# (1+ε)-Approximation for Facility Location in Data Streams

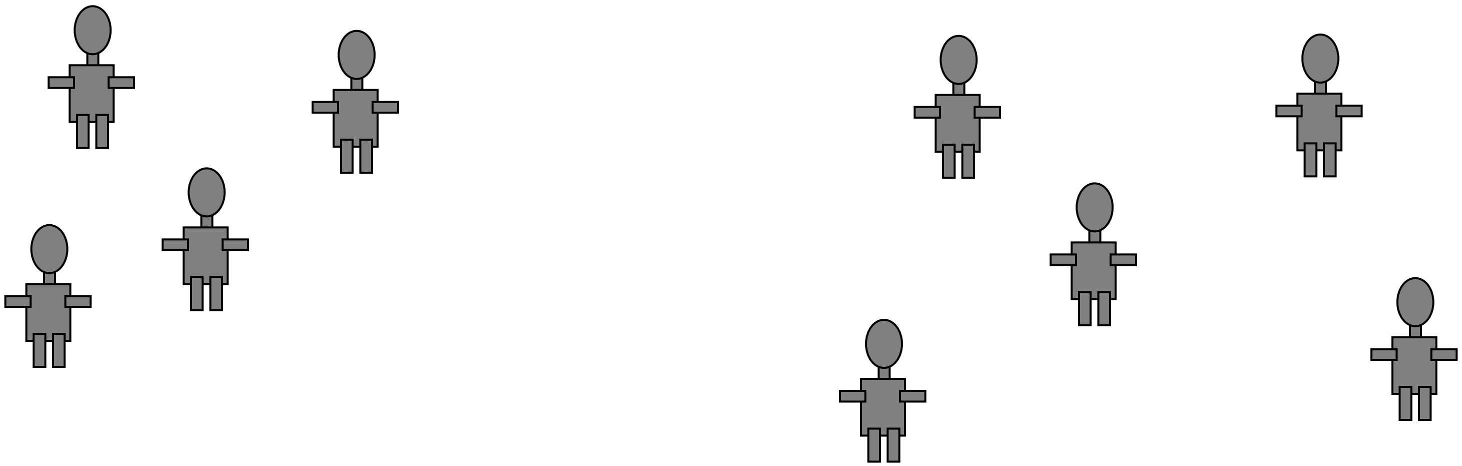
Artur Czumaj
Morteza Monemizadeh
Christiane Lammersen
Christian Sohler

SFU SIMON FRASER UNIVERSITY
COMPUTING SCIENCE

# Dynamic Geometric Data Streams

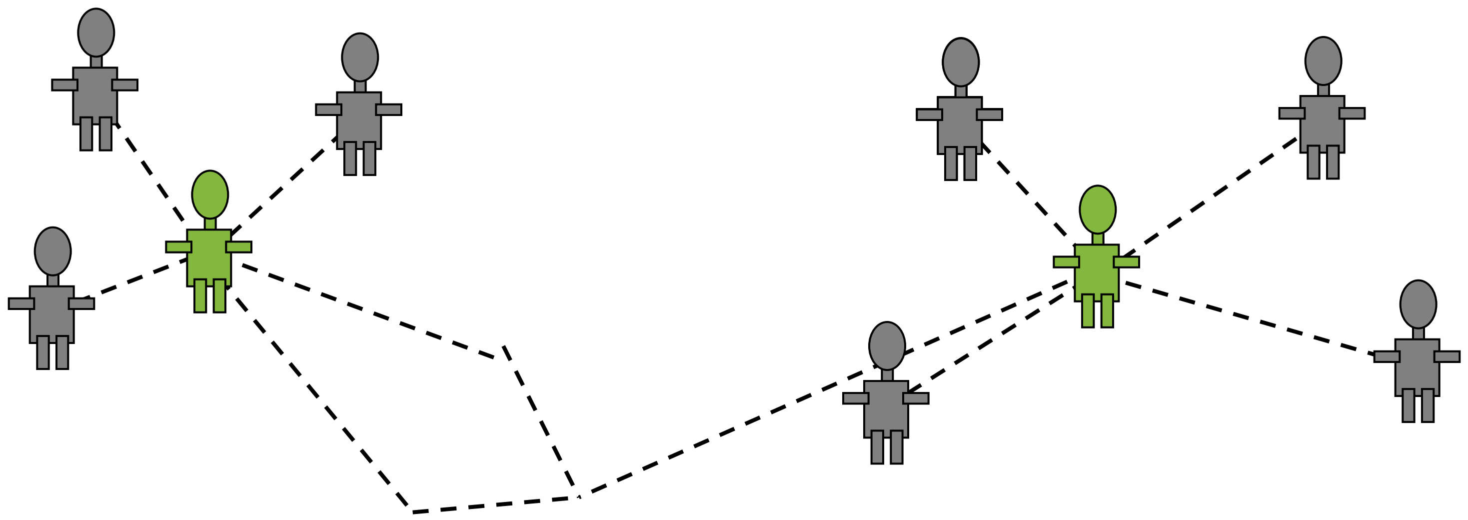
- Streams of geometric data arise in
  - Mobile networks
  - Sensor networks
  - …
- Continuously changing data
  - Mobile networks: position of nodes
  - Sensor networks: measured data
- Communication in form of update operations
  - Update consists of ID of node, old value, new value

# Hierarchical Communication Systems



- Upper layer offers lower layer a certain service
- Each node can be a server
- Cost for server ↔ cost for access to server

# Hierarchical Communication Systems

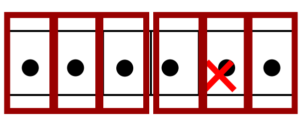


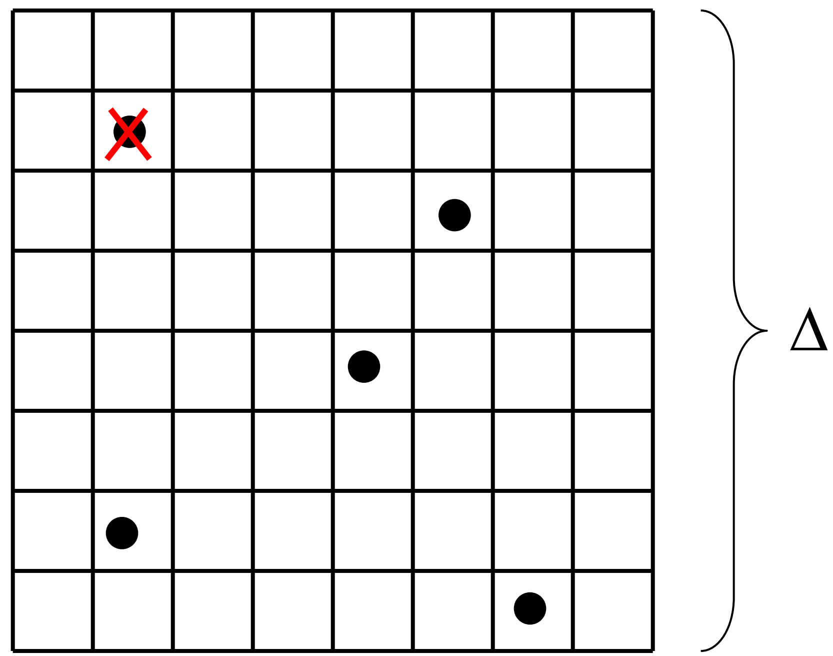
- Upper layer offers lower layer a certain service
- Each node can be a server
- Cost for server ↔ cost for access to server

# Dynamic Geometric Data Streams

- $m$ insert and delete operations
- Points in discrete Euclidean space $\{1, \ldots, \Delta\}^2$
- $\text{polylog}(\Delta, m)$ memory space, one pass

[Indyk '04]



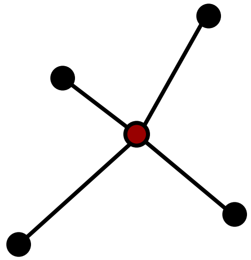


# Dynamic Uniform FLP

- Point set $P \subseteq \{1,\ldots,\Delta\}^2$
- Facilities have uniform opening cost $f$
- Goal: maintaining $F \subseteq \{1,\ldots,\Delta\}^2$ so as to minimize

$$f \cdot |F| + \sum_{p \in P} \min_{q \in F} \|p - q\|_2$$

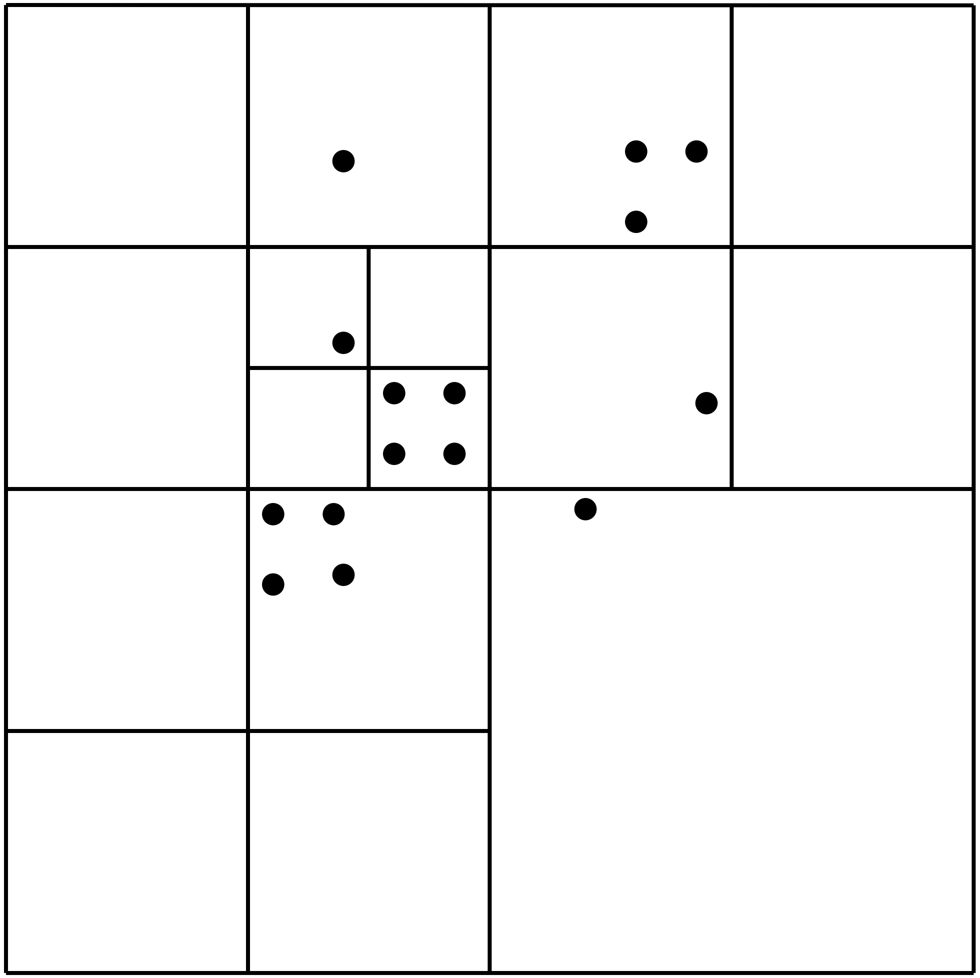
FLP related to $k$-median **but**
$|F|$ can be $\Theta(|P|)$
$\Rightarrow$ problem in streaming
$\Rightarrow$ approx of the cost



# Related Work

- Fotakis (STACS ’06):
  - Non-uniform, metric FLP in insertion-only model
  - $O(1)$-approx
  - Space linear in number of open facilities
- Indyk (STOC ’04):
  - Uniform FLP in dynamic geometric data streams
  - $O(\log^2\Delta)$-approx for cost of FLP
- Lammersen, Sohler (ESA '08):
  - Uniform FLP in dynamic geometric data streams
  - $O(1)$-approx for cost of FLP

# Idea: Decomposition into Small Subsets



**PTAS:**

- Decompose into cells with few open facilities
- Solve FLP in each cell
- Combine solutions of all cells

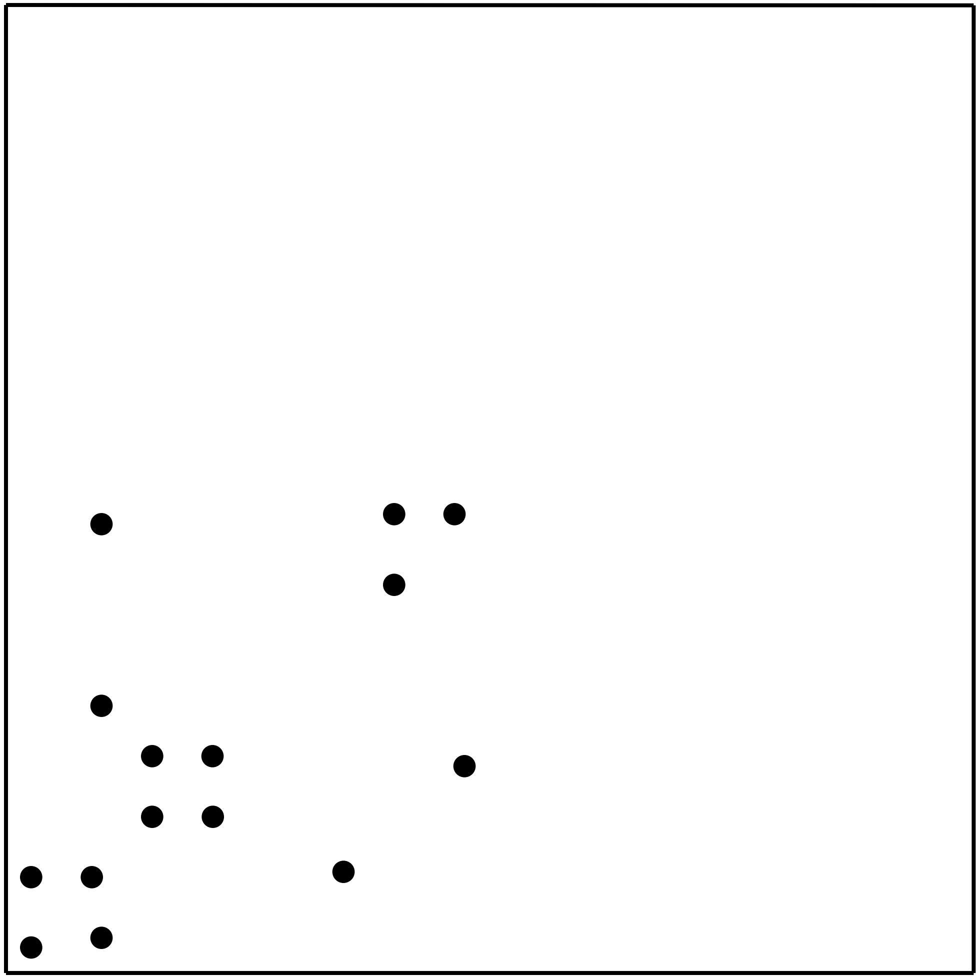
**Streaming:**

- Sample from cells
- Estimate cost from sample

# PTAS

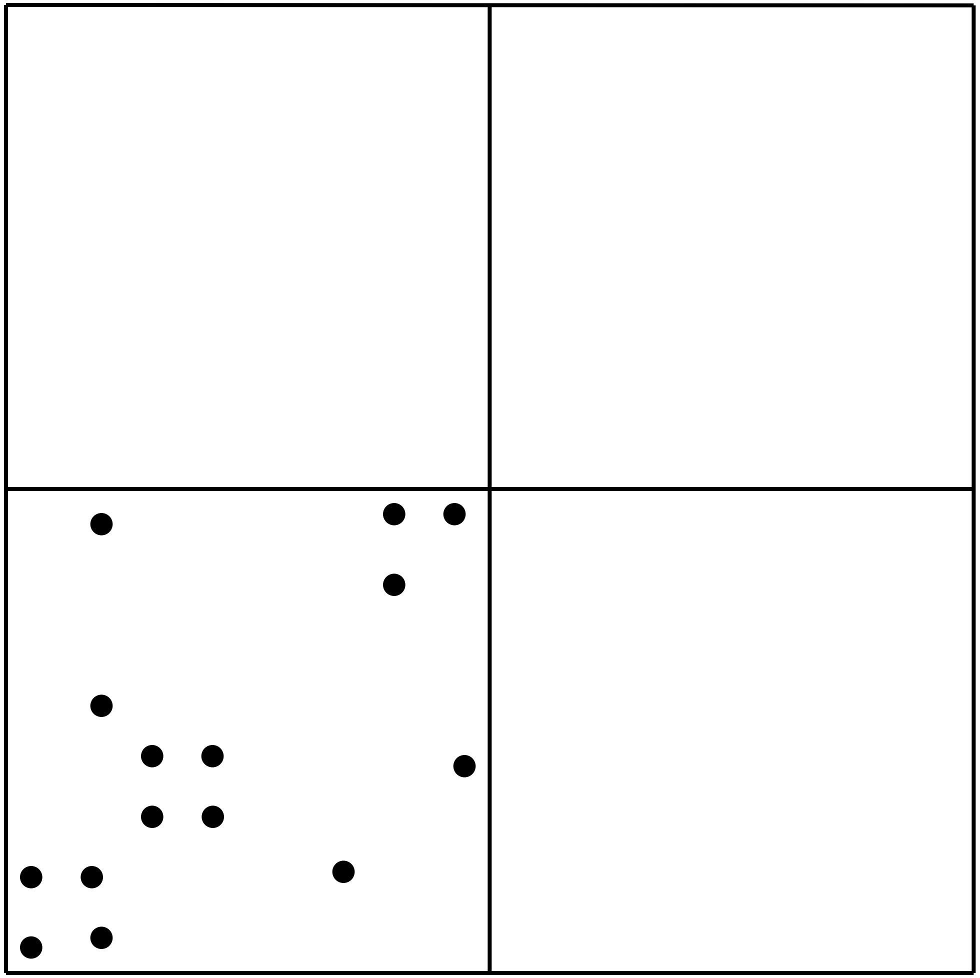
- Impose $\log(\Delta)+2$ randomly shifted nested grids
- Start partition with biggest cell
- Recursively split each heavy cell
- For each cell in partition, find $(1+\varepsilon)$-approx for FLP of cell
- Return union of solutions

# Randomly Shifted Nested Grids



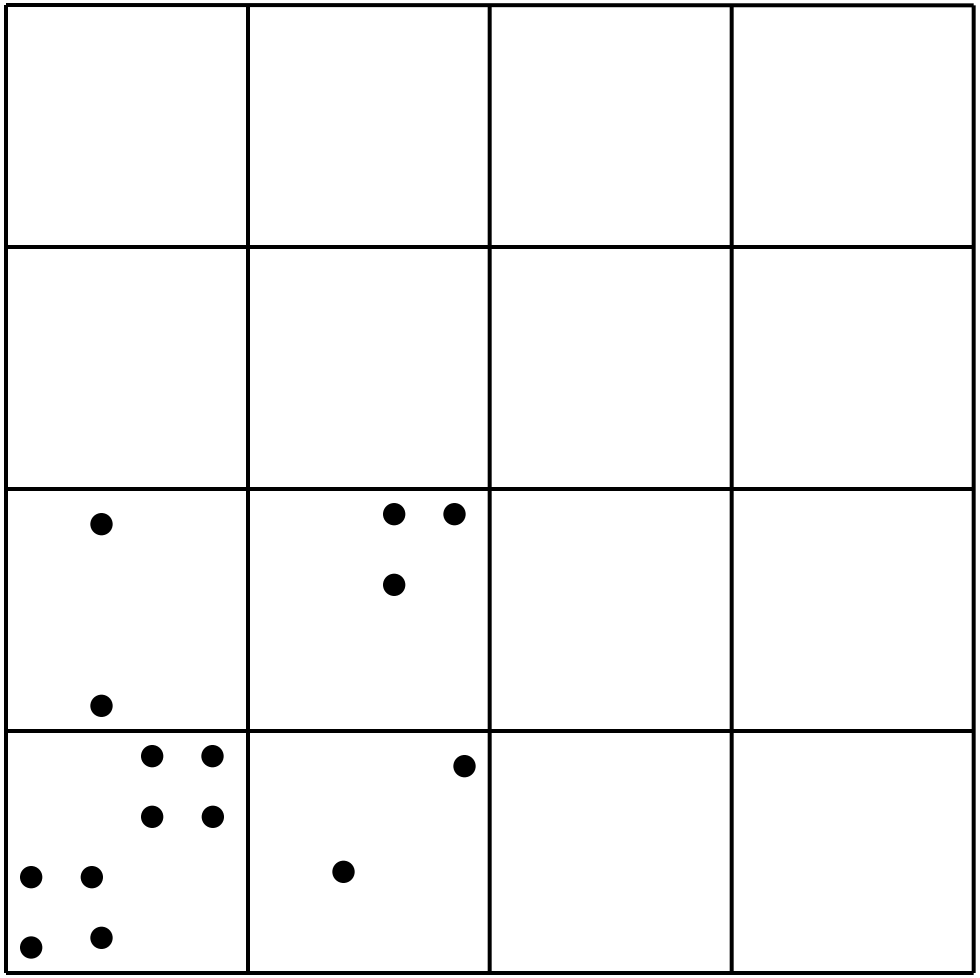
$\Delta = 8$
Level: 4

# Randomly Shifted Nested Grids



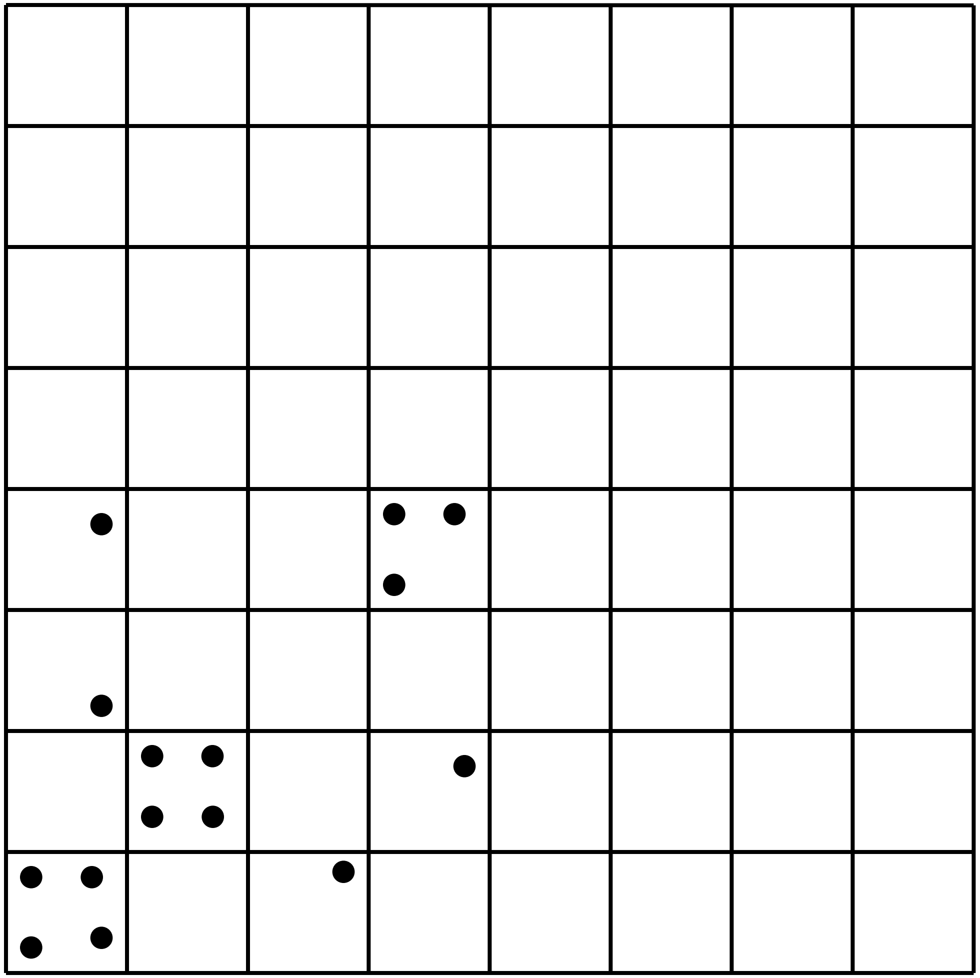
$\Delta = 8$
Level: 3

# Randomly Shifted Nested Grids



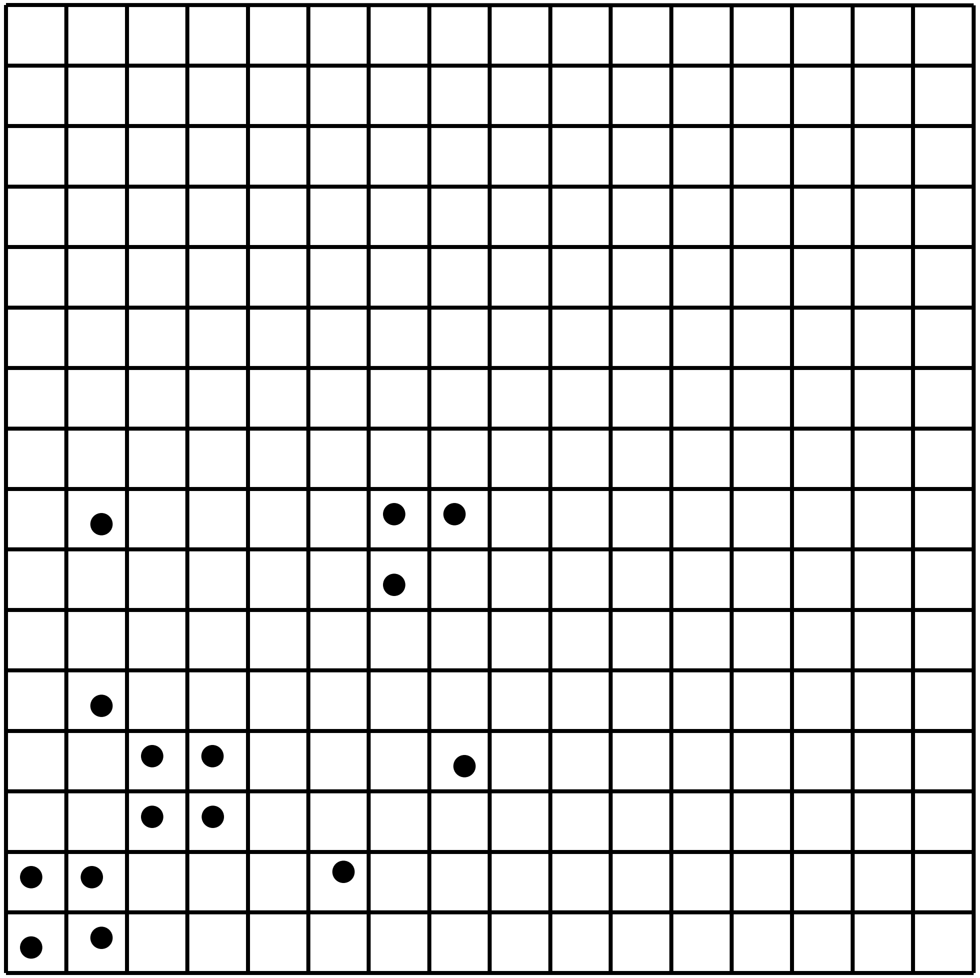
$\Delta = 8$
Level: 2

# Randomly Shifted Nested Grids



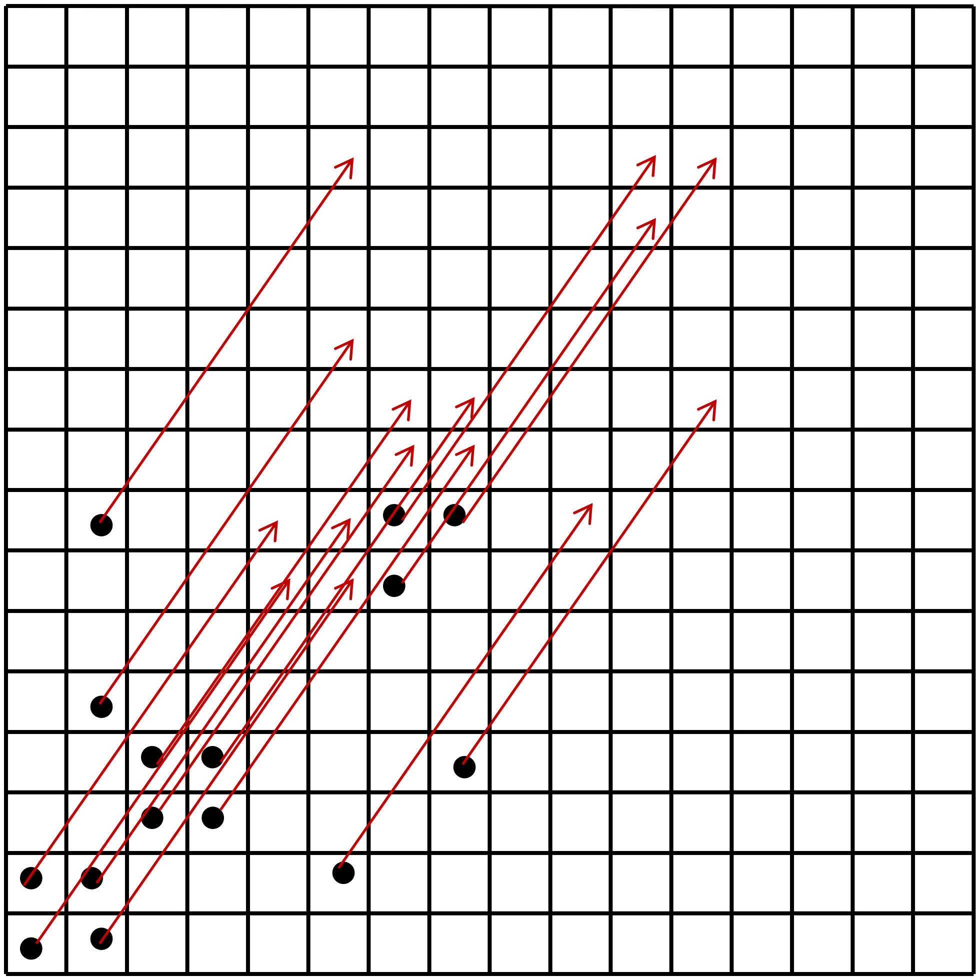
$\Delta = 8$
Level: 1

# Randomly Shifted Nested Grids



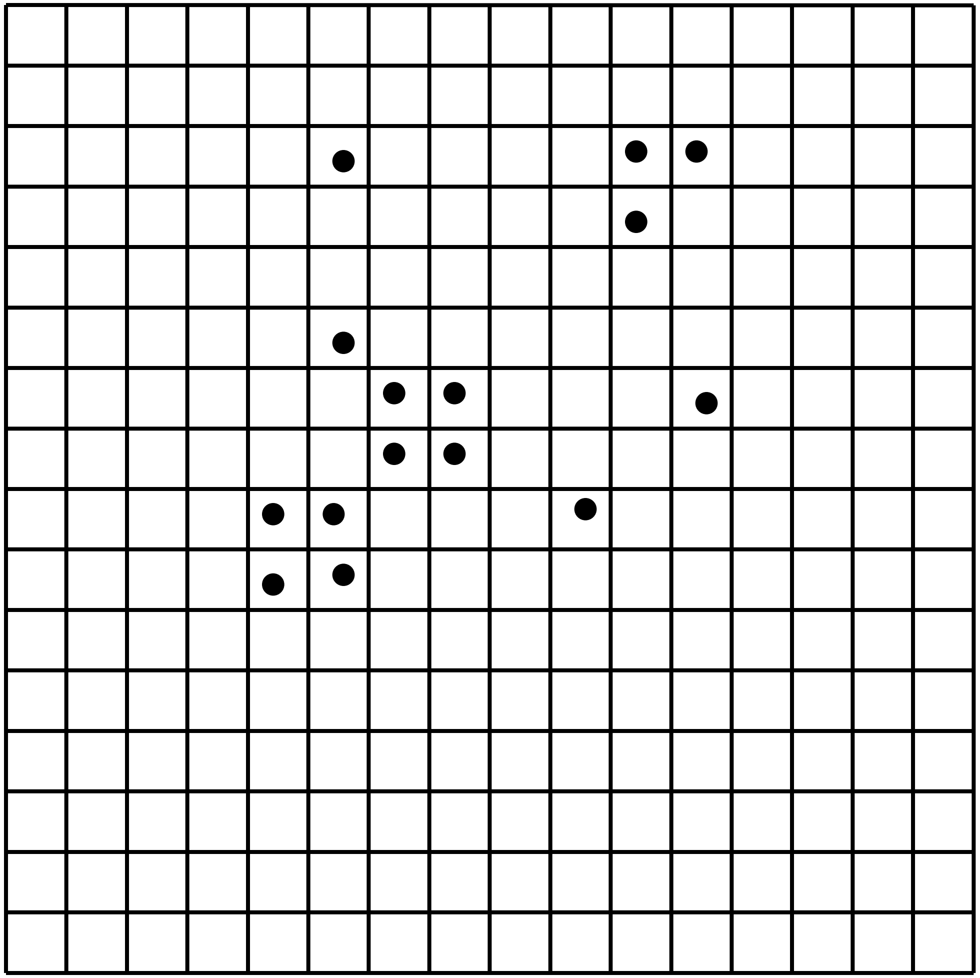
$\Delta = 8$
Level: 0

# Randomly Shifted Nested Grids



$\Delta = 8$
Level: 0

# Randomly Shifted Nested Grids

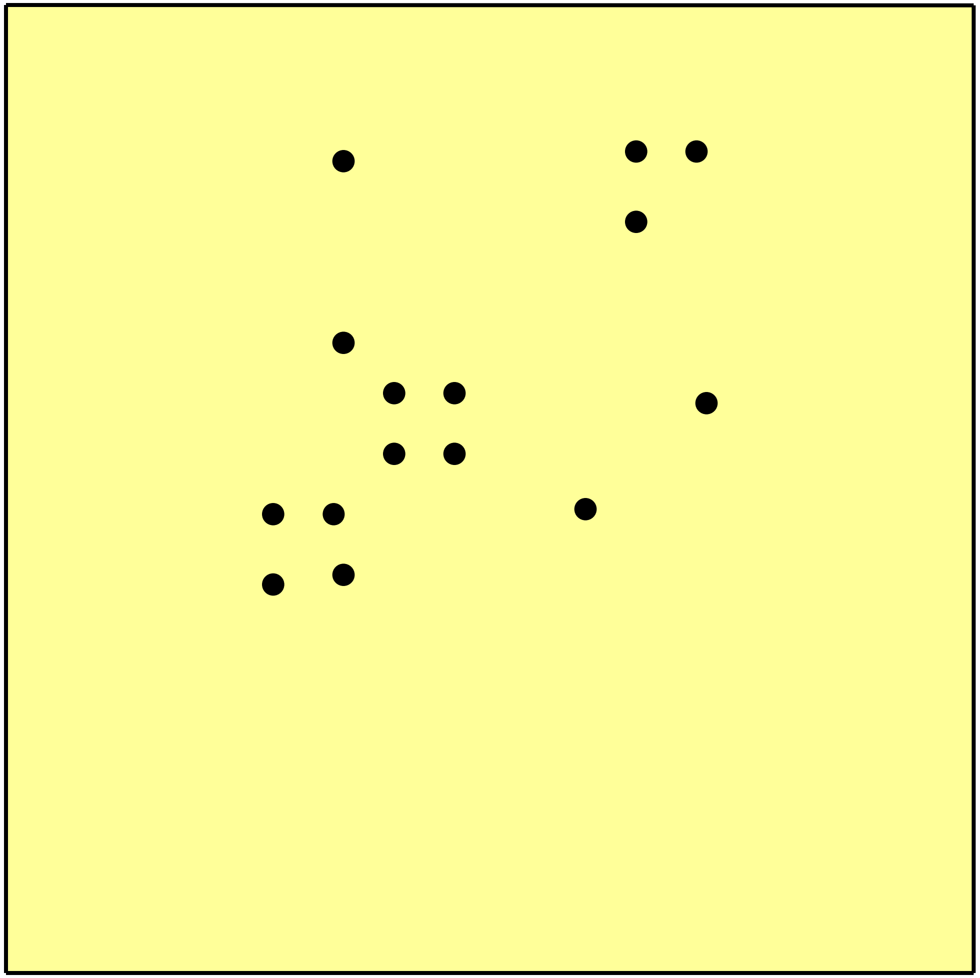


$\Delta = 8$
Level: 0

# PTAS

- Impose $\log(\Delta)+2$ randomly shifted nested grids
- Start partition with biggest cell
- Recursively split each heavy cell
- For each cell in partition, find $(1+\varepsilon)$-approx for FLP of cell
- Return union of solutions

# Space Partition



- Start partition with biggest cell
- Recursively split each heavy cell

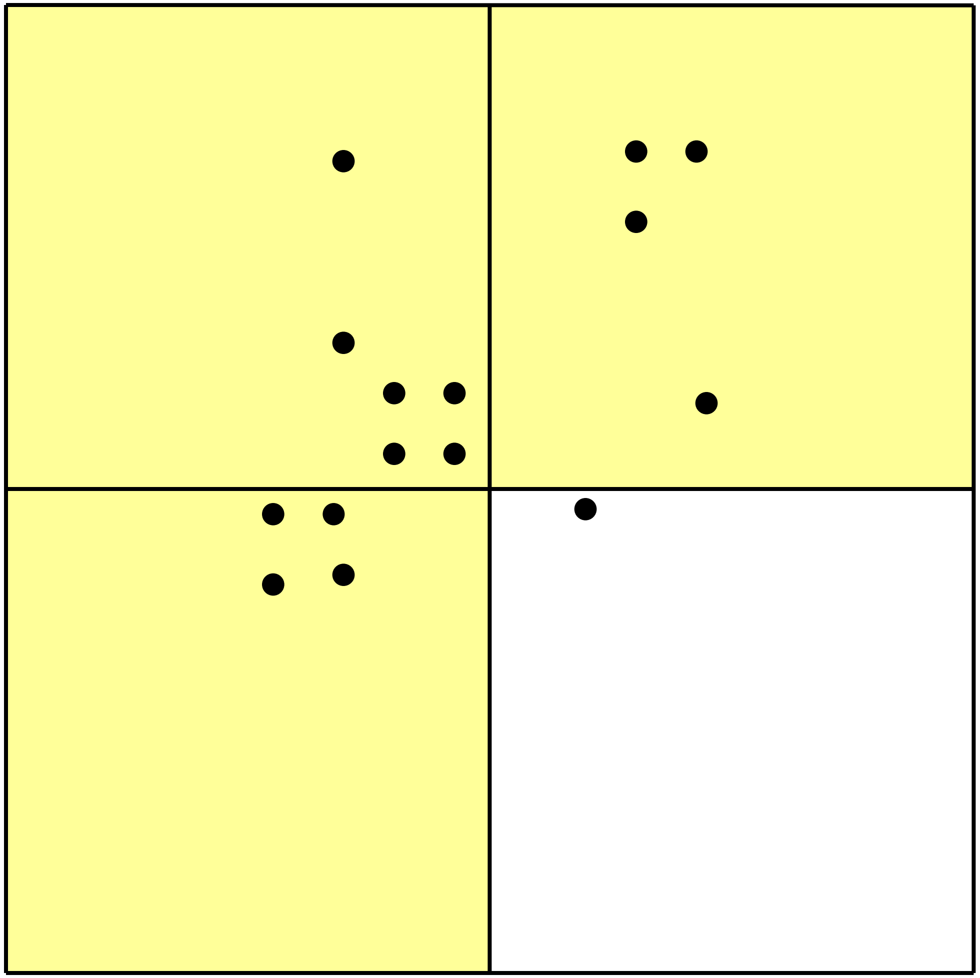
Cell is heavy if O(1)-approx outputs cost ≥ $T$.

$T := \text{poly}(\log(\Delta)/\varepsilon) \cdot f$

$\Delta = 8$

Level: 4

# Space Partition



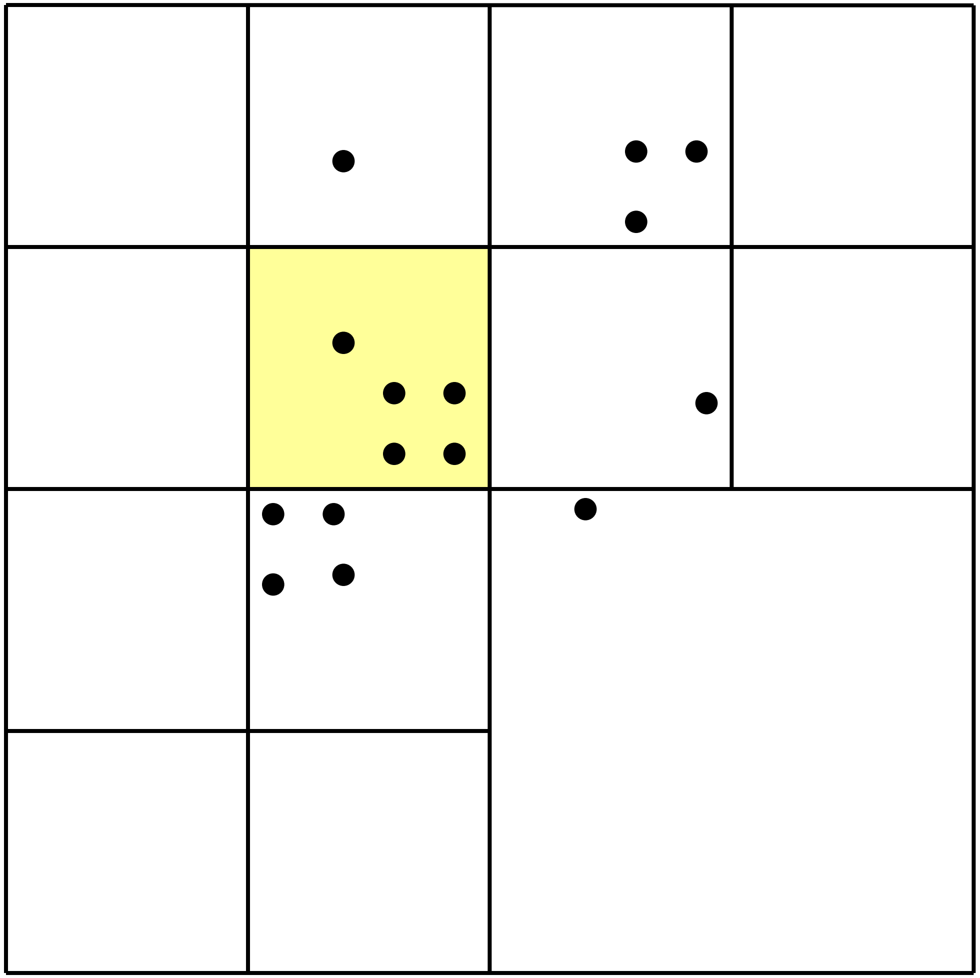
- Start partition with biggest cell
- Recursively split each heavy cell

Cell is heavy if O(1)-approx outputs cost ≥ $T$.

$T := \text{poly}(\log(\Delta)/\varepsilon) \cdot f$

$\Delta = 8$
Level: 3

# Space Partition



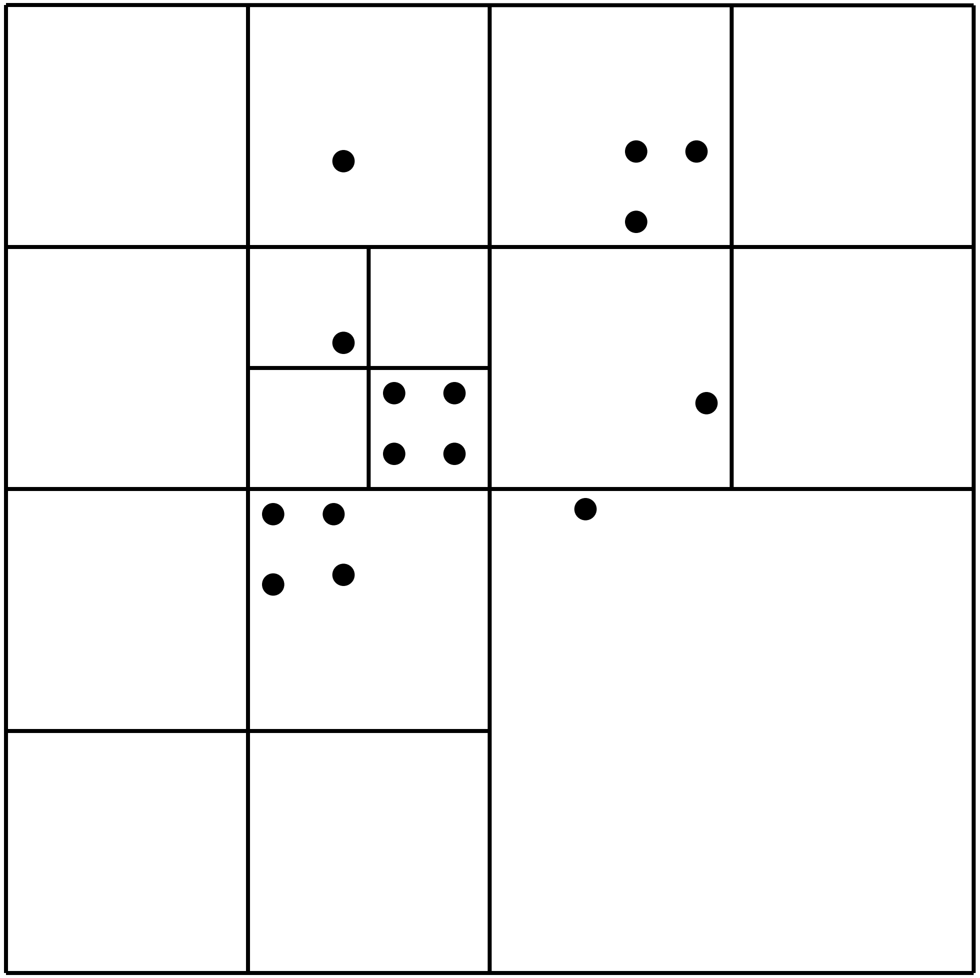
- Start partition with biggest cell
- Recursively split each heavy cell

Cell is heavy if O(1)-approx outputs cost ≥ $T$.

$T := \text{poly}(\log(\Delta)/\varepsilon) \cdot f$

$\Delta = 8$
Level: 2

# Space Partition



- Start partition with biggest cell
- Recursively split each heavy cell

Cell is heavy if O(1)-approx outputs cost ≥ $T$.

$T := \text{poly}(\log(\Delta)/\varepsilon) \cdot f$

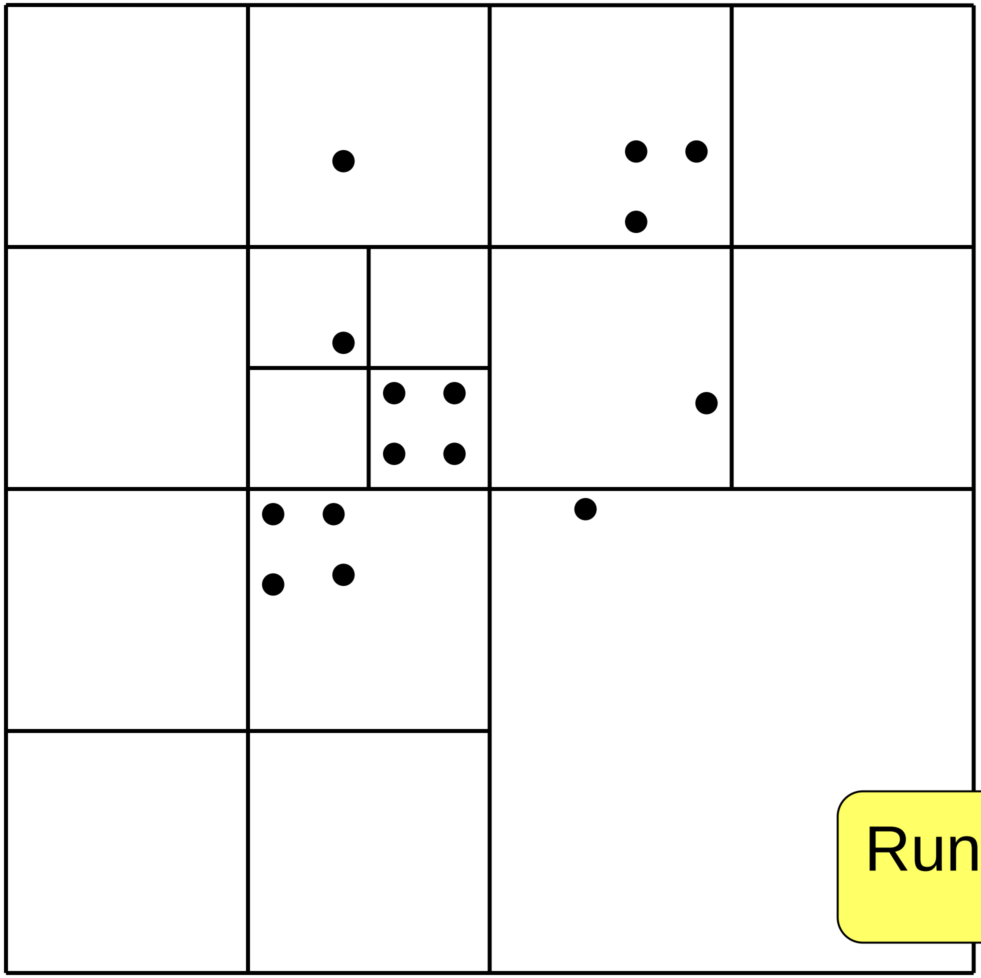
$\Delta = 8$

Level: $1$

# PTAS

- Impose $\log(\Delta)+2$ randomly shifted nested grids
- Start partition with biggest cell
- Recursively split each heavy cell
- For each cell in partition, find $(1+\varepsilon)$-approx for FLP of cell
- Return union of solutions

# Compute Solution

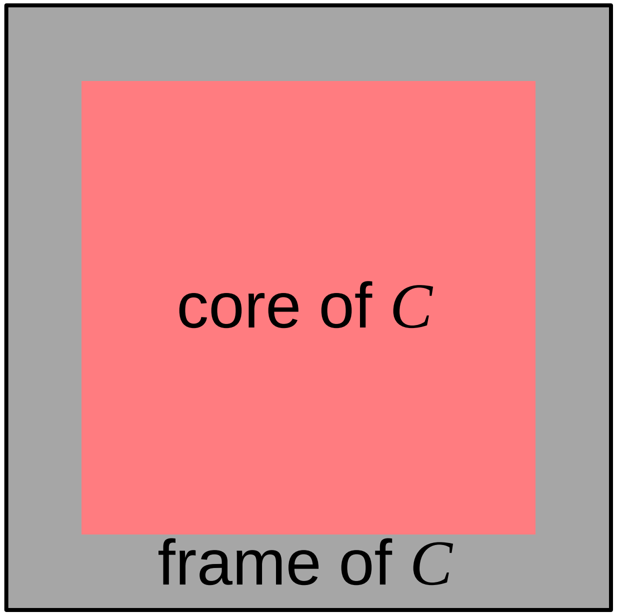


- For each cell, find $(1+\varepsilon)$-approx for FLP
- Return union of solutions

Cost of cell $\text{poly}(\log(\Delta)/\varepsilon) \cdot f$
Reduction to $k$-median
[Har-Peled, Mazumdar '04]

Runtime: $O(n \log \Delta \log n \log \log n)$

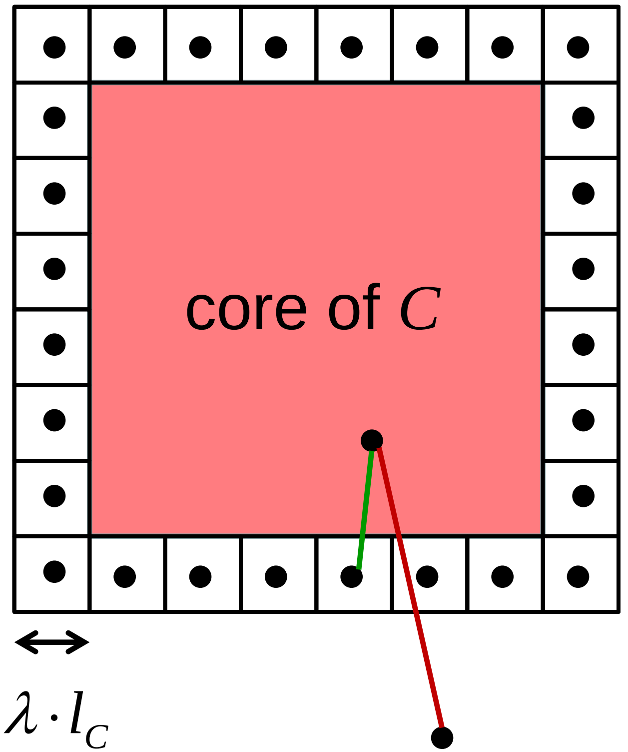
# Cores and Frames




$$\sum_{\text{max light } C} \text{Cost}(C) \le \sum_{\text{max light } C} \text{Cost}(\text{core of } C) + \text{Cost}(\text{frame of } C)$$

$$\le (1+\varepsilon) \cdot \text{OPT}$$

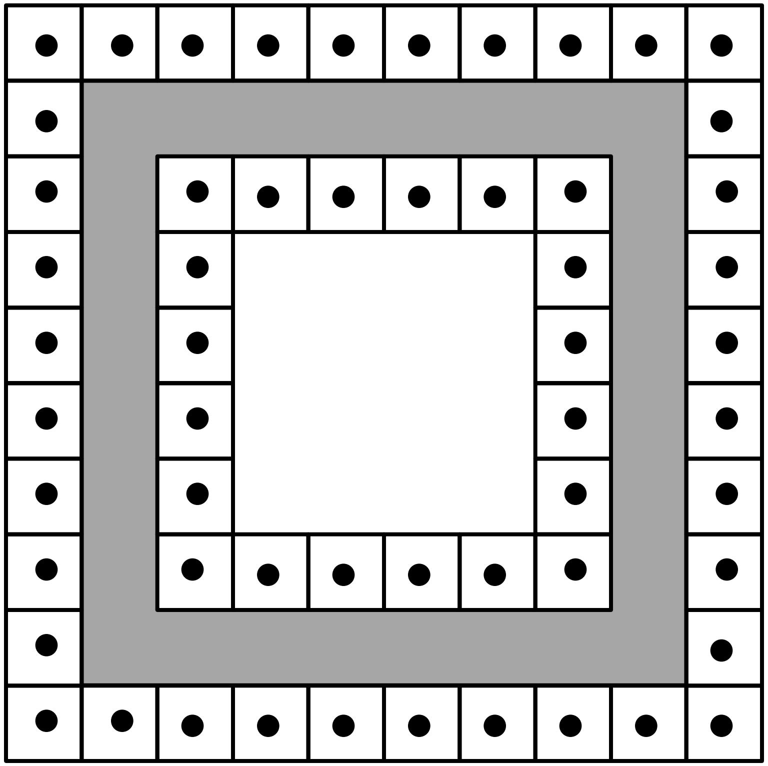
# Core and Buffer-Net of Core




$$4\left(\frac{1}{\lambda}-1\right) \text{ points in buffer-net}$$

$$\text{Cost}(\text{core of } C) \leq \text{Cost}_{\text{OPT}}(C) + \frac{4}{\lambda} \cdot f$$

# Frame and Buffer-Net of Frame



$$8\left(\frac{1}{\lambda}-1\right) \text{ points in buffer-net}$$

$$\text{Cost}(\text{frame of } C) \le \text{Cost}_{\text{OPT}}(\text{frame and buffer of } C) + \frac{8}{\lambda} \cdot f$$

$$E\left[\sum_{\text{max light } C} \text{Cost}_{\text{OPT}}(\text{frame and buffer of } C)\right] \le 12 \cdot \lambda \cdot (\log \Delta + 2) \cdot \text{OPT}$$

# Analysis of Cost Estimator

$$\sum_{\text{max light } C} \mathrm{Cost}(C)$$

$$\leq \sum_{\text{max light } C} \mathrm{Cost}(\text{core of } C) + \mathrm{Cost}(\text{frame of } C)$$

$$\leq (1 + c \cdot \lambda \cdot (\log \Delta + 2)) \cdot \mathrm{OPT} + \#\text{max light cells} \frac{12}{\lambda} f$$

O(OPT/$T$) heavy cells without heavy subcells.
Each cell creates ≤ 4(log Δ+2) max light cells.

# Approximating the Cost

FL($P$)

Cost($P$) $\leftarrow$ 0

Apx($P$) $\leftarrow$ O(1)-approx of FL cost of $P$

**for** each cell $C$ in each level **do**

Sample $C$ with prob. $s$ / Apx($P$)

**if** $C$ is maximal light **then**

Cost($C$) $\leftarrow$ (1+ε)-approx of FL cost of $C$

Cost($P$) $\leftarrow$ Cost($P$) + Cost($C$)

**return** Cost($P$) · Apx($P$) / $s$

# Approximating the Cost

## FL-Sampling $(P, s)$

$\text{Cost}(P) \leftarrow 0$

$\text{Apx}(P) \leftarrow$ O(1)-approx of FL cost of $P$

**for** each cell $C$ in each level **do**

  Sample $C$ with prob. $s / \text{Apx}(P)$

  **if** $C$ is sampled and $\delta$-detectable **then**

    $\text{Cost}(C) \leftarrow (1+\varepsilon)$-approx of FL cost of $C$

    $\text{Cost}(P) \leftarrow \text{Cost}(P) + \text{Cost}(C)$

**return** $\text{Cost}(P) \cdot \text{Apx}(P) / s$

Cost of $C \leq (1-\delta)\cdot T$ and cost of parent of $C \geq (1+\delta)\cdot T$

# Approximating the Cost

## **FL-Sampling** $(P, s)$

Cost($P$) $\leftarrow 0$

Apx($P$) $\leftarrow$ O(1)-approx of FL cost of $P$

**for** each cell $C$ in each level **do**

  Sample $C$ with prob. $s$ / A x($P$)

  **if** $C$ is sampled and $\delta$-detec **then**

    Cost($C$) $\leftarrow$ (1+$\varepsilon$)-approx of t of $C$

    Cost($P$) $\leftarrow$ Cost($P$) + C

**return** Cost($P$) · Apx($P$) / $s$

- O(log $\Delta$) guesses of Apx($P$)
- Maintain O(1)-approx

[Lammersen, Sohler '08]

# Approximating the Cost

## **FL-Sampling** $(P, s)$

$\mathrm{Cost}(P) \leftarrow 0$

$\mathrm{Apx}(P) \leftarrow$ O(1)-approx of FL cost of $P$

**for** each cell $C$ in each level **do**

  Sample $C$ with prob. $s$ / $\mathrm{Apx}(P)$

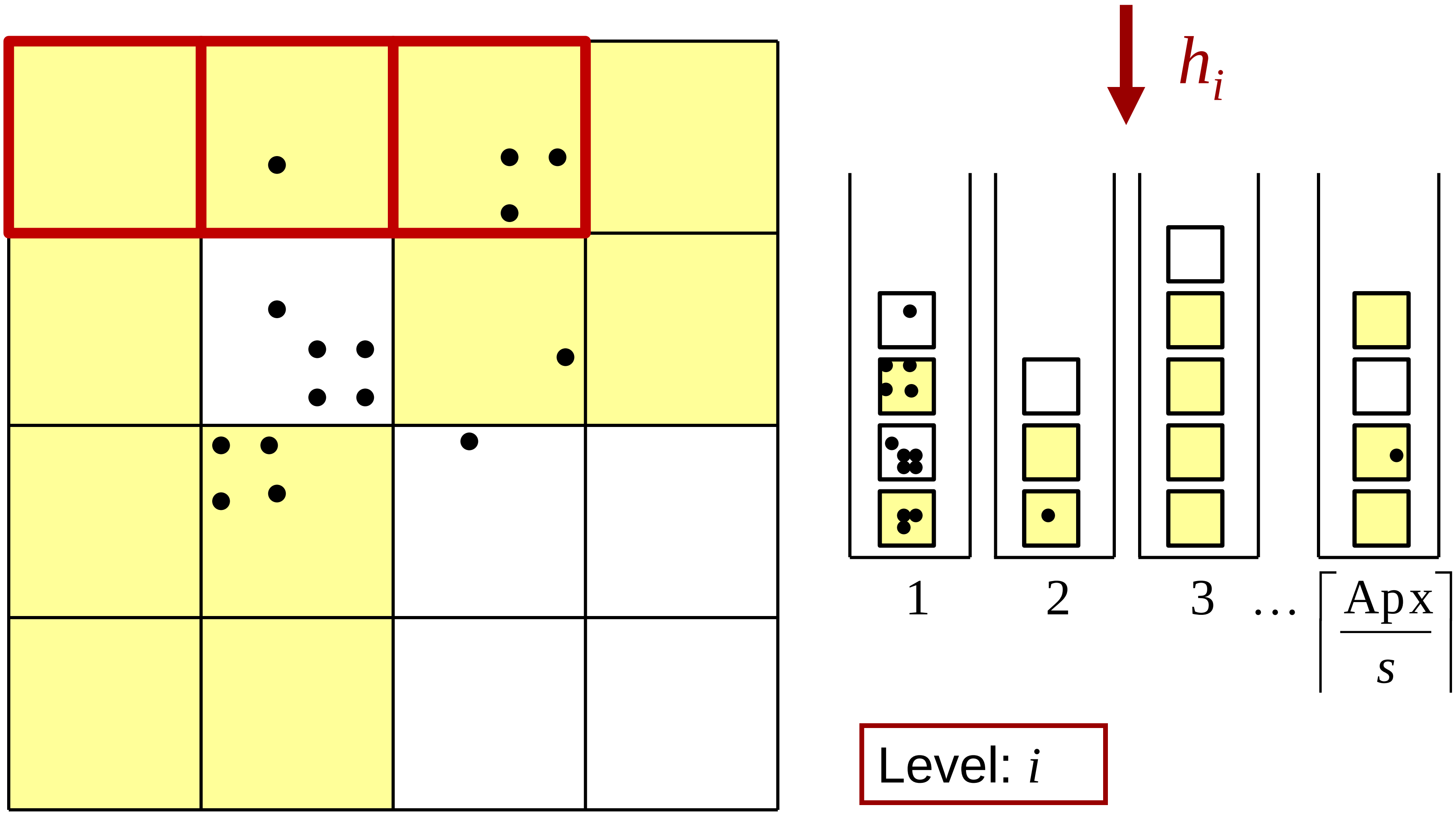
  **if** $C$ is sampled and $\delta$-detectable **then**

    $\mathrm{Cost}(C) \leftarrow (1+\varepsilon)$-approx of FL cost of $C$

    $\mathrm{Cost}(P) \leftarrow \mathrm{Cost}(P) + \mathrm{Cost}(C)$

**return** $\mathrm{Cost}(P) \cdot \mathrm{Apx}(P) / s$

# Sampling of Cells



# Approximating the Cost

## **FL-Sampling** $(P, s)$

$\text{Cost}(P) \leftarrow 0$

$\text{Apx}(P) \leftarrow$ O(1)-approx of FL cost of $P$

**for** each cell $C$ in each level **do**

Sample $C$ with prob. $s / \text{Apx}(P)$

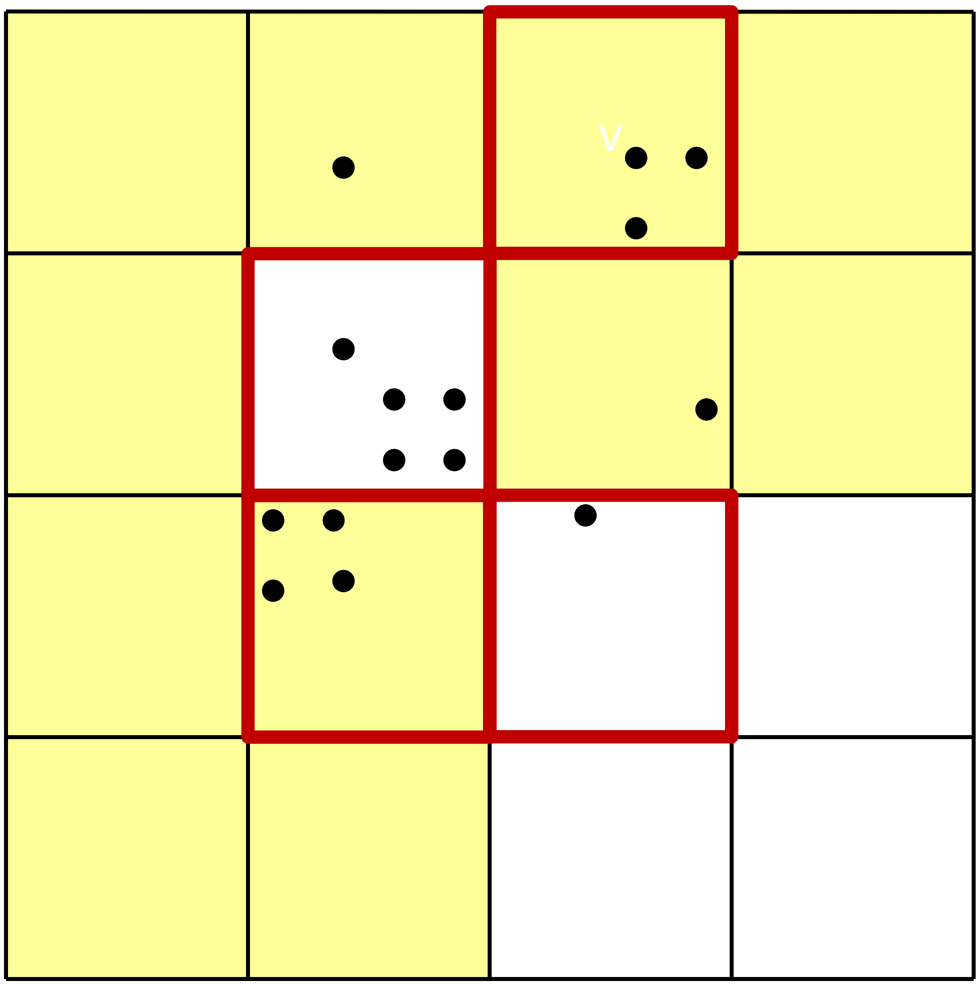
**if** $C$ is sampled and $\delta$-detectable **then**

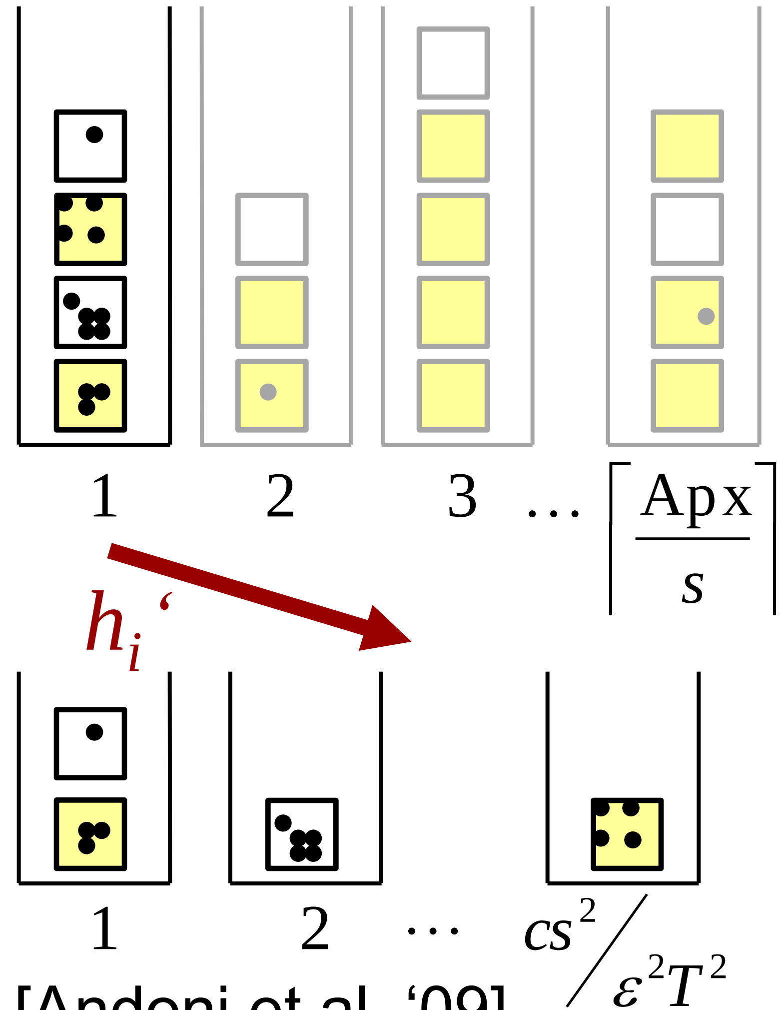
$\text{Cost}(C) \leftarrow (1+\varepsilon)$-approx of FL cost of $C$

$\text{Cost}(P) \leftarrow \text{Cost}(P) + \text{Cost}(C)$

**return** $\text{Cost}(P) \cdot \text{Apx}(P) / s$

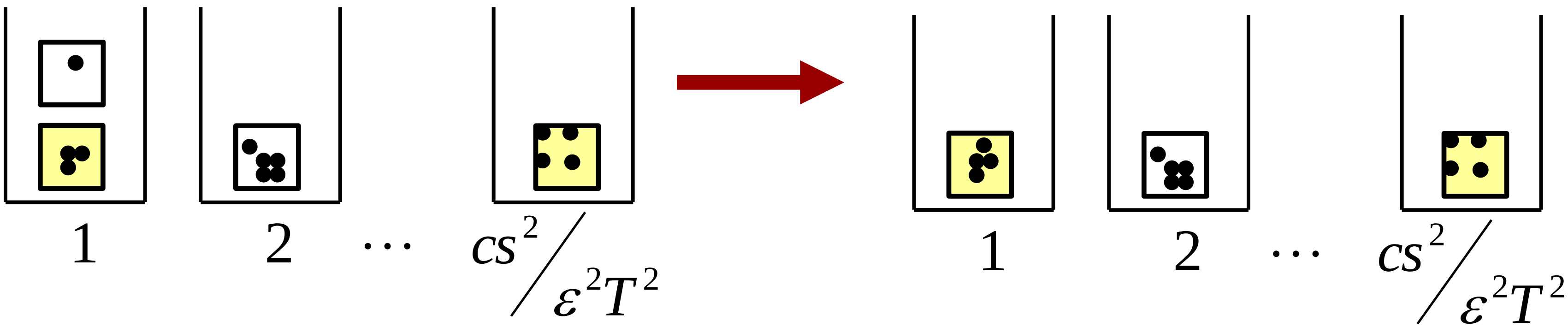
# Isolating δ-Detectable Cells




[Andoni et al. '09]

# δ-Detectable Cells




- δ-detectable:
  - cost of $C \leq (1-\delta)\cdot T$
  - cost of parent of $C \geq (1+\delta)\cdot T$
- Take minimum over all $k \in \{1,\ldots,\lceil T/f \rceil\}$ of $k$-median cost plus $k \cdot f$
- Maintain $(k, \varepsilon)$-coreset [Frahling, Sohler '05]

# Conclusion & Future Work

## Conclusion:

- $(1+\varepsilon)$- approx of cost for dynamic FLP in $\{1,\ldots,\Delta\}^2$
- Space: $(\log(\Delta)/\varepsilon)^{O(1)}$

## Future Work:

- $(1+\varepsilon)$-approx of cost for dynamic FLP in $\{1,\ldots,\Delta\}^d$
- Max independent set in connected disk graphs
- Max matching in connected unit disk graphs

# Thank you for your attention!

*School of Computing Science*
*Simon Fraser University*
*8888 University Drive*
*Burnaby, B.C., V5A 1S6, Canada*

*Phone: +1.778.782.7331*
*E-mail: christiane_lammersen@sfu.ca*

SFU SIMON FRASER UNIVERSITY COMPUTING SCIENCE