# Continuous Inverse Ranking Queries in Uncertain Streams

Thomas Bernecker*, Hans-Peter Kriegel*,
Nikos Mamoulis**, Matthias Renz* and Andreas Zuefle*

*)
Ludwig-Maximilians-Universität München (LMU)
Munich, Germany
http://www.dbs.ifi.lmu.de
{bernecker, kriegel, renz, zuefle}@dbs.ifi.lmu.de

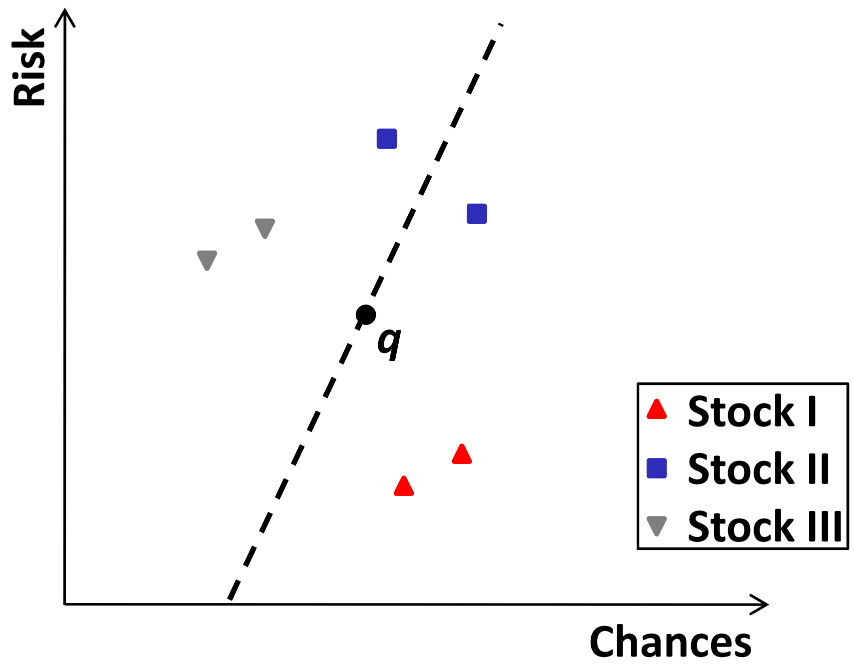
**)
University of Hong Kong (HKU)
Hong Kong
http://www.cs.hku.hk
nikos@cs.hku.hk

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF
COMPUTER SCIENCE

# Outline

# Probabilistic Inverse Ranking

- Identification of the significance of objects among peers
  - *Inverse Ranking:* Return the position of the query object ***q*** w.r.t. the score function ***S***
  - *Probabilistic Inverse Ranking:* Find all possible positions of ***q***
- Example: Stock rating system




**S = Chances - Risk**

***q*** →
- Rank **1**? → 0 %
- Rank **2**? → 50 %
- Rank **3**? → 50 %
- Rank **4**? → 0 %

# Probabilistic Inverse Ranking



- Probabilistic Inverse Ranking (PIR) Query
  - Probabilistic database ***DB*** where |***DB***| = ***n***
  - Uncertain object ***o***: ***m*** alternative locations (discrete uncertainty) or pdf (continuous uncertainty)
  - Query object ***q***
  - Score function ***S*** : ***DB*** → $R_0^+$
  - Definition: ∀ ***i*** = 1, ..., ***k*** : P(***q*** is on rank ***i*** w.r.t. ***S*** ) = $P_q^t(i)$
  - There exist exactly ***i*** - 1 objects ***o*** ϵ ***DB*** with ***S(o)*** > ***S(q)***

- Challenge: Application to dynamic data
  - General stream model with location updates retrieved at a time ***t***
  - P(***q*** is on rank ***i*** at time ***t***) = $P_q^t(i)$
  - Initial computation
  - Incremental processing

# Initial Computation (1)

- Initial time ***t***: Compute $P_q^t(i)$ $\forall$ ***i*** = 1,…,***k***
  - Object ***o*** ϵ ***DB*** : $p_o^t$ = P(***S*(*o*)** > ***S*(*q*)** at time ***t***)
  - ***j*** objects have been processed so far (**o*j*** is the latest)
  - Successive processing by the *Poisson Binomial Recurrence (PBR)*:

$$P_{i,j}^t = \begin{cases} 1 & \text{if } i = 0 \wedge j = 0 \\ 0 & \text{if } i < 0 \vee i > j \\ P_{i-1,j-1}^t \cdot p_{o_j}^t + P_{i,j-1}^t \cdot \left(1 - p_{o_j}^t\right) & \text{else} \end{cases}$$

***i*** out of ***j***:
***S*(*o*)** > ***S*(*q*)**

***i-1*** out of ***j-1***:
***S*(*o*)** > ***S*(*q*)**
and
***S*(*oj*)** > ***S*(*q*)**

***i*** out of ***j-1***:
***S*(*o*)** > ***S*(*q*)**
and
***S*(*oj*)** ≤ ***S*(*q*)**

# Initial Computation (2)



- ***j* = *n* (∀*i* = 0,...,*k-1*)**: $P_{i,j}^t = P_{i,n}^t = P_q^t(i+1)$

  ⇒ PIR result for ***q*** ⇒ runtime: O(***k·n***)

- Optimizations:
  - $p_o^t = 0$ ⇒ ***o*** has no effect on the rank of ***q***
  - $p_o^t = 1$ ⇒ increment counter $C^t$

- General case ($0 < p_o^t < 1$) ⇒ process ***o*** by PBR: ∀***i*** = 0,...,***k-1*** :

  P(***i*** objects processed by PBR have a higher score than ***q***) = $P_{PBR}^t(i)$

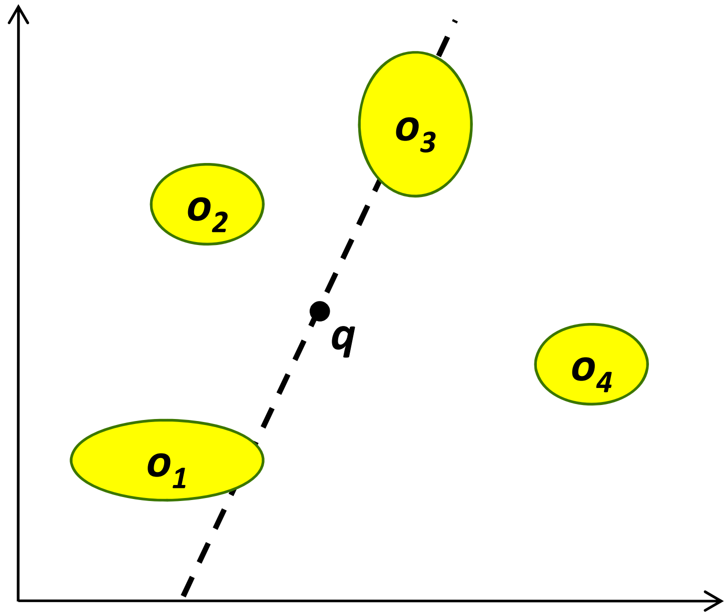
- Initial PIR result:

$$P_q^t(i) = \begin{cases} P_{PBR}^t\left(i-1-C^t\right) & \text{if } C^t+1 \le i \le C^t+1+k \\ 0 & \text{else} \end{cases}$$

# Initial Computation (3)

- Example: ***n*** = 4, ***k*** = 2

$p^t_{o_1} = 0.1 \quad p^t_{o_2} = 0 \quad p^t_{o_3} = 0.6 \quad p^t_{o_4} = 1 \quad\quad C^t = 0$



# Initial Computation (3)

DATABASE SYSTEMS GROUP

LMU

- Example: ***n*** = 4, ***k*** = 2, ***j*** = 1

$p_{o_1}^t = 0.1 \quad p_{o_2}^t = 0 \quad p_{o_3}^t = 0.6 \quad p_{o_4}^t = 1 \qquad C^t = 0$

- ***j*** = 1: $P_{0,1}^t = P_{-1,0}^t \cdot p_{o_1}^t + P_{0,0}^t \cdot \left(1 - p_{o_1}^t\right) = 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

$$P_{1,1}^t = P_{0,0}^t \cdot p_{o_1}^t + P_{1,0}^t \cdot \left(1 - p_{o_1}^t\right) = 1 \cdot 0.1 + 0 \cdot 0.9 = 0.1$$

# Initial Computation (3)



- Example: ***n*** = 4, ***k*** = 2, ***j*** = 2

$p_{o_1}^t = 0.1 \quad p_{o_2}^t = 0 \quad p_{o_3}^t = 0.6 \quad p_{o_4}^t = 1 \qquad C^t = 0$

- ***j*** = 1: $P_{0,1}^t = P_{-1,0}^t \cdot p_{o_1}^t + P_{0,0}^t \cdot \left(1 - p_{o_1}^t\right) = 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

$$P_{1,1}^t = P_{0,0}^t \cdot p_{o_1}^t + P_{1,0}^t \cdot \left(1 - p_{o_1}^t\right) = 1 \cdot 0.1 + 0 \cdot 0.9 = 0.1$$

# Initial Computation (3)



- Example: ***n*** = 4, ***k*** = 2, ***j*** = 3

$$p_{o_1}^t = 0.1 \qquad p_{o_2}^t = 0 \qquad p_{o_3}^t = 0.6 \qquad p_{o_4}^t = 1 \qquad C^t = 0$$

- ***j*** = 1: $P_{0,1}^t = P_{-1,0}^t \cdot p_{o_1}^t + P_{0,0}^t \cdot \left(1 - p_{o_1}^t\right) = 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

$$P_{1,1}^t = P_{0,0}^t \cdot p_{o_1}^t + P_{1,0}^t \cdot \left(1 - p_{o_1}^t\right) = 1 \cdot 0.1 + 0 \cdot 0.9 = 0.1$$

- ***j*** = 3: $P_{0,2}^t = P_{-1,1}^t \cdot p_{o_3}^t + P_{0,1}^t \cdot \left(1 - p_{o_3}^t\right) = 0 \cdot 0.6 + 0.9 \cdot 0.4 = 0.36$

$$P_{1,2}^t = P_{0,1}^t \cdot p_{o_3}^t + P_{1,1}^t \cdot \left(1 - p_{o_3}^t\right) = 0.9 \cdot 0.6 + 0.1 \cdot 0.4 = 0.58$$

# Initial Computation (3)



- Example: ***n*** = 4, ***k*** = 2, ***j*** = 4

$$p_{o_1}^t = 0.1 \qquad p_{o_2}^t = 0 \qquad p_{o_3}^t = 0.6 \qquad p_{o_4}^t = 1 \qquad C^t = 1$$

- ***j*** = 1: $P_{0,1}^t = P_{-1,0}^t \cdot p_{o_1}^t + P_{0,0}^t \cdot \left(1 - p_{o_1}^t\right) = 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

$$P_{1,1}^t = P_{0,0}^t \cdot p_{o_1}^t + P_{1,0}^t \cdot \left(1 - p_{o_1}^t\right) = 1 \cdot 0.1 + 0 \cdot 0.9 = 0.1$$

- ***j*** = 3: $P_{0,2}^t = P_{-1,1}^t \cdot p_{o_3}^t + P_{0,1}^t \cdot \left(1 - p_{o_3}^t\right) = 0 \cdot 0.6 + 0.9 \cdot 0.4 = 0.36$

$$P_{1,2}^t = P_{0,1}^t \cdot p_{o_3}^t + P_{1,1}^t \cdot \left(1 - p_{o_3}^t\right) = 0.9 \cdot 0.6 + 0.1 \cdot 0.4 = 0.58$$

# Initial Computation (3)



- Example: ***n*** = 4, ***k*** = 2

$$p^t_{o_1} = 0.1 \qquad p^t_{o_2} = 0 \qquad p^t_{o_3} = 0.6 \qquad p^t_{o_4} = 1 \qquad C^t = 1$$

- ***j*** = 1: $P^t_{0,1} = P^t_{-1,0} \cdot p^t_{o_1} + P^t_{0,0} \cdot \left(1 - p^t_{o_1}\right) = 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

$$P^t_{1,1} = P^t_{0,0} \cdot p^t_{o_1} + P^t_{1,0} \cdot \left(1 - p^t_{o_1}\right) = 1 \cdot 0.1 + 0 \cdot 0.9 = 0.1$$

- ***j*** = 3: $P^t_{0,2} = P^t_{-1,1} \cdot p^t_{o_3} + P^t_{0,1} \cdot \left(1 - p^t_{o_3}\right) = 0 \cdot 0.6 + 0.9 \cdot 0.4 = 0.36 = P^t_{PBR}(0)$

$$P^t_{1,2} = P^t_{0,1} \cdot p^t_{o_3} + P^t_{1,1} \cdot \left(1 - p^t_{o_3}\right) = 0.9 \cdot 0.6 + 0.1 \cdot 0.4 = 0.58 = P^t_{PBR}(1)$$

- Initial PIR result:

$$P^t_q(1) = P^t_{PBR}(1-1-1) = P^t_{PBR}(-1) = 0$$

$$P^t_q(2) = P^t_{PBR}(2-1-1) = P^t_{PBR}(0) = 0.36$$

# Incremental Processing (1)

- Location update of one alternative location of object ***o***: compute $P_q^t(i)$ $\forall$ ***i*** = 1,...,***k***
- Naive solution: Apply PBR $\Rightarrow$ O(***n***) $\forall$ ***i*** = 1,...,***k***
- Enhanced solution: just consider update of ***o*** $\Rightarrow$ O(***1***) $\forall$ ***i*** = 1,...,***k***
  - Phase 1
    - Remove effect of old value $p_o^t$ from $P_{PBR}^t(i)$ $\forall$ ***i*** = 0,...,***k-1***
    - Obtain intermediate result $\hat{P}_{PBR}^{t+1}(i)$
  - Phase 2
    - Incorporate effect of new value $p_o^{t+1}$ in $\hat{P}_{PBR}^{t+1}(i)$
    - Obtain new PIR result $P_q^{t+1}(i)$

# Incremental Processing (2)



- Phase 1: Three cases
  1. $p_o^t = 0 \quad \Rightarrow \quad \hat{P}_{PBR}^t(i) = P_{PBR}^t(i)$
  2. $p_o^t = 1 \quad \Rightarrow \quad \hat{P}_{PBR}^t(i) = P_{PBR}^t(i)$ and $\mathrm{C}^{t+1} = \mathrm{C}^{t} - 1$
  3. $0 < p_o^t < 1 \Rightarrow$ remove $p_o^t$ from $P_{PBR}^t(i)$

$$P_{PBR}^t(i) = \hat{P}_{PBR}^t(i-1) \cdot p_o^t + \hat{P}_{PBR}^t(i) \cdot (1 - p_o^t)$$

$$\hat{P}_{PBR}^t(i) = \frac{P_{PBR}^t(i) - \hat{P}_{PBR}^t(i-1) \cdot p_o^t}{1 - p_o^t}$$

$$\hat{P}_{PBR}^t(0) = \frac{P_{PBR}^t(0)}{1 - p_o^t}$$

# Incremental Processing (3)



- Phase 2: Three cases
  1. $p_o^{t+1} = 0 \quad \Rightarrow \quad P_{PBR}^{t+1}(i) = \hat{P}_{PBR}^{t}(i)$
  2. $p_o^{t+1} = 1 \quad \Rightarrow \quad P_{PBR}^{t+1}(i) = \hat{P}_{PBR}^{t}(i)$ and $C^{t+1} = C^{t} + 1$
  3. $0 < p_o^{t+1} < 1 \Rightarrow$ compute $P_{PBR}^{t+1}(i)$ applying PBR

$$P_{PBR}^{t+1}(i) = \hat{P}_{PBR}^{t}(i-1) \cdot p_o^{t+1} + \hat{P}_{PBR}^{t}(i) \cdot \left(1 - p_o^{t+1}\right)$$

- New PIR result:

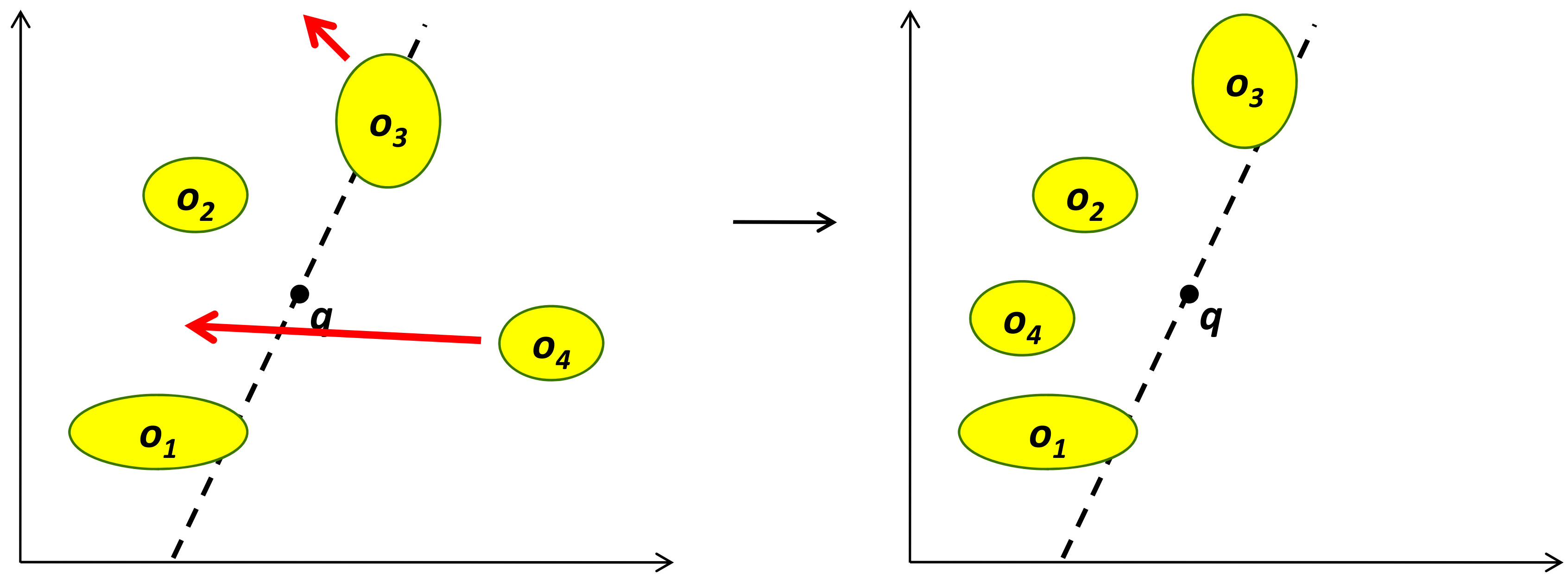
$$P_q^{t+1}(i) = \begin{cases} P_{PBR}^{t+1}\left(i - 1 - C^{t+1}\right) & \text{if } C^{t+1} + 1 \leq i \leq C^{t+1} + 1 + k \\ 0 & \text{else} \end{cases}$$

# Incremental Processing (4)



- Example: ***n*** = 4, ***k*** = 2

$$p^t_{o_1} = 0.1 \quad p^t_{o_2} = 0 \quad p^t_{o_3} = 0.6 \rightarrow p^{t+1}_{o_3} = 0.2 \quad p^t_{o_4} = 1 \rightarrow p^{t+2}_{o_4} = 0 \quad C^t = 1$$



# Incremental Processing (4)



- Example: ***n*** = 4, ***k*** = 2

$$p_{o_1}^t = 0.1 \quad p_{o_2}^t = 0 \quad p_{o_3}^t = 0.6 \rightarrow p_{o_3}^{t+1} = 0.2 \quad p_{o_4}^t = 1 \rightarrow p_{o_4}^{t+2} = 0 \quad C^t = 1$$

  – Phase 1 (Case 3): $\hat{P}_{PBR}^t(0) = \frac{P_{PBR}^t(0)}{1 - p_{o_3}^t} = \frac{0.36}{0.4} = 0.9$

$$\hat{P}_{PBR}^t(1) = \frac{P_{PBR}^t(1) - \hat{P}_{PBR}^t(0) \cdot p_{o_3}^t}{1 - p_{o_3}^t} = \frac{0.58 - 0.9 \cdot 0.6}{0.4} = 0.1$$

  – Phase 2 (Case 3):

$$P_{PBR}^{t+1}(0) = \hat{P}_{PBR}^t(-1) \cdot p_{o_3}^{t+1} + \hat{P}_{PBR}^t(0) \cdot \left(1 - p_{o_3}^{t+1}\right) = 0 \cdot 0.2 + 0.9 \cdot 0.8 = 0.72$$

$$P_{PBR}^{t+1}(1) = \hat{P}_{PBR}^t(0) \cdot p_{o_3}^{t+1} + \hat{P}_{PBR}^t(1) \cdot \left(1 - p_{o_3}^{t+1}\right) = 0.9 \cdot 0.2 + 0.1 \cdot 0.8 = 0.26$$

  – PIR result: $P_q^t(1) = 0 \rightarrow P_q^{t+1}(1) = 0 \quad P_q^t(2) = 0.36 \rightarrow P_q^{t+1}(2) = 0.72$

# Incremental Processing (4)



- Example: ***n*** = 4, ***k*** = 2

$p_{o_1}^{t}=0.1 \quad p_{o_2}^{t}=0 \quad p_{o_3}^{t}=0.6 \rightarrow p_{o_3}^{t+1}=0.2 \quad p_{o_4}^{t}=1 \rightarrow p_{o_4}^{t+2}=0 \quad C^{t}=1$

- Phase 1 (Case 1):

$$\hat{P}_{PBR}^{t+1}(0)=P_{PBR}^{t+1}(0)=0.72$$
$$\hat{P}_{PBR}^{t+1}(1)=P_{PBR}^{t+1}(1)=0.26$$
$$C^{t}=0$$

- Phase 2 (Case 2):

$$P_{PBR}^{t+2}(0)=\hat{P}_{PBR}^{t+1}(0)=0.72$$
$$P_{PBR}^{t+2}(1)=\hat{P}_{PBR}^{t+1}(1)=0.26$$

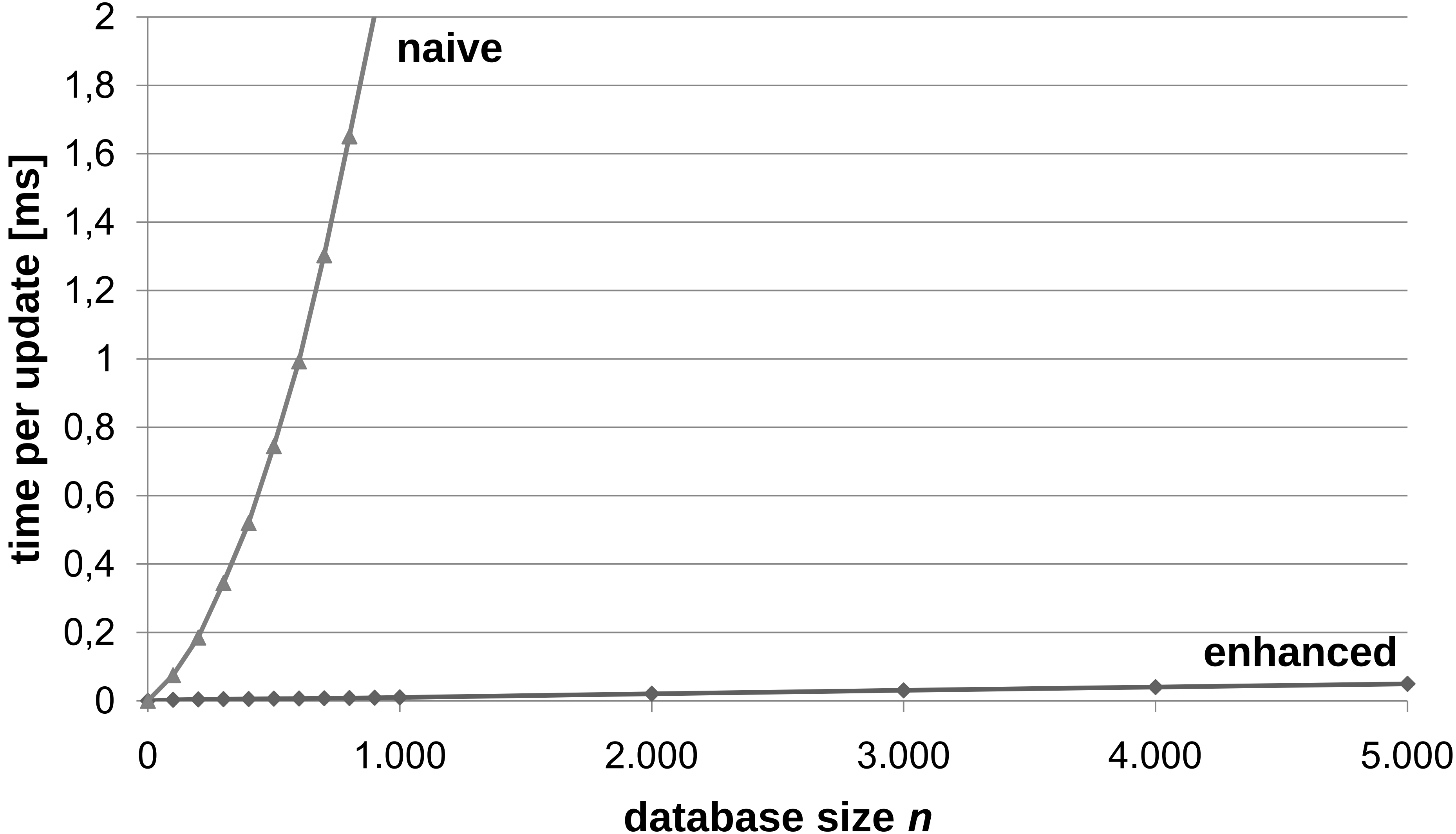
- PIR result:

$$P_{q}^{t+1}(1)=0 \rightarrow P_{q}^{t+2}(1)=P_{PBR}^{t+2}(1-1-0)=P_{PBR}^{t+2}(0)=0.72$$
$$P_{q}^{t+1}(2)=0.72 \rightarrow P_{q}^{t+2}(2)=P_{PBR}^{t+2}(2-1-0)=P_{PBR}^{t+2}(1)=0.26$$

# Experiments (1)



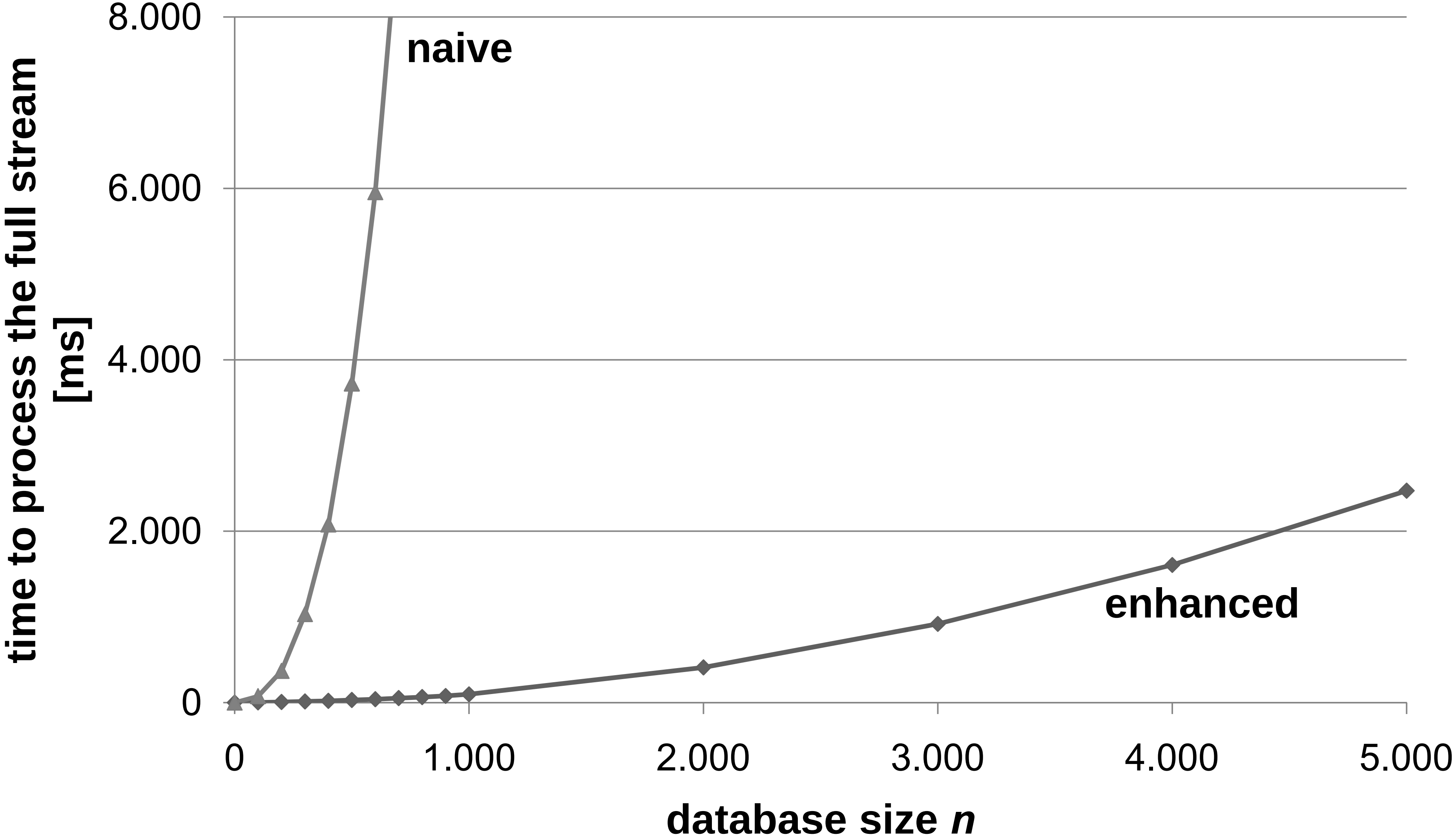
- dimensions = 2, ***m*** = 10, ***σ*** = 5, ***k*** = ***n***, buffer = 3



# Experiments (2)



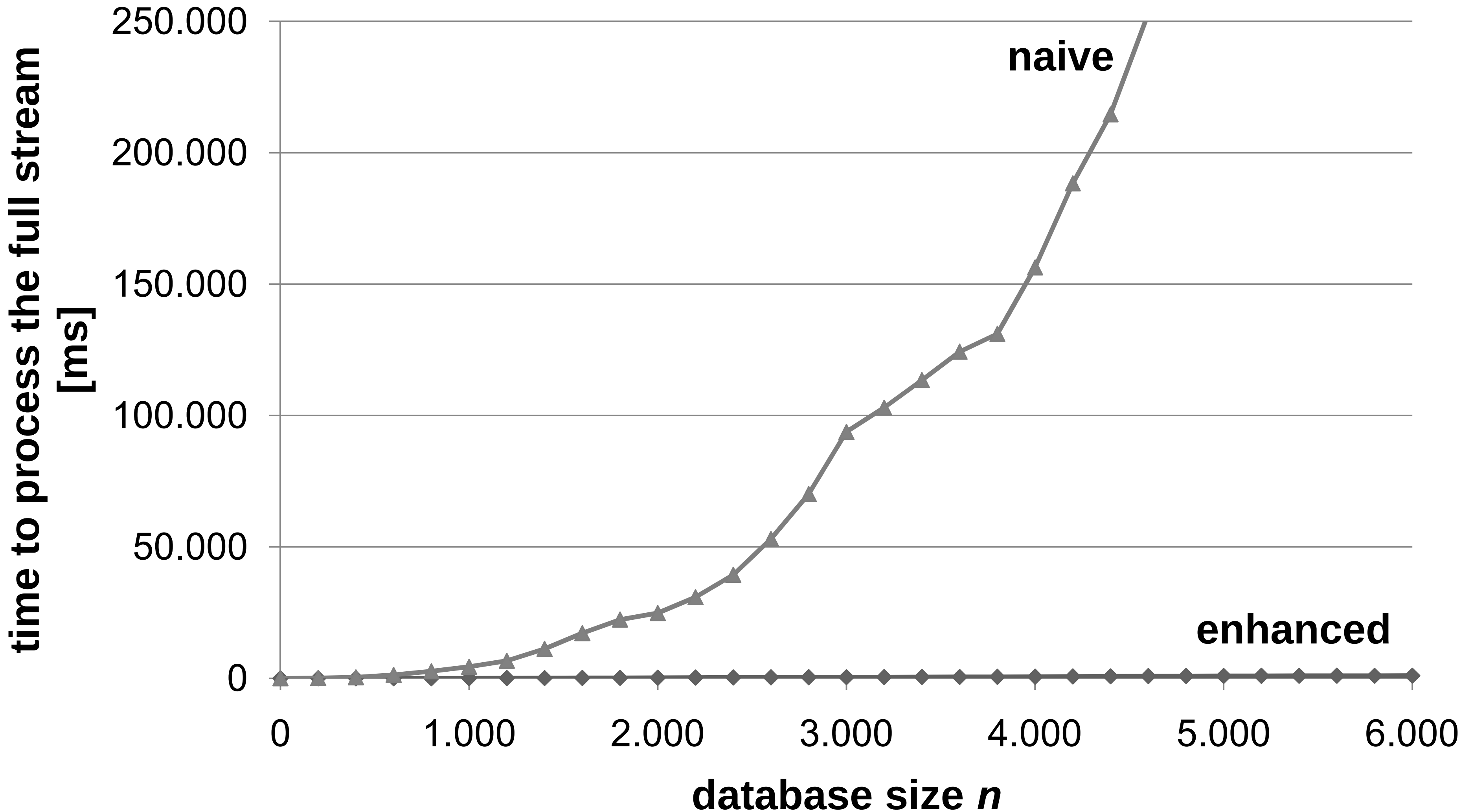
- dimensions = 2, ***m*** = 10, ***σ*** = 5, ***k*** = ***n***, buffer = 3



# Experiments (3)

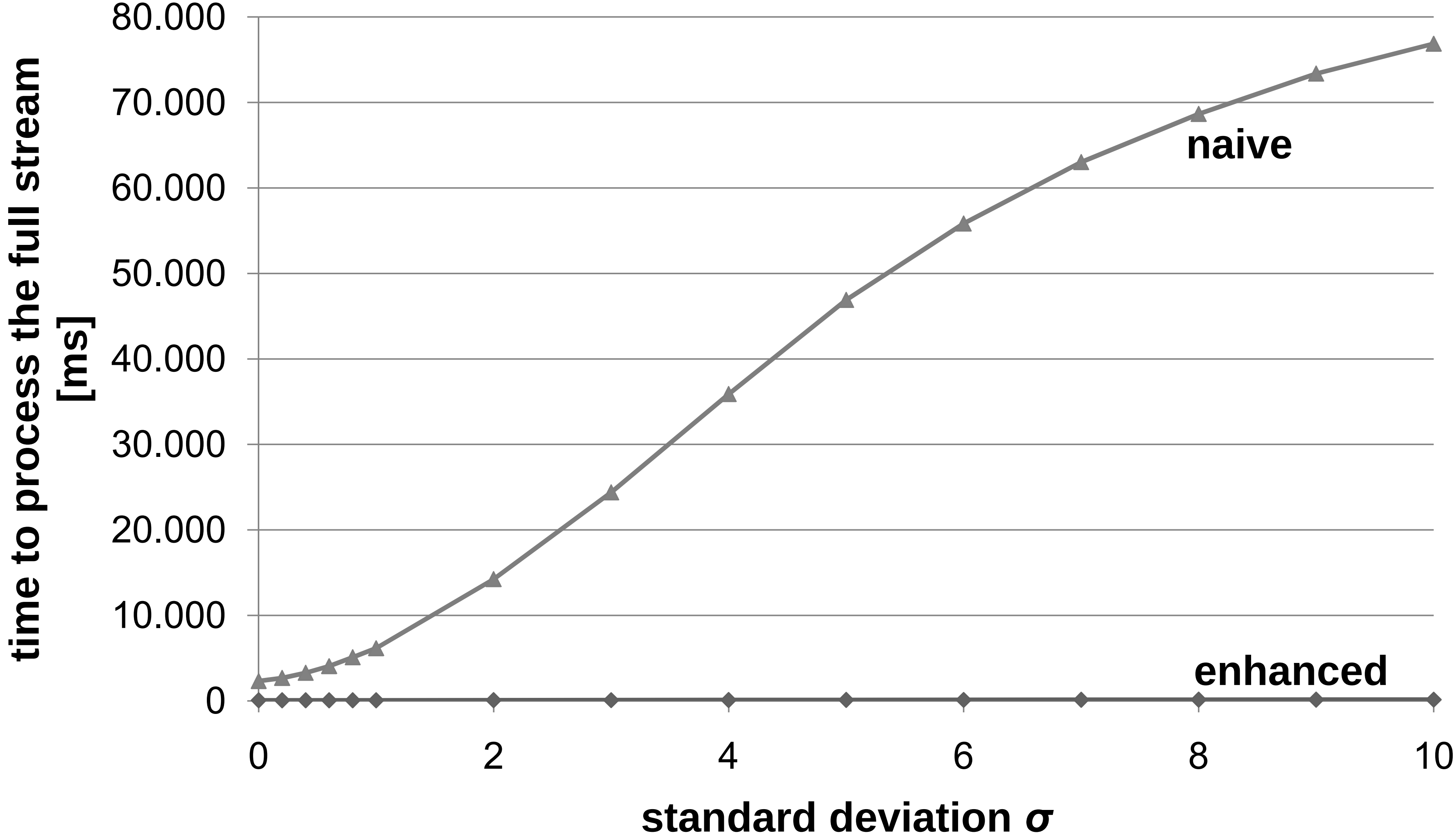


- IIP dataset, dimensions = 2, ***m*** = 10, ***k*** = ***n***, buffer = 3



# Experiments (4)

- ***n*** = 10,000, dimensions = 2, ***m*** = 10, ***k*** = ***n***, buffer = 3



DATABASE SYSTEMS GROUP

# Summary

LMU

- Efficient solution for PIR queries on continuous data yielding update costs of O($\boldsymbol{k}$) instead of O($\boldsymbol{k \cdot n}$)
- The framework can be adapted to other query types, e.g. the probabilistic threshold inverse ranking query
- Future work: approximate approach using lower and upper bounds for the probabilities and applying the concept of Generating Functions