

Space-Time Trade-offs for Conjunctive Queries with Access Patterns

Hangdong Zhao[†], Shaleen Deep[‡] and Paraschos Koutris[†]

[†]University of Wisconsin - Madison, [‡]Microsoft Jim Gray Systems Lab

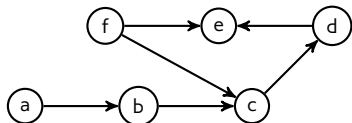
June 17, 2023 (PODS 2023 @ Seattle, WA)

Motivating Example

3-reachability query with **access pattern**

R encodes a directed graph G

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability workload on G

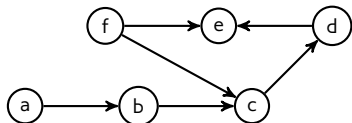
is there a 3-path from *start node* x_1 to *end node* x_4 ?

Motivating Example

3-reachability query with **access pattern**

R encodes a directed graph G

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability workload on G

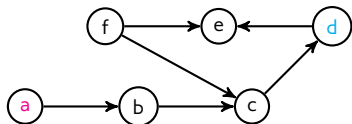
*is there a 3-path from **start node** x_1 to **end node** x_4 ?*

Motivating Example

3-reachability query with **access pattern**

R encodes a directed graph G

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability workload on G

is there a 3-path from **start node** $x_1 = a$ to **end node** $x_4 = d$?

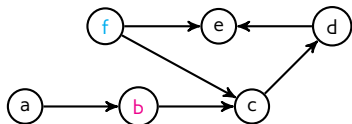
$$TRUE = \Pi_{\emptyset} R(x_1 = a, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4 = d)$$

Motivating Example

3-reachability query with **access pattern**

R encodes a directed graph G

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



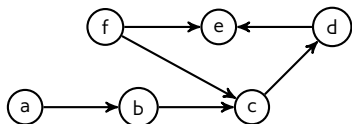
3-reachability workload on G

is there a 3-path from **start node** $x_1 = b$ to **end node** $x_4 = f$?

$$FALSE = \Pi_{\emptyset} R(x_1 = b, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4 = f)$$

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- Answer on-the-fly

small space = $O(|R|)$ LONG TIME = $O(|R|)$

- Materialize the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

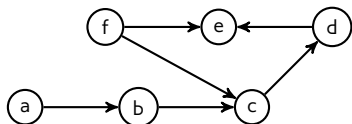
S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- **Answer on-the-fly**

small space = $O(|R|)$ **LONG TIME** = $O(|R|)$

- **Materialize** the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

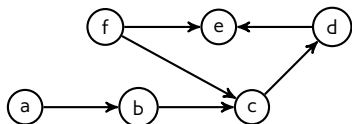
S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- **Answer on-the-fly**

small space = $O(|R|)$ **LONG TIME** = $O(|R|)$

- **Materialize** the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

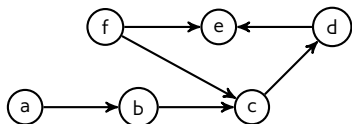
S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- **Answer on-the-fly**

small space = $O(|R|)$ **LONG TIME** = $O(|R|)$

- **Materialize** the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

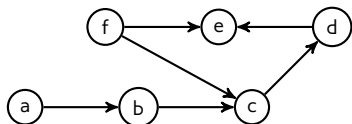
S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- **Answer on-the-fly**

small space = $O(|R|)$ **LONG TIME** = $O(|R|)$

- **Materialize** the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

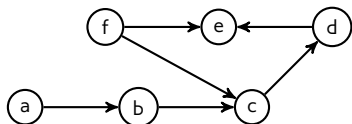
S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Motivating Example

R	x_1	x_2
	a	b
	b	c
	c	d
	d	e
	f	e
	f	c



3-reachability query with access pattern

$$\varphi(| \underbrace{x_1, x_4}_{\text{access pattern}} |) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

- **Answer on-the-fly**

small space = $O(|R|)$ **LONG TIME** = $O(|R|)$

- **Materialize** the 3-paths table $S = \Pi_{x_1, x_4} R(x_1, x_2) \bowtie R(x_2, x_3) \bowtie R(x_3, x_4)$

S	start: x_1	end: x_4
	a	d
	b	e
	f	e

LARGE SPACE = $O(|R|^2)$ short time = $O(1)$

- Can we develop a space-efficient materialization strategy?

Talk Outline

- (1) Problem Setup Conjunctive Queries with Access Patterns (CQAP)
- (2) Main Techniques Tree Decompositions and Entropic Bounds
- (3) Main Contribution A General Framework for Space-time Trade-offs of CQAPs

Talk Outline

- (1) **Problem Setup** Conjunctive Queries with Access Patterns (CQAP)
- (2) Main Techniques Tree Decompositions and Entropic Bounds
- (3) Main Contribution A General Framework for Space-time Trade-offs of CQAPs

Conjunctive Query (CQ) with Access Patterns

$$\varphi\left(\underbrace{x_H}_{\text{output vars}} \mid \underbrace{x_A}_{\text{access pattern}}\right) \leftarrow \underbrace{\bigwedge_i R_i(x_i)}_{\text{full join}}$$

An access request $Q_A(x_A)$ asks the CQ

$$\varphi\left(\underbrace{x_H}_{\text{output vars}}\right) \leftarrow \underbrace{Q_A(x_A)}_{\text{access request}} \wedge \bigwedge_i R_i(x_i)$$

Conjunctive Query with Access Patterns (CQAP)

$$\varphi\left(\underbrace{x_H}_{\text{output vars}} \mid \underbrace{x_A}_{\text{access pattern}}\right) \leftarrow \bigwedge_i R_i(x_i)$$

full join

An access request $Q_A(x_A)$ asks the CQ

$$\varphi\left(\underbrace{x_H}_{\text{output vars}}\right) \leftarrow \overbrace{Q_A(x_A)}^{\text{access request}} \wedge \bigwedge_i R_i(x_i)$$

Conjunctive Query with Access Patterns (CQAP)

$$\varphi\left(\underbrace{\mathbf{x}_H}_{\text{output vars}} \mid \underbrace{\mathbf{x}_A}_{\text{access pattern}}\right) \leftarrow \bigwedge_i R_i(\mathbf{x}_i)$$

full join

An access request $Q_A(\mathbf{x}_A)$ asks the CQ

$$\varphi\left(\underbrace{\mathbf{x}_H}_{\text{output vars}}\right) \leftarrow \overbrace{Q_A(\mathbf{x}_A)}^{\text{access request}} \wedge \bigwedge_i R_i(\mathbf{x}_i)$$

3-reachability CQAP

Conjunctive Query with Access Patterns (CQAP)

$$\varphi(\mid \underbrace{x_1, x_4}_{\text{access pattern}}) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

An access request $Q_{14}(x_1, x_4)$ asks the CQ

$$\varphi() \leftarrow Q_{14}(x_1, x_4) \wedge R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

3-reachability CQAP

Conjunctive Query with Access Patterns (CQAP)

$$\varphi(\mid \underbrace{x_1, x_4}_{\text{access pattern}}) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

An access request $Q_{14}(x_1, x_4)$ asks the CQ

$$\varphi() \leftarrow Q_{14}(x_1, x_4) \wedge R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

Why express access requests $Q_A(x_A)$ as a relation?

$$\varphi(\mid \underbrace{x_1, x_4}_{\text{access pattern}}) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

More semantics, e.g. batch processing of N access requests $\mid Q_{14} \mid = N$

$$\varphi(a, b) \leftarrow R(x_1 = a, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = b)$$

$$\varphi(f, c) \leftarrow R(x_1 = f, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = c)$$

$$\varphi(e, d) \leftarrow R(x_1 = e, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = d)$$

\Downarrow

Q_{14}	x_1	x_4
	a	b
	f	c
	e	d

$$\varphi(x_1, x_4) \leftarrow Q_{14}(x_1, x_4) \wedge R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

This talk focuses on *single requests* $N = 1$.

Why express access requests $Q_A(x_A)$ as a relation?

$$\varphi(\mid \underbrace{x_1, x_4}_{\text{access pattern}}) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

More semantics, e.g. batch processing of N access requests $|Q_{14}| = N$

$$\varphi(a, b) \leftarrow R(x_1 = a, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = b)$$

$$\varphi(f, c) \leftarrow R(x_1 = f, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = c)$$

$$\varphi(e, d) \leftarrow R(x_1 = e, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4 = d)$$

Q_{14}

x_1	x_4
a	b
f	c
e	d

\Downarrow

$$\varphi(x_1, x_4) \leftarrow Q_{14}(x_1, x_4) \wedge R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

This talk focuses on *single requests* $N = 1$.

Problem Setup



database D

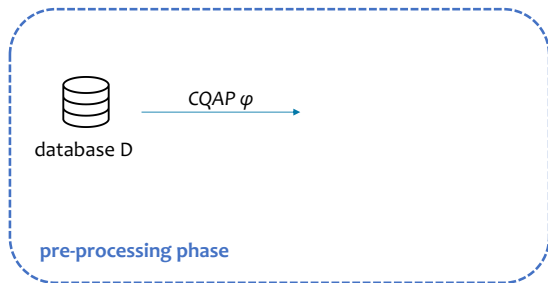
Problem Setup



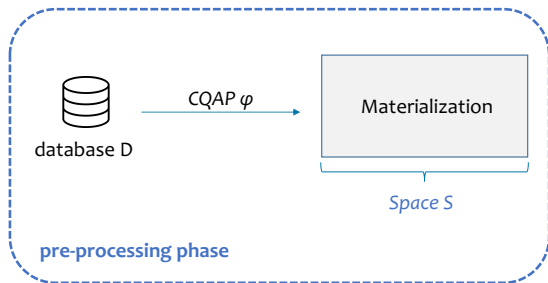
database D

pre-processing phase

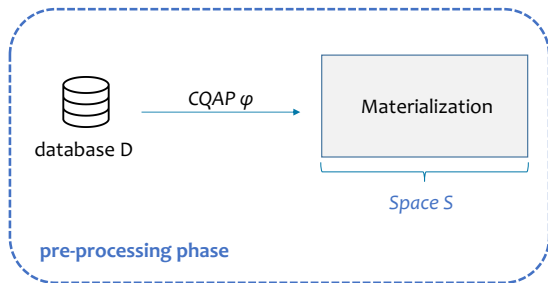
Problem Setup



Problem Setup

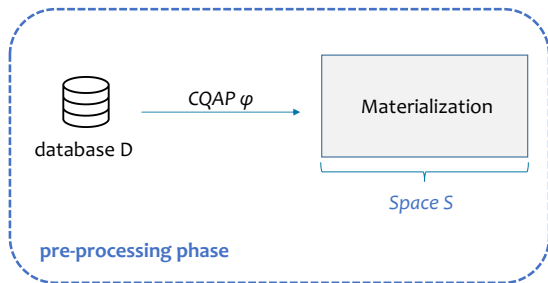


Problem Setup

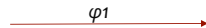


online phase

Problem Setup



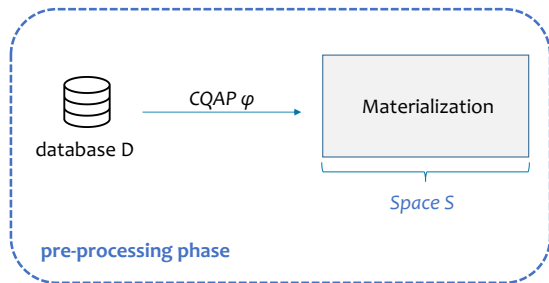
online phase



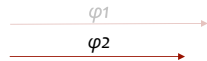
answering time T

(for the worst-case query instance)

Problem Setup

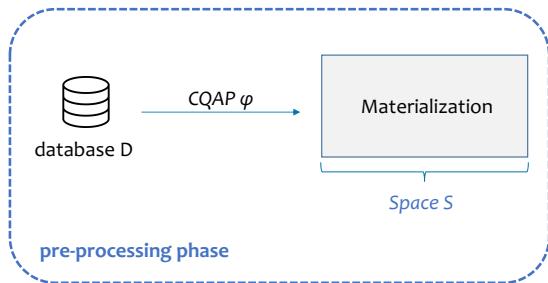


online phase

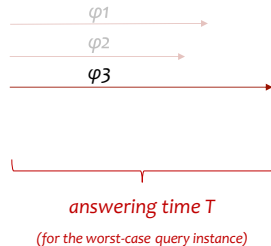


answering time T
(for the worst-case query instance)

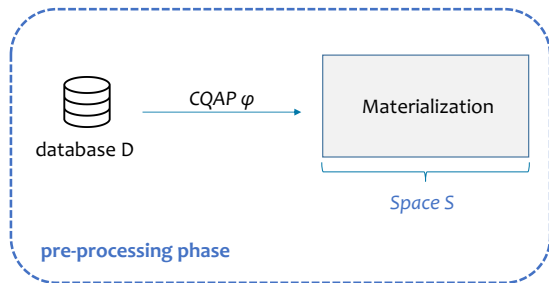
Problem Setup



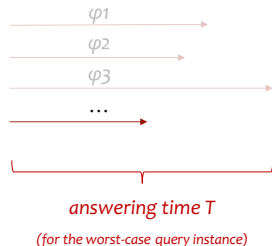
online phase



Problem Setup



online phase



Space-time trade-offs

- pre-processing phase materialization **space S**
- online phase answering **time T** for the worst-case access request

Talk Outline

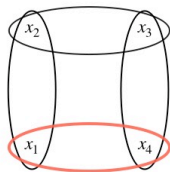
- (1) Problem Setup Conjunctive Queries with Access Patterns (CQAP)
- (2) Main Techniques Tree Decompositions and Entropic Bounds
- (3) Main Contribution A General Framework for Space-time Trade-offs of CQAPs

Tree Decompositions

$$\varphi() \leftarrow Q(x_1, x_4) \wedge R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

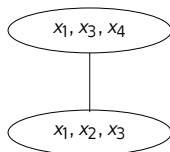
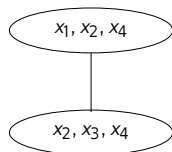
Tree Decomposition of a CQ φ

- Tree $\mathcal{T} = (V, E)$
- Bag $\chi(t)$ for each tree-node $t \in V$



A TD must satisfy

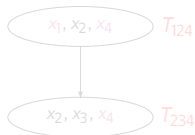
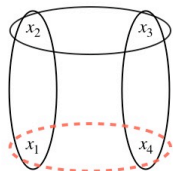
- (1) Running intersection property
 - (2) Containment property
- every hyperedge is covered by some bag



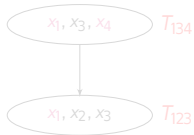
Partially Materialized Tree Decompositions (PMTDs)

$$\varphi(x_1, x_4) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

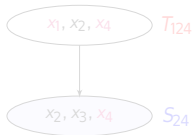
PMTD of a CQAP = TD with materialized subtrees



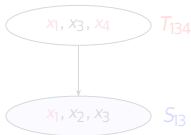
PMTD #1



PMTD #2



PMTD #3



PMTD #4

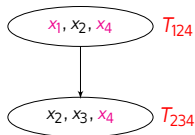
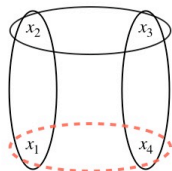


PMTD #5

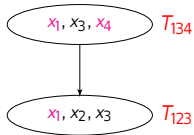
Partially Materialized Tree Decompositions (PMTDs)

$$\varphi(| x_1, x_4) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

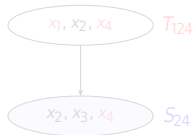
PMTD of a CQAP = TD with materialized subtrees



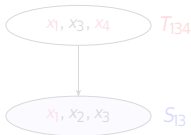
PMTD #1



PMTD #2



PMTD #3



PMTD #4

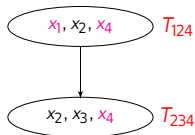
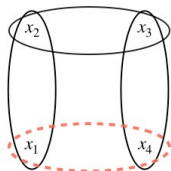


PMTD #5

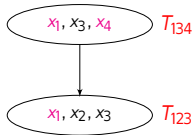
Partially Materialized Tree Decompositions (PMTDs)

$$\varphi(| x_1, x_4) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$

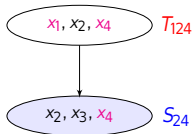
PMTD of a CQAP = TD with materialized subtrees



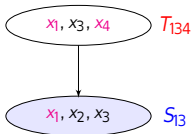
PMTD #1



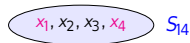
PMTD #2



PMTD #3



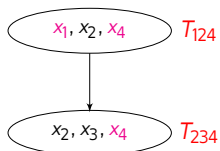
PMTD #4



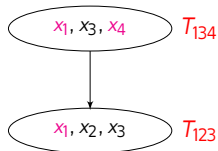
PMTD #5

PMTDs = 2-phase Query Plans

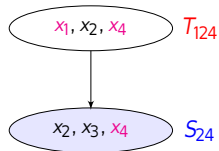
We have 5 PMTDs in total



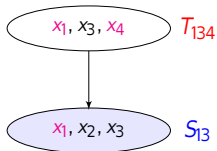
PMTD #1



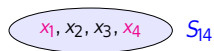
PMTD #2



PMTD #3



PMTD #4

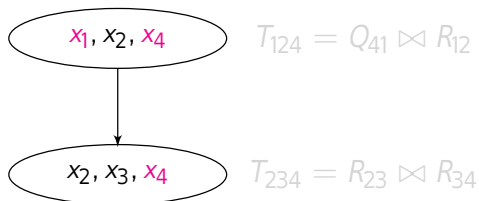


PMTD #5

PMTDs and 2-phase Query Plans

$$\text{OUT} = T_{124} \bowtie T_{234}$$

↑ Yannakakis



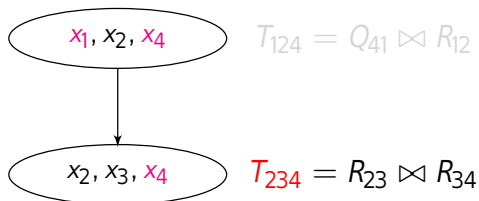
PMTD #1

[On-the-fly evaluation]

PMTDs and 2-phase Query Plans

$$\text{OUT} = T_{124} \bowtie T_{234}$$

↑ Yannakakis



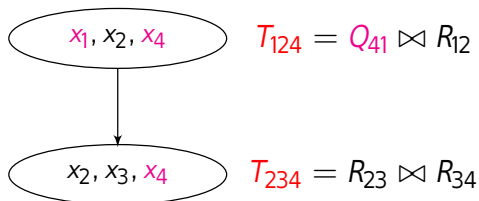
PMTD #1

[On-the-fly evaluation]

PMTDs and 2-phase Query Plans

$$\text{OUT} = T_{124} \bowtie T_{234}$$

↑ Yannakakis



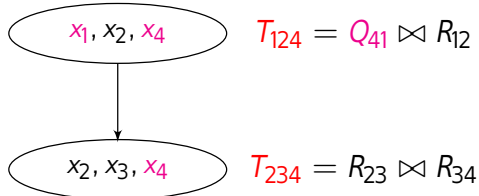
PMTD #1

[On-the-fly evaluation]

PMTDs and 2-phase Query Plans

$$\text{OUT} = T_{124} \bowtie T_{234}$$

↑ Yannakakis



PMTD #1

[On-the-fly evaluation]

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(Q_{14} \bowtie S_{14})$$

↑ Yannakakis

$$x_1, x_2, x_3, x_4$$

$$S_{14} = \Pi_{14}(R_{12} \bowtie R_{23} \bowtie R_{34})$$

PMTD #5

[Materialize the 3-paths table]

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(Q_{14} \bowtie S_{14})$$

↑ Yannakakis

$$x_1, x_2, x_3, x_4$$

$$S_{14} = \Pi_{14}(R_{12} \bowtie R_{23} \bowtie R_{34})$$

PMTD #5

[Materialize the 3-paths table]

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(Q_{14} \bowtie S_{14})$$

↑ Yannakakis

x_1, x_2, x_3, x_4 $S_{14} = \Pi_{14}(R_{12} \bowtie R_{23} \bowtie R_{34})$

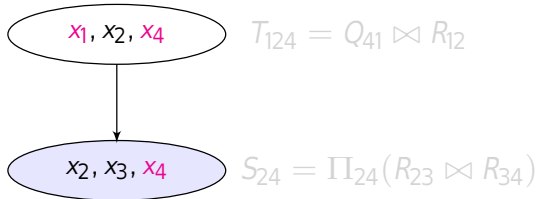
PMTD #5

[Materialize the 3-paths table]

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(T_{124} \bowtie S_{24})$$

↑ Yannakakis

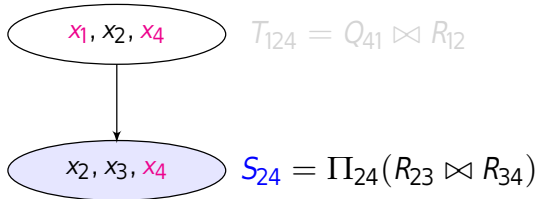


PMTD #3

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(T_{124} \bowtie S_{24})$$

↑ Yannakakis

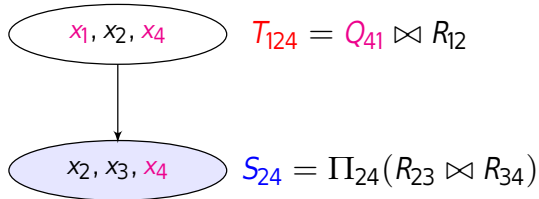


PMTD #3

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(T_{124} \bowtie S_{24})$$

↑ Yannakakis

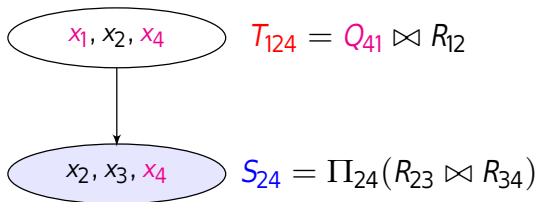


PMTD #3

PMTDs = 2-phase Query Plans

$$\text{OUT} = \Pi_{\emptyset}(T_{124} \bowtie S_{24})$$

↑ Yannakakis

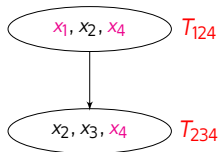


PMTD #3

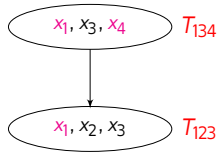
Talk Outline

- (1) Problem Setup Conjunctive Queries with Access Patterns (CQAP)
- (2) Main Techniques Tree Decompositions and Entropic Bounds
- (3) Main Contribution A General Framework for Space-time Trade-offs of CQAPs

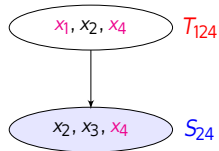
Can we use all PMTDs? Yes!



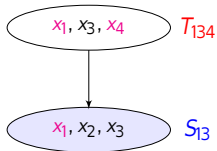
PMTD #1



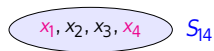
PMTD #2



PMTD #3

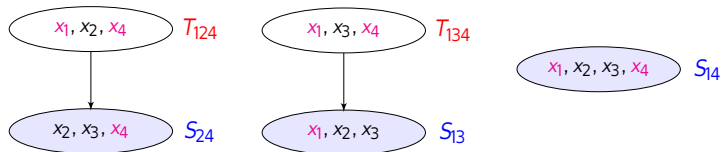


PMTD #4



PMTD #5

(1/4) List out all PMTDs of the CQAP



The Disjunctive Normal Form (DNF)

$$(T_{124} \wedge S_{24}) \vee (T_{134} \wedge S_{13}) \vee (T_{124} \wedge T_{234}) \vee (T_{134} \wedge T_{123}) \vee S_{14}$$

(2/4) DNF to 2-phase Disjunctive Rules

DNF

$$(T_{124} \wedge S_{24}) \vee (T_{134} \wedge S_{13}) \vee (T_{124} \wedge T_{234}) \vee (T_{134} \wedge T_{123}) \vee S_{14}$$

Equivalent CNF (use distributivity)

$$(S_{14} \vee T_{124} \vee T_{134}) \wedge (S_{13} \vee S_{14} \vee T_{124} \vee T_{123}) \wedge (S_{24} \vee S_{14} \vee T_{134} \vee T_{234}) \\ \wedge (S_{24} \vee S_{13} \vee S_{14} \vee T_{234} \vee T_{123})$$

Evaluation of a 2-phase Disjunctive Rule

$$\underbrace{S_{14}}_{\text{pre-processing outputs}} \vee \overbrace{(T_{124} \vee T_{134})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

We want tables

- $S_{14} \leq S$ during pre-processing phase
- T_{124}, T_{134} during online phase

such that every output tuple is placed in one of the output tables.

(2/4) DNF to 2-phase Disjunctive Rules

DNF

$$(T_{124} \wedge S_{24}) \vee (T_{134} \wedge S_{13}) \vee (T_{124} \wedge T_{234}) \vee (T_{134} \wedge T_{123}) \vee S_{14}$$

Equivalent CNF (use distributivity)

$$(S_{14} \vee T_{124} \vee T_{134}) \wedge (S_{13} \vee S_{14} \vee T_{124} \vee T_{123}) \wedge (S_{24} \vee S_{14} \vee T_{134} \vee T_{234}) \\ \wedge (S_{24} \vee S_{13} \vee S_{14} \vee T_{234} \vee T_{123})$$

Evaluation of a 2-phase Disjunctive Rule

$$\underbrace{S_{13} \vee S_{14}}_{\text{pre-processing outputs}} \vee \underbrace{(T_{124} \vee T_{123})}_{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

We want tables

- $S_{13}, S_{14} \leq S$ during pre-processing phase
- T_{124}, T_{123} during online phase

such that every output tuple is placed in one of the output tables.

(2/4) DNF to 2-phase Disjunctive Rules

DNF

$$(T_{124} \wedge S_{24}) \vee (T_{134} \wedge S_{13}) \vee (T_{124} \wedge T_{234}) \vee (T_{134} \wedge T_{123}) \vee S_{14}$$

Equivalent CNF (use distributivity)

$$(S_{14} \vee T_{124} \vee T_{134}) \wedge (S_{13} \vee S_{14} \vee T_{124} \vee T_{123}) \wedge (S_{24} \vee S_{14} \vee T_{134} \vee T_{234}) \\ \wedge (S_{24} \vee S_{13} \vee S_{14} \vee T_{234} \vee T_{123})$$

Evaluation of a 2-phase Disjunctive Rule

$$\underbrace{S_{24} \vee S_{14}}_{\text{pre-processing outputs}} \vee \overbrace{(T_{134} \vee T_{234})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

We want tables

- $S_{24}, S_{14} \leq S$ during pre-processing phase
- T_{134}, T_{234} during online phase

such that every output tuple is placed in one of the output tables.

(2/4) DNF to 2-phase Disjunctive Rules

DNF

$$(T_{124} \wedge S_{24}) \vee (T_{134} \wedge S_{13}) \vee (T_{124} \wedge T_{234}) \vee (T_{134} \wedge T_{123}) \vee S_{14}$$

Equivalent CNF (use distributivity)

$$(S_{14} \vee T_{124} \vee T_{134}) \wedge (S_{13} \vee S_{14} \vee T_{124} \vee T_{123}) \wedge (S_{24} \vee S_{14} \vee T_{134} \vee T_{234}) \\ \wedge (S_{24} \vee S_{13} \vee S_{14} \vee T_{234} \vee T_{123})$$

Evaluation of a 2-phase Disjunctive Rule

$$\underbrace{S_{24} \vee S_{13} \vee S_{14}}_{\text{pre-processing outputs}} \vee \overbrace{(T_{234} \vee T_{123})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

We want tables

- $S_{24}, S_{13}, S_{14} \leq S$ during **pre-processing phase**
- T_{234}, T_{123} during **online phase**

such that every output tuple is placed in one of the output tables.

(3/4) The PANDA Algorithm¹

Proof-Assisted eNtropic Degree-Aware

use **entropic bounds** to guide join processing

¹Khamis, Ngo, Suciu, "What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?" PODS'17

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$\underbrace{S_{14}}_{\text{pre-processing outputs}} \quad \underbrace{\vee (T_{124} \vee T_{134})}_{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &+ 2 \cdot h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \left(\overbrace{T_{124} \vee T_{134}}^{\text{online outputs}} \right) \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

pre-processing outputs

entropic bound	two-phase query plan
$h(x_1 x_2) + h(x_3 x_4) = h(x_1) + h(x_2 x_1) + h(x_3 x_4)$ $= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4)$ $\geq h(x_1 x_4) + h(x_4 x_1) + h(x_3 x_4)$	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$+ 2 \cdot h(x_1 x_4)$ $= h(x_1 x_4) + h(x_1 x_2 x_4) + h(x_3 x_1 x_4) + h(x_1 x_4)$ $= h(x_1 x_4) + h(x_1 x_2 x_4) + h(x_1 x_3 x_4)$	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1 x_2) + h(x_3 x_4) + 2 \cdot h(x_1 x_4) \geq h(x_1 x_4) + h(x_1 x_2 x_4) + h(x_1 x_3 x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \left(\underbrace{\quad}_{\text{pre-processing outputs}} \right) \vee \left(\underbrace{T_{124} \vee T_{134}}_{\text{online outputs}} \right) \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$h(x_1x_2) + h(x_3x_4) = h(x_1) + h(x_2 x_1) + h(x_3x_4)$ $= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4)$ $\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4)$	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$+ 2 \cdot h(x_1x_4)$ $= h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4)$ $= h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \underbrace{\qquad\qquad\qquad}_{\text{pre-processing outputs}} \quad \overbrace{\left(T_{124} \vee T_{134} \right)}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &\quad + 2 \cdot h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \quad \underbrace{\vee (T_{124} \vee T_{134})}_{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

pre-processing outputs

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &+ 2 \cdot h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \quad \overbrace{\vee (T_{124} \vee T_{134})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

pre-processing outputs

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &+ 2 \cdot h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \underbrace{\quad}_{\text{pre-processing outputs}} \quad \overbrace{(\underbrace{T_{124} \vee T_{134}}_{\text{online outputs}})} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &\quad + 2 \cdot h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ &= h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

- Evaluation of 2-phase Disjunctive Rule

$$S_{14} \underbrace{\quad}_{\text{pre-processing outputs}} \vee \overbrace{(T_{124} \vee T_{134})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$ \begin{aligned} h(x_1x_2) + h(x_3x_4) &= h(x_1) + h(x_2 x_1) + h(x_3x_4) \\ &= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4) \\ &\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4) \end{aligned} $	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$ \begin{aligned} &+ 2 \cdot h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4) \\ = &h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4) \end{aligned} $	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

$$h(x_1x_2) + h(x_3x_4) + 2 \cdot h(x_1x_4) \geq h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$$

$$|R_{12}| \cdot |R_{34}| \cdot |Q_{14}|^2 \geq S \cdot T^2 \quad (\text{space-time trade-off})$$

(3/4) 2-phase PANDA

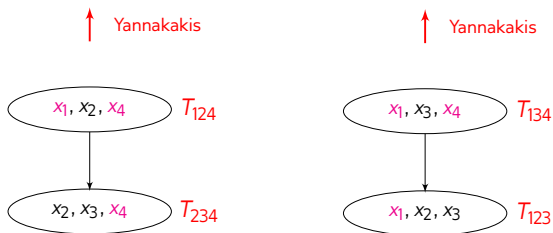
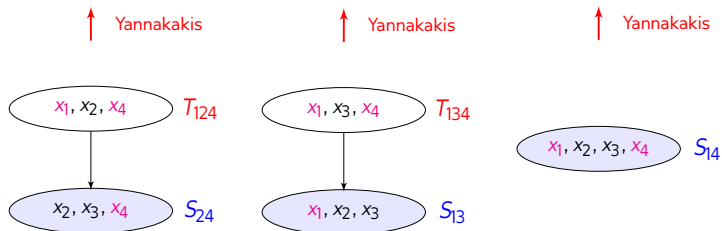
- Evaluation of 2-phase Disjunctive Rule

$$\underbrace{S_{14}}_{\text{pre-processing outputs}} \vee \overbrace{(T_{124} \vee T_{134})}^{\text{online outputs}} \leftarrow Q_{41} \wedge R_{12} \wedge R_{23} \wedge R_{34}$$

entropic bound	two-phase query plan
$h(x_1x_2) + h(x_3x_4) = h(x_1) + h(x_2 x_1) + h(x_3x_4)$ $= h(x_1) + h(x_2 x_1) + h(x_4) + h(x_3 x_4)$ $\geq h(x_1x_4) + h(x_4 x_1) + h(x_3 x_4)$	<p><i>pre-processing phase</i></p> <p>partition $R_{12}(x_1, x_2)$ into $R_1(x_1), R_2(x_1, x_2)$</p> <p>partition $R_{34}(x_3, x_4)$ into $R_4(x_4), R_3(x_3, x_4)$</p> <p>materialize $S_{14} = R_1(x_1) \bowtie R_4(x_4)$</p>
$+ 2 \cdot h(x_1x_4)$ $= h(x_1x_4) + h(x_1x_2x_4) + h(x_3 x_1x_4) + h(x_1x_4)$ $= h(x_1x_4) + h(x_1x_2x_4) + h(x_1x_3x_4)$	<p><i>online phase</i></p> <p>receive access request $Q_{14}(x_1, x_4)$</p> <p>join $T_{124} = Q_{14}(x_1, x_4) \bowtie R_2(x_1, x_2)$</p> <p>join $T_{134} = Q_{14}(x_1, x_4) \bowtie R_3(x_3, x_4)$</p>

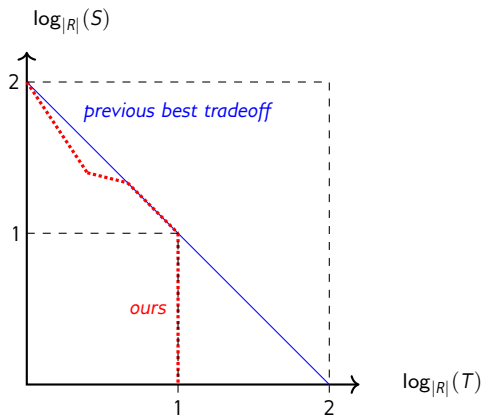
TIGHTEST entropic bound = OPTIMAL two-phase query plan = BEST trade-off

(4/4) Yannakakis on all PMTDs and union the outputs



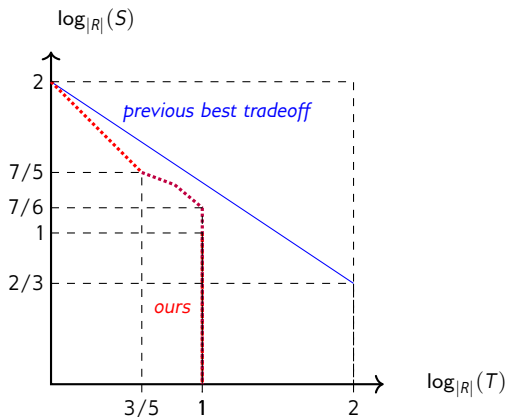
Results for the 3-reachability CQAP

$$\varphi(|x_1, x_4) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4)$$



Results for the 4-reachability CQAP

$$\varphi_4(|x_1, x_5|) \leftarrow R(x_1, x_2) \wedge R(x_2, x_3) \wedge R(x_3, x_4) \wedge R(x_4, x_5)$$



Main Contribution

A General Framework for Space-time Trade-offs of ANY CQAPs

Given a **CQAP**

- (1) List out all **PMTDs**
- (2) Turn the DNF over all **PMTDs** into a set of **2-phase Disjunctive Rules** (use distributivity)
- (3) Run **2-phase PANDA** on each **2-phase D.R.** to get tables for bags
- (4) Run Yannakakis on every **PMTD** and union the outputs