Mean-Based Trace Reconstruction over Oblivious Synchronization Channels

Alexandra Veliche

joint work with Mahdi Cheraghchi, Joseph Downs, and João Ribeiro

Trace Reconstruction

Problem: using as few i.i.d. traces $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t$ as possible, recover $\boldsymbol{x}$ with high probability.

Trace Reconstruction

message $\boxed{x} \longrightarrow \boxed{\text{Channel}} \longrightarrow \boxed{y}$ trace

Problem: using as few i.i.d. traces $\boldsymbol{y}_1, \dots, \boldsymbol{y}_t$ as possible, recover $\boldsymbol{x}$ with high probability via algorithm $\mathcal{A}$.

(worst-case) For any $\boldsymbol{x} \in \{-1,1\}^n$, $\mathbb{P}_{y_1,\dots,y_t}[\mathcal{A}(y_1, \dots, y_t) = x] \approx 1.$

(average-case) $\mathbb{P}_{\substack{\boldsymbol{x} \leftarrow \{-1,1\}^n \\ y_1,\dots,y_t}}[\mathcal{A}(y_1, \dots, y_t) = x] \approx 1.$

Channel Models for Best Known Bounds

discrete memoryless synchronization

∪

oblivious synchronization

∪ ∪

deletion

geometric insertion-deletion

Upper Bounds on Trace Number

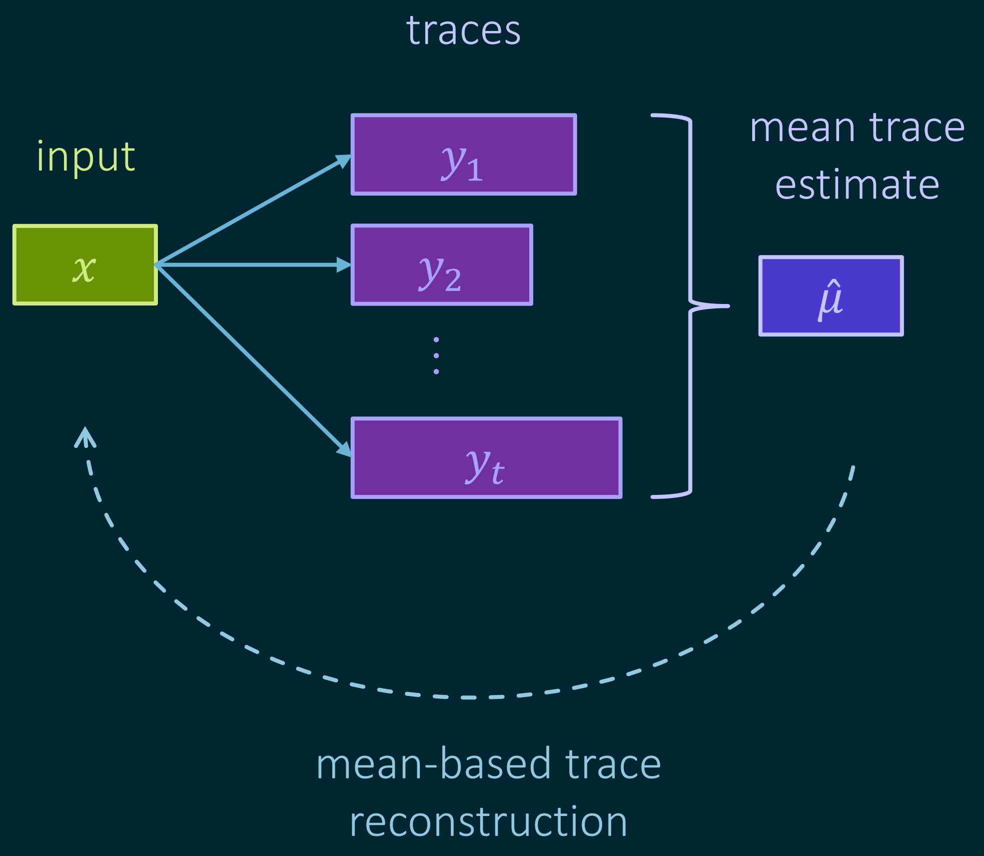
Work	Channel Model	Sufficient Number of Traces
[Nazarov, Peres, 2017]	geometric insertion-deletion	$\exp\left(O\left(n^{1/3}\right)\right)$
[De, O'Donnell, Servedio, 2017]	deletion	$\exp\left(O\left(n^{1/3}\right)\right)$
[Chase, 2022]	deletion	$\exp\left(\tilde{O}\left(n^{1/5}\right)\right)$
this work	oblivious synchronization	$\exp\left(O\left(n^{1/3}\right)\right)$
?	discrete memoryless synchronization	?

Upper Bounds on Trace Number

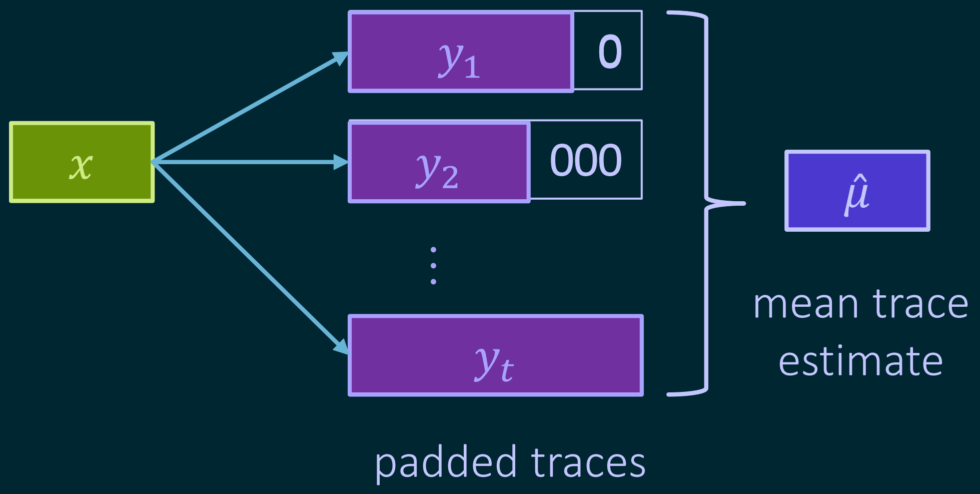
Work	Channel Model	Sufficient Number of Traces
[Nazarov, Peres, 2017]	geometric insertion-deletion	$\exp\left(O\left(n^{1/3}\right)\right)$
[De, O'Donnell, Servedio, 2017]	deletion	$\exp\left(O\left(n^{1/3}\right)\right)$
~~[Chase, 2022]~~	~~deletion~~	$\cancel{\exp\left(\tilde{O}\left(n^{1/5}\right)\right)}$
this work	oblivious synchronization	$\exp\left(O\left(n^{1/3}\right)\right)$
?	discrete memoryless synchronization	?

These all use mean-based trace reconstruction!

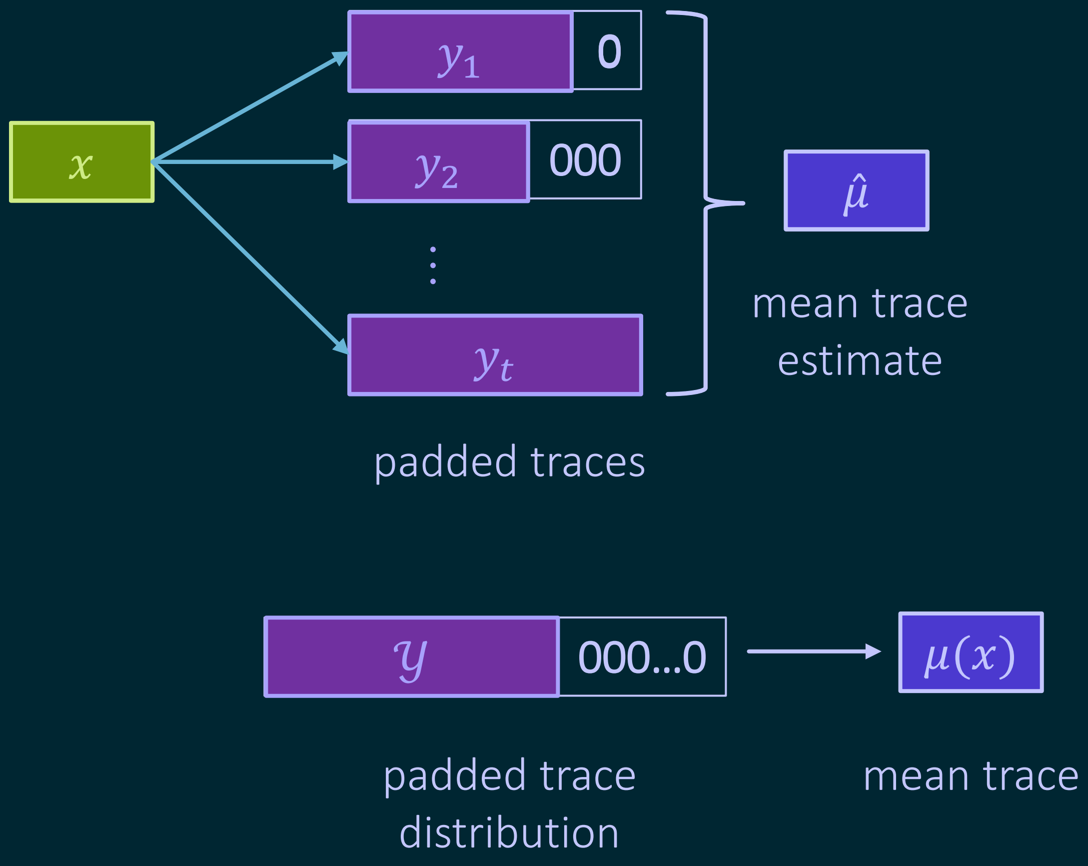
Mean-Based Trace Reconstruction (MBTR)



Mean-Based Trace Reconstruction



Mean-Based Trace Reconstruction



Mean-Based Trace Reconstruction

For $\boldsymbol{x} \in \{-1,1\}^n$, there is a corresponding padded trace distribution: $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2, \ldots)||00 \ldots 0$.

The *mean trace* of x is $\mu(x) = (\mathbb{E}[\mathcal{Y}_1], \mathbb{E}[\mathcal{Y}_2], \ldots)$.

MBTR: Given t i.i.d. traces $y_1, \ldots, y_t \leftarrow \mathcal{Y}$,

1. Compute the estimate $\hat{\mu} = (\widehat{\mu_1}, \widehat{\mu_2}, \ldots)$ of $\mu(x)$, where $\widehat{\mu_i} \coloneqq \frac{1}{t}\sum_{j=1}^{t} y_j$.
2. Find $x^* \in \{-1,1\}^n$ that minimizes $||\mu(x^*) - \hat{\mu}||_1$. Output x^*.

Mean-Based Trace Reconstruction

For $\boldsymbol{x} \in \{-1,1\}^n$, there is a corresponding padded trace distribution: $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2, \dots)||00 \dots 0$.

The *mean trace* of x is $\mu(x) = (\mathbb{E}[\mathcal{Y}_1], \mathbb{E}[\mathcal{Y}_2], \dots)$.

MBTR: Given t i.i.d. traces $y_1, \dots, y_t \leftarrow \mathcal{Y}$,

1. Compute the estimate $\hat{\mu} = (\widehat{\mu_1}, \widehat{\mu_2}, \dots)$ of $\mu(x)$, where $\widehat{\mu_i} := \frac{1}{t}\sum_{j=1}^{t} y_j$.

2. Find $x^* \in \{-1,1\}^n$ that minimizes $||\mu(x^*) - \hat{\mu}||_1$. Output x^*.

For large enough $t = t(n)$, we get $x^* = x$ with high probability (over randomness of traces).

Oblivious Synchronization Channel (OSC)

$Ch_M \in \mathrm{OSC}$ characterized by a random variable M over $\mathbb{Z}_{\geq 0}$ where $\mathbb{P}[M > 0] > 0$.

M corresponds to a collection of randomized functions $\mathcal{F}_M = \{f: \{-1,1\} \to \{-1,1\}^M\}$.

Given input $x \in \{-1,1\}^n$, for each bit x_i:

Sample $m \leftarrow M$.

Sample $f: \{-1,1\} \to \{-1,1\}^m$ from $\mathcal{F}_m$.

Evaluate $f(x_i)$.

Output $f(x_1) || \ldots || f(x_n)$.

Oblivious Synchronization Channel

$Ch_M \in \mathrm{OSC}$ characterized by a random variable M over $\mathbb{Z}_{\geq 0}$ where $\mathbb{P}[M > 0] > 0$.

M corresponds to a collection of randomized functions $\mathcal{F}_M = \{f : \{-1,1\} \to \{-1,1\}^M\}$.

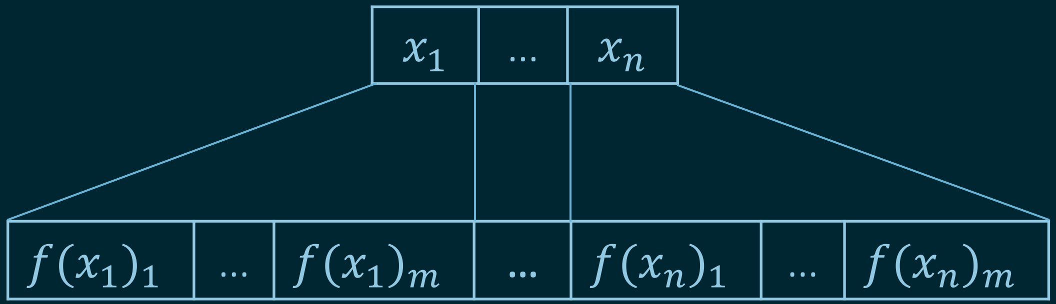
Given input $x \in \{-1,1\}^n$, for each bit x_i:

Sample $m \leftarrow M$.

Sample $f : \{-1,1\} \to \{-1,1\}^m$ from $\mathcal{F}_m$.

Evaluate $f(x_i)$.

Output $f(x_1) || \ldots || f(x_n)$.



Oblivious Synchronization Channel

For each x_i, the function f partitions the output space $[m]$ into

$\mathrm{Rep} \coloneqq \{j : f(-1)_j = -1, f(1)_j = 1\}$ replicated bits

$\mathrm{Flip} \coloneqq \{j : f(-1)_j = 1, f(1)_j = -1\}$ flipped bits

$\mathrm{C}_+ \coloneqq \{j : f(-1)_j = 1, f(1)_j = 1\}$ constant positive bits

$\mathrm{C}_- \coloneqq \{j : f(-1)_j = -1, f(1)_j = -1\}$ constant negative bits

Oblivious Synchronization Channel

For each x_i, the function f partitions the output space $[m]$ into

$\mathrm{Rep} \coloneqq \{j : f(-1)_j = -1, f(1)_j = 1\}$ replicated bits

$\mathrm{Flip} \coloneqq \{j : f(-1)_j = 1, f(1)_j = -1\}$ flipped bits

$\mathrm{C}_+ \coloneqq \{j : f(-1)_j = 1, f(1)_j = 1\}$ constant positive bits

$\mathrm{C}_- \coloneqq \{j : f(-1)_j = -1, f(1)_j = -1\}$ constant negative bits

Ex: $(m = 7)$

$f(-1) =$	1	-1	1	-1	-1	1	1
$f(1) =$	1	1	-1	-1	-1	1	-1

$\{2\} \cup \{3,7\} \cup \{1,6\} \cup \{4,5\} = [7]$

Main Results

Theorem: (informal) Under some mild conditions, $\exp\left(O\left(n^{1/3}\right)\right)$ traces suffice for MBTR over the oblivious synchronization channel with high probability.

Main Results

Theorem: (informal) Under some mild conditions, $\exp\left(O\left(n^{1/3}\right)\right)$ traces suffice for MBTR over the oblivious synchronization channel with high probability.

1. M subexponential.
2. Rep and $Flip$ must differ in a particular way.

Main Results

Theorem: (informal) Under some mild conditions, $\exp\left(O\left(n^{1/3}\right)\right)$ traces suffice for MBTR over the oblivious synchronization channel with high probability.

1. M subexponential. Unknown if necessary. Open problem!
2. Rep and $Flip$ must differ in a particular way. Necessary! Reconstruction impossible otherwise.

Main Results

Theorem: Let $Ch_M \in \mathrm{OSC}$, where M is subexponential. Define random variables W_R, W_F with p.m.f.s

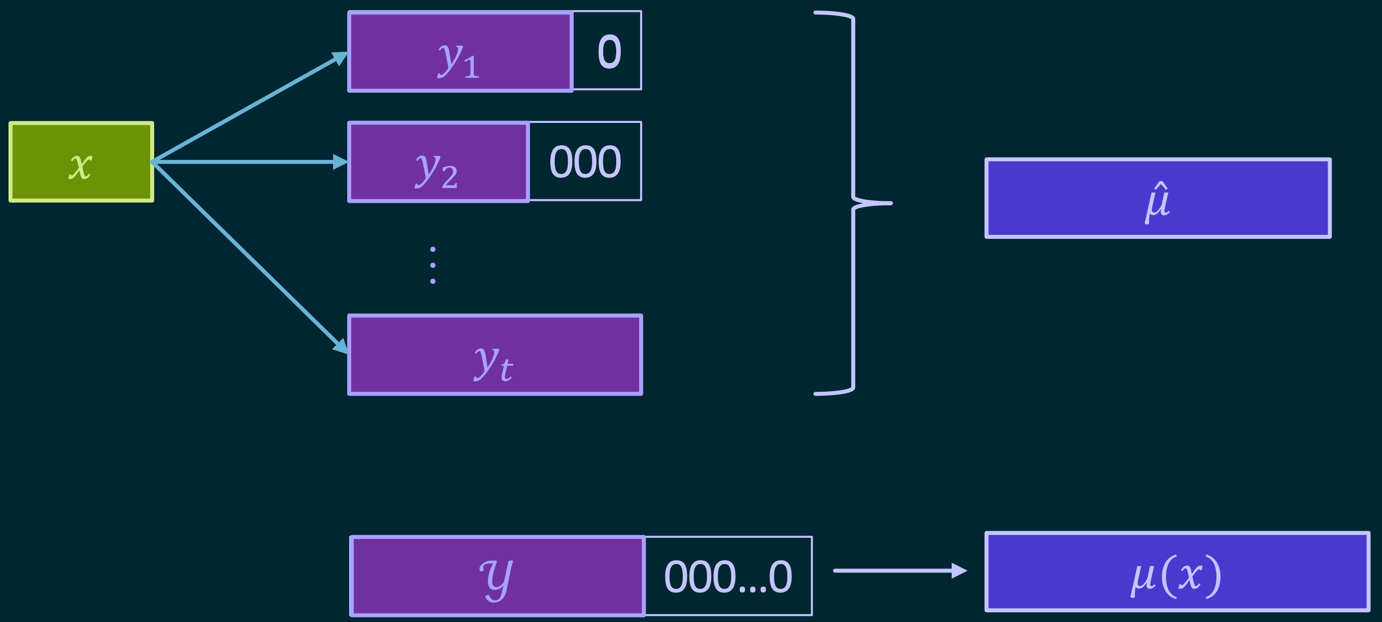
$$W_R(j) \coloneqq \frac{\mathbb{P}[j+1 \in Rep]}{\mathbb{E}[|Rep|]}, W_F(j) \coloneqq \frac{\mathbb{P}[j+1 \in Flip]}{\mathbb{E}[|Flip|]},$$

where $j \in \mathbb{Z}_{\geq 0}$, and p.g.f.s g_{W_R}, g_{W_F}. If $Rep = \emptyset$ (resp. $Flip = \emptyset$), define $g_{W_R} = 0$ (resp. $g_{W_F} = 0$).

If $\mathbb{E}[|Rep|] \cdot g_{W_R}(z) \neq \mathbb{E}[|Flip|] \cdot g_{W_F}(z)$ for some $z \in \mathbb{C}$, then $\exp\left(O\left(n^{1/3}\right)\right)$ traces suffice for MBTR over Ch_M with probability $1 - e^{-\Omega(n)}$. Otherwise, MBTR is impossible.

Proof Sketch

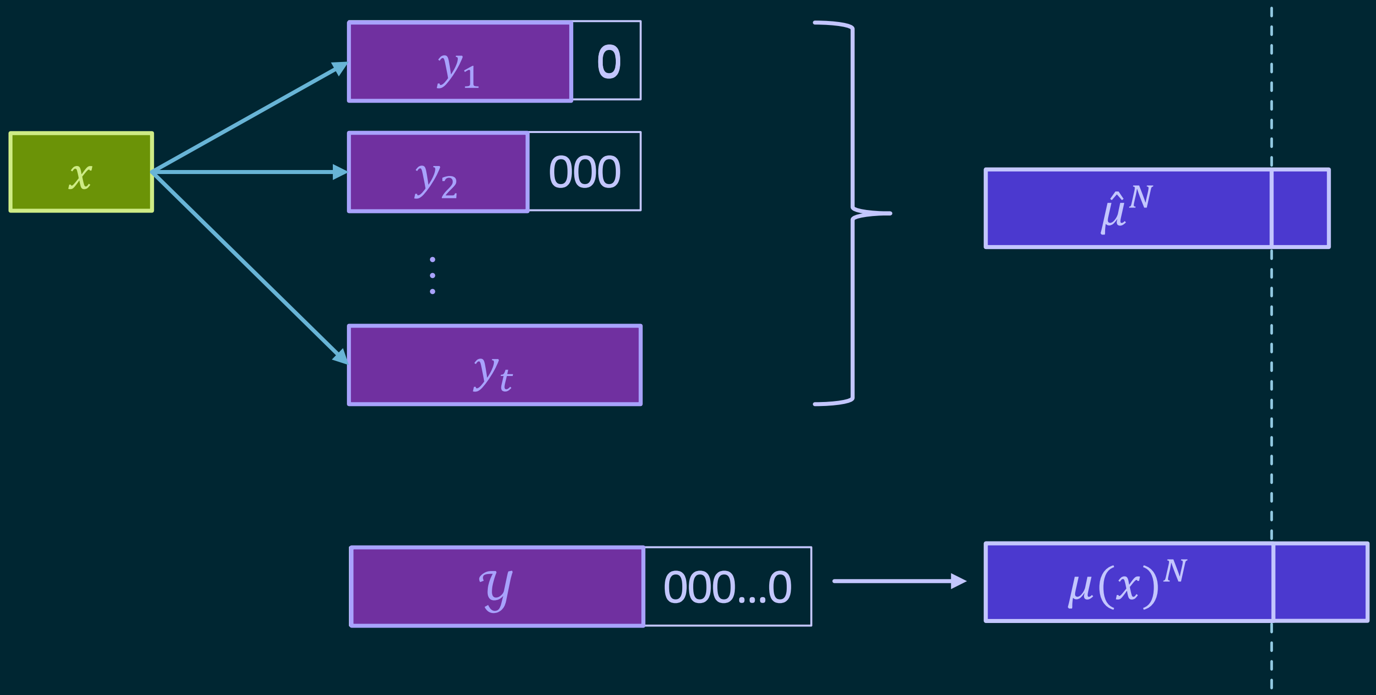
Let x be the input message. Suppose that we are given $t = \exp\left(O\left(n^{1/3}\right)\right)$ traces.



Proof Sketch

Let x be the input message. Suppose that we are given $t = \exp\left(O(n^{1/3})\right)$ traces.

Consider truncation of mean trace $\mu(x)^N$ and its estimate $\hat{\mu}^N$, for some $N \in \mathbb{Z}_+$.



Proof Sketch

Naïve recovery of x from $\hat{\mu}$:

Compute $\mu(x')^N$ for all possible x' and output $\mathbf{argmin}_{x'}\{\|\hat{\mu}^N - \mu(x')^N\|_1\}$.

Proof Sketch

Naïve recovery of x from $\hat{\mu}$:

Compute $\mu(x')^N$ for all possible x' and output $\mathrm{argmin}_{x'}\{\|\hat{\mu}^N - \mu(x')^N\|_1\}$.

Claim: For any $x' \neq x$, $\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \frac{3}{4}\exp\left(-Cn^{1/3}\right)$ with probability $1 - \exp\left(-\Omega(n)\right)$.

Proof Sketch

Naïve recovery of x from $\hat{\mu}$:

Compute $\mu(x')^N$ for all possible x' and output $\mathrm{argmin}_{x'}\{\|\hat{\mu}^N - \mu(x')^N\|_1\}$.

Claim: For any $x' \neq x$, $\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \frac{3}{4}\underbrace{\exp\left(-Cn^{1/3}\right)}_{\delta(n)}$ with probability $1 - \exp\left(-\Omega(n)\right)$.

Proof Sketch

Claim: For any $x' \neq x$, $\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \frac{3}{4}\delta(n)$ with probability $1 - \exp\big(-\Omega(n)\big)$.

$$\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \|\hat{\mu}^N - \mu(x)^N\|_1 - \|\mu(x)^N - \mu(x')^N\|_1 \geq \delta(n) - \frac{1}{4}\delta(n) = \frac{3}{4}\delta(n).$$

Proof Sketch

Claim: For any $x' \neq x$, $\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \frac{3}{4}\delta(n)$ with probability $1 - \exp\big(-\Omega(n)\big)$.

$$\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \|\hat{\mu}^N - \mu(x)^N\|_1 - \|\mu(x)^N - \mu(x')^N\|_1 \geq \delta(n) - \frac{1}{4}\delta(n) = \frac{3}{4}\delta(n).$$

Lemma 1: If x is the input and t traces are given, then

$$\mathbb{P}\left[\, \|\hat{\mu}^N - \mu(x)^N\|_1 \leq \frac{1}{4}\delta(n) \right] \geq 1 - \exp\big(-\Omega(n)\big)$$

Proof Sketch

Claim: For any $x' \neq x$, $\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \frac{3}{4}\delta(n)$ with probability $1 - \exp(-\Omega(n))$.

$$\|\hat{\mu}^N - \mu(x')^N\|_1 \geq \|\hat{\mu}^N - \mu(x)^N\|_1 - \|\mu(x)^N - \mu(x')^N\|_1 \geq \delta(n) - \frac{1}{4}\delta(n) = \frac{3}{4}\delta(n).$$

Lemma 1: If x is the input and t traces are given, then

$$\mathbb{P}\left[\|\hat{\mu}^N - \mu(x)^N\|_1 \leq \frac{1}{4}\delta(n)\right] \geq 1 - \exp(-\Omega(n))$$

Lemma 2: There exists a constant $C > 0$ such that for n large enough, $N = O(n)$, for any $x \neq x'$,

$$\|\mu(x)^N - \mu(x')^N\|_1 \geq \delta(n).$$

Proof Sketch

Lemma 1: If x is the input and t traces are given, then

$$\mathbb{P}\left[\|\hat{\mu}^N - \mu(x)^N\|_1 \leq \frac{1}{4}\exp\left(-Cn^{1/3}\right)\right] \geq 1 - \exp\left(-\Omega(n)\right).$$

Lemma 2: There exists a constant $C > 0$ such that for n large enough, $N = O(n)$, for any $x \neq x'$,

$$\|\mu(x)^N - \mu(x')^N\|_1 \geq \exp\left(-Cn^{1/3}\right).$$

Proof Sketch

Lemma 1: If x is the input and t traces are given, then

$$\mathbb{P}\left[\, \|\hat{\mu}^N - \mu(x)^N\|_1 \leq \frac{1}{4}\exp\left(-Cn^{1/3}\right)\right] \geq 1 - \exp\left(-\Omega(n)\right).$$

Chernoff bound + union bound

Lemma 2: There exists a constant $C > 0$ such that for n large enough, $N = O(n)$, for any $x \neq x'$,

$$\|\mu(x)^N - \mu(x')^N\|_1 \geq \exp\left(-Cn^{1/3}\right).$$

Proof Sketch

Lemma 1: If x is the input and t traces are given, then

$$\mathbb{P}\left[\,\|\hat{\mu}^N - \mu(x)^N\|_1 \le \tfrac{1}{4}\exp\left(-Cn^{1/3}\right)\right] \ge 1 - \exp\left(-\Omega(n)\right).$$

Chernoff bound + union bound

Lemma 2: There exists a constant $C > 0$ such that for n large enough, $N = O(n)$, for any $x \neq x'$,

$$\|\mu(x)^N - \mu(x')^N\|_1 \ge \exp\left(-Cn^{1/3}\right).$$

Complex analysis! For any $z \in \mathbb{C}$, $|z| \ge 1$,

$$\|\mu(x)^N - \mu(x')^N\|_1 \ge |z|^{-N}\left(|\overline{P_x}(z) - \overline{P_{x'}}(z)| - \sum_{i=N+1}^{\infty} |\mu(x)_i - \mu(x')_i| \cdot |z|^{i-1}\right).$$

Complex Analytic Techniques

1. Number of traces required to accurately estimate $\hat{\mu}$ is determined by bounding

$$\min_{\substack{x \neq x' \\ x, x' \in \{-1,1\}^n}} \{\|\mu(x) - \mu(x')\|_1\}.$$

Complex Analytic Techniques

1. Number of traces required to accurately estimate $\hat{\mu}$ is determined by bounding

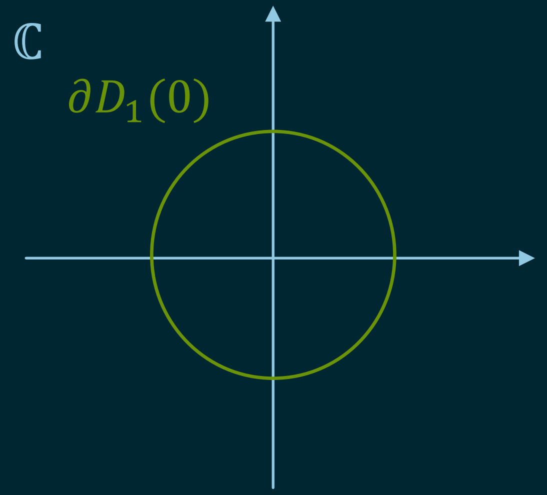
$$\min_{\substack{x \neq x' \\ x, x' \in \{-1,1\}^n}} \{\|\mu(x) - \mu(x')\|_1\}.$$

$$\boxed{x} \dashrightarrow \boxed{P_x(z)} = \sum_{i=1}^{n} x_i z^{i-1} \in \mathbb{C}[z]$$

$$\boxed{\mu(x)} \dashrightarrow \boxed{\bar{P}_x(z)} = \sum_{i=1}^{\infty} \mu(x)_i z^{i-1}$$ mean trace power series

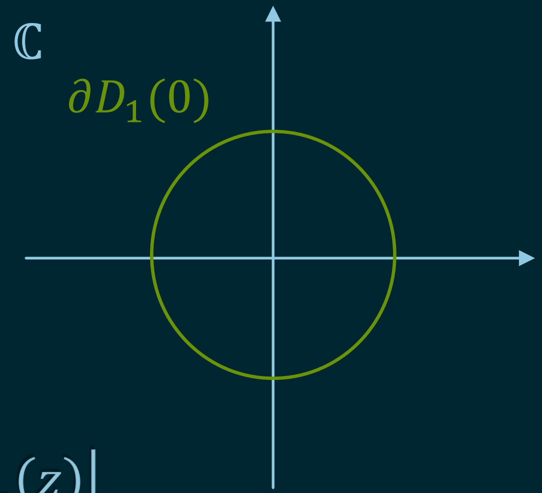
Complex Analytic Techniques

2. $$\|\mu(x) - \mu(x')\|_1 \geq \max_{z \in \partial D_1(0)} \{|\overline{P}_x(z) - \overline{P_{x'}}(z)|\}$$



Complex Analytic Techniques

2. $\|\mu(x) - \mu(x')\|_1 \geq \max_{z \in \partial D_1(0)} \{|\overline{P_x}(z) - \overline{P_{x'}}(z)|\}$



change of variable:

$$|\overline{P_x}(z) - \overline{P_{x'}}(z)| = |P_x(z) - P_{x'}(z)| \cdot \left|\mathbb{E}[|Rep|] \cdot g_{W_R}(z) - \mathbb{E}[|Flip|] \cdot g_{W_F}(z)\right|$$

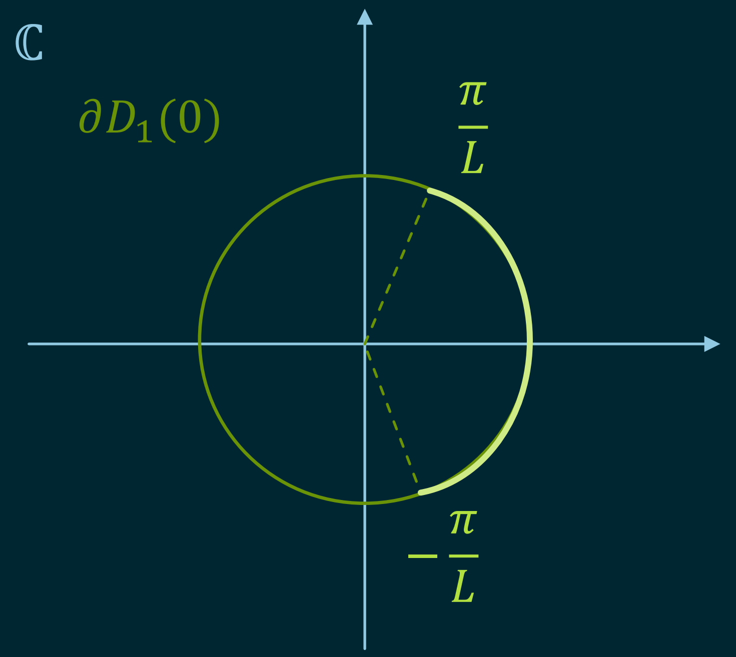
for any x, x', and z in the disk of convergence of all g power series.

Complex Analytic Techniques

3. For any $x \neq x', P_x(z) - P_{x'}(z) = |2A(z)|$ for some Littlewood polynomial

$$A(z) = \sum_{i=1}^{n} a_i \cdot z^i, a_i \in \{-1,0,1\}$$

and $\max_{z \in arc}\{|A(z)|\} \geq e^{-cL}$ for large L.



Proof Sketch

Claim: If $\mathbb{E}[|Rep|] \cdot g_{W_R}(z) = \mathbb{E}[|Flip|] \cdot g_{W_F}(z)$, then MBTR is impossible.

Proof Sketch

Claim: If $\mathbb{E}[|Rep|] \cdot g_{W_R}(z) = \mathbb{E}[|Flip|] \cdot g_{W_F}(z)$, then MBTR is impossible.

Let $x, x' \in \{-1,1\}^n$. By assumption, for any z

$$|\overline{P_x}(z) - \overline{P_{x'}}(z)| = |P_x(z) - P_{x'}(z)| \cdot \left|\mathbb{E}[|Rep|] \cdot g_{W_R}(z) - \mathbb{E}[|Flip|] \cdot g_{W_F}(z)\right| = 0.$$

Hence $\overline{P_x}(z) - \overline{P_{x'}}(z) = 0$ has all coefficients 0, so $\mu(x_i) = \mu(x_i')$ for all i.

Thus, $\mu(x) = \mu(x')$ for all x, x', so mean-based trace reconstruction is impossible.

$\square$

Example

Claim: $\exists\ Ch_M \in \mathrm{OSC}$ where $\mathbb{E}[|Rep|] \cdot g_{W_R}(z) = \mathbb{E}[|Flip|] \cdot g_{W_F}(z)$.

Example

Claim: $\exists\ Ch_M \in$ OSC where $\mathbb{E}[|Rep|] \cdot g_{W_R}(z) = \mathbb{E}[|Flip|] \cdot g_{W_F}(z)$.

$M = 2$. Let $C_- = \emptyset$ and $Rep, Flip, C_+$ be jointly distributed among 3 equally likely outcomes:

	Outcome 1	Outcome 2	Outcome 3
	1 2	1 2	1 2
$-1 \mapsto$	-1 -1	1 1	1 1
$1 \mapsto$	1 1	-1 1	1 -1

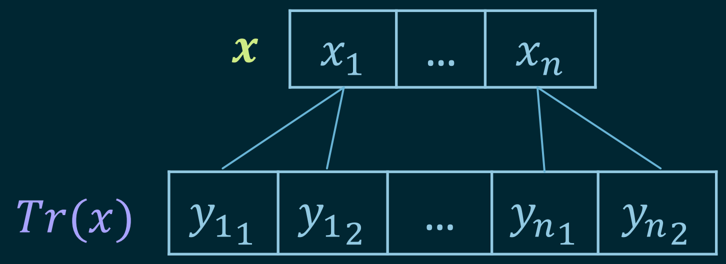
Regardless of input, each output bit has expected value $\frac{1}{3}$. So MBTR is impossible!

Example

(cont.) But trace reconstruction is easy!

$-1 \mapsto$ $\boxed{-1}\boxed{-1}$ with probability $\frac{1}{3}$

$1 \mapsto$ $\boxed{-1}\boxed{-1}$ with probability 0

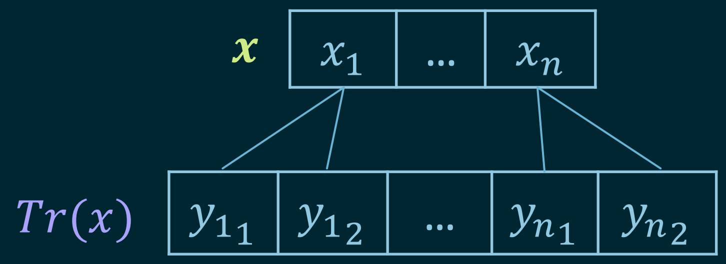


Example

(cont.) But trace reconstruction is easy!

$-1 \mapsto$ $\boxed{\text{-1}}\boxed{\text{-1}}$ with probability $\frac{1}{3}$

$1 \mapsto$ $\boxed{\text{-1}}\boxed{\text{-1}}$ with probability 0



Use this to distinguish any $x \neq x'$:

W.l.o.g. assume $x_i = -1$ and $x_i' = 1$ for some i. Consider $Tr(x), Tr(x')$ traces of x, x'.

$$\mathbb{P}[Tr(x)_{2i-1} = Tr(x)_{2i} = -1] = \frac{1}{3}$$

$$\mathbb{P}[Tr(x')_{2i-1} = Tr(x')_{2i} = -1] = 0$$

Example

(cont.) Algorithm:

Given t traces of input $z \in \{-1,1\}^n$, reconstruct a guess z^*:

1. If any trace has $Tr(z)_{2i-1} = Tr(z)_{2i} = -1$, set $z_i^* := -1$.

   Else, set $z_i^* := 1$.

2. Output z^*.

Example

(cont.) Algorithm:

Given t traces of input $z \in \{-1,1\}^n$, reconstruct a guess z^*:

1. If any trace has $Tr(z)_{2i-1} = Tr(z)_{2i} = -1$, set $z_i^* \coloneqq -1$.

   Else, set $z_i^* \coloneqq 1$.

2. Output z^*.

Correctness: $\mathbb{P}[z^* \neq z] \leq \sum_{i=1}^{n} \mathbb{P}[z_i^* \neq z_i] = n\left(\frac{2}{3}\right)^t \leq \delta$ small if we take $t = O\left(\log\left(\frac{n}{\delta}\right)\right)$.

Future Work

- Remove subexponential condition or prove it is necessary for MBTR with $\exp\left(O(n^{1/3})\right)$ traces.
- $\exp\left(\Omega(n^{1/3})\right)$ traces is required for deletion channel, so our upper bound is tight.
- Generalize this result further to all discrete memoryless synchronization channels.
- [Chase, 2022] Best known upper bound for deletion channel, analyses appearance of short sequences instead of bits. Extend this technique to oblivious synchronization channel?

Future Work

- Remove subexponential condition or prove it is necessary for MBTR with $\exp\left(O(n^{1/3})\right)$ traces.
- $\exp\left(\Omega(n^{1/3})\right)$ traces is required for deletion channel, so our upper bound is tight.
- Generalize this result further to all discrete memoryless synchronization channels.
- [Chase, 2022] Best known upper bound for deletion channel, analyses appearance of short sequences instead of bits. Extend this technique to oblivious synchronization channel?

Thank you! Questions?