Error-Correcting Codes for Nanopore Sequencing

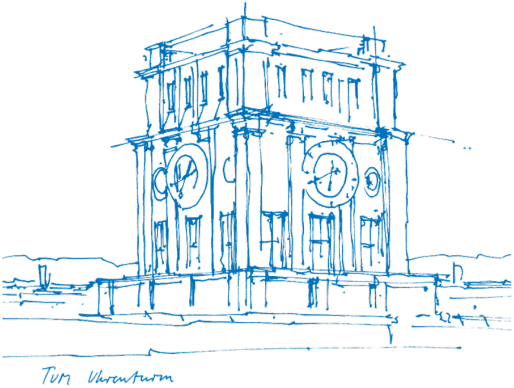
Anisha Banerjee[1]

Joint work with
Yonatan Yehezkeally[1], Antonia Wachter-Zeh[1], and Eitan Yaakobi[2]

[1]Technical University of Munich
Institute for Communications Engineering

[2]Technion – Israel Institute of Technology
Department of Computer Science

June 26, 2023



Outline

Motivation

- Need for dense, reliable, robust storage media
  - Molecular storage paradigms *e.g., DNA storage.*[1]

[1] R. Heckel *et al.*, "A characterization of the DNA data storage channel," *Scientific Reports*, vol. 9, no. 9663, Jul. 2019.

[2] D. Deamer *et al.*, "Three decades of nanopore sequencing," *Nature Biotechnology*, vol. 34, no. 5, pp. 518–524, May 2016; A. H. Laszlo *et al.*, "Decoding long nanopore sequencing reads of natural DNA," *Nature Biotechnology*, vol. 32, no. 8, pp. 829–833, Aug. 2014.

Motivation

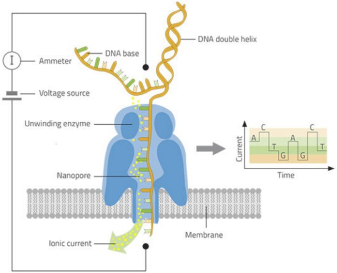
- Need for dense, reliable, robust storage media
  - Molecular storage paradigms *e.g., DNA storage.*[1]
- Faster, cheaper sequencers in development.

[1] R. Heckel *et al.*, "A characterization of the DNA data storage channel," *Scientific Reports*, vol. 9, no. 9663, Jul. 2019.

[2] D. Deamer *et al.*, "Three decades of nanopore sequencing," *Nature Biotechnology*, vol. 34, no. 5, pp. 518–524, May 2016; A. H. Laszlo *et al.*, "Decoding long nanopore sequencing reads of natural DNA," *Nature Biotechnology*, vol. 32, no. 8, pp. 829–833, Aug. 2014.

Motivation

- Need for dense, reliable, robust storage media
  - Molecular storage paradigms *e.g., DNA storage.*[1]
- Faster, cheaper sequencers in development.
- Nanopore sequencing[2]:
  - \+ Can read longer DNA strands
  - \+ More portable
  - \+ Low cost
  - – High error rates



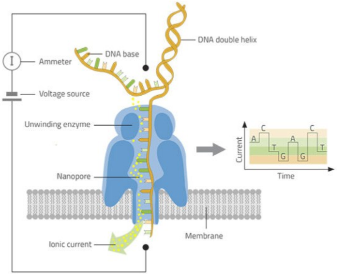
Source: *"Decoding DNA with a pocket-sized sequencer," Science in School.* `https://www.scienceinschool.org/article/2018/decoding-dna-pocket-sized-sequencer/`

[1] R. Heckel *et al.*, "A characterization of the DNA data storage channel," *Scientific Reports*, vol. 9, no. 9663, Jul. 2019.

[2] D. Deamer *et al.*, "Three decades of nanopore sequencing," *Nature Biotechnology*, vol. 34, no. 5, pp. 518–524, May 2016; A. H. Laszlo *et al.*, "Decoding long nanopore sequencing reads of natural DNA," *Nature Biotechnology*, vol. 32, no. 8, pp. 829–833, Aug. 2014.

Motivation

- Need for dense, reliable, robust storage media
  - Molecular storage paradigms *e.g., DNA storage.*[1]
- Faster, cheaper sequencers in development.
- Nanopore sequencing[2]:
  - \+ Can read longer DNA strands
  - \+ More portable
  - \+ Low cost
  - — High error rates
- *Aim:* Design coding techniques tailored for nanopore sequencing!

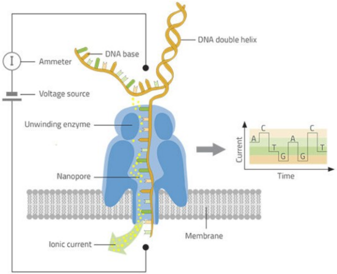


Source: *"Decoding DNA with a pocket-sized sequencer," Science in School.* https://www.scienceinschool.org/article/2018/decoding-dna-pocket-sized-sequencer/

[1] R. Heckel *et al.*, "A characterization of the DNA data storage channel," *Scientific Reports*, vol. 9, no. 9663, Jul. 2019.

[2] D. Deamer *et al.*, "Three decades of nanopore sequencing," *Nature Biotechnology*, vol. 34, no. 5, pp. 518–524, May 2016; A. H. Laszlo *et al.*, "Decoding long nanopore sequencing reads of natural DNA," *Nature Biotechnology*, vol. 32, no. 8, pp. 829–833, Aug. 2014.

Nanopore Sequencing



Source: *"Decoding DNA with a pocket-sized sequencer," Science in School.* `https://www.scienceinschool.org/article/2018/decoding-dna-pocket-sized-sequencer/`

- Sources of noise [MDK18] :
  - Nanopore holds $\ell > 1$ nucleotides at a time $\rightarrow$ Intersymbol interference (ISI)!
  - Strand moves irregularly $\rightarrow$ backtracking & skipping.
  - Noisy measurements

[MDK18] W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Prior Work

- [MDK18] → Introduced a mathematical model.
  → Established bounds on capacity.
- [HCW21] → Algorithm to compute capacity of an abstracted, deterministic channel model.
  → Coding schemes.
- [MVS22] → Finite-state Markov channel model incorporating major noise sources.
  → Generalized MAP detection algorithms.
- [CVVY21] considered a similar channel for racetrack memories:
  → Computed information limits.
  → Proposed error-correcting codes.

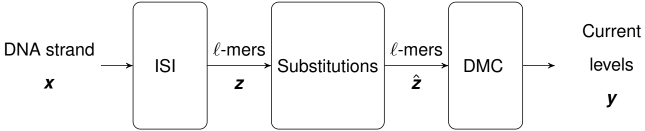
[MDK18] W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

[HCW21] R. Hulett *et al.*, "On Coding for an Abstracted Nanopore Channel for DNA Storage," in *IEEE Intl. Symp. on Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2465–2470

[MVS22] B. McBain *et al.*, "Finite-state semi-markov channels for nanopore sequencing," in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Espoo, Finland, Jun. 2022, pp. 216–221

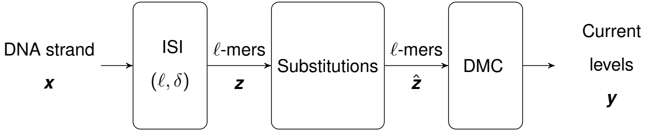
[CVVY21] Y. M. Chee *et al.*, "Coding for transverse-reads in domain wall memories," in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2924–2929

Simplified Model



W. Mao *et al.*, “Models and information-theoretic bounds for nanopore sequencing,” *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model

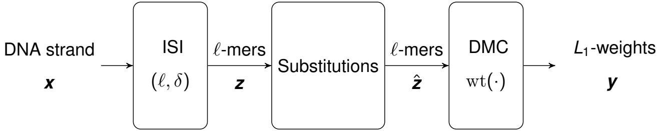


We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



We model

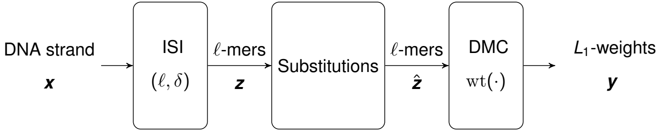
- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



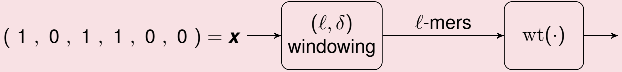
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

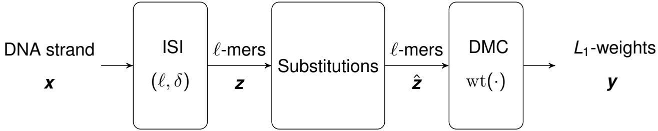
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



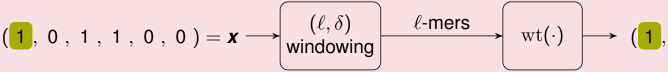
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

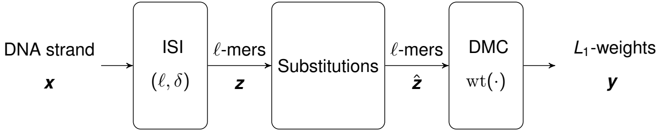
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, “Models and information-theoretic bounds for nanopore sequencing,” *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



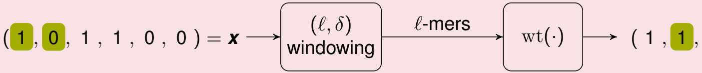
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

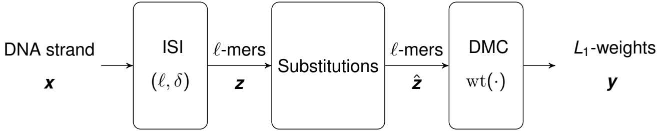
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



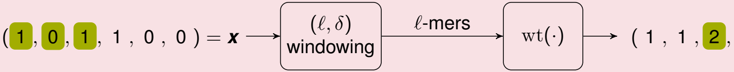
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

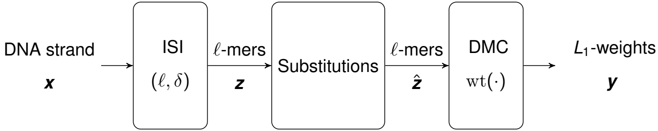
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



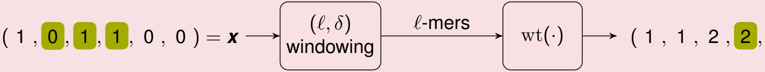
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

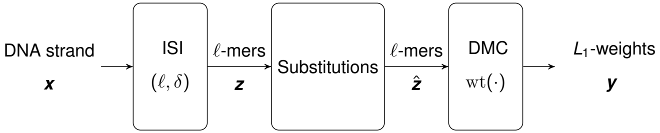
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, “Models and information-theoretic bounds for nanopore sequencing,” *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



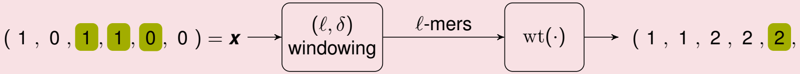
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

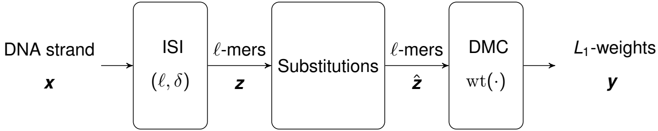
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



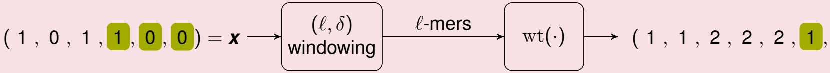
We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

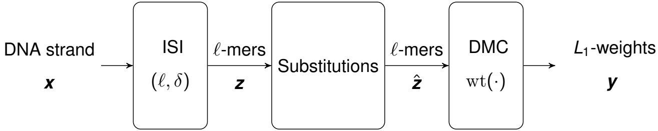
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)



W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

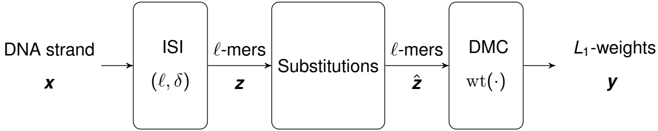
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)

$(1, 0, 1, 1, \boxed{0}, \boxed{0}) = \boldsymbol{x} \longrightarrow \boxed{(\ell, \delta) \text{ windowing}} \xrightarrow{\ell\text{-mers}} \boxed{\mathrm{wt}(\cdot)} \longrightarrow (1, 1, 2, 2, 2, 1, \boxed{0},$

W. Mao *et al.*, “Models and information-theoretic bounds for nanopore sequencing,” *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

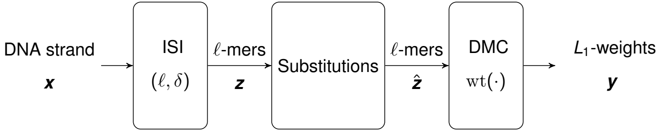
- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)

$$(1, 0, 1, 1, 0, \boxed{0}) = \boldsymbol{x} \longrightarrow \boxed{\begin{matrix}(\ell,\delta)\\ \text{windowing}\end{matrix}} \xrightarrow{\ell\text{-mers}} \boxed{\mathrm{wt}(\cdot)} \longrightarrow (1, 1, 2, 2, 2, 1, 0, \boxed{0})$$

W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model



We model

- ISI as (ℓ, δ)-sliding window
- Measurement noise as substitutions

Assumptions:

- Input $\boldsymbol{x} \in \{0, 1\}^n$
- Output: L_1-weights of ℓ-mers

Example ($\ell = 3$, $\delta = 1$)

W. Mao *et al.*, "Models and information-theoretic bounds for nanopore sequencing," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018

Simplified Model

Definition

The (ℓ, δ)-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is

$$\mathcal{R}_{\ell,\delta}(\boldsymbol{x}) = (\mathrm{wt}(\boldsymbol{x}_{\delta-\ell+1}^{\delta}), \mathrm{wt}(\boldsymbol{x}_{2\delta-\ell+1}^{2\delta}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\delta+1}^{n+\ell-\delta})),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathrm{wt}(\cdot)$ denotes L_1-weight.

Y. M. Chee *et al.*, “Coding for transverse-reads in domain wall memories,” in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2924–2929

Simplified Model

Definition

The (ℓ, δ)-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is

$$\mathcal{R}_{\ell,\delta}(\boldsymbol{x}) = (\mathrm{wt}(\boldsymbol{x}_{\delta-\ell+1}^{\delta}), \mathrm{wt}(\boldsymbol{x}_{2\delta-\ell+1}^{2\delta}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\delta+1}^{n+\ell-\delta})),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathrm{wt}(\cdot)$ denotes L_1-weight.

Our Results

- $\log\log n - o(1)$ min redundancy to correct 1 substitution in $(\ell \geq 3, \delta = 1)$-read vectors
- Introduce a 1-substitution correcting code that is optimal up to a constant.

Y. M. Chee *et al.*, “Coding for transverse-reads in domain wall memories,” in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2924–2929

Read Vector

Definition

The (ℓ, δ)-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $(n+\ell)/\delta - 1$ and denoted by

$$\mathcal{R}_{\ell,\delta}(\boldsymbol{x}) = (\mathrm{wt}(\boldsymbol{x}_{\delta-\ell+1}^{\delta}), \mathrm{wt}(\boldsymbol{x}_{2\delta-\ell+1}^{2\delta}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\delta+1}^{n+\ell-\delta})),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}_{\ell,\delta}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i\delta-\ell+1}^{i\delta})$.

Read Vector

Definition

The (ℓ, δ)-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $(n + \ell)/\delta - 1$ and denoted by

$$\mathcal{R}_{\ell,\delta}(\boldsymbol{x}) = (\mathrm{wt}(\boldsymbol{x}_{\delta-\ell+1}^{\delta}), \mathrm{wt}(\boldsymbol{x}_{2\delta-\ell+1}^{2\delta}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\delta+1}^{n+\ell-\delta})),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}_{\ell,\delta}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i\delta-\ell+1}^{i\delta})$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.

Read Vector

Definition

The $(\ell, 1)$-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $n + \ell - 1$ and denoted by

$$\mathcal{R}(\boldsymbol{x}) = (\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots, \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(x_n))),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i-\ell+1}^i)$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.

Read Vector

Definition

The $(\ell, 1)$-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $n + \ell - 1$ and denoted by

$$\mathcal{R}(\boldsymbol{x}) = (\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots, \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(x_n))),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i-\ell+1}^i)$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.
- Read vectors have special properties.
  - $|\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{x})_{i-1}| \leq 1 \impliedby \mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{x})_{i-1} = x_i - x_{i-\ell}$

Read Vector

Definition

The $(\ell, 1)$-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $n + \ell - 1$ and denoted by

$$\mathcal{R}(\boldsymbol{x}) = (\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots, \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(x_n))),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i-\ell+1}^i)$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.
- Read vectors have special properties.
  - $|\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{x})_{i-1}| \leq 1$,
  - $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0 \Longleftarrow \sum_i \mathcal{R}(\boldsymbol{x})_i = \ell \cdot \mathrm{wt}(\boldsymbol{x})$

Read Vector

Definition

The $(\ell, 1)$-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $n + \ell - 1$ and denoted by

$$\mathcal{R}(\boldsymbol{x}) = (\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots, \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(x_n))),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i-\ell+1}^i)$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.
- Read vectors have special properties.
  - $|\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{x})_{i-1}| \leq 1$,
  - $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0$, etc.
- Can correct errors of abs. value ≥ 2.

Read Vector

Definition

The $(\ell, 1)$-*read vector* of any $\boldsymbol{x} \in \Sigma_2^n$ is of length $n + \ell - 1$ and denoted by

$$\mathcal{R}(\boldsymbol{x}) = (\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots, \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(x_n))),$$

where for $i \notin [n]$, let $x_i = 0$ and $\mathcal{R}(\boldsymbol{x})_i = \mathrm{wt}(\boldsymbol{x}_{i-\ell+1}^i)$.

- We limit focus to $\ell \geq 3$, $\delta = 1$.
- Read vectors have special properties.
  - $|\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{x})_{i-1}| \leq 1$,
  - $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0$, etc.
- Can correct errors of abs. value ≥ 2.
  - Suffices to **consider ± 1 errors**.

Error-correcting Code

Definition

$\mathcal{C}$ is a t-substitution $(\ell, 1)$-read code if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$,

$$d_H\big(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\big) > 2t,$$

where $d_H(\cdot, \cdot)$ denotes Hamming distance.

Error-correcting Code

Definition

$\mathcal{C}$ is a t-substitution $(\ell, 1)$-read code if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$,

$$d_H\big(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\big) > 2t,$$

where $d_H(\cdot, \cdot)$ denotes Hamming distance.

- *Aim:* Find a 1-substitution $(\ell, 1)$-read code.

Error-correcting Code

Definition

$\mathcal{C}$ is a t-substitution $(\ell, 1)$-read code if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$,

$$d_H\big(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\big) > 2t,$$

where $d_H(\cdot, \cdot)$ denotes Hamming distance.

- *Aim:* Find a 1-substitution $(\ell, 1)$-read code.
  - Minimum redundancy required?

Error-correcting Code

Definition

$\mathcal{C}$ is a t-substitution $(\ell, 1)$-read code if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$,

$$d_H\big(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\big) > 2t,$$

where $d_H(\cdot, \cdot)$ denotes Hamming distance.

- *Aim:* Find a 1-substitution $(\ell, 1)$-read code.
  - Minimum redundancy required?
  - When does $d_H\big(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\big) \leq 2$ occur?

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) \leq 2$ occur?

Lemma

For $\ell > 1$, $\delta = 1$ and any distinct $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}^n$, $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) \neq 1$.

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) \leq 2$ occur?

Lemma

For $\ell > 1$, $\delta = 1$ and any distinct $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$, $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) \neq 1$.

Proof idea:

- Use read-vector properties:
  - $\sum_i \mathcal{R}_i \bmod \ell = 0$.
  - $|\mathcal{R}_{i+1} - \mathcal{R}_i| \leq 1$.

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$ occur?

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$ occur?

Theorem

For $\ell \geq 3$ and any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, the following are equivalent:

1. $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.

Example ($\ell = 3$, $\delta = 1$)

$$\boldsymbol{x} = (1,0,1,1,0,0), \quad \mathcal{R}(\boldsymbol{x}) = (\boxed{1}, 1, 2, 2, 2, 1, \boxed{0}, 0)$$

$$\boldsymbol{y} = (0,1,1,0,1,0), \quad \mathcal{R}(\boldsymbol{y}) = (\boxed{0}, 1, 2, 2, 2, 1, \boxed{1}, 0)$$

$$d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$$

Confusable Read Vectors

- When does $d_H\left(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})\right) = 2$ occur?

Theorem

For $\ell \geq 3$ and any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, the following are equivalent:

1. $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
2. There exist distinct $i, j \in [n + \ell - 1]$, $j = i \pmod{\ell}$, s.t.

Example ($\ell = 3$, $\delta = 1$)

$$\boldsymbol{x} = (1, 0, 1, 1, 0, 0), \quad \mathcal{R}(\boldsymbol{x}) = (\ 1\ ,\ 1\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$$

$$\boldsymbol{y} = (0, 1, 1, 0, 1, 0), \quad \mathcal{R}(\boldsymbol{y}) = (\ 0\ ,\ 1\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 1\ ,\ 0\ )$$

$$d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$$

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$ occur?

Theorem

For $\ell \geq 3$ and any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, the following are equivalent:

1. $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
2. There exist distinct $i, j \in [n + \ell - 1]$, $j = i \pmod{\ell}$, s.t. $\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{y})_i = \mathcal{R}(\boldsymbol{y})_j - \mathcal{R}(\boldsymbol{x})_j = 1$ and $\mathcal{R}(\boldsymbol{x})_r = \mathcal{R}(\boldsymbol{y})_r$ for all $r \notin \{i, j\}$.

Example ($\ell = 3$, $\delta = 1$)

$$\boldsymbol{x} = (1,0,1,1,0,0), \quad \mathcal{R}(\boldsymbol{x}) = (\boxed{1}, 1, 2, 2, 2, 1, \boxed{0}, 0)$$

$$\boldsymbol{y} = (0,1,1,0,1,0), \quad \mathcal{R}(\boldsymbol{y}) = (\boxed{0}, 1, 2, 2, 2, 1, \boxed{1}, 0)$$

$$d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$$

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$ occur?

Theorem

For $\ell \geq 3$ and any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, the following are equivalent:

1. $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
2. There exist distinct $i, j \in [n+\ell-1]$, $j = i \pmod{\ell}$, s.t. $\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{y})_i = \mathcal{R}(\boldsymbol{y})_j - \mathcal{R}(\boldsymbol{x})_j = 1$ and $\mathcal{R}(\boldsymbol{x})_r = \mathcal{R}(\boldsymbol{y})_r$ for all $r \notin \{i, j\}$.
3. There exist $p \geq 1$, $i \in [n-(p-1)\ell-1]$ s.t.

Example ($\ell = 3$, $\delta = 1$)

$$\boldsymbol{x} = (1,0,1,1,0,0), \quad \mathcal{R}(\boldsymbol{x}) = (\ 1\ ,\ 1\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$$
$$\boldsymbol{y} = (0,1,1,0,1,0), \quad \mathcal{R}(\boldsymbol{y}) = (\ 0\ ,\ 1\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 1\ ,\ 0\ )$$
$$d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$$

Confusable Read Vectors

- When does $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$ occur?

Theorem

For $\ell \geq 3$ and any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, the following are equivalent:

1. $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
2. There exist distinct $i, j \in [n + \ell - 1]$, $j = i \pmod{\ell}$, s.t. $\mathcal{R}(\boldsymbol{x})_i - \mathcal{R}(\boldsymbol{y})_i = \mathcal{R}(\boldsymbol{y})_j - \mathcal{R}(\boldsymbol{x})_j = 1$ and $\mathcal{R}(\boldsymbol{x})_r = \mathcal{R}(\boldsymbol{y})_r$ for all $r \notin \{i, j\}$.
3. There exist $p \geq 1$, $i \in [n - (p-1)\ell - 1]$ s.t. $\forall\, m \in \Sigma_p$, $\boldsymbol{x}_{i+m\ell}^{i+m\ell+1} = (1, 0)$, $\boldsymbol{y}_{i+m\ell}^{i+m\ell+1} = (0, 1)$ (or vice versa), and $x_r = y_r$ for all $r \notin \bigcup_{m \in \Sigma_p} \{i + m\ell, i + m\ell + 1\}$.

Further, if these conditions hold, then $j = i + p\ell$ in the above notation.

Example ($\ell = 3$, $\delta = 1$)

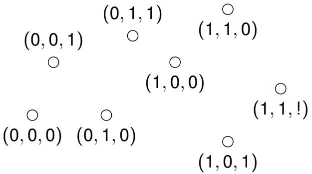
$$\boldsymbol{x} = (\boxed{1, 0}, 1, \boxed{1, 0}, 0), \quad \mathcal{R}(\boldsymbol{x}) = (1, 1, 2, 2, 2, 1, 0, 0)$$

$$\boldsymbol{y} = (\boxed{0, 1}, 1, \boxed{0, 1}, 0), \quad \mathcal{R}(\boldsymbol{y}) = (0, 1, 2, 2, 2, 1, 1, 0)$$

$$d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$$

Minimum Redundancy

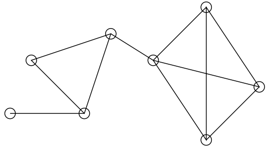
- Consider graph with all vertices in $\{0, 1\}^n$.



[K94] D. E. Knuth, “The sandwich theorem,” *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994
[CKY22] J. Chrisnata *et al.*, “Correcting deletions with multiple reads,” *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

- Consider graph with all vertices in $\{0,1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.

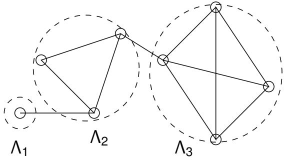


[K94] D. E. Knuth, "The sandwich theorem," *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994

[CKY22] J. Chrisnata *et al.*, "Correcting deletions with multiple reads," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

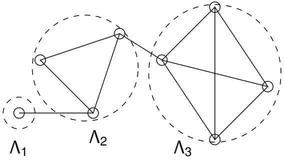
- Consider graph with all vertices in $\{0,1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
  - Clique $\Lambda \subset \{0,1\}^n \rightarrow$ all vertices pairwise-adjacent.



[K94] D. E. Knuth, “The sandwich theorem,” *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994
[CKY22] J. Chrisnata *et al.*, “Correcting deletions with multiple reads,” *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

- Consider graph with all vertices in $\{0,1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
  - Clique $\Lambda \subset \{0,1\}^n \rightarrow$ all vertices pairwise-adjacent.
  - Clique cover $\mathcal{Q} \rightarrow$ collection of cliques that covers $\{0,1\}^n$.



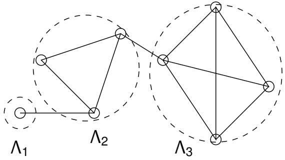
$\mathcal{Q} = \{\Lambda_1, \Lambda_2, \Lambda_3\}$

[K94] D. E. Knuth, “The sandwich theorem,” *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994

[CKY22] J. Chrisnata *et al.*, “Correcting deletions with multiple reads,” *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

- Consider graph with all vertices in $\{0, 1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
  - Clique $\Lambda \subset \{0, 1\}^n \rightarrow$ all vertices pairwise-adjacent.
  - Clique cover $\mathcal{Q} \rightarrow$ collection of cliques that covers $\{0, 1\}^n$.
- $|\mathcal{Q}|$ is an upper bound on size of code [K94] .

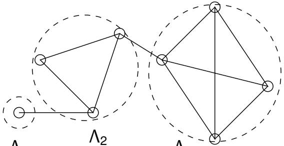


$\mathcal{Q} = \{\Lambda_1, \Lambda_2, \Lambda_3\}$

[K94] D. E. Knuth, "The sandwich theorem," *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994
[CKY22] J. Chrisnata *et al.*, "Correcting deletions with multiple reads," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

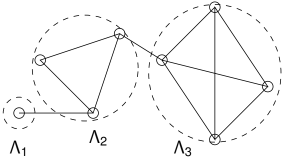
- Consider graph with all vertices in $\{0,1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
  - Clique $\Lambda \subset \{0,1\}^n \to$ all vertices pairwise-adjacent.
  - Clique cover $\mathcal{Q} \to$ collection of cliques that covers $\{0,1\}^n$.
- $|\mathcal{Q}|$ is an upper bound on size of code [K94] .
- Adapt clique-cover defined in [CKY22] .



[K94] D. E. Knuth, "The sandwich theorem," *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994
[CKY22] J. Chrisnata *et al.*, "Correcting deletions with multiple reads," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Minimum Redundancy

- Consider graph with all vertices in $\{0,1\}^n$.
  - $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$ adjacent iff $d_H(\mathcal{R}(\boldsymbol{x}), \mathcal{R}(\boldsymbol{y})) = 2$.
  - Clique $\Lambda \subset \{0,1\}^n \to$ all vertices pairwise-adjacent.
  - Clique cover $\mathcal{Q} \to$ collection of cliques that covers $\{0,1\}^n$.
- $|\mathcal{Q}|$ is an upper bound on size of code [K94] .
- Adapt clique-cover defined in [CKY22] .



$\mathcal{Q} = \{\Lambda_1, \Lambda_2, \Lambda_3\}$

Theorem

The redundancy of a 1-substitution $(\ell, 1)$-read code is bounded from below by

$$\log_2 \log_2(n) - o(1).$$

[K94] D. E. Knuth, "The sandwich theorem," *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994
[CKY22] J. Chrisnata *et al.*, "Correcting deletions with multiple reads," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022

Introduction

Channel Model

Minimum Redundancy

Error-correcting Code

Conclusion

Reconstruction from Read Vectors

$$\mathcal{R}(\boldsymbol{x}) = \left(\mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^n), \mathrm{wt}(x_n)\right)$$

Reconstruction from Read Vectors

$$\big(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^{\ell}), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots \ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^{n}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^{n}), \mathrm{wt}(x_n), 0\big)$$

Reconstruction from Read Vectors

$$\left(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^n), \mathrm{wt}(x_n), 0\right)$$

$$\Big\downarrow \text{Derivative}$$

$$\left(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\right)$$

Reconstruction from Read Vectors

$$\left(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^n), \mathrm{wt}(x_n), 0\right)$$

$$\Big\downarrow \text{ Derivative}$$

$$\left(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\right)$$

$$\Big\downarrow \text{ Sub-sequence}$$

$$\left(x_1, x_{\ell+1} - x_1, x_{2\ell+1} - x_{\ell+1}, \ldots, -x_{\lfloor \frac{n-1}{\ell} \rfloor \ell+1}\right)$$

Reconstruction from Read Vectors

$$\left(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^n), \mathrm{wt}(x_n), 0\right)$$

$$\Big\downarrow \text{Derivative}$$

$$\left(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\right)$$

$$\Big\downarrow \text{Sub-sequence}$$

$$\left(x_1, x_{\ell+1} - x_1, x_{2\ell+1} - x_{\ell+1}, \ldots, -x_{\lfloor \frac{n-1}{\ell} \rfloor \ell+1}\right)$$

- Call this **read sub-derivative $\Delta^0(\boldsymbol{x})$**.

Reconstruction from Read Vectors

$$\big(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^{\ell}), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots \ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^{n}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^{n}), \mathrm{wt}(x_n), 0\big)$$

$$\Big\downarrow \text{ Derivative}$$

$$\big(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\big)$$

$$\Big\downarrow \text{ Sub-sequence}$$

$$\big(x_1, x_{\ell+1} - x_1, x_{2\ell+1} - x_{\ell+1}, \ldots, -x_{\lfloor \frac{n-1}{\ell} \rfloor \ell+1}\big)$$

[Sums to 0]

- Call this **read sub-derivative** $\Delta^0(\boldsymbol{x})$.

Reconstruction from Read Vectors

$$\left(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^{\ell}), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^{n}), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^{n}), \mathrm{wt}(x_n), 0\right)$$

$$\Big\downarrow \text{Derivative}$$

$$\left(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\right)$$

$$\Big\downarrow \text{Sub-sequence}$$

$$\left(x_{\alpha+1}, x_{\ell+\alpha+1} - x_{\alpha+1}, x_{2\ell+\alpha+1} - x_{\ell+\alpha+1}, \ldots, -x_{\lfloor \frac{n-\alpha-1}{\ell} \rfloor \ell+\alpha+1}\right)$$

[Sums to 0]

- Call this **read sub-derivative** $\Delta^{\alpha}(\boldsymbol{x})$, $\alpha \in \{0, \ldots, \ell - 1\}$.

Reconstruction from Read Vectors

$$\left(0, \mathrm{wt}(x_1), \mathrm{wt}(\boldsymbol{x}_1^2), \ldots \mathrm{wt}(\boldsymbol{x}_1^\ell), \mathrm{wt}(\boldsymbol{x}_2^{\ell+1}), \ldots\ldots \mathrm{wt}(\boldsymbol{x}_{n-\ell+1}^n), \ldots, \mathrm{wt}(\boldsymbol{x}_{n-1}^n), \mathrm{wt}(x_n), 0\right)$$

$$\Big\downarrow \text{Derivative}$$

$$\left(x_1, x_2, \ldots, x_\ell, x_{\ell+1} - x_1, x_{\ell+2} - x_2, \ldots, x_n - x_{n-\ell}, -x_{n-\ell+1}, \ldots, -x_{n-1}, -x_n\right)$$

$$\Big\downarrow \text{Sub-sequence}$$

$$\left(x_{\alpha+1}, x_{\ell+\alpha+1} - x_{\alpha+1}, x_{2\ell+\alpha+1} - x_{\ell+\alpha+1}, \ldots, -x_{\lfloor \frac{n-\alpha-1}{\ell} \rfloor \ell+\alpha+1}\right) \qquad \text{[Sums to 0]}$$

- Call this **read sub-derivative** $\Delta^\alpha(\boldsymbol{x})$, $\alpha \in \{0, \ldots, \ell - 1\}$.
- Helps reconstruct subsequence $(x_{\alpha+1}, x_{\ell+\alpha+1}, x_{2\ell+\alpha+1}, \ldots)$.

Read Sub-derivative

Sub-derivative of read vector

For any $\alpha \in \{0, \ldots, \ell - 1\}$, the α-th read sub-derivative of $\mathcal{R}(\boldsymbol{x})$ is

$$\begin{aligned}\Delta^{\alpha}(\boldsymbol{x}) =& \left(\mathcal{R}(\boldsymbol{x})_{\alpha+1} - \mathcal{R}(\boldsymbol{x})_{\alpha}, \mathcal{R}(\boldsymbol{x})_{\alpha+\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+\ell}, \ldots, \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell}\right) \\ =& \left(x_{\alpha+1}, x_{\alpha+\ell+1} - x_{\alpha+1}, \ldots, x_{\alpha+(k-1)\ell+1} - x_{\alpha+(k-2)\ell+1}, -x_{\alpha+(k-1)\ell+1}\right),\end{aligned}$$

It holds that $\sum_i \Delta^{\alpha}(\boldsymbol{x})_i = 0$.

Read Sub-derivative

Sub-derivative of read vector

For any $\alpha \in \{0, \ldots, \ell - 1\}$, the α-th read sub-derivative of $\mathcal{R}(\boldsymbol{x})$ is

$$\begin{aligned}\Delta^{\alpha}(\boldsymbol{x}) =& (\mathcal{R}(\boldsymbol{x})_{\alpha+1} - \mathcal{R}(\boldsymbol{x})_{\alpha}, \mathcal{R}(\boldsymbol{x})_{\alpha+\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+\ell}, \ldots, \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell}) \\ =& (x_{\alpha+1}, x_{\alpha+\ell+1} - x_{\alpha+1}, \ldots, x_{\alpha+(k-1)\ell+1} - x_{\alpha+(k-2)\ell+1}, -x_{\alpha+(k-1)\ell+1}),\end{aligned}$$

It holds that $\sum_i \Delta^{\alpha}(\boldsymbol{x})_i = 0$.

- *Note:* If $\mathcal{R}(\boldsymbol{x})_i \xrightarrow{\text{substitute}} \mathcal{R}(\boldsymbol{x})'_i$, two read sub-derivatives are affected.

$$\begin{aligned}\mathcal{R}(\boldsymbol{x})'_i - \mathcal{R}(\boldsymbol{x})_i &= \sum_k \Delta^{i-1 \bmod \ell}(\boldsymbol{x})'_k \\ &= -\sum_k \Delta^{i \bmod \ell}(\boldsymbol{x})'_k \neq 0\end{aligned}$$

Read Sub-derivative

Sub-derivative of read vector

For any $\alpha \in \{0, \dots, \ell - 1\}$, the α-th read sub-derivative of $\mathcal{R}(\boldsymbol{x})$ is

$$\begin{aligned}\Delta^{\alpha}(\boldsymbol{x}) =& (\mathcal{R}(\boldsymbol{x})_{\alpha+1} - \mathcal{R}(\boldsymbol{x})_{\alpha}, \mathcal{R}(\boldsymbol{x})_{\alpha+\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+\ell}, \dots, \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell+1} - \mathcal{R}(\boldsymbol{x})_{\alpha+k\ell}) \\ =& (x_{\alpha+1}, x_{\alpha+\ell+1} - x_{\alpha+1}, \dots, x_{\alpha+(k-1)\ell+1} - x_{\alpha+(k-2)\ell+1}, -x_{\alpha+(k-1)\ell+1}),\end{aligned}$$

It holds that $\sum_i \Delta^{\alpha}(\boldsymbol{x})_i = 0$.

- *Note:* If $\mathcal{R}(\boldsymbol{x})_i \xrightarrow{\text{substitute}} \mathcal{R}(\boldsymbol{x})'_i$, two read sub-derivatives are affected.

$$\begin{aligned}\mathcal{R}(\boldsymbol{x})'_i - \mathcal{R}(\boldsymbol{x})_i &= \sum_k \Delta^{i-1 \bmod \ell}(\boldsymbol{x})'_k \\ &= -\sum_k \Delta^{i \bmod \ell}(\boldsymbol{x})'_k \neq 0\end{aligned}$$

Takeaway

Given $\mathcal{R}(\boldsymbol{x})'$, can determine **error value** & **position upto** $\bmod \ell$.

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$$

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\ &= (\quad 1 \quad , \quad 1 \quad , \quad -1 \quad)\end{aligned}$$

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1,1,2,3,2,1,0,0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1,4,7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\ &= (\quad 1 \quad , \quad 1 \quad , \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\ &= (\quad 1 \quad , \quad 1 \quad , \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_4 = 2 \notin \{0, 1\}$ ✗

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1,1,2,3,2,1,0,0)$$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad, \quad x'_4 - x'_1 \quad, \quad -x'_4 \quad) \\ &= (\quad 1 \quad, \quad 1 \quad, \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_4 = 2 \notin \{0, 1\}$ ✗
- Right to left reconstruction:
  - $x'_4 = 1$.

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad, \quad x'_4 - x'_1 \quad, \quad -x'_4 \quad) \\ &= (\quad 1 \quad, \quad 1 \quad, \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_4 = 2 \notin \{0, 1\}$ ✗
- Right to left reconstruction:
  - $x'_4 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_1 = 0$

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1,1,2,3,2,1,0,0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad, \quad x'_4 - x'_1 \quad, \quad -x'_4 \quad) \\ &= (\quad 1 \quad, \quad 1 \quad, \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_4 = 2 \notin \{0, 1\}$ ✗
- Right to left reconstruction:
  - $x'_4 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_1 = 0$
  - $x'_1 = 1$ ✗

Reconstruction with Noisy Sub-derivative

- Can an erroneous read sub-derivative reveal region of error?

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_1} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 3, 2, 1, 0, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_1 \bmod \ell = 1 \implies i_1 \in \{1, 4, 7\} \implies i_1 \in \{1, 4\}$

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad, \quad x'_4 - x'_1 \quad, \quad -x'_4 \quad) \\ &= (\quad 1 \quad, \quad 1 \quad, \quad -1 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_4 = 2 \notin \{0, 1\}$ ✗
- Right to left reconstruction:
  - $x'_4 = 1$.
  - $x'_4 - x'_1 = 1 \implies x'_1 = 0$
  - $x'_1 = 1$ ✗

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}
\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\
&= (\quad x'_1 \quad , \qquad x'_4 - x'_1 \qquad , \qquad -x'_4 \qquad) \\
&= (\quad 1 \quad , \qquad 0 \qquad , \qquad 0 \qquad)
\end{aligned}$$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}
\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\
&= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\
&= (\quad 1 \quad , \quad 0 \quad , \quad 0 \quad)
\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}
\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\
&= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\
&= (\quad 1 \quad , \quad 0 \quad , \quad 0 \quad)
\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 0 \implies x'_4 = 1$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\ &= (\quad 1 \quad , \quad 0 \quad , \quad 0 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 0 \implies x'_4 = 1$
  - $-x'_4 = 0$ ✗

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\ x'_1\ ,\ x'_4 - x'_1\ ,\ -x'_4\ ) \\ &= (\ 1\ ,\ 0\ ,\ 0\ )\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 0 \implies x'_4 = 1$
  - $-x'_4 = 0$ ✗
- Right to left reconstruction:
  - $x'_4 = 0$.

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}
\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\
&= (\quad x'_1 \quad , \quad x'_4 - x'_1 \quad , \quad -x'_4 \quad) \\
&= (\quad 1 \quad , \quad 0 \quad , \quad 0 \quad)
\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 0 \implies x'_4 = 1$
  - $-x'_4 = 0$ ✗

- Right to left reconstruction:
  - $x'_4 = 0$.
  - $x'_4 - x'_1 = 0 \implies x'_1 = 0$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3, \delta = 1$)

$$\mathcal{R}(\boldsymbol{x}) \xrightarrow{\text{error at } i_2} \mathcal{R}(\boldsymbol{x})' = (1, 1, 2, 2, 2, 1, 1, 0)$$

- $\sum_k \Delta^0(\boldsymbol{x})'_k = -\sum_k \Delta^1(\boldsymbol{x})'_k = 1 \implies i_2 \bmod \ell = 1 \implies i_2 \in \{1, 4, 7\}$.

$$\begin{aligned}\Delta^0(\boldsymbol{x})' &= (\mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_4 - \mathcal{R}(\boldsymbol{x})'_1, \mathcal{R}(\boldsymbol{x})'_7 - \mathcal{R}(\boldsymbol{x})'_6) \\ &= (\quad x'_1 \quad, \quad x'_4 - x'_1 \quad, \quad -x'_4 \quad) \\ &= (\quad 1 \quad, \quad 0 \quad, \quad 0 \quad)\end{aligned}$$

- Left to right reconstruction:
  - $x'_1 = 1$.
  - $x'_4 - x'_1 = 0 \implies x'_4 = 1$
  - $-x'_4 = 0$ ✗

- Right to left reconstruction:
  - $x'_4 = 0$.
  - $x'_4 - x'_1 = 0 \implies x'_1 = 0$
  - $x'_1 = 1$ ✗

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3$, $\delta = 1$)

$$\begin{array}{ccc} & \mathcal{R}(\boldsymbol{x}) & \\ \swarrow \text{error at } i_1 & & \searrow \text{error at } i_2 \\ \Delta^0(\boldsymbol{x})' = (1,1,-1) & & \Delta^0(\boldsymbol{x})' = (1,0,0) \\ \downarrow & & \downarrow \\ i_1 \in \{1,4\} & & i_2 \in \{1,4,7\} \end{array}$$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3$, $\delta = 1$)

$$\begin{array}{ccc} & \mathcal{R}(\boldsymbol{x}) & \\ \swarrow{\scriptstyle \text{error at } i_1} & & \searrow{\scriptstyle \text{error at } i_2} \\ \Delta^0(\boldsymbol{x})' = (1,1,-1) & & \Delta^0(\boldsymbol{x})' = (1,0,0) \\ \downarrow & & \downarrow \\ i_1 \in \{1,4\} & & i_2 \in \{1,4,7\} \end{array}$$

$$\begin{array}{rcccccccc} \Delta^0(\boldsymbol{x}) = (& 1 & 0 & 0 & 0 & \cdots\cdots & 0 & -1) \\ & x_1 = 1 & x_4 = 1 & x_7 = 1 & x_{11} = 1 & \cdots\cdots & x_{\lfloor \frac{n-1}{3} \rfloor 3+1} = 1 & \end{array}$$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3$, $\delta = 1$)

$$\begin{array}{ccc}
 & \mathcal{R}(\boldsymbol{x}) & \\
\text{error at } i_1 \swarrow & & \searrow \text{error at } i_2 \\
\Delta^0(\boldsymbol{x})' = (1, 1, -1) & & \Delta^0(\boldsymbol{x})' = (1, 0, 0) \\
\downarrow & & \downarrow \\
i_1 \in \{1, 4\} & & i_2 \in \{1, 4, 7\}
\end{array}$$

$$\begin{array}{cccccccc}
\Delta^0(\boldsymbol{x})' = (& 1 & 0 & -1 & 0 & \cdots\cdots & 0 & -1) \\
 & x_1 = 1 & x_4 = 1 & x_7 = 1 & x_{11} = 1 & \cdots\cdots & x_{\lfloor \frac{n-1}{3} \rfloor 3+1} = 1 & \\
 & x'_1 = 1 & x'_4 = 1 & x'_7 = 0 & x'_{11} = 0 & \cdots\cdots & x'_{\lfloor \frac{n-1}{3} \rfloor 3+1} = 0 &
\end{array}$$

Reconstruction with Noisy Sub-derivative

Example ($\ell = 3$, $\delta = 1$)

$$
\begin{array}{ccc}
 & \mathcal{R}(\boldsymbol{x}) & \\
\swarrow \text{error at } i_1 & & \searrow \text{error at } i_2 \\
\Delta^0(\boldsymbol{x})' = (1, 1, -1) & & \Delta^0(\boldsymbol{x})' = (1, 0, 0) \\
\downarrow & & \downarrow \\
i_1 \in \{1, 4\} & & i_2 \in \{1, 4, 7\}
\end{array}
$$

$$
\begin{array}{lcccccc}
\Delta^0(\boldsymbol{x})' = (\; 1 & 0 & -1 & 0 & \cdots\cdots & 0 & -1) \\
x_1 = 1 & x_4 = 1 & x_7 = 1 & x_{11} = 1 & \cdots\cdots & x_{\lfloor \frac{n-1}{3} \rfloor 3+1} = 1 & \\
x'_1 = 1 & x'_4 = 1 & x'_7 = 0 & x'_{11} = 0 & \cdots\cdots & x'_{\lfloor \frac{n-1}{3} \rfloor 3+1} = 0 &
\end{array}
$$

- Run of 0s delays error detection $\rightarrow$ **restrict run-length of 0s** in all $\Delta^\alpha(\boldsymbol{x})$!

Error-correcting Code

- restrict run-length of 0s in all read sub-derivatives,

Construction

$$\mathcal{C}(n,\ell) = \{\boldsymbol{x} \in \Sigma_2^n : \Delta(\boldsymbol{x}) \bmod 2 \in RLL(\log 2(n+\ell)),$$

where $\Delta(\boldsymbol{x}) = \Delta^0(\boldsymbol{x}) \circ \Delta^1(\boldsymbol{x}) \circ \cdots \circ \Delta^{\ell-1}(\boldsymbol{x})$

C. Schoeny *et al.*, "Codes correcting a burst of deletions or insertions," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1971–1985, Apr. 2017

Error-correcting Code

- restrict run-length of 0s in all read sub-derivatives,
- Correct 1 substitution in length-a window with folded Hamming code $\mathcal{H}(n, a)$:

$$\underbrace{\begin{bmatrix}\mathbf{H}_a & \mathbf{H}_a & \cdots & \mathbf{H}_a\end{bmatrix}}_{\frac{n}{2^a-1}\text{ times}}$$

where $\mathbf{H}_a$ corresponds to Hamming code of order a.

Construction

$$\begin{aligned}\mathcal{C}(n,\ell) = \{&\boldsymbol{x} \in \Sigma_2^n : \Delta(\boldsymbol{x}) \bmod 2 \in RLL(\log 2(n+\ell)),\\ &\mathcal{R}^{\pi}(\boldsymbol{x}) \bmod 2 \in \mathcal{H}(n+\ell-1, \log\log 8(n+\ell)+1)\},\end{aligned}$$

where $\Delta(\boldsymbol{x}) = \Delta^0(\boldsymbol{x}) \circ \Delta^1(\boldsymbol{x}) \circ \cdots \circ \Delta^{\ell-1}(\boldsymbol{x})$ and $\mathcal{R}^{\pi}(\boldsymbol{x})$ is a permutation of $\mathcal{R}(\boldsymbol{x})$.

C. Schoeny *et al.*, "Codes correcting a burst of deletions or insertions," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1971–1985, Apr. 2017

Error-correcting Code

- restrict run-length of 0s in all read sub-derivatives,
- Correct 1 substitution in length-a window with folded Hamming code $\mathcal{H}(n, a)$:

$$\underbrace{\begin{bmatrix}\mathbf{H}_a & \mathbf{H}_a & \cdots & \mathbf{H}_a\end{bmatrix}}_{\frac{n}{2^a-1}\text{ times}}$$

where $\mathbf{H}_a$ corresponds to Hamming code of order a.

Construction

$$\begin{aligned}\mathcal{C}(n, \ell) = &\{\boldsymbol{x} \in \Sigma_2^n : \Delta(\boldsymbol{x}) \bmod 2 \in RLL(\log 2(n+\ell)),\\ &\mathcal{R}^\pi(\boldsymbol{x}) \bmod 2 \in \mathcal{H}(n+\ell-1, \log\log 8(n+\ell)+1)\},\end{aligned}$$

where $\Delta(\boldsymbol{x}) = \Delta^0(\boldsymbol{x}) \circ \Delta^1(\boldsymbol{x}) \circ \cdots \circ \Delta^{\ell-1}(\boldsymbol{x})$ and $\mathcal{R}^\pi(\boldsymbol{x})$ is a permutation of $\mathcal{R}(\boldsymbol{x})$.

$\rightarrow$ Requires $\log\log n + o(1)$ redundant bits (optimal up to a constant).

C. Schoeny *et al.*, “Codes correcting a burst of deletions or insertions,” *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1971–1985, Apr. 2017

Introduction

Channel Model

Minimum Redundancy

Error-correcting Code

Conclusion

Summary

Results

For a simplified model of nanopore sequencing,

- Established that $\log \log n$ redundancy needed to correct 1 substitution
- Proposed an error-correcting construction that is optimal up to a constant.

Future work

- Multiple substitutions
- Deletions
- Extension to non-binary, $\delta > 1$.

Thank you!

References I

[1] R. Heckel, G. Mikutis, and R. N. Grass, “A characterization of the DNA data storage channel,” *Scientific Reports*, vol. 9, no. 9663, Jul. 2019.

[2] D. Deamer, M. Akeson, and D. Branton, “Three decades of nanopore sequencing,” *Nature Biotechnology*, vol. 34, no. 5, pp. 518–524, May 2016.

[3] A. H. Laszlo *et al.*, “Decoding long nanopore sequencing reads of natural DNA,” *Nature Biotechnology*, vol. 32, no. 8, pp. 829–833, Aug. 2014.

[4] W. Mao, S. N. Diggavi, and S. Kannan, “Models and information-theoretic bounds for nanopore sequencing,” *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3216–3236, Apr. 2018.

[5] R. Hulett, S. Chandak, and M. Wootters, “On Coding for an Abstracted Nanopore Channel for DNA Storage,” in *IEEE Intl. Symp. on Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2465–2470.

[6] B. McBain, E. Viterbo, and J. Saunderson, “Finite-state semi-markov channels for nanopore sequencing,” in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Espoo, Finland, Jun. 2022, pp. 216–221.

[7] Y. M. Chee, A. Vardy, V. K. Vu, and E. Yaakobi, “Coding for transverse-reads in domain wall memories,” in *IEEE Intl. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 2924–2929.

References II

[8] D. E. Knuth, “The sandwich theorem,” *The Electronic Journal of Combinatorics*, vol. 1, no. 1, A1, Apr. 1994.

[9] J. Chrisnata, H. M. Kiah, and E. Yaakobi, “Correcting deletions with multiple reads,” *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7141–7158, Nov. 2022.

[10] C. Schoeny, A. Wachter-Zeh, R. Gabrys, and E. Yaakobi, “Codes correcting a burst of deletions or insertions,” *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1971–1985, Apr. 2017.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ 1\ ,\ 3\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (0 , \boxed{1} , \boxed{3} , 2 , 2 , 2 , 1 , 0 , 0)$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\boxed{0}, 1, 3, 2, 2, 2, 1, 0, 0)$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \rightarrow x_1 = 0$

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ \boxed{1},\ 3\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \rightarrow x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \rightarrow x_2 = 1$

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ 1\ ,\ \boxed{3}\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \rightarrow x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \rightarrow x_2 = 1$
  - $\mathcal{R}(\boldsymbol{x})_3 = \mathrm{wt}(\boldsymbol{x}_1^3) = 3 \rightarrow x_3 = 2 \notin \{0, 1\} \rightarrow$ Error pos. 3.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ 1\ ,\ \circledast, \ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \rightarrow x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \rightarrow x_2 = 1$
  - $\mathcal{R}(\boldsymbol{x})_3 = \mathrm{wt}(\boldsymbol{x}_1^3) = 3 \rightarrow x_3 = 2 \notin \{0, 1\} \rightarrow$ Error pos. 3.
- Consider erasure.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ 1\ ,\ \boxed{\circledast}\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \rightarrow$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \rightarrow x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \rightarrow x_2 = 1$
  - $\mathcal{R}(\boldsymbol{x})_3 = \mathrm{wt}(\boldsymbol{x}_1^3) = 3 \rightarrow x_3 = 2 \notin \{0, 1\} \rightarrow$ Error pos. 3.
- Consider erasure.
- $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0$

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = ($ 0 , 1 , 1 , 2 , 2 , 2 , 1 , 0 , 0 $)$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \to$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \to x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \to x_2 = 1$
  - $\mathcal{R}(\boldsymbol{x})_3 = \mathrm{wt}(\boldsymbol{x}_1^3) = 3 \to x_3 = 2 \notin \{0, 1\} \to$ Error pos. 3.
- Consider erasure.
- $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0 \to \mathcal{R}(\boldsymbol{x})_3 = 1$.

Correctable Substitutions

Example ($\ell = 3, \delta = 1$)

We receive $\mathcal{R}(\boldsymbol{x})' = (\ 0\ ,\ 1\ ,\ 1\ ,\ 2\ ,\ 2\ ,\ 2\ ,\ 1\ ,\ 0\ ,\ 0\ )$.

- $|\mathcal{R}(\boldsymbol{x})'_3 - \mathcal{R}(\boldsymbol{x})'_2| \not\leq 1 \to$ Error pos. 2 or 3.
- Left-to-right reconstruction:
  - $\mathcal{R}(\boldsymbol{x})_1 = \mathrm{wt}(x_1) = 0 \to x_1 = 0$
  - $\mathcal{R}(\boldsymbol{x})_2 = \mathrm{wt}(\boldsymbol{x}_1^2) = 1 \to x_2 = 1$
  - $\mathcal{R}(\boldsymbol{x})_3 = \mathrm{wt}(\boldsymbol{x}_1^3) = 3 \to x_3 = 2 \notin \{0, 1\} \to$ Error pos. 3.
- Consider erasure.
- $\sum_i \mathcal{R}(\boldsymbol{x})_i \bmod \ell = 0 \to \mathcal{R}(\boldsymbol{x})_3 = 1$.

Takeaway

If error magnitude>1, can correct immediately.
$\to$ Sufficient to consider ± 1 errors!