MADALGO

CENTER FOR MASSIVE DATA ALGORITHMICS

PhD Defence

Jakob Truelsen

Danmarks
Grundforskningsfond
Danish National
Research Foundation

AARHUS
UNIVERSITY

Welcome

Overview

Overview

Theoretical Computer Science

Overview

Theoretical Computer Science

Large Datasets

Overview

Theoretical Computer Science

Large Datasets

- Decomposition Simplification

Overview

Theoretical Computer Science

Large Datasets

- Decomposition Simplification
- Multiresolution rasters

Overview

- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures

Overview

- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures
  - Working Set Dictonary

Overview

- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures
  - Working Set Dictonary
  - Finger Search Dictonary

Overview

- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures
  - Working Set Dictonary
  - Finger Search Dictonary
  - Lower Bounds

Overview

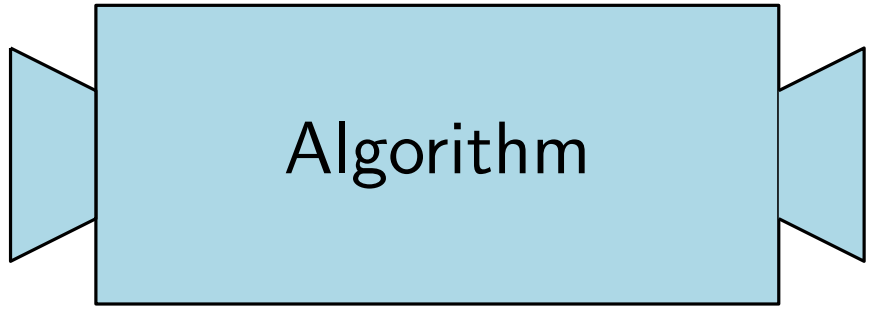
- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures
  - Working Set Dictonary
  - Finger Search Dictonary
  - Lower Bounds
  - Priority Queue

Overview

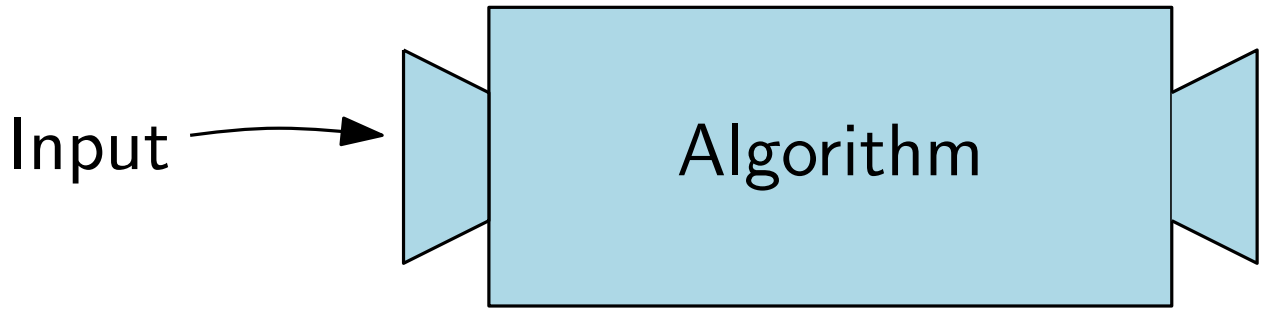
- Theoretical Computer Science
- Large Datasets
  - Decomposition Simplification
  - Multiresolution rasters
- Implicite Structures
  - Working Set Dictonary
  - Finger Search Dictonary
  - Lower Bounds
  - Priority Queue
- Range Mode

Theoretical Computer Science

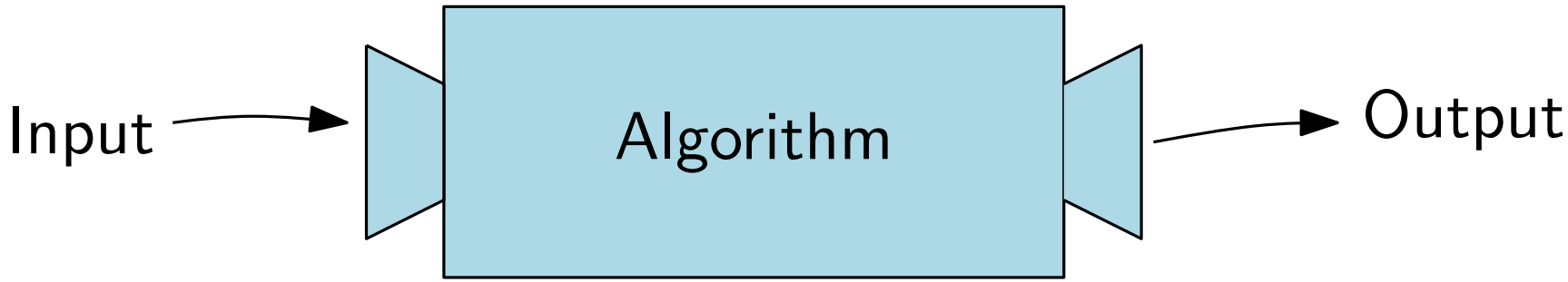
Theoretical Computer Science



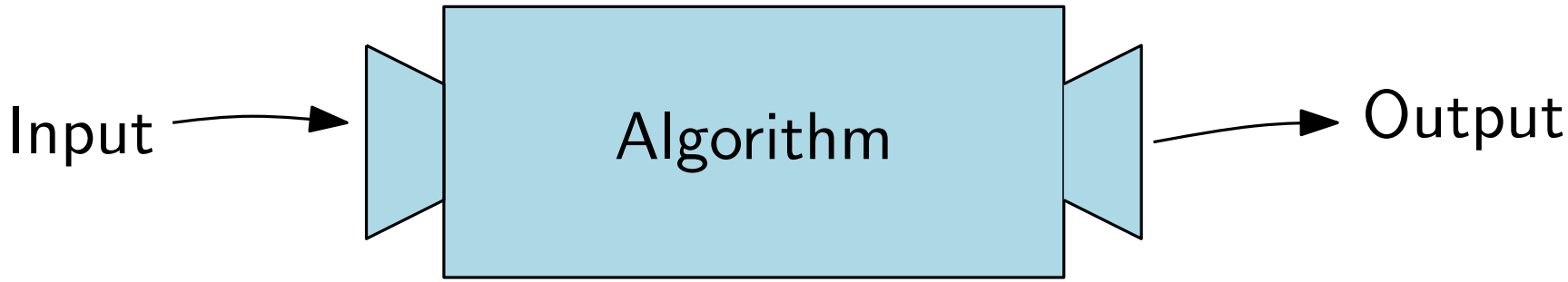
Theoretical Computer Science



Theoretical Computer Science

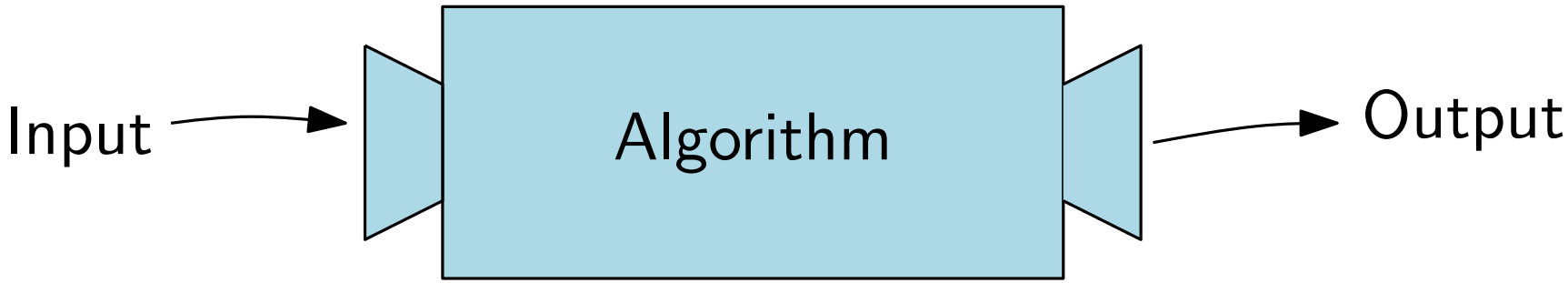


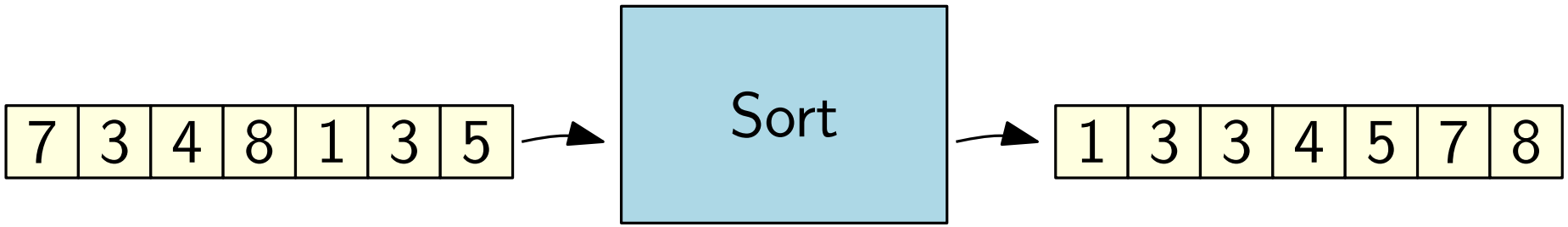
Theoretical Computer Science

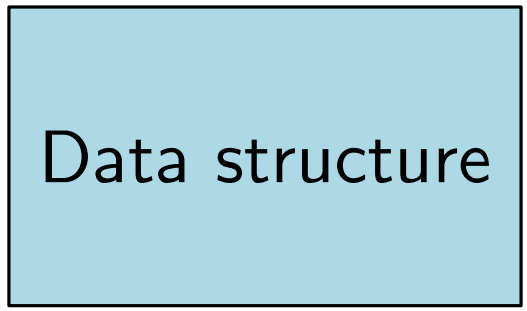


7	3	4	8	1	3	5

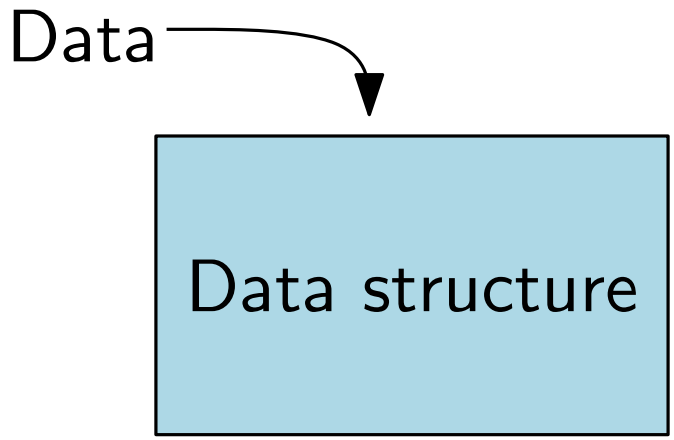
Theoretical Computer Science



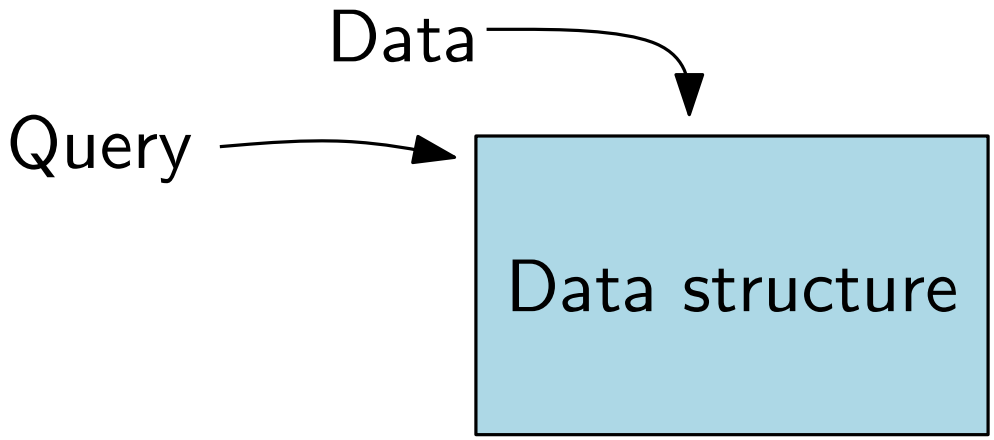



Data structure

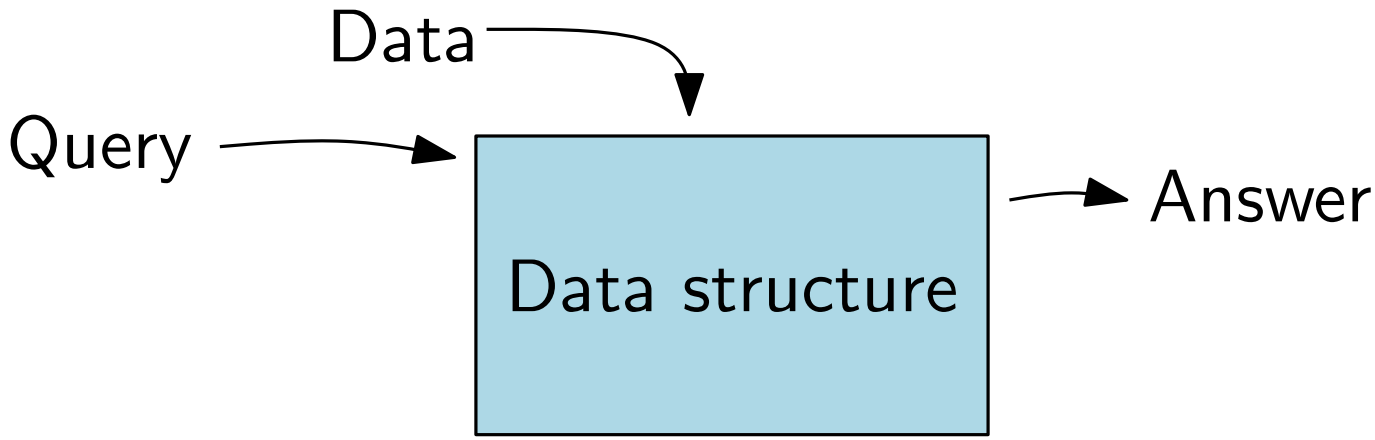
Theoretical Computer Science



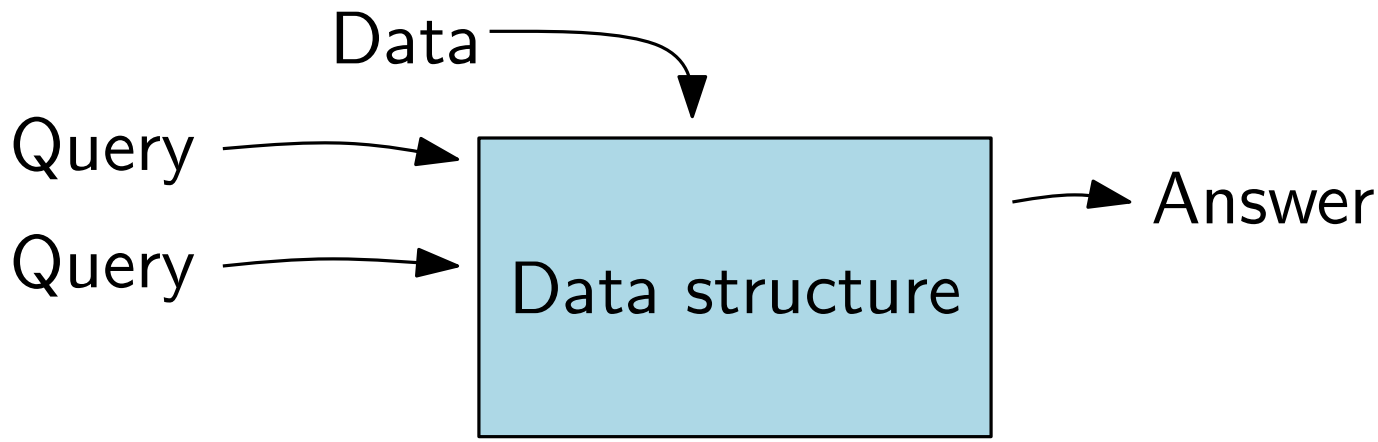
Theoretical Computer Science



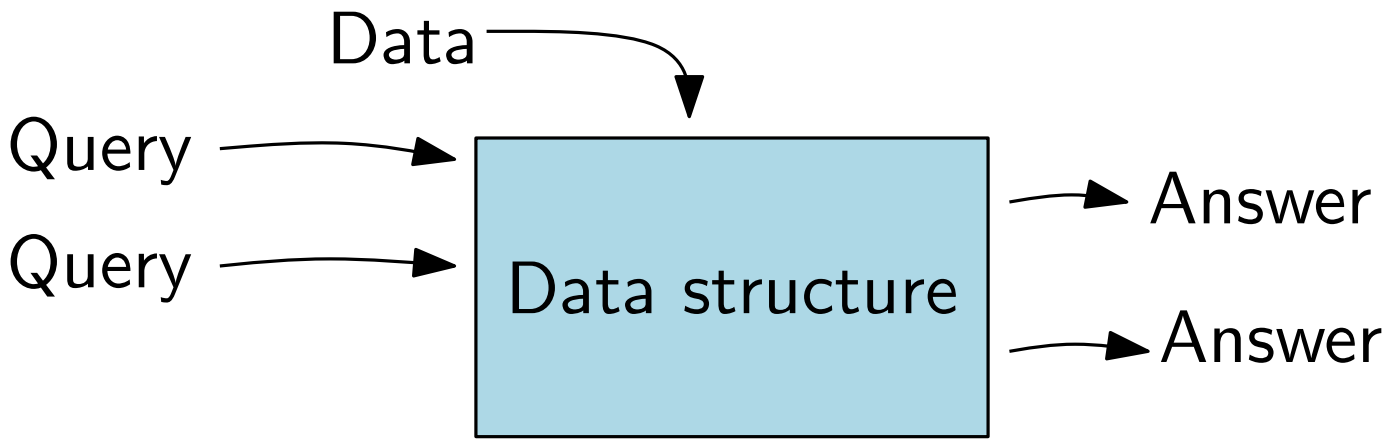
Theoretical Computer Science



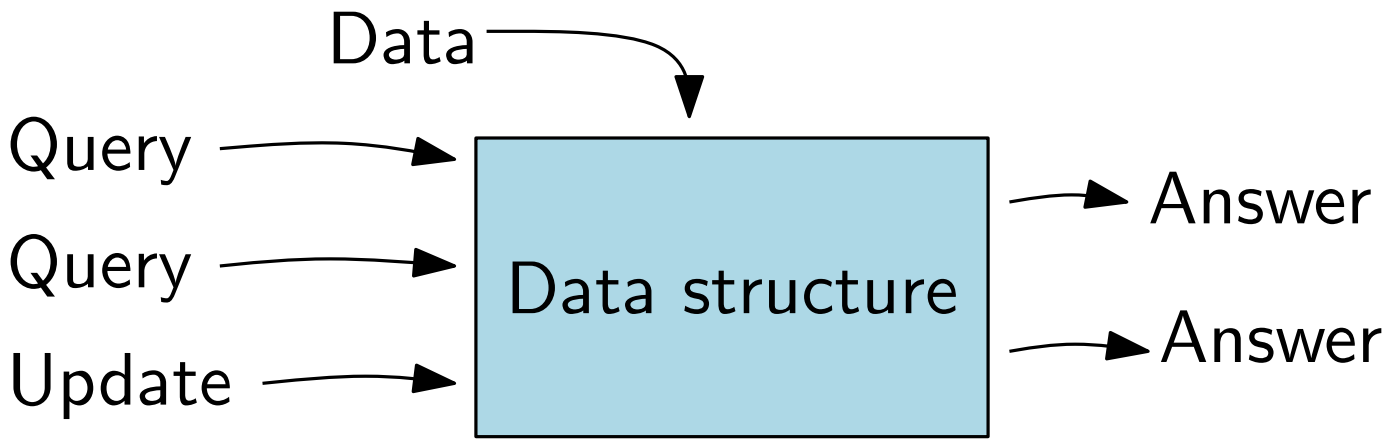
Theoretical Computer Science



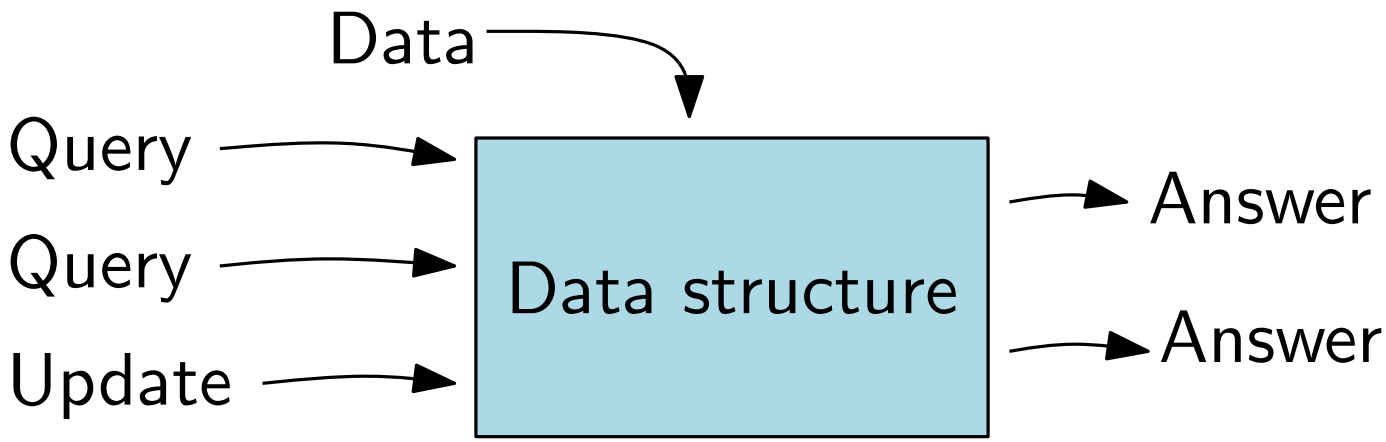
Theoretical Computer Science



Theoretical Computer Science

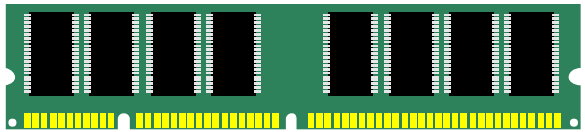


Theoretical Computer Science

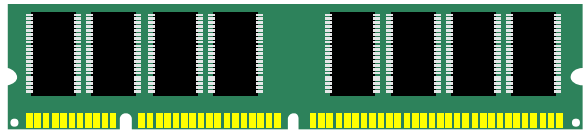


Alice	1324
Bob	4315
Carlos	6875
Eve	3542

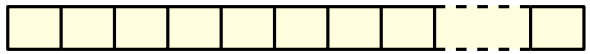
Large Datasets



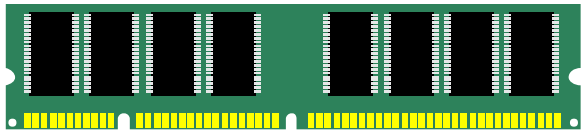
Large Datasets

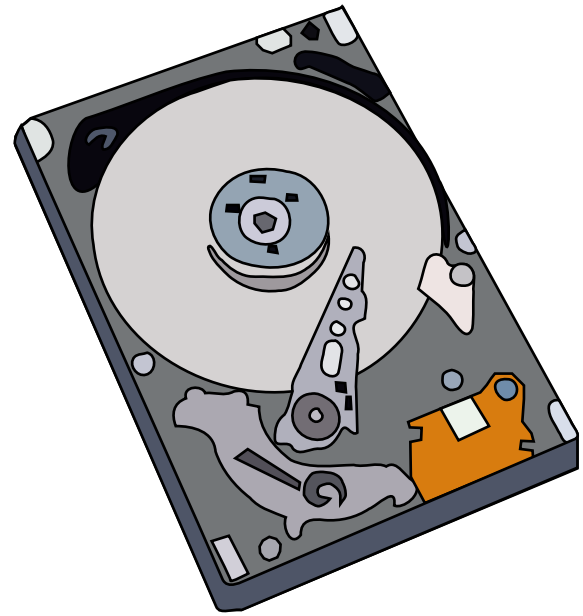


0 1 2 3 4 5 6 7 $\cdots$ 6442450944

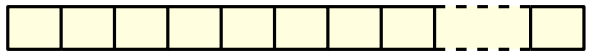


Large Datasets

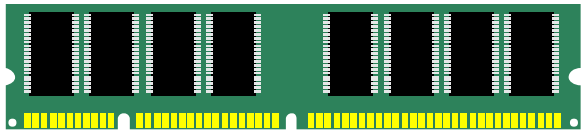


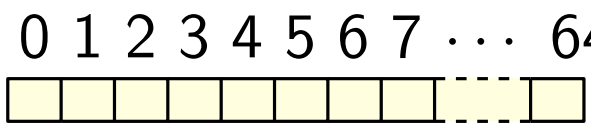


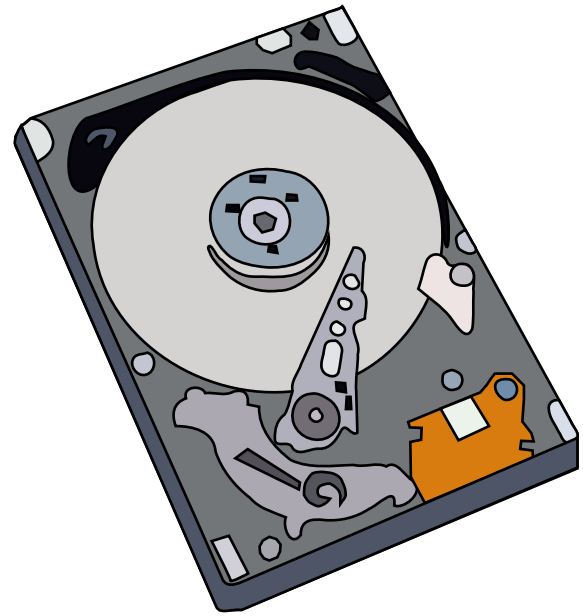
0 1 2 3 4 5 6 7 $\cdots$ 6442450944



Large Datasets

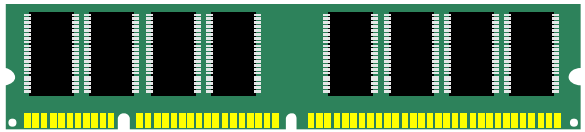


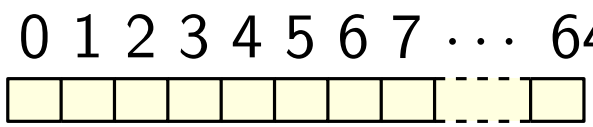


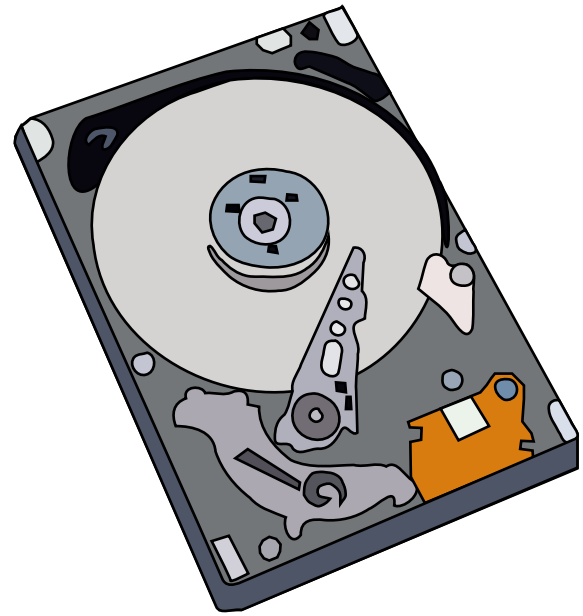


Size:	48 GiB	21 TiB	($\times$ 500)

Large Datasets

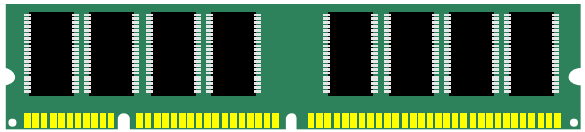


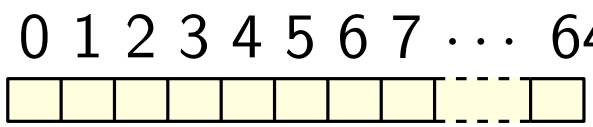


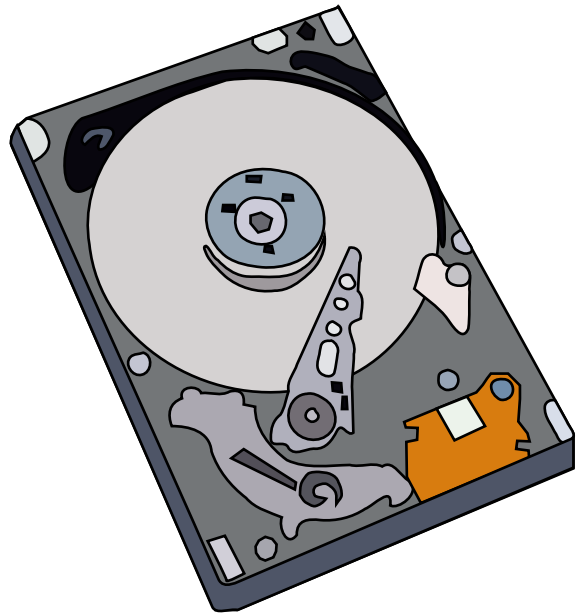


Size:	48 GiB	21 TiB	($\times$ 500)
Access Time:	9 ns	9 ms	($\times$ 1000000)

Large Datasets



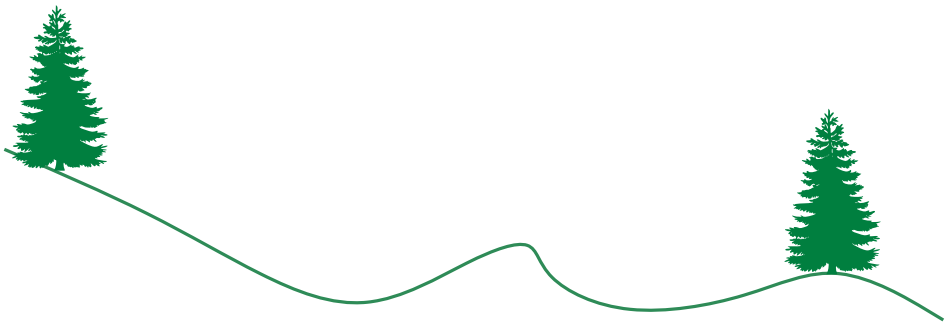




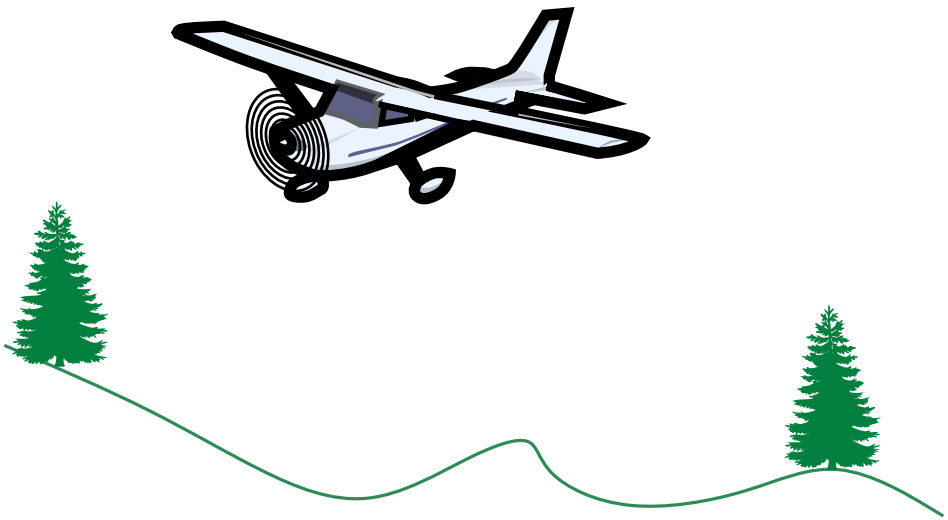
Size:	48 GiB	21 TiB	($\times$ 500)
Access Time:	9 ns	9 ms	($\times$ 1000000)
Throughput:	8 GiB/s	450 MiB/s	($\div$ 18)

Large Datasets

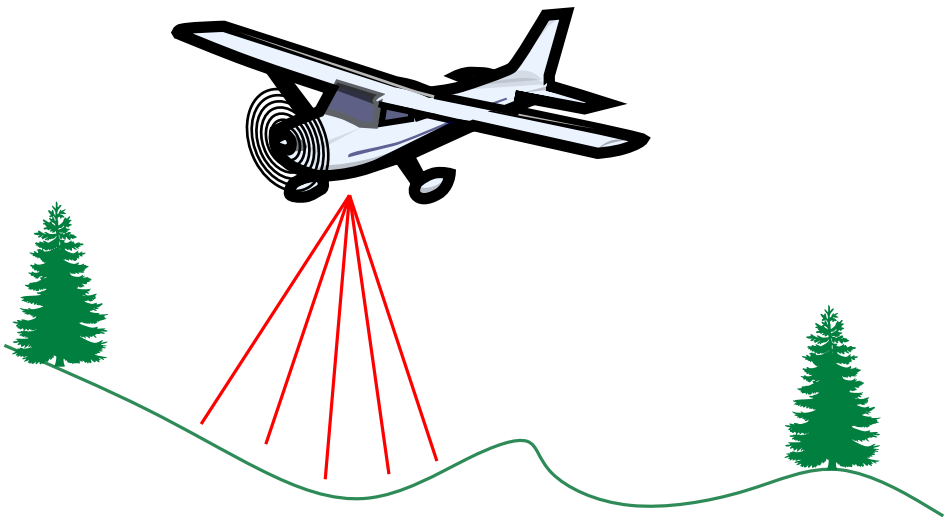
Large Datasets



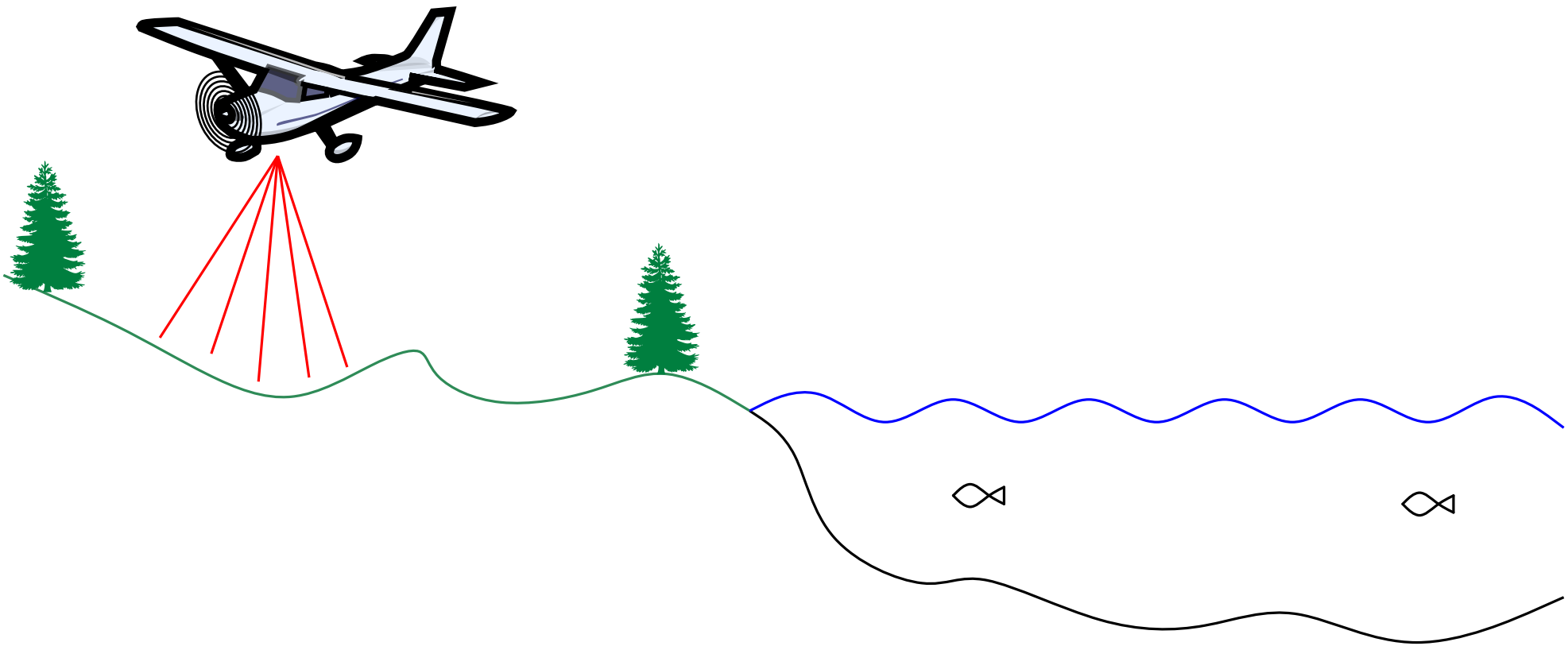
Large Datasets



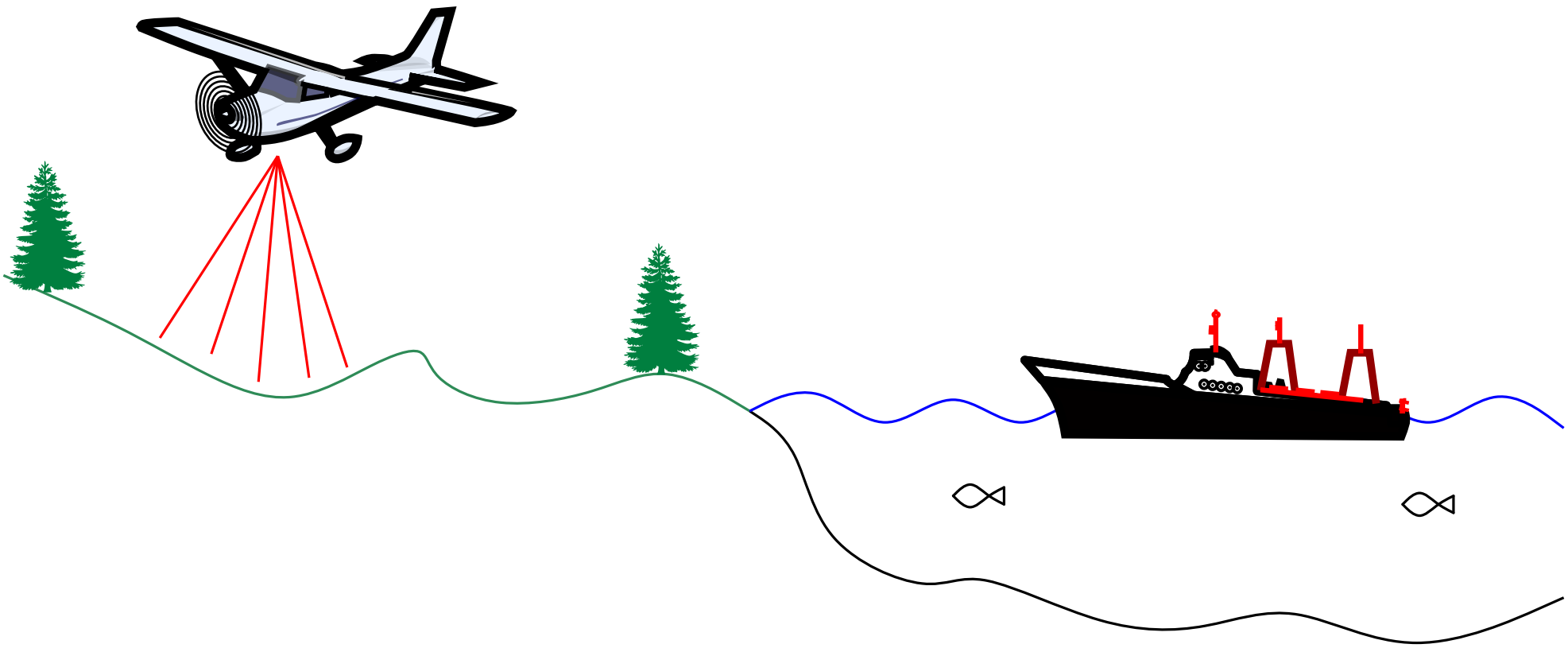
Large Datasets



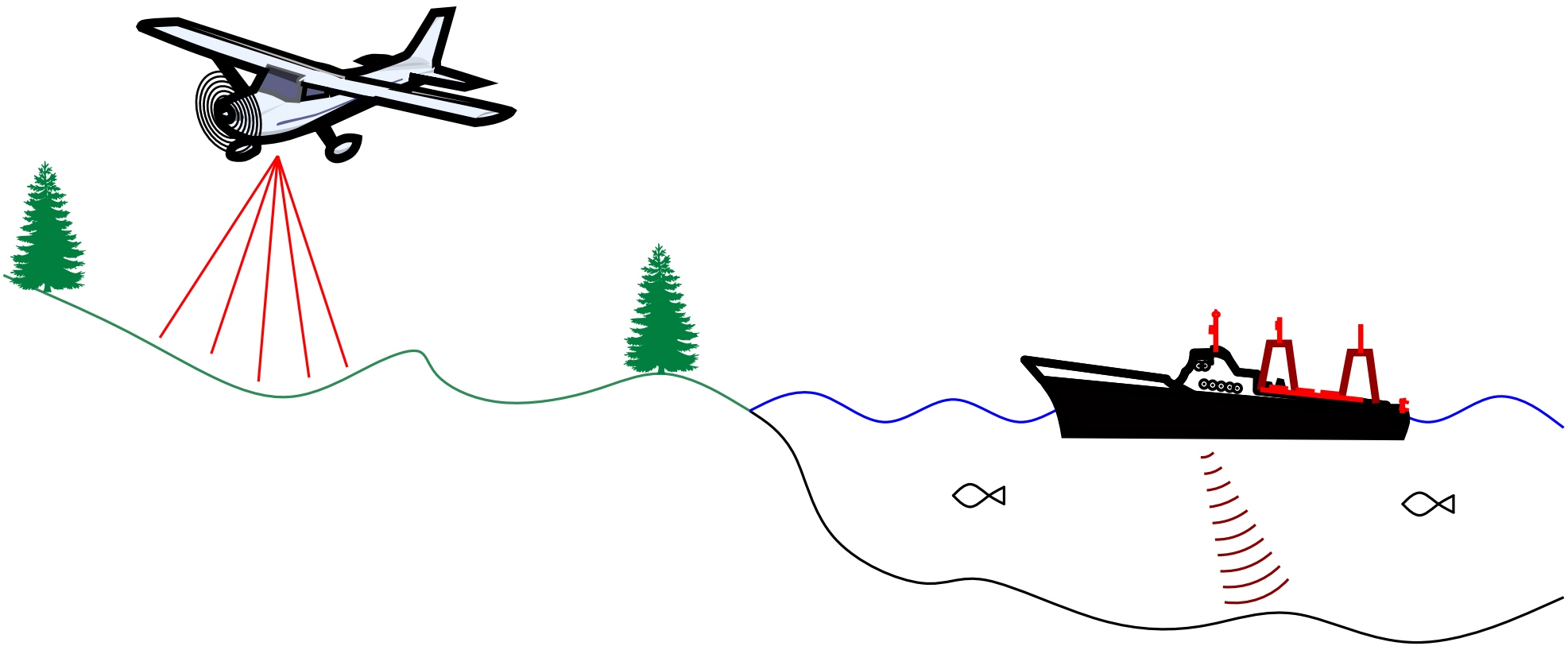
Large Datasets



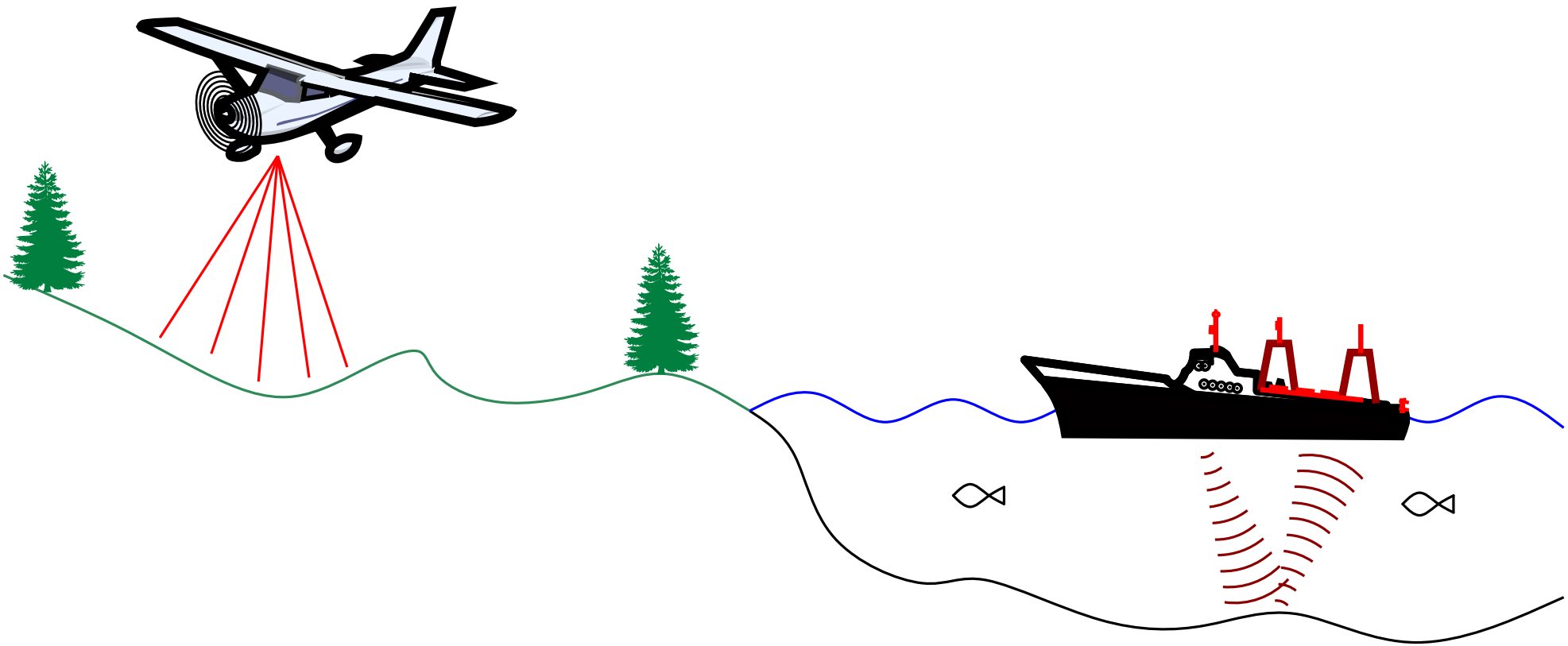
Large Datasets

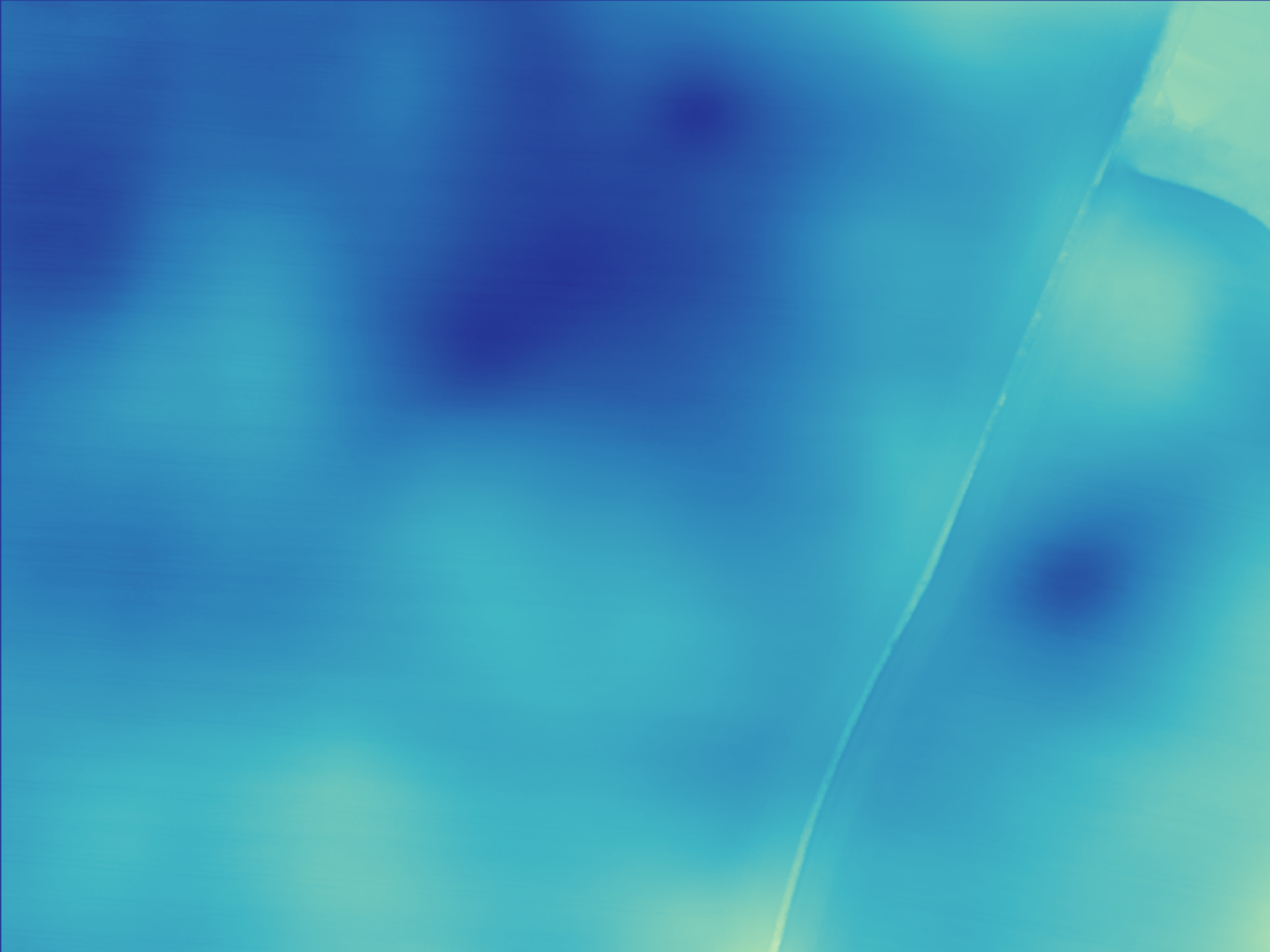


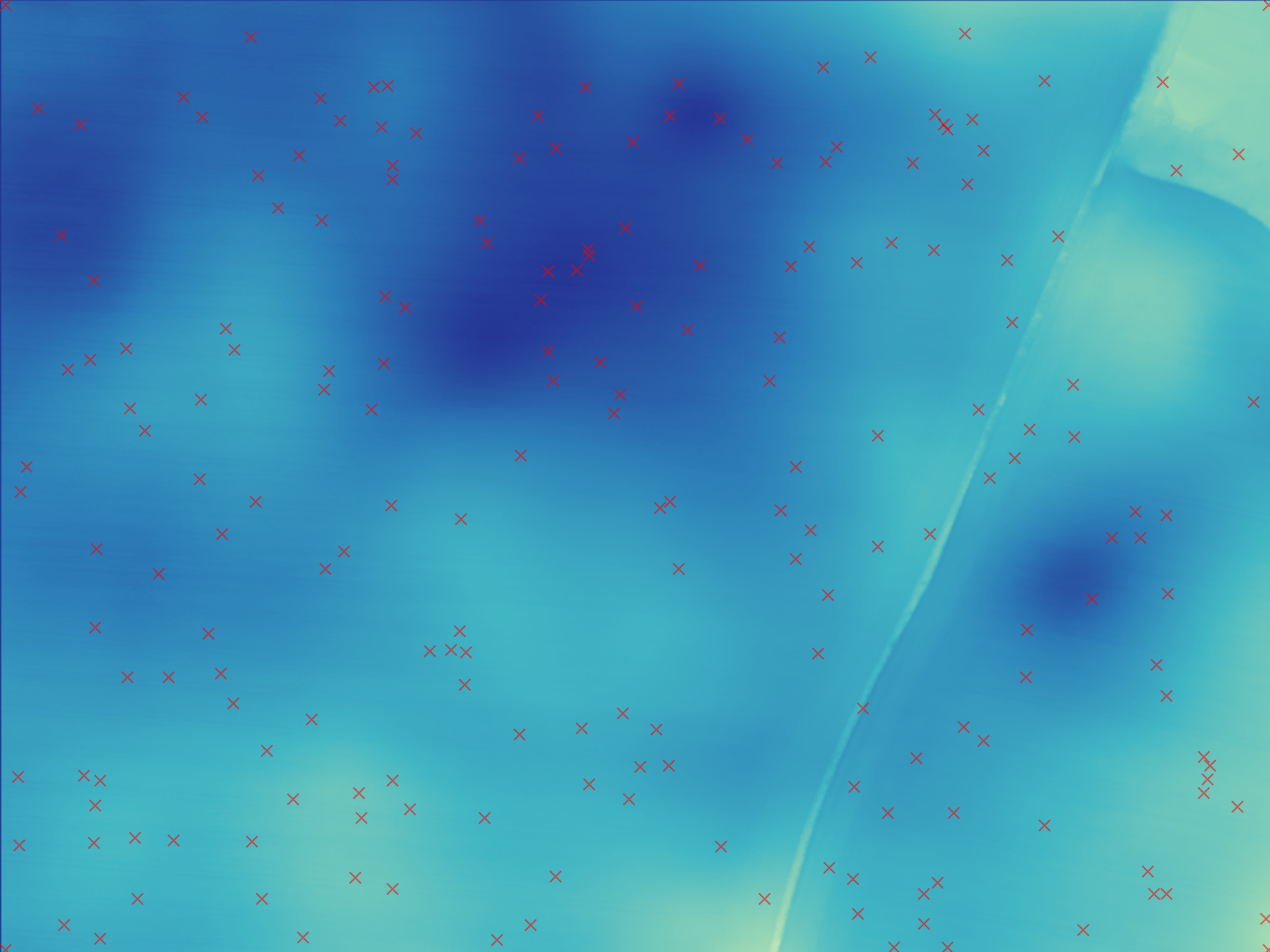
Large Datasets

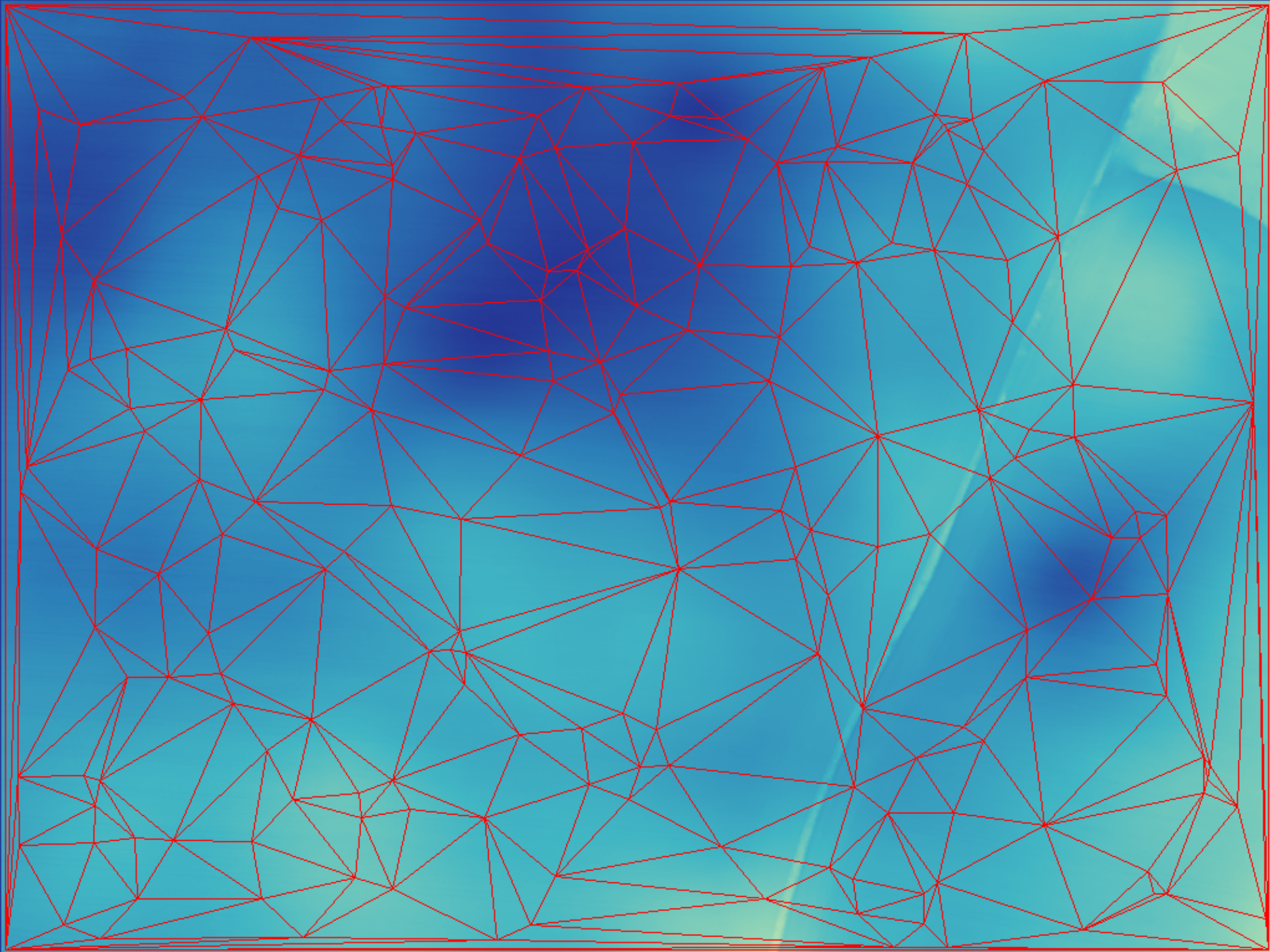


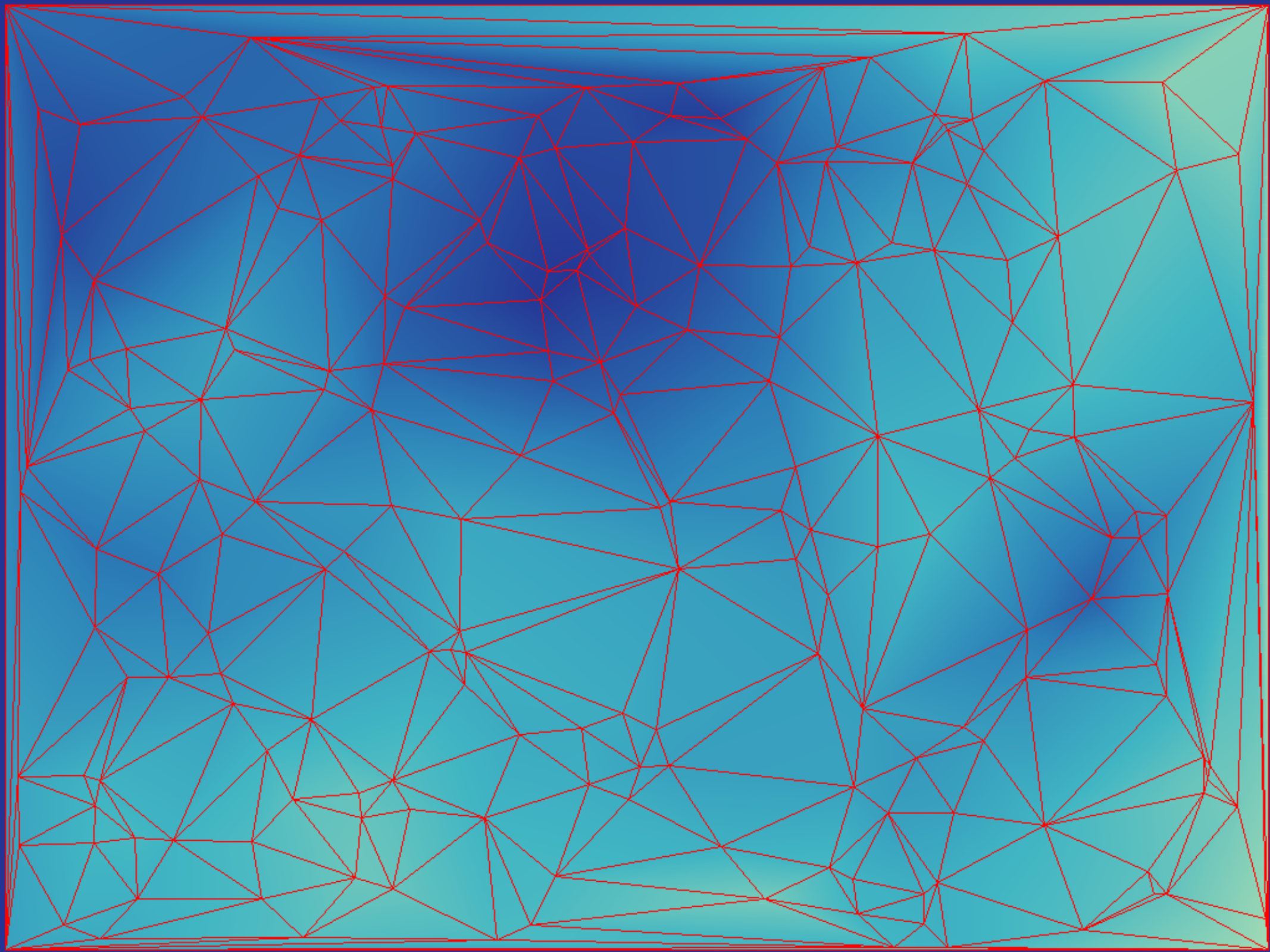
Large Datasets

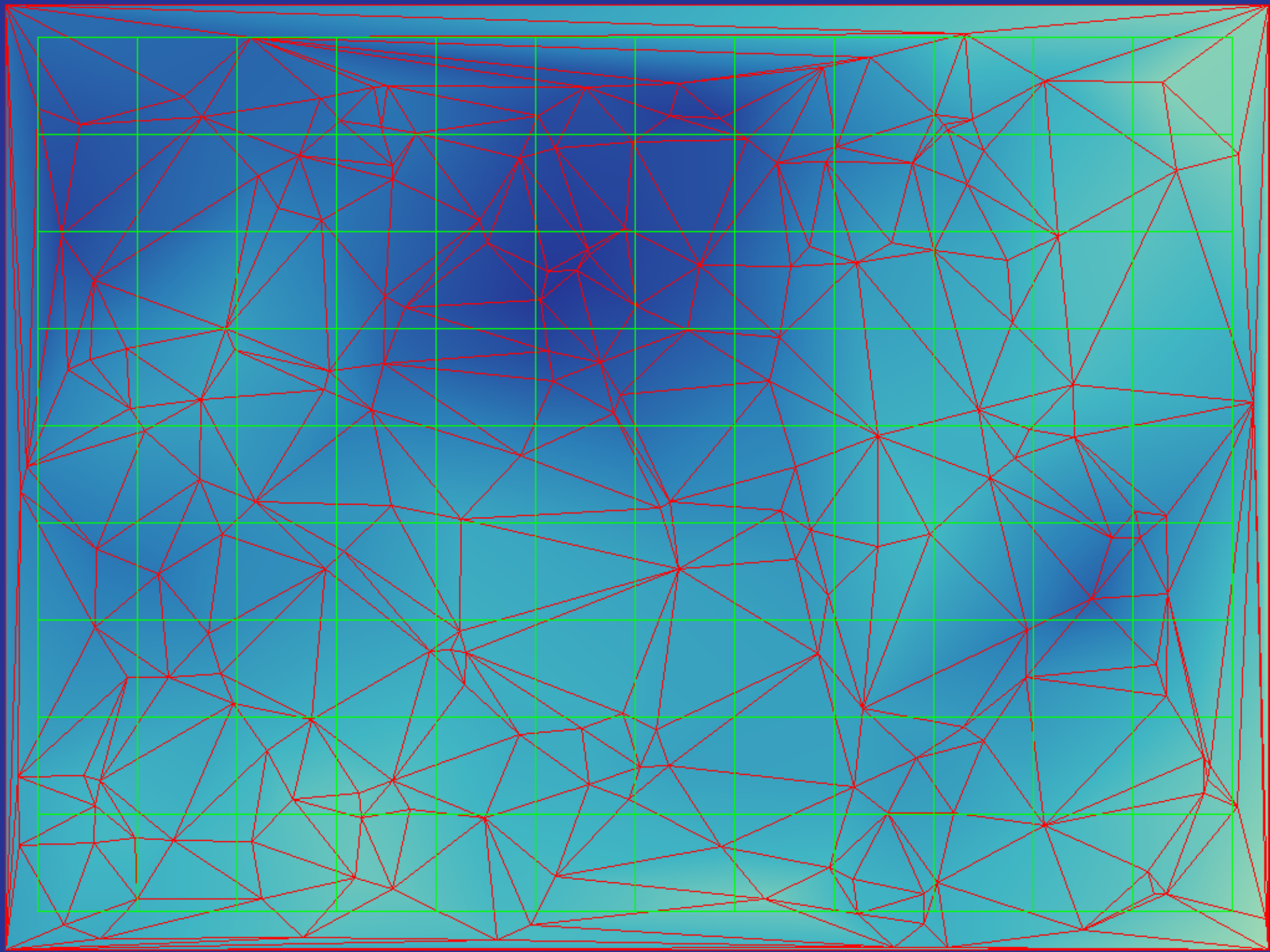


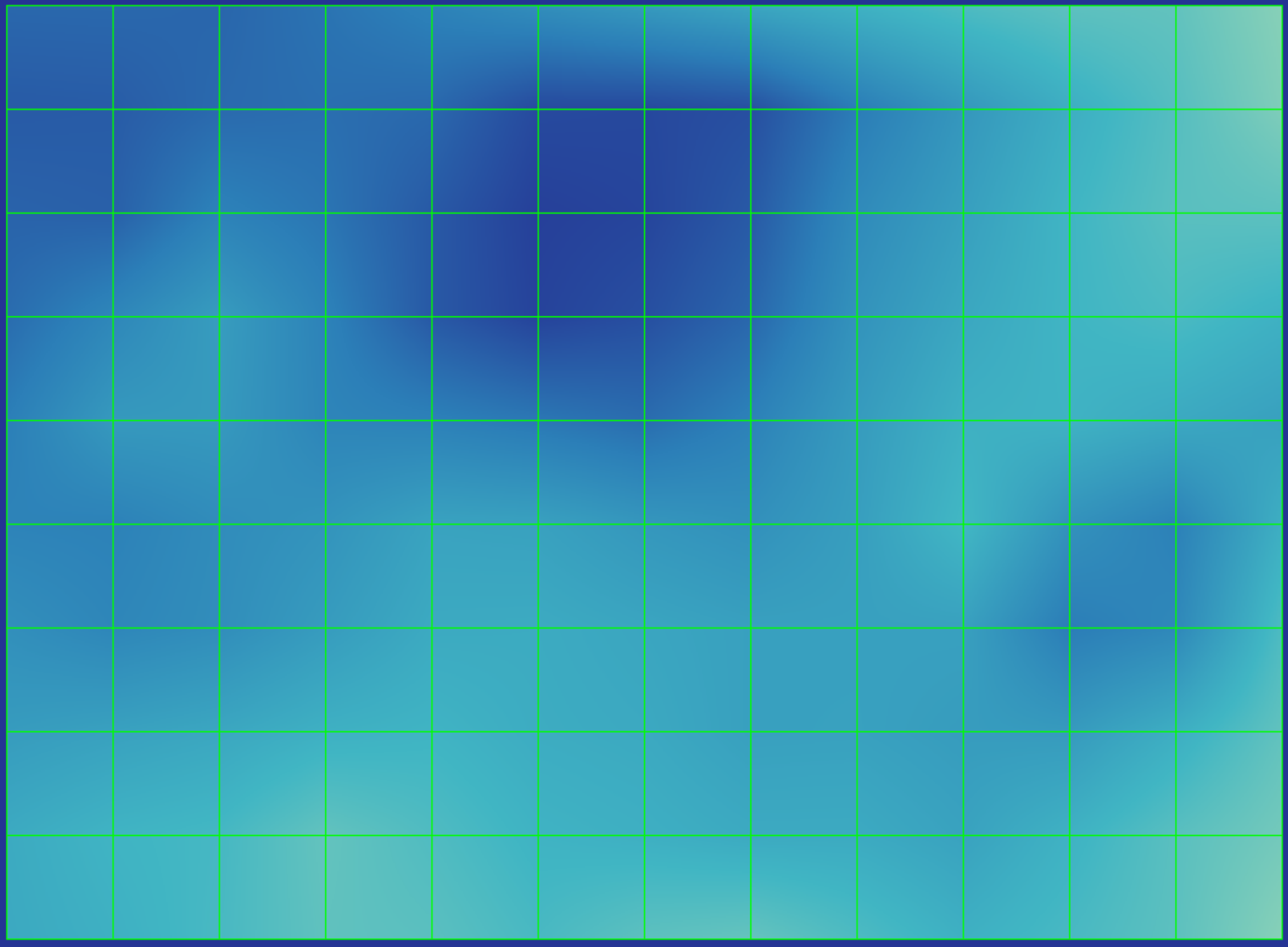












Decomposition Simplification

Simplifying Massive Contour Maps
Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk
ESA 2012

Decomposition Simplification

Simplifying Massive Contour Maps
Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk
ESA 2012

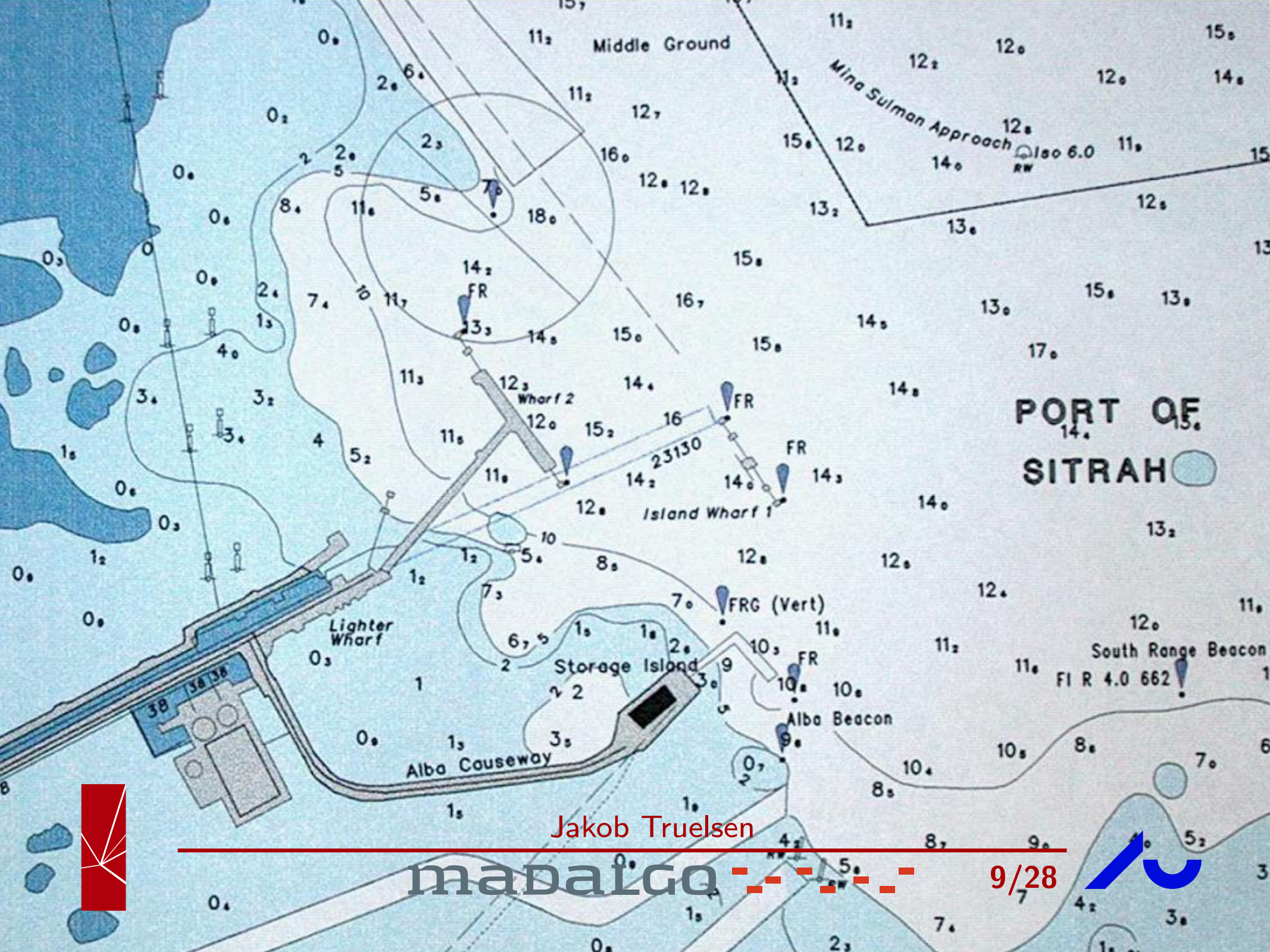
Simplifying massive planar decompositions
Lars Arge and Jungwoo Yang
Alenex 14

Decomposition Simplification

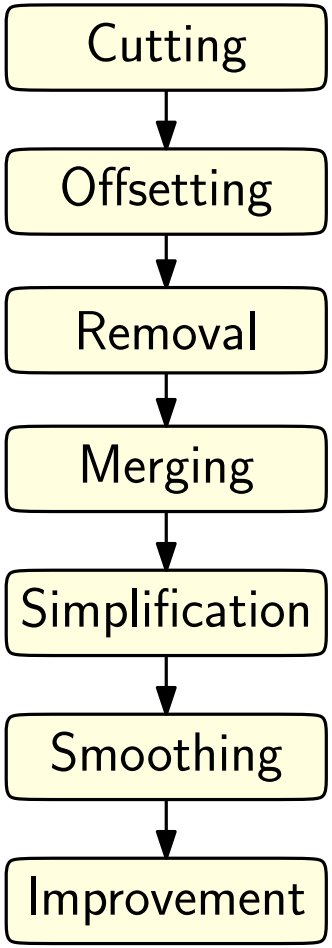
Simplifying Massive Contour Maps
Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk
ESA 2012

Simplifying massive planar decompositions
Lars Arge and Jungwoo Yang
Alenex 14

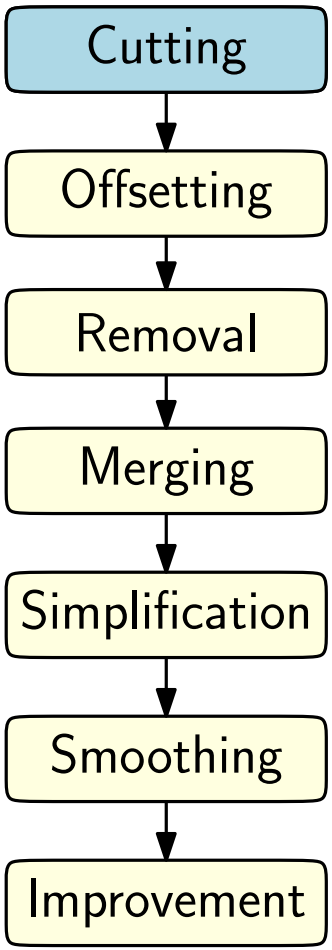
Ongoing work with SCALGO and GST

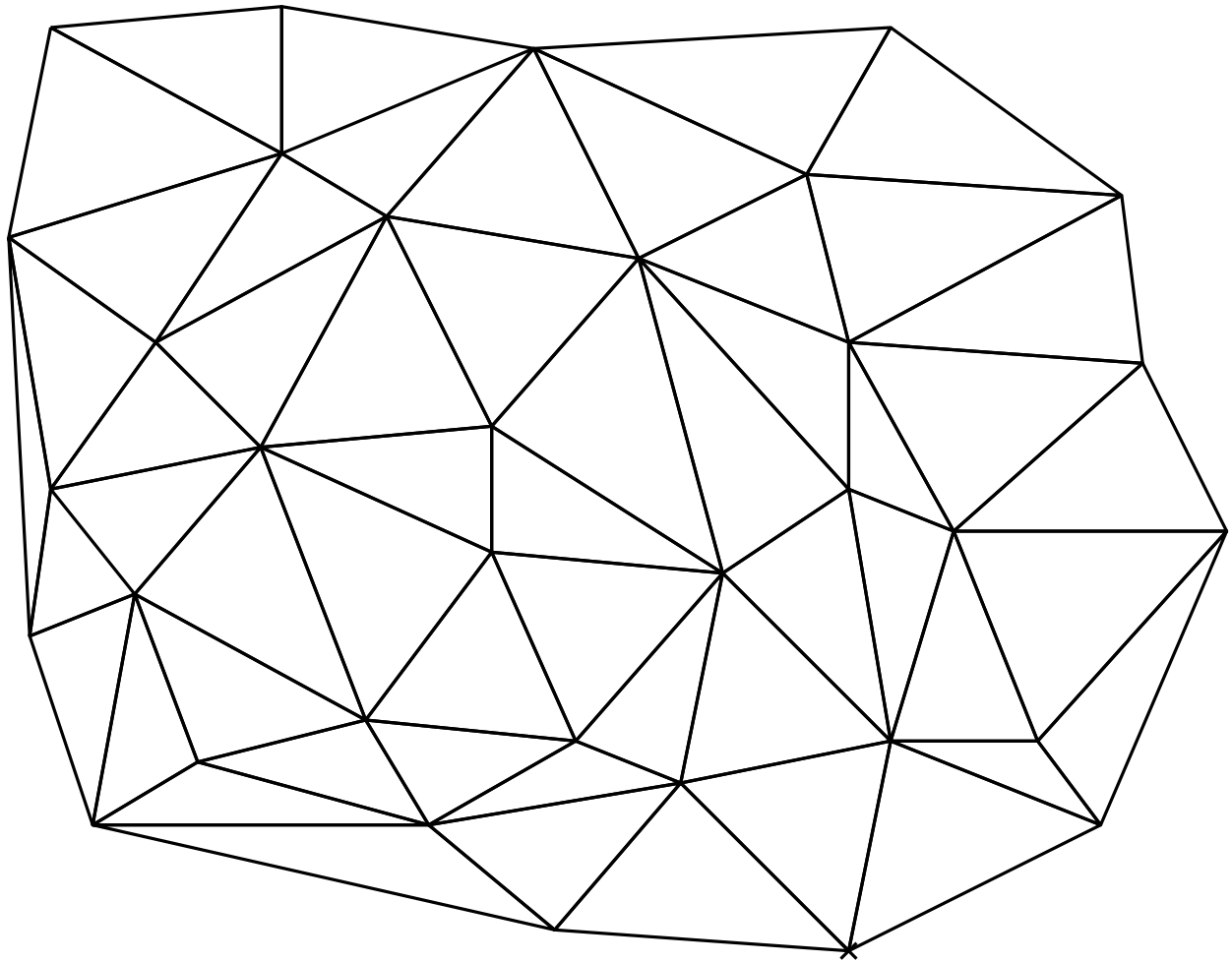

Middle Ground
Mina Sulman Approach
Iso 6.0
PORT OF
SITRAH
Wharf 2
Island Wharf 1
FR
FRG (Vert)
23130
Lighter
Wharf
Storage Island
Alba Causeway
Alba Beacon
South Range Beacon
Fl R 4.0 662
38

Decomposition Simplification

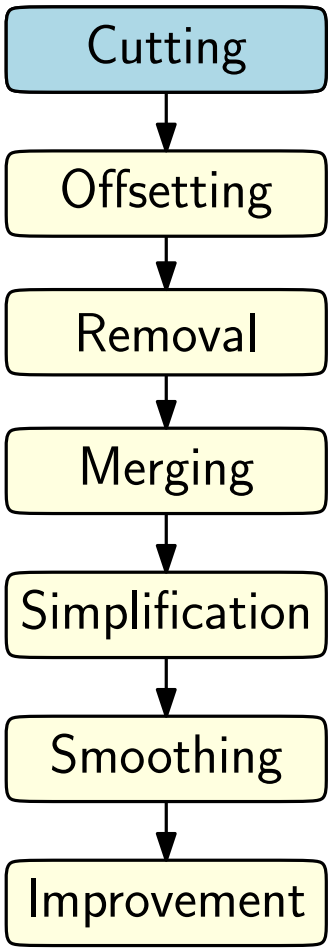


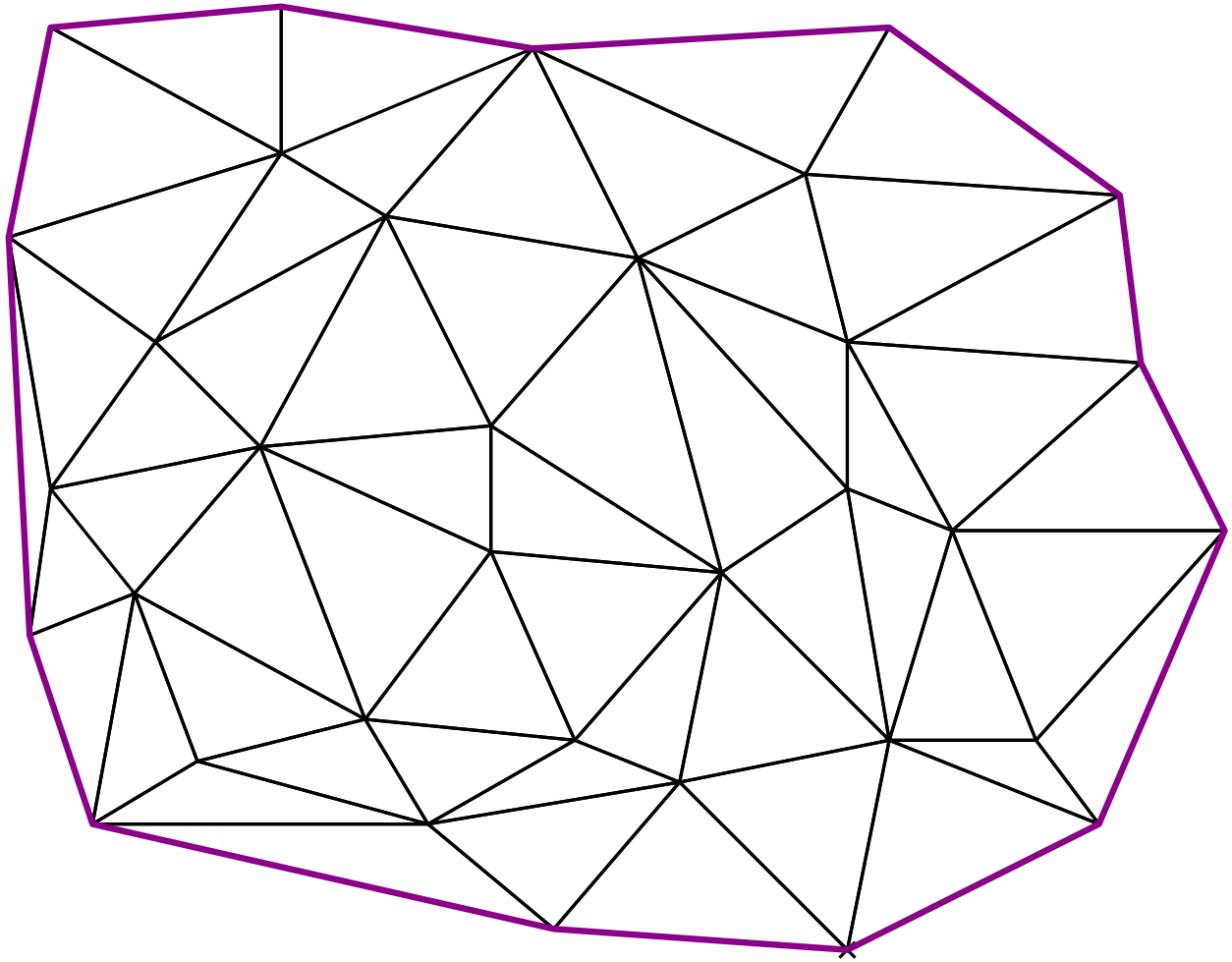
Decomposition Simplification



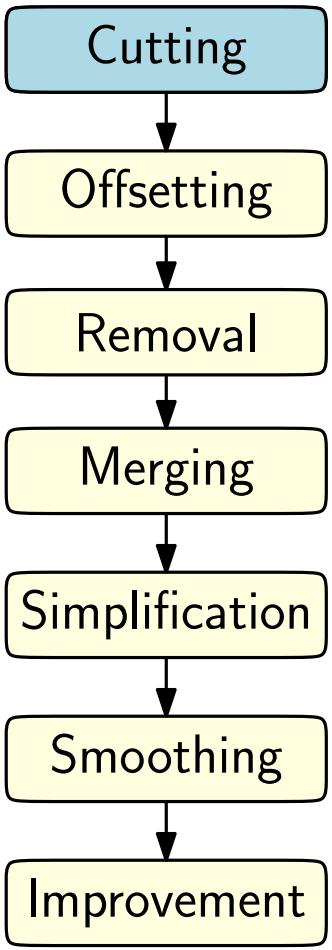


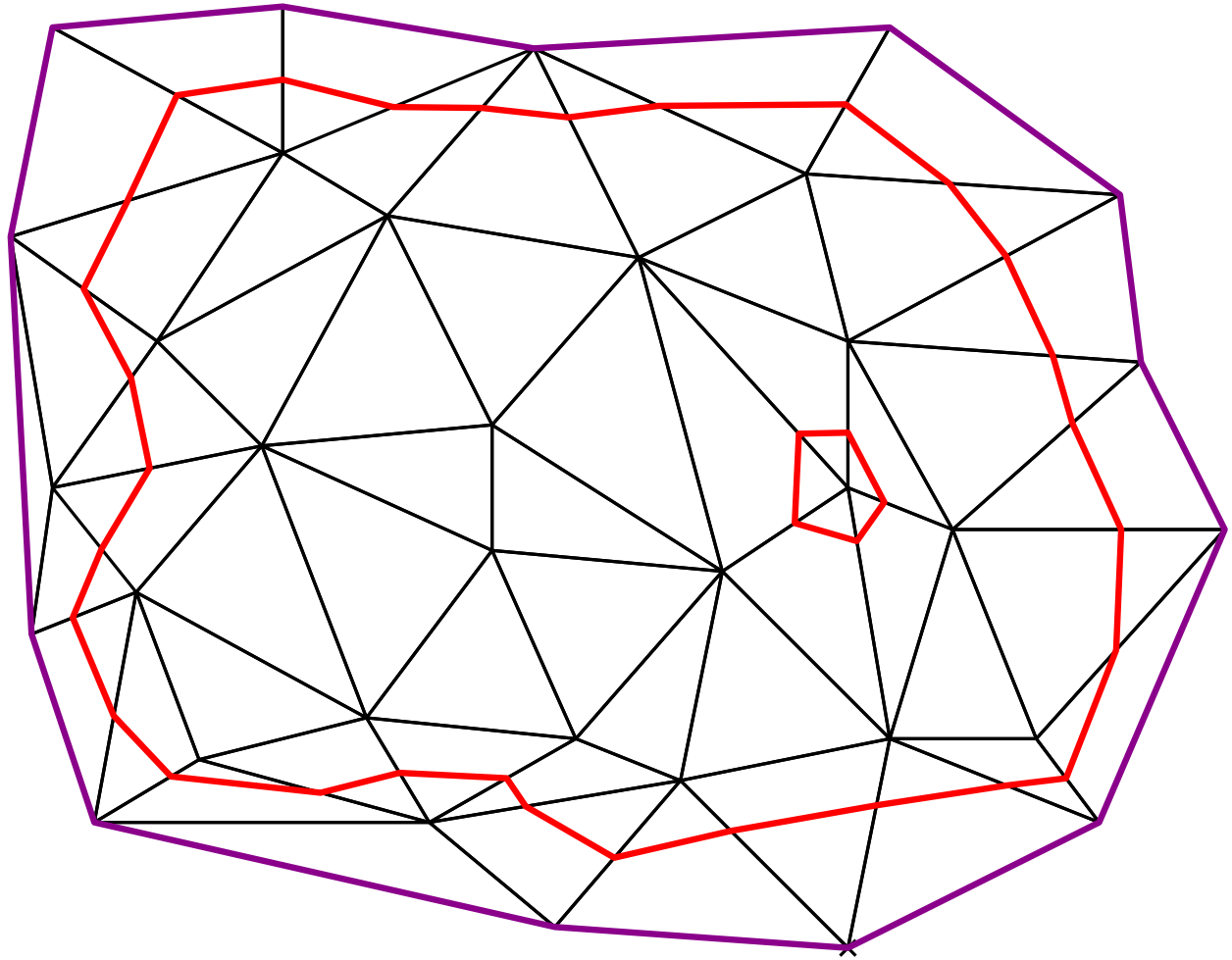
Decomposition Simplification



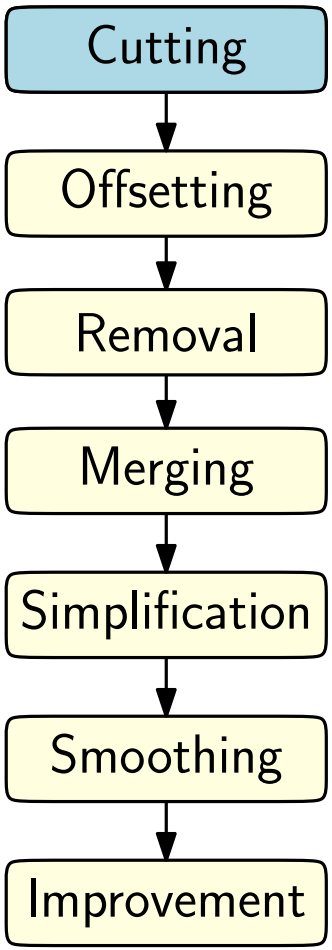


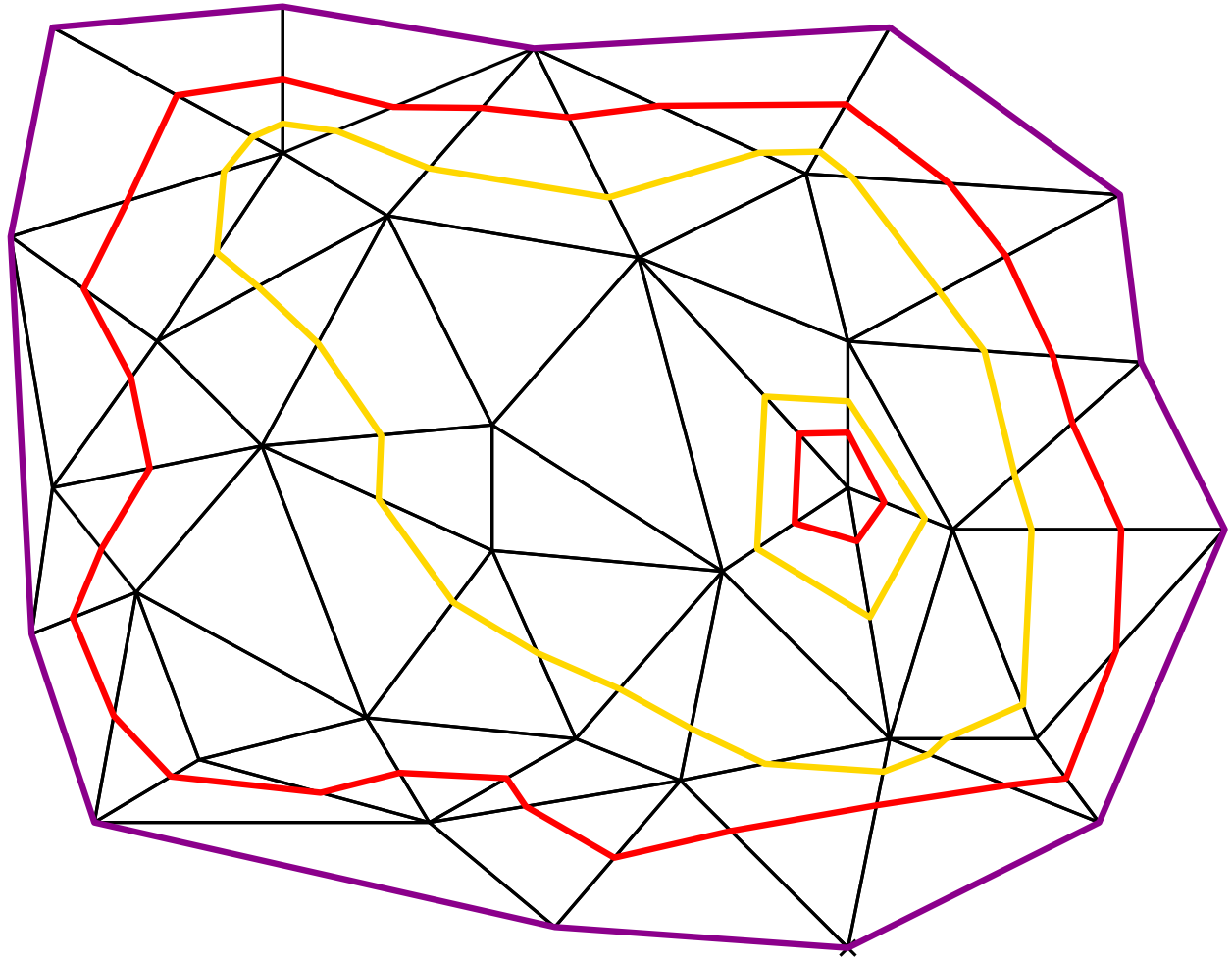
Decomposition Simplification



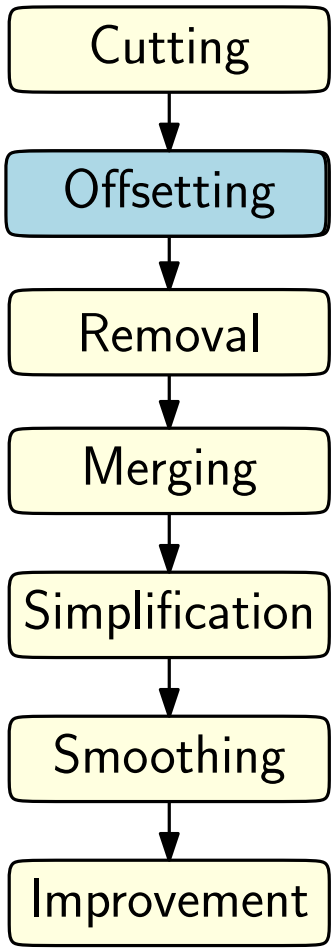


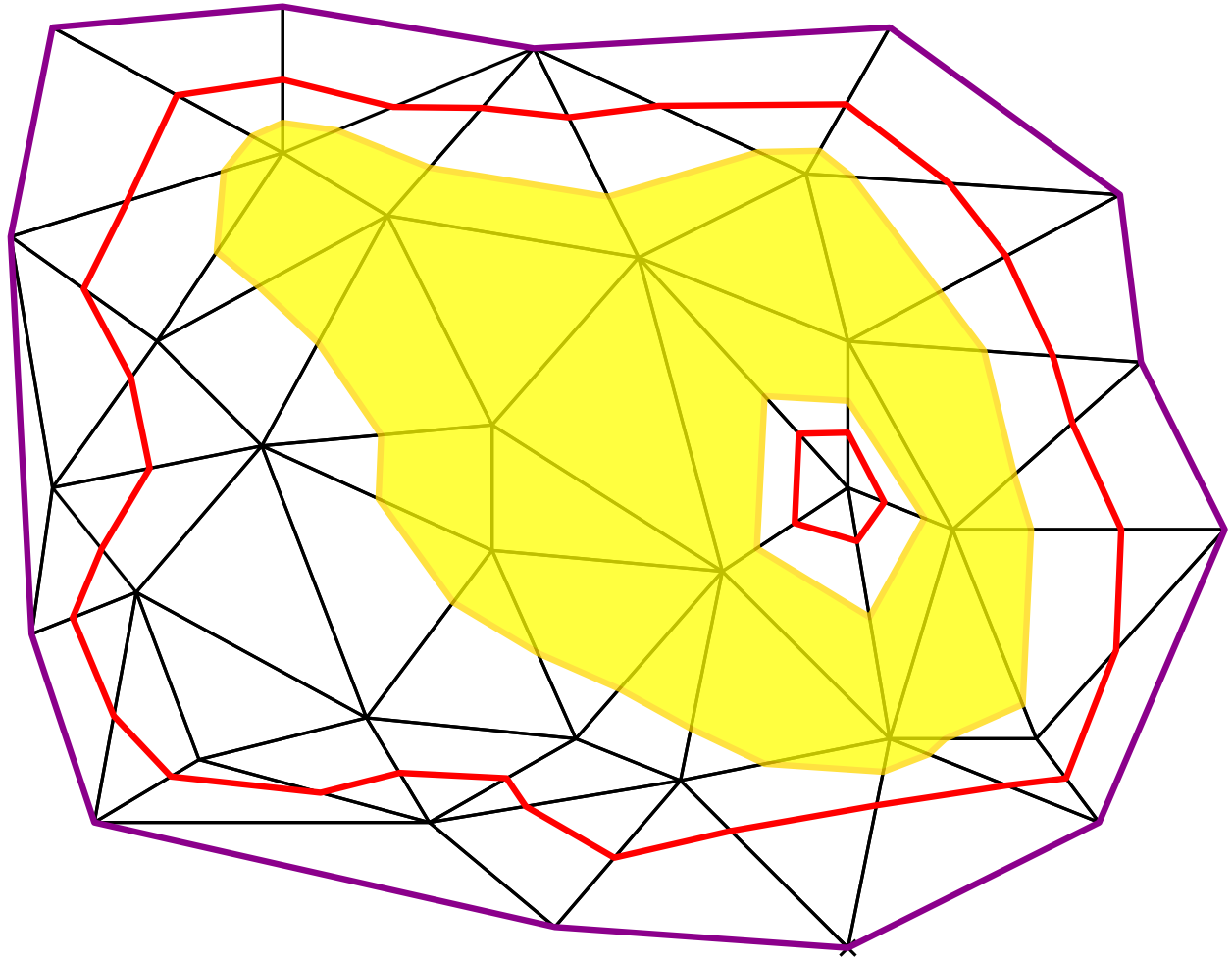
Decomposition Simplification



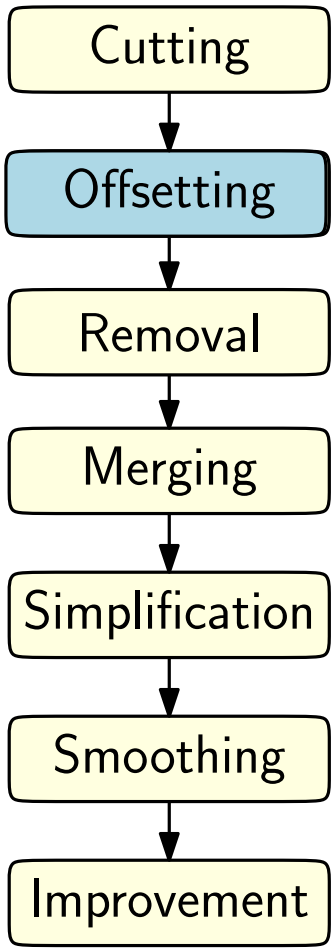


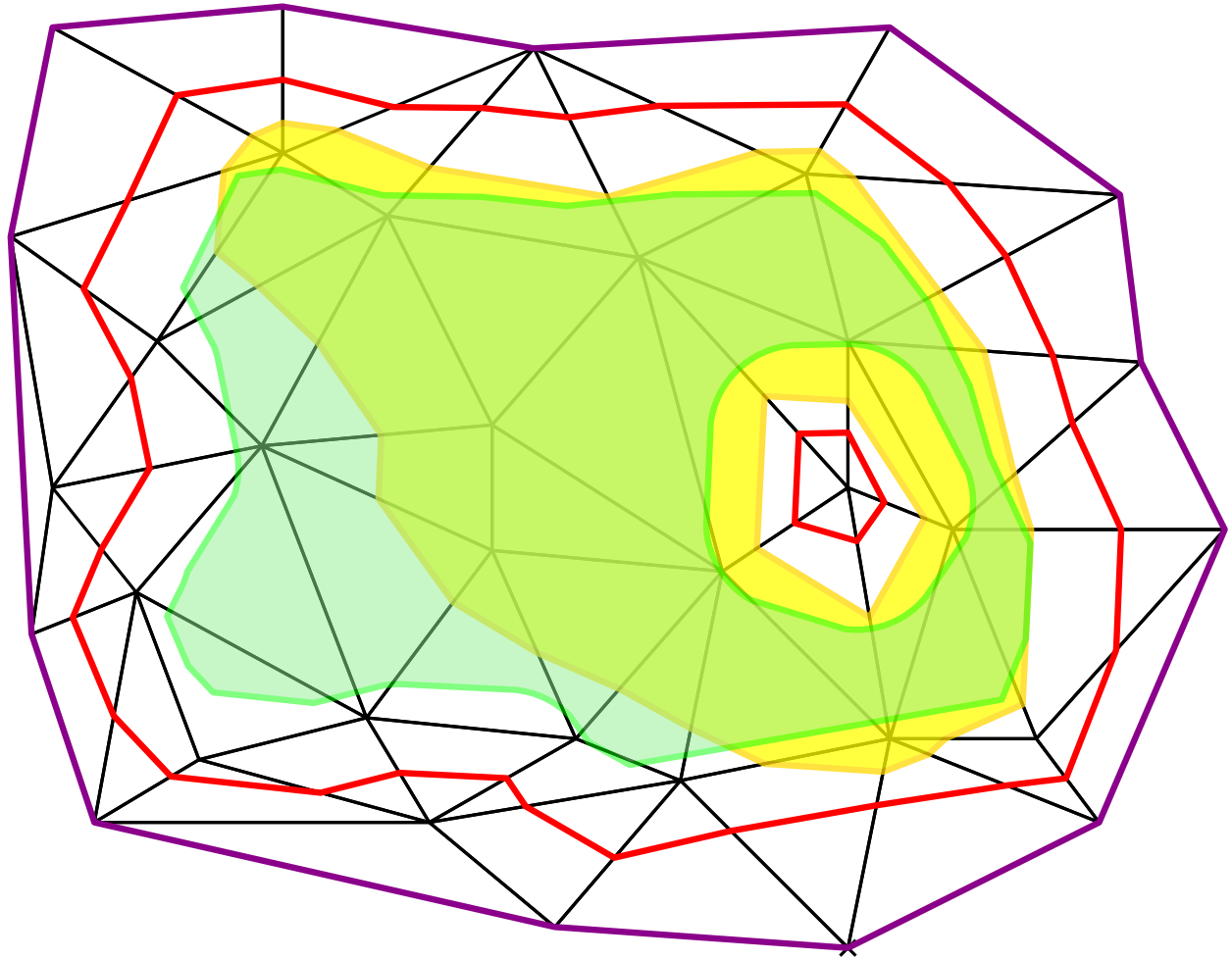
Decomposition Simplification



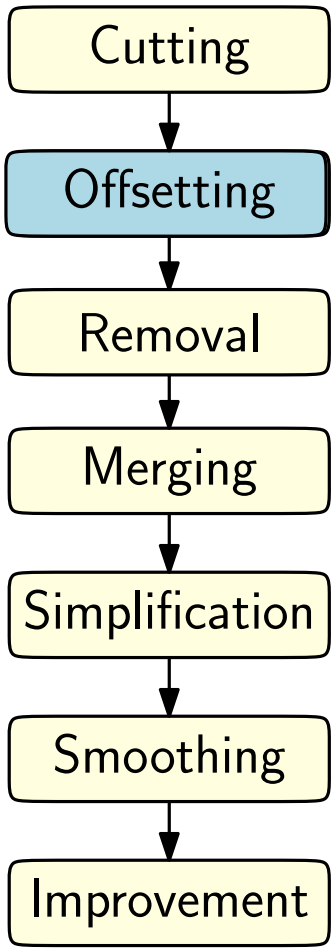


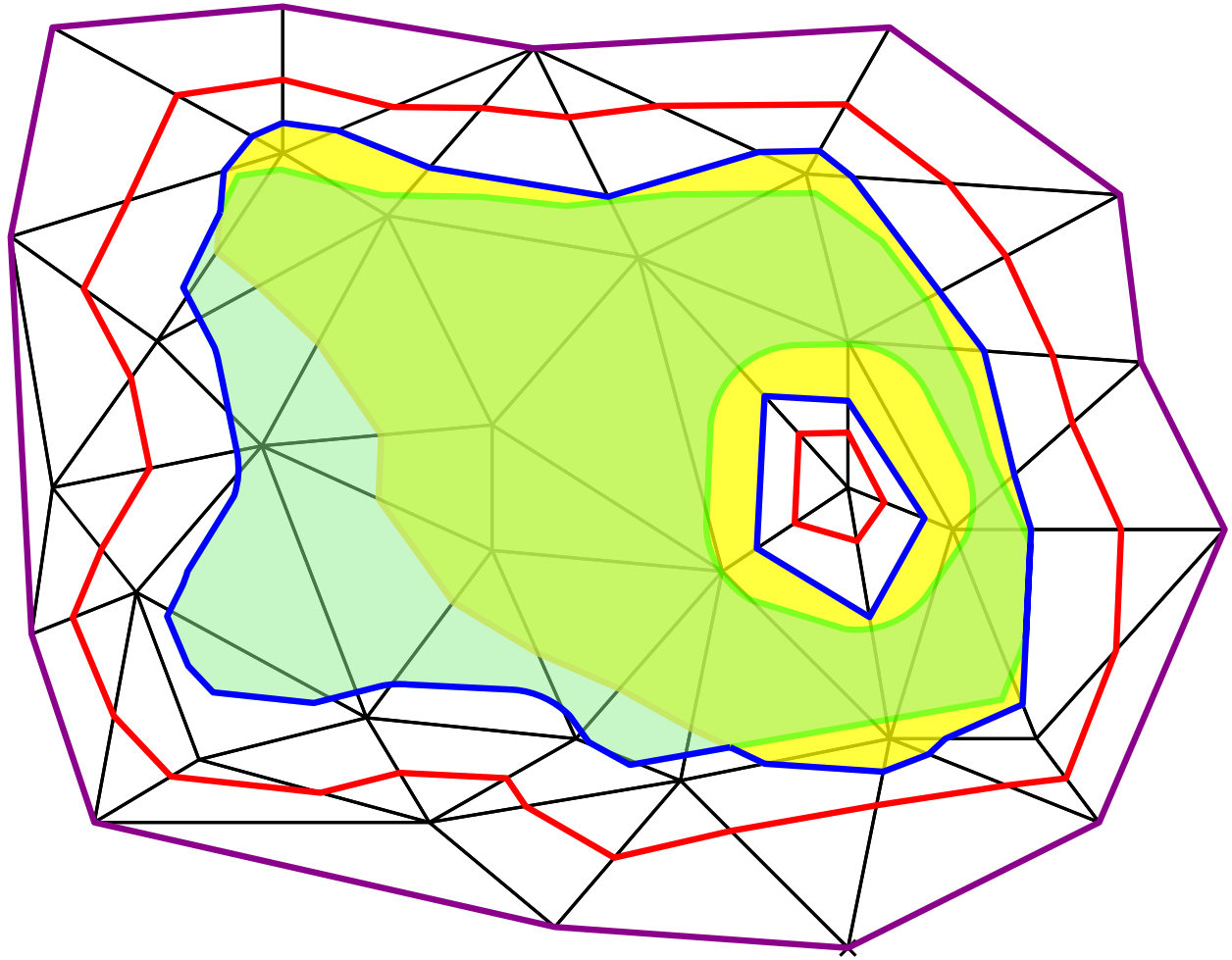
Decomposition Simplification



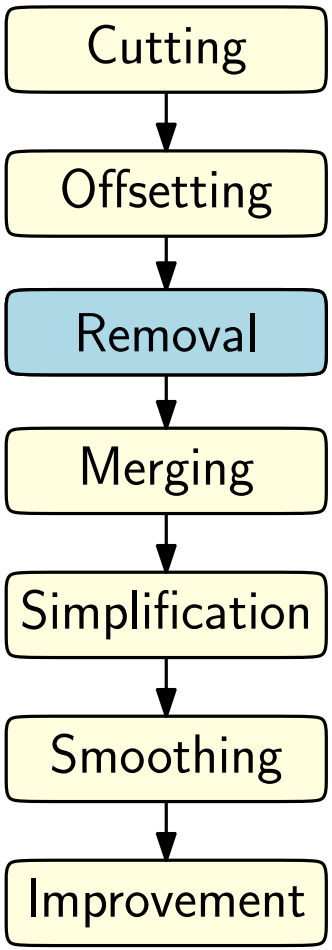


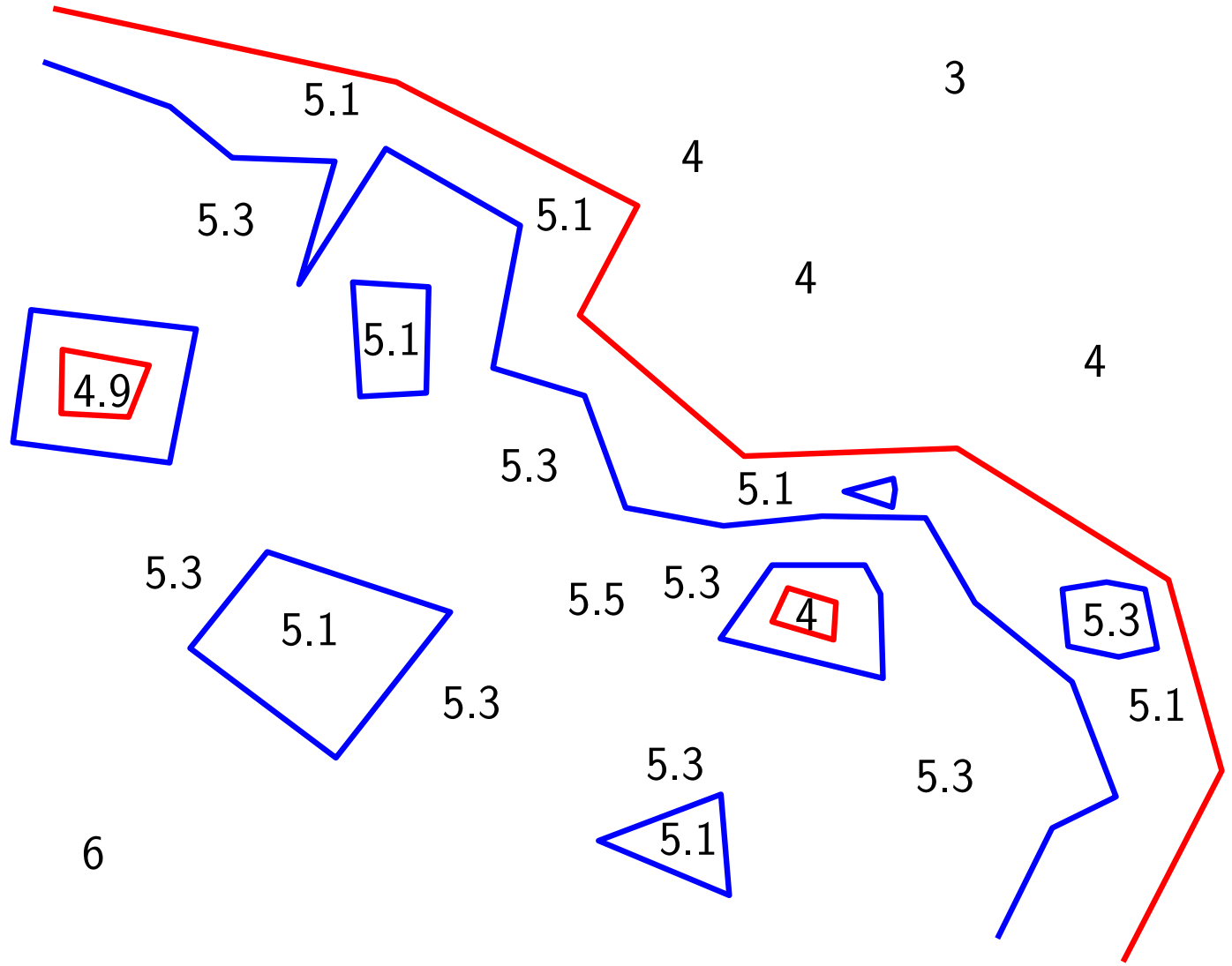
Decomposition Simplification



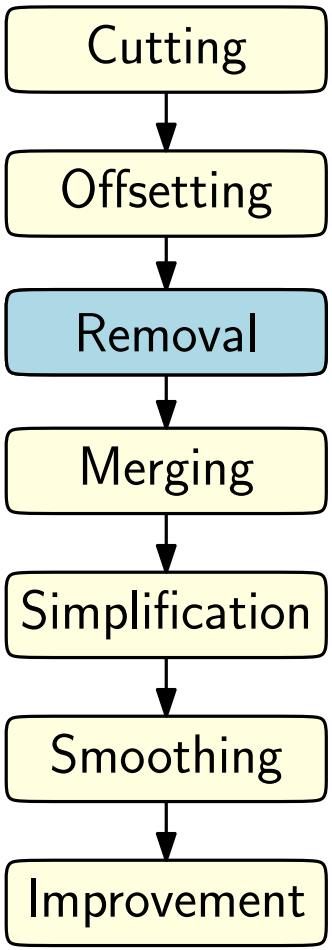


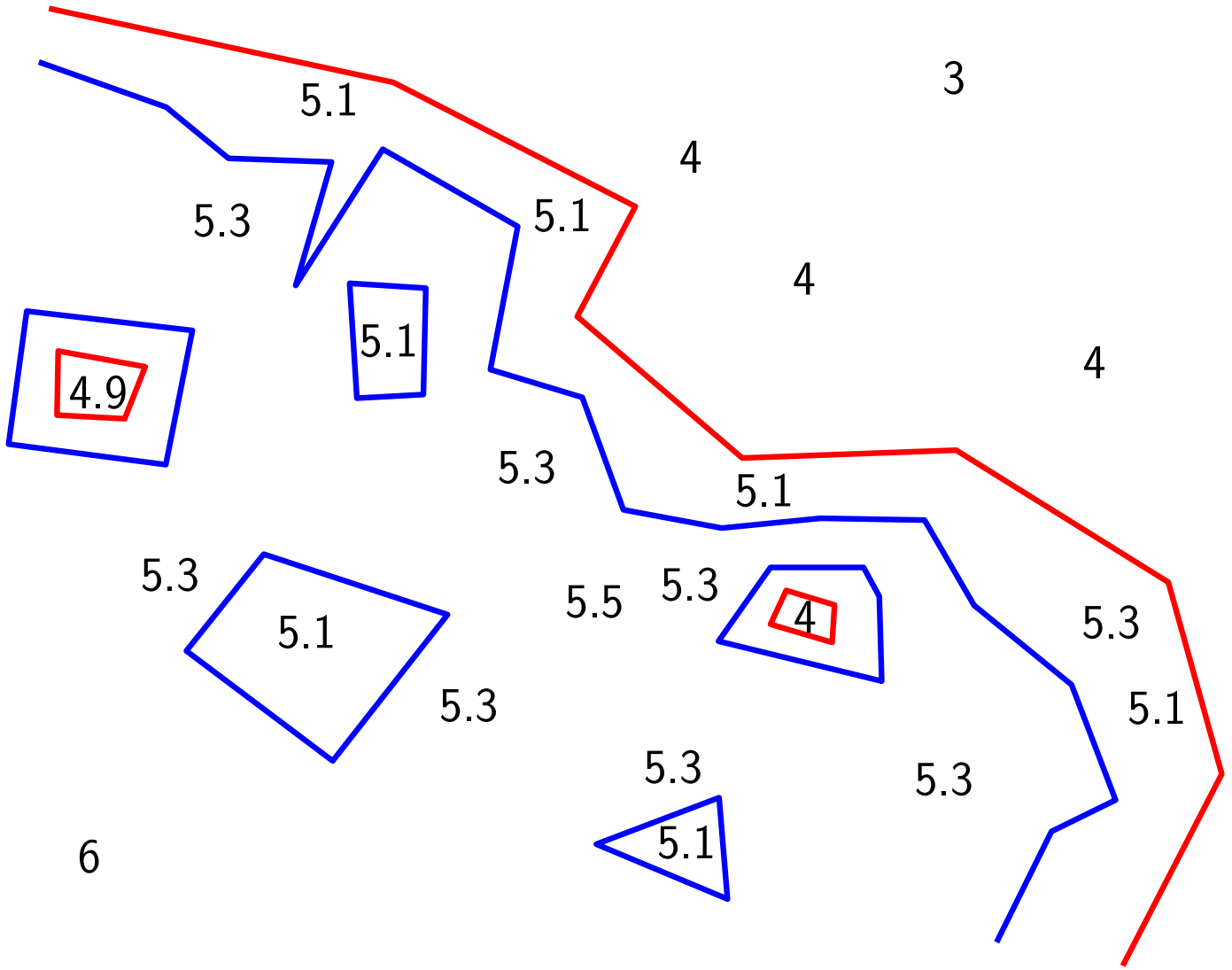
Decomposition Simplification



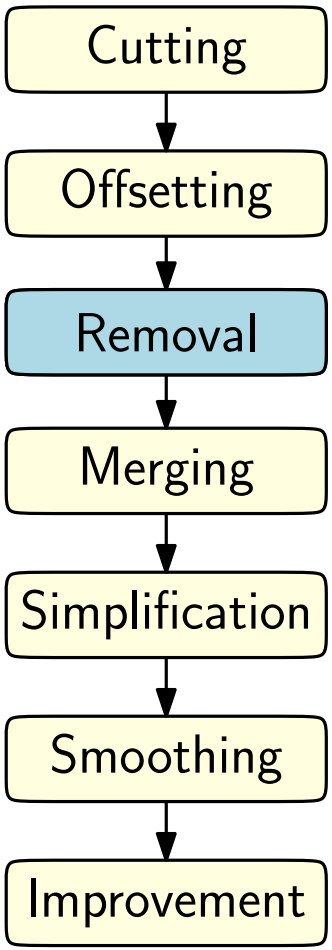


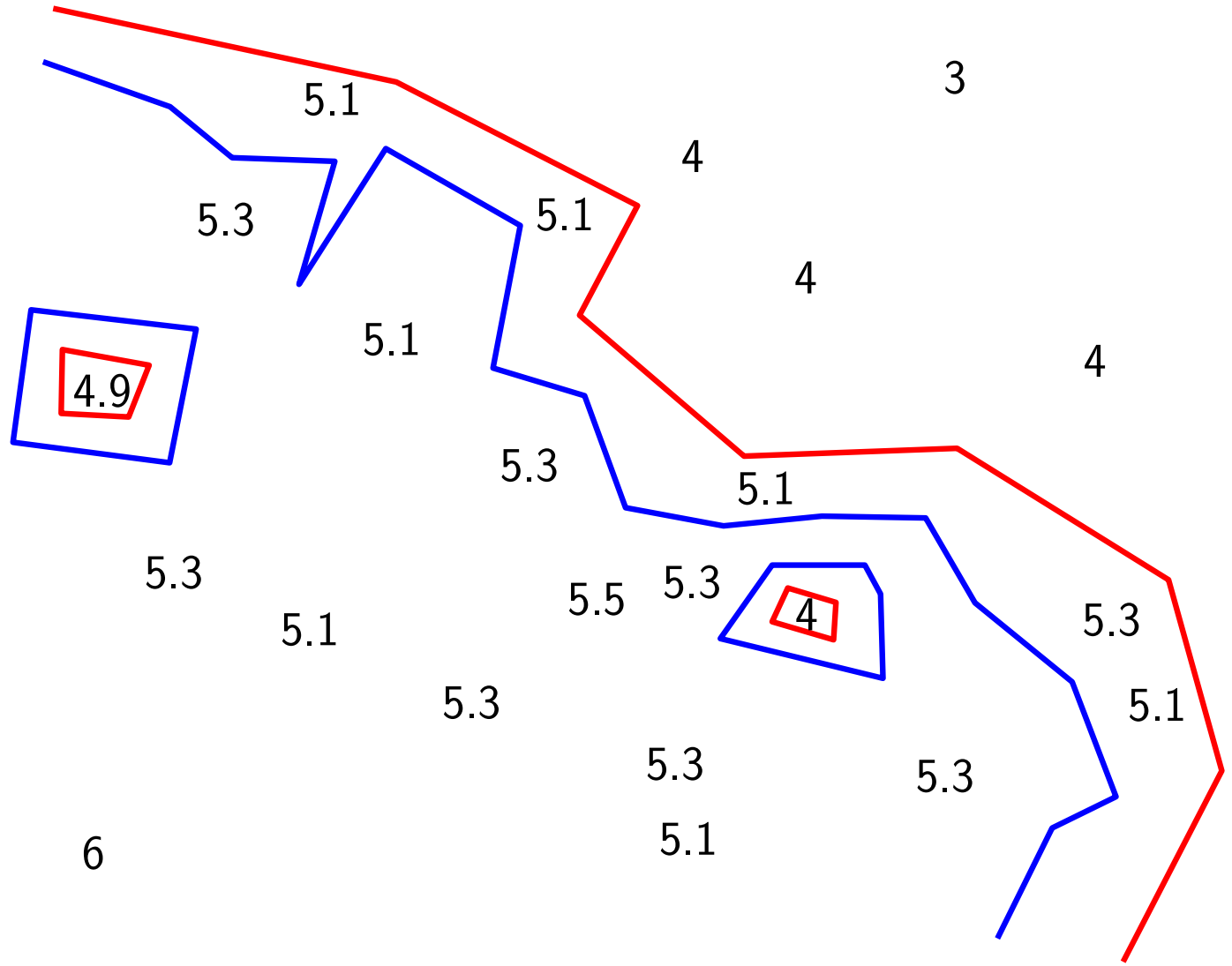
Decomposition Simplification



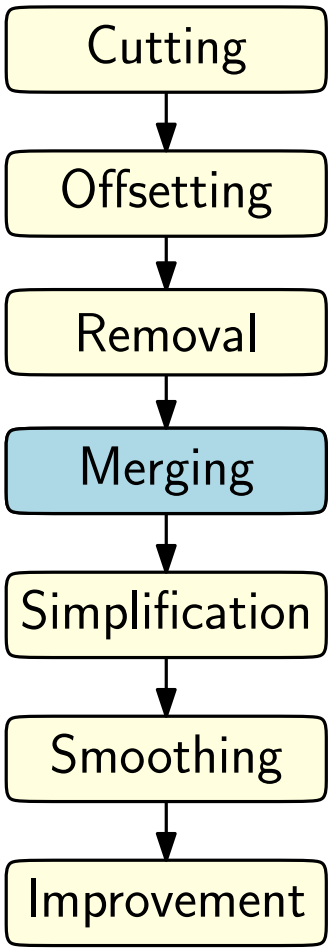


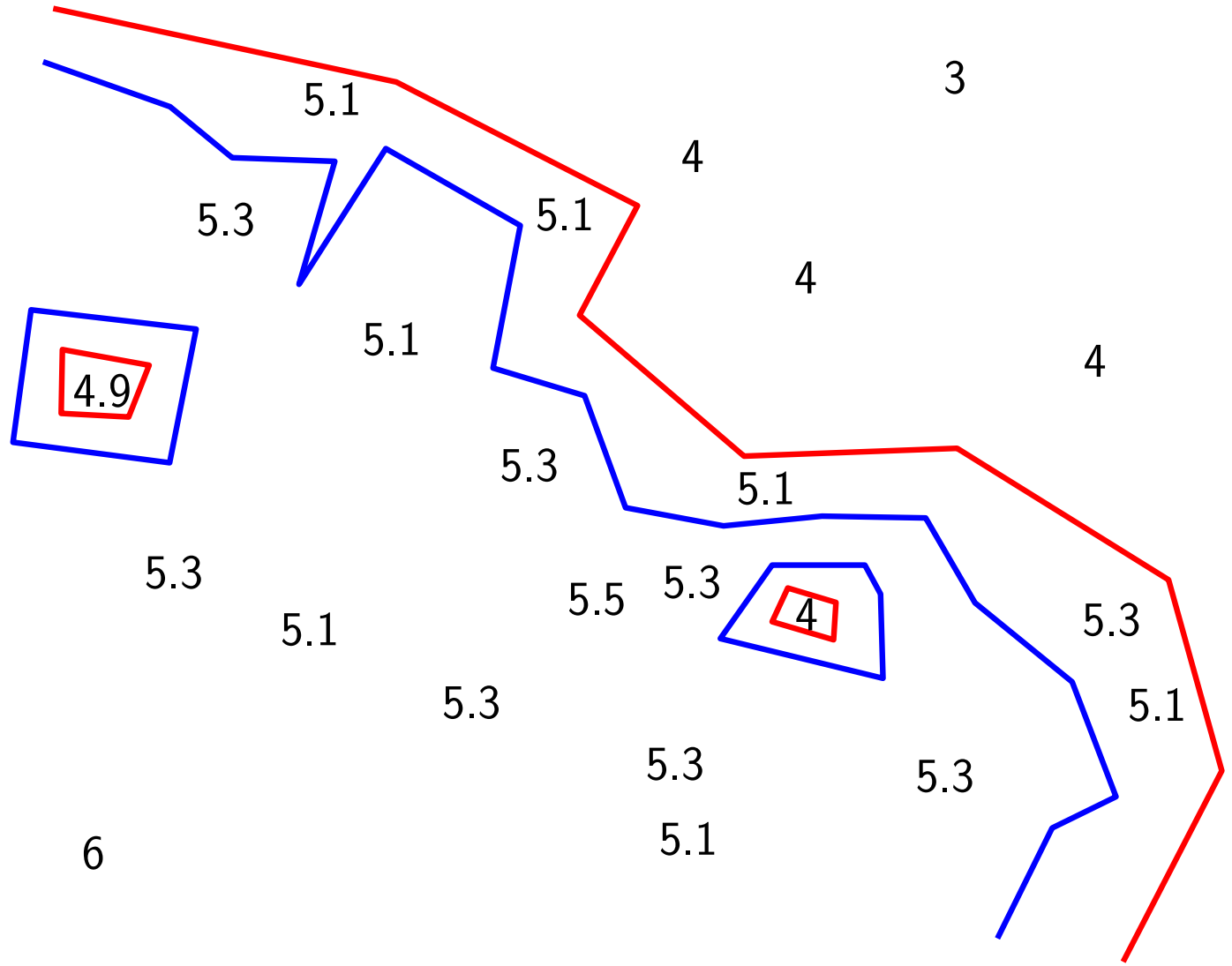
Decomposition Simplification



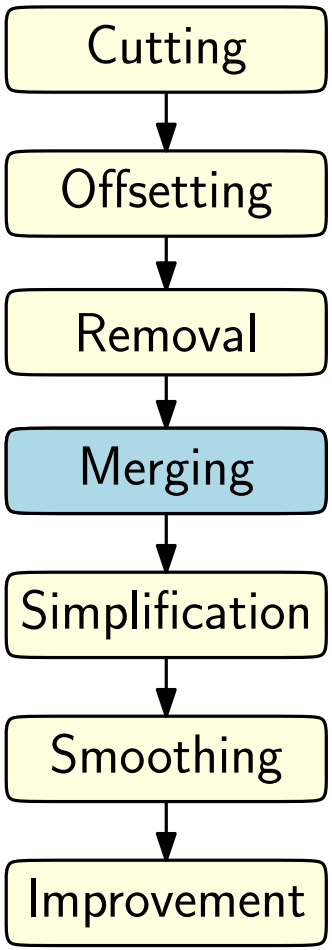


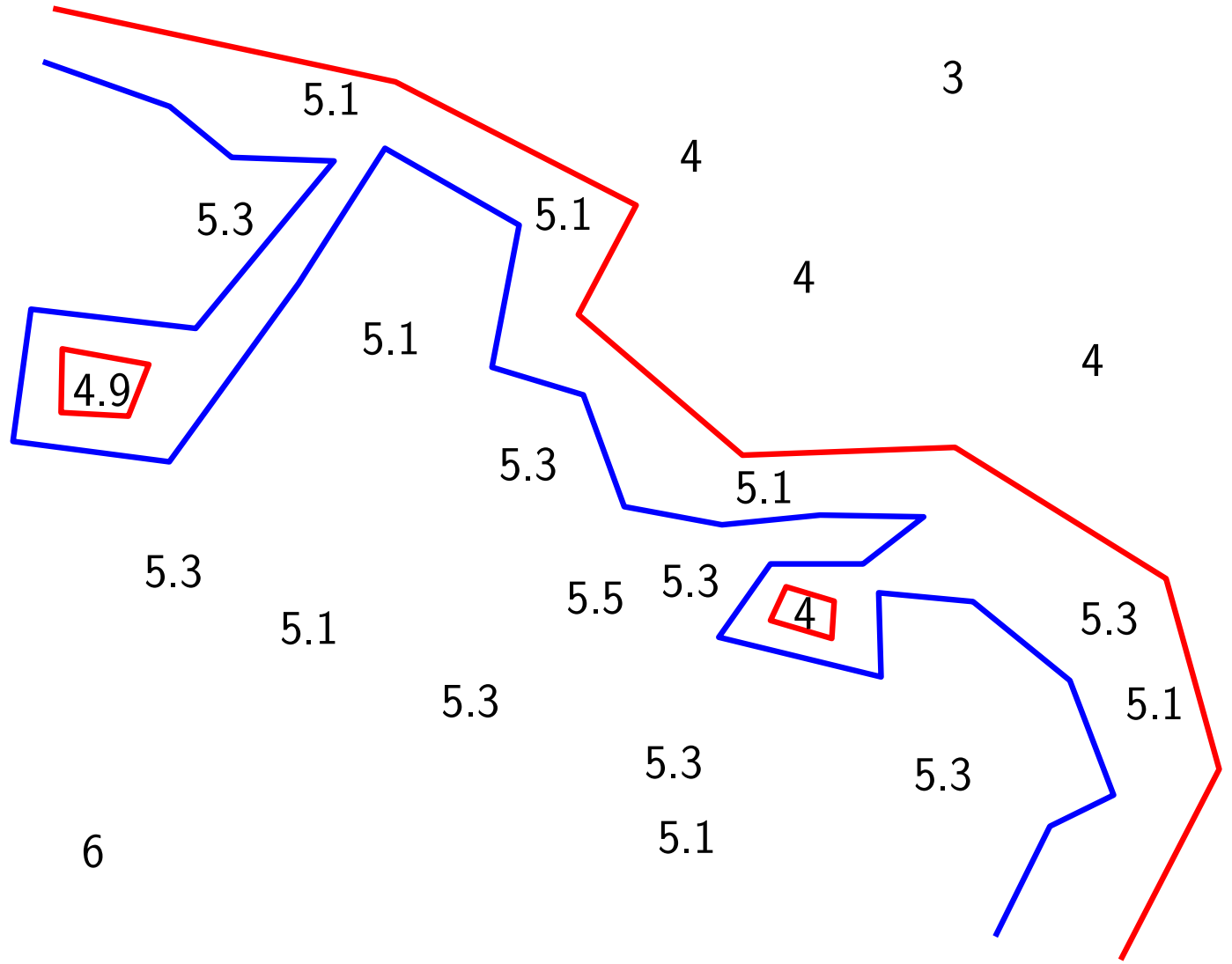
Decomposition Simplification



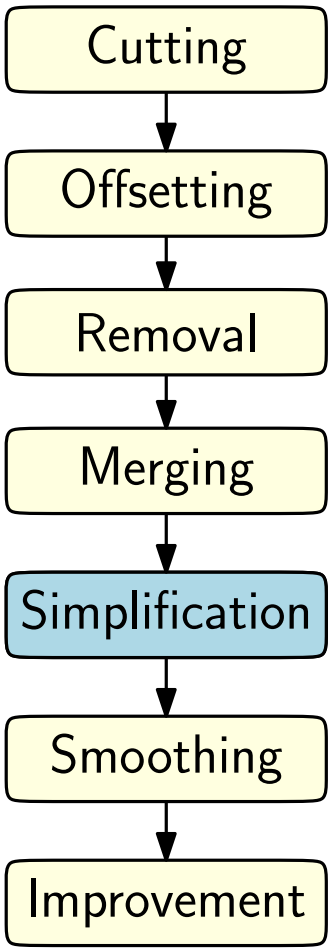


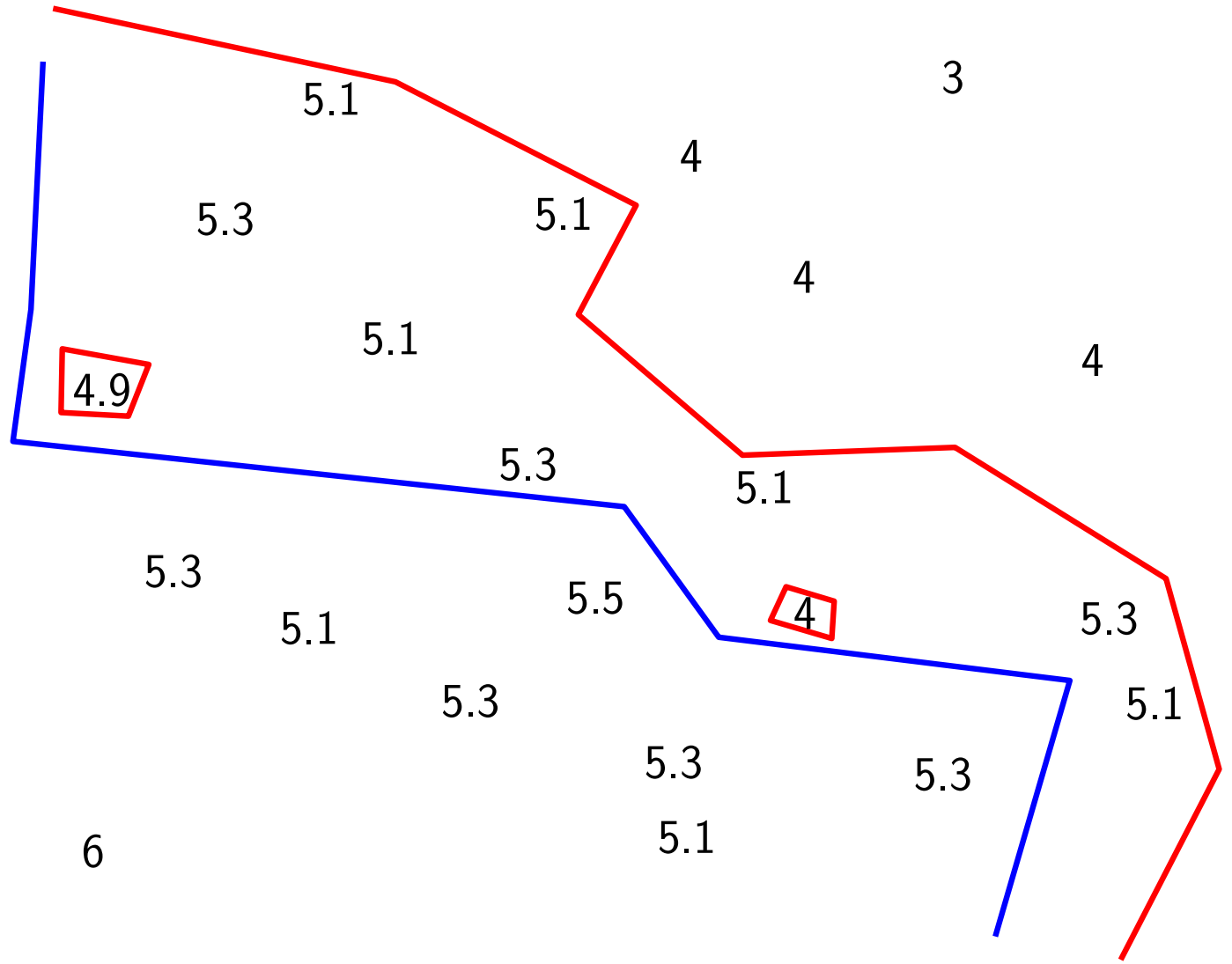
Decomposition Simplification



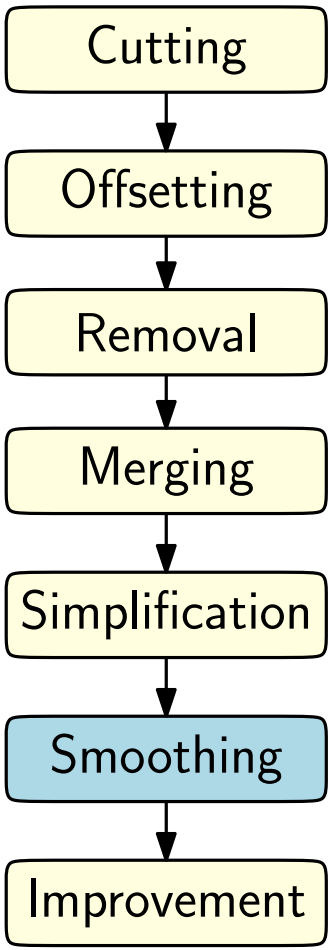


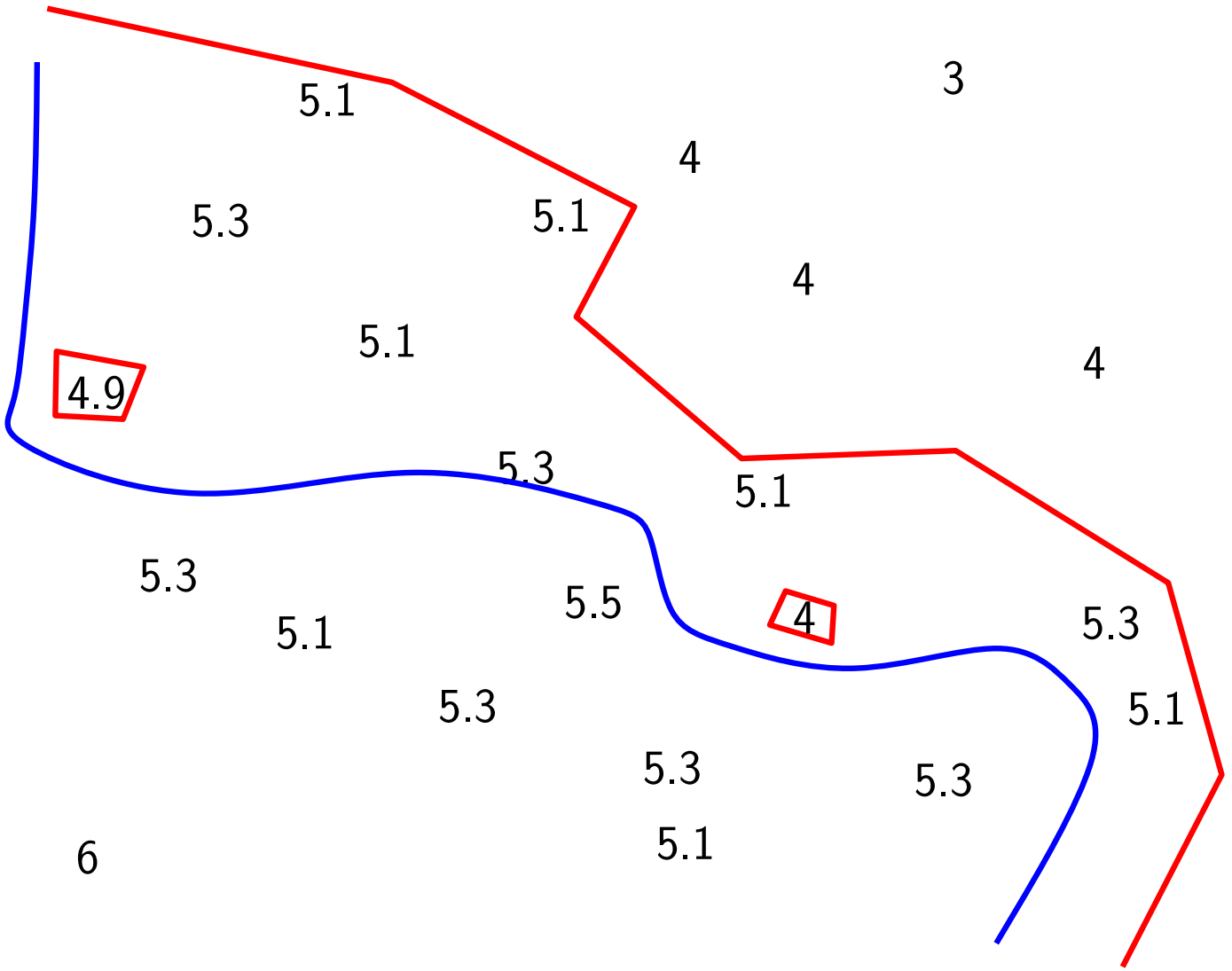
Decomposition Simplification



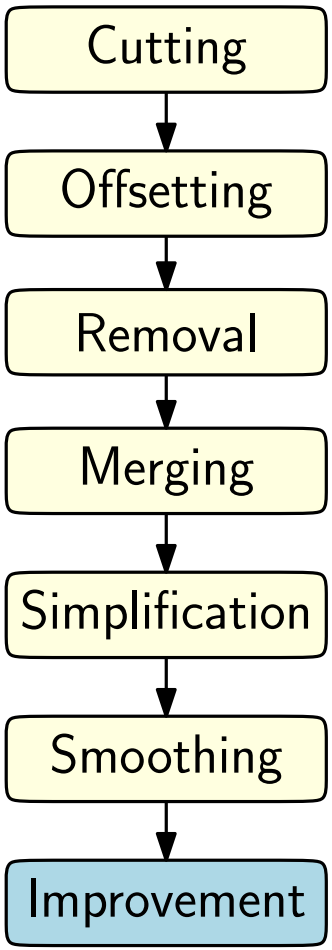


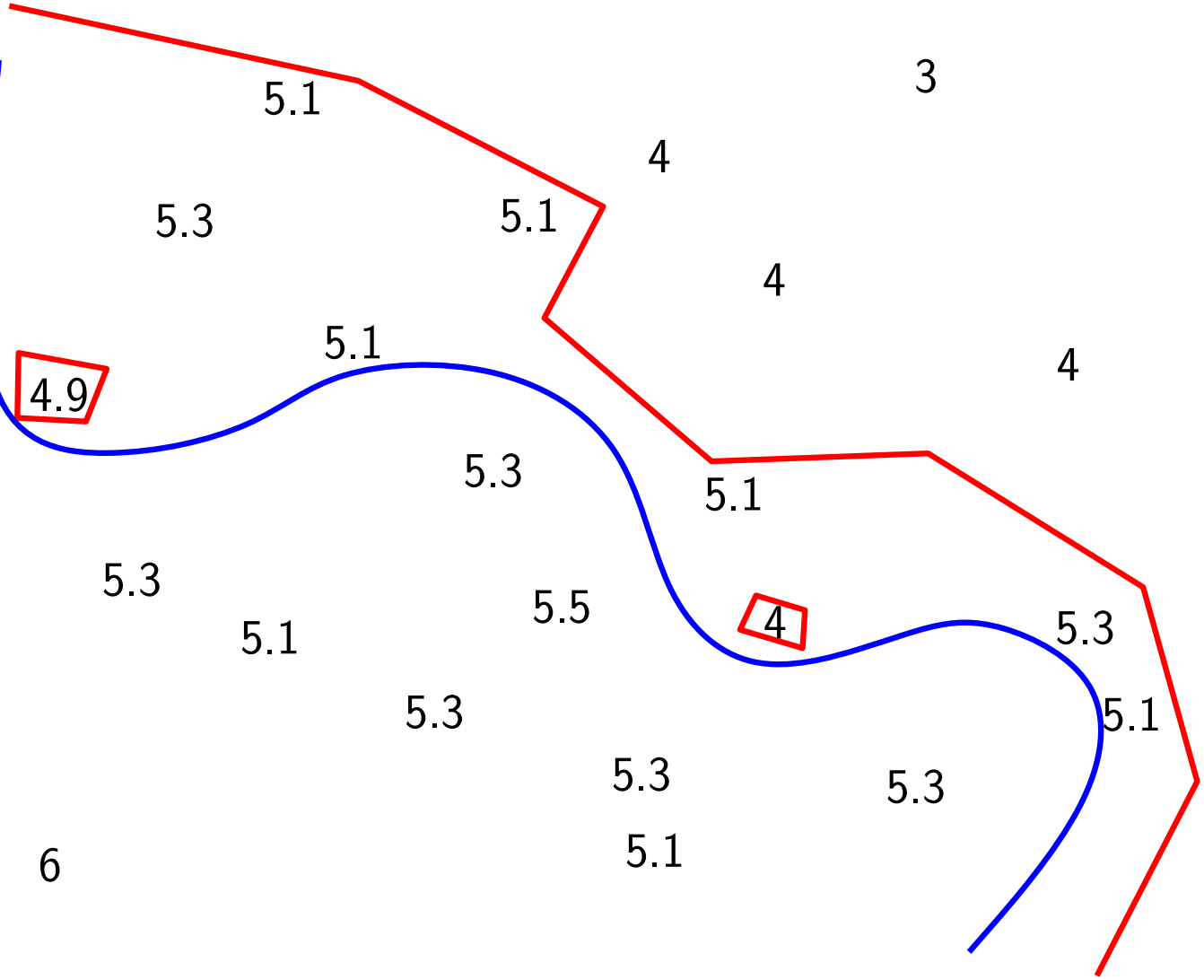
Decomposition Simplification

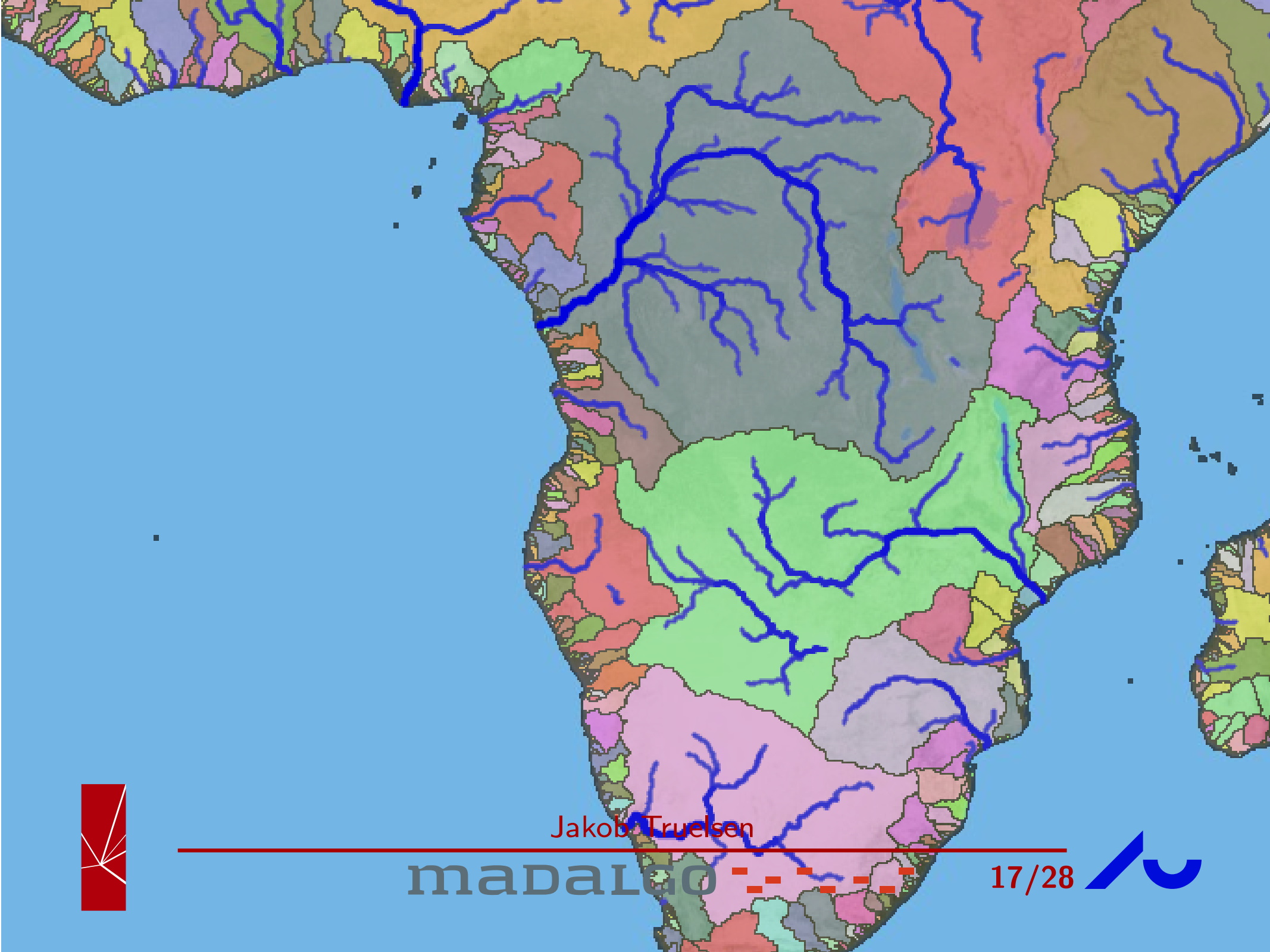


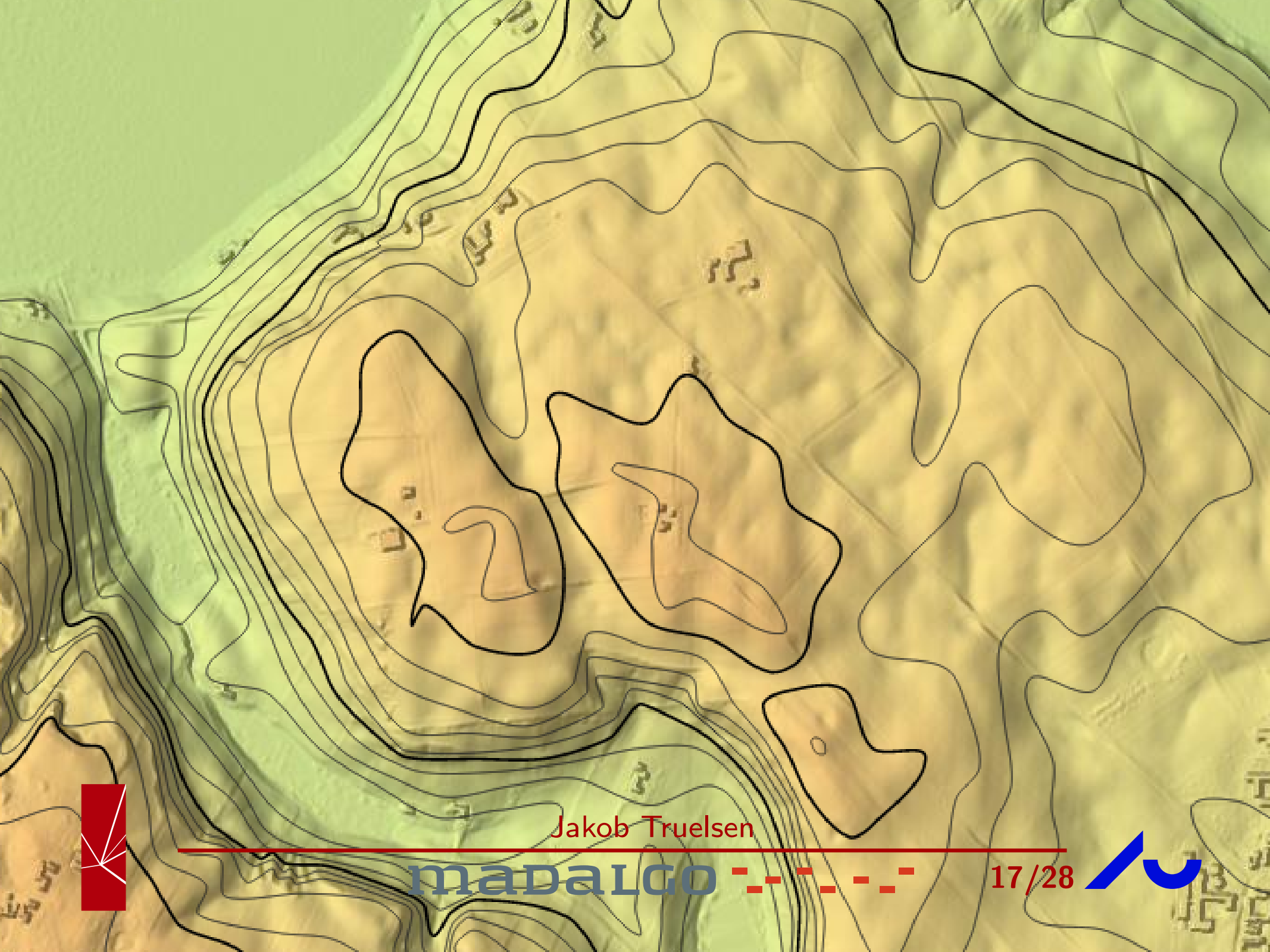


Decomposition Simplification

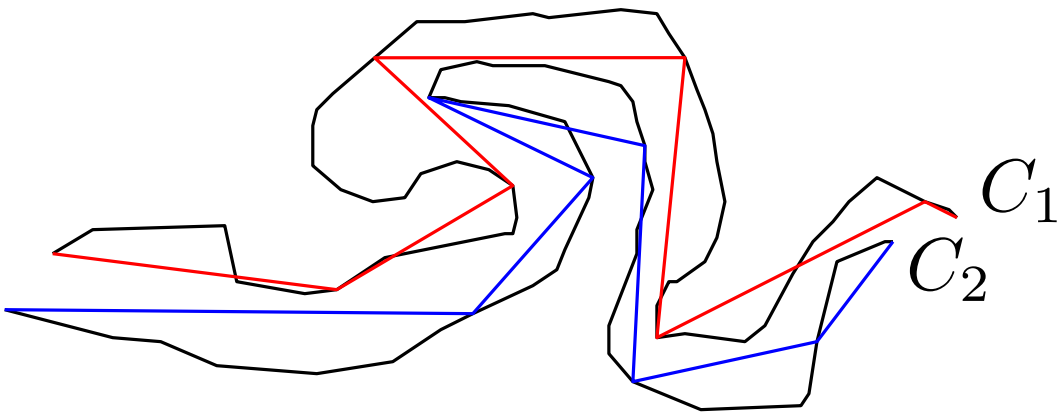




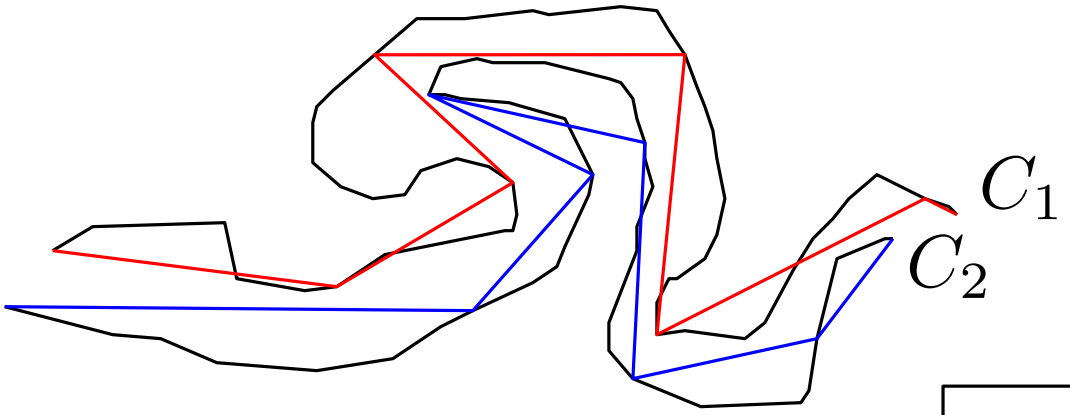


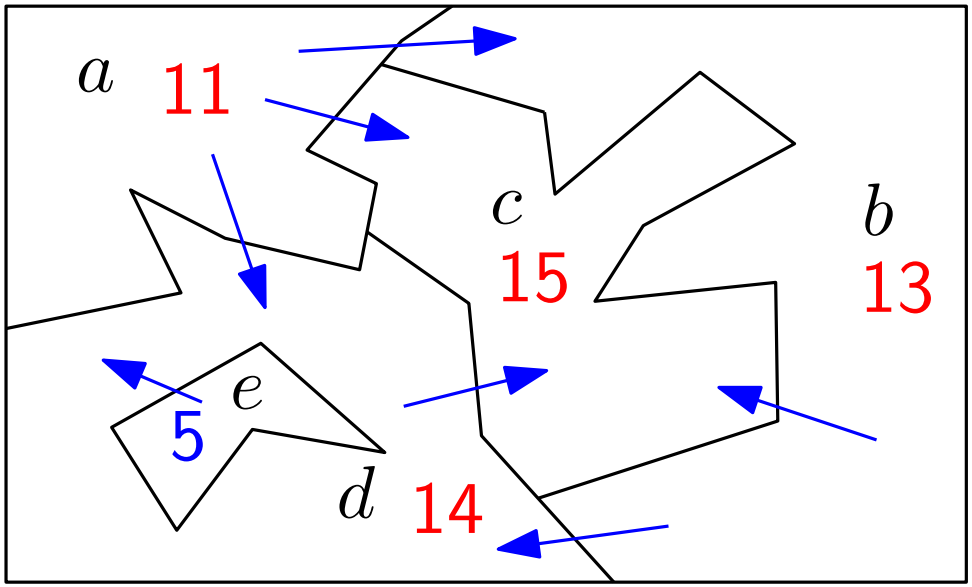


Decomposition Simplification



Decomposition Simplification



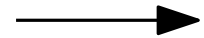


Multiresolution Rasters

1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	8	8
3	4	5	6	7	8	9	8	7
4	5	6	7	8	9	8	7	6
5	6	7	8	8	8	7	6	5
6	7	8	9	8	7	6	5	4
7	8	9	8	7	6	5	4	3
8	8	8	7	6	5	4	3	2
9	8	7	6	5	4	3	2	1

Multiresolution Rasters

1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	8	8
3	4	5	6	7	8	9	8	7
4	5	6	7	8	9	8	7	6
5	6	7	8	8	8	7	6	5
6	7	8	9	8	7	6	5	4
7	8	9	8	7	6	5	4	3
8	8	8	7	6	5	4	3	2
9	8	7	6	5	4	3	2	1

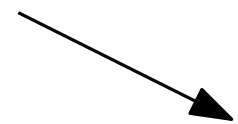


3	6	8
6	8	6
8	6	3

Multiresolution Rasters

1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	8	8
3	4	5	6	7	8	9	8	7
4	5	6	7	8	9	8	7	6
5	6	7	8	8	8	7	6	5
6	7	8	9	8	7	6	5	4
7	8	9	8	7	6	5	4	3
8	8	8	7	6	5	4	3	2
9	8	7	6	5	4	3	2	1

3	6	8
6	8	6
8	6	3



2	4	6	$7\frac{3}{4}$
4	6	8	8
6	8	$7\frac{3}{4}$	6
$7\frac{3}{4}$	8	6	4

Multiresolution Rasters

"An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters"
Lars Arge, Gerth Stølting Brodal and Constantinos Tsirogiannis.
Constantinos at ESA 2013

Multiresolution Rasters

"An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters"
Lars Arge, Gerth Stølting Brodal and Constantinos Tsirogiannis.
Constantinos at ESA 2013

$\mathcal{O}(\text{scan}(N))$ IO's cache oblivious.

Multiresolution Rasters

"An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters"
Lars Arge, Gerth Stølting Brodal and Constantinos Tsirogiannis.
Constantinos at ESA 2013

$\mathcal{O}(\text{scan}(N))$ IO's cache oblivious.

Height model of Denmark:

Naive $\mathcal{O}(\text{sort}(N))$ IO algorithm: $13\frac{1}{4}$ Hours.

Multiresolution Rasters

"An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters"
Lars Arge, Gerth Stølting Brodal and Constantinos Tsirogiannis.
Constantinos at ESA 2013

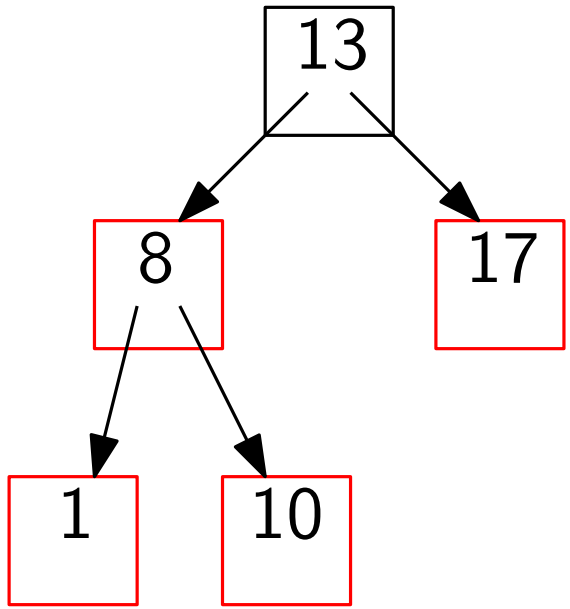
$\mathcal{O}(\text{scan}(N))$ IO's cache oblivious.

Height model of Denmark:

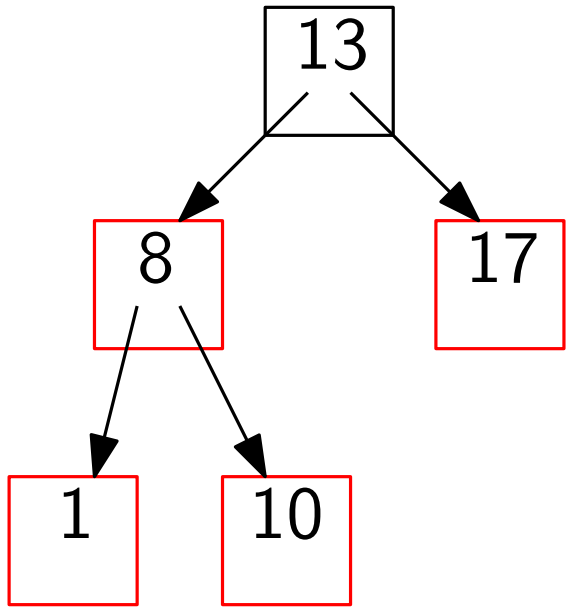
Naive $\mathcal{O}(\text{sort}(N))$ IO algorithm: $13\frac{1}{4}$ Hours.

Our algorithm: $2\frac{1}{4}$ Hours.

Implicite Structures

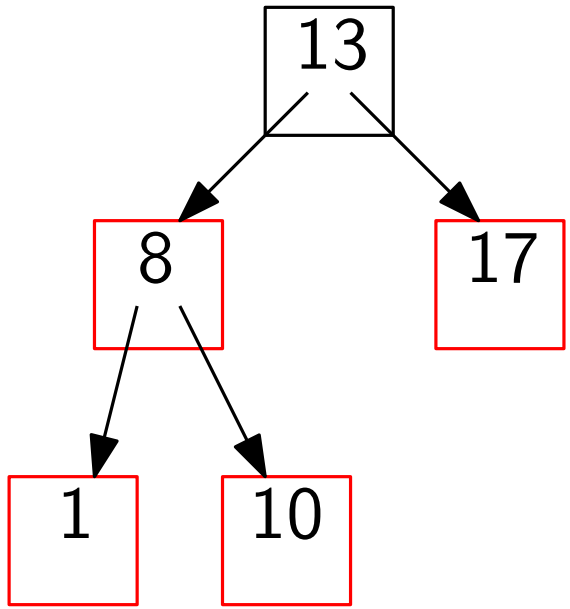


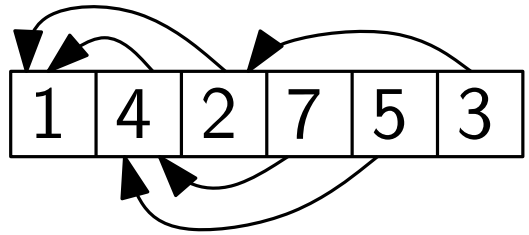
Implicite Structures



1	4	2	7	5	3

Implicite Structures





Workingset Structure

1	4	22	24	25	30	37	42	45	55	56	57	60	63	69	74	88	97	99

Workingset Structure

1	4	22	24	25	30	37	42	45	55	56	57	60	63	69	74	88	97	99

	Find	Predecessor	Update
Our	$\mathcal{O}(\log \ell)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$

A Cache-Oblivious Implicit Dictionary with the Working Set Property
Gerth Stølting Brodal, Casper Kejlberg-Rasmussen
ISAAC 2010

Workingset Structure

1	4	22	24	25	30	37	42	45	55	56	57	60	63	69	74	88	97	99

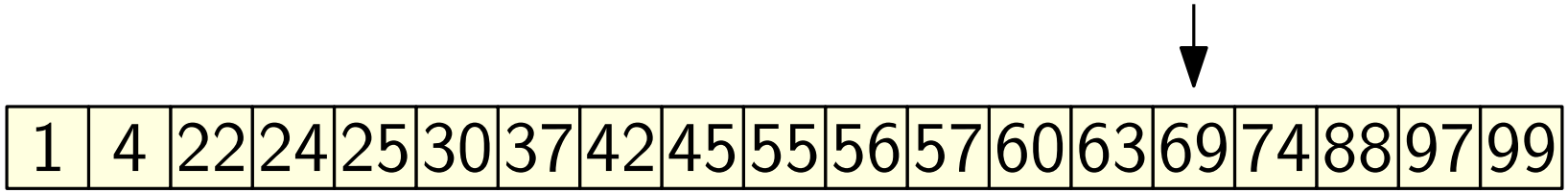
	Find	Predecessor	Update
Our	$\mathcal{O}(\log \ell)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
Better	$\mathcal{O}(\log \ell)$	$\mathcal{O}(\log \ell)$	$\mathcal{O}(\log n)$

A Cache-Oblivious Implicit Dictionary with the Working Set Property
Gerth Stølting Brodal, Casper Kejlberg-Rasmussen
ISAAC 2010

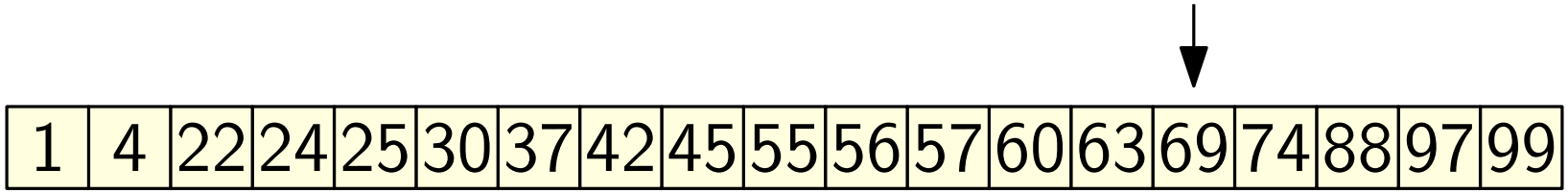
Fingersearch structure

1	4	22	24	25	30	37	42	45	55	56	57	60	63	69	74	88	97	99

Fingersearch structure

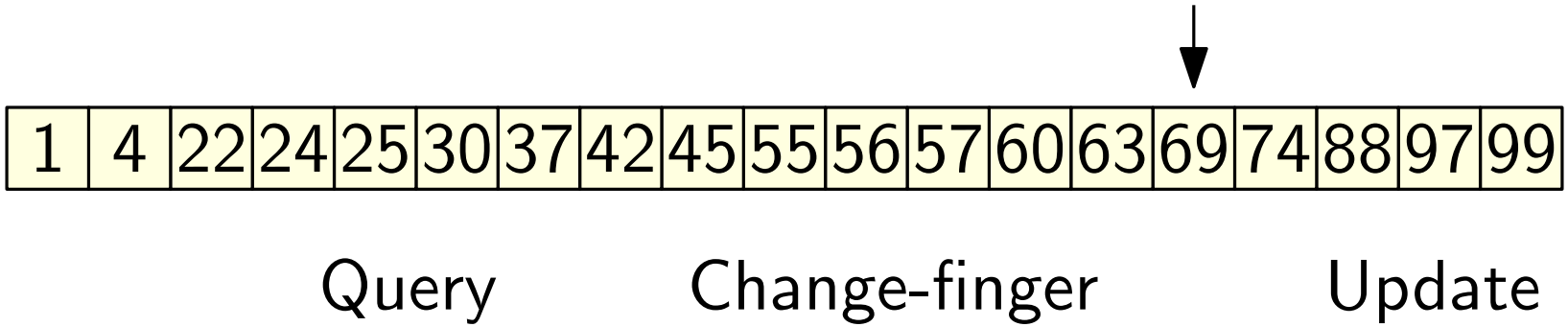


Fingersearch structure



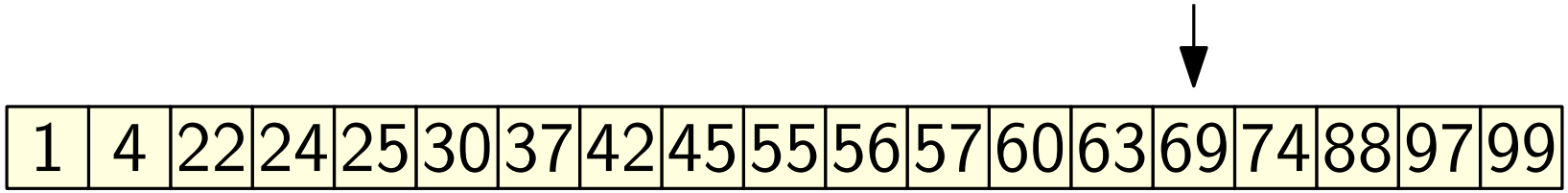
Query Change-finger Update

Fingersearch structure



Finger Search in the Implicit Model
Gerth Stølting Brodal and Jesper Sindahl Nielsen
IAAAC 2012

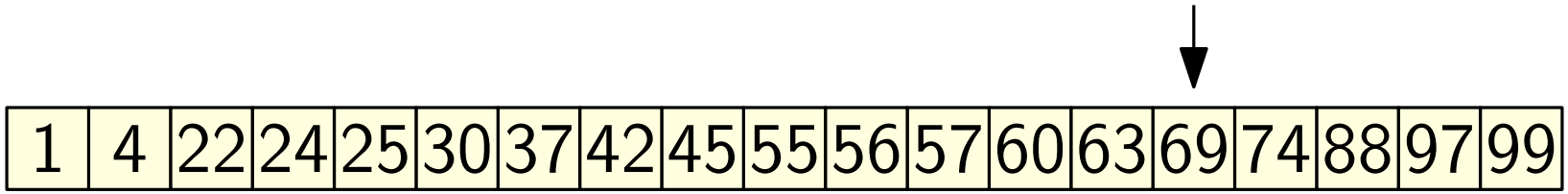
Fingersearch structure



	Query	Change-finger	Update
Static	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n))$	

Finger Search in the Implicit Model
Gerth Stølting Brodal and Jesper Sindahl Nielsen
IAAAC 2012

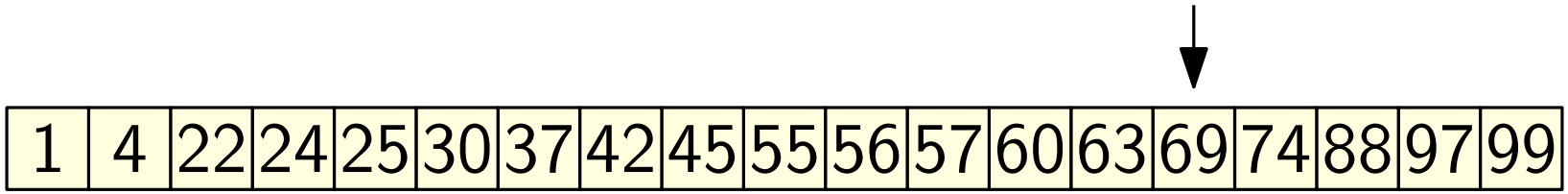
Fingersearch structure



	Query	Change-finger	Update
Static	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n))$	
Dynamic	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n) \log n)$	$\mathcal{O}(\log n)$

Finger Search in the Implicit Model
Gerth Stølting Brodal and Jesper Sindahl Nielsen
IAAAC 2012

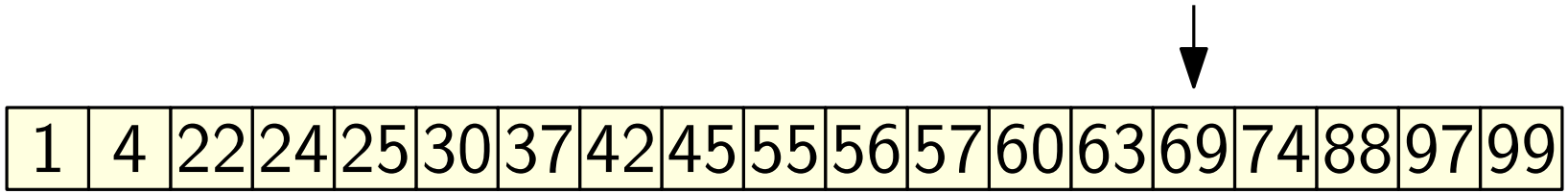
Fingersearch structure



	Query	Change-finger	Update
Static	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n))$	
Dynamic	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n)\log n)$	$\mathcal{O}(\log n)$
	$\mathcal{O}(\log t)$	$\mathcal{O}(n^{\varepsilon})$	$\mathcal{O}(\log n)$

Finger Search in the Implicit Model
Gerth Stølting Brodal and Jesper Sindahl Nielsen
IAAAC 2012

Fingersearch structure



	Query	Change-finger	Update
Static	$\mathcal{O}(q(t))^*$	$\mathcal{O}(q^{-1}(\log n))^*$	
Dynamic	$\mathcal{O}(q(t))$	$\mathcal{O}(q^{-1}(\log n)\log n)$	$\mathcal{O}(\log n)$
	$\mathcal{O}(\log t)$	$\mathcal{O}(n^\varepsilon)$	$\mathcal{O}(\log n)$

Finger Search in the Implicit Model
Gerth Stølting Brodal and Jesper Sindahl Nielsen
IAAAC 2012

Priority Queues

Insert(v) ExtractMin()

Priority Queues

	Insert(v)	ExtractMin()
Williams	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$

Priority Queues

	Insert(v)	ExtractMin()
Williams	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
Gerth Stølting Brodal And Jesper Sindahl Nielsen
WADS 2015

Priority Queues

	Insert(v)	ExtractMin()
Williams	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
	$\mathcal{O}(1)^*$	$\mathcal{O}(\log n)^*$

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
Gerth Stølting Brodal And Jesper Sindahl Nielsen
WADS 2015

Priority Queues

	Insert(v)	ExtractMin()
Williams	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
	$\mathcal{O}(1)^*$	$\mathcal{O}(\log n)^*$
	$\mathcal{O}(1)$	$\mathcal{O}(\log n)$

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
Gerth Stølting Brodal And Jesper Sindahl Nielsen
WADS 2015

Priority Queues

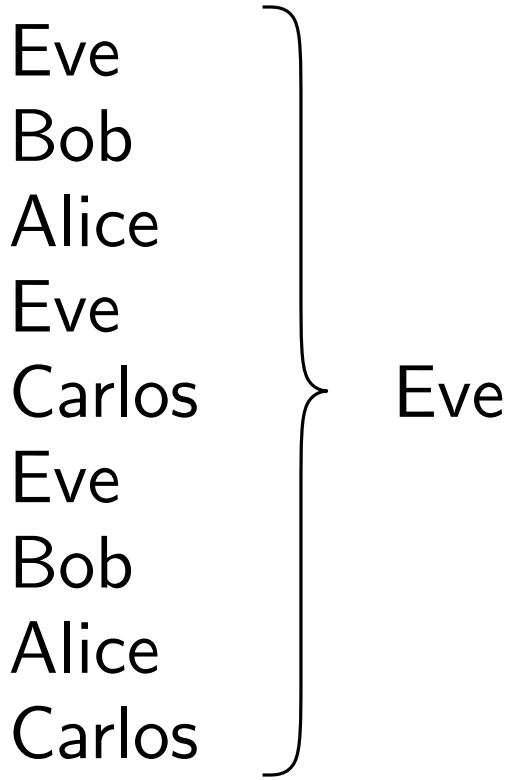
	Insert(v)	ExtractMin()	Moves
Williams	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
	$\mathcal{O}(1)^*$	$\mathcal{O}(\log n)^*$	$\mathcal{O}(1)^*$
	$\mathcal{O}(1)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
Gerth Stølting Brodal And Jesper Sindahl Nielsen
WADS 2015

Range Mode

Eve
Bob
Alice
Eve
Carlos
Eve
Bob
Alice
Carlos

Range Mode



Range Mode

Eve
Bob
Alice
Eve
Carlos
Eve
Bob
Alice
Carlos

} Carlos

Range Mode

Eve
Bob
Alice } Bob
Eve
Carlos
Eve
Bob
Alice
Carlos

Range Mode

Cell Probe Lower Bounds and Approximations for Range Mode
Mark Greve, Allan G. Jørgensen and Kasper G. Larsen
ICALP 2010

Range Mode

Cell Probe Lower Bounds and Approximations for Range Mode
Mark Greve, Allan G. Jørgensen and Kasper G. Larsen
ICALP 2010

	Query	Space
3-approximation	$\mathcal{O}(1)$	$\mathcal{O}(n)$

Range Mode

Cell Probe Lower Bounds and Approximations for Range Mode
Mark Greve, Allan G. Jørgensen and Kasper G. Larsen
ICALP 2010

	Query	Space
3-approximation	$\mathcal{O}(1)$	$\mathcal{O}(n)$
$(1+\varepsilon)$-approximation	$\mathcal{O}(\log \frac{1}{\varepsilon})$	$\mathcal{O}(\frac{n}{\varepsilon})$

Range Mode

Cell Probe Lower Bounds and Approximations for Range Mode
Mark Greve, Allan G. Jørgensen and Kasper G. Larsen
ICALP 2010

	Query	Space
3-approximation	$\mathcal{O}(1)$	$\mathcal{O}(n)$
$(1+\varepsilon)$-approximation	$\mathcal{O}(\log \frac{1}{\varepsilon})$	$\mathcal{O}(\frac{n}{\varepsilon})$

Lowerbound $\Omega\left(\frac{\log n}{\log(Sw/n)}\right)$

Summary

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode

ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Finger Search in the Implicit Model
ISAAC 2012 Gerth Stølting Brodal, Jesper Sindahl Nielsen, Jakob Truelsen

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Finger Search in the Implicit Model
ISAAC 2012 Gerth Stølting Brodal, Jesper Sindahl Nielsen, Jakob Truelsen

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
WADS 2015 Gerth Stølting Brodal, Jesper Sindahl Nielsen

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Finger Search in the Implicit Model
ISAAC 2012 Gerth Stølting Brodal, Jesper Sindahl Nielsen, Jakob Truelsen

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
WADS 2015 Gerth Stølting Brodal, Jesper Sindahl Nielsen

Practical

Simplifying Massive Contour Maps
ESA 2012 Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk, Jakob Truelsen

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Finger Search in the Implicit Model
ISAAC 2012 Gerth Stølting Brodal, Jesper Sindahl Nielsen, Jakob Truelsen

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
WADS 2015 Gerth Stølting Brodal, Jesper Sindahl Nielsen

Practical

Simplifying Massive Contour Maps
ESA 2012 Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk, Jakob Truelsen

An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters
ESA 2013 Lars Arge, Gerth Stølting Brodal, Jakob Truelsen, Constantinos Tsirogiannis

Summary

Theory

Cell Probe Lower Bounds and Approximations for Range Mode
ICALP 2010 Mark Greve, Allan Grønlund Jørgensen, Kasper Green Larsen, Jakob Truelsen

A Cache-Oblivious Implicit Dictionary with the Working Set Property
ISAAC 2010 Gerth Stølting Brodal, Casper Kejlberg-Rasmussen, Jakob Truelsen

Finger Search in the Implicit Model
ISAAC 2012 Gerth Stølting Brodal, Jesper Sindahl Nielsen, Jakob Truelsen

Strictly Implicit Priority Queues: On the Number of Moves and Worst-Case Time
WADS 2015 Gerth Stølting Brodal, Jesper Sindahl Nielsen

Practical

Simplifying Massive Contour Maps
ESA 2012 Lars Arge, Lasse Deleuran, Thomas Mølhave, Morten Revsbæk, Jakob Truelsen

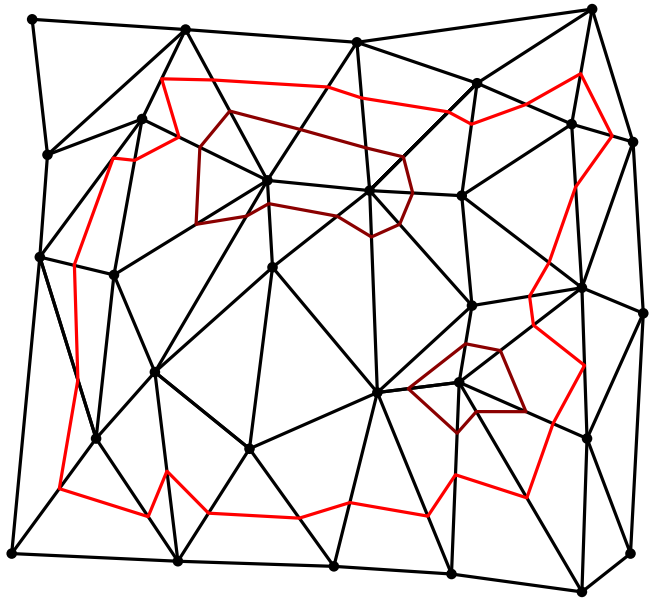
An Optimal and Practical Cache-Oblivious Algorithm for Computing Multiresolution Rasters
ESA 2013 Lars Arge, Gerth Stølting Brodal, Jakob Truelsen, Constantinos Tsirogiannis

Simplifying massive planar decompositions
Alenex 2014 Lars Arge, Jakob Truelsen, Jungwoo Yang

Thank You!

Decomposition Simplification

Decomposition Simplification



Lower Bounds

Let A be an operation running in time τ worst case. Then there exists a set $\mathcal{X}$ of size 2^τ, such that executing A will touch only cells from $\mathcal{X}$.

Lower Bounds

Let A be an operation running in time τ worst case. Then there exists a set $\mathcal{X}$ of size 2^τ, such that executing A will touch only cells from $\mathcal{X}$.

Fingersearch
Quiry time $q(t) \Rightarrow$ change-finger time $\Omega(q^{-1}(\log n))$.