

CMU SCS

# 15-826: Multimedia Databases and Data Mining

*Multi-key and Spatial Access Methods - II*  
C. Faloutsos

---

---

---

---

---

---

---

---

CMU SCS

## Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- ➔ • Indexing - similarity search
- Data Mining

15-826 Copyright: C. Faloutsos (2007) 2

---

---

---

---

---

---

---

---

CMU SCS

## Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- ➔ • spatial access methods
  - problem defn
  - z-ordering
  - R-trees
  - ...
- text
- ...

15-826 Copyright: C. Faloutsos (2007) 3

---

---

---

---

---

---

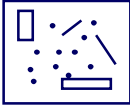
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer spatial queries (like??)



15-826 Copyright: C. Faloutsos (2007) 4

---

---

---

---

---

---

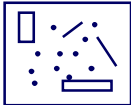
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer
  - point queries
  - range queries
  - k-nn queries
  - spatial joins ('all pairs' queries)



15-826 Copyright: C. Faloutsos (2007) 5

---

---

---

---

---

---

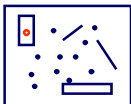
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer
  - point queries
  - range queries
  - k-nn queries
  - spatial joins ('all pairs' queries)



15-826 Copyright: C. Faloutsos (2007) 6

---

---

---

---

---

---

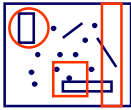
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer
  - point queries
  - range queries
  - k-nn queries
  - spatial joins ('all pairs' queries)



15-826 Copyright: C. Faloutsos (2007) 7

---

---

---

---

---

---

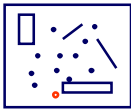
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer
  - point queries
  - range queries
  - k-nn queries
  - spatial joins ('all pairs' queries)



15-826 Copyright: C. Faloutsos (2007) 8

---

---

---

---

---

---

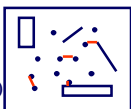
---

---

CMU SCS

### Spatial Access Methods - problem

- Given a collection of geometric objects (points, lines, polygons, ...)
- organize them on disk, to answer
  - point queries
  - range queries
  - k-nn queries
  - spatial joins ('all pairs' within  $\epsilon$ )



15-826 Copyright: C. Faloutsos (2007) 9

---

---

---

---

---

---

---

---

CMU SCS

## SAMs - motivation

- Q: applications?

15-826 Copyright: C. Faloutsos (2007) 10

---

---

---

---

---

---

---

---

CMU SCS

## SAMs - motivation

traditional DB      GIS

age

salary

15-826 Copyright: C. Faloutsos (2007) 11

---

---

---

---

---

---

---

---

CMU SCS

## SAMs - motivation

traditional DB      GIS

age

salary

15-826 Copyright: C. Faloutsos (2007) 12

---

---

---

---

---

---

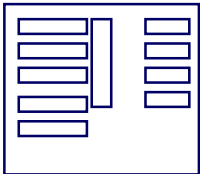
---

---

CMU SCS

### SAMs - motivation

CAD/CAM



find elements too close to each other

15-826 Copyright: C. Faloutsos (2007) 13

---

---

---

---

---

---

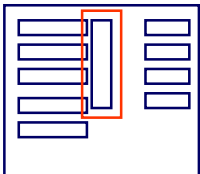
---

---

CMU SCS

### SAMs - motivation

CAD/CAM



15-826 Copyright: C. Faloutsos (2007) 14

---

---

---

---

---

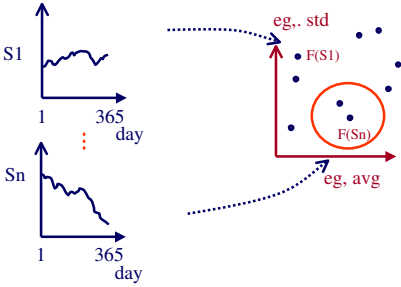
---

---

---

CMU SCS

### SAMs - motivation



15-826 Copyright: C. Faloutsos (2007) 15

---

---

---

---

---

---

---

---

CMU SCS

## Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
  - problem defn
  - ➔ – z-ordering
  - R-trees
  - ...
- text
- ...

15-826 Copyright: C. Faloutsos (2007) 16

---

---

---

---

---

---

---


---

CMU SCS

## SAMs: solutions

- z-ordering
- R-trees
- (grid files)

Q: how would you organize, e.g.,  $n$ -dim points, on disk? ( $C$  points per disk page)



15-826 Copyright: C. Faloutsos (2007) 17

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering


Q: how would you organize, e.g.,  $n$ -dim points, on disk? ( $C$  points per disk page)

Hint: reduce the problem to 1-d points (!!)

Q1: why?

A:

Q2: how?



15-826 Copyright: C. Faloutsos (2007) 18

---

---

---

---

---

---

---

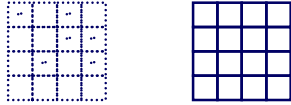
---



CMU SCS

### z-ordering

Q2.1: how to map  $n$ -d cells to 1-d cells?



15-826 Copyright: C. Faloutsos (2007) 22

---

---

---

---

---

---

---

---

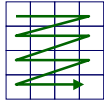
CMU SCS

### z-ordering

Q2.1: how to map  $n$ -d cells to 1-d cells?

A: row-wise

Q: is it good?



15-826 Copyright: C. Faloutsos (2007) 23

---

---

---

---

---

---

---

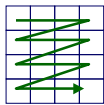
---

CMU SCS

### z-ordering

Q: is it good?

A: great for 'x' axis; bad for 'y' axis



15-826 Copyright: C. Faloutsos (2007) 24

---

---

---

---

---

---

---

---

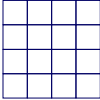




CMU SCS

## z-ordering

Q: solution? (w/ good clustering, and easy to compute, for 2-d and  $n$ -d?)



15-826 Copyright: C. Faloutsos (2007) 28

---

---

---

---

---

---

---

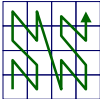
---

CMU SCS

## z-ordering

Q: solution? (w/ good clustering, and easy to compute, for 2-d and  $n$ -d?)

A: z-ordering/bit-shuffling/linear-quadtrees



'looks' better:

- few long jumps;
- scoops out the whole quadrant before leaving it
- a.k.a. space filling curves

15-826 Copyright: C. Faloutsos (2007) 29

---

---

---

---

---

---

---

---

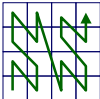
CMU SCS

## z-ordering

z-ordering/bit-shuffling/linear-quadtrees

Q: How to generate this curve ( $z = f(x,y)$ )?

A: 3 (equivalent) answers!



15-826 Copyright: C. Faloutsos (2007) 30

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

**z-ordering/bit-shuffling/linear-quadtrees**

Q: How to generate this curve ( $z = f(x,y)$ )?

A1: 'z' (or 'N') shapes, RECURSIVELY

order-1    order-2    ...    order (n+1)

15-826                      Copyright: C. Faloutsos (2007)                      31

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

Notice:

- self similar (we'll see about fractals, soon)
- method is hard to use:  $z = ? f(x,y)$

order-1    order-2    ...    order (n+1)

15-826                      Copyright: C. Faloutsos (2007)                      32

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

**z-ordering/bit-shuffling/linear-quadtrees**

Q: How to generate this curve ( $z = f(x,y)$ )?

A: 3 (equivalent) answers!

Method #2?

15-826                      Copyright: C. Faloutsos (2007)                      33

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

**bit-shuffling**

x

0 0

1 1

$z = (0101)_2 = 5$

15-826 Copyright: C. Faloutsos (2007) 34

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

**bit-shuffling**

x

0 0

1 1

$z = (0101)_2 = 5$

How about the reverse:  
 $(x,y) = g(z)$ ?

15-826 Copyright: C. Faloutsos (2007) 35

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

**bit-shuffling**

x

0 0

1 1

$z = (0101)_2 = 5$

How about  $n$ -d spaces?

15-826 Copyright: C. Faloutsos (2007) 36

---

---

---

---

---

---

---

---

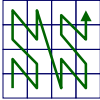
CMU SCS

## z-ordering

z-ordering/bit-shuffling/**linear-quadtrees**

Q: How to generate this curve ( $z = f(x,y)$ )?

A: 3 (equivalent) answers!



Method #3?

15-826 Copyright: C. Faloutsos (2007) 37

---

---

---

---

---

---

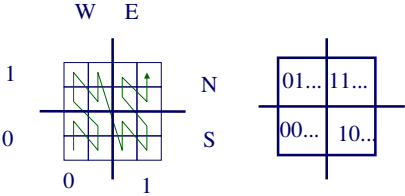
---

---

CMU SCS

## z-ordering

**linear-quadtrees** : assign N->1, S->0 e.t.c.



15-826 Copyright: C. Faloutsos (2007) 38

---

---

---

---

---

---

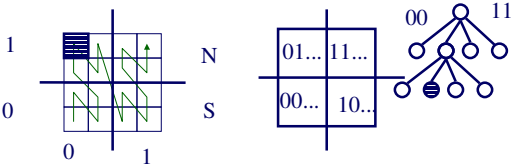
---

---

CMU SCS

## z-ordering

... and repeat recursively. Eg.:  $z_{\text{blue-cell}} = WN; WN = (0101)_2 = 5$



15-826 Copyright: C. Faloutsos (2007) 39

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

Drill: z-value of magenta cell, with the three methods?

15-826 Copyright: C. Faloutsos (2007) 40

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

Drill: z-value of magenta cell, with the three methods?

method#1: 14  
method#2: shuffle(11;10)=  
 $(1110)_2 = 14$

15-826 Copyright: C. Faloutsos (2007) 41

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering

Drill: z-value of magenta cell, with the three methods?

method#1: 14  
method#2: shuffle(11;10)=  
 $(1110)_2 = 14$   
method#3: EN;ES = ... = 14

15-826 Copyright: C. Faloutsos (2007) 42

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering - Detailed outline

- spatial access methods
  - z-ordering
    - main idea - 3 methods
    - use w/ B-trees; algorithms (range, knn queries ...)
    - non-point (eg., region) data
    - analysis; variations
  - R-trees
  - ...

15-826 Copyright: C. Faloutsos (2007) 43

---

---

---

---

---

---

---

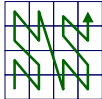
---

CMU SCS

## z-ordering - usage & algo's

Q1: How to store on disk?  
A:

Q2: How to answer range queries etc



15-826 Copyright: C. Faloutsos (2007) 44

---

---

---

---

---

---

---

---

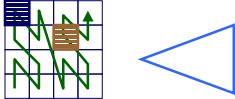
CMU SCS

## z-ordering - usage & algo's

Q1: How to store on disk?  
A: treat z-value as primary key; feed to B-tree

PGH

SF



z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 45

---

---

---

---

---

---

---

---

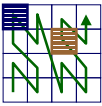
CMU SCS

### z-ordering - usage & algo's

MAJOR ADVANTAGES w/ B-tree:

- already inside commercial systems (no coding/debugging!)
- concurrency & recovery is ready

PGH

SF 

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 46

---

---

---

---

---

---

---

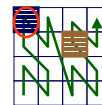
---

CMU SCS

### z-ordering - usage & algo's

Q2: queries? (eg.: *find city at (0,3)* )?

PGH

SF 

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 47

---

---

---

---

---

---

---

---

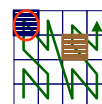
CMU SCS

### z-ordering - usage & algo's

Q2: queries? (eg.: *find city at (0,3)* )?

A: find z-value; search B-tree

PGH

SF 

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 48

---

---

---

---

---

---

---

---



CMU SCS

## z-ordering - usage & algo's

Q2: range queries?

SF

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 49

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering - usage & algo's

Q2: range queries?  
A: compute ranges of z-values; use B-tree

SF

9,11-15

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 50

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering - usage & algo's

Q2': range queries - how to reduce # of qualifying of ranges?

SF

9,11-15

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 51

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q2': range queries - how to reduce # of qualifying of ranges?

A: Augment the query!

PGH 9,11-15 -> 8-15

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 52

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q2'': range queries - how to break a query into ranges?

9,11-15

15-826 Copyright: C. Faloutsos (2007) 53

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q2'': range queries - how to break a query into ranges?

A: recursively, quadtree-style; decompose only non-full quadrants

12-15                      9,11-15

15-826 Copyright: C. Faloutsos (2007) 54

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q2'': range queries - how to break a query into ranges?  
 A: recursively, quadtree-style; decompose only non-full quadrants

15-826 Copyright: C. Faloutsos (2007) 55

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - Detailed outline

- spatial access methods
  - z-ordering
    - main idea - 3 methods
    - use w/ B-trees; algorithms (range, knn queries ...)
    - non-point (eg., region) data
    - analysis; variations
  - R-trees
  - ...

15-826 Copyright: C. Faloutsos (2007) 56

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q3: k-nn queries? (say, 1-nn)?

SF PGH

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 57

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

Q3: k-nn queries? (say, 1-nn)?  
 A: traverse B-tree; find nn wrt z-values and ...

PGH

SF

z	cname	etc
5	SF	
12	PGH	

15-826 Copyright: C. Faloutsos (2007) 58

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

... ask a range query.

PGH

SF

nn wrt z-value

15-826 Copyright: C. Faloutsos (2007) 59

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - usage & algo's

... ask a range query.

PGH

SF

nn wrt z-value

15-826 Copyright: C. Faloutsos (2007) 60

---

---

---

---

---

---

---

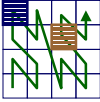
---

CMU SCS

## z-ordering - usage & algo's

Q4: all-pairs queries? ( *all pairs of cities within 10 miles from each other?* )

PGH

SF  (we'll see 'spatial joins' later: *find all PA counties that intersect a lake*)

15-826 Copyright: C. Faloutsos (2007) 61

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering - Detailed outline

- spatial access methods
  - z-ordering
    - main idea - 3 methods
    - use w/ B-trees; algorithms (range, knn queries ...)
    - non-point (eg., region) data
    - analysis; variations
  - R-trees
  - ...

15-826 Copyright: C. Faloutsos (2007) 62

---

---

---

---

---

---

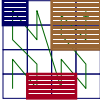
---

---

CMU SCS

## z-ordering - regions

Q: z-value for a region?

A 

B

C

$z_B = ??$

$z_C = ??$

15-826 Copyright: C. Faloutsos (2007) 63

---

---

---

---

---

---

---

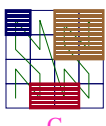
---

CMU SCS

### z-ordering - regions

Q: z-value for a region?  
 A: 1 or more z-values; by quadtree decomposition

A



C

$z_B = ??$   
 $z_C = ??$

15-826 Copyright: C. Faloutsos (2007) 64

---

---

---

---

---

---

---

---

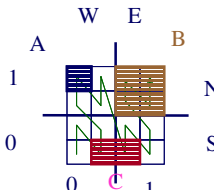
CMU SCS

### z-ordering - regions

Q: z-value for a region?

W E

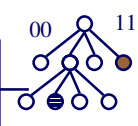
A B



C

$z_B = 11^{**}$  ← "don't care"

$z_C = ??$



15-826 Copyright: C. Faloutsos (2007) 65

---

---

---

---

---

---

---

---

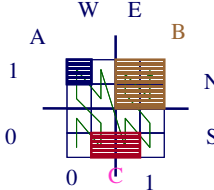
CMU SCS

### z-ordering - regions

Q: z-value for a region?

W E

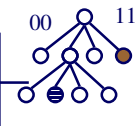
A B



C

$z_B = 11^{**}$  ← "don't care"

$z_C = \{0010; 1000\}$



15-826 Copyright: C. Faloutsos (2007) 66

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Q: How to store in B-tree?  
Q: How to search (range etc queries)

15-826 Copyright: C. Faloutsos (2007) 67

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Q: How to store in B-tree? A: sort (\*<0<1)  
Q: How to search (range etc queries)

z	obj-id	etc
0010	C	
0101	A	
1000	C	
11**	B	

15-826 Copyright: C. Faloutsos (2007) 68

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Q: How to search (range etc queries) - eg 'red' range query

z	obj-id	etc
0010	C	
0101	A	
1000	C	
11**	B	

15-826 Copyright: C. Faloutsos (2007) 69

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Q: How to search (range etc queries) - eg 'red' range query

A: break query in z-values; check B-tree

z	obj-id	etc
0010	C	
0101	A	
1000	C	
11**	B	

15-826 Copyright: C. Faloutsos (2007) 70

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Almost identical to range queries for point data, except for the "don't cares" - i.e.,

1100 ?? 11\*\*

z	obj-id	etc
0010	C	
0101	A	
1000	C	
11**	B	

15-826 Copyright: C. Faloutsos (2007) 71

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Almost identical to range queries for point data, except for the "don't cares" - i.e.,

$z1 = 1100 \text{ ?? } 11^{**} = z2$

Specifically: does z1 contain/avoid/intersect z2?

Q: what is the criterion to decide?

15-826 Copyright: C. Faloutsos (2007) 72

---

---

---

---

---

---

---

---



CMU SCS

### z-ordering - regions

$z1 = 1100 \quad ?? \quad 11^{**} = z2$

Specifically: does  $z1$  contain/avoid/intersect  $z2$ ?

Q: what is the criterion to decide?

A: **Prefix property:** let  $r1, r2$  be the corresponding regions, and let  $r1$  be the smallest ( $\Rightarrow z1$  has fewest '\*'s). Then:

15-826 Copyright: C. Faloutsos (2007) 73

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

- $r2$  will either contain completely, or avoid completely  $r1$ .
- it will contain  $r1$ , if  $z2$  is the prefix of  $z1$

$1100 \quad ?? \quad 11^{**}$

region of  $z1$ :  
completely contained in  
region of  $z2$

15-826 Copyright: C. Faloutsos (2007) 74

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Drill (True/False). Given:

- $z1 = 011001^{**}$
- $z2 = 01^{*****}$
- $z3 = 0100^{****}$

T/F  $r2$  contains  $r1$

T/F  $r3$  contains  $r1$

T/F  $r3$  contains  $r2$

15-826 Copyright: C. Faloutsos (2007) 75

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Drill (True/False). Given:

- $z1 = 011001**$
- $z2 = 01*****$
- $z3 = 0100****$

T/F  $r2$  contains  $r1$  - TRUE (prefix property)  
 T/F  $r3$  contains  $r1$  - FALSE (disjoint)  
 T/F  $r3$  contains  $r2$  - FALSE ( $r2$  contains  $r3$ )

15-826 Copyright: C. Faloutsos (2007) 76

---

---

---

---

---

---

---

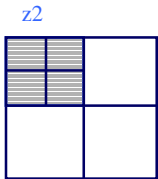
---

CMU SCS

### z-ordering - regions

Drill (True/False). Given:

- $z1 = 011001**$
- $z2 = 01*****$
- $z3 = 0100****$



15-826 Copyright: C. Faloutsos (2007) 77

---

---

---

---

---

---

---

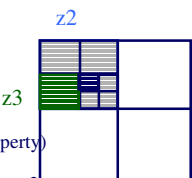
---

CMU SCS

### z-ordering - regions

Drill (True/False). Given:

- $z1 = 011001**$
- $z2 = 01*****$
- $z3 = 0100****$



T/F  $r2$  contains  $r1$  - TRUE (prefix property)  
 T/F  $r3$  contains  $r1$  - FALSE (disjoint)  
 T/F  $r3$  contains  $r2$  - FALSE ( $r2$  contains  $r3$ )

15-826 Copyright: C. Faloutsos (2007) 78

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

**Spatial joins:** find (quickly) all  
counties intersecting lakes

15-826 Copyright: C. Faloutsos (2007) 79

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

**Spatial joins:** find (quickly) all  
counties intersecting lakes

Naive algorithm:  $O(N * M)$   
 Something faster?

15-826 Copyright: C. Faloutsos (2007) 80

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

**Spatial joins:** find (quickly) all  
counties intersecting lakes

z	obj-id	etc
0010	ALG	
...	...	
1000	WAS	
11**	ALG	

z	obj-id	etc
0011	Erie	
0101	Erie	
...		
10**	Ont.	

15-826 Copyright: C. Faloutsos (2007) 81

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - regions

Spatial joins: find (quickly) all  
**counties** intersecting lakes

Solution: merge the lists of (sorted) z-values,  
looking for the prefix property

footnote#1: '\*' needs careful treatment  
footnote#2: need dup. elimination

15-826 Copyright: C. Faloutsos (2007) 82

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - Detailed outline

- spatial access methods
  - z-ordering
    - main idea - 3 methods
    - use w/ B-trees; algorithms (range, knn queries ...)
    - non-point (eg., region) data
    - analysis; variations
  - R-trees
  - ...



---

---

---

---

---

---

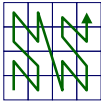
---

---

CMU SCS

### z-ordering - variations

Q: is z-ordering the best we can do?



15-826 Copyright: C. Faloutsos (2007) 84

---

---

---

---

---

---

---

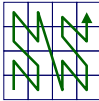
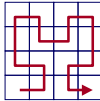
---



CMU SCS

### z-ordering - variations

'Looks' better (never long jumps). How to derive it?

15-826 Copyright: C. Faloutsos (2007) 88

---

---

---

---

---

---


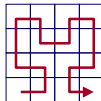
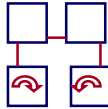
---

---

CMU SCS

### z-ordering - variations

'Looks' better (never long jumps). How to derive it?

order-1                  order-2                  ...          order (n+1)

15-826 Copyright: C. Faloutsos (2007) 89

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - variations

Q: function for the Hilbert curve ( $h = f(x,y)$ )?  
 A: bit-shuffling, followed by post-processing, to account for rotations. Linear on # bits.  
 See textbook, for pointers to code/algorithms (eg., [Jagadish, 90])

15-826 Copyright: C. Faloutsos (2007) 90

---

---

---

---

---

---


---

---


CMU SCS

### z-ordering - variations


Q: how about Hilbert curve in 3-d? n-d?  
 A: Exists (and is not unique!). Eg., 3-d, order-1 Hilbert curves (Hamiltonian paths on cube)



#1



#2



15-826 Copyright: C. Faloutsos (2007) 91

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - Detailed outline

- spatial access methods
  - z-ordering
    - main idea - 3 methods
    - use w/ B-trees; algorithms (range, knn queries ...)
    - non-point (eg., region) data
    - analysis; variations
  - R-trees
  - ...

15-826 Copyright: C. Faloutsos (2007) 92

---

---

---

---

---

---


---

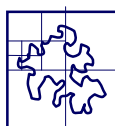
---

CMU SCS

### z-ordering - analysis

Q: How many pieces ('quad-tree blocks') per region?  
 A: proportional to perimeter (surface etc)





15-826 Copyright: C. Faloutsos (2007) 93

---

---

---

---

---

---

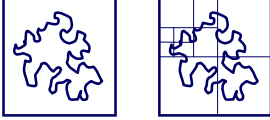
---

---

CMU SCS

### z-ordering - analysis

(How long is the coastline, say, of England?  
Paradox: The answer changes with the yardstick -> fractals ...)



15-826 Copyright: C. Faloutsos (2007) 94

---

---

---

---

---

---

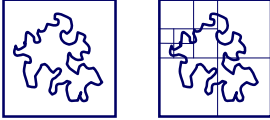
---

---

CMU SCS

### z-ordering - analysis

Q: Should we decompose a region to full detail (and store in B-tree)?



15-826 Copyright: C. Faloutsos (2007) 95

---

---

---

---

---

---

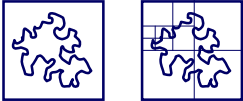
---

---

CMU SCS

### z-ordering - analysis

Q: Should we decompose a region to full detail (and store in B-tree)?  
A: NO! approximation with 1-3 pieces/z-values is best [Orenstein90]



15-826 Copyright: C. Faloutsos (2007) 96

---

---

---

---

---

---

---

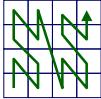
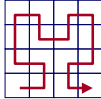
---



CMU SCS

### z-ordering - analysis

Q: how to measure the 'goodness' of a curve?

15-826 Copyright: C. Faloutsos (2007) 97

---

---

---

---

---

---

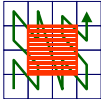
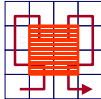
---

---

CMU SCS

### z-ordering - analysis

Q: how to measure the 'goodness' of a curve?  
 A: e.g., avg. # of runs, for range queries

4 runs
3 runs

(#runs ~ #disk accesses on B-tree)

15-826 Copyright: C. Faloutsos (2007) 98

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - analysis

Q: So, is Hilbert really better?  
 A: 27% fewer runs, for 2-d (similar for 3-d)

Q: are there formulas for #runs, #of quadtree blocks etc?  
 A: Yes ([Jagadish; Moon+ etc] see textbook)

15-826 Copyright: C. Faloutsos (2007) 99

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - fun observations

Hilbert and z-ordering curves: "space filling curves": eventually, they visit every point in n-d space - therefore:

order-1      order-2      ... order (n+1)

15-826      Copyright: C. Faloutsos (2007)      100

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - fun observations

... they show that the plane has as many points as a line (-> headaches for 1900's mathematics/topology). (fractals, again!)

order-1      order-2      ... order (n+1)

15-826      Copyright: C. Faloutsos (2007)      101

---

---

---

---

---

---

---

---

CMU SCS

### z-ordering - fun observations

Observation #2: Hilbert (like) curve for video encoding [Y. Matias+, CRYPTO '87]:  
Given a frame, visit its pixels in randomized hilbert order; compress; and transmit

15-826      Copyright: C. Faloutsos (2007)      102

---

---

---

---

---

---

---

---

CMU SCS

## z-ordering - fun observations

In general, Hilbert curve is great for preserving distances, clustering, vector quantization etc

15-826 Copyright: C. Faloutsos (2007) 103

---

---

---

---

---

---

---

---

CMU SCS

## Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
  - problem defn
  - z-ordering
  - ➡ - R-trees
  - ...
- text
- ...

15-826 Copyright: C. Faloutsos (2007) 104

---

---

---

---

---

---

---

---

CMU SCS

## Conclusions

- z-ordering is a great idea (n-d points -> 1-d points; feed to B-trees)
- used by TIGER system and (most probably) by other GIS products
- works great with low-dim points

15-826 Copyright: C. Faloutsos (2007) 105

---

---

---

---

---

---

---

---