Expected case for projecting points

Sergio Cabello
Matt deVos
Bojan Mohar

} IMFM, Ljubljana, Slovenia

FMF, Ljubljana, Slovenia

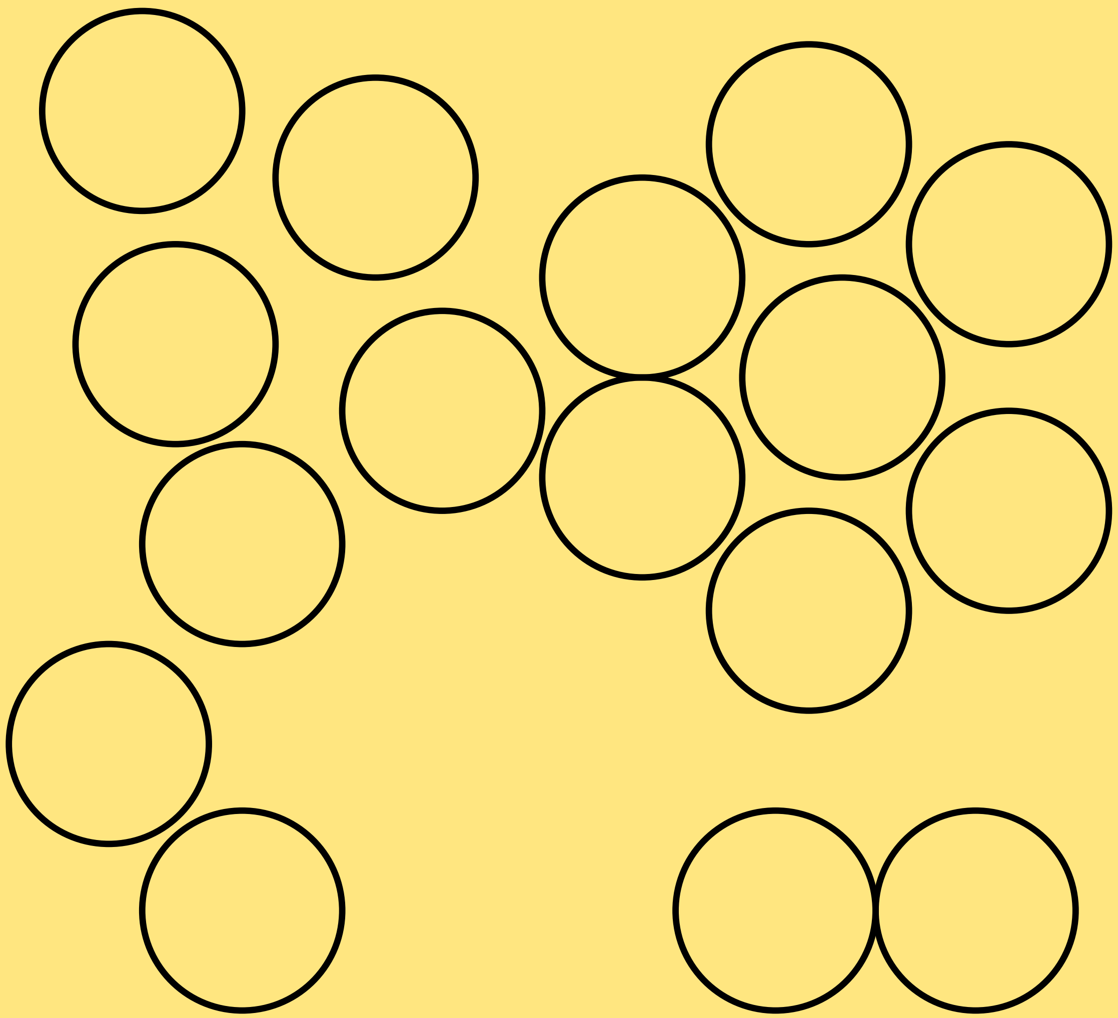
S.C. partially supported by an EC/FP6 Marie Curie Fellowship.

Motivation

Sweep-line among unit-size disjoint disks

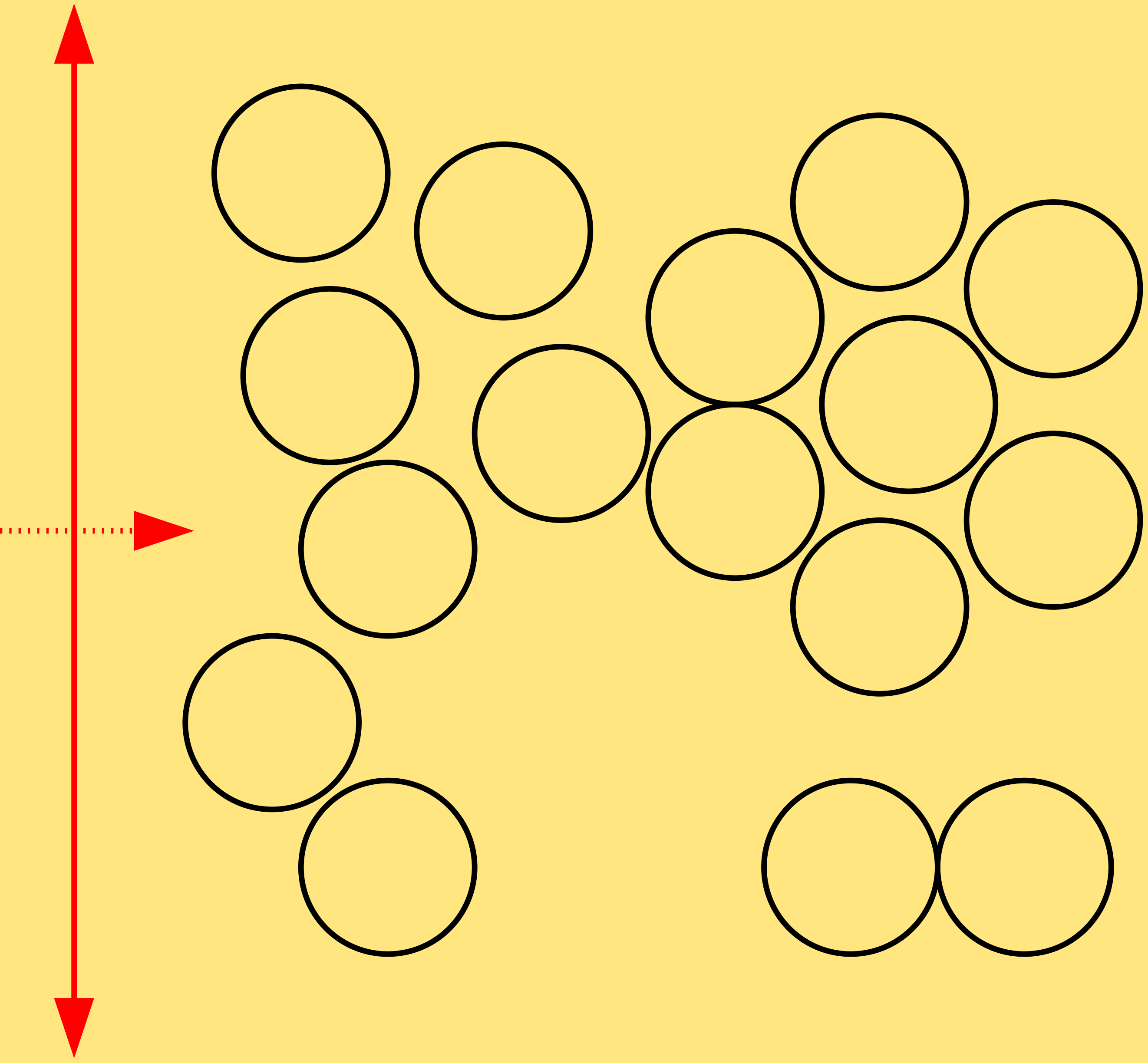
Motivation

Sweep-line among unit-size disjoint disks



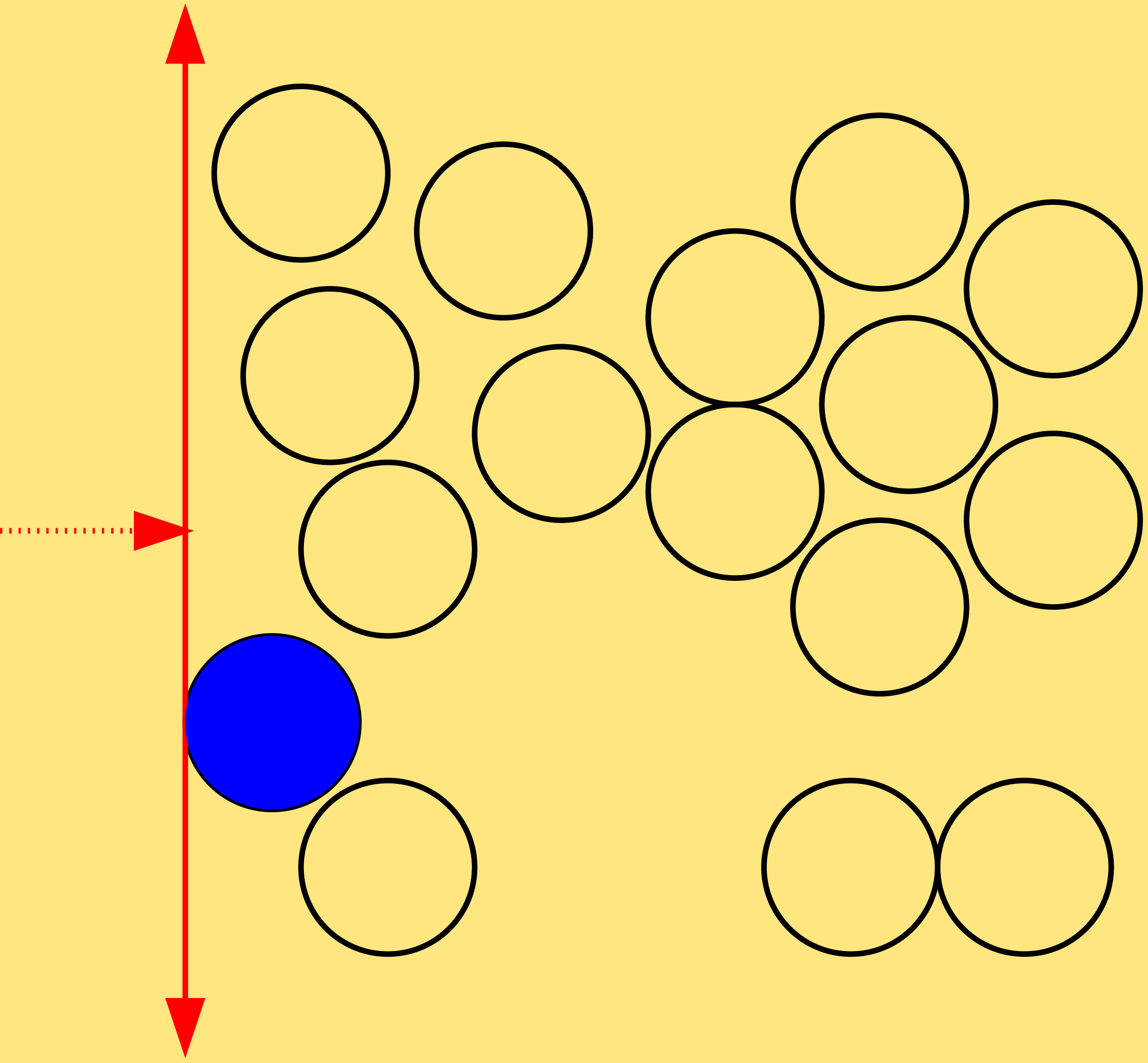
Motivation

Sweep-line among unit-size disjoint disks



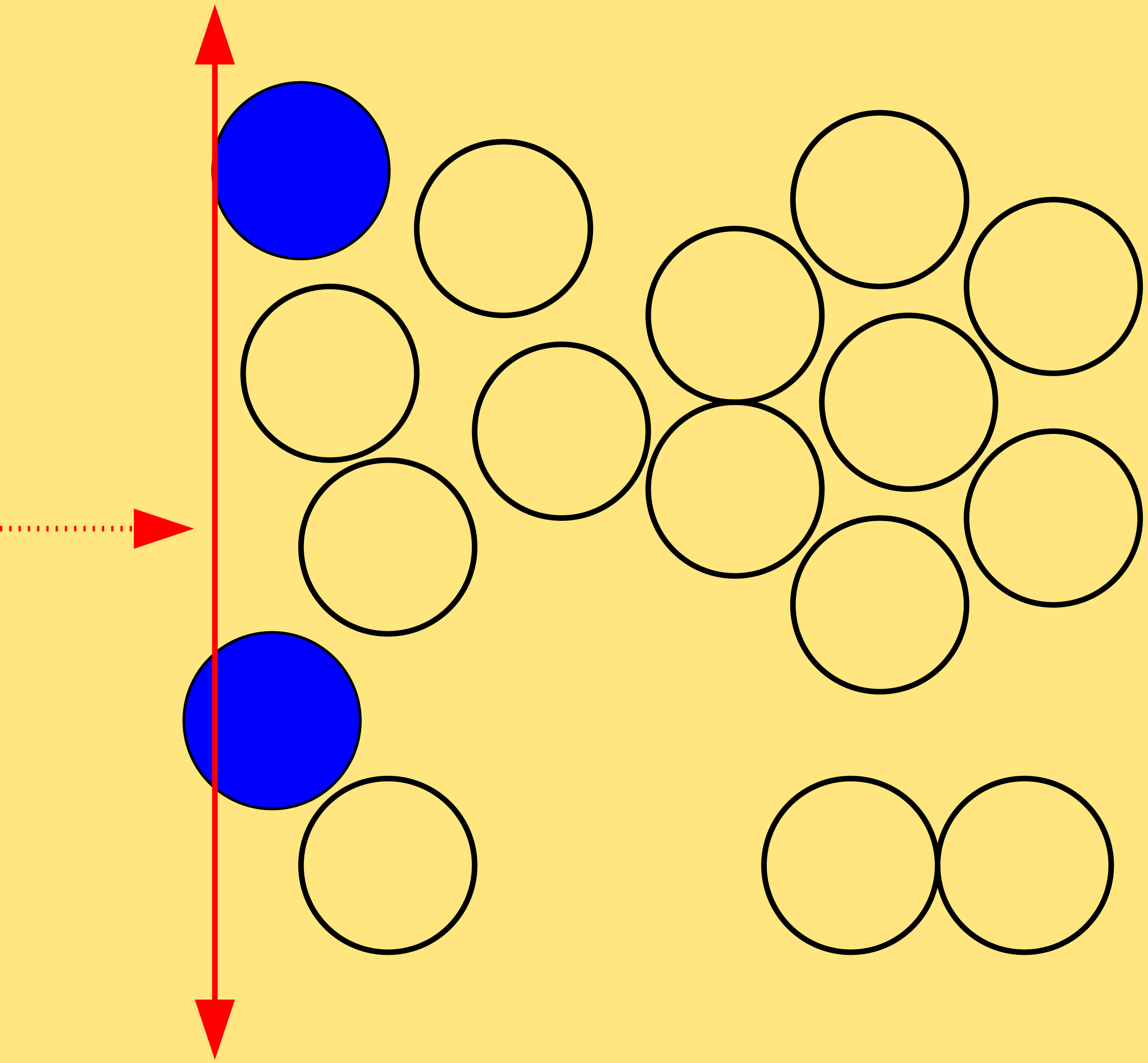
Motivation

Sweep-line among unit-size disjoint disks



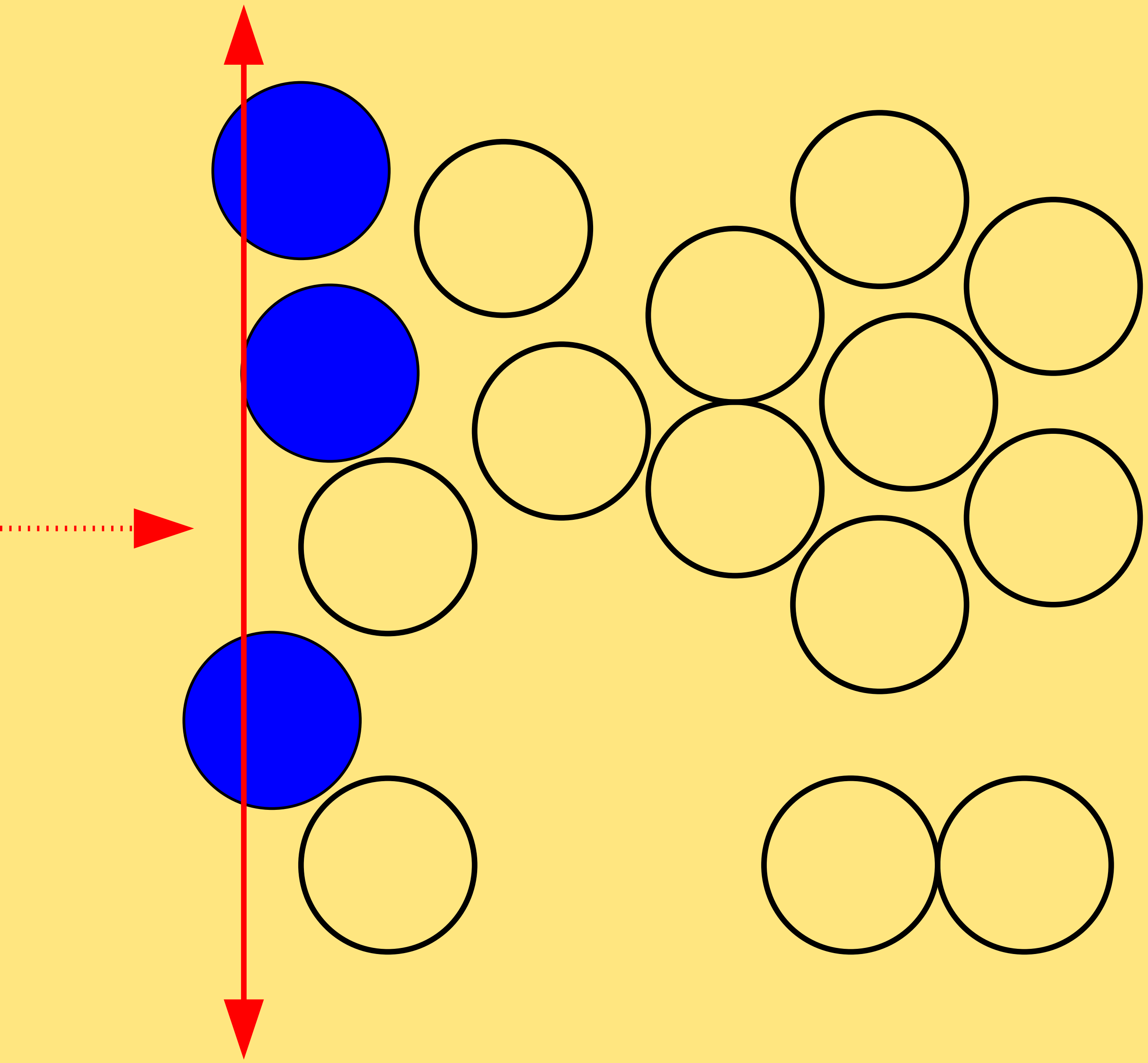
Motivation

Sweep-line among unit-size disjoint disks



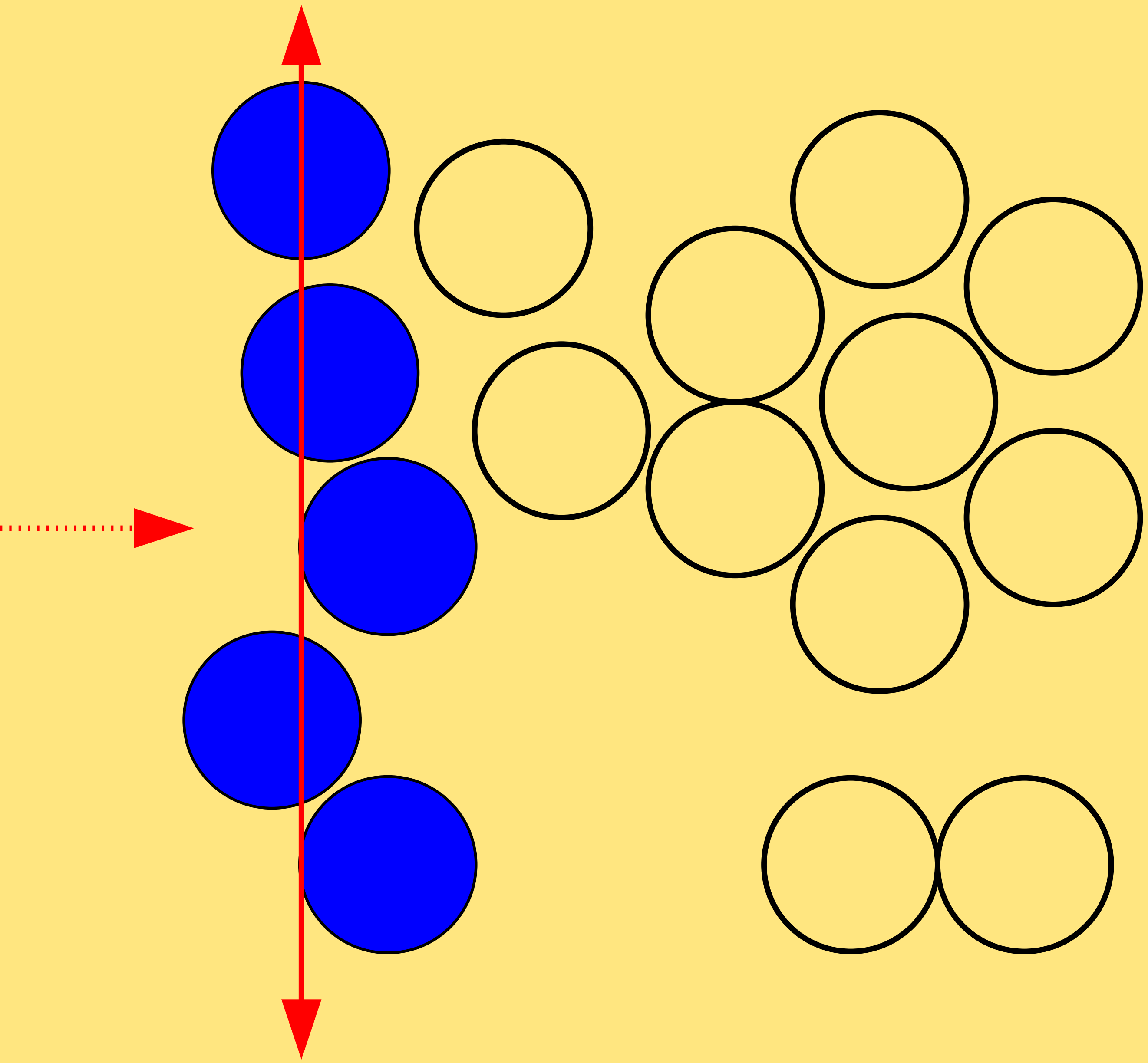
Motivation

Sweep-line among unit-size disjoint disks



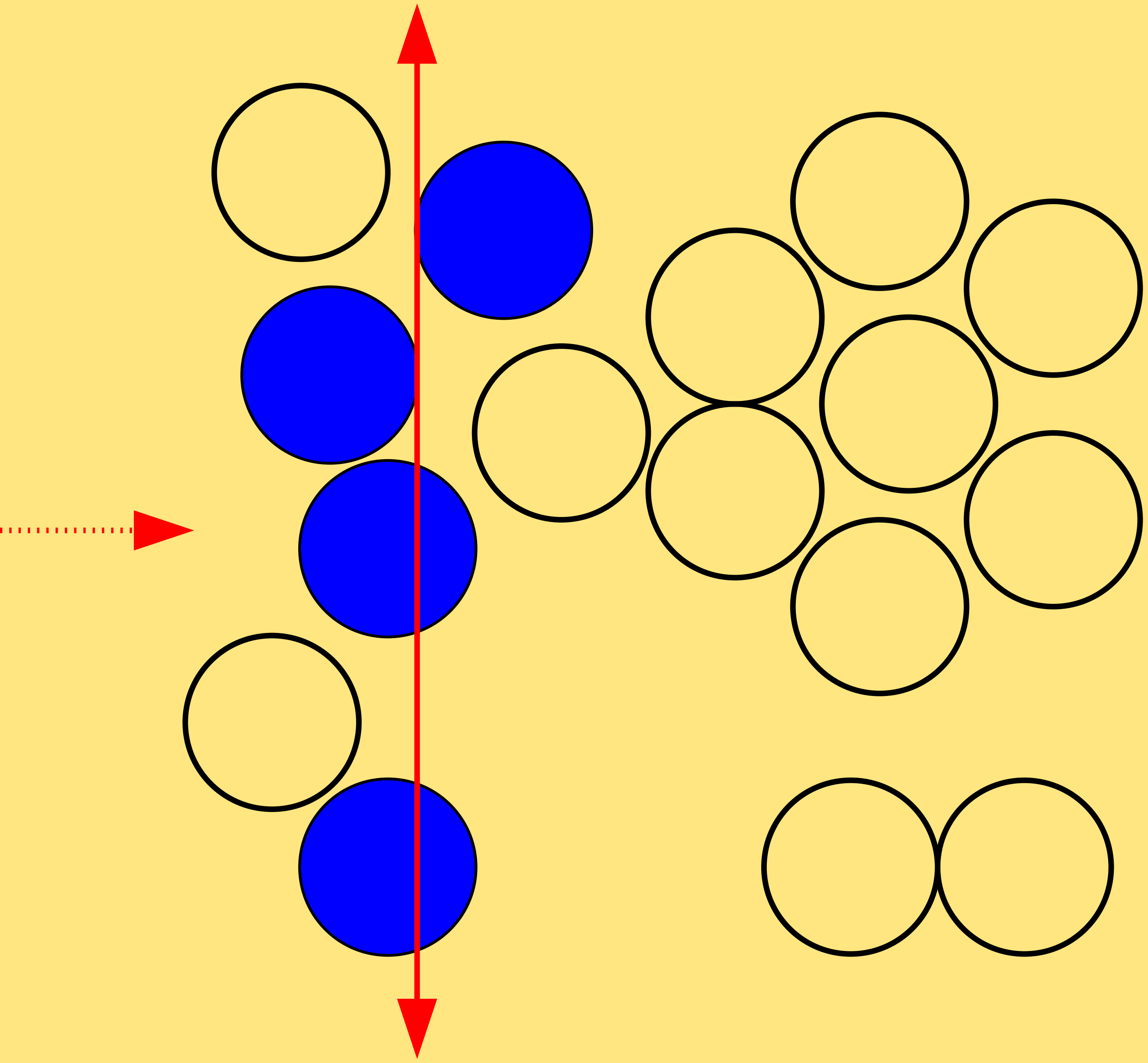
Motivation

Sweep-line among unit-size disjoint disks



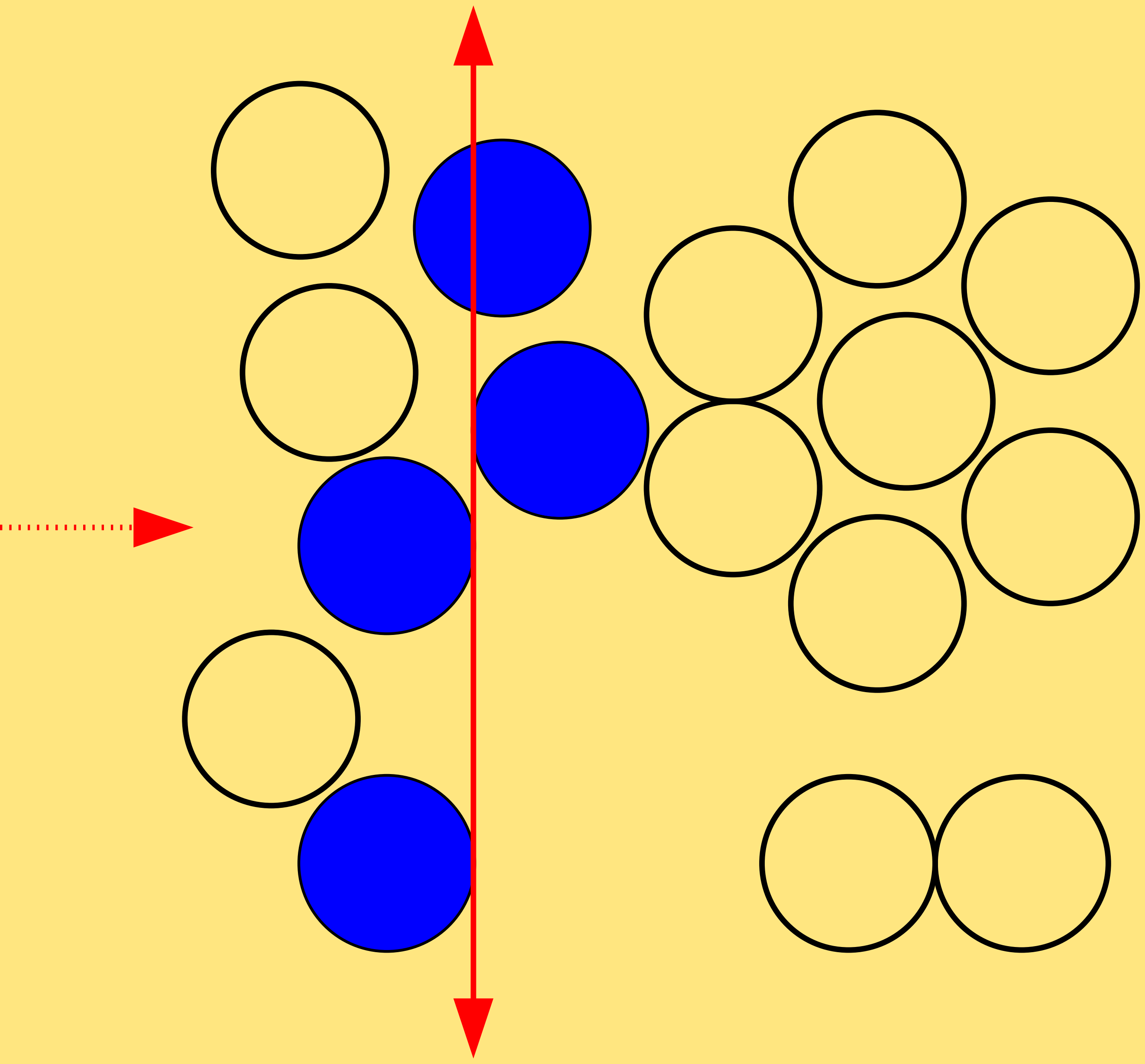
Motivation

Sweep-line among unit-size disjoint disks



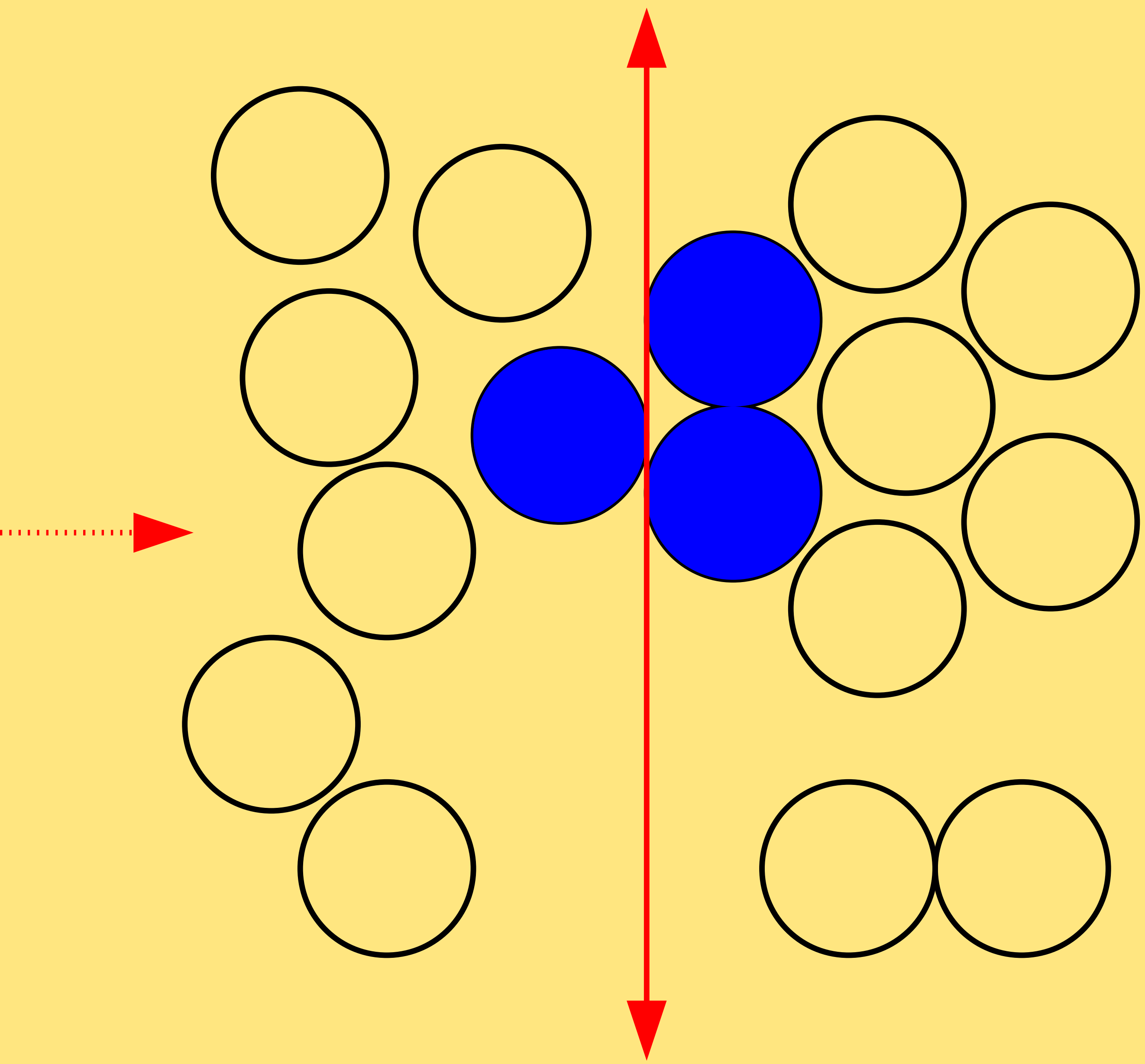
Motivation

Sweep-line among unit-size disjoint disks



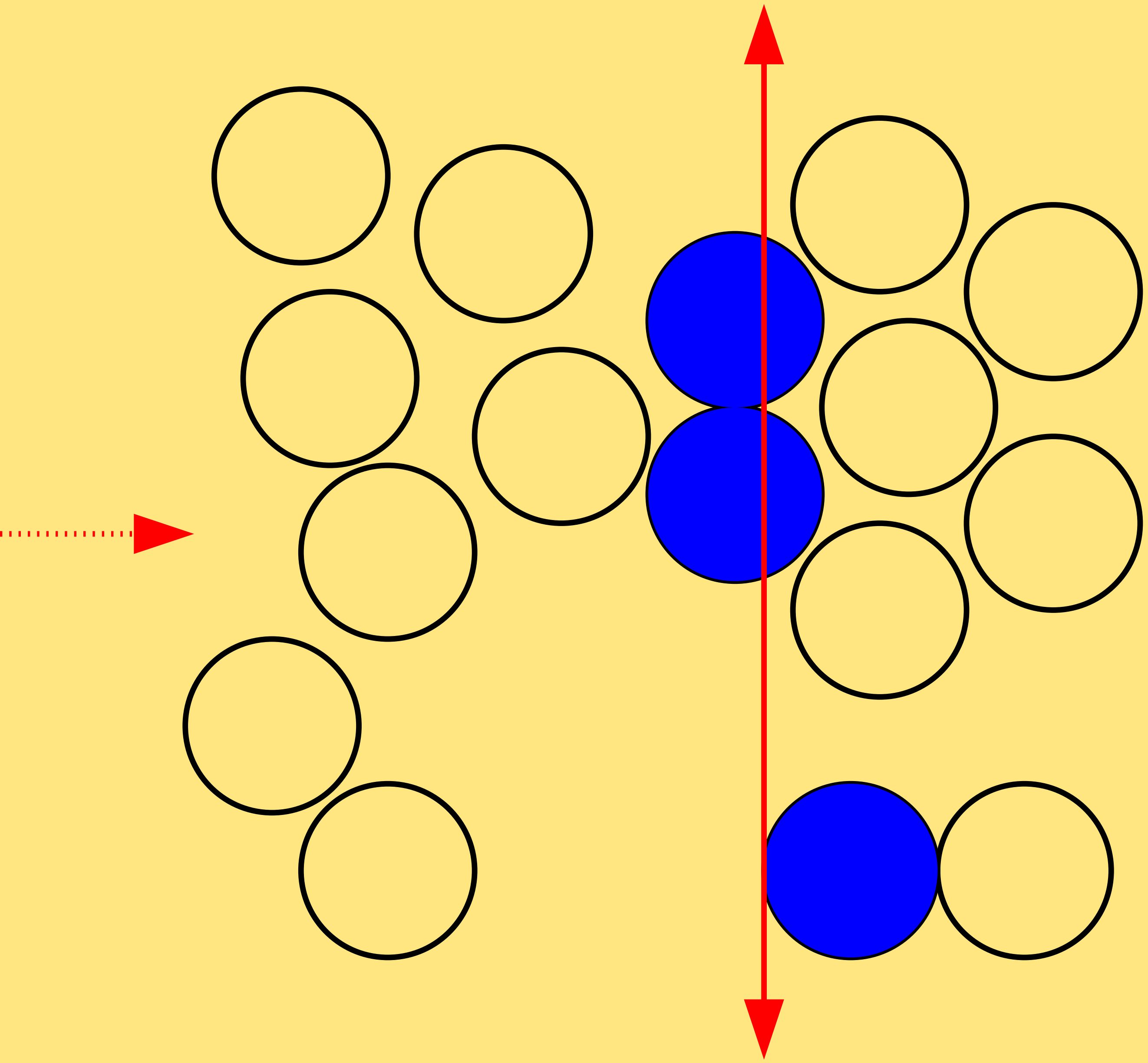
Motivation

Sweep-line among unit-size disjoint disks



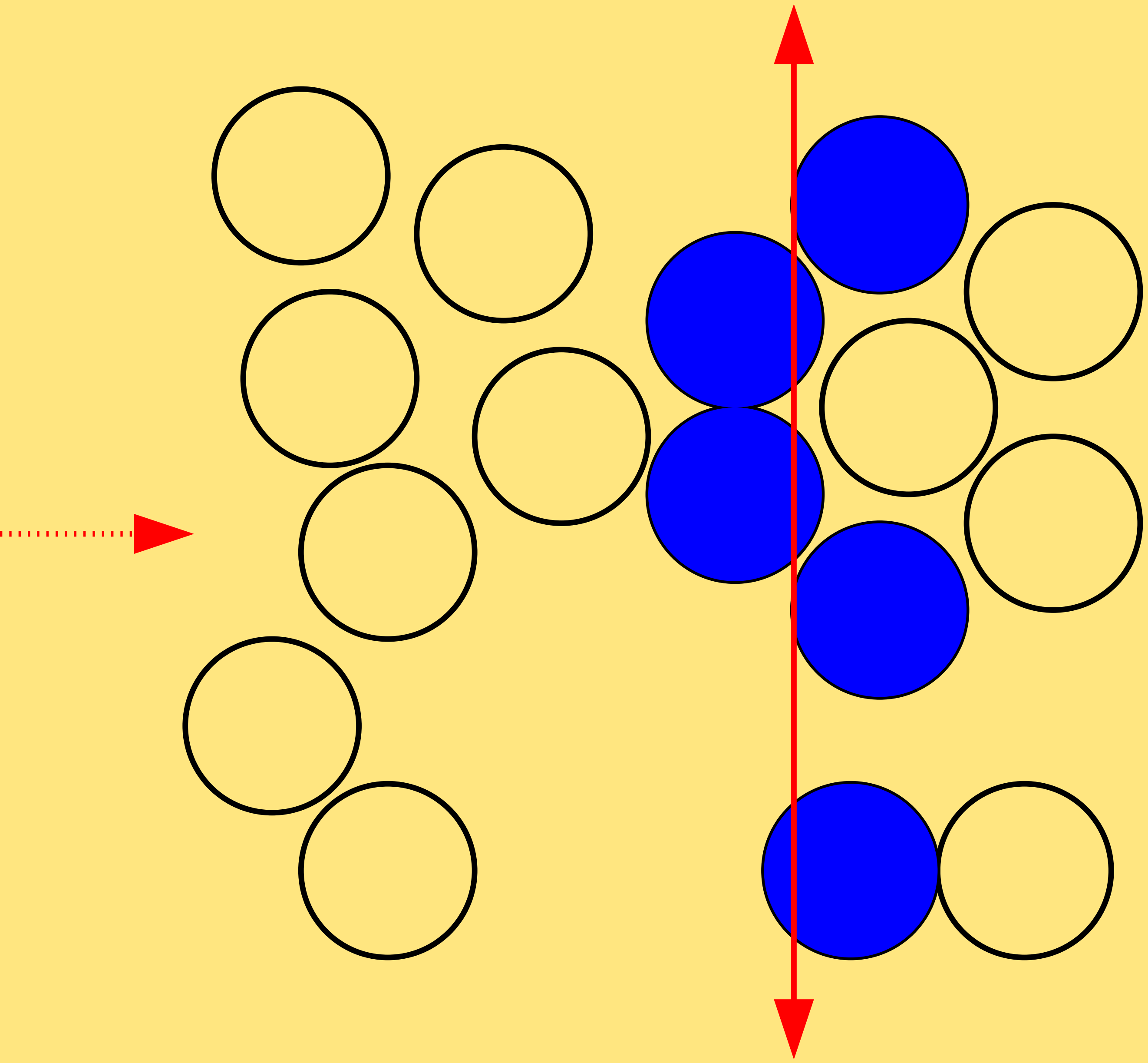
Motivation

Sweep-line among unit-size disjoint disks



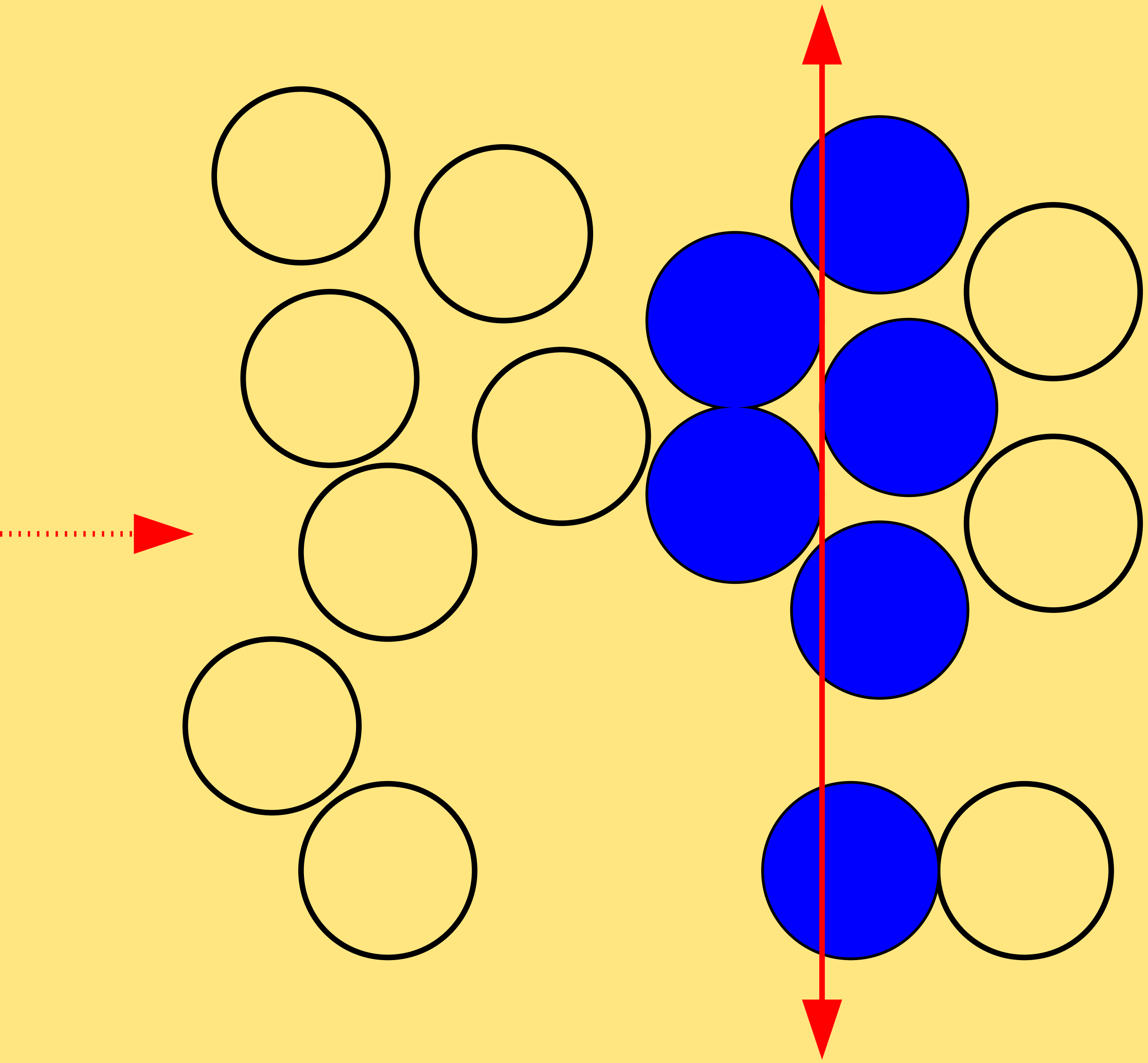
Motivation

Sweep-line among unit-size disjoint disks



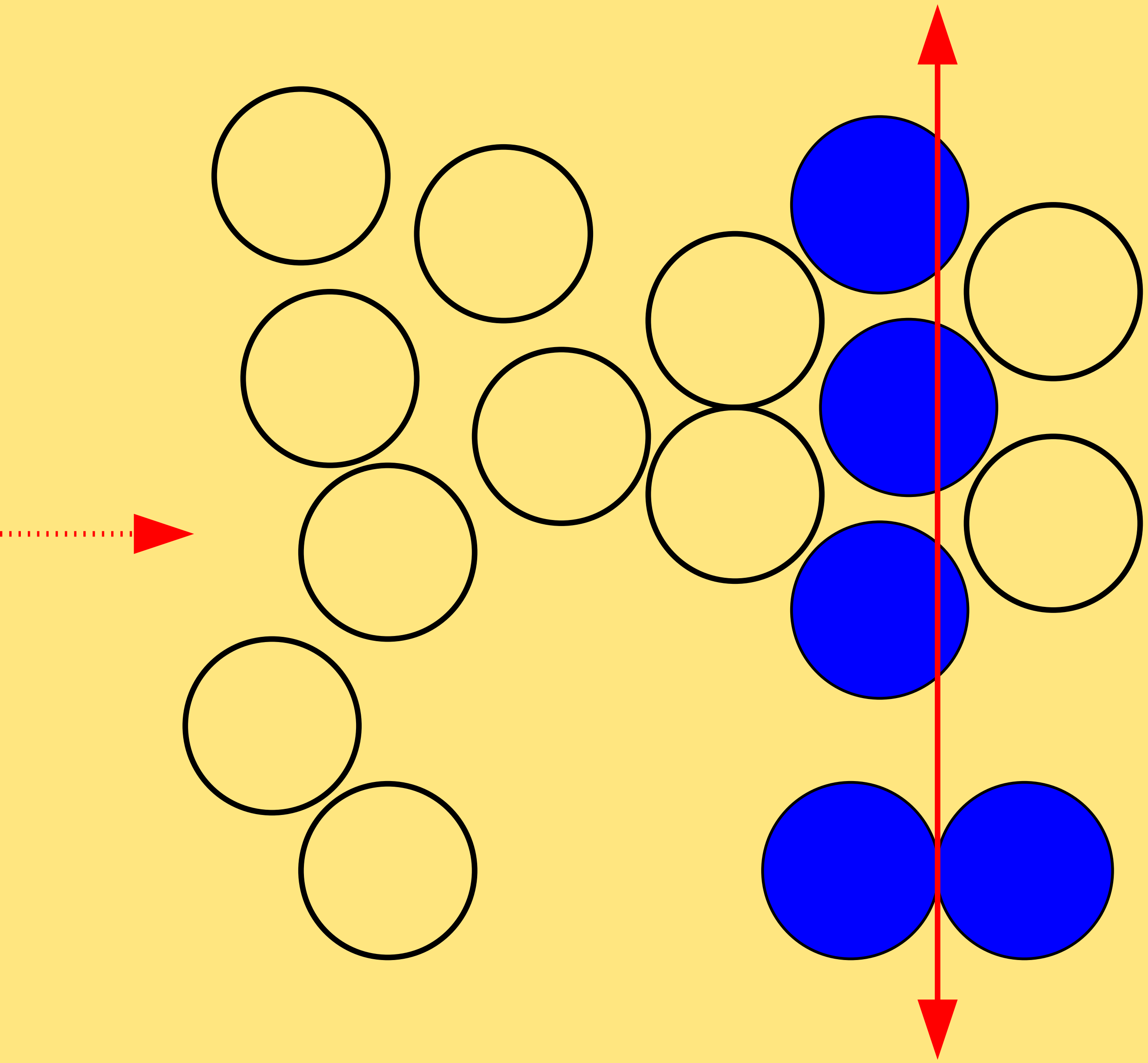
Motivation

Sweep-line among unit-size disjoint disks



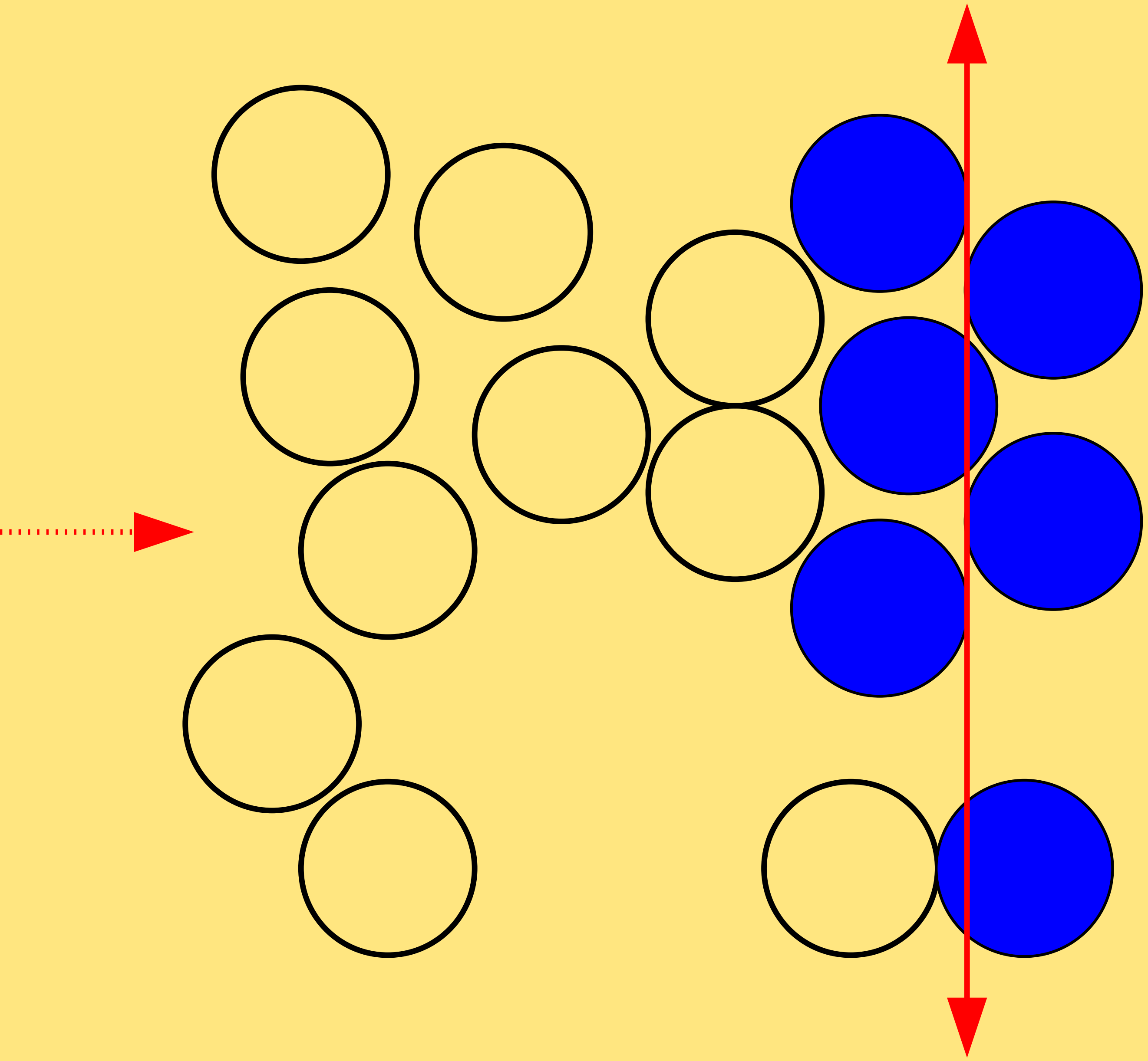
Motivation

Sweep-line among unit-size disjoint disks



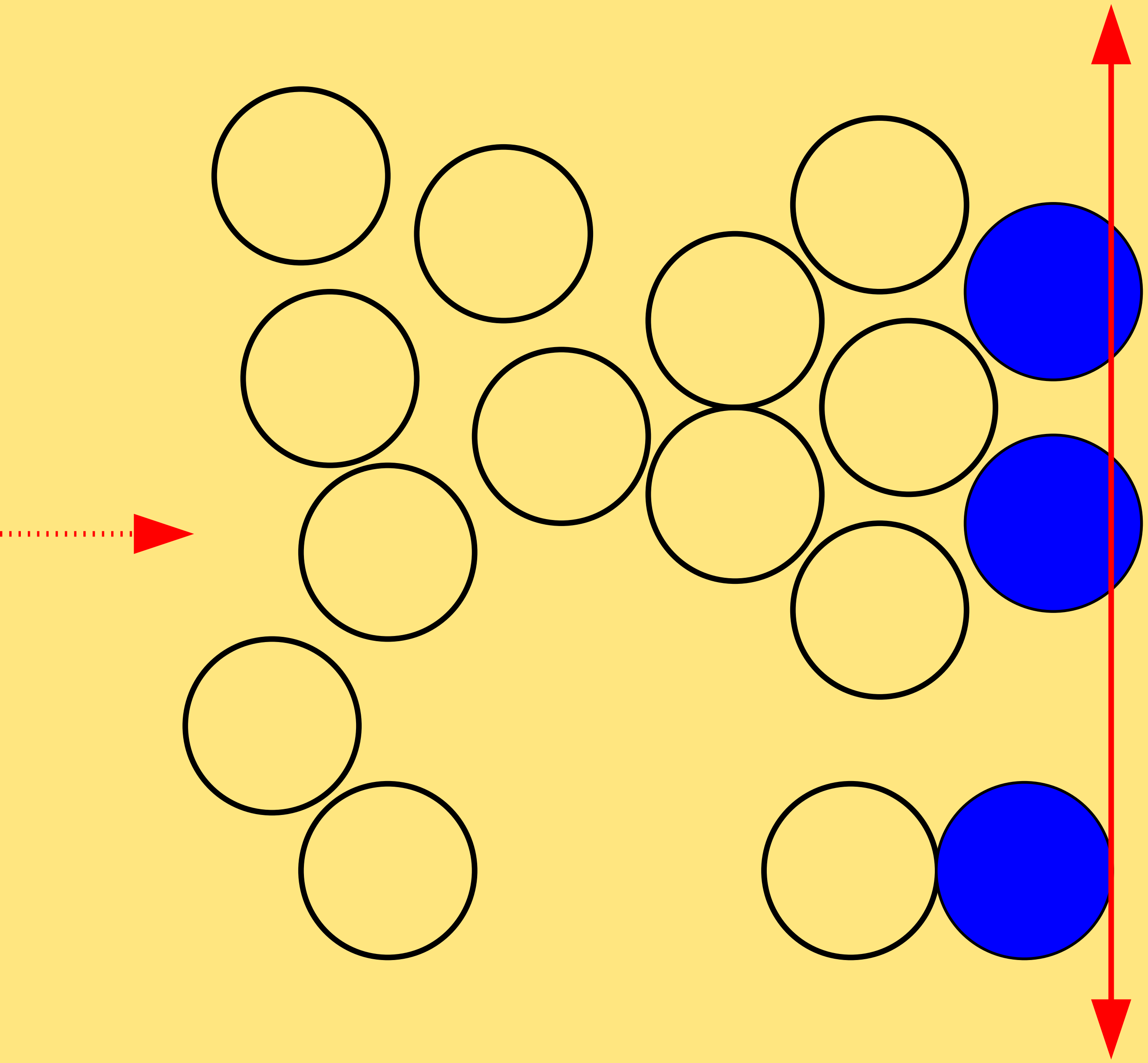
Motivation

Sweep-line among unit-size disjoint disks



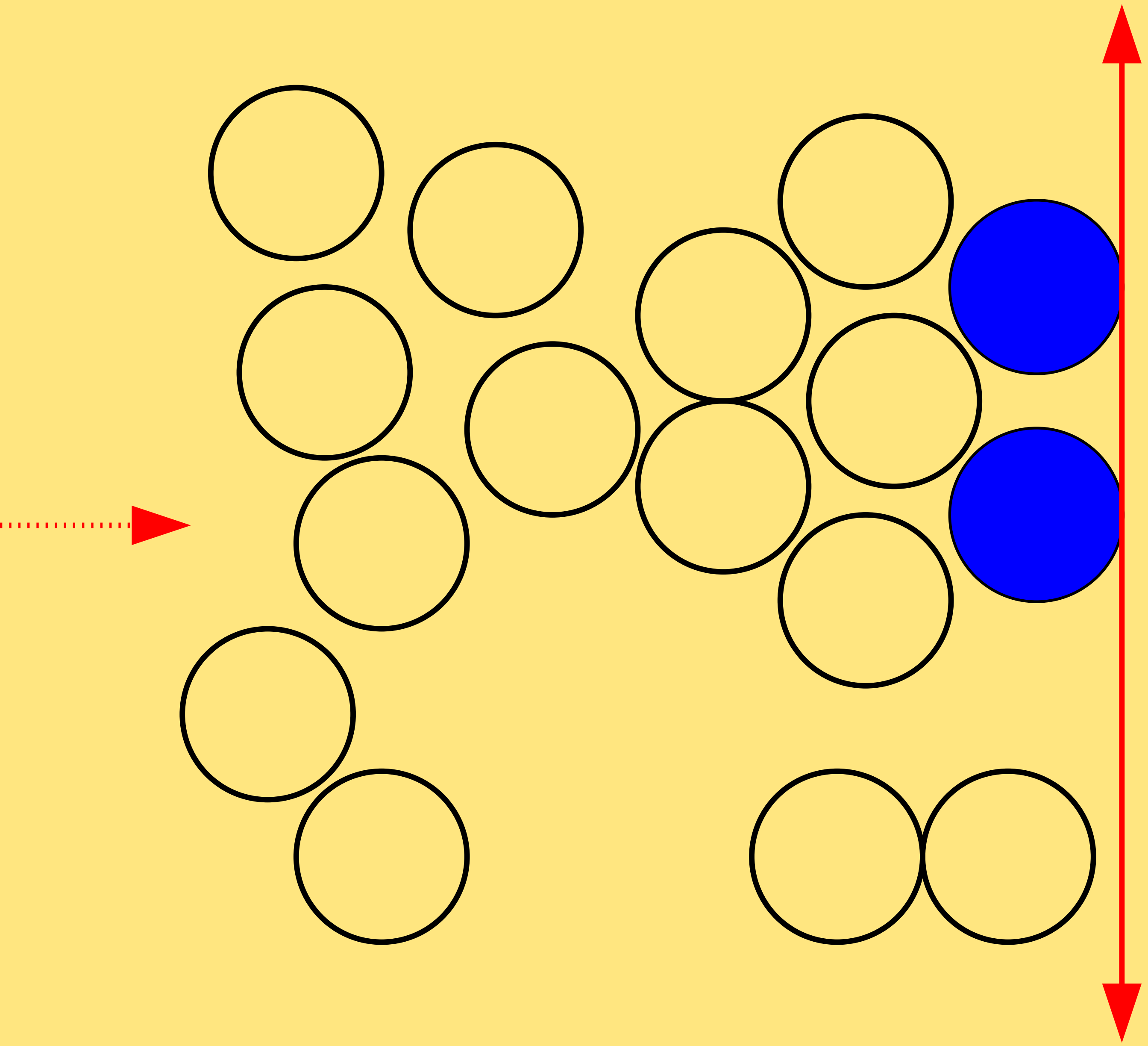
Motivation

Sweep-line among unit-size disjoint disks



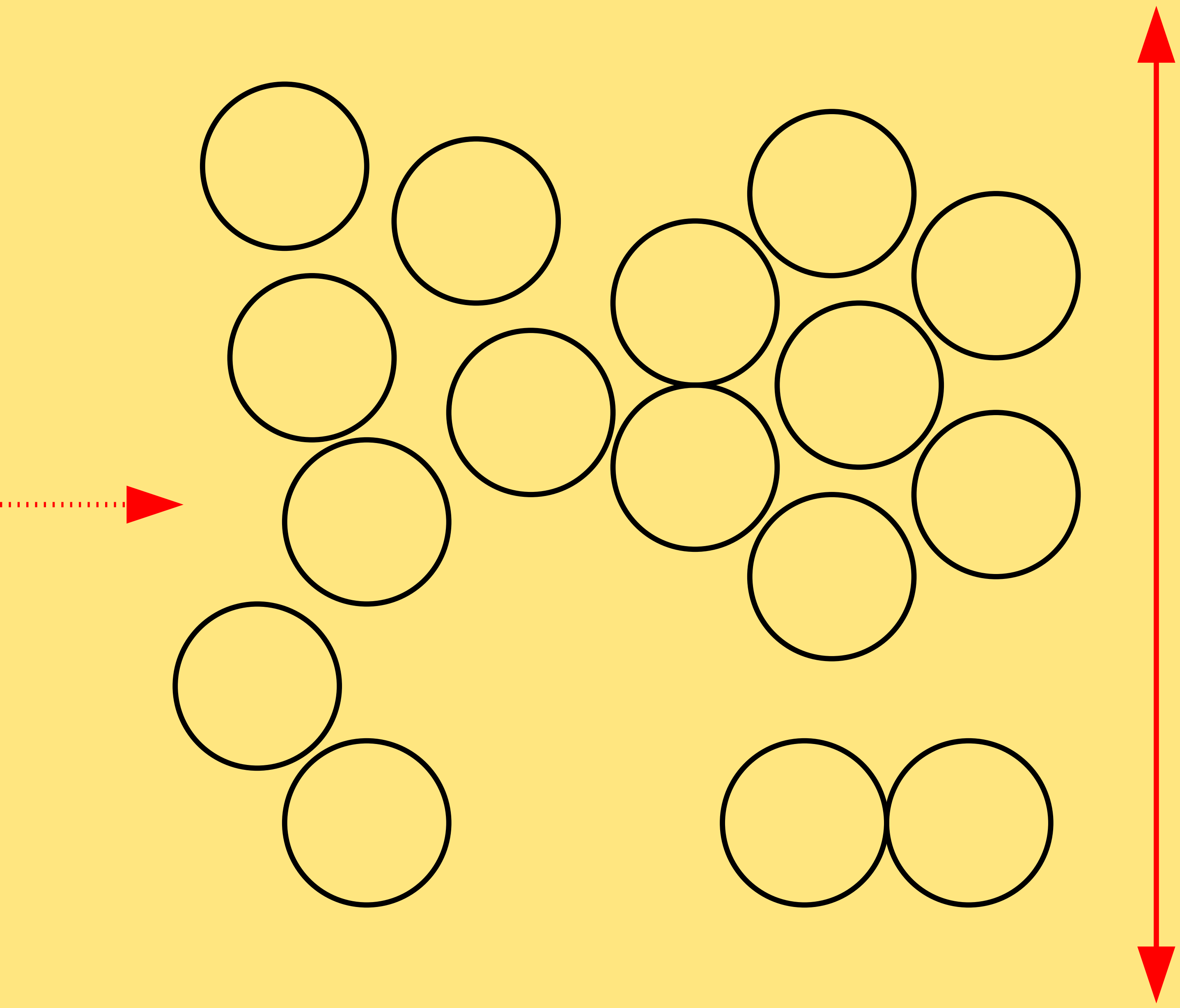
Motivation

Sweep-line among unit-size disjoint disks



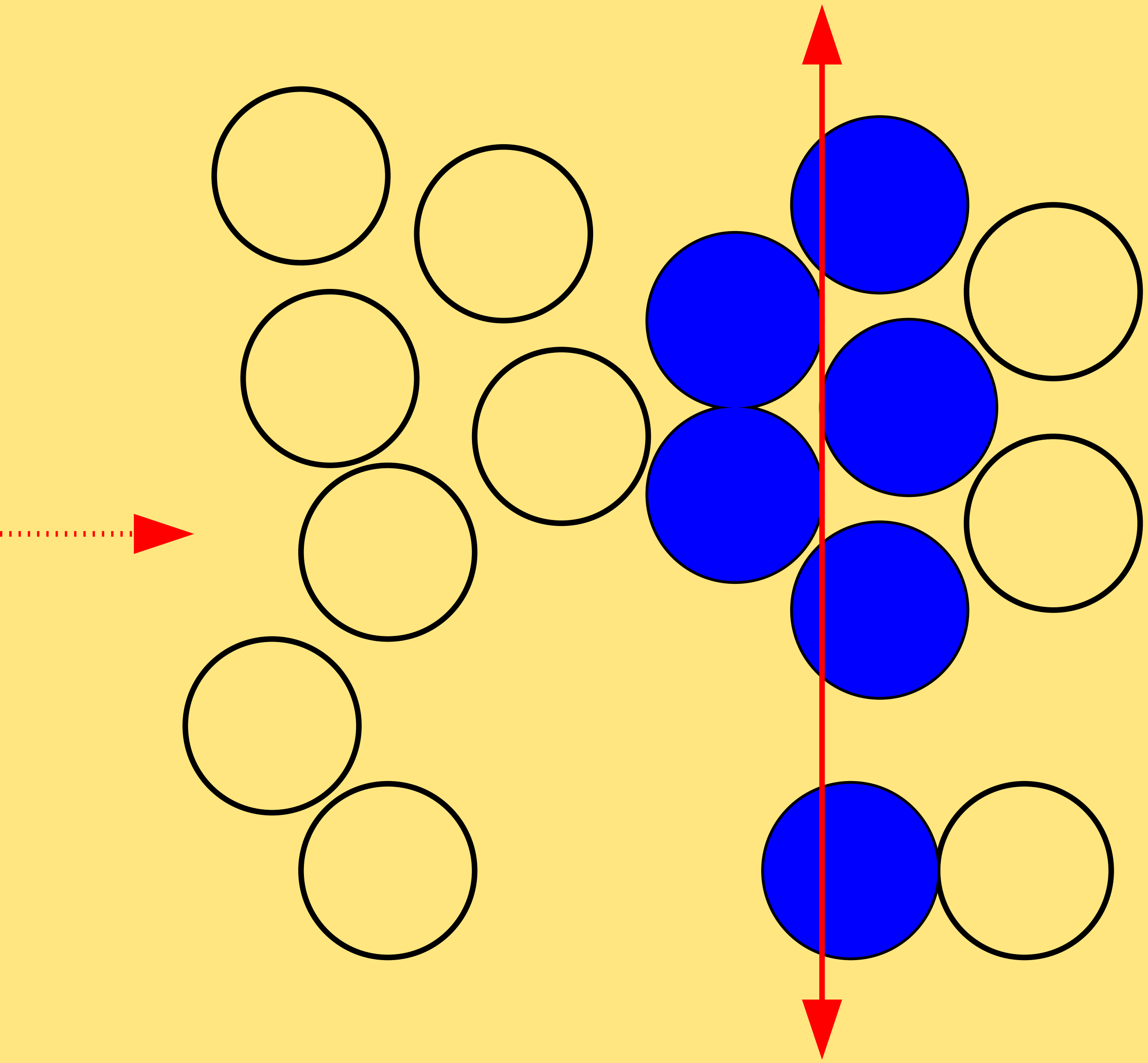
Motivation

Sweep-line among unit-size disjoint disks



Motivation

Sweep-line among unit-size disjoint disks

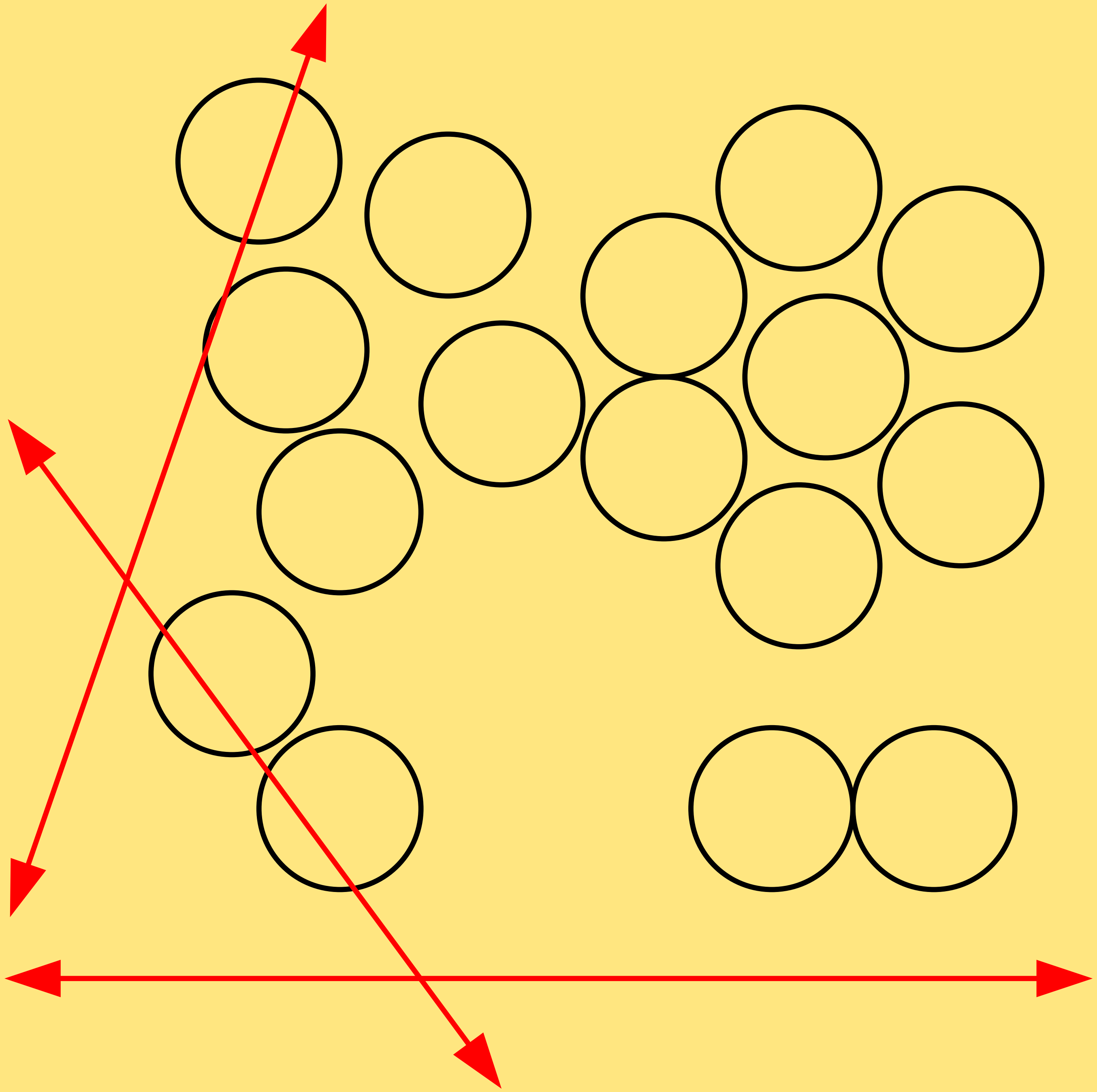


Often, running time depends on

$\sharp$ active disks

Motivation

Sweep-line among unit-size disjoint disks



Often, running time depends on

$\sharp$ active disks

Problem: choose a good sweeping direction

Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line

$\leadsto$ Project P onto L

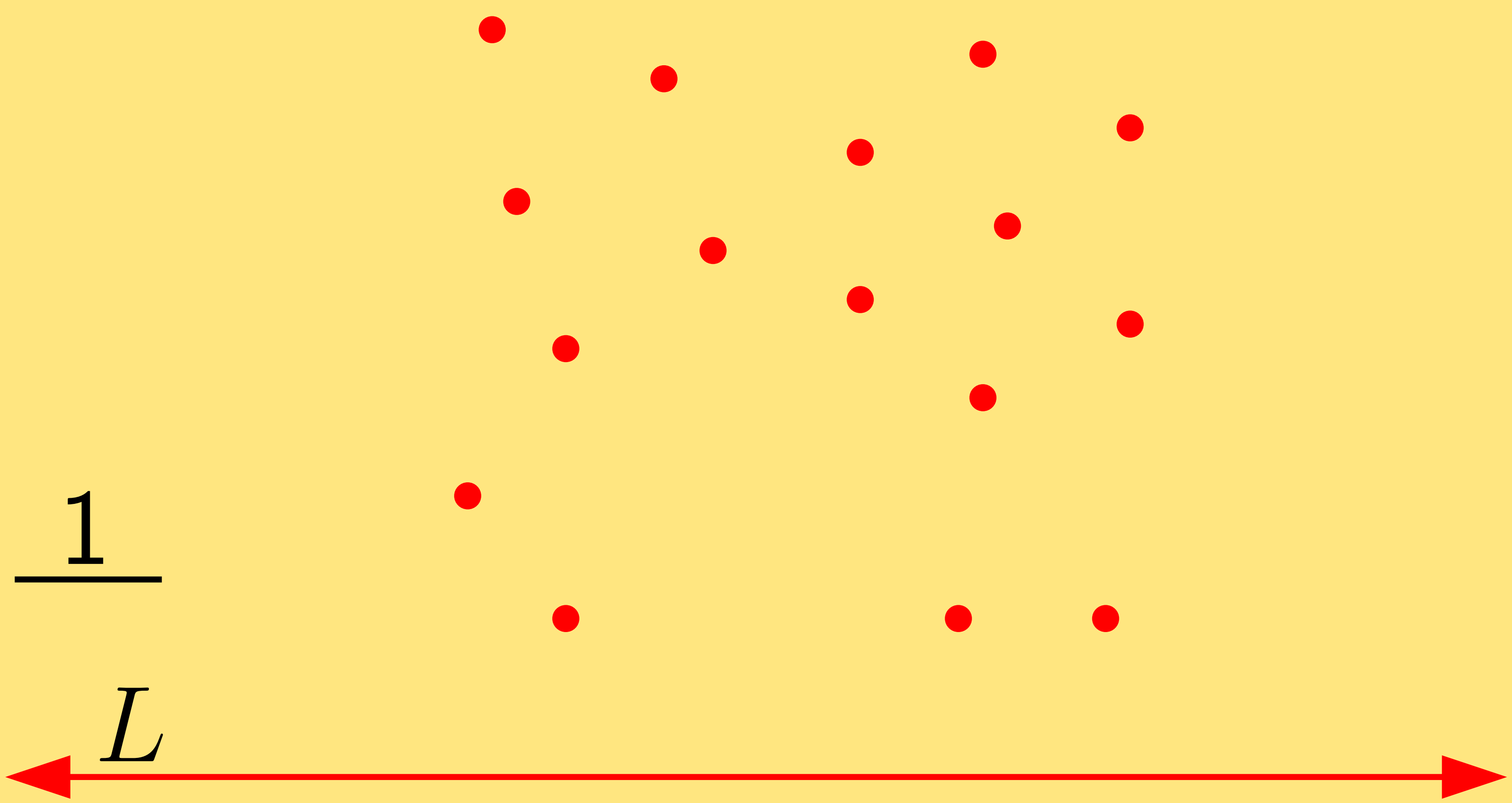
Maximum $\sharp$ points in a unit-length interval?

Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line

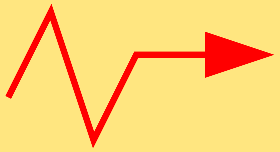
$\leadsto$ Project P onto L

Maximum $\sharp$ points in a unit-length interval?



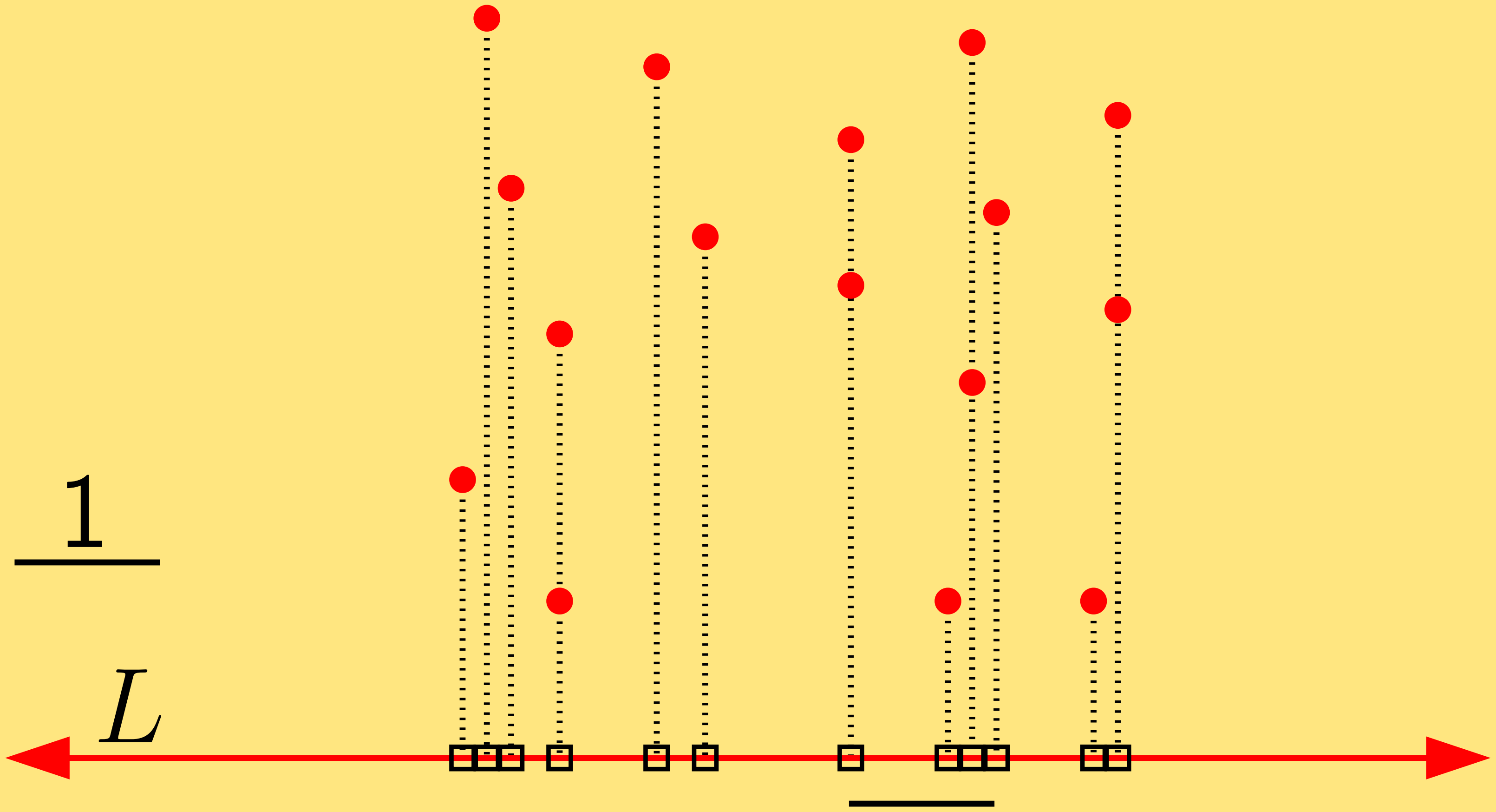
Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line



Project P onto L

Maximum $\sharp$ points in a unit-length interval?

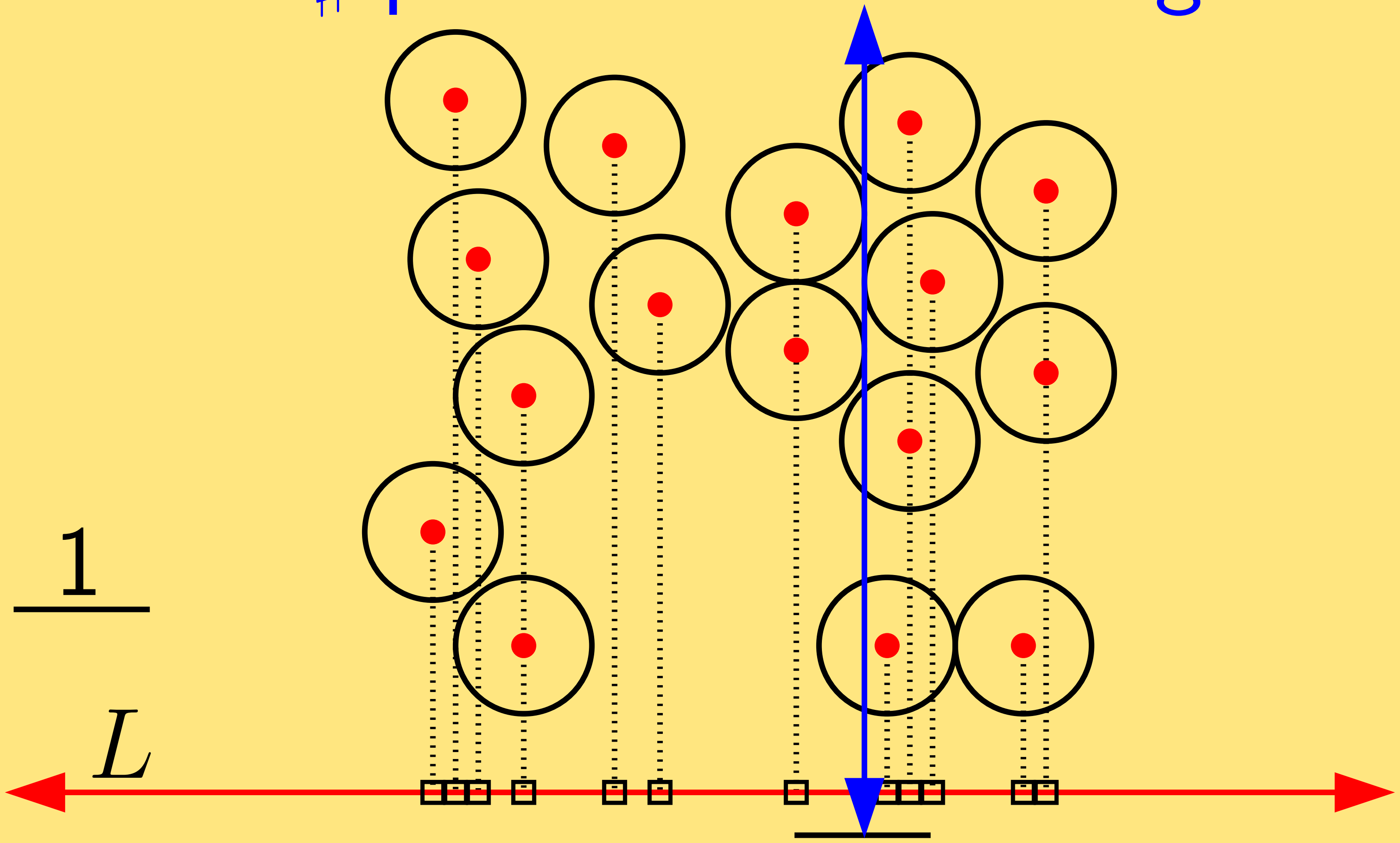


Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line

Project P onto L

Maximum $\sharp$ points in a unit-length interval?



Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line

$\rightsquigarrow$ Project P onto L

$C(L) =$ max $\sharp$ points in a unit-length interval

Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line

$\leadsto$ Project P onto L

$C(L) =$ max $\sharp$ points in a unit-length interval

Problem:

Choose a good $L/$ sweeping-line

Problem Definition

$P \subset \mathbb{R}^2$ 1-apart
L a line $\leadsto$ Project P onto L

$C(L) =$ max $\sharp$ points in a unit-length interval

Problem:

Choose a good L/ sweeping-line

$C(L)$ only depends on the slope α of L

$\Rightarrow \ C(\alpha) := C(L(\alpha))$

Previous Results

$P \subset \mathbb{R}^2$ 1-apart, $|P| = n$

- [Kučera, Mehlhorn, Preis, Schwarzenecker STACS'93]
  - ∗ Always a line L with $C(L) = O(\sqrt{n \log n})$
  - ∗ P with $C(L) = \Omega(\sqrt{n \log n})$ for all L (Besicovitch's sets)
  - ∗ Best L in polynomial time
- [Díaz, Hurtado, López, Sellarès, ISAAC'03]
  - ∗ Computing a 2-approx line in $O(n\Delta \log n\Delta)$ time ($\Delta \equiv$ diameter P, arbitrary P)

Our Results

$P \subset \mathbb{R}^2$ 1-apart, $|P| = n$

Choose a random line $L(\alpha)$, $\alpha \sim U[0, \pi]$

How good or bad is $C(\alpha)$?

What is the expected value of $C(\alpha)$?

Our Results

$P \subset \mathbb{R}^2$ 1-apart, $|P| = n$

Choose a random line $L(\alpha)$, $\alpha \sim U[0, \pi]$

How good or bad is $C(\alpha)$?

What is the expected value of $C(\alpha)$?

Theorem: $\mathbb{E}[C(\alpha)] = O(n^{2/3})$

Our Results

$P \subset \mathbb{R}^2$ 1-apart, $|P| = n$

Choose a random line $L(\alpha)$, $\alpha \sim U[0, \pi]$

How good or bad is $C(\alpha)$?

What is the expected value of $C(\alpha)$?

Theorem: $\mathbb{E}[C(\alpha)] = O(n^{2/3})$

Problem: We only know $\mathbb{E}[C(\alpha)] = \Omega(\sqrt{n \log n})$

Obs: If each point has $\leq k$ points at distance ≤ 1

$$\Rightarrow \quad \mathbb{E}[C(\alpha)] = O(kn^{2/3})$$

Alg. Obs: We can take or discard $L(\alpha)$ in $O(n \log n)$ time.

Our Approach

$P = \{p_1, \ldots, p_n\}$

$X_i(\alpha) :=$ $\sharp$ points at distance ≤ 1 from p_i

$X_{max}(\alpha) := \max\{X_1(\alpha), \ldots, X_n(\alpha)\}$

$$\Rightarrow \quad C(\alpha) \leq X_{max}(\alpha) \leq 2C(\alpha)$$

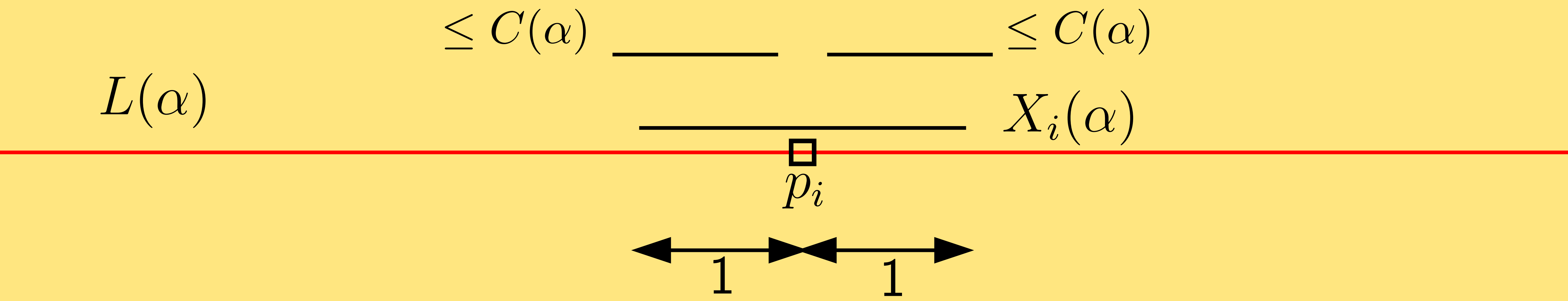
Our Approach

$P = \{p_1, \ldots, p_n\}$

$X_i(\alpha) :=$ ♯ points at distance ≤ 1 from p_i

$X_{max}(\alpha) := \max\{X_1(\alpha), \ldots, X_n(\alpha)\}$

$$\Rightarrow \quad C(\alpha) \leq X_{max}(\alpha) \leq 2C(\alpha)$$



Our Approach

$P = \{p_1, \ldots, p_n\}$

$X_i(\alpha) :=$ $\sharp$ points at distance ≤ 1 from p_i

$X_{max}(\alpha) := \max\{X_1(\alpha), \ldots, X_n(\alpha)\}$

$$\Rightarrow \quad C(\alpha) \leq X_{max}(\alpha) \leq 2C(\alpha)$$

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$

A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$

Proof: Assume without loss of generality that $d_{i,j} \geq d_{i,k}$ whenever $j > k$; that is, the points are indexed according to their distance from p_i. Like above, we assume that the line $L(\alpha)$ passes through p_i. We have

$$\mathbb{E}[X_i^2] = \mathbb{E}\left[\sum_{j,k\in[n]} X_{i,j}X_{i,k}\right] \leq \mathbb{E}\left[2\sum_j \sum_{k\leq j} X_{i,j}X_{i,k}\right] = 2\sum_j \mathbb{E}\left[X_{i,j}\sum_{k\leq j} X_{i,k}\right]$$

We claim that $\mathbb{E}\left[X_{i,j}\sum_{k\leq j} X_{i,k}\right] = O(1)$, and so the result follows. To prove the claim, observe that if $X_{i,j}(\alpha) = 1$, then all the points p_k that have $X_{i,k}(\alpha) = 1$ need to be in the strip (or slab) of width two having $L(\alpha + \pi/2)$ as axis. Because of a packing argument, in this strip there are $O(d_{i,j})$ points p_k that satisfy $d_{i,j} \geq d_{i,k}$. Therefore, by the way we indexed the points, we conclude that, if $X_{i,j}(\alpha) = 1$, then $\left(\sum_{k\leq j} X_{i,k}\right)(\alpha) = O(d_{i,j})$. In any case, we always have $\left(X_{i,j}\sum_{k\leq j} X_{i,k}\right)(\alpha) = O(d_{i,j})$. Therefore

$$\mathbb{E}\left[X_{i,j}\sum_{k\leq j} X_{i,k}\right] = \sum_{t=1}^{n} t\cdot \Pr\left[X_{i,j}\sum_{k\leq j} X_{i,k} = t\right] \leq \sum_{t=1}^{n} O(d_{i,j})\cdot \Pr\left[X_{i,j}\sum_{k\leq j} X_{i,k} = t\right]$$

$$= O(d_{i,j})\sum_{t=1}^{n} \Pr\left[X_{i,j}\sum_{k\leq j} X_{i,k} = t\right] \leq O(d_{i,j})\cdot \Pr\left[X_{i,j} = 1\right] = O(d_{i,j})\frac{2\arcsin 1/d_{i,j}}{\pi} = O(1).$$

A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$ Corollary: $\mathbb{E}[T(\alpha)] = O(n^2)$

Main idea: If $X_{max}(\alpha) \geq t$, then $T(\alpha) \geq t^3/8$

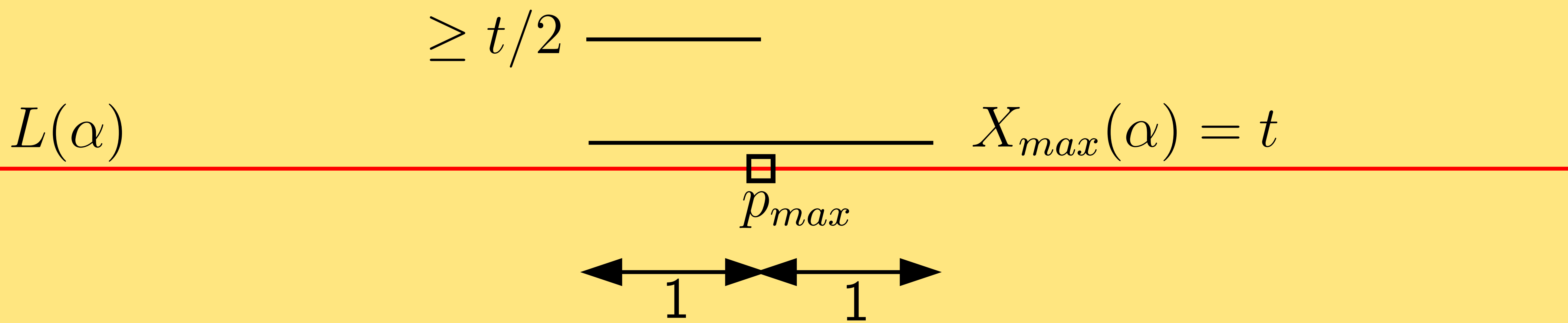
A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$ Corollary: $\mathbb{E}[T(\alpha)] = O(n^2)$

Main idea: If $X_{max}(\alpha) \geq t$, then $T(\alpha) \geq t^3/8$



A couple of ideas

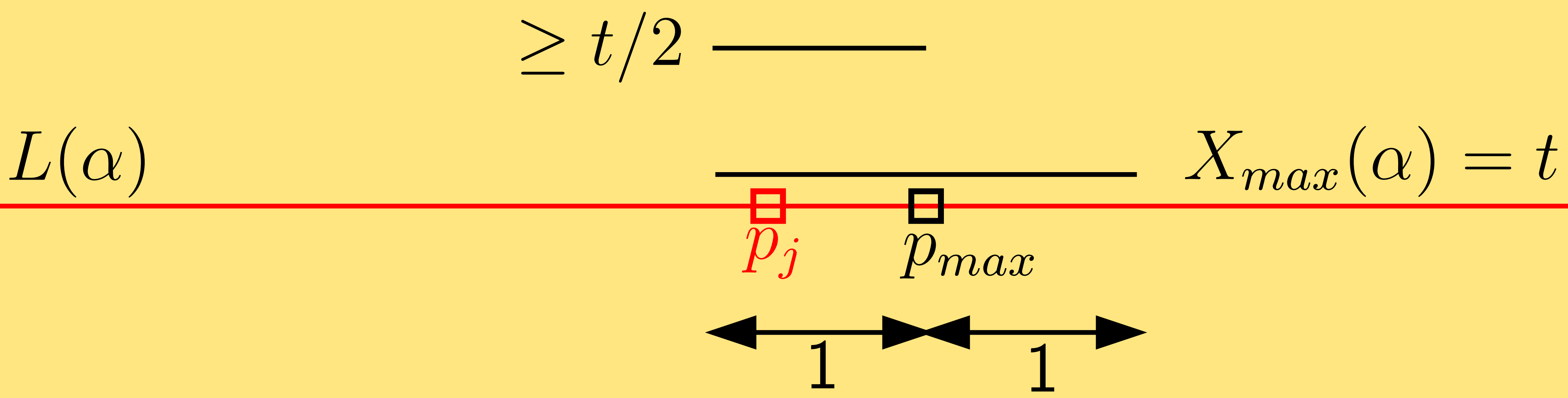
Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$ Corollary: $\mathbb{E}[T(\alpha)] = O(n^2)$

Main idea: If $X_{max}(\alpha) \geq t$, then $T(\alpha) \geq t^3/8$

$$X_j(\alpha) \geq t/2 \quad \Rightarrow \quad X_j^2(\alpha) \geq t^2/4$$



A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$ Corollary: $\mathbb{E}[T(\alpha)] = O(n^2)$

Main idea: If $X_{max}(\alpha) \geq t$, then $T(\alpha) \geq t^3/8$

A couple of ideas

Task: Bound $\mathbb{E}[X_{max}(\alpha)]$

$T(\alpha) := \sum_i X_i^2(\alpha)$.

Lemma: $\mathbb{E}[X_i^2(\alpha)] = O(n)$ Corollary: $\mathbb{E}[T(\alpha)] = O(n^2)$

Main idea: If $X_{max}(\alpha) \geq t$, then $T(\alpha) \geq t^3/8$

$$\mathbb{E}[X_{max}] = \sum_{t=1}^{n} \Pr[X_{max} \geq t] = \sum_{t=1}^{n^{2/3}} \Pr[X_{max} \geq t] + \sum_{t=n^{2/3}+1}^{n} \Pr[X_{max} \geq t]$$

$$\leq \sum_{t=1}^{n^{2/3}} 1 + \sum_{t=n^{2/3}+1}^{n} \Pr[T \geq t^3/8] \leq n^{2/3} + \sum_{t=n^{2/3}+1}^{n} \frac{O(n^2)}{t^3}$$

$$\leq n^{2/3} + O(n^2) \int_{n^{2/3}}^{n} \frac{1}{t^3} dt \leq n^{2/3} + O(n^2) \left(\frac{2}{(n^{2/3})^2} - \frac{2}{n^2} \right) = O(n^{2/3}).$$

Summary

- Projection of non-dense point sets onto lines.
- Relation to sweep-line algorithms.
- Expensive: choose a best sweeping-line.
- Cheap: choose a random line?
- How good a random line is? $O(n^{2/3})$.

Summary

- Projection of non-dense point sets onto lines.
- Relation to sweep-line algorithms.
- Expensive: choose a best sweeping-line.
- Cheap: choose a random line?
- How good a random line is? $O(n^{2/3})$.

Close the gap!!

Summary

- Projection of non-dense point sets onto lines.
- Relation to sweep-line algorithms.
- Expensive: choose a best sweeping-line.
- Cheap: choose a random line?
- How good a random line is? $O(n^{2/3})$.

Close the gap!!

THANKS!

Sergio Cabello, IMFM, cabello@imfm.uni-lj.si