Information-Theoretic Limits for Streaming Communication

Vincent Y. F. Tan

Joint work with Si-Hyeon Lee and Ashish Khisti (Toronto)

National University of Singapore (NUS)

Beyond IID, 2016 (Barcelona)

Outline

Outline

1. Background and Streaming Setup

2. Achievability Results and Proof Sketches

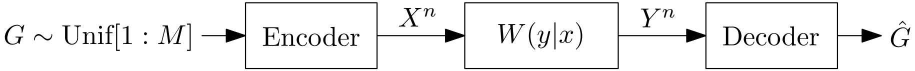
Outline

Outline

Outline

Outline

Block-Coding Formalism



Block channel coding

- Consider a discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$

Block-Coding Formalism

$$G \sim \mathrm{Unif}[1:M] \longrightarrow \boxed{\text{Encoder}} \xrightarrow{X^n} \boxed{W(y|x)} \xrightarrow{Y^n} \boxed{\text{Decoder}} \longrightarrow \hat{G}$$

Block channel coding

- Consider a discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$
- Fundamental problem in information theory is the interplay between n, R and ε

Block-Coding Formalism

$$G \sim \mathrm{Unif}[1:M] \longrightarrow \boxed{\text{Encoder}} \xrightarrow{X^n} \boxed{W(y|x)} \xrightarrow{Y^n} \boxed{\text{Decoder}} \longrightarrow \hat{G}$$

Block channel coding

- Consider a discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$
- Fundamental problem in information theory is the interplay between n, R and ε
- Shannon (1948) showed that the maximum rate of communication with $\varepsilon \to 0$ as $n \to \infty$ is

$$C(W) = \max_{P_X} I(X;Y) = \max_{P_X} I(P_X, W)$$

Refined Asymptotics

Regime	Large deviations [Gallager '65] [Shannon et al. '67]	Moderate deviations [Altüg, Wagner '14] [Polyanskiy, Verdú '10]	Central limit [Strassen '62] [Hayashi '09] [Polyanskiy et al. '10]
Code Rate	$R < C$	$R = C - \rho_n$ $\rho_n \to 0$ and $n\rho_n^2 \to \infty$	$R \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\varepsilon)$
Error Prob.	Exponential $\varepsilon \le \exp\{-nE_r(R)\}$	Subexponential $\varepsilon \approx \exp\left\{-\frac{n\rho_n^2}{2V}\right\}$	Non-vanishing $\varepsilon > 0$

Refined Asymptotics

Regime	Large deviations [Gallager '65] [Shannon et al. '67]	Moderate deviations [Altüg, Wagner '14] [Polyanskiy, Verdú '10]	Central limit [Strassen '62] [Hayashi '09] [Polyanskiy et al. '10]
Code Rate	$R < C$	$R = C - \rho_n$ $\rho_n \to 0$ and $n\rho_n^2 \to \infty$	$R = C - \frac{L}{\sqrt{n}}$ $L > 0$
Error Prob.	Exponential $\varepsilon \le \exp\{-nE_r(R)\}$	Subexponential $\varepsilon \approx \exp\left\{-\frac{n\rho_n^2}{2V}\right\}$	Non-vanishing $\varepsilon \approx Q\left(\frac{L}{\sqrt{V}}\right)$

Refined Asymptotics

Regime	Large deviations [Gallager '65] [Shannon et al. '67]	Moderate deviations [Altüg, Wagner '14] [Polyanskiy, Verdú '10]	Central limit [Strassen '62] [Hayashi '09] [Polyanskiy et al. '10]
Code Rate	$R < C$	$R = C - \rho_n$ $\rho_n \to 0$ and $n\rho_n^2 \to \infty$	$R = C - \frac{L}{\sqrt{n}}$ $L > 0$
Error Prob.	Exponential $\varepsilon \le \exp\{-nE_r(R)\}$	Subexponential $\varepsilon \approx \exp\left\{-\frac{n\rho_n^2}{2V}\right\}$	Non-vanishing $\varepsilon \approx Q\left(\frac{L}{\sqrt{V}}\right)$

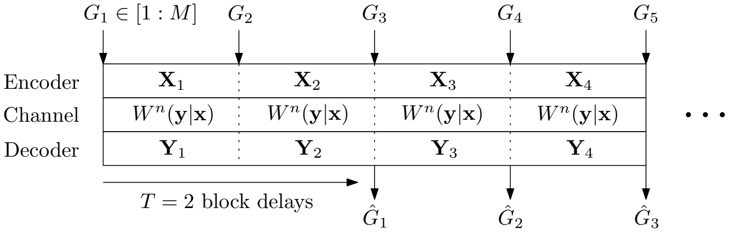
- Channel dispersion

$$V = V(W) = \min_{P_X : I(P_X, W) = C(W)} \mathrm{var}(i(X;Y))$$

- Smaller dispersion is better.

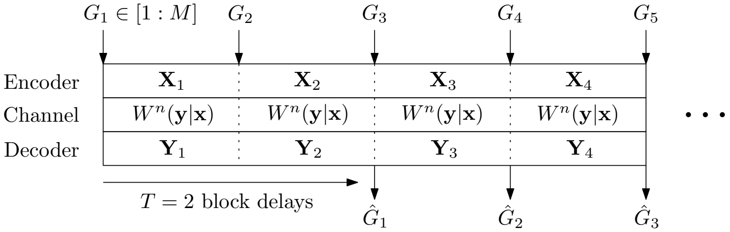
Streaming Services

An Information-Theoretic Model



Streaming setup

An Information-Theoretic Model



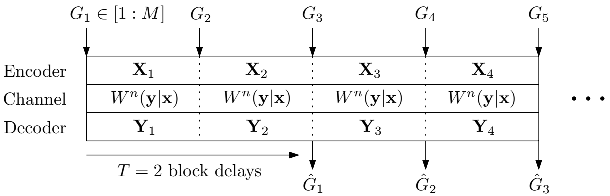
Streaming setup

- An (n, M, ε, T)-streaming code consists of
  1. a sequence of messages $\{G_k\}_{k \ge 1}$ each uniform over $\mathcal{G} = [1:M]$;
  2. a sequence of encoders $\phi_k : \mathcal{G}^k \to \mathcal{X}^n$ s.t. $\phi_k(G^k) = \mathbf{X}_k$;
  3. a sequence of decoders $\psi_k : \mathcal{Y}^{(k+T-1)n} \to \mathcal{G}$ s.t. $\psi_k(\mathbf{Y}^{k+T-1}) = \hat{G}_k$.

that satisfies

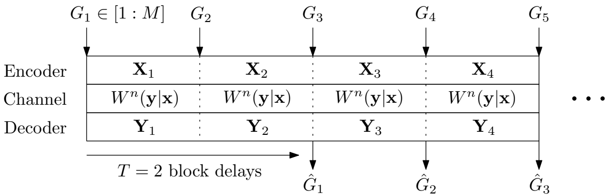
$$\limsup_{N \to \infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k)}{N} \le \varepsilon.$$

Streaming Setup



Streaming setup

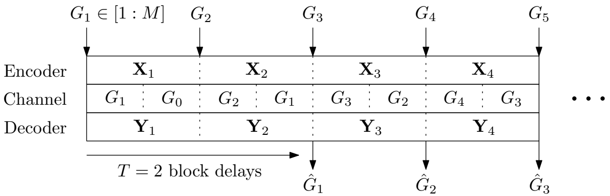
Streaming Setup



Streaming setup

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?

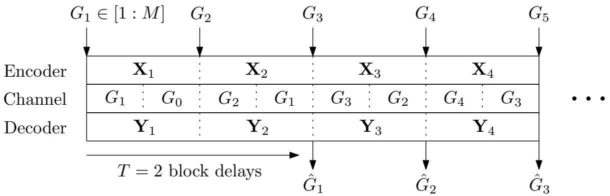
Streaming Setup



Time sharing

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?
- Time sharing?

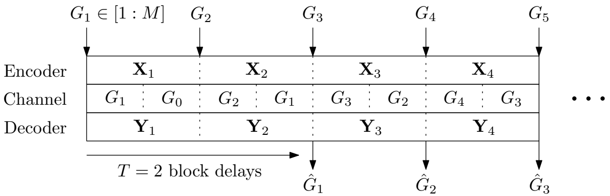
Streaming Setup



Time sharing

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?
- Time sharing?
  - For block fading channels with constant fading gain for each block, this attains the optimal diversity-multiplexing tradeoff [Khisti, Draper '14]

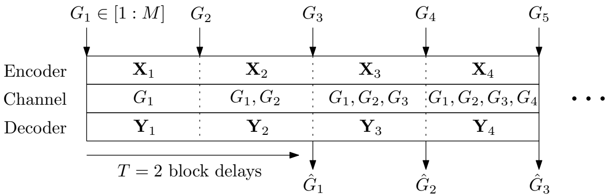
Streaming Setup



Time sharing

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?
- Time sharing?
  - For block fading channels with constant fading gain for each block, this attains the optimal diversity-multiplexing tradeoff [Khisti, Draper '14]
  - No gain in our setup due to the memoryless nature of W^n

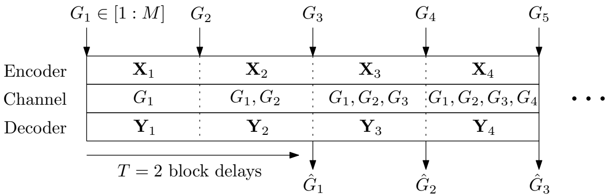
Streaming Setup



Joint encoding

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?
- Time sharing?
  - For block fading channels with constant fading gain for each block, this attains the optimal diversity-multiplexing tradeoff [Khisti, Draper '14]
  - No gain in our setup due to the memoryless nature of W^n
- Joint encoding?

Streaming Setup



Joint encoding

- Inherent tension in utilizing a block:
  Use codeword only for fresh msg vs. also for previous msges?
- Time sharing?
  - For block fading channels with constant fading gain for each block, this attains the optimal diversity-multiplexing tradeoff [Khisti, Draper '14]
  - No gain in our setup due to the memoryless nature of W^n
- Joint encoding?
  - Does improve

Outline

First Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

First Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

There exists a sequence of $(n, M_n, \varepsilon_n, T)$*-streaming codes such that*

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \leq -\frac{T}{2V}$$

Interpretation of Moderate Deviations Result

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \le -\frac{T}{2V}$$

- In block coding,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n^* = -\frac{1}{2V}$$

Interpretation of Moderate Deviations Result

$$\overline{\lim_{n\to\infty}} \frac{1}{n\rho_n^2} \log \varepsilon_n \le -\frac{T}{2V}$$

- In block coding,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n^* = -\frac{1}{2V}$$

- Hence, moderate deviations constant improves (increases) by a factor of T

Interpretation of Moderate Deviations Result

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \le -\frac{T}{2V}$$

- In block coding,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n^* = -\frac{1}{2V}$$

- Hence, moderate deviations constant improves (increases) by a factor of T
- Dispersion V is reduced by a factor of T

Second Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

For any $L > 0$, let the message size grow as

$$\log M_n = n \left(C - \frac{L}{\sqrt{n}} \right).$$

Second Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

For any $L > 0$, let the message size grow as

$$\log M_n = n\left(C - \frac{L}{\sqrt{n}}\right).$$

Then there exists a sequence of $(n, M_n, \varepsilon_n, T)$-streaming codes s.t.

$$\varepsilon_n \lesssim c \cdot Q\left(\sqrt{\frac{T}{V}}L\right) \qquad c \approx 1.$$

Second Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

For any $L > 0$, let the message size grow as

$$\log M_n = n\left(C - \frac{L}{\sqrt{n}}\right).$$

Then there exists a sequence of $(n, M_n, \varepsilon_n, T)$-streaming codes s.t.

$$\varepsilon_n \lesssim c \cdot Q\left(\sqrt{\frac{T}{V}}L\right) \qquad c \approx 1.$$

- In block coding,

$$\lim_{n\to\infty} \varepsilon_n^* = Q\left(\frac{L}{\sqrt{V}}\right)$$

Second Main Result (Achievability)

Theorem (Lee-T.-Khisti (2015))

For any $L > 0$, let the message size grow as

$$\log M_n = n\left(C - \frac{L}{\sqrt{n}}\right).$$

Then there exists a sequence of $(n, M_n, \varepsilon_n, T)$-streaming codes s.t.

$$\varepsilon_n \lesssim c \cdot Q\left(\sqrt{\frac{T}{V}}L\right) \qquad c \approx 1.$$

- In block coding,

$$\lim_{n\to\infty} \varepsilon_n^* = Q\left(\frac{L}{\sqrt{V}}\right)$$

- Dispersion V is approx. reduced by a factor of T

Summary of Main Results

Regime	Moderate deviations	Central limit
Operating rate	$R = C - \rho_n$ $\rho_n \to 0$ and $n\rho_n^2 \to \infty$	$R = C - \frac{L}{\sqrt{n}}$ $L > 0$
Error Prob.	$\varepsilon \approx \exp\left\{-\frac{Tn\rho_n^2}{2V}\right\}$	$\varepsilon \approx Q\left(L\sqrt{\frac{T}{V}}\right)$
	$V \to V/T$	
Encoding	Joint encoding of previous and fresh msges	
Decoding	Sequential decoding of previous and new msges	
	Accumulation of error probabilities	
Key innovation	Non-asymptotic moderate deviations theorem	Truncated memory structure

Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving P_X. For each message $g \in [1 : M]$, generate $\mathbf{x}(g)$ indep. according to P_X^n.

Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving P_X. For each message $g \in [1 : M]$, generate $\mathbf{x}(g)$ indep. according to P_X^n.
- Encoding: If g is the message, send $\mathbf{x}(g)$

Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving P_X. For each message $g \in [1:M]$, generate $\mathbf{x}(g)$ indep. according to P_X^n.
- Encoding: If g is the message, send $\mathbf{x}(g)$
- Decoding: If there exists a unique $g \in [1:M]$ such that

$$i(\mathbf{x}(g);\mathbf{y}) > \log M, \quad \text{where} \quad i(\mathbf{x};\mathbf{y}) = \log \frac{W^n(\mathbf{y}|\mathbf{x})}{(P_X W)^n(\mathbf{y})}$$

  let $\hat{G} = g$.

Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving P_X. For each message $g \in [1:M]$, generate $\mathbf{x}(g)$ indep. according to P_X^n.
- Encoding: If g is the message, send $\mathbf{x}(g)$
- Decoding: If there exists a unique $g \in [1:M]$ such that

$$i(\mathbf{x}(g);\mathbf{y}) > \log M, \quad \text{where} \quad i(\mathbf{x};\mathbf{y}) = \log \frac{W^n(\mathbf{y}|\mathbf{x})}{(P_X W)^n(\mathbf{y})}$$

  let $\hat{G} = g$.
- Error analysis:

$$\varepsilon \le \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2)$$

  where

$$\mathcal{E}_1 := \{i(\mathbf{X}(G);\mathbf{Y}) \le \log M\}$$
$$\mathcal{E}_2 := \{\exists\, \tilde{g} \ne G \text{ s.t. } i(\mathbf{X}(\tilde{g});\mathbf{Y}) > \log M\}$$

  Note that $\mathcal{E}_1$ is dominant in both regimes

Analysis of Error Probability

- Probability of error

$$\varepsilon_n \approx \Pr\left(i(\mathbf{X}(G);\mathbf{Y}) \le M\right) = \Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right)$$

where

$$Z_l := \log \frac{W(Y_l|X_l)}{P_X W(Y_l)}, \quad l = 1, \ldots, n$$

are i.i.d. random variables.

Analysis of Error Probability

- Probability of error

$$\varepsilon_n \approx \Pr\left(i(\mathbf{X}(G);\mathbf{Y}) \le M\right) = \Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right)$$

where

$$Z_l := \log \frac{W(Y_l|X_l)}{P_X W(Y_l)}, \quad l = 1, \ldots, n$$

are i.i.d. random variables.

- Note that for all $l \in [1:n]$,

$$\mathbb{E}[Z_l] = C \quad \text{and} \quad \mathrm{var}[Z_l] = V$$

Analysis of Error Probability

- Probability of error

$$\varepsilon_n \approx \Pr\left(i(\mathbf{X}(G);\mathbf{Y}) \le M\right) = \Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right)$$

where

$$Z_l := \log \frac{W(Y_l|X_l)}{P_X W(Y_l)}, \quad l = 1, \ldots, n$$

are i.i.d. random variables.

- Note that for all $l \in [1:n]$,

$$\mathbb{E}[Z_l] = C \quad \text{and} \quad \mathrm{var}[Z_l] = V$$

- Under various regimes, analyze

$$\Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right).$$

Analysis of Error: Moderate Deviations Regime

Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))

Under regularity conditions on Z_l, *and* $\log M = n(C - \rho_n)$,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right) = -\frac{1}{2V}.$$

Analysis of Error: Moderate Deviations Regime

Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))

Under regularity conditions on Z_l, and $\log M = n(C - \rho_n)$,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{n} Z_l \le \log M \right) = -\frac{1}{2V}.$$

- Thus, we have

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \le -\frac{1}{2V}.$$

Analysis of Error: Moderate Deviations Regime

Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))

Under regularity conditions on Z_l, and $\log M = n(C - \rho_n)$,

$$\lim_{n\to\infty} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right) = -\frac{1}{2V}.$$

- Thus, we have

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \le -\frac{1}{2V}.$$

- However, note that the standard MD theorem is asymptotic in nature
- We need a non-asymptotic version in the streaming scenario

Analysis of Error: Central Limit Regime

Theorem (Berry-Esseen Theorem)

Under regularity conditions on Z_l, and

$$\log M = nC - \sqrt{n}L,$$

we have

$$\Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right) = Q\left(\frac{L}{\sqrt{V}}\right) \pm \frac{\tau}{\sqrt{n}}.$$

where τ is a constant (depending on Z_1).

Analysis of Error: Central Limit Regime

Theorem (Berry-Esseen Theorem)

Under regularity conditions on Z_l, and

$$\log M = nC - \sqrt{n}L,$$

we have

$$\Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right) = Q\left(\frac{L}{\sqrt{V}}\right) \pm \frac{\tau}{\sqrt{n}}.$$

where τ is a constant (depending on Z_1).

- Thus, we have

$$\varepsilon_n \le Q\left(\frac{L}{\sqrt{V}}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

Analysis of Error: Central Limit Regime

Theorem (Berry-Esseen Theorem)

Under regularity conditions on Z_l, and

$$\log M = nC - \sqrt{n}L,$$

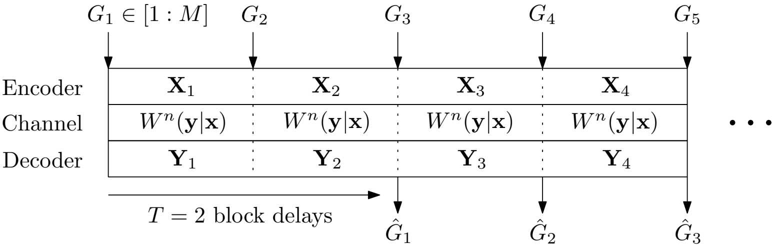
we have

$$\Pr\left(\sum_{l=1}^{n} Z_l \le \log M\right) = Q\left(\frac{L}{\sqrt{V}}\right) \pm \frac{\tau}{\sqrt{n}}.$$

where τ is a constant (depending on Z_1).

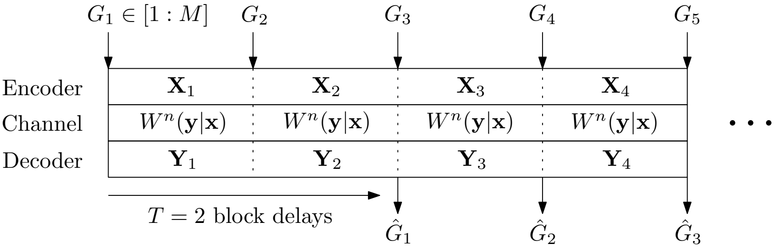
- Thus, we have
$$\varepsilon_n \le Q\left(\frac{L}{\sqrt{V}}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$
- However, note that the Berry-Esseen residual terms hurt us in the streaming setup

[Streaming setup] Moderate deviations regime



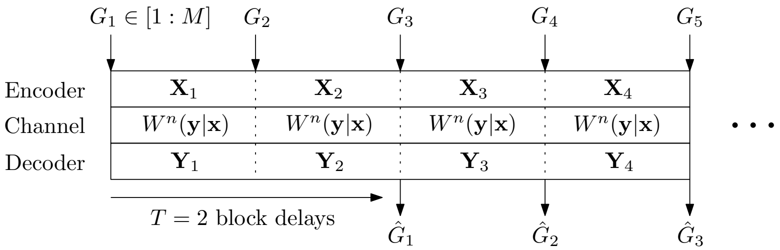
- Consider block delay $T = 2$.

[Streaming setup] Moderate deviations regime



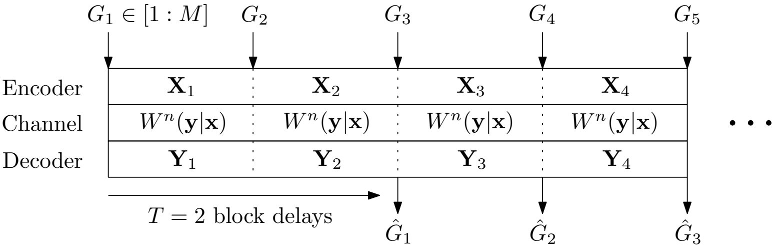
- Consider block delay $T = 2$.
- Codebook generation for block k: For each $g^k \in [1:M]^k$, generate $\mathbf{x}_k(g^k)$ in an i.i.d. manner according to P_X that achieves the dispersion.

[Streaming setup] Moderate deviations regime



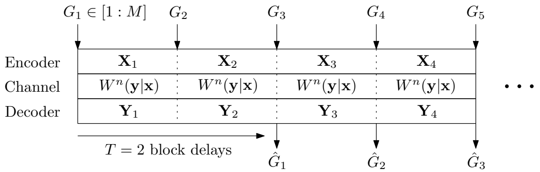
- Consider block delay $T = 2$.
- Codebook generation for block k: For each $g^k \in [1:M]^k$, generate $\mathbf{x}_k(g^k)$ in an i.i.d. manner according to P_X that achieves the dispersion.
- Encoding at block k: Send $\mathbf{x}_k(G_1, \cdots, G_k)$.

[Streaming setup] Moderate deviations regime



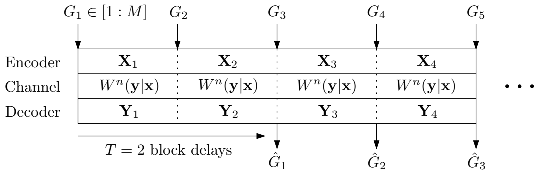
- Consider block delay $T = 2$.
- Codebook generation for block k: For each $g^k \in [1:M]^k$, generate $\mathbf{x}_k(g^k)$ in an i.i.d. manner according to P_X that achieves the dispersion.
- Encoding at block k: Send $\mathbf{x}_k(G_1, \cdots, G_k)$.
- Decoding at block $k+1$:
  - Target message: G_k
  - Due to joint encoding, G_k is in error if any of $\hat{G}_1, \cdots, \hat{G}_{k-1}$ is in error.
  - Sequentially decode $G_1, \cdots, G_k$.

[Streaming Setup] Moderate Deviations Regime



- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.

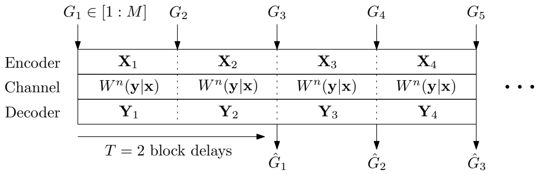
[Streaming Setup] Moderate Deviations Regime



- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.
- Re-decode G_1: Choose $\hat{G}_1'$ as a unique $g_1 \in [1:M]$ such that

$$i([\mathbf{x}_1(g_1)\ \mathbf{x}_2(g_1, g_2)\ \mathbf{x}_3(g_1, g_2, g_3)], [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]) > 3\log M, \quad \text{for some} \quad g_2, g_3$$

[Streaming Setup] Moderate Deviations Regime



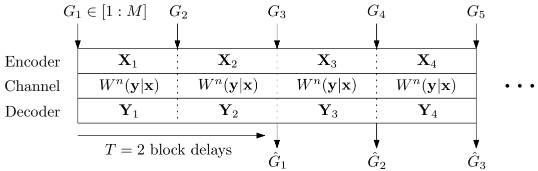
- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.
- Re-decode G_1: Choose $\hat{G}_1'$ as a unique $g_1 \in [1:M]$ such that

$$i([\mathbf{x}_1(g_1)\ \mathbf{x}_2(g_1, g_2)\ \mathbf{x}_3(g_1, g_2, g_3)], [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]) > 3\log M, \quad \text{for some} \quad g_2, g_3$$

- Decode G_2: Choose $\hat{G}_2$ as a unique $g_2 \in [1:M]$ such that

$$i([\mathbf{x}_2(\hat{G}_1', g_2)\ \mathbf{x}_3(\hat{G}_1', g_2, g_3)], [\mathbf{y}_2\ \mathbf{y}_3]) > 2\log M \quad \text{for some} \quad g_3$$

[Streaming Setup] Moderate Deviations Regime



- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.
- Re-decode G_1: Choose $\hat{G}_1'$ as a unique $g_1 \in [1:M]$ such that

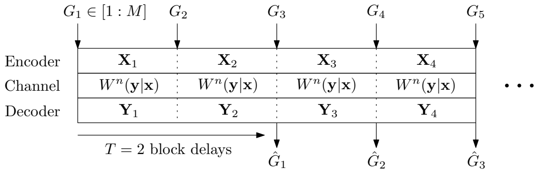
$$i([\mathbf{x}_1(g_1)\ \mathbf{x}_2(g_1, g_2)\ \mathbf{x}_3(g_1, g_2, g_3)], [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]) > 3\log M, \quad \text{for some} \quad g_2, g_3$$

- Decode G_2: Choose $\hat{G}_2$ as a unique $g_2 \in [1:M]$ such that

$$i([\mathbf{x}_2(\hat{G}_1', g_2)\ \mathbf{x}_3(\hat{G}_1', g_2, g_3)], [\mathbf{y}_2\ \mathbf{y}_3]) > 2\log M \quad \text{for some} \quad g_3$$

- $\Pr(\hat{G}_2 \neq G_2) \leq \Pr((\hat{G}_1' \neq G_1) \cup (\hat{G}_2 \neq G_2))$
  $\approx \Pr(\sum_{l=1}^{3n} Z_l \leq 3\log M) + \Pr(\sum_{l=1}^{2n} Z_l \leq 2\log M)$

[Streaming Setup] Moderate Deviations Regime



- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.
- Re-decode G_1: Choose $\hat{G}_1'$ as a unique $g_1 \in [1:M]$ such that

$$i([\mathbf{x}_1(g_1)\ \mathbf{x}_2(g_1, g_2)\ \mathbf{x}_3(g_1, g_2, g_3)], [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]) > 3 \log M, \quad \text{for some} \quad g_2, g_3$$

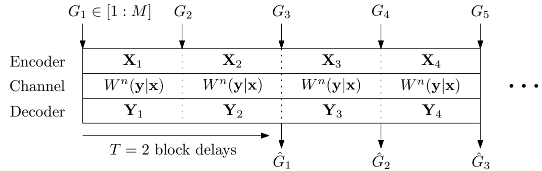
- Decode G_2: Choose $\hat{G}_2$ as a unique $g_2 \in [1:M]$ such that

$$i([\mathbf{x}_2(\hat{G}_1', g_2)\ \mathbf{x}_3(\hat{G}_1', g_2, g_3)], [\mathbf{y}_2\ \mathbf{y}_3]) > 2 \log M \quad \text{for some} \quad g_3$$

- $\Pr(\hat{G}_2 \neq G_2) \leq \Pr((\hat{G}_1' \neq G_1) \cup (\hat{G}_2 \neq G_2))$
  $\approx \Pr(\sum_{l=1}^{3n} Z_l \leq 3 \log M) + \Pr(\sum_{l=1}^{2n} Z_l \leq 2 \log M)$
- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=2}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)$$

[Streaming Setup] Moderate Deviations Regime



- $T = 2$: At the end block 3, sequentially decode G_1 and G_2.
- Re-decode G_1: Choose $\hat{G}_1'$ as a unique $g_1 \in [1:M]$ such that

$$i([\mathbf{x}_1(g_1)\ \mathbf{x}_2(g_1, g_2)\ \mathbf{x}_3(g_1, g_2, g_3)], [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]) > 3 \log M, \quad \text{for some} \quad g_2, g_3$$

- Decode G_2: Choose $\hat{G}_2$ as a unique $g_2 \in [1:M]$ such that

$$i([\mathbf{x}_2(\hat{G}_1', g_2)\ \mathbf{x}_3(\hat{G}_1', g_2, g_3)], [\mathbf{y}_2\ \mathbf{y}_3]) > 2 \log M \quad \text{for some} \quad g_3$$

- $\Pr(\hat{G}_2 \neq G_2) \leq \Pr((\hat{G}_1' \neq G_1) \cup (\hat{G}_2 \neq G_2))$
  $\approx \Pr(\sum_{l=1}^{3n} Z_l \leq 3 \log M) + \Pr(\sum_{l=1}^{2n} Z_l \leq 2 \log M)$
- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)$$

[Streaming Analysis] Moderate Deviations Regime

- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M \right)$$

[Streaming Analysis] Moderate Deviations Regime

- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)$$

- However, recall that the standard moderate deviations theorem is asymptotic, i.e.,

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right) \leq -\frac{j}{2V}$$

[Streaming Analysis] Moderate Deviations Regime

- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M \right)$$

- However, recall that the standard moderate deviations theorem is asymptotic, i.e.,

$$\overline{\lim_{n\to\infty}} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M \right) \leq -\frac{j}{2V}$$

- Cannot "exchange limits"

[Streaming Analysis] Moderate Deviations Regime

- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M \right)$$

- However, recall that the standard moderate deviations theorem is asymptotic, i.e.,

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M \right) \leq -\frac{j}{2V}$$

- Cannot "exchange limits"

- Need to develop a non-asymptotic upper bound for moderate deviations theorem [Altüg and Wagner (2014)]

[Streaming Analysis] Moderate Deviations Regime

Lemma

Under regularity conditions on Z_l, for any positive ρ_n satisfying $\rho_n \to 0$ and $n\rho_n^2 \to \infty$,

$$\Pr\left(\frac{1}{n}\sum_{l=1}^{n} Z_l \geq \rho_n\right) \leq \exp\left\{-n\left(\frac{\rho_n^2}{2\sigma^2} - \frac{\rho_n^3}{6\sigma^6}K\right)\right\}$$

where K is a constant that only depends on Z_1.

[Streaming Analysis] Moderate Deviations Regime

Lemma

Under regularity conditions on Z_l, for any positive ρ_n satisfying $\rho_n \to 0$ and $n\rho_n^2 \to \infty$,

$$\Pr\left(\frac{1}{n}\sum_{l=1}^{n} Z_l \geq \rho_n\right) \leq \exp\left\{-n\left(\frac{\rho_n^2}{2\sigma^2} - \frac{\rho_n^3}{6\sigma^6}K\right)\right\}$$

where K is a constant that only depends on Z_1.

Using the lemma, we conclude that for all $k \in \mathbb{N}$,

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log\left[\varlimsup_{n\to\infty} \frac{1}{N}\sum_{k=1}^{N} \Pr(\hat{G}_k \neq G_k)\right] \leq -\frac{T}{2V}.$$

[Streaming Analysis] Central limit regime

- For all $k \in \mathbb{N}$,

$$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)$$
$$= \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j(nR - L\sqrt{n})\right)$$

- Compared to the block coding case, $n \to jn$, $L \to \sqrt{j}L$.

[Streaming Analysis] Central limit regime

- For all $k \in \mathbb{N}$,

$$\begin{aligned}\Pr(\hat{G}_k \neq G_k) &\leq \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)\\ &= \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j(nR - L\sqrt{n})\right)\\ &\leq \sum_{j=T}^{\infty} \left(Q\left(\frac{\sqrt{j}L}{\sqrt{V}}\right) + \frac{\tau}{\sqrt{jn}}\right)\end{aligned}$$

- Compared to the block coding case, $n \to jn$, $L \to \sqrt{j}L$.

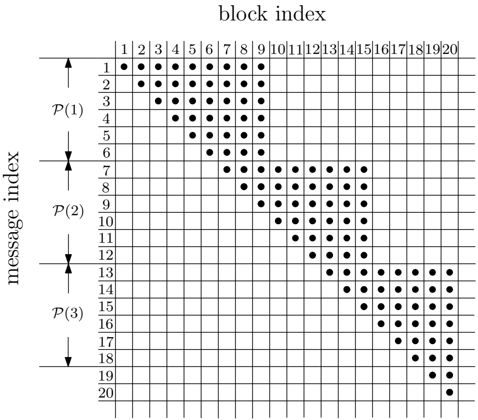
[Streaming Analysis] Central limit regime

- For all $k \in \mathbb{N}$,

$$
\begin{aligned}
\Pr(\hat{G}_k \neq G_k) &\le \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \le j \log M\right) \\
&= \sum_{j=T}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \le j(nR - L\sqrt{n})\right) \\
&\le \sum_{j=T}^{\infty} \left(Q\left(\frac{\sqrt{j}L}{\sqrt{V}}\right) + \frac{\tau}{\sqrt{jn}}\right) \\
&= \sum_{j=T}^{\infty} Q\left(\frac{\sqrt{j}L}{\sqrt{V}}\right) + \sum_{j=T}^{\infty} \frac{\tau}{\sqrt{jn}} \\
&\le c \cdot Q\left(\frac{\sqrt{T}L}{\sqrt{V}}\right) + \sum_{j=T}^{\infty} \frac{\tau}{\sqrt{jn}},
\end{aligned}
$$

  where $c \approx 1$ for a wide range of channel parameters.

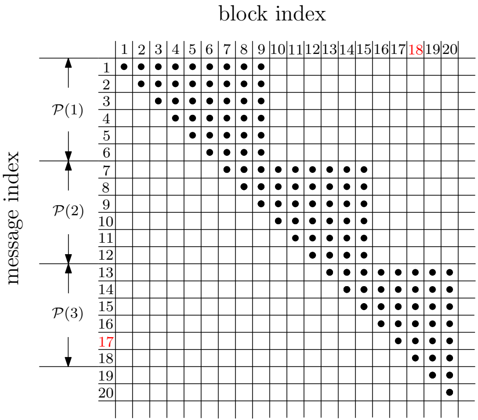
- Compared to the block coding case, $n \to jn$, $L \to \sqrt{j}L$.
- The remainder terms from the Berry-Esseen theorem diverge!



Memory structure with $A = 9$, $B = 4$

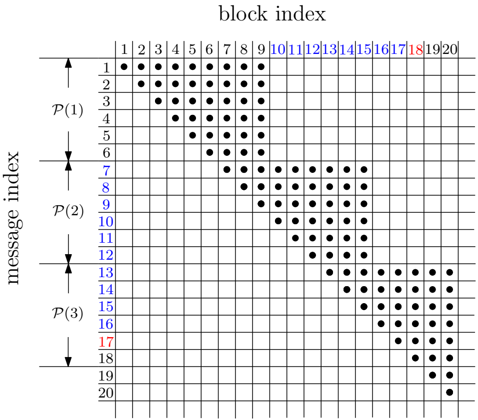
- A, B: Max/Min memories
- Decode all msgs in the previous group and all previous msgs in the current group
- Example of $A = 9$, $B = 4$, $T = 2$ To decode G_{17} at the end of block 18, decodes $G_7, \cdots, G_{17}$ by considering codewords in blocks $10, \cdots, 18$.
- Judiciously choose A and B as functions of n to balance
  - Rate penalty ($\frac{B}{A} \downarrow$)
  - Contributions to error probability
    Remainder terms ($A \downarrow$)
    Previous group ($B \uparrow$)

[Streaming Analysis] Truncated Memory



Memory structure with $A = 9$, $B = 4$

- A, B: Max/Min memories
- Decode all msgs in the previous group and all previous msgs in the current group
- Example of $A = 9$, $B = 4$, $T = 2$
  To decode G_{17} at the end of block 18, decodes $G_7, \cdots, G_{17}$ by considering codewords in blocks $10, \cdots, 18$.
- Judiciously choose A and B as functions of n to balance
  - Rate penalty ($\frac{B}{A} \downarrow$)
  - Contributions to error probability
    Remainder terms ($A \downarrow$)
    Previous group ($B \uparrow$)



Memory structure with $A = 9$, $B = 4$

- A, B: Max/Min memories
- Decode all msgs in the previous group and all previous msgs in the current group
- Example of $A = 9$, $B = 4$, $T = 2$
  To decode G_{17} at the end of block 18, decodes $G_7, \cdots, G_{17}$ by considering codewords in blocks $10, \cdots, 18$.
- Judiciously choose A and B as functions of n to balance
  - Rate penalty ($\frac{B}{A} \downarrow$)
  - Contributions to error probability
    Remainder terms ($A \downarrow$)
    Previous group ($B \uparrow$)

Outline

Extension 1: Erasure Option

- An $(n, M, \varepsilon, \varepsilon', T)$-streaming code with an erasure option is the same as the usual streaming code except that

Extension 1: Erasure Option

- An $(n, M, \varepsilon, \varepsilon', T)$-streaming code with an erasure option is the same as the usual streaming code except that
  1. the decoding functions

$$\psi_k : \mathcal{Y}^{(k+T-1)n} \to \mathcal{G} \cup \{0\}$$

     where 0 denotes the erasure option

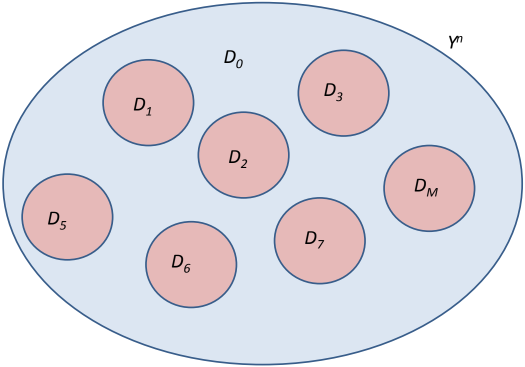
Extension 1: Erasure Option

- An $(n, M, \varepsilon, \varepsilon', T)$-streaming code with an erasure option is the same as the usual streaming code except that
  1. the decoding functions
     $$\psi_k : \mathcal{Y}^{(k+T-1)n} \to \mathcal{G} \cup \{0\}$$
     where 0 denotes the erasure option
  2. the total error probability does not exceed ε, i.e.,
     $$\limsup_{N\to\infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k)}{N} \leq \varepsilon.$$

Extension 1: Erasure Option

- An $(n, M, \varepsilon, \varepsilon', T)$-streaming code with an erasure option is the same as the usual streaming code except that
  1. the decoding functions
  $$\psi_k : \mathcal{Y}^{(k+T-1)n} \to \mathcal{G} \cup \{0\}$$
  where 0 denotes the erasure option
  2. the total error probability does not exceed ε, i.e.,
  $$\limsup_{N\to\infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k)}{N} \leq \varepsilon.$$
  3. the erasure error probability does not exceed ε', i.e.,
  $$\limsup_{N\to\infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k, \hat{G}_k \neq 0)}{N} \leq \varepsilon'.$$
- Seek upper bounds on ε and ε' when M is the moderate deviations regime.

Illustration of the Erasure Option



Decoding with an erasure option

Result for Erasure Option

Theorem

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

Result for Erasure Option

Theorem

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

There exists a sequence of $(n, M_n, \varepsilon_n, \varepsilon_n', T)$*-streaming codes with the erasure option such that*

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \leq -\frac{T(1-\gamma)^2}{2V}$$

$$\varlimsup_{n\to\infty} \frac{1}{n\rho_n} \log \varepsilon_n' \leq -T\gamma$$

for any $0 < \gamma < 1$.

Discussion of Result for Erasure Option

- The undetected error probability is

$$\varepsilon_n \le \exp\left\{-n\rho_n^2 \cdot \frac{T(1-\gamma)^2}{2V} + o(n\rho_n^2)\right\}$$

  and the total error probability is

$$\varepsilon_n' \le \exp\left\{-n\rho_n \cdot T\gamma + o(n\rho_n)\right\}$$

Discussion of Result for Erasure Option

- The undetected error probability is

$$\varepsilon_n \le \exp\left\{-n\rho_n^2 \cdot \frac{T(1-\gamma)^2}{2V} + o(n\rho_n^2)\right\}$$

  and the total error probability is

$$\varepsilon_n' \le \exp\left\{-n\rho_n \cdot T\gamma + o(n\rho_n)\right\}$$

- Total error probability is much larger than undetected error probability

Discussion of Result for Erasure Option

- The undetected error probability is

$$\varepsilon_n \le \exp\left\{-n\rho_n^2 \cdot \frac{T(1-\gamma)^2}{2V} + o(n\rho_n^2)\right\}$$

  and the total error probability is

$$\varepsilon_n' \le \exp\left\{-n\rho_n \cdot T\gamma + o(n\rho_n)\right\}$$

- Total error probability is much larger than undetected error probability
- When $T = 1$, this reduces to Theorem 1 in Hayashi-T. (Dec. 2015)

Discussion of Result for Erasure Option

- The undetected error probability is

$$\varepsilon_n \le \exp\left\{-n\rho_n^2 \cdot \frac{T(1-\gamma)^2}{2V} + o(n\rho_n^2)\right\}$$

  and the total error probability is

$$\varepsilon_n' \le \exp\left\{-n\rho_n \cdot T\gamma + o(n\rho_n)\right\}$$

- Total error probability is much larger than undetected error probability
- When $T = 1$, this reduces to Theorem 1 in Hayashi-T. (Dec. 2015)
- With $T > 1$, streaming boosts both exponents by a factor of T

Extension 2: Decoding with Variable Delay

- An (n, M, ε, T)-streaming code with an average delay constraint is the same as the usual streaming code except that

Extension 2: Decoding with Variable Delay

- An (n, M, ε, T)-streaming code with an average delay constraint is the same as the usual streaming code except that
  1. the sequence of decoding functions

$$\psi_k : \mathcal{Y}^{kn} \to (\mathcal{G} \cup \{0\})^k$$

Extension 2: Decoding with Variable Delay

- An (n, M, ε, T)-streaming code with an average delay constraint is the same as the usual streaming code except that

  1. the sequence of decoding functions

  $$\psi_k : \mathcal{Y}^{kn} \to (\mathcal{G} \cup \{0\})^k$$

  2. the average error probability is upper bounded as

  $$\varlimsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_{k+D_k+1} \neq G_k) \leq \varepsilon$$

  where $D_k := \min\{d \in \mathbb{N} : \hat{G}_{k+d-1,k} \neq 0\}$ denotes the random decoding delay of the k-th message and

Extension 2: Decoding with Variable Delay

- An (n, M, ε, T)-streaming code with an average delay constraint is the same as the usual streaming code except that
  1. the sequence of decoding functions

$$\psi_k : \mathcal{Y}^{kn} \to (\mathcal{G} \cup \{0\})^k$$

  2. the average error probability is upper bounded as

$$\varlimsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_{k+D_k+1} \neq G_k) \leq \varepsilon$$

  where $D_k := \min\{d \in \mathbb{N} : \hat{G}_{k+d-1,k} \neq 0\}$ denotes the random decoding delay of the k-th message and

  3. the average delay satisfies

$$\varlimsup_{N\to\infty} \sum_{k=1}^{N} \frac{\mathbb{E}[D_k]}{N} \leq T.$$

Result for Decoding with Variable Delay

Theorem

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

Result for Decoding with Variable Delay

Theorem

Let the message size grow as

$$\log M_n = n(C - \rho_n)$$

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$

There exists a sequence of $(n, M_n, \varepsilon_n, T_n)$*-streaming codes with average delay constraint such that*

$$\lim_{n\to\infty} T_n = T$$

$$\overline{\lim_{n\to\infty}} \frac{1}{n\rho_n} \log \varepsilon_n \leq -T$$

Discussion of Result for Decoding with Variable Delay

- For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved

Discussion of Result for Decoding with Variable Delay

- For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved
- Without variable delay,

$$\varepsilon_n \le \exp(-\Theta(n\rho_n^2))$$

Discussion of Result for Decoding with Variable Delay

- For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved
- Without variable delay,

$$\varepsilon_n \le \exp(-\Theta(n\rho_n^2))$$

- With variable delay

$$\varepsilon_n \le \exp(-n\rho_n T + o(n\rho_n))$$

Discussion of Result for Decoding with Variable Delay

- For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved
- Without variable delay,

$$\varepsilon_n \le \exp(-\Theta(n\rho_n^2))$$

- With variable delay

$$\varepsilon_n \le \exp(-n\rho_n T + o(n\rho_n))$$

- A significant gain in the can be achieved in the moderate deviations regime with streaming and variable delay without feedback

Outline

A Slightly Different Streaming Setup

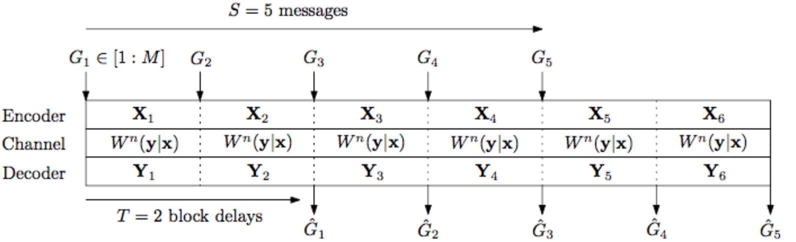
- To derive lower bounds to error probability, we consider a slightly different setup.

A Slightly Different Streaming Setup

- To derive lower bounds to error probability, we consider a slightly different setup.
- An $(n, M, \varepsilon, T, S)$-streaming code consists of
  1. a sequence of messages $\{G_k\}_{k=1}^{S}$ each uniformly distributed over $\mathcal{G} = [1:M]$
  2. a sequence of encoding functions
  $$\phi_k : \mathcal{G}^{\min\{k,S\}} \to \mathcal{X}^n \quad \text{for} \quad k \in [1:S+T-1]$$
  3. a sequence of decoding functions
  $$\psi_k : \mathcal{Y}^{(k+T-1)n} \to \mathcal{G}, \quad \text{for} \quad k \in [1:S]$$

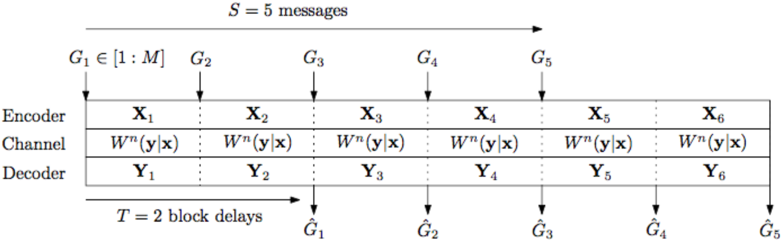
  s.t. the maximum error probability over all S msgs satisfies
  $$\max_{k \in [1:S]} \Pr(\hat{G}_k \neq G_k) \leq \varepsilon.$$

A Slightly Different Streaming Setup



$T = 2$ and $S = 5$. A total of five msges ($S = 5$) are sequentially encoded and are sequentially decoded after the delay of two blocks ($T = 2$).

A Slightly Different Streaming Setup



$T = 2$ and $S = 5$. A total of five msges ($S = 5$) are sequentially encoded and are sequentially decoded after the delay of two blocks ($T = 2$).

Fundamental limit on error probability

$$\varepsilon^*(n, M, T, S) := \min\{\varepsilon : \exists \text{ an } (n, M, \varepsilon, T, S)\text{-streaming code}\}$$

Main Converse Result

Theorem (Lee-T.-Khisti (2016))

For an output symmetric DMC with $V > 0$, consider sequences M_n and S_n such that

$$\log M_n = n(C - n^{-t}), \quad \textit{with} \quad 0 < t < 1/3,$$

and

$$S_n = \omega(n^t) \cap \exp(o(n^{1-2t})).$$

Main Converse Result

Theorem (Lee-T.-Khisti (2016))

For an output symmetric DMC with $V > 0$, consider sequences M_n and S_n such that

$$\log M_n = n(C - n^{-t}), \quad \textit{with} \quad 0 < t < 1/3,$$

and

$$S_n = \omega(n^t) \cap \exp(o(n^{1-2t})).$$

Then

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

Main Converse Result

Theorem (Lee-T.-Khisti (2016))

For an output symmetric DMC with $V > 0$, consider sequences M_n and S_n such that

$$\log M_n = n(C - n^{-t}), \quad \textit{with} \quad 0 < t < 1/3,$$

and

$$S_n = \omega(n^t) \cap \exp(o(n^{1-2t})).$$

Then

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

Matches previous moderate deviations achievability result

Discussion of Main Converse Result

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

Discussion of Main Converse Result

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

- Need to restrict to output symmetric channels because

$$E^+(R;W) = E_{sp}(R;W)$$

for output symmetric channels, where

$$E^+(R;W) := \min_{V:C(V)\le R} \max_P D(V\|W|P) \qquad \text{(Haroutunian)}$$

$$E_{sp}(R;W) := \max_P \min_{V:I(P,V)\le R} D(V\|W|P) \qquad \text{(Sphere Packing)}$$

Discussion of Main Converse Result

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

- Need to restrict to output symmetric channels because

$$E^+(R;W) = E_{sp}(R;W)$$

for output symmetric channels, where

$$E^+(R;W) := \min_{V:C(V)\le R} \max_P D(V\|W|P) \qquad \text{(Haroutunian)}$$

$$E_{sp}(R;W) := \max_P \min_{V:I(P,V)\le R} D(V\|W|P) \qquad \text{(Sphere Packing)}$$

- Range $0 < t < 1/3$ is more restrictive than the usual $0 < t < 1/2$

Discussion of Main Converse Result

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}$$

- Need to restrict to output symmetric channels because

$$E^+(R;W) = E_{sp}(R;W)$$

for output symmetric channels, where

$$E^+(R;W) := \min_{V:C(V)\le R} \max_P D(V\|W|P) \qquad \text{(Haroutunian)}$$

$$E_{sp}(R;W) := \max_P \min_{V:I(P,V)\le R} D(V\|W|P) \qquad \text{(Sphere Packing)}$$

- Range $0 < t < 1/3$ is more restrictive than the usual $0 < t < 1/2$
- Range of $S_n = \omega(n^t) \cap \exp(o(n^{1-2t}))$ is rather extensive

Proof Ideas for the Converse Part

$$\underline{\lim}_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) \geq -\frac{T}{2V}$$

Proof Ideas for the Converse Part

$$\varliminf_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) \geq -\frac{T}{2V}$$

- Step 1: Assume a feedforward decoder (genie-aided decoder)

Proof Ideas for the Converse Part

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) \geq -\frac{T}{2V}$$

- Step 1: Assume a feedforward decoder (genie-aided decoder)
- Step 2: Lower bound maximal error probability over a certain number of messages S_n^* using an auxiliary channel V_n^*

Proof Ideas for the Converse Part

$$\lim_{n\to\infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) \geq -\frac{T}{2V}$$

- Step 1: Assume a feedforward decoder (genie-aided decoder)
- Step 2: Lower bound maximal error probability over a certain number of messages S_n^* using an auxiliary channel V_n^*
- Step 3: Lower bound error probability of the maximal error message under true channel W using a change-of-measure idea due to Sahai (2008)

Step 1 of the Converse Part

Definition

A feedforward decoder consists of a sequence of decoding function $\psi_k^f : \mathcal{G}^{k-1} \times \mathcal{Y}^{(k+T-1)n} \to \mathcal{G}$ for $k \in [1 : S_n]$, i.e.,

$$\psi_k^f(G^{k-1}, \mathbf{Y}^{k+T-1}) = \hat{G}_k$$

Step 1 of the Converse Part

Definition

A feedforward decoder consists of a sequence of decoding function $\psi_k^f : \mathcal{G}^{k-1} \times \mathcal{Y}^{(k+T-1)n} \to \mathcal{G}$ for $k \in [1 : S_n]$, i.e.,

$$\psi_k^f(G^{k-1}, \mathbf{Y}^{k+T-1}) = \hat{G}_k$$

Suffices for a feedforward decoder to consider decoding functions that utilize the channel output sequences only in recent T blocks.

Step 1 of the Converse Part

Definition

A feedforward decoder consists of a sequence of decoding function $\psi_k^f : \mathcal{G}^{k-1} \times \mathcal{Y}^{(k+T-1)n} \to \mathcal{G}$ for $k \in [1 : S_n]$, i.e.,

$$\psi_k^f(G^{k-1}, \mathbf{Y}^{k+T-1}) = \hat{G}_k$$

Suffices for a feedforward decoder to consider decoding functions that utilize the channel output sequences only in recent T blocks.

Lemma

For a feedforward decoder, there exists a sequence of decoding functions $\psi_k^* : \mathcal{G}^{k-1} \times \mathcal{Y}^{Tn} \to \mathcal{G}$ for $k \in [1 : S_n]$, i.e.,

$$\psi_k^*(G^{k-1}, \mathbf{Y}_k^{k+T-1}) = \hat{G}_k \quad \textit{and satisfies}$$

$$\Pr(G_k \neq \psi_k^*(G^{k-1}, \mathbf{Y}_k^{k+T-1})) \leq \Pr(G_k \neq \psi_k^f(G^{k-1}, \mathbf{Y}^{k+T-1}))$$

Step 2 of the Converse Part

Lemma

Let V_n^* be an auxiliary channel defined as

$$V_n^* := \min_{V : C(V) \le R_n - \delta_n} \max_P D(V \| W | P)$$

for appropriately chosen $R_n = C - n^{-t}$ and $\delta_n = o(n^{-t})$.

Step 2 of the Converse Part

Lemma

Let V_n^* *be an auxiliary channel defined as*

$$V_n^* := \min_{V:C(V)\le R_n-\delta_n} \max_P D(V\|W|P)$$

for appropriately chosen $R_n = C - n^{-t}$ *and* $\delta_n = o(n^{-t})$*. Then there exists* $\delta_n' \approx \delta_n$ *s.t. if* ψ_k^* *is applied to* V_n^*

$$\max_{k\in[1:S_n]} \Pr\left(\hat{G}_k \neq G_k\right) \ge \delta_n'.$$

Step 2 of the Converse Part

Lemma

Let V_n^* be an auxiliary channel defined as

$$V_n^* := \min_{V : C(V) \le R_n - \delta_n} \max_P D(V \| W | P)$$

for appropriately chosen $R_n = C - n^{-t}$ and $\delta_n = o(n^{-t})$. Then there exists $\delta_n' \approx \delta_n$ s.t. if ψ_k^* is applied to V_n^*

$$\max_{k \in [1:S_n]} \Pr\left(\hat{G}_k \neq G_k\right) \ge \delta_n'.$$

Thus $\exists$ at least a fraction of $\delta_n'/2$ messages s.t.

$$(V_n^*)^{Tn}\left(\{\textit{bad channel outputs }\} \middle| \textit{ cwd given message}\right) \ge \frac{\delta_n'}{2}$$

Step 3 of the Converse Part

Lemma

If for some $x^{Tn} \in \mathcal{X}^{Tn}$ *with type* $\hat{P}_{x^{Tn}}$*,*

$$(V_n^*)^{Tn}(\mathcal{A}|x^{Tn}) \geq \frac{\delta_n'}{2}, \quad \textit{for some} \quad \mathcal{A} \subset \mathcal{Y}^{Tn},$$

and $0 < t < 1/3$,

Step 3 of the Converse Part

Lemma

If for some $x^{Tn} \in \mathcal{X}^{Tn}$ *with type* $\hat{P}_{x^{Tn}}$*,*

$$(V_n^*)^{Tn}(\mathcal{A}|x^{Tn}) \geq \frac{\delta_n'}{2}, \quad \textit{for some} \quad \mathcal{A} \subset \mathcal{Y}^{Tn},$$

and $0 < t < 1/3$, *then*

$$W^{Tn}(\mathcal{A}|x^{Tn}) \geq \frac{\delta_n'}{4} \exp\left\{-Tn\left(D(V_n^* \| W | \hat{P}_{x^{Tn}}) + \eta_n\right)\right\}$$

where $\eta_n = o(n^{-2t})$.

Step 3 of the Converse Part

Lemma

If for some $x^{Tn} \in \mathcal{X}^{Tn}$ *with type* $\hat{P}_{x^{Tn}}$,

$$(V_n^*)^{Tn}(\mathcal{A}|x^{Tn}) \geq \frac{\delta_n'}{2}, \quad \text{for some} \quad \mathcal{A} \subset \mathcal{Y}^{Tn},$$

and $0 < t < 1/3$, *then*

$$W^{Tn}(\mathcal{A}|x^{Tn}) \geq \frac{\delta_n'}{4} \exp\left\{-Tn\left(D(V_n^*\|W|\hat{P}_{x^{Tn}}) + \eta_n\right)\right\}$$

where $\eta_n = o(n^{-2t})$.

Finally approximate

$$\max_P D(V_n^*\|W|P) = E^+(R_n - \delta_n) = E_{sp}(C - n^{-t} - o(n^{-t})) \leq \frac{n^{-2t}}{2V} + o(n^{-2t}).$$

Outline

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)
- Joint encoding and decoding of fresh and previous messages

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)
- Joint encoding and decoding of fresh and previous messages
- Error probabilities associated with the previous messages add up

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)
- Joint encoding and decoding of fresh and previous messages
- Error probabilities associated with the previous messages add up
- Sequential decoding and truncation of memory if necessary

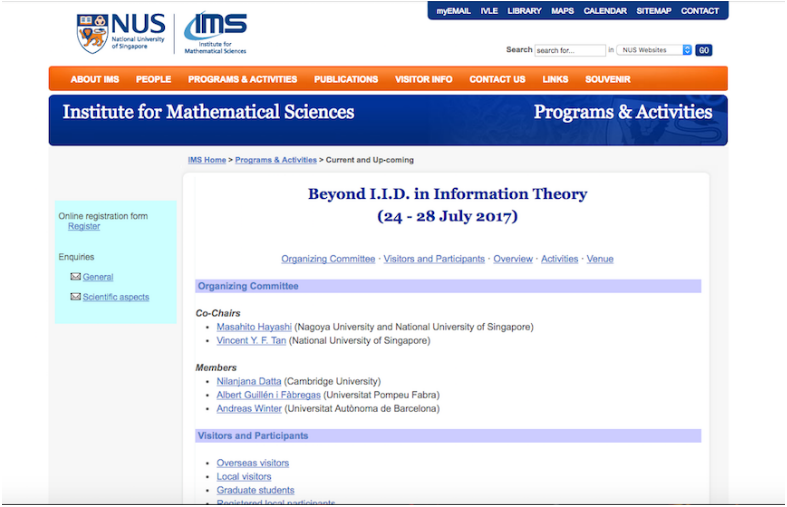
Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)
- Joint encoding and decoding of fresh and previous messages
- Error probabilities associated with the previous messages add up
- Sequential decoding and truncation of memory if necessary
- Also provided a converse in the MD regime under some conditions

Summary

- Information-theoretic streaming model with a delay of T blocks for the MD and CL regimes
- For both regimes, $V \to \frac{V}{T}$ (approximately for the CL regime)
- Joint encoding and decoding of fresh and previous messages
- Error probabilities associated with the previous messages add up
- Sequential decoding and truncation of memory if necessary
- Also provided a converse in the MD regime under some conditions
- See arXiv 1512.06298 for achievability and arXiv 1604.06848 for the converse

Invitation to Beyond IID 2017



Beyond IID 2017 will be at NUS, Singapore