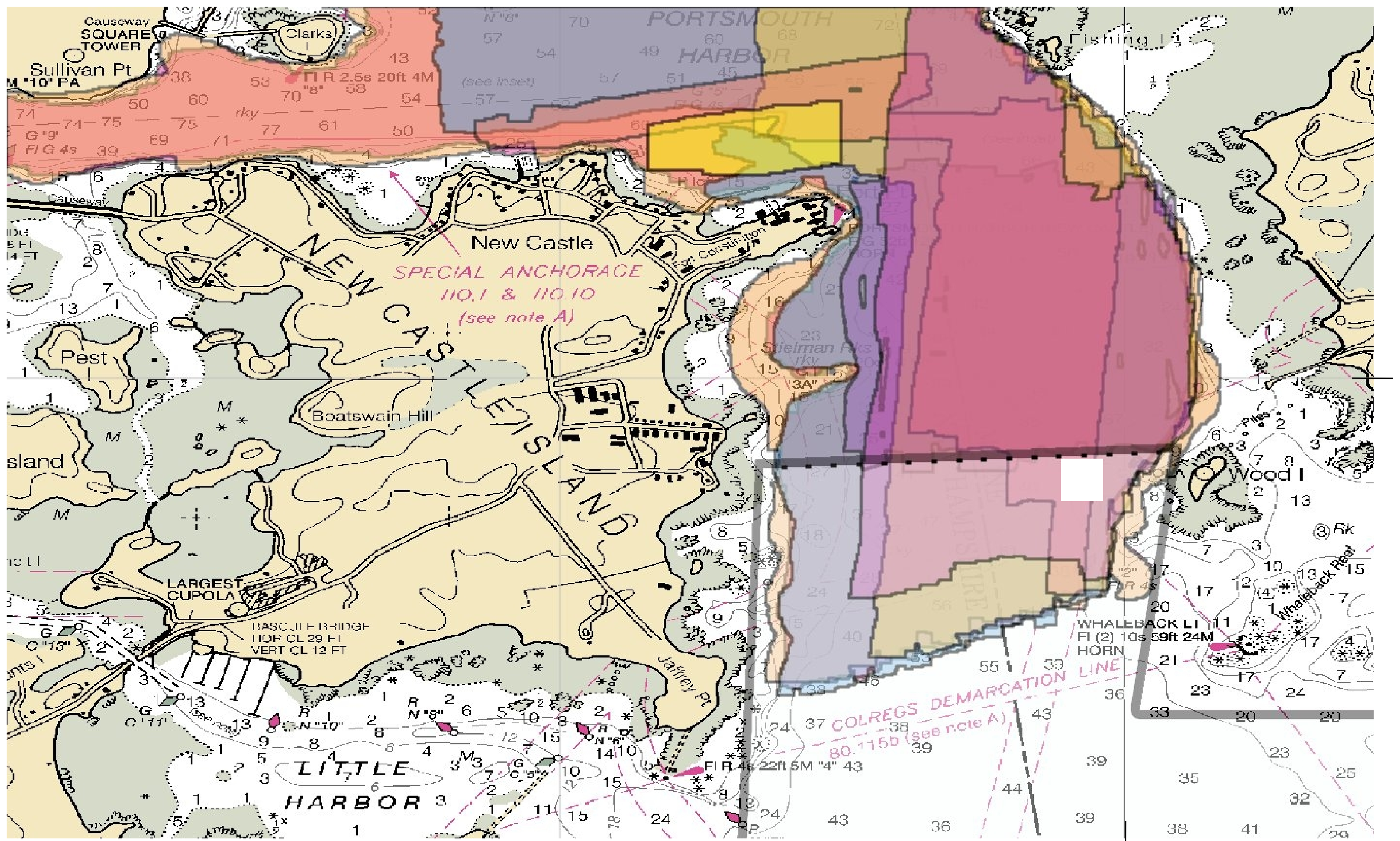
# Storage of Versioned Data Across Polygonal Regions

***Stuart A. MacGillivray and Bradford G. Nickerson***
Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada

## Motivation

- Multiple versions of spatial data, such as from bathymetric surveys, can have distinct overlapping coverage areas.
- Window queries on data must be able to distinguish what portions of space are covered by multiple versions to determine which versions of the data are to be considered.
- We need to store a planar subdivision consisting of connected subregions arising from overlapping polygons, along with the data from each version of each region.




From Shallow Survey 2008. http://www.shallowsurvey2008.org/

## Results

- Number of subregions dependent on polygon arrangement.
- At most $O(km^2)$ or $O(k^2m^2)$ subregions exist in arrangements of m k-sided convex or simple polygons, respectively.
- Each subregion can have as most O(mk) edges, and worst-case average O(k).
- Index structure supporting orthogonal range search requires minimum space equal to the number of edges and must process all edges of a given subregion in the worst case.
- Search returning T subregions from m k-sided simple polygons can be done in RAM in $O(k^2m^2)$ space, O(km(log km) + kT) time. [1]
- Data storage cost is linear in the total number of points N.
- Range search is an aggregate of queries on intersected subregions.
- Total search cost to retrieve S points from R subregions in the I/O model with block size B is $O(\sqrt{RN/B} + S/B)$ I/Os.

## Example Subdivisions

- Two classes of polygon sets: all polygons convex, or all simple.
- Each of the following examples is a worst case for k and m.

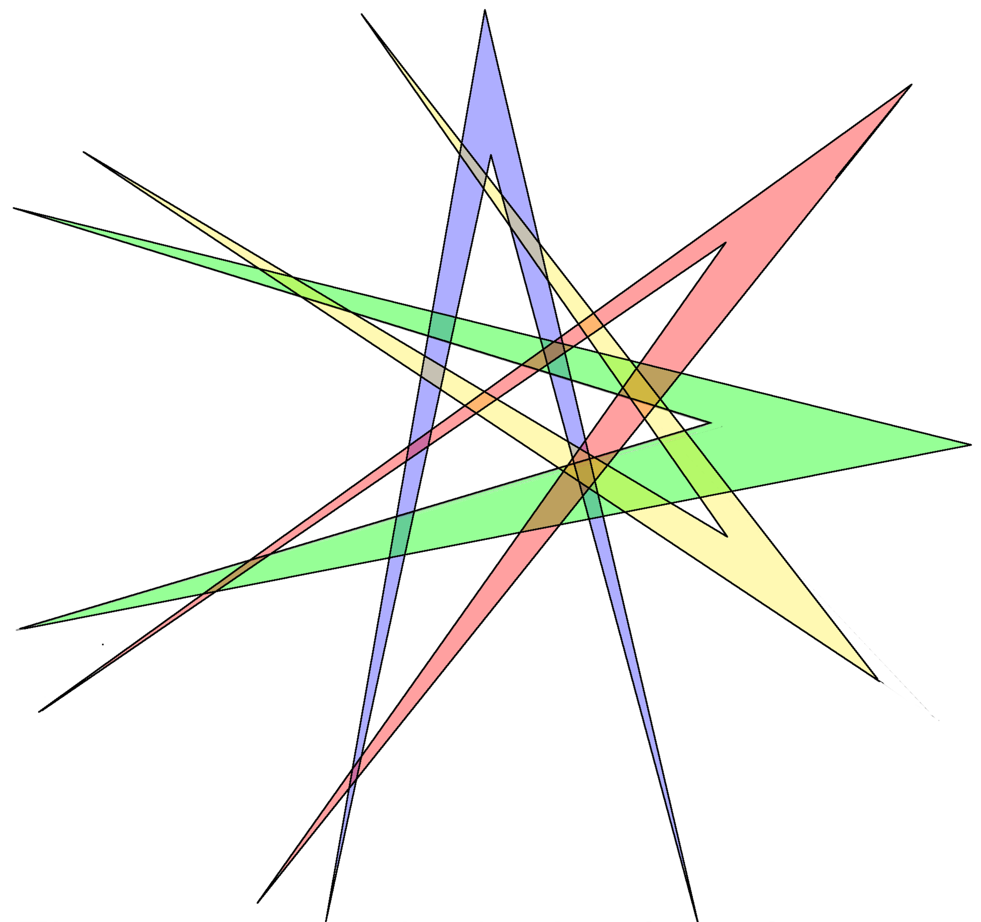


Figure 1: m=4, k=4, simple polygons

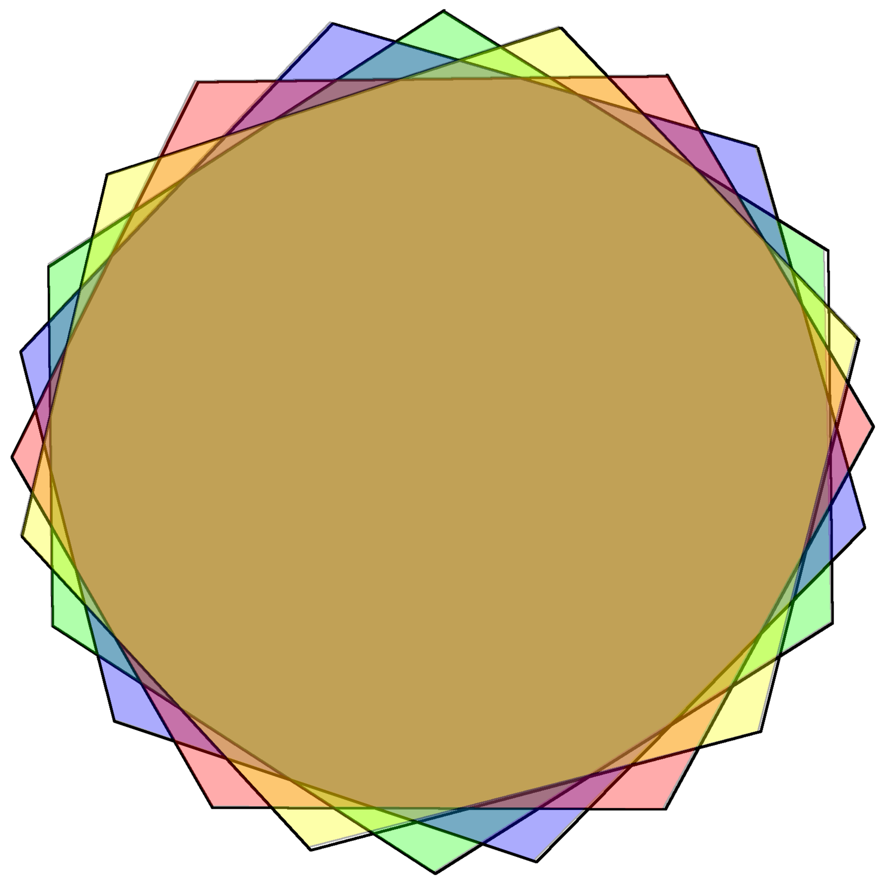
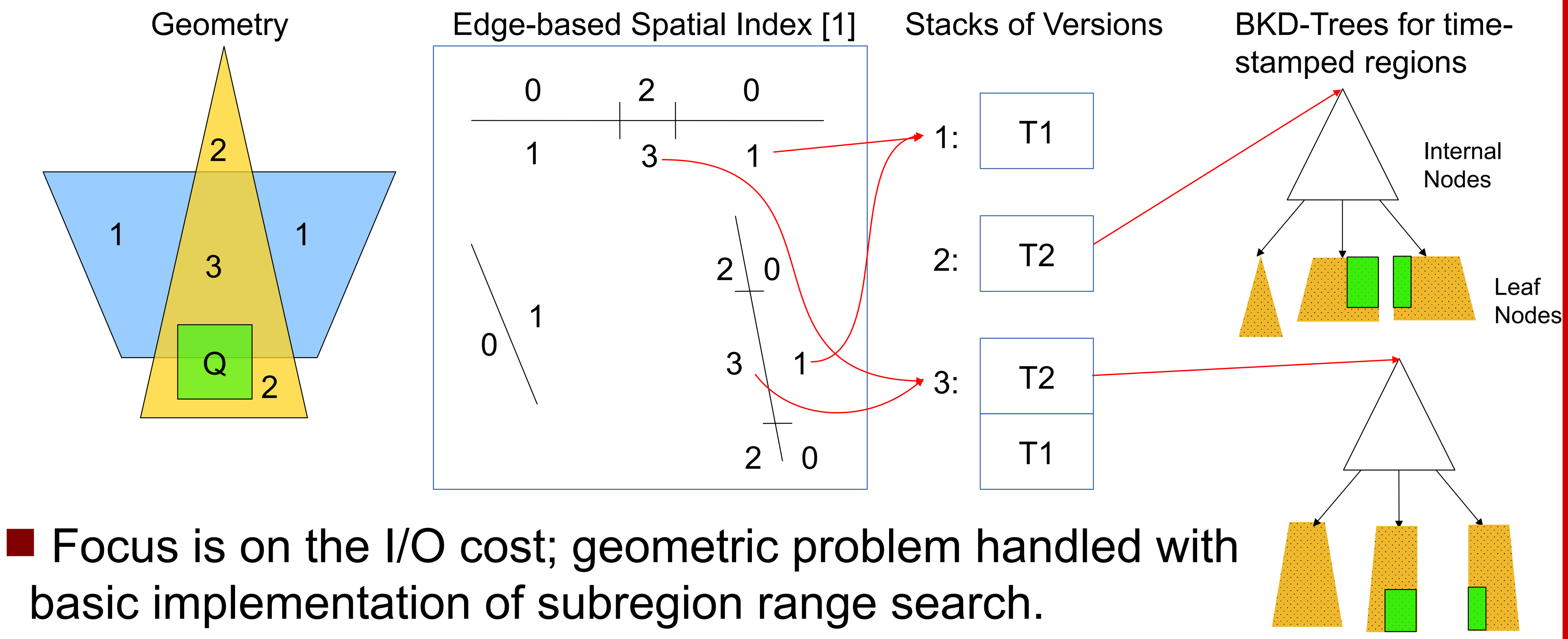


Figure 2: m=4, k=6, convex polygons

| Figure | $\|E\|$ | $\|S_0\|$ | $\|S_1\|$ | $\|S_2\|$ | $\|S_3\|$ | $\|S_4\|$ |
|---|---|---|---|---|---|---|
| 1 | 208 | 17 | 42 | 31 | 7 | 1 |
| 2 | 168 | 1 | 24 | 24 | 24 | 1 |

Definitions: E is the set of edges, $S_n$ is the set of subregions with n layers.

## Data Structure




- Focus is on the I/O cost; geometric problem handled with basic implementation of subregion range search.
- Disk-based point storage implemented as BKD-trees using STXXL[2].
- Multi-layer structure: geometric structure points to temporal stacks with pointers to BKD-trees that index blocks on disk.

References: [1] Stuart A. MacGillivray, Bradford G. Nickerson: Window Queries On Planar Subdivisions Arising From Overlapping Polygons. Proceedings of the 26th Canadian Conference on Computational Geometry, CCCG 2014, Halifax, Nova Scotia, Canada, August 11-13, 2014, pages 281-286
[2] STXXL: Standard Template Library for Extra Large Data Sets. http://stxxl.sourceforge.net/

**Sponsored by:** UNB Faculty of Computer Science, Mitacs, Caris, NSERC CRSNG