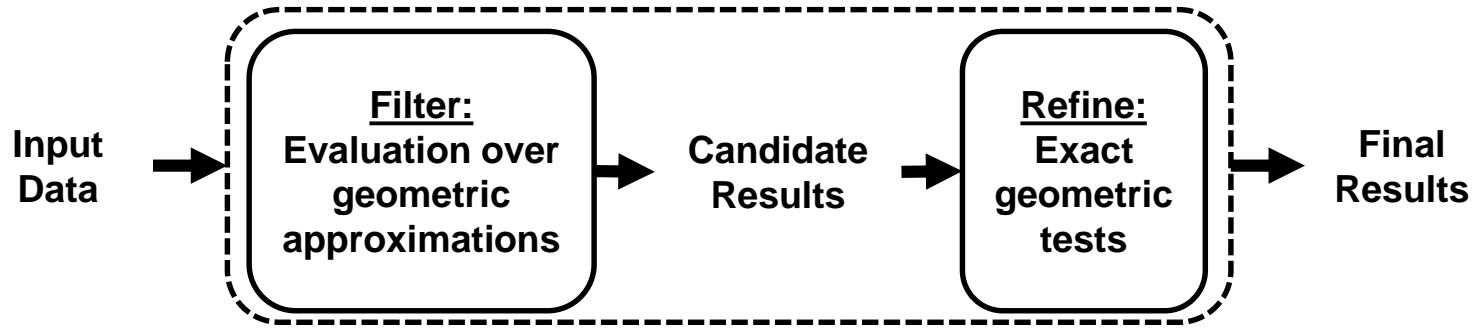


Distribution of taxi pickup locations per neighborhood in Manhattan

- Need: interactivity
- Approximate visualizations
- “Level-of-detail” exploration
- Imprecise GPS positions
- Fuzzy region boundaries

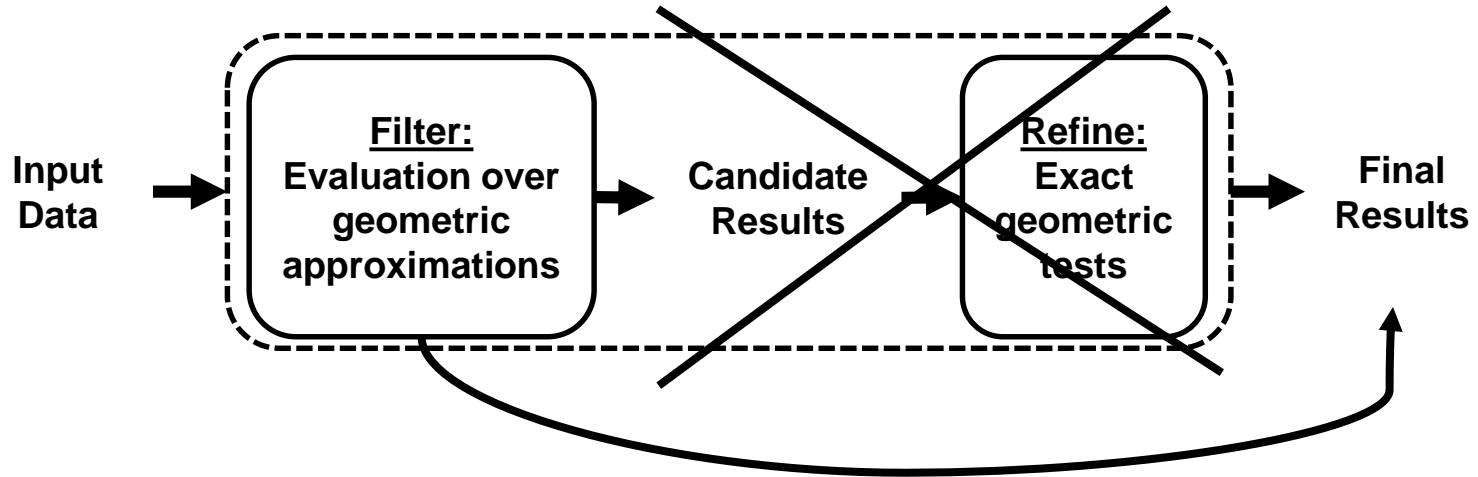
Paradigm shift: approximate spatial data processing

Traditional Spatial Query Evaluation



Imprecise geometric approximations
Expensive geometric tests

Spatial Query Evaluation Revisited



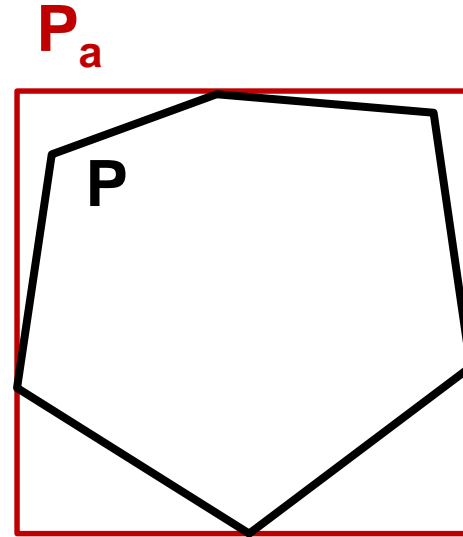
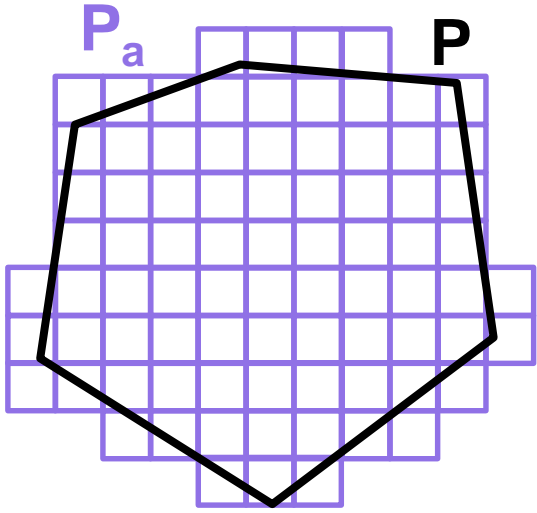
Increase the geometric approximation precision

Make the precision tunable

Distance Bound

- Bound on the Hausdorff distance between the approximate (purple) and the original polygon:

$$H(P_a, P) \leq \varepsilon$$

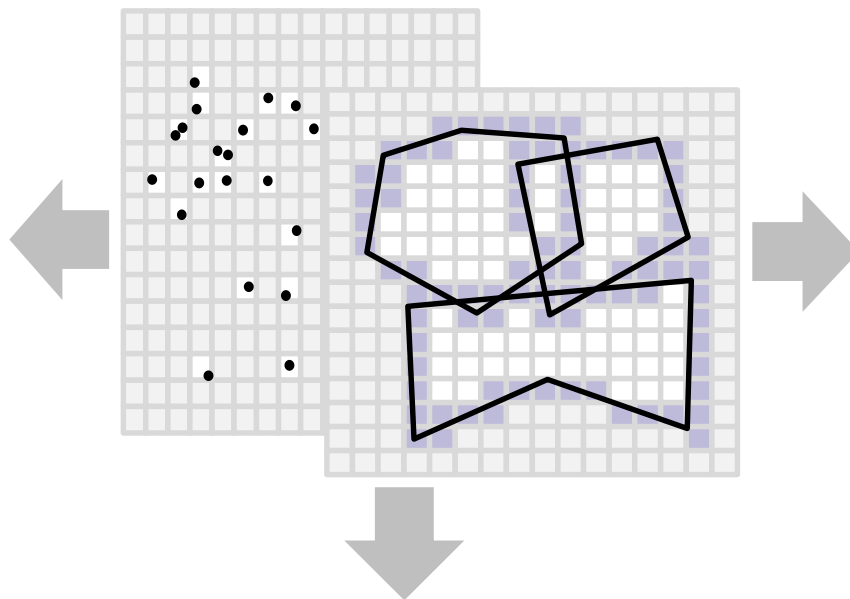


Geometry-independent, tunable precision

Vision: Raster Approximations at the Core

Data Access

- Novel indexes



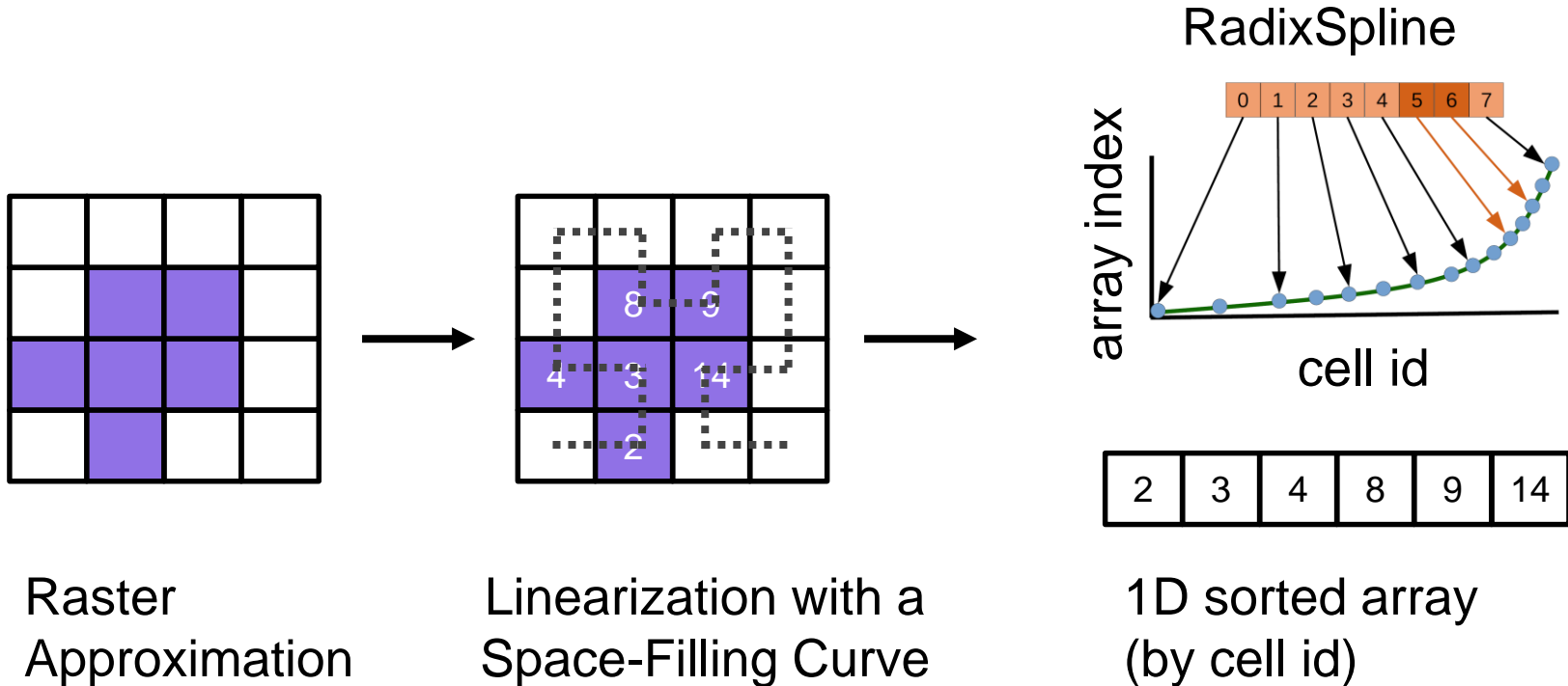
Query Execution

- Novel evaluation strategies

Query Optimization

- Novel data representation & GPU-friendly operators
- Enabling fine-grained optimization

A Learned Index for Rasterized Points

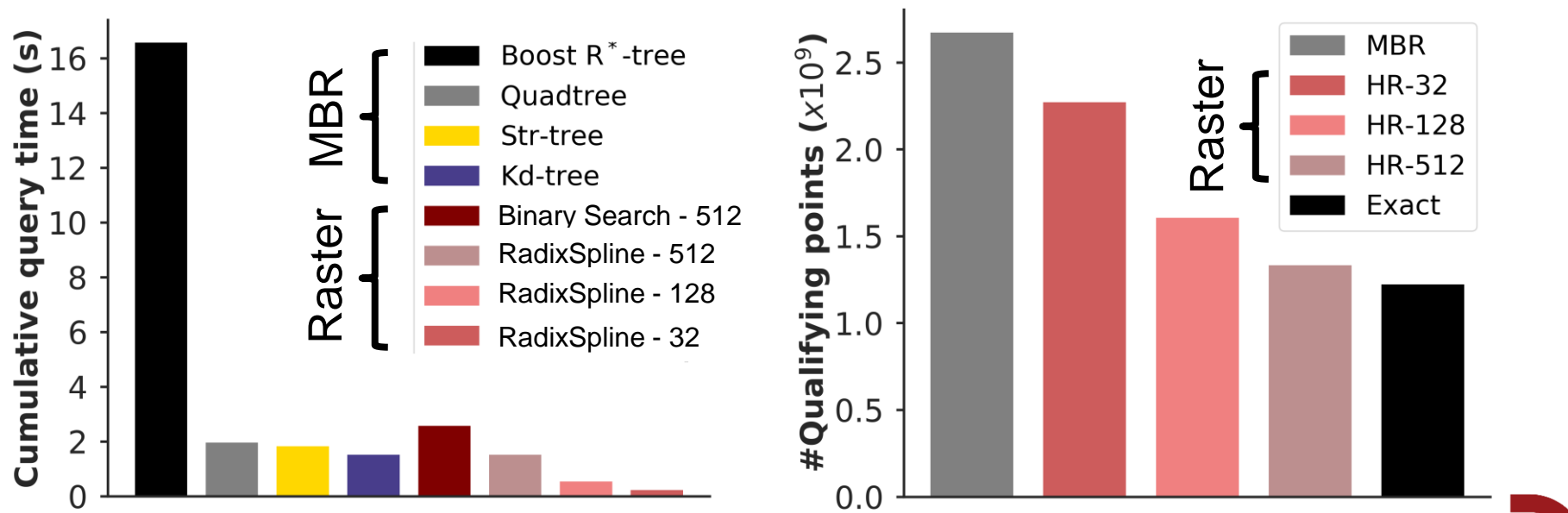


Map raster cells to an array and *learn* their position

Containment Query Performance

COUNT Taxi rides (1.2 B points) WITHIN NYC census region (39,200 polygons)

Single-threaded, Intel Xeon Gold 6230 CPU @ 3.9 GHz turbo, 256 GB RAM

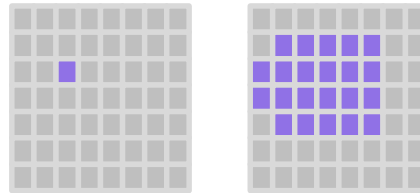


Sweet spot in the trade-off between precision and query time

Spatial Data Representation & Query Operators

- **Filter & Refine – based processing**
 - Monolithic operators
 - Geometry-specific implementations
 - Limited optimization options
- **Approximate rasterization – based processing**

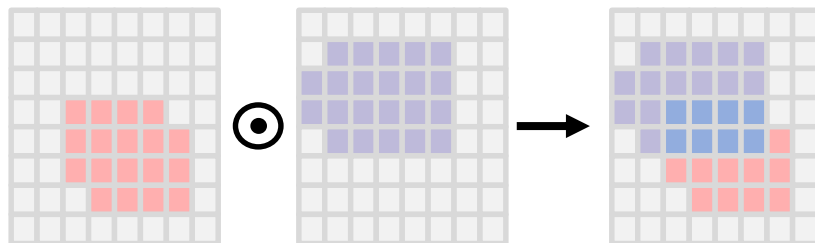
- Geometry-agnostic



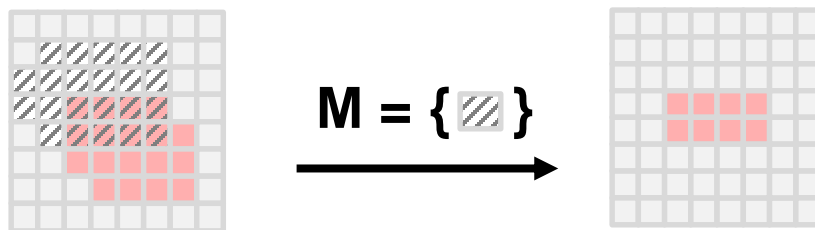
Unified representation of geometric objects

Geometry - Independent Operators

- Blend



- Mask



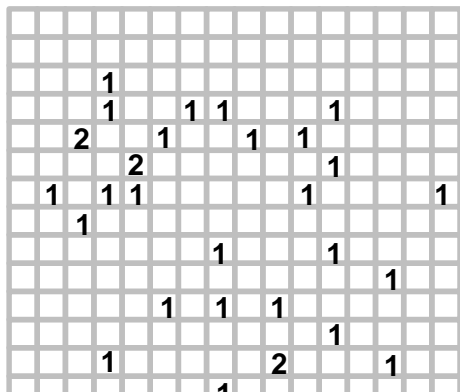
- Affine transformations

Common graphics operations supported by GPUs
Enable finer-grained optimization over a wider set of options

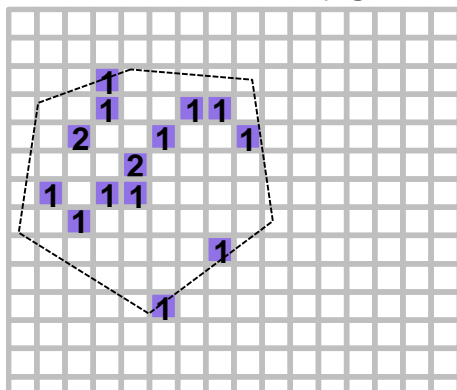
Spatial Aggregation Query Evaluation

```
SELECT COUNT(*)  
FROM taxi ride T, neighborhoods N  
WHERE T.pickup INSIDE N.geometry  
GROUP BY N.id
```

Blend points



Blend & Mask
points and polygons



Transform & Blend
to get the aggregate

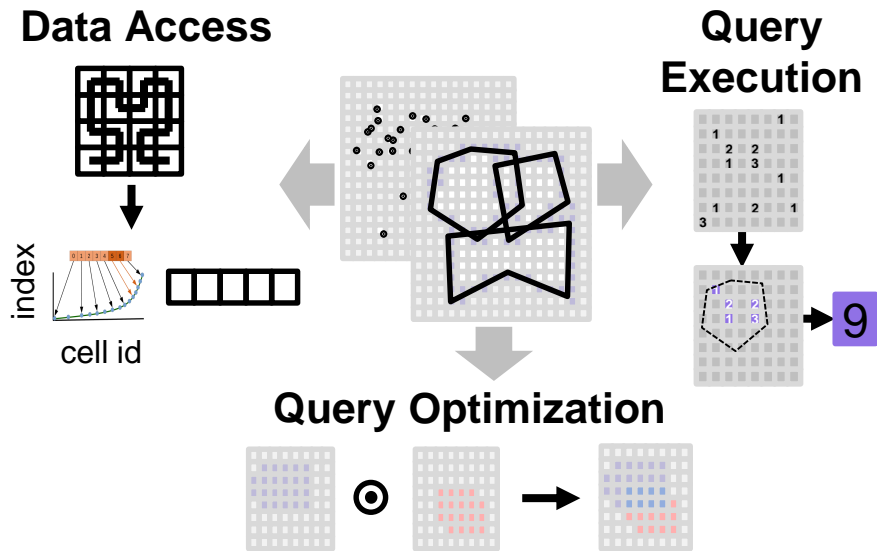
16

Orders of magnitude speedup over typical evaluation strategies

The Case for Distance-Bounded Spatial Approximations

Eleni Tzirita Zacharitou, Andreas Kipf, Ibrahim Sabek, Varun Pandey, Harish Doraiswamy, Volker Markl

- Approximate spatial data processing techniques need a distance bound
- Trade precision for interactivity
- Express spatial operators as graphics primitives and use modern GPUs



We envision novel spatial systems that employ distance-bounded spatial approximations at their core