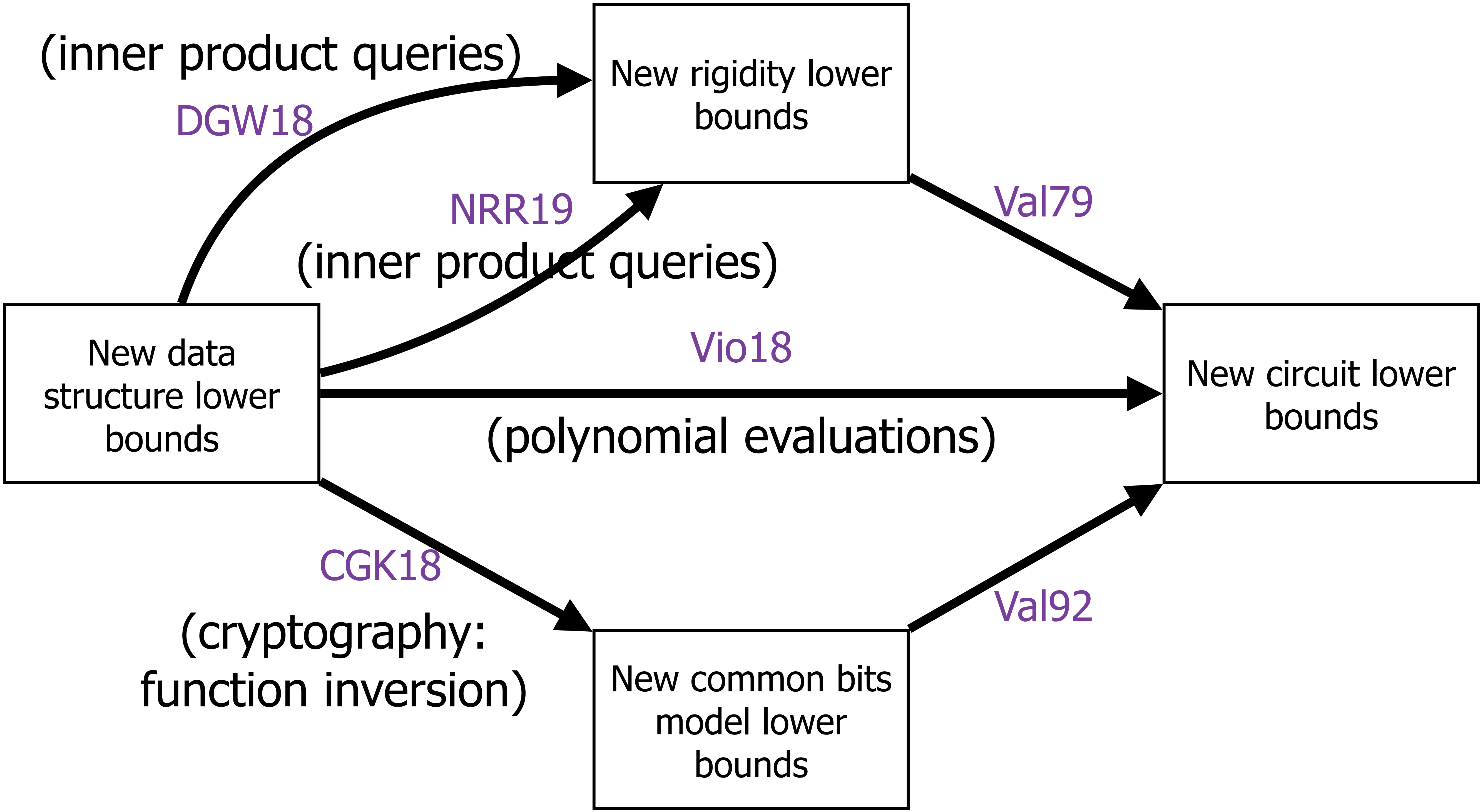
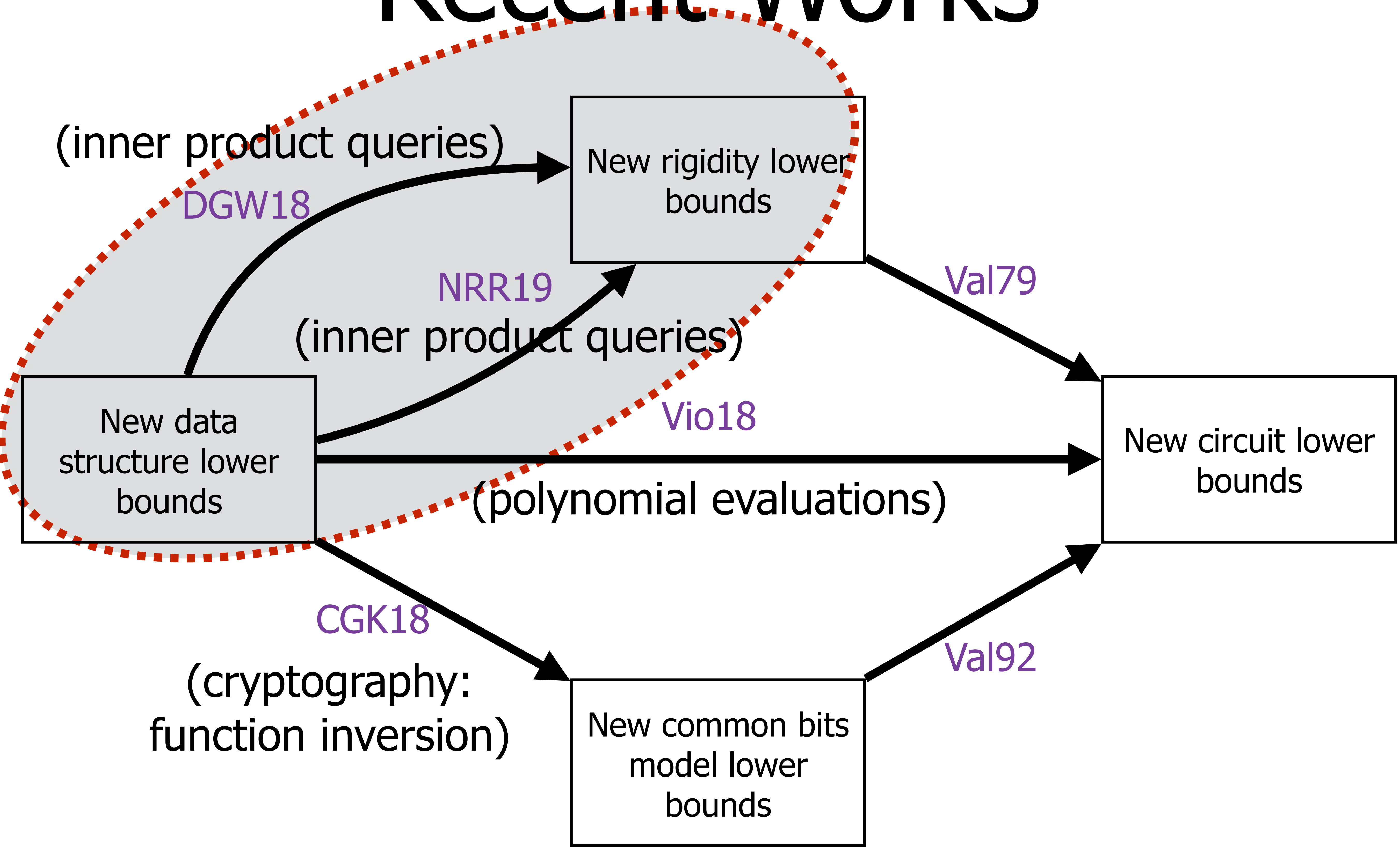
On Linear Data Structures and Matrix Rigidity

Sivaramakrishnan Natarajan Ramamoorthy
Theorem LP

Recent Works



Recent Works



Outline

- Linear data structures
- Rectangular rigidity
- Linear data structure lower bounds imply rigidity lower bounds — a result of [DGW18]
- Systematic linear data structures
  - Equivalence to rectangular rigidity
- Rigidity lower bound for Vector-Matrix-Vector problem

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Query

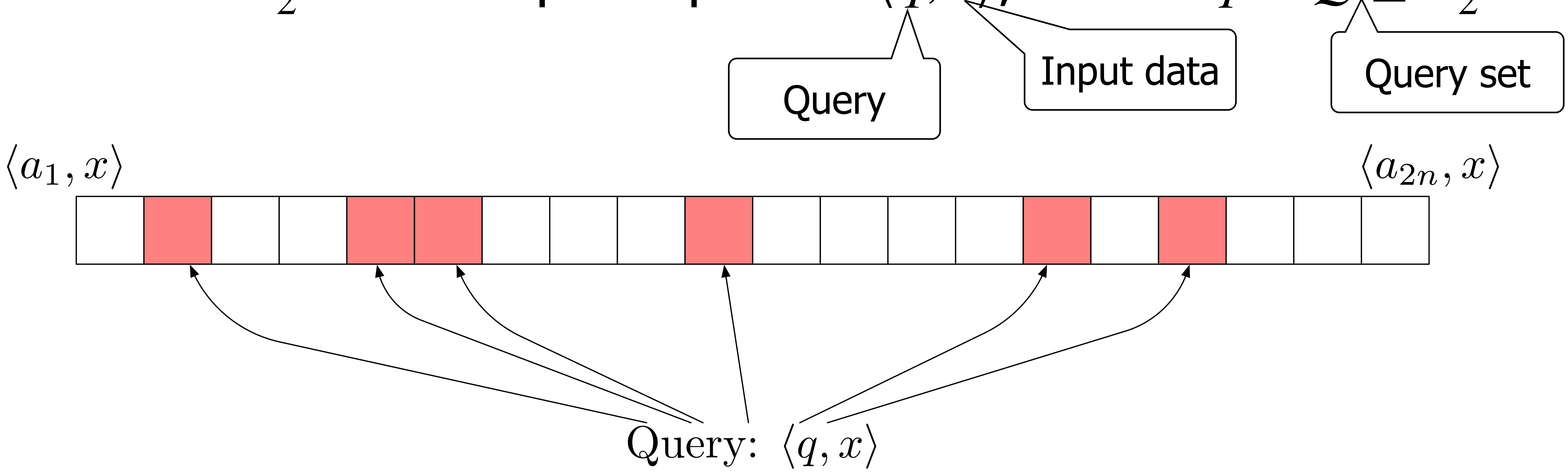
Input data

Query set

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



Store: $\langle a_1, x \rangle, \ldots, \langle a_{2n}, x \rangle$

Query time: number of accesses on ay query

Linear Data Structures

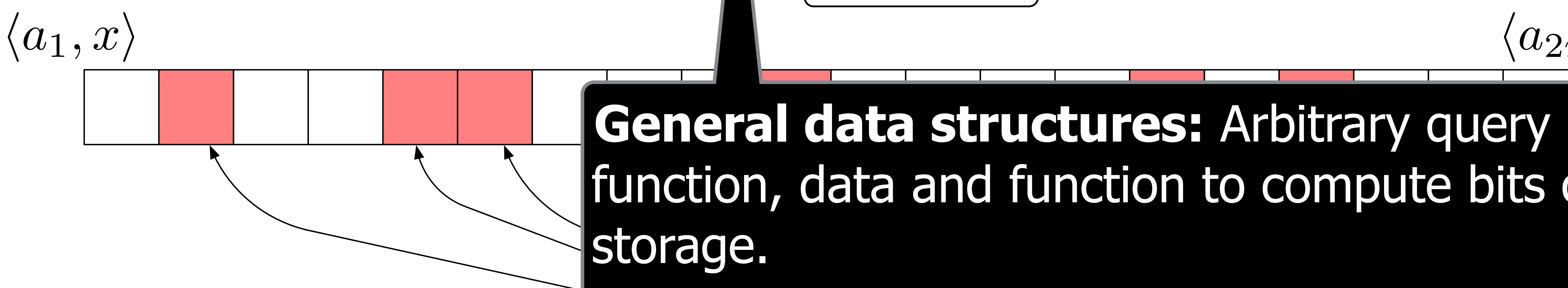
[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Query

Input data

Query set



General data structures: Arbitrary query function, data and function to compute bits of storage.

For example, the input data can be a graph and the query corresponds to finding shortest paths between pairs of vertices.

Store: $\langle a_1, x \rangle, \ldots, \langle a_{2n}, x \rangle$

Query time: number of accesses on ay query

Linear Data Structures

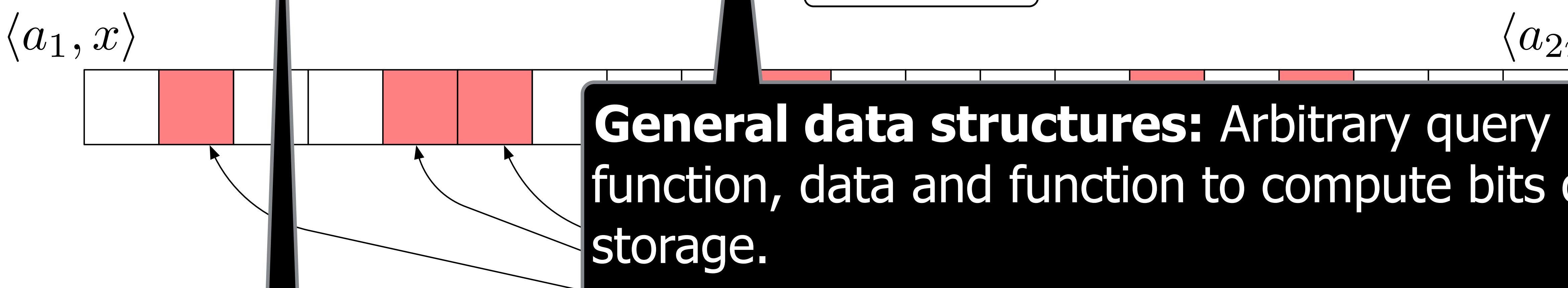
[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Query

Input data

Query set



General data structures: Arbitrary query function, data and function to compute bits of storage.

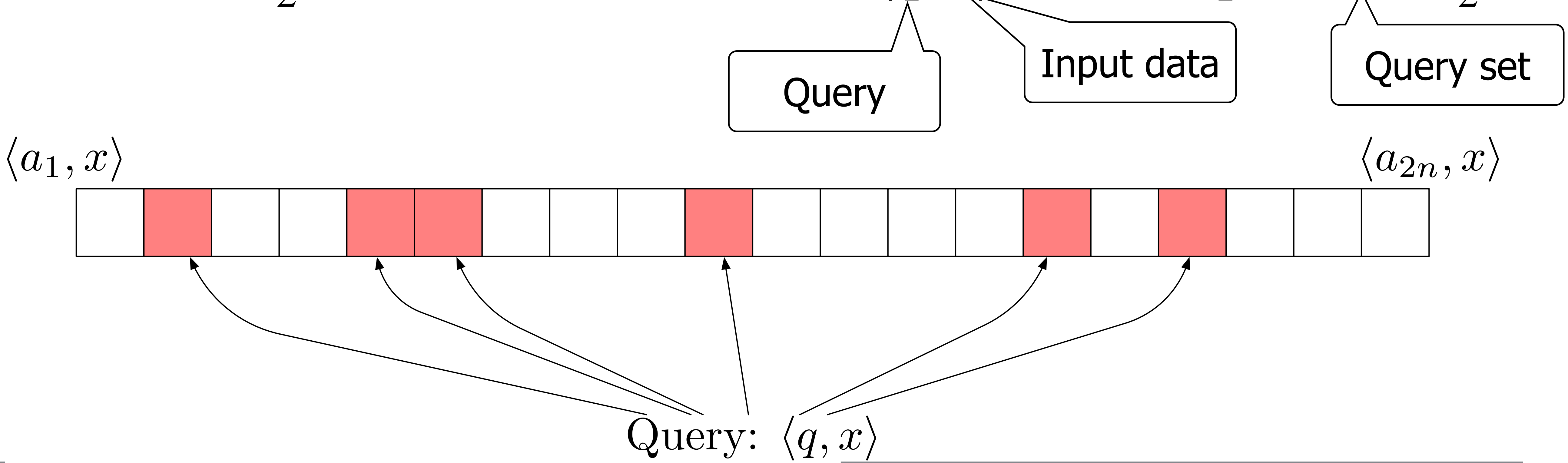
For example, the input data can be a graph and the query corresponds to finding shortest paths between pairs of vertices.

Store: $\langle a_1, x$ … $\ldots, \langle a_{2n}, x \rangle$

Field can be arbitrary and the **space** is not necessarily 2n.

Two Simple Data Structures

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



Store: $\langle q_1, x \rangle, \ldots, \langle q_{|Q|}, x \rangle$

Query time: 1

Store: $\langle e_1, x \rangle, \ldots, \langle e_n, x \rangle$

Query time: n

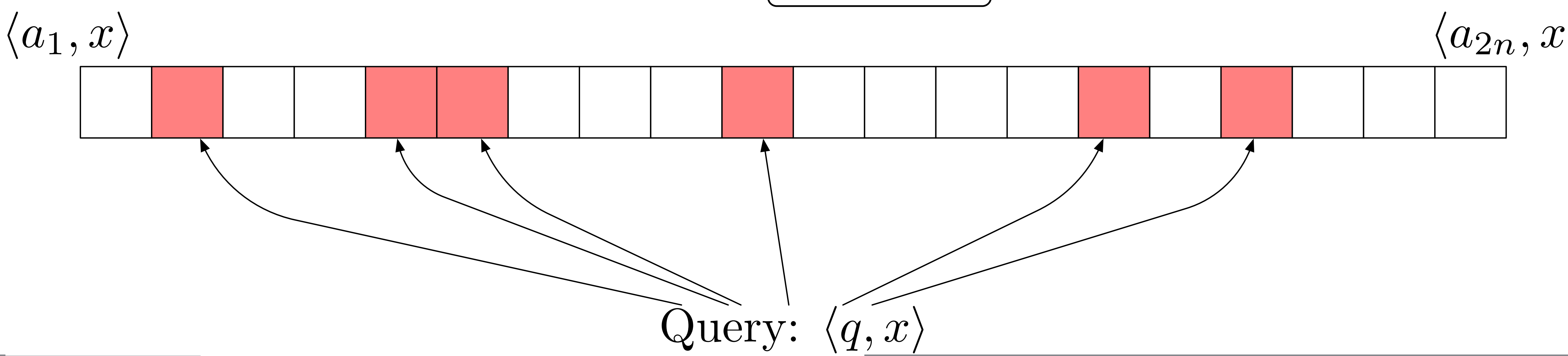
Example Query Sets

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Query

Input data

Query set



$Q = \mathbb{F}_2^n$

Q: Set of all $\sqrt{n} \times \sqrt{n}$ matrices

Q: Hamming weight 2 vectors

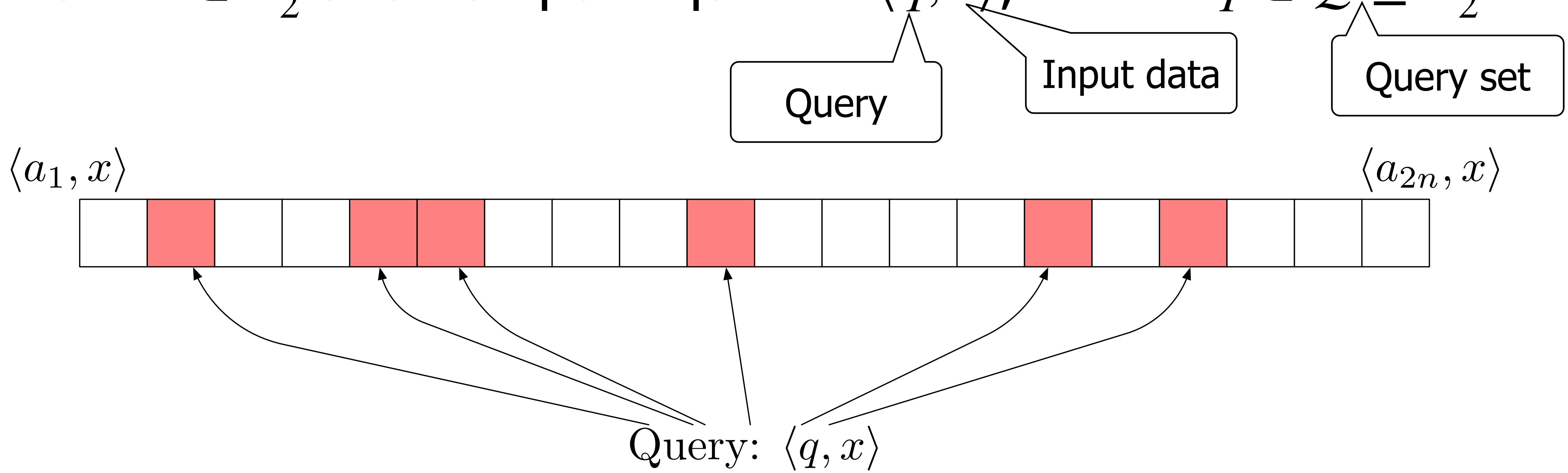
$Q : \{v : v_{\leq i} = 1,\ v_{>i} = 0,\ i \in [n]\}$

(Partial Sums)

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



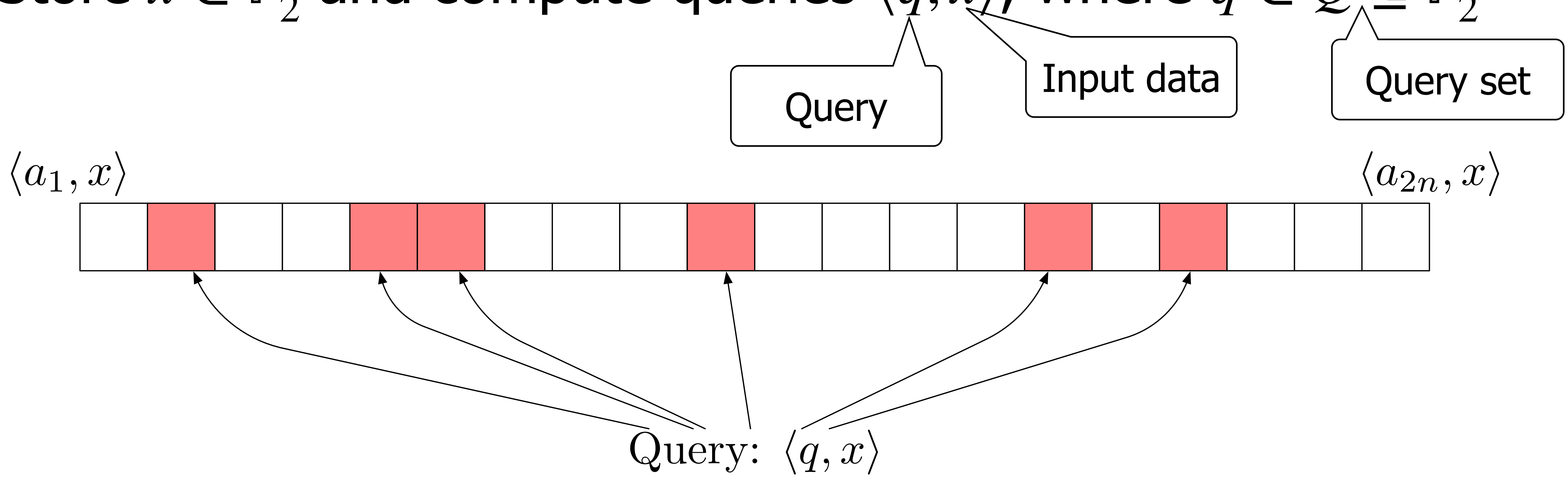
LT(Q, s): smallest query time among space s data structures

Question: relationship between LT(Q, s), $|Q|$, n, s

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



LT(Q, s): smallest query time among space s data structures

Question: relationship between LT(Q, s), $|Q|$, n, s

Best known lower bound on LT(Q, s) for explicit Q: $\dfrac{\log |Q|}{\log(s/n)}$

Linear Data Structures

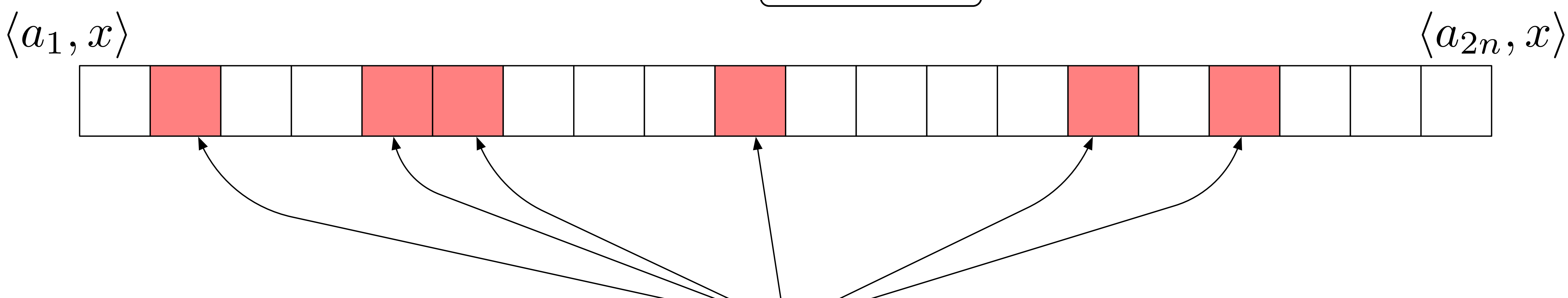
[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

Query

Input data

Query set



This is the case for general data structures as well!
[Sie04, Pat08, PTW10, Lar12]

LT(Q, s): smallest

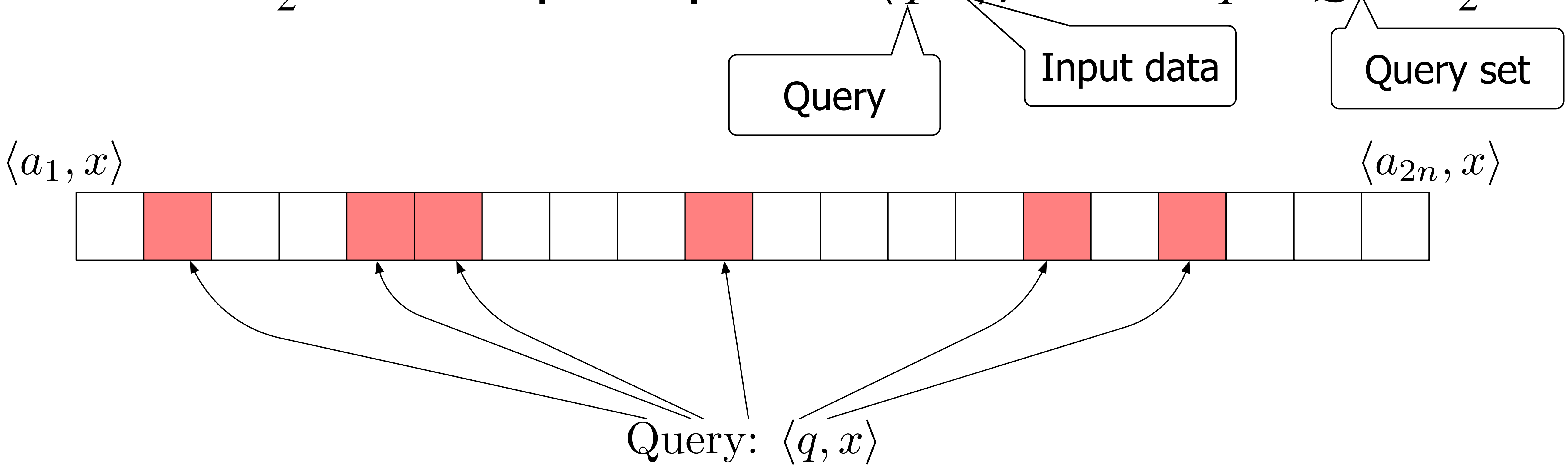
Question: relationship between LT(Q, s), $|Q|$, n, s

Best known lower bound on LT(Q, s) for explicit Q: $\dfrac{\log |Q|}{\log(s/n)}$

Linear Data Structures

[Yao81, Fre81]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



LT(Q, s): smallest query time among space s data structure

Question: relationship between LT(Q, s), $|Q|$, n, s

Best known lower bound on LT$(Q, 2n)$ for explicit Q: $\log |Q|$

Rectangular Rigidity [APY09]

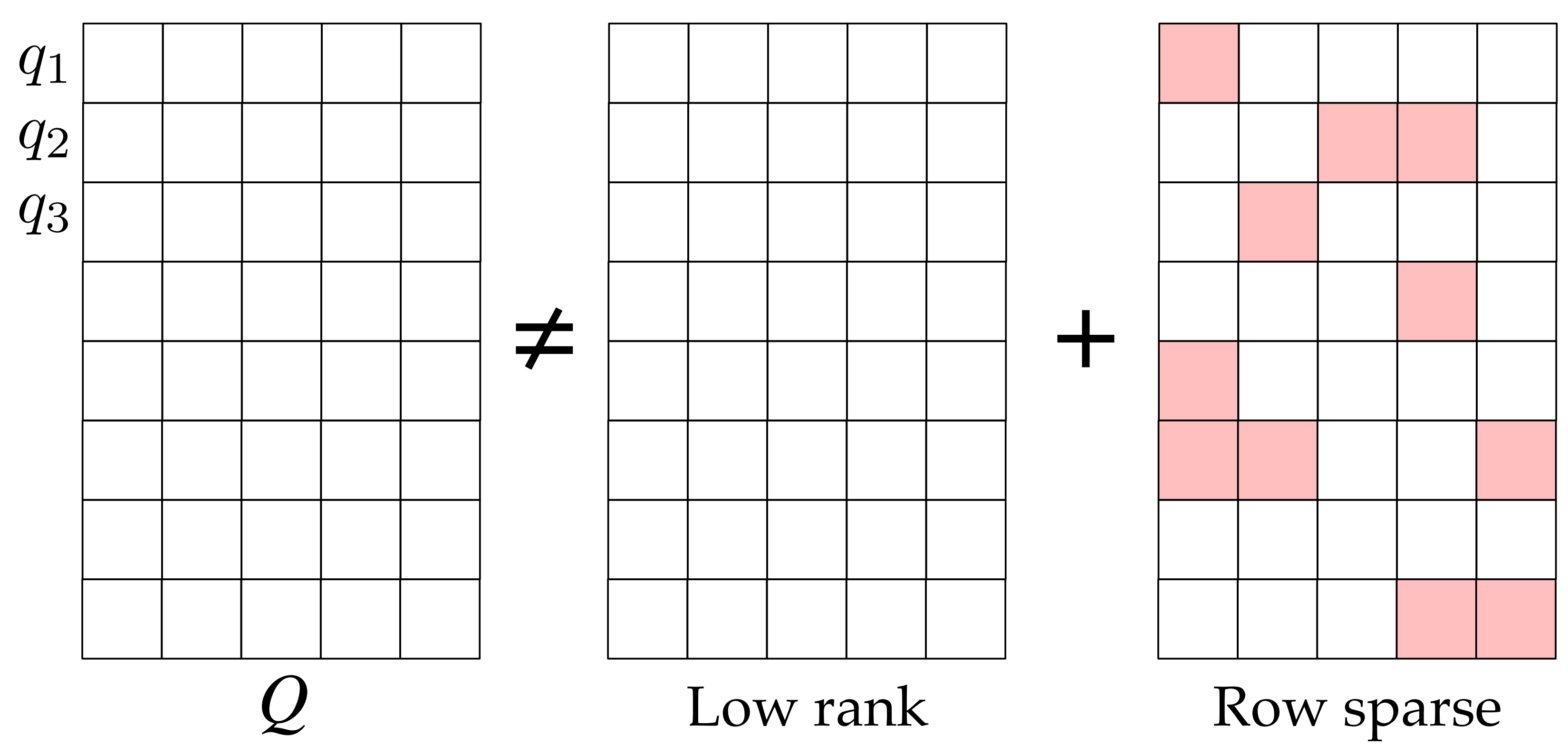
Want a subset $Q \subseteq \mathbb{F}_2^n$ that is "far-away" from any low dimensional subspace

- "small" random sets have this property
- interested in explicit sets

Rectangular Rigidity [APY09]

Want a subset $Q \subseteq \mathbb{F}_2^n$ that is "far-away" from any low dimensional subspace

- "small" random sets have this property
- interested in explicit sets

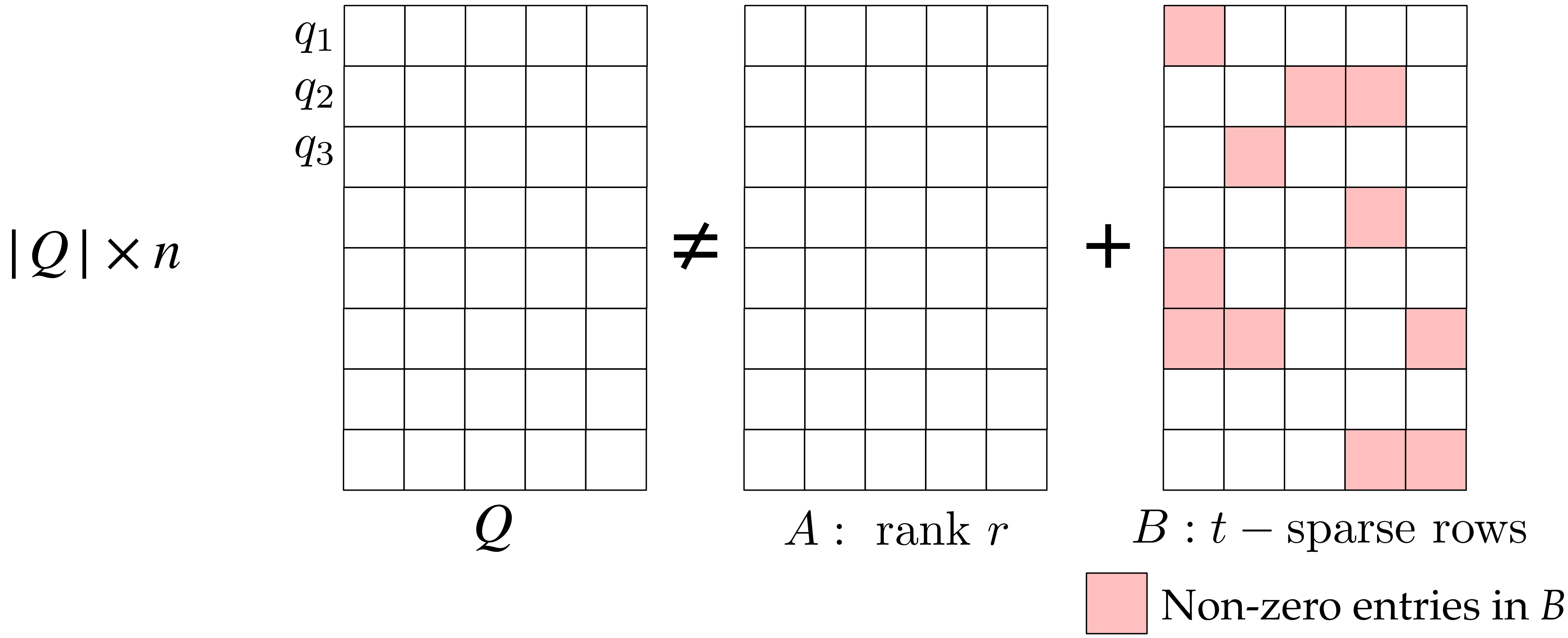


Rectangular Rigidity [APY09]

$Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid if for every r-dim subspace V, there is a point $q \in Q$ at a Hamming distance of at least t from V.

Rectangular Rigidity [APY09]

$Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid if for every r-dim subspace V, there is a point $q \in Q$ at a Hamming distance of at least t from V.



Rectangular Rigidity [APY09]

$Q \subseteq \mathbb{F}_2^n$ is (r,t)-rigid if for every r-dim subspace V, there is a point $q \in Q$ at a Hamming distance of at least t from V.

Question: relationship between $r, t, n, |Q|$

Rectangular Rigidity [APY09]

$Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid if for every r-dim subspace V, there is a point $q \in Q$ at a Hamming distance of at least t from V.

Question: relationship between $r, t, n, |Q|$

Best known lower bounds for explicit Q [Fri93, SSS97, APY09, AC15]:

$$t \geq \min \left\{ \frac{n}{r} \cdot \log\left(\frac{|Q|}{r}\right), n \right\}$$

Rectangular Rigidity [APY09]

$Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid if for every r-dim subspace V, there is a point $q \in Q$ at a Hamming distance of at least t from V.

Question: relationship between $r, t, n, |Q|$

Best known lower bounds for explicit Q [Fri93, SSS97, APY09, AC15]:

$$t \geq \log|Q|/n \text{ for } r = n/2.$$

[DGW18]: Linear DS Lower Bounds imply Rigidity Lower Bounds

Theorem:
If $Q \subseteq \mathbb{F}_2^n$ is explicit and $\mathsf{LT}(Q, 2n) = \omega\left(\log|Q| \cdot \log n\right)$, then there exists a semi-explicit $\left(\mathsf{P}^{\mathsf{NP}}\right)$ rigid set $Q' \subseteq \mathbb{F}_2^{n'}$ that is $\left(n'/2, \omega\left(\log|Q|\right)\right)$ rigid.

[DGW18]: Linear DS Lower Bounds imply Rigidity Lower Bounds

Theorem:
If $Q \subseteq \mathbb{F}_2^n$ is explicit and $\mathsf{LT}(Q, 2n) = \omega\left(\log|Q| \cdot \log n\right)$, then there exists a semi-explicit $\left(\mathsf{P}^{\mathsf{NP}}\right)$ rigid set $Q' \subseteq \mathbb{F}_2^{n'}$ that is $\left(n'/2,\ \omega\left(\log|Q|\right)\right)$ rigid.

$\log|Q|$ is the best known lower bound for explicit sets

$\log|Q|$ is the best known lower bound for explicit sets

Key Lemma:

For $t \geq \log n$, if $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exits Q' such that

(a) dim(Q') $\leq n/2$

(b) $\mathsf{LT}(Q,\ 2n) < \mathsf{LT}(Q',\ n) + t/\log n$.

Key Lemma:

For $t \geq \log n$, if $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exits Q' such that

(a) dim(Q') $\leq n/2$

(b) LT$(Q,\ 2n) \leq$ LT$(Q',\ n) + t/\log n$.

1. Recurse until Q' is a constant dimension set.
2. In the process, either a rigid set is found, OR
3. LT$(Q,\ 2n) \leq t + O(1)$

Key Lemma:

For $t \geq \log n$, if $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exits Q' such that

(a) dim(Q') $\leq n/2$

(b) LT$(Q,\ 2n) \leq$ LT$(Q',\ n) + t/\log n$.

1. Recurse until Q’ is a constant dimension set.
2. In the process, either a rigid set is found, OR
3. LT$(Q,\ 2n) \leq t + O(1)$

Exercise to show Q’ can be obtained by a poly time algorithm with access to an NP oracle

Key Lemma:

If $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exists a Q' such that

(a) dim(Q') $\leq n/2$

(b) LT$(Q,\ 2n) <$ LT$(Q',\ n) + t/\log n$.

Proof:

Key Lemma:
If $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exists a Q' such that
(a) dim(Q') $\leq n/2$
(b) LT$(Q,\ 2n) <$ LT$(Q',\ n) + t/\log n$.

Proof:

Since Q not rigid, there is a subspace V of dim n/2 such that for $q \in Q$, $q = v_q + u_q$ satisfying $v_q \in V$ and $d_H(u_q) < t/\log n$.

Key Lemma:

If $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exists a Q' such that

(a) dim(Q') $\leq n/2$

(b) LT$(Q,\ 2n) <$ LT$(Q',\ n) + t/\log n$.

Proof:

Since Q not rigid, there is a subspace V of dim n/2 such that for $q \in Q$, $q = v_q + u_q$ satisfying $v_q \in V$ and $d_H(u_q) < t/\log n$.

Define $Q' = \{v_q \,|\, q \in Q\}$

Key Lemma:

If $Q \subseteq \mathbb{F}_2^n$ is not $\left(n/2,\ t/\log n\right)$-rigid, then there exists a Q' such that

(a) dim(Q') $\leq n/2$

(b) LT$(Q,\ 2n) <$ LT$(Q',\ n) + t/\log n$.

Proof:

Since Q not rigid, there is a subspace V of dim n/2 such that for $q \in Q$, $q = v_q + u_q$ satisfying $v_q \in V$ and $d_H(u_q) < t/\log n$.

Define $Q' = \{v_q \,|\, q \in Q\}$

Compute

$\langle q, x\rangle = \langle v_q, x\rangle + \langle u_q, x\rangle$

→ $\langle v_q, x\rangle$ with LT$(Q',\ n)$ queries and n space

→ $\langle u_q, x\rangle$ with $t/\log n$ queries and n space storing $\langle e_1, x\rangle, \ldots, \langle e_n, x\rangle$

[DGW18]: Linear DS Lower Bounds imply Rigidity Lower Bounds

Theorem:
If $Q \subseteq \mathbb{F}_2^n$ is explicit and $\mathsf{LT}(Q, 2n) = \omega\left(\log|Q| \cdot \log n\right)$, then there exists a semi-explicit $\left(\mathsf{P}^{\mathsf{NP}}\right)$ rigid set $Q' \subseteq \mathbb{F}_2^{n'}$ that is $\left(n'/2,\, \omega\left(\log|Q|\right)\right)$ rigid.

Remarks:

1. [DGW18]'s proof is more general. Shows a connection between **inner and outer dimension** of matrices, measures defined by [PP06].
2. [DGW18]'s prove connections between other types of data structures and rigidity.

Systematic Linear Model

[Val92, GM07, CKL18]

Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$

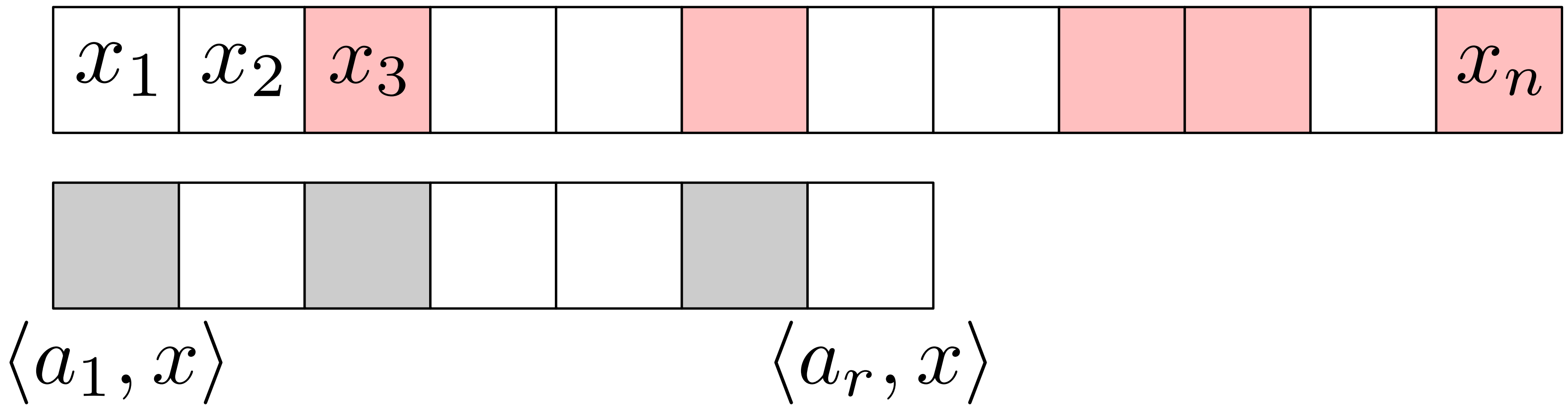
Query

Input data

Systematic Linear Model

[Val92, GM07, CKL18]

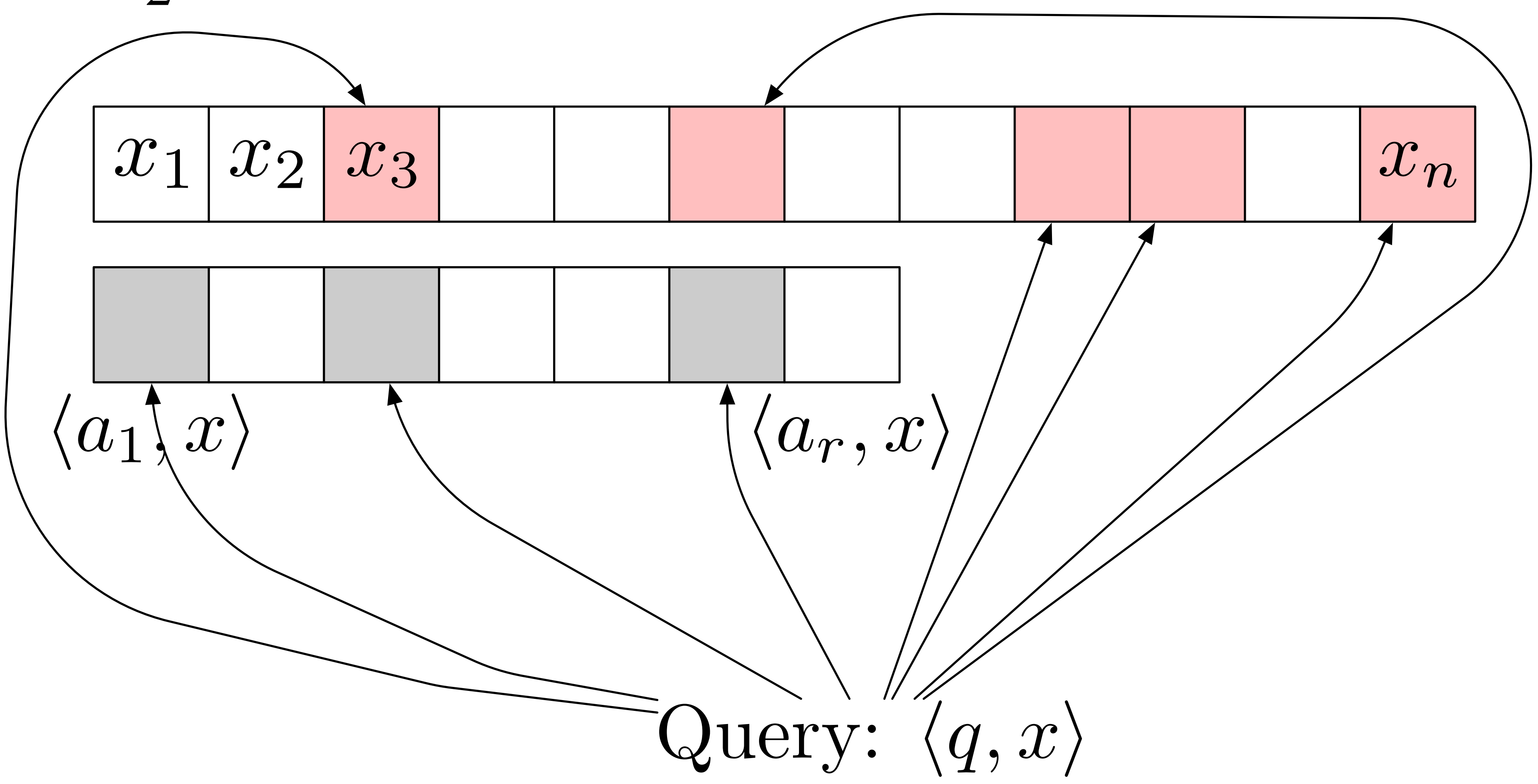
Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



Systematic Linear Model

[Val92, GM07, CKL18]

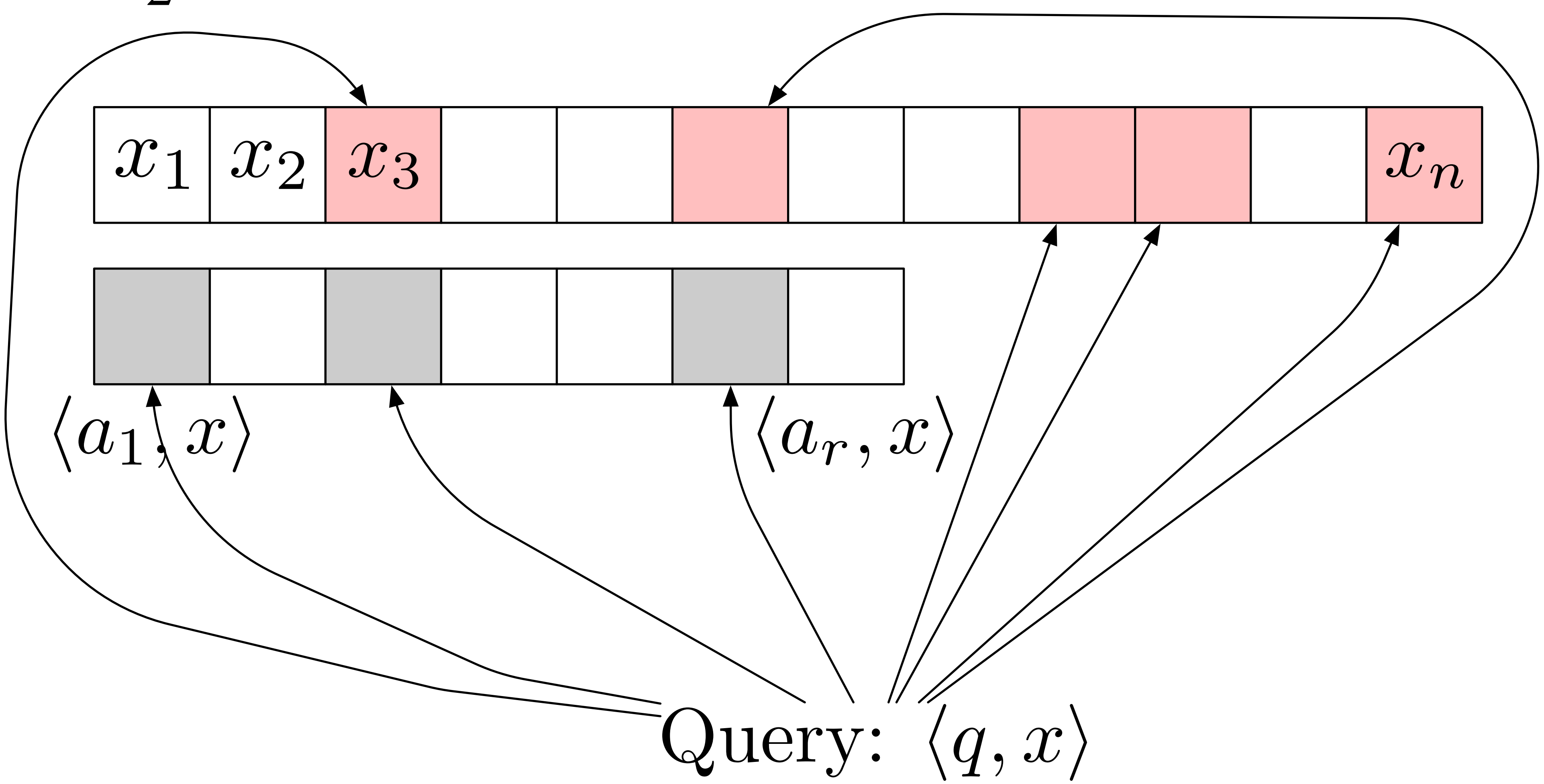
Store $x \in \mathbb{F}_2^n$ and **compute queries** $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



Systematic Linear Model

[Val92, GM07, CKL18]

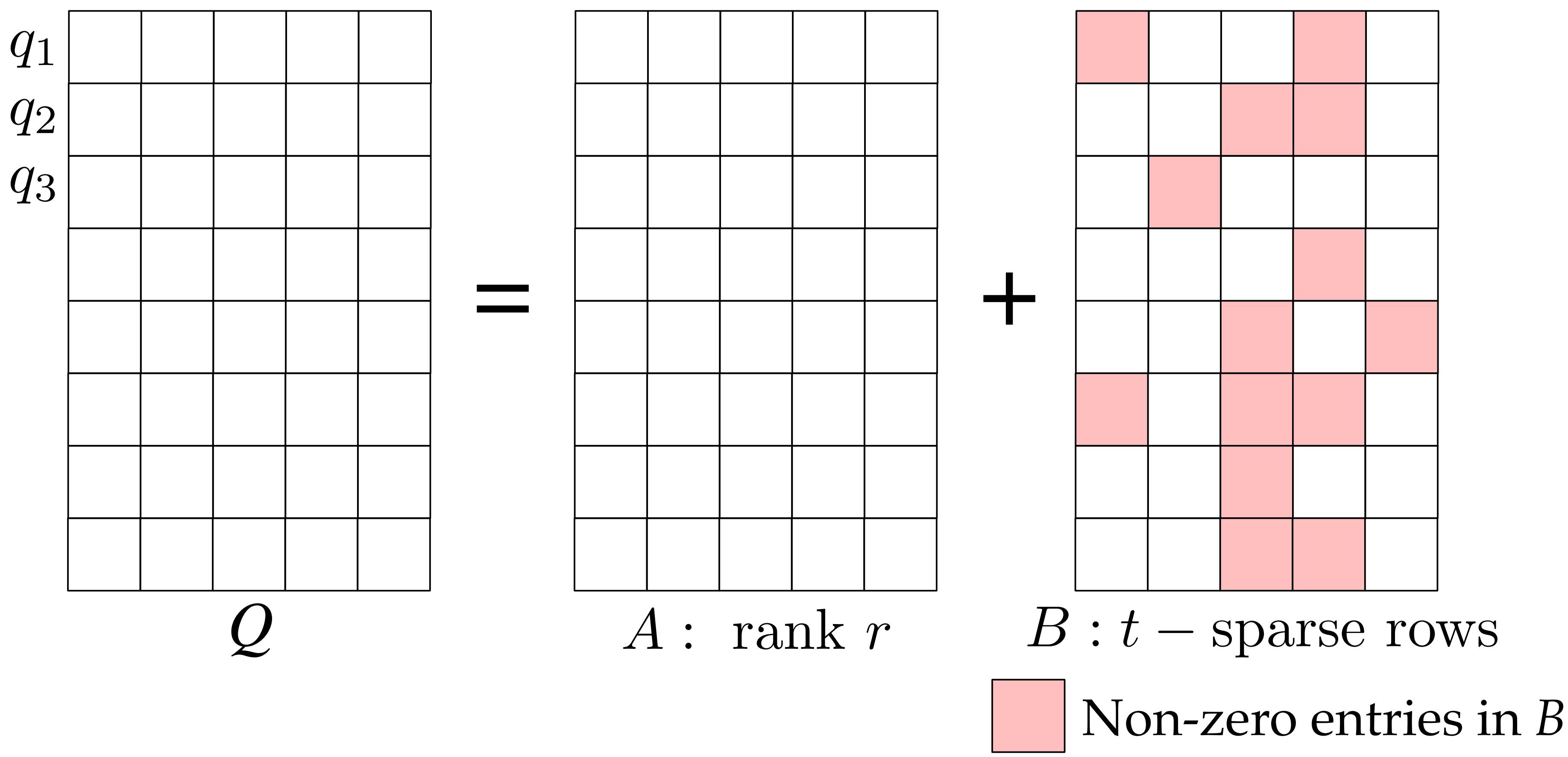
Store $x \in \mathbb{F}_2^n$ and compute queries $\langle q, x \rangle$, where $q \in Q \subseteq \mathbb{F}_2^n$



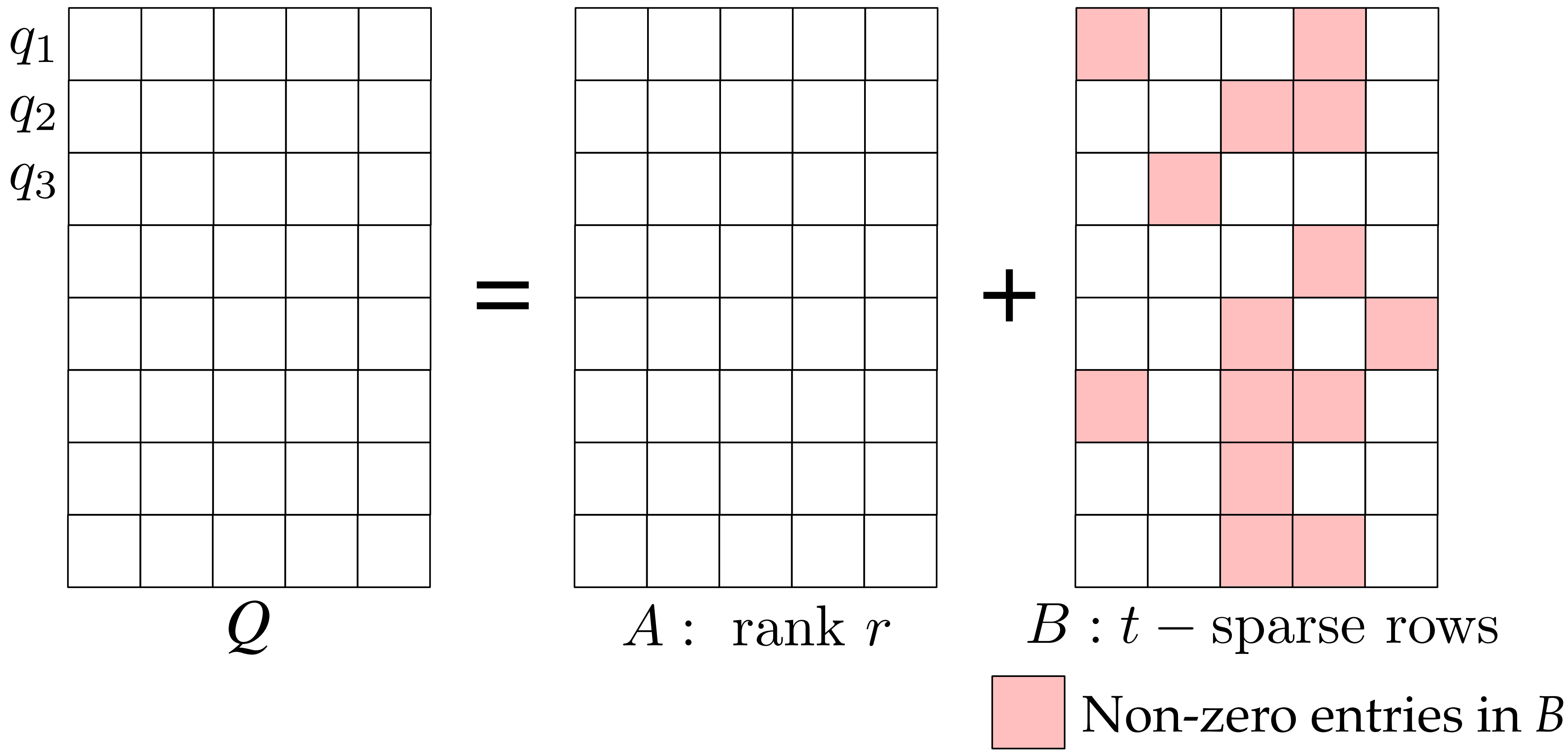
Redundant bits: $\langle a_1, x \rangle, \ldots, \langle a_r, x \rangle$

Query time: number of accesses to x

A Data Structure Upper Bound

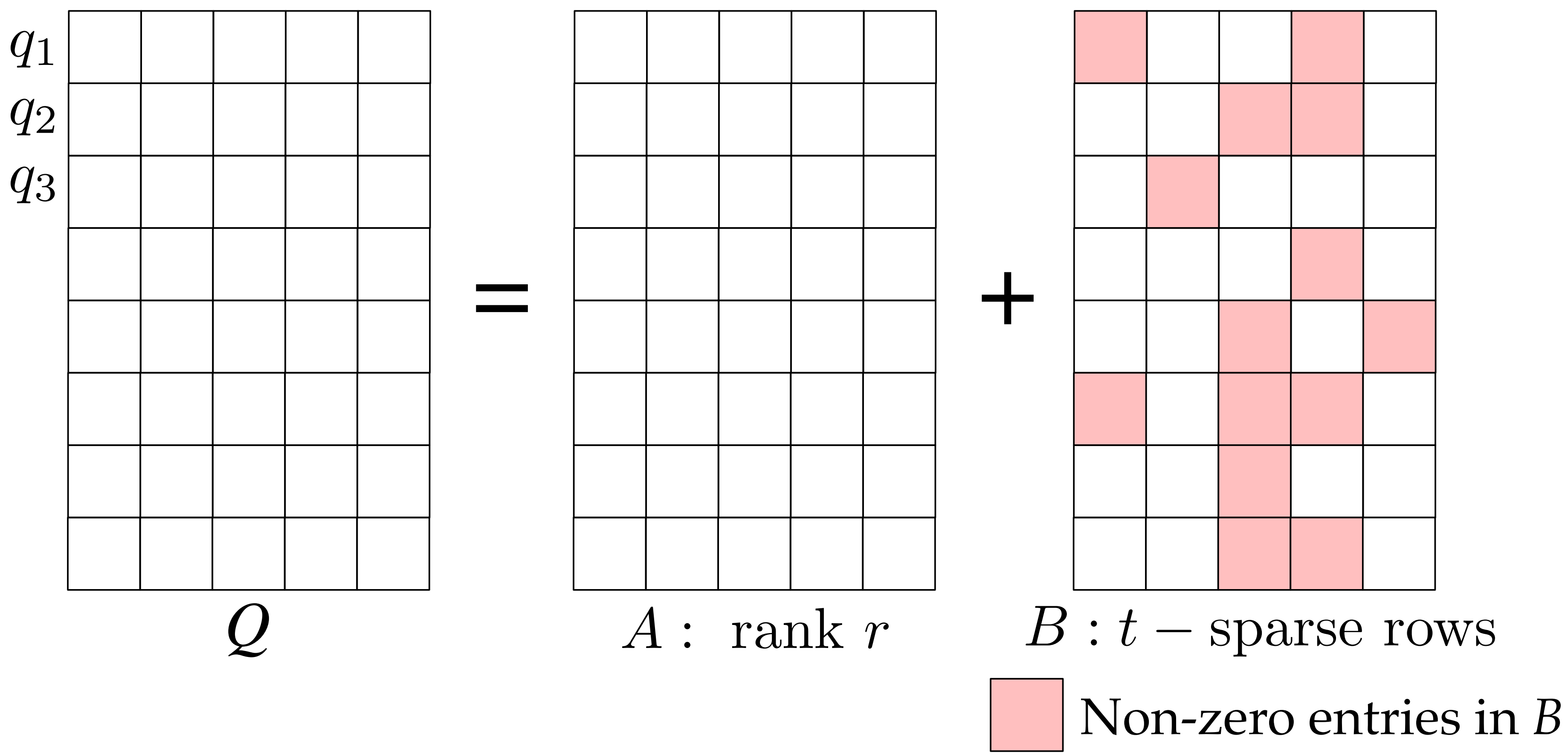


A Data Structure Upper Bound



Redundant bits: $\langle a_1^*, x\rangle, \ldots, \langle a_r^*, x\rangle$, where $a_1^*, \ldots, a_r^*$ basis of A

A Data Structure Upper Bound



Redundant bits: $\langle a_1^*, x\rangle, \ldots, \langle a_r^*, x\rangle$, where $a_1^*, \ldots, a_r^*$ basis of A

Queries: $q_i = a_i + b_i$, and $\langle q_i, x\rangle = \langle a_i, x\rangle + \langle b_i, x\rangle$

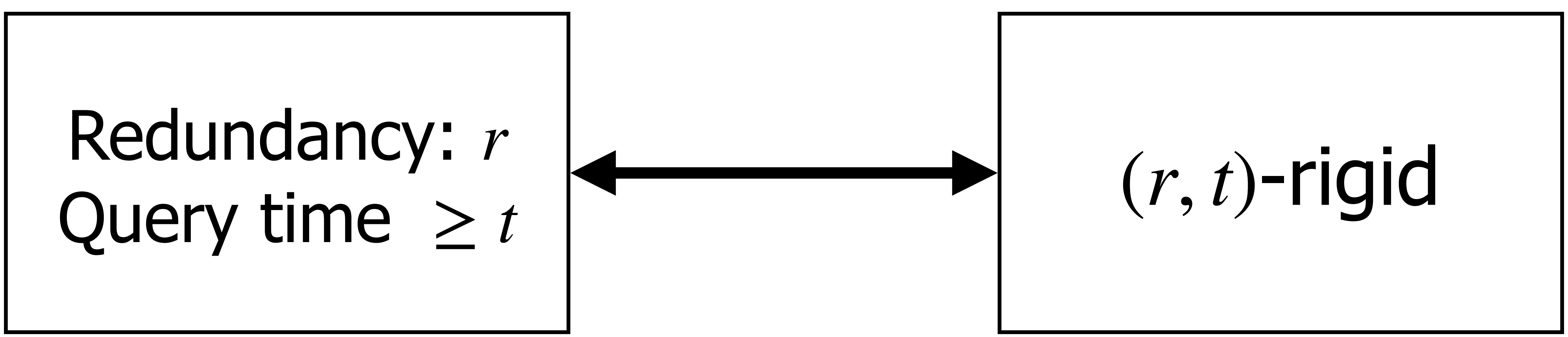
Query time: t accesses to compute $\langle b_i, x\rangle$ (t-sparse)

The Equivalence

Theorem **[NRR19]**: $Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid **iff** every systematic linear data structure with redundancy r has query time at least t

The Equivalence

Theorem **[NRR19]**: $Q \subseteq \mathbb{F}_2^n$ is (r,t)-rigid **iff** every systematic linear data structure with redundancy r has query time at least t



The Equivalence

Theorem **[NRR19]**: $Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid **iff** every systematic linear data structure with redundancy r has query time at least t

Best known systematic linear lower bounds [GM07, CKL18]: $t \geq \min\left\{\frac{n}{r} \cdot \log\left(\frac{|Q|}{r}\right), n\right\}$

The Lower Bound

Theorem **[NRR19]**: If $Q \subseteq \mathbb{F}_2^n$ is (r,t)-rigid then every systematic linear data structure with redundancy r has query time at least t

Proof:

The Lower Bound

Theorem **[NRR19]**: If $Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid then every systematic linear data structure with redundancy r has query time at least t

Proof:

Redundant bits given by $\langle a_1, x\rangle, \ldots, \langle a_r, x\rangle$, and $U = \text{span}(a_1, \ldots, a_r)$

q^* at a distance of t from $U = \text{span}(a_1, \ldots, a_r)$

The Lower Bound

Theorem [NRR19]: If $Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid then every systematic linear data structure with redundancy r has query time at least t

Proof:

Redundant bits given by $\langle a_1, x \rangle, \ldots, \langle a_r, x \rangle$, and $U = \text{span}(a_1, \ldots, a_r)$

q^* at a distance of t from $U = \text{span}(a_1, \ldots, a_r)$

Claim: Query time to compute $\langle q^*, x \rangle$ is at least t

The Lower Bound

Theorem **[NRR19]**: If $Q \subseteq \mathbb{F}_2^n$ is (r, t)-rigid then every systematic linear data structure with redundancy r has query time at least t

Proof:

Redundant bits given by $\langle a_1, x \rangle, \ldots, \langle a_r, x \rangle$, and $U = \text{span}(a_1, \ldots, a_r)$

q^* at a distance of t from $U = \text{span}(a_1, \ldots, a_r)$

Claim: Query time to compute $\langle q^*, x \rangle$ is at least t

Query accesses $\langle a_1, x \rangle, \ldots, \langle a_r, x \rangle$ and $\langle e_{i_1}, x \rangle, \ldots, \langle e_{i_k}, x \rangle$

If $U' = \text{span}(a_1, \ldots, a_r, e_{i_1}, \ldots, e_{i_k})$, then $d(q^*, U') = 0$. So $k \geq t$

Consequences for Linear Data Structures

Theorem [NRR19]: If $Q \subseteq \mathbb{F}_2^n$ and $\mathsf{LT}(Q, 2n) \geq \omega\left(\sqrt{\log|Q| \cdot n}\right)$, then there exists a $Q' \subseteq \mathbb{F}_2^k$ that is $(k/2, \log|Q'|)$-rigid.

Consequences for Linear Data Structures

Theorem [NRR19]: If $Q \subseteq \mathbb{F}_2^n$ and $\mathsf{LT}(Q, 2n) \geq \omega\left(\sqrt{\log |Q| \cdot n}\right)$, then there exists a $Q' \subseteq \mathbb{F}_2^k$ that is $(k/2, \log |Q'|)$-rigid.

In contrast to [DGW18], need to prove polynomial lower bounds on the query time but will result in explicit rigid matrices.

Consequences for Linear Data Structures

Theorem **[NRR19]**: If $Q \subseteq \mathbb{F}_2^n$ and $\mathrm{LT}(Q, 2n) \geq \omega\left(\sqrt{\log |Q| \cdot n}\right)$, then there exists a $Q' \subseteq \mathbb{F}_2^k$ that is $(k/2, \log |Q'|)$-rigid.

Observation: If there is a systematic linear data structure for Q with redundancy r and query time t, then $\mathrm{LT}(Q, n + r) \leq t$.

Vector-Matrix-Vector Problem

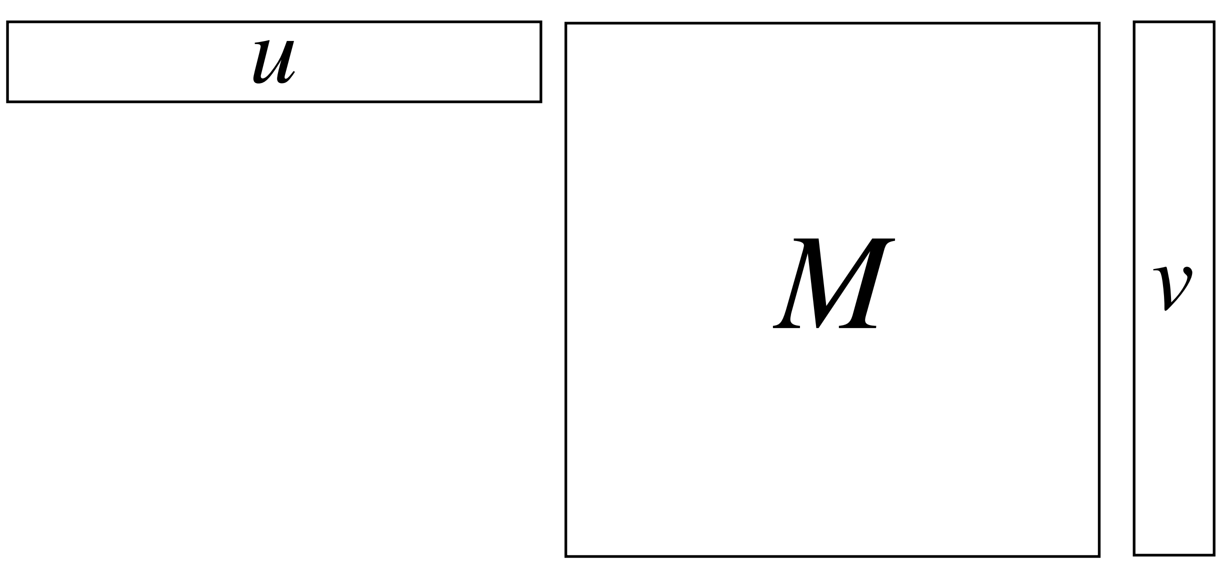
Goal is to design an efficient data structure that

a) stores a $\sqrt{n} \times \sqrt{n}$-bit matrix M
b) compute queries $u^{\intercal} M v \pmod 2$, where u, v are $\sqrt{n}$-bit vectors

Vector-Matrix-Vector Problem

Goal is to design an efficient data structure that

a) stores a $\sqrt{n} \times \sqrt{n}$-bit matrix M
b) compute queries $u^\intercal M v \pmod{2}$, where u, v are $\sqrt{n}$-bit vectors

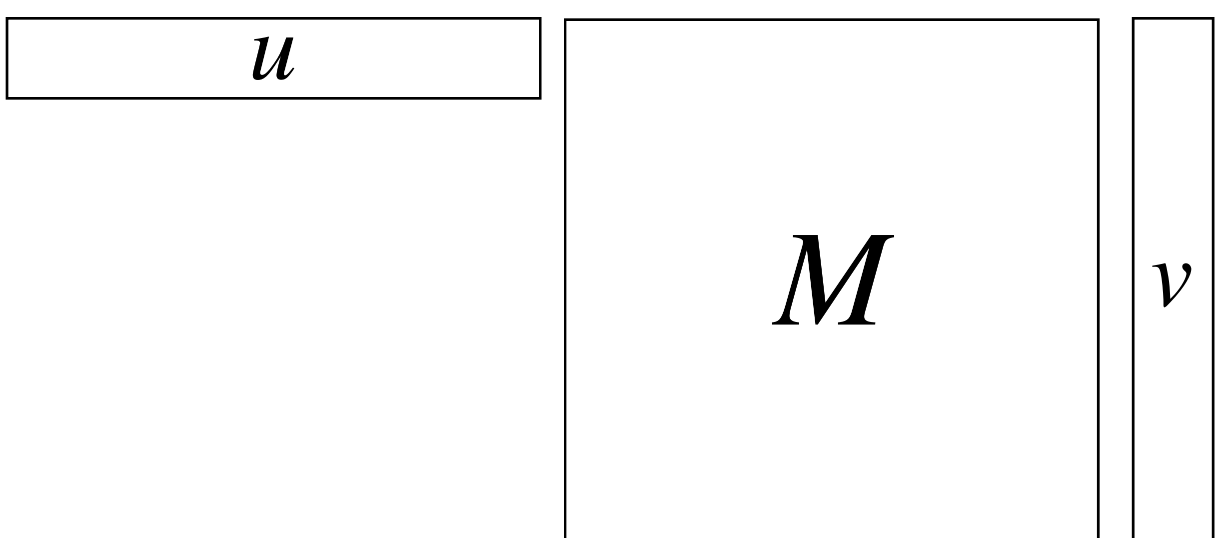


Vector-Matrix-Vector Problem

Goal is to design an efficient data structure that

a) stores a $\sqrt{n} \times \sqrt{n}$-bit matrix M

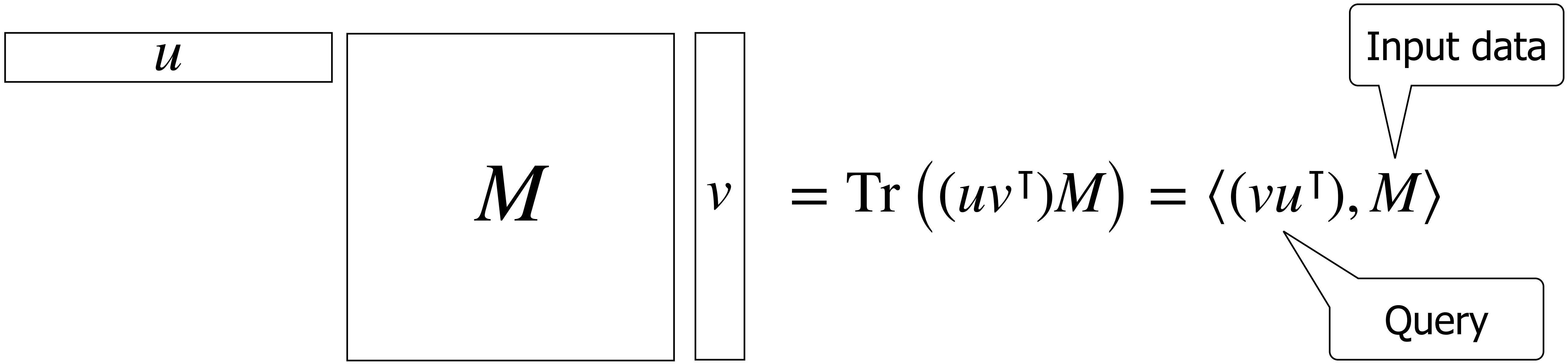
b) compute queries $u^{\intercal}Mv \pmod{2}$, where u, v are $\sqrt{n}$-bit vectors



$$= \mathrm{Tr}\left((uv^{\intercal})M\right) = \langle (vu^{\intercal}), M \rangle$$

Systematic Linear Model

[Val92,GM07,CKL18]



Redundant bits: $\langle A_1, M\rangle, \ldots, \langle A_r, M\rangle$

Query time: number of accesses to M

Rigidity Lower Bound for Vec-Mat-Vec

Q = the set of vectors corresponding to $\{uv^\intercal\}$.

Theorem **[NRR19]**: If Q is (r, t)-rigid, then

$$t \geq \min\left\{n^{1.5}/r, n\right\}$$

(A $\log n$ factor improvement over [CKL18])

Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \leq 2r/\sqrt{n}$, and

(b) $d(M, V) \geq \Omega(n)$

Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \le 2r/\sqrt{n}$, and

(b) $d(M, V) \ge \Omega(n)$

1. $M = M_1 + \ldots + M_k$, where $k = 2r/\sqrt{n}$ and each M_i is rank 1.
2. Since $d(M, V) \ge \Omega(n)$, there is an M_i such that $d(M_i, V) \ge \Omega(n^{1.5}/r)$

Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \leq 2r/\sqrt{n}$, and

(b) $d(M, V) \geq \Omega(n)$

Proof:

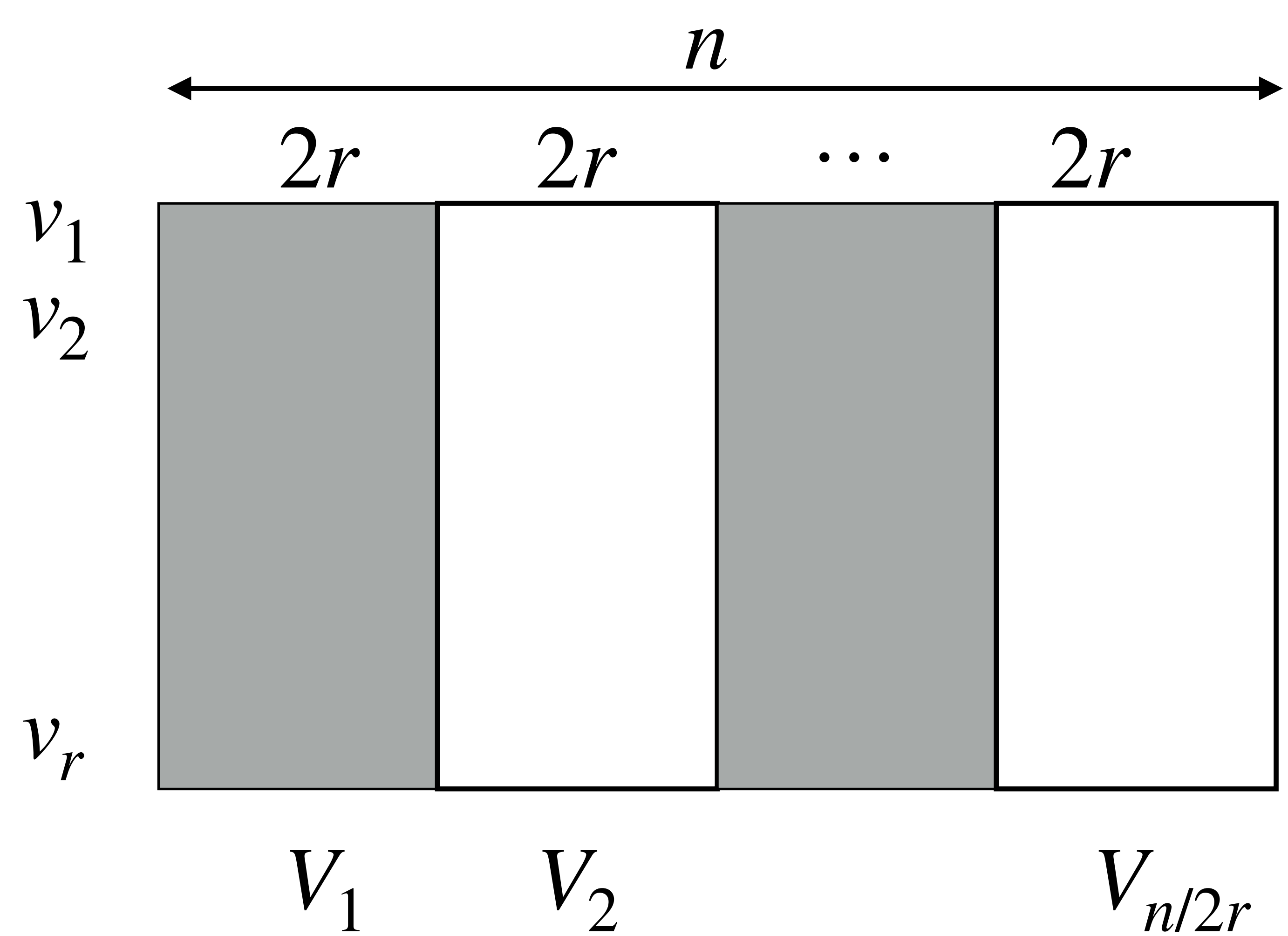
V = $\text{span}(v_1, \ldots, v_r)$

Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with
(a) $\text{rank}(M) \leq 2r/\sqrt{n}$, and
(b) $d(M, V) \geq \Omega(n)$

Proof:

V = $\text{span}(v_1, \ldots, v_r)$

Each V_i is a projection of V to 2r coordinates,
And $\dim(V_i) \leq r$

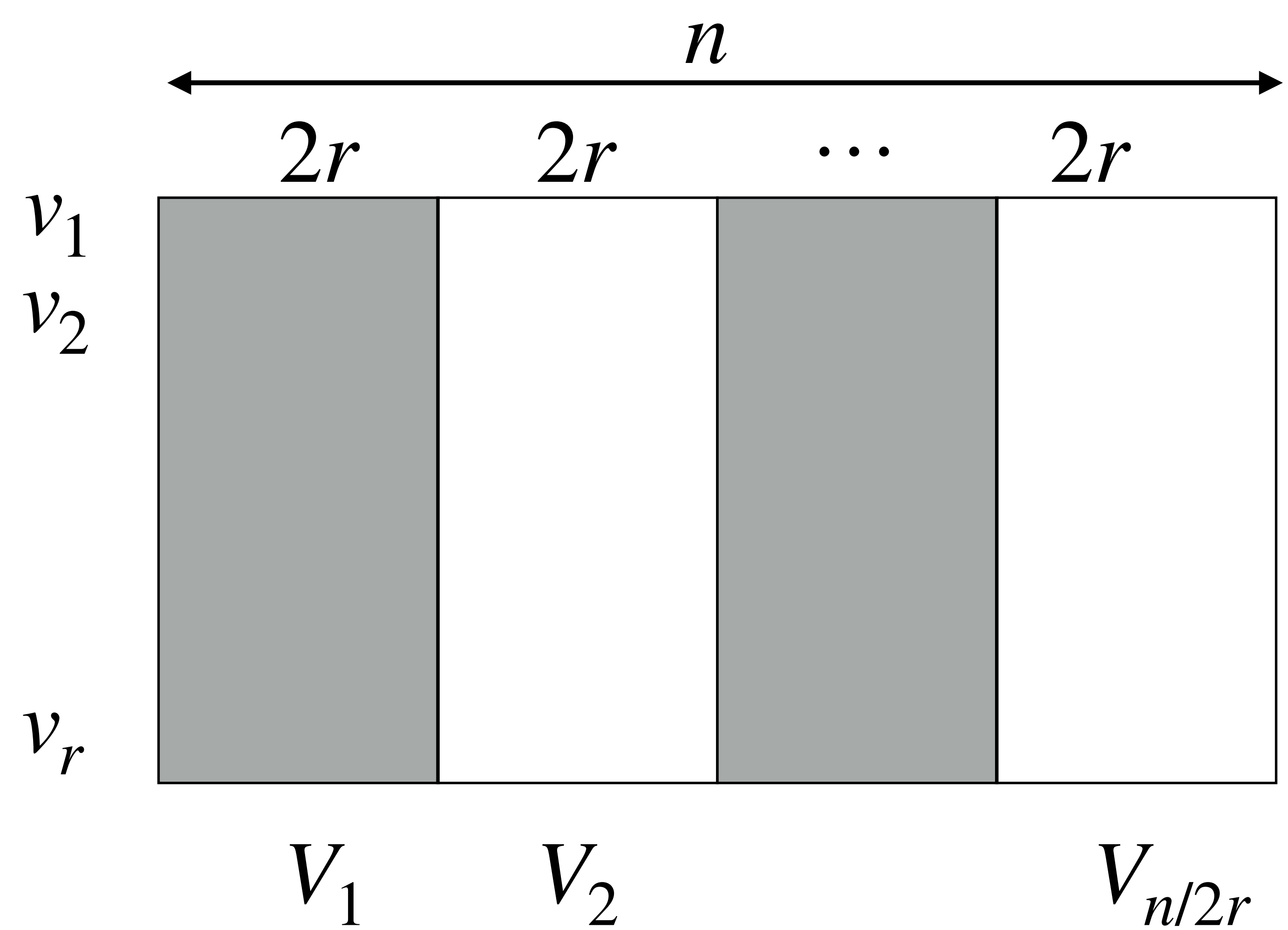


Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \le 2r/\sqrt{n}$, and

(b) $d(M, V) \ge \Omega(n)$

Proof:



V = $\text{span}(v_1, \ldots, v_r)$

Each V_i is a projection of V to 2r coordinates,
And $\dim(V_i) \le r$

Claim: There exists a $v' \in \mathbb{F}_2^{2r}$ such that $d(v', V_i) \ge \Omega(r)$

Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \leq 2r/\sqrt{n}$, and

(b) $d(M, V) \geq \Omega(n)$

Proof:

There exists a $v' \in \mathbb{F}_2^{2r}$ such that $d(v', V_i) \geq \Omega(r)$

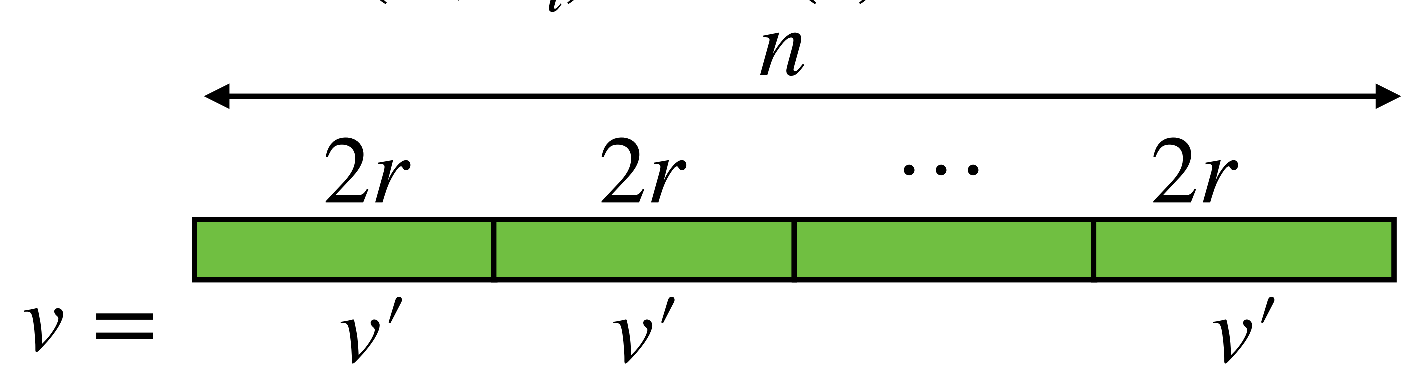
Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \le 2r/\sqrt{n}$, and

(b) $d(M, V) \ge \Omega(n)$

Proof:

There exists a $v' \in \mathbb{F}_2^{2r}$ such that $d(v', V_i) \ge \Omega(r)$



By definition of v, $d(v, V) \ge (n/2r) \cdot \Omega(r) \ge \Omega(n)$

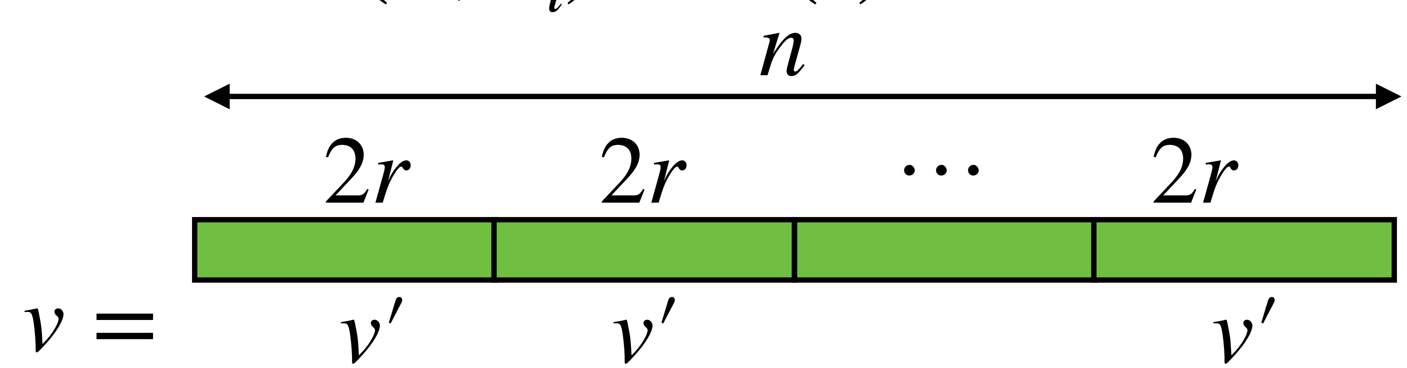
Key Lemma: For every r-dim vector space V, there is a $\sqrt{n} \times \sqrt{n}$ matrix M with

(a) $\text{rank}(M) \leq 2r/\sqrt{n}$, and

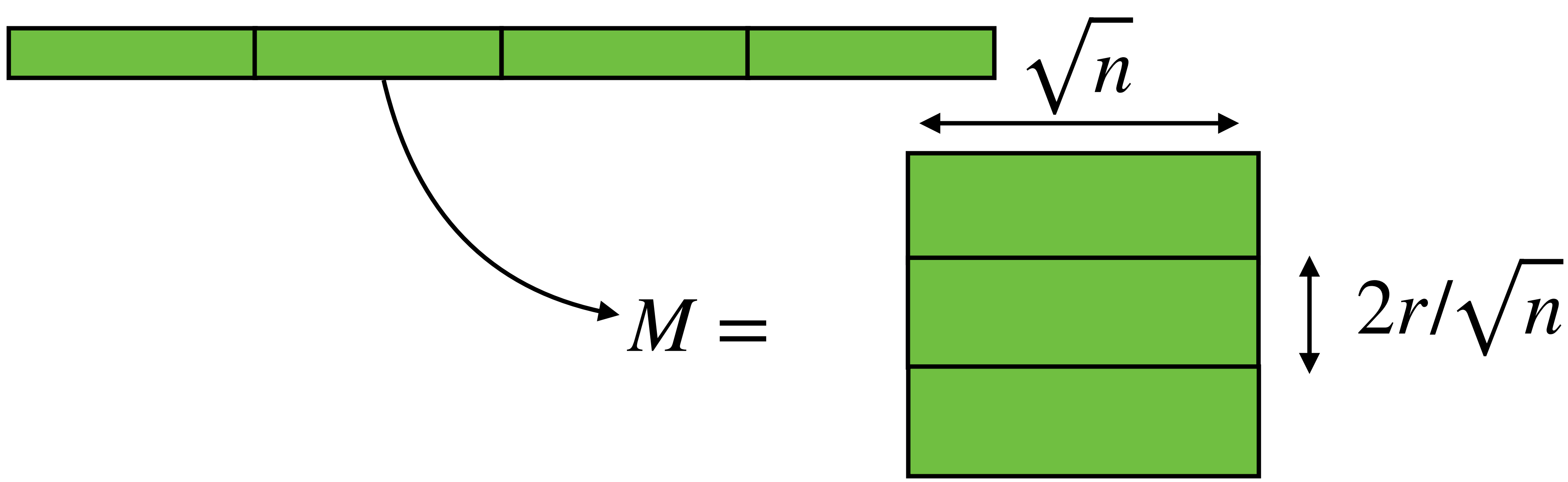
(b) $d(M, V) \geq \Omega(n)$

Proof:

There exists a $v' \in \mathbb{F}_2^{2r}$ such that $d(v', V_i) \geq \Omega(r)$



By definition of v, $d(v, V) \geq (n/2r) \cdot \Omega(r) \geq \Omega(n)$



Rigidity Lower Bound for Vec-Mat-Vec

Q = the set of vectors corresponding to $\{uv^\intercal\}$.

Theorem **[NRR19]**: If Q is (r, t)-rigid, then

$$t \geq \min\left\{n^{1.5}/r, n\right\}$$

(A $\log n$ factor improvement over [CKL18])

Open Questions

(a) Prove better rigidity bounds for $\{uv^\intercal\}$

(b) Improve explicitness guarantee in [DGW18]

(c) What are the rigidity implications for $\omega\left(\log|Q|\right)$ lower bounds for linear data structures?

Thank you