Report all rectangles that intersect rectangular query range
Report all rectangles that intersect rectangular query range

Query time $\approx$ seek time $\approx$ disk blocks retrieved
Space-filling curves for spatial data structures

Index in memory (block bounding boxes)

Report all rectangles that intersect rectangular query range

Query time $\approx$ seek time $\approx$ disk blocks retrieved
Space-filling curves for spatial data structures

Report all rectangles that intersect rectangular query range

Query time \approx \text{seek time} \approx \text{disk blocks retrieved}
Space-filling curves for spatial data structures

Report all rectangles that intersect rectangular query range

Query time ≈ seek time ≈ disk blocks retrieved

Only read blocks whose bounding boxes intersect query range
Space-filling curves for spatial data structures

Report all rectangles that intersect rectangular query range

Query time ≈ seek time ≈ disk blocks retrieved

Only read blocks whose bounding boxes intersect query range
Space-filling curves for spatial data structures

Report all rectangles that intersect rectangular query range

Query time $\approx$ seek time $\approx$ disk blocks retrieved

Only read blocks whose bounding boxes intersect query range
Algorithms to make blocks

Theory:

blocks visited in ... for ... rectangle query point query
KDB-trees (point data) $\Theta(\sqrt{n/B} + t/B) \quad 1$
PR-trees (rectangle data) $\Theta(\sqrt{n/B} + t/B) \quad \Theta(\sqrt{n/B})$

($n$ is input size; $t$ is output size; $B$ is disk block size)

Practice:

Use space-filling curve.
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE

Group data into blocks in order along curve (example: block size = 4)
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE

Group data into blocks in order along curve (example: block size = 4)
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE

Group data into blocks in order along curve (example: block size = 4)
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE

Group data into blocks in order along curve (example: block size = 4)
Example: Z-order for point data

Space-filling curve: mapping from unit square to unit interval
Z-order: map quadrants recursively in order NW, NE, SW, SE

Group data into blocks in order along curve (example: block size = 4)

Question: what to do with rectangle data?
Question: what is the best curve?
What curve?

Author A: “Use the Peano curve!”

What curve?

Author A: “Use the Peano curve!”

What curve?

Author A: “Use the Peano curve!”

Author B: “Use the Peano curve!”

Giuseppe Peano:
Sur une courbe, qui remplit toute une aire plane.
What curve?

Author A: “Use the Peano curve!”

Author B: “Use the Peano curve!”

Giuseppe Peano:
Sur une courbe, qui remplit toute une aire plane.
What curve?

Author A: "Use the Peano curve!"

Author B: "Use the Peano curve!"

Author C: "Use the Peano curve!"

What curve?

Author A: “Use the Peano curve!”

Author B: “Use the Peano curve!”


Author C: “Use the Peano curve!”
What curve?

Author A: "Use the Peano curve!"

Author B: "Use the Peano curve!"

Giuseppe Peano:
Sur une courbe, qui remplit toute une aire plane.

Author C: "Use the Peano curve!"
What curve?

Author A: “Use the Peano curve!”

Author B: “Use the Peano curve!”

Author C: “Use the Peano curve!”


Z-order

Hilbert curve

a.k.a. Lebesgue order, Morton indexing

Peano’s curve

H-order
What curve?

Z-order
Hilbert curve
H-order
Peano’s curve
What curve?

Z-order

Hilbert curve

H-order

Peano's curve

$\mathbb{AR}^2\mathbb{W}^2$-curve

$\beta\Omega$-curve

Balanced Peano
Measuring curves

worst-case dilation := \( \max_{p,q \in \text{unit square}} \) squared distance between \( p \) and \( q \)

For a curve section of fixed size, how far can the endpoints be apart?

\( \implies \) how big can the bounding box be?
### Measuring curves

<table>
<thead>
<tr>
<th>Worst dilation</th>
<th>$L_\infty$</th>
<th>$L_2$</th>
<th>$L_1$</th>
<th>$L_\infty$</th>
<th>$L_2$</th>
<th>$L_1$</th>
<th>$L_\infty$</th>
<th>$L_2$</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peano</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>closed: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balanced</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>closed: 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hilbert</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>closed: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta\Omega$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>4</td>
<td>closed: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-order</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>8</td>
<td>closed: 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR$^2$W$^2$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>4</td>
<td>closed: 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-order</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>8</td>
<td>closed: 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table compares the worst dilation of curves for different orderings and metrics. The lower bound values are given for each case.
## Measuring curves

<table>
<thead>
<tr>
<th>dilation</th>
<th>Peano</th>
<th>Balanced</th>
<th>Hilbert</th>
<th>$\beta\Omega$</th>
<th>Z-order</th>
<th>AR$^2$W$^2$</th>
<th>H-order</th>
<th>lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_\infty$</td>
<td>$L_2$</td>
<td>$L_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worst</td>
<td></td>
<td>8</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>5.40</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td>dilation</td>
<td></td>
<td>8</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>6.05</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td>$L_1$</td>
<td></td>
<td>$10^{\frac{2}{3}}$</td>
<td>9</td>
<td>9.00</td>
<td>$\infty$</td>
<td>12.00</td>
<td>8</td>
<td>closed: 8</td>
</tr>
</tbody>
</table>
Measuring curves

worst-case bbox area ratio := $\max_{p,q \in \text{unit } \square} \frac{\text{bbox area of curve between } p \text{ and } q}{\text{area filled by curve between } p \text{ and } q}$

For a curve section of fixed size, how big can the bounding box be?
### Measuring curves

<table>
<thead>
<tr>
<th></th>
<th>Peano</th>
<th>Balanced</th>
<th>Hilbert</th>
<th>$\beta\Omega$</th>
<th>Z-order</th>
<th>$AR^2W^2$</th>
<th>H-order</th>
<th>lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$L_\infty$</strong></td>
<td>8</td>
<td>4.62</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>5.40</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td><strong>$L_2$</strong></td>
<td>8</td>
<td>4.62</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>6.05</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td><strong>Worst dilation</strong></td>
<td>$10\frac{2}{3}$</td>
<td>8.62</td>
<td>9</td>
<td>9.00</td>
<td>$\infty$</td>
<td>12.00</td>
<td>8</td>
<td>closed: 8</td>
</tr>
<tr>
<td><strong>Worst box area</strong></td>
<td>2.00</td>
<td>2.00</td>
<td>2.40</td>
<td>2.22</td>
<td>$\infty$</td>
<td>3.05</td>
<td>3.00</td>
<td>rectangles /triangles: 2</td>
</tr>
</tbody>
</table>

**Notes:**
- $L_\infty$ and $L_2$ are metrics for measuring curves.
- $\beta\Omega$ represents a specific ordering or algorithm.
- $AR^2W^2$ and H-order are likely related to geometric or spatial properties.
- The lower bound is the minimum value for the measured curves.
Measuring curves

random subdivision bbox area =
sum of section bbox areas of random subdivision of curve into sections
Measuring curves

random subdivision bbox area = sum of section bbox areas of random subdivision of curve into sections
Measuring curves

random subdivision bbox area =
sum of section bbox areas of random subdivision of curve into sections
Measuring curves

random subdivision bbox area =
sum of section bbox areas of random subdivision of curve into sections
Measuring curves

random subdivision bbox area =
sum of section bbox areas of random subdivision of curve into sections

etc. etc. etc.
<table>
<thead>
<tr>
<th>Method</th>
<th>Peano</th>
<th>Balanced</th>
<th>Hilbert</th>
<th>$\beta\Omega$</th>
<th>Z-order</th>
<th>AR$^2$W$^2$</th>
<th>H-order</th>
<th>Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_\infty$</td>
<td>8</td>
<td>4.62</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>5.40</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td>$L_2$</td>
<td>8</td>
<td>4.62</td>
<td>6</td>
<td>5.00</td>
<td>$\infty$</td>
<td>6.05</td>
<td>4</td>
<td>closed: 4</td>
</tr>
<tr>
<td>worst dilation $L_1$</td>
<td>$10^\frac{2}{3}$</td>
<td>8.62</td>
<td>9</td>
<td>9.00</td>
<td>$\infty$</td>
<td>12.00</td>
<td>8</td>
<td>closed: 8</td>
</tr>
<tr>
<td>worst box area</td>
<td>2.00</td>
<td>2.00</td>
<td>2.40</td>
<td>2.22</td>
<td>$\infty$</td>
<td>3.05</td>
<td>3.00</td>
<td>rectangles /triangles: 2</td>
</tr>
<tr>
<td>total box area</td>
<td>1.42</td>
<td>1.42</td>
<td>1.41</td>
<td>1.40</td>
<td>2.86</td>
<td>1.47</td>
<td>1.69</td>
<td></td>
</tr>
</tbody>
</table>
What to do with rectangle data on Hilbert curves?

Three ways to order rectangles: map rectangle to:

- **CENTRE**: \((c_x, c_y)\), order along 2D curve
- **MIN-MAX**: \((a_x, a_y, z_x, z_y)\), order along 4D curve
- **CTR-SIZE**: \((c_x, c_y, w, h)\), order along 4D curve
What to do with rectangle data on Hilbert curves?

Three ways to order rectangles: map rectangle to:

- CENTRE: \((c_x, c_y)\), order along 2D curve
- MIN-MAX: \((a_x, a_y, z_x, z_y)\), order along 4D curve
- CTR-SIZE: \((c_x, c_y, w, h)\), order along 4D curve

Relative number of blocks visited by queries of size \(q\) in data Kamel and Faloutsos
What to do with rectangle data on Hilbert curves?

Three ways to order rectangles: map rectangle to:
- CENTRE: \((c_x, c_y)\), order along 2D curve
- MIN-MAX: \((a_x, a_y, z_x, z_y)\), order along 4D curve
- CTR-SIZE: \((c_x, c_y, w, h)\), order along 4D curve

Relative number of blocks visited by queries of size \(q\) in data Kamel and Faloutsos

Their conclusion: 4D curves are no good

Our conclusion: You used the wrong curve!
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D
Making Hilbert curves in 2D

There is only one way to do this
Making Hilbert curves in 3D
Making Hilbert curves in 3D
Making Hilbert curves in 3D
Making Hilbert curves in 3D
Making Hilbert curves in 3D
Making Hilbert curves in 3D
Making Hilbert curves in 3D

There are 1536 different 3D Hilbert curves...
Making Hilbert curves in 4D

There are 9782750684860313493469 different 4D Hilbert curves (more or less)

What does 4D implementation of Butz/Thomas/Moore do with point data?

- ordering by \((c_x, c_y, 0, 0)\) is not the same as by \((c_x, 0, c_y, 0)\) etc.
- ordering by \((c_x, c_y, 0, 0)\) is not as effective as by \((c_x, c_y)\) on 2D curve
Making Hilbert curves in 4D

There are 9782750684860313493469 different 4D Hilbert curves (more or less)

What does 4D implementation of Butz/Thomas/Moore do with point data?

- ordering by \((c_x, c_y, 0, 0)\) is not the same as by \((c_x, 0, c_y, 0)\) etc.
- ordering by \((c_x, c_y, 0, 0)\) is not as effective as by \((c_x, c_y)\) on 2D curve

Our curve: \((c_x, c_y, 0, 0)\) on 4D curve in same order as \((c_x, c_y)\) on 2D curve

TIGER road data

PR: PR-tree

With old curves:
C: \((c_x, c_y)\)
CS: \((c_x, c_y, w, h)\)
MM: \((a_x, a_y, z_x, z_y)\)

\[ q = 0.01° \quad 0.1° \quad 1° \]

our curve = 1
Making Hilbert curves in 4D

There are 9782750684860313493469 different 4D Hilbert curves (more or less)

What does 4D implementation of Butz/Thomas/Moore do with point data?

- ordering by \((c_x, c_y, 0, 0)\) is not the same as by \((c_x, 0, c_y, 0)\) etc.
- ordering by \((c_x, c_y, 0, 0)\) is not as effective as by \((c_x, c_y)\) on 2D curve

Our curve: \((c_x, c_y, 0, 0)\) on 4D curve in same order as \((c_x, c_y)\) on 2D curve

TIGER road data

- PR
- MM
- CS

VLSI data

- C
- PR
- MM

Our curve = 1

Ours is as good as the best for both types of data
Problems

2D: Can a recursive plane-filling curve have worst-case bbox area ratio $< 2$ ?
What if we rotate the bounding boxes?
What curve is best if we allow, say, octagonal bounding boxes?
2D: Can a recursive plane-filling curve have worst-case bbox area ratio $< 2$?
   What if we rotate the bounding boxes?
   What curve is best if we allow, say, octagonal bounding boxes?

4D: Our 4D Hilbert curve is weird and a rather random choice. Why this one?
   What (other) conditions should the curve fulfill?
   To get good bounding boxes? And good caching?
   Is there a better or a simpler curve? 4D-Peano?
Problems

2D: Can a recursive plane-filling curve have worst-case bbox area ratio < 2? What if we rotate the bounding boxes? What curve is best if we allow, say, octagonal bounding boxes?

4D: Our 4D Hilbert curve is weird and a rather random choice. Why this one? What (other) conditions should the curve fulfill? To get good bounding boxes? And good caching? Is there a better or a simpler curve? 4D-Peano?

References: mail cs.herman@haverkort.net and ask for our manuscripts:

Bulk-loading R-trees with four-dimensional space-filling curves
Space-filling curve properties for efficient spatial index structures (EuroCG 2008)
A comparison of space-filling curves for data structures (survey/notes)
Problems

2D: Can a recursive plane-filling curve have worst-case bbox area ratio $< 2$? What if we rotate the bounding boxes? What curve is best if we allow, say, octagonal bounding boxes?

4D: Our 4D Hilbert curve is weird and a rather random choice. Why this one? What (other) conditions should the curve fulfill? To get good bounding boxes? And good caching? Is there a better or a simpler curve? 4D-Peano?

References: mail cs.herman@haverkort.net and ask for our manuscripts:

*Bulk-loading R-trees with four-dimensional space-filling curves*

*Space-filling curve properties for efficient spatial index structures* (EuroCG 2008)

*A comparison of space-filling curves* (survey/notes)

That’s all folks