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)
1 Background and Streaming Setup
Outline

1. Background and Streaming Setup

2. Achievability Results and Proof Sketches
1. Background and Streaming Setup

2. Achievability Results and Proof Sketches

3. Achievability Extensions
1. Background and Streaming Setup

2. Achievability Results and Proof Sketches

3. Achievability Extensions

4. Converse Result and the Proof Sketch
Outline

1. Background and Streaming Setup
2. Achievability Results and Proof Sketches
3. Achievability Extensions
4. Converse Result and the Proof Sketch
5. Conclusion and an Announcement
Consider a discrete memoryless channel $W : \mathcal{X} \rightarrow \mathcal{Y}$
Block channel coding

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

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

Fundamental problem in information theory is the interplay between $n$, $R$ and $\varepsilon$

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Rate</td>
<td>$R &lt; C$</td>
<td>$R = C - \rho_n$ $\rho_n \to 0$ and $n\rho_n^2 \to \infty$</td>
<td>$R \approx C - \sqrt{\frac{V}{n} Q^{-1}(\varepsilon)}$</td>
</tr>
<tr>
<td>Error Prob.</td>
<td>Exponential</td>
<td>Subexponential</td>
<td>Non-vanishing</td>
</tr>
<tr>
<td>$\varepsilon \leq \exp{-nE_r(R)}$</td>
<td>$\varepsilon \approx \exp\left{-\frac{n\rho_n^2}{2V}\right}$</td>
<td>$\varepsilon &gt; 0$</td>
<td></td>
</tr>
</tbody>
</table>
## Refined Asymptotics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Rate</td>
<td>$R &lt; C$</td>
<td>$R = C - \rho_n$ for $\rho_n \to 0$ and $n\rho_n^2 \to \infty$</td>
<td>$R = C - \frac{L}{\sqrt{n}}$ for $L &gt; 0$</td>
</tr>
<tr>
<td>Error Prob.</td>
<td>Exponential</td>
<td>Subexponential</td>
<td>Non-vanishing</td>
</tr>
<tr>
<td></td>
<td>$\varepsilon \leq \exp{-nE_r(R)}$</td>
<td>$\varepsilon \approx \exp\left{-\frac{n\rho_n^2}{2V}\right}$</td>
<td>$\varepsilon \approx Q\left(\frac{L}{\sqrt{V}}\right)$</td>
</tr>
</tbody>
</table>

Vincent Tan (NUS)  Streaming Communication  Beyond IID, 2016  5 / 43
### Refined Asymptotics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Rate</td>
<td>$R &lt; C$</td>
<td>$R = C - \rho_n$</td>
<td>$R = C - \frac{L}{\sqrt{n}}$</td>
</tr>
<tr>
<td>Error Prob.</td>
<td>Exponential $\varepsilon \leq \exp{-nE_r(R)}$</td>
<td>Subexponential $\varepsilon \approx \exp{-\frac{n\rho_n^2}{2V}}$</td>
<td>Non-vanishing $\varepsilon \approx Q\left(\frac{L}{\sqrt{V}}\right)$</td>
</tr>
</tbody>
</table>

- **Channel dispersion**

  \[ V = V(W) = \min_{P_X: I(P_X; W) = C(W)} \text{var}(i(X; Y)) \]

- **Smaller dispersion is better.**
An Information-Theoretic Model

Streaming setup

Streaming communication beyond IID, 2016 7/43
An Information-Theoretic Model

An \((n, M, \varepsilon, T)\)-streaming code consists of

1. a sequence of messages \(\{G_k\}_{k \geq 1}\) each uniform over \(\mathcal{G} = [1 : M]\);
2. a sequence of encoders \(\phi_k : \mathcal{G}^k \rightarrow \mathcal{X}^n\) s.t. \(\phi_k(G^k) = X^k_k\);
3. a sequence of decoders \(\psi_k : \mathcal{Y}^{(k+T-1)n} \rightarrow \mathcal{G}\) s.t. \(\psi_k(Y^{k+T-1}) = \hat{G}_k\).

that satisfies

\[
\limsup_{N \to \infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k)}{N} \leq \varepsilon.
\]
Streaming Setup

$G_1 \in [1 : M] \quad G_2 \quad G_3 \quad G_4 \quad G_5$

Encoder

$X_1 \quad X_2 \quad X_3 \quad X_4$

Channel

$W^n(y|x) \quad W^n(y|x) \quad W^n(y|x) \quad W^n(y|x)$

Decoder

$Y_1 \quad Y_2 \quad Y_3 \quad Y_4$

$T = 2$ block delays

$\hat{G}_1 \quad \hat{G}_2 \quad \hat{G}_3$

Streaming setup
Streaming Setup

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

\[ G_1 \in [1 : M] \]

\[ G_2 \]

\[ G_3 \]

\[ G_4 \]

\[ G_5 \]

\[ \hat{G}_1 \]

\[ \hat{G}_2 \]

\[ \hat{G}_3 \]

\[ T = 2 \text{ block delays} \]

Encoder

Channel

Decoder

Time sharing

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

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

Inherent tension in utilizing a block:
- Use codeword only for fresh msg vs. also for previous msgs?
- 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?
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
Theorem (Lee-T.-Khisti (2015))

Let the message size grow as

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

where

$$\rho_n \geq 0, \quad \rho_n \to 0, \quad n\rho_n^2 \to \infty.$$
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 $\left(n, M_n, \varepsilon_n, T\right)$-streaming codes such that

$$\lim_{n \to \infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \leq -\frac{T}{2V}$$
Interpretation of Moderate Deviations Result

\[ \lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \varepsilon_n \leq -\frac{T}{2V} \]

- In block coding,

\[ \lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \varepsilon_n^* = -\frac{1}{2V} \]
Interpretation of Moderate Deviations Result

\[ \lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \varepsilon_n \leq -\frac{T}{2V} \]

- 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**

\[
\lim_{n \to \infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \leq -\frac{T}{2V}
\]

- 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 \(T\)
For any $L > 0$, let the message size grow as

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

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

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

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.

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

- 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

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

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

Error analysis:

$$\varepsilon \leq \Pr(E_1) + \Pr(E_2)$$

where

$E_1 := \{ i(x(G); Y) \leq \log M \}$

$E_2 := \{ \exists \tilde{g} \neq G \text{s.t.} i(x(\tilde{g}); Y) > \log M \}$

Note that $E_1$ is dominant in both regimes.
Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving $P_X$. For each message $g \in [1 : M]$, generate $x(g)$ indep. according to $P^n_X$.
- Encoding: If $g$ is the message, send $x(g)$
Recap of Coding Scheme for Block Coding

- Codebook generation: Fix dispersion-achieving $P_X$. For each message $g \in [1:M]$, generate $x(g)$ indep. according to $P^n_X$.
- Encoding: If $g$ is the message, send $x(g)$
- Decoding: If there exists a unique $g \in [1:M]$ such that
  \[
  i(x(g); y) > \log M, \quad \text{where} \quad i(x; y) = \log \frac{W^n(y|x)}{(P_X W)^n(y)}
  \]
  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 $x(g)$ indep. according to $P^n_X$.

- **Encoding:** If $g$ is the message, send $x(g)$

- **Decoding:** If there exists a unique $g \in [1 : M]$ such that

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

  let $\hat{G} = g$.

- **Error analysis:**

  \[
  \varepsilon \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2)
  \]

  where

  \[
  \mathcal{E}_1 := \{i(X(G); Y) \leq \log M\}
  \]

  \[
  \mathcal{E}_2 := \{\exists \tilde{g} \neq G \text{ s.t. } i(X(\tilde{g}); Y) > \log M\}
  \]

  Note that $\mathcal{E}_1$ is dominant in both regimes.
Analysis of Error Probability

- Probability of error

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

where

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

are i.i.d. random variables.
Probability of error

\[ \varepsilon_n \approx \Pr \left( \text{i}(X(G); Y) \leq M \right) = \Pr \left( \sum_{l=1}^{n} Z_l \leq \log M \right) \]

where

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

are i.i.d. random variables.

Note that for all \( l \in [1 : n] \),

\[ \mathbb{E}[Z_l] = C \quad \text{and} \quad \text{var}[Z_l] = V \]
Analysis of Error Probability

- **Probability of error**

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

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** \( l \in [1 : n], \)

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

- **Under various regimes, analyze**

\[ \Pr \left( \sum_{l=1}^{n} Z_l \leq \log M \right). \]
Analysis of Error: Moderate Deviations Regime

**Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))**

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

\[
\lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \Pr \left( \sum_{l=1}^{n} Z_l \leq \log M \right) = -\frac{1}{2V}.
\]
Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))

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

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

Thus, we have

\[
\lim_{n \to \infty} \frac{1}{n\rho_n^2} \log \varepsilon_n \leq -\frac{1}{2V}.
\]
Theorem (Moderate Deviations Theorem (Dembo and Zeitouni))

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

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

Thus, we have

\[
\lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \varepsilon_n \leq -\frac{1}{2V}.
\]

However, note that the standard MD theorem is asymptotic in nature.

We need a non-asymptotic version in the streaming scenario.
Theorem (Berry-Esseen Theorem)

Under regularity conditions on $Z_l$, and

$$\log M = nC - \sqrt{nL},$$

we have

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

where $\tau$ is a constant (depending on $Z_1$).
Theorem (Berry-Esseen Theorem)

Under regularity conditions on $Z_l$, and

$$\log M = nC - \sqrt{nL},$$

we have

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

where $\tau$ is a constant (depending on $Z_1$).

Thus, we have

$$\varepsilon_n \leq Q \left( \frac{L}{\sqrt{V}} \right) + O \left( \frac{1}{\sqrt{n}} \right).$$
Analysis of Error: Central Limit Regime

Theorem (Berry-Esseen Theorem)

Under regularity conditions on $Z_l$, and

$$\log M = nC - \sqrt{nL},$$

we have

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

where $\tau$ is a constant (depending on $Z_1$).

- Thus, we have

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

- However, note that the Berry-Esseen residual terms hurt us in the streaming setup
Consider block delay $T = 2$. 

![Diagram of streaming setup with moderate deviations regime](image-url)
Consider block delay $T = 2$.

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

- Codebook generation for block $k$: For each $g^k \in [1 : M]^k$, generate $x_k(g^k)$ in an i.i.d. manner according to $P_X$ that achieves the dispersion.

- Encoding at block $k$: Send $x_k(G_1, \cdots, G_k)$. 
Consider block delay $T = 2$.

Codebook generation for block $k$: For each $g^k \in [1 : M]^k$, generate $x_k(g^k)$ in an i.i.d. manner according to $P_X$ that achieves the dispersion.

Encoding at block $k$: Send $x_k(G_1, \cdots, G_k)$.

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

\begin{align*}
G_1 \in [1:M] & \quad G_2 & \quad G_3 & \quad G_4 & \quad G_5 \\
\text{Encoder} & \quad X_1 & \quad X_2 & \quad X_3 & \quad X_4 \\
\text{Channel} & \quad W^n(y|x) & \quad W^n(y|x) & \quad W^n(y|x) & \quad W^n(y|x) \\
\text{Decoder} & \quad Y_1 & \quad Y_2 & \quad Y_3 & \quad Y_4 \\
& \quad T = 2\text{ block delays} & \quad \hat{G}_1 & \quad \hat{G}_2 & \quad \hat{G}_3
\end{align*}
\[ G_1 \in [1 : M] \]

- **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

\[
i([x_1(g_1) x_2(g_1, g_2) x_3(g_1, g_2, g_3)], [y_1 y_2 y_3]) > 3 \log M, \quad \text{for some} \quad g_2, g_3
\]
\[ T = 2: \text{At the end block 3, sequentially decode } G_1 \text{ and } G_2. \]

- **Re-decode } G_1: \text{Choose } \hat{G}_1' \text{ as a unique } g_1 \in [1 : M] \text{ such that}

\[
i([x_1(g_1) \quad x_2(g_1, g_2) \quad x_3(g_1, g_2, g_3)], [y_1 \quad y_2 \quad y_3]) > 3 \log M, \quad \text{for some } g_2, g_3
\]

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

\[
i([x_2(\hat{G}_1', g_2) \quad x_3(\hat{G}_1', g_2, g_3)], [y_2 \quad y_3]) > 2 \log M \quad \text{for some } g_3
\[ T = 2: \text{At the end block 3, sequentially decode } G_1 \text{ and } G_2. \]

- **Re-decode** \( G_1 \): Choose \( \hat{G}_1' \) as a unique \( g_1 \in [1 : M] \) such that
  \[ i([x_1(g_1) \ x_2(g_1, g_2) \ x_3(g_1, g_2, g_3)] , [y_1 \ y_2 \ y_3]) > 3 \log M, \quad \text{for some } g_2, g_3 \]

- **Decode** \( G_2 \): Choose \( \hat{G}_2 \) as a unique \( g_2 \in [1 : M] \) such that
  \[ i([x_2(\hat{G}_1', g_2) \ x_3(\hat{G}_1', g_2, g_3)] , [y_2 \ y_3]) > 2 \log M \quad \text{for some } g_3 \]

\[ \Pr(\hat{G}_2 \neq G_2) \leq \Pr((\hat{G}_1' \neq G_1) \cup (\hat{G}_2 \neq G_2)) \]
\[ \approx \Pr(\sum_{l=1}^{3n} Z_l \leq 3 \log M) + \Pr(\sum_{l=1}^{2n} Z_l \leq 2 \log M) \]
**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
  
  $$i([x_1(g_1), x_2(g_1, g_2), x_3(g_1, g_2, g_3)], [y_1, y_2, y_3]) > 3 \log M, \text{ for some } g_2, g_3$$

- **Decode $G_2$:** Choose $\hat{G}_2$ as a unique $g_2 \in [1 : M]$ such that
  
  $$i([x_2(\hat{G}_1', g_2), x_3(\hat{G}_1', g_2, g_3)], [y_2, y_3]) > 2 \log M, \text{ for some } g_3$$

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

- **For all $k \in \mathbb{N}$,**
  
  $$\Pr(\hat{G}_k \neq G_k) \leq \sum_{j=2}^{\infty} \Pr\left(\sum_{l=1}^{jn} Z_l \leq j \log M\right)$$
$T = 2$: 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
  \[
i([x_1(g_1) \ x_2(g_1, g_2) \ x_3(g_1, g_2, g_3)], [y_1 \ y_2 \ y_3]) > 3 \log M, \quad \text{for some } g_2, g_3\]

- Decode $G_2$: Choose $\hat{G}_2$ as a unique $g_2 \in [1 : M]$ such that
  \[
i([x_2(\hat{G}_1', g_2) \ x_3(\hat{G}_1', g_2, g_3)], [y_2 \ y_3]) > 2 \log M \quad \text{for some } g_3\]

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

- For all $k \in \mathbb{N}$,
  \[
  \Pr(\hat{G}_k \neq G_k) \leq \sum_{j=T}^{\infty} \Pr \left( \sum_{l=1}^{j} Z_l \leq j \log M \right)
  \]
For all $k \in \mathbb{N}$,

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

$$\lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \Pr \left( \sum_{l=1}^{jn} Z_l \leq j \log M \right) \leq -\frac{j}{2V}$$
For all $k \in \mathbb{N}$,

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

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

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

Cannot “exchange limits”
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.,

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

Cannot “exchange limits”

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

Under regularity conditions on $Z_l$, for any positive $\rho_n$ 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 $Z_1$. 
Lemma

Under regularity conditions on $Z_l$, for any positive $\rho_n$ 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 $k \in \mathbb{N}$,

$$\lim_{n \to \infty} \frac{1}{n \rho_n^2} \log \left[ \lim_{n \to \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_k \neq G_k) \right] \leq -\frac{T}{2V}.$$
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, $n \to jn$, $L \to \sqrt{jL}$. 
For all $k \in \mathbb{N}$,

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

$$= \sum_{j=T}^{\infty} \Pr \left( \sum_{l=1}^{jn} Z_l \leq j(nR - L \sqrt{n}) \right)$$

$$\leq \sum_{j=T}^{\infty} \left( Q \left( \frac{\sqrt{jL}}{\sqrt{V}} \right) + \frac{\tau}{\sqrt{jn}} \right)$$

Compared to the block coding case, $n \to jn$, $L \to \sqrt{jL}$. 
For all $k \in \mathbb{N}$,

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

$$= \sum_{j=T}^{\infty} \Pr \left( \sum_{l=1}^{jn} Z_l \leq j(nR - L\sqrt{n}) \right)$$

$$\leq \sum_{j=T}^{\infty} \left( Q \left( \frac{\sqrt{jL}}{\sqrt{V}} \right) + \frac{\tau}{\sqrt{jn}} \right)$$

$$= \sum_{j=T}^{\infty} Q \left( \frac{\sqrt{jL}}{\sqrt{V}} \right) + \sum_{j=T}^{\infty} \frac{\tau}{\sqrt{jn}}$$

$$\leq c \cdot Q \left( \frac{\sqrt{TL}}{\sqrt{V}} \right) + \sum_{j=T}^{\infty} \frac{\tau}{\sqrt{jn}},$$

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

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

The remainder terms from the Berry-Esseen theorem diverge!
[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 ($B \uparrow$)
- \( A, B \): 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 \( \left( \frac{B}{A} \downarrow \right) \)
  - Contributions to error probability
    - Remainder terms \( (A \downarrow) \)
    - Previous group \( (B \uparrow) \)

Memory structure with \( A = 9, B = 4 \)
1. Background and Streaming Setup

2. Achievability Results and Proof Sketches

3. Achievability Extensions

4. Converse Result and the Proof Sketch

5. Conclusion and an Announcement
An \((n, M, \varepsilon, \varepsilon', T)\)-streaming code with an erasure option is the same as the usual streaming code except that the total error probability does not exceed \(\varepsilon\), i.e.,

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

The erasure error probability does not exceed \(\varepsilon'\), i.e.,

\[
\limsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_k \neq G_k, \hat{G}_k \neq 0) \leq \varepsilon'.
\]

Seek upper bounds on \(\varepsilon\) and \(\varepsilon'\) when \(M\) is the moderate deviations regime.
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 the decoding functions

\[
\psi_k : \mathcal{Y}^{(k+T-1)n} \rightarrow \mathcal{G} \cup \{0\}
\]

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 \(\varepsilon\), i.e.,
     \[
     \limsup_{N \to \infty} \sum_{k=1}^{N} \frac{\Pr(\hat{G}_k \neq G_k)}{N} \leq \varepsilon.
     \]
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

\[
\psi_k : \mathcal{Y}^{(k+T-1)n} \rightarrow \mathcal{G} \cup \{0\}
\]

where 0 denotes the erasure option.

2. The total error probability does not exceed \(\varepsilon\), 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 \(\varepsilon'\), 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 \(\varepsilon\) and \(\varepsilon'\) when \(M\) is the moderate deviations regime.
Illustration of the Erasure Option

Decoding with an erasure option
Theorem

Let the message size grow as

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

where

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

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

$$\lim_{n \to \infty} \frac{1}{n\rho_n} \log \varepsilon'_n \leq -T\gamma$$

for any \(0 < \gamma < 1\).
Discussion of Result for Erasure Option

- The undetected error probability is

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

and the total error probability is

\[ \varepsilon'_{n} \leq \exp \left\{ -n \rho_n \cdot T \gamma + o(n \rho_n) \right\} \]

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 \( T \).
The undetected error probability is

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

and the total error probability is

\[ \varepsilon'_n \leq \exp \left\{ -n\rho_n \cdot T\gamma + o(n\rho_n) \right\} \]

Total error probability is much larger than undetected error probability
The undetected error probability is

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

and the total error probability is

$$\varepsilon'_n \leq \exp \left\{ -n \rho_n \cdot T\gamma + o(n \rho_n) \right\}$$

Total error probability is much larger than undetected error probability.

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

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

and the total error probability is

$$\varepsilon'_n \leq \exp \left\{ -n\rho_n \cdot T\gamma + o(n\rho_n) \right\}$$

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 $T$
An \((n, M, \varepsilon, T)\)-streaming code with an average delay constraint is the same as the usual streaming code except that the sequence of decoding functions \(\psi_k: Y^n \rightarrow (G \cup \{0\})^k\) and the average error probability is upper bounded as

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_k + D_k + 1 \neq G_k) \leq \varepsilon
\]

where \(D_k := \min\{d \in \mathbb{N}: \hat{G}_k + d - 1 \neq 0\}\) denotes the random decoding delay of the \(k\)-th message and the average delay satisfies

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[D_k] \leq T.
\]
An \((n, M, \varepsilon, T)\)-streaming code with an average delay constraint is the same as the usual streaming code except that the sequence of decoding functions

\[
\psi_k : \mathcal{Y}^{kn} \rightarrow (\mathcal{G} \cup \{0\})^k
\]
Extension 2: Decoding with Variable Delay

- An \((n, M, \varepsilon, T)\)-streaming code with an average delay constraint is the same as the usual streaming code except that
  1. the sequence of decoding functions
     \[
     \psi_k : Y^{kn} \rightarrow (G \cup \{0\})^k
     \]
  2. the average error probability is upper bounded as
     \[
     \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_{k+D_k+1} \neq G_k) \leq \varepsilon
     \]

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

1. the sequence of decoding functions

\[ \psi_k : \mathcal{Y}^{kn} \rightarrow (G \cup \{0\})^k \]

2. the average error probability is upper bounded as

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{G}_{k+D_{k+1}} \neq G_k) \leq \varepsilon
\]

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

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

Let the message size grow as

\[ \log M_n = n(C - \rho_n) \]

where

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

\[ \lim_{n \to \infty} T_n = T \]

\[ \lim_{n \to \infty} \frac{1}{n\rho_n} \log \varepsilon_n \leq -T \]
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,

\[ \varepsilon_n \leq \exp(-\Theta(n\rho_n^2)) \]

A significant gain can be achieved in the moderate deviations regime with streaming and variable delay without feedback.
For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved without variable delay,

\[ \varepsilon_n \leq \exp(-\Theta(n\rho_n^2)) \]

With variable delay,

\[ \varepsilon_n \leq \exp(-n\rho_n T + o(n\rho_n)) \]
For block coding with one-bit feedback (ARQ), Forney (1968) showed that the reliability function can be significantly improved.

Without variable delay,

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

With variable delay

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

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

1. Background and Streaming Setup
2. Achievability Results and Proof Sketches
3. Achievability Extensions
4. Converse Result and the Proof Sketch
5. Conclusion and an Announcement
To derive lower bounds to error probability, we consider a slightly different setup.

A (n, M, \varepsilon, T, S)-streaming code consists of:

1. a sequence of messages \{G_k\}_{S_k=1}:
   - Each uniformly distributed over \(G = [1 : M]\).
2. a sequence of encoding functions \(\varphi_k: G_{\min\{k, S\}} \to X_n\) for \(k \in [1 : S + T - 1]\).
3. a sequence of decoding functions \(\psi_k: Y(k + T - 1)n \to G\), for \(k \in [1 : S]\)

Such that the maximum error probability over all \(S\) messages satisfies:

\[
\max_{k \in [1 : S]} \Pr(\hat{G}_k \neq G_k) \leq \varepsilon.
\]
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 \(G = [1 : M]\)

2. A sequence of encoding functions

\[\phi_k : G^{\min\{k, S\}} \rightarrow \mathcal{X}^n \quad \text{for} \quad k \in [1 : S + T - 1]\]

3. A sequence of decoding functions

\[\psi_k : \mathcal{Y}^{(k+T-1)n} \rightarrow G, \quad \text{for} \quad k \in [1 : S]\]

Such that 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 messages (\( S = 5 \)) 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 msgs ($S = 5$) are sequentially encoded and are sequentially decoded after the delay of two blocks ($T = 2$).

Fundamental limit on error probability

$$\varepsilon^* (n, M, T, S) := \min \{ \varepsilon : \exists \text{ an } (n, M, \varepsilon, T, S)\text{-streaming code} \}$$
Theorem (Lee-T.-Khisti (2016))

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

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

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

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

and

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

Then

\[
\lim_{n \to \infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) = -\frac{T}{2V}
\]
For an output symmetric DMC with $V > 0$, consider sequences $M_n$ and $S_n$ such that

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

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

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

\[
E^+(R; W) := \min_{V: \mathcal{C}(V) \leq R} \max_P D(V || W | P) \quad \text{(Haroutunian)}
\]

\[
E_{sp}(R; W) := \max_P \min_{V: \mathcal{I}(P, V) \leq R} D(V || W | P) \quad \text{(Sphere Packing)}
\]
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

\[ E^+(R; W) := \min_{V: C(V) \leq R} \max_P D(V||W|P) \quad \text{(Haroutunian)} \]

\[ E_{sp}(R; W) := \max_P \min_{V: I(P,V) \leq R} D(V||W|P) \quad \text{(Sphere Packing)} \]

- 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

\[
E^+(R; W) := \min_{V: C(V) \leq R} \max_P D(V||W|P) \quad \text{(Haroutunian)}
\]

\[
E_{sp}(R; W) := \max_P \min_{V: I(P,V) \leq R} D(V||W|P) \quad \text{(Sphere Packing)}
\]

- 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

\[
\lim_{n \to \infty} \frac{1}{n^{1-2t}} \log \varepsilon^*(n, M_n, T, S_n) \geq -\frac{T}{2V}
\]
Proof Ideas for the Converse Part

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

- Step 1: Assume a feedforward decoder (genie-aided decoder)
Proof Ideas for the Converse Part

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

- **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)
A feedforward decoder consists of a sequence of decoding function
\[ \psi_f^k : \mathcal{G}_{k-1} \times \mathcal{Y}^{(k+T-1)n} \rightarrow \mathcal{G} \text{ for } k \in [1 : S_n], \text{ i.e.,} \]
\[ \psi_f^k (G_{k-1}^{k-1}, Y_{k+T-1}^{k+T-1}) = \hat{G}_k \]
A feedforward decoder consists of a sequence of decoding function
\[ \psi^f_k : G^{k-1} \times Y^{(k+T-1)n} \rightarrow G \text{ for } k \in [1 : S_n], \] i.e.,
\[ \psi^f_k(G^{k-1}, Y^{k+T-1}) = \hat{G}_k \]

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

A **feedforward decoder** consists of a sequence of decoding functions

\[ \psi_f^k : \mathcal{G}^{k-1} \times \mathcal{Y}^{(k+T-1)n} \rightarrow \mathcal{G} \text{ for } k \in [1 : S_n], \text{ i.e.,} \]

\[ \psi_f^k(G^{k-1}, Y^{k+T-1}) = \hat{G}_k \]

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** \( \psi^*_k : \mathcal{G}^{k-1} \times \mathcal{Y}^{Tn} \rightarrow \mathcal{G} \text{ for } k \in [1 : S_n], \text{ i.e.,} \)

\[ \psi^*_k(G^{k-1}, Y^{k+T-1}) = \hat{G}_k \quad \text{and satisfies} \]

\[ \Pr(G_k \neq \psi^*_k(G^{k-1}, Y^{k+T-1})) \leq \Pr(G_k \neq \psi_f^k(G^{k-1}, Y^{k+T-1})) \]
Lemma

Let $V^*_n$ be an auxiliary channel defined as

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

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

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

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

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 $\psi_k^*$ is applied to $V_n^*$

$$\max_{k \in [1:S_n]} \Pr\left( \hat{G}_k \neq G_k \right) \geq \delta'_n.$$
Step 2 of the Converse Part

Lemma

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

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

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 $\psi_k^*$ is applied to $V_n^*$

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

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

$$\left( V_n^* \right)^T \left( \{ \text{bad channel outputs} \} \middle| \text{cwd given message} \right) \geq \frac{\delta'_n}{2}.$$
Lemma

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

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

and $0 < t < 1/3$,
Lemma

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

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

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

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

where $\eta_n = o(n^{-2t})$. 
Step 3 of the Converse Part

**Lemma**

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

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

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

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

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

1 Background and Streaming Setup
2 Achievability Results and Proof Sketches
3 Achievability Extensions
4 Converse Result and the Proof Sketch
5 Conclusion and an Announcement
Information-theoretic streaming model with a delay of $T$ blocks for the MD and CL regimes

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
Summary

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

For both regimes, $V \rightarrow \frac{V}{T}$ (approximately for the CL regime)

Joint encoding and decoding of fresh and previous messages

Error probabilities associated with the previous messages add up
Information-theoretic streaming model with a delay of $T$ blocks for the MD and CL regimes

For both regimes, $V \rightarrow \frac{V}{T}$ (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, $V \rightarrow \frac{V}{T}$ (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.
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
Beyond I.I.D. in Information Theory
(24 - 28 July 2017)

Organizing Committee
Visitors and Participants
Overview
Activities
Venue

Co-Chairs
- Masahito Hayashi (Nagoya University and National University of Singapore)
- Vincent Y. F. Tan (National University of Singapore)

Members
- Nilanjana Datta (Cambridge University)
- Albert Guillén i Fàbregas (Universitat Pompeu Fabra)
- Andreas Winter (Universitat Autònoma de Barcelona)

Visitors and Participants
- Overseas visitors
- Local visitors
- Graduate students
- Registered local participants

Beyond IID 2017 will be at NUS, Singapore