Time-Division Transmission is Optimal for Covert Communication over Some Broadcast Channels

Vincent Y. F. Tan (NUS)
Joint work with Prof. Si-Hyeon Lee (POSTECH)

National University of Singapore (NUS)
ITW, Guangzhou (Nov 2018)
Problem Setting

Consider a traditional two-user broadcast channel.
Consider a traditional two-user broadcast channel

\[(W_1, W_2) \rightarrow (X^n, P_{Y_1Y_2Z|X}^n) \rightarrow \hat{W}_1, \hat{W}_2\]
Problem Setting

- Consider a traditional two-user broadcast channel

\[
(W_1, W_2) \xrightarrow{f} X^n \xrightarrow{P_{Y_1Y_2Z|X}^{\times n}} Y_1^n, Y_2^n, Z^n \xrightarrow{\phi_1, \phi_2} \hat{W}_1, \hat{W}_2
\]

Vincent Tan (NUS)
Consider a traditional two-user broadcast channel
Consider a traditional two-user broadcast channel.
Problem Setting

- Consider a traditional two-user broadcast channel

\[(W_1, W_2) \xrightarrow{f} X^n \quad P_{Y_1Y_2Z|X}^{X^n} \rightarrow \phi_1 \rightarrow Y_1^n \rightarrow \hat{W}_1 \]

\[\phi_2 \rightarrow Y_2^n \rightarrow \hat{W}_2 \]

\[Z^n \rightarrow \text{Warden} \rightarrow H_0 : Q_0^{X^n} \quad H_1 : \hat{Q_z^n} \]

- Covert Communication $\iff$ Warden should not be able to know whether there is communication
Problem Setting

Consider a traditional two-user broadcast channel

(W₁, W₂) → f → Xⁿ → Pₓⁿᵧ₁ᵧ₂zₓ → Y₁ⁿ → ϕ₁ → W₁

Y₂ⁿ → ϕ₂ → W₂

Zⁿ → Warden → H₀ : Q₀ₓⁿ

H₁ : ̂Qₓⁿ

Covert Communication ⇔ Warden should not be able to know whether there is communication

Should not be able to distinguish between ̂Qₓⁿ (output dist. induced by a code) and the “no-communication” output dist. Q₀ₓⁿ.
Innocent symb. \(0 \in \mathcal{X}\) inducing warden output dist. \(Q_0 = P_{Z|X}(\cdot|0)\).
Innocent symb. $0 \in \mathcal{X}$ inducing warden output dist. $Q_0 = P_{Z|X}(\cdot | 0)$

Symb. $1 \in \mathcal{X}$ inducing warden output dist. $Q_1 = P_{Z|X}(\cdot | 0)$

Assume $Q_1 \ll Q_0$

Warden attempts to design optimal detector to distinguish $H_0$: observed distribution is $Q \times n_0$ (no communication)

$H_1$: observed distribution is $\hat{Q}_Z n$ (communication active)

Optimal performance

$$\pi_1 | 0 + \pi_0 | 1 = 1 - \frac{1}{2} \| Q \times n_0 - \hat{Q}_Z n \|_1 \geq 1 - \sqrt{D(\hat{Q}_Z n \| Q \times n_0)}$$

For covert communication, we want to make $D(\hat{Q}_Z n \| Q \times n_0)$ small.
More on Covert Communications over Noisy Channels

- Innocent symb. $0 \in \mathcal{X}$ inducing warden output dist. $Q_0 = P_{Z|X}(\cdot | 0)$
- Symb. $1 \in \mathcal{X}$ inducing warden output dist. $Q_1 = P_{Z|X}(\cdot | 0)$
- Assume $Q_1 \ll Q_0$

Vincent Tan (NUS)
More on Covert Communications over Noisy Channels

- Innocent symb. $0 \in \mathcal{X}$ inducing warden output dist. $Q_0 = P_{Z|X}(\cdot|0)$
- Symb. $1 \in \mathcal{X}$ inducing warden output dist. $Q_1 = P_{Z|X}(\cdot|0)$
- Assume $Q_1 \ll Q_0$

Warden attempts to design optimal detector to distinguish

- $H_0$: observed distribution is $Q_0^\times n$ (no communication)
- $H_1$: observed distribution is $\hat{Q}_Z^n$ (communication active)
More on Covert Communications over Noisy Channels

- Innocent symb. \( 0 \in \mathcal{X} \) inducing warden output dist. \( Q_0 = P_{Z|X}(\cdot|0) \)

- Symb. \( 1 \in \mathcal{X} \) inducing warden output dist. \( Q_1 = P_{Z|X}(\cdot|0) \)

- Assume \( Q_1 \ll Q_0 \)

- Warden attempts to design optimal detector to distinguish
  - \( H_0 \): observed distribution is \( Q_0^x \) (no communication)
  - \( H_1 \): observed distribution is \( \hat{Q}_Z^x \) (communication active)

- Optimal performance

\[
\pi_1|0 + \pi_0|1 = 1 - \frac{1}{2} \left\| Q_0^x - \hat{Q}_Z^x \right\|_1 \geq 1 - \sqrt{D(\hat{Q}_Z^x \parallel Q_0^x)}
\]
Innocent symb. $0 \in \mathcal{X}$ inducing warden output dist. $Q_0 = P_{Z|X}(\cdot|0)$

Symb. $1 \in \mathcal{X}$ inducing warden output dist. $Q_1 = P_{Z|X}(\cdot|0)$

Assume $Q_1 \ll Q_0$

Warden attempts to design optimal detector to distinguish

- $H_0$: observed distribution is $Q_0^\times n$ (no communication)
- $H_1$: observed distribution is $\hat{Q}_Z^n$ (communication active)

Optimal performance

$$\pi_{1|0} + \pi_{0|1} = 1 - \frac{1}{2} \left\| Q_0^\times n - \hat{Q}_Z^n \right\|_1 \geq 1 - \sqrt{D(\hat{Q}_Z^n \| Q_0^\times n)}$$

For covert communication, we want to make $D(\hat{Q}_Z^n \| Q_0^\times n)$ small.
Growing concern for privacy and confidentiality
Growing concern for privacy and confidentiality

Renewed interest in fundamental limits of covert communications:

- Secure space-time codes [Hero ’03]
- Secure stegosystems [Korzhik et al. ’05]
- $O(\sqrt{n})$ bits over $n$ ch. uses with $O(\sqrt{n \log n})$ key [Bash et al. ’12]
- Similar to square-root law in steganography [Cachin ’04]

- Several extensions and related results:
  - Constants in $O(\sqrt{n})$ term [Wang, Wornell, Zheng ’16 and Bloch ’16]
  - Second-order [Tahmasbi-Bloch ’16]
  - Error exponents [Tahmasbi-Bloch-Tan ’17]
  - Multi-user [Arumugam-Bloch ’16, ’17]
Growing concern for privacy and confidentiality

Renewed interest in fundamental limits of covert communications:

- Secure space-time codes [Hero '03]
- Secure stegosystems [Korzhik et al. '05]
- $O(\sqrt{n})$ bits over $n$ ch. uses with $O(\sqrt{n \log n})$ key [Bash et al. '12]
- Similar to square-root law in steganography [Cachin '04]

Several extensions and related results:

- Constants in $O(\sqrt{n})$ term [Wang, Wornell, Zheng '16 and Bloch '16]
- Second-order [Tahmasbi-Bloch '16]
- Error exponents [Tahmasbi-Bloch-Tan '17]
- Multi-user [Arumugam-Bloch '16, '17]
Definition of a code

An \((n, M_1^n, M_2^n, \varepsilon, \delta)\)-code for the broadcast channel with a warden \(P_{Y_1, Y_2, Z|X}\) consists of

- Two message sets \(\mathcal{M}_j := \{1, \ldots, M_{jn}\}\) for \(j = 1, 2\);
- Two independent messages uniformly distributed over their respective message sets, i.e., \(W_j \sim \text{Unif}(\mathcal{M}_j)\) for \(j = 1, 2\);
Definition of a code

An \((n, M_{1n}, M_{2n}, \varepsilon, \delta)\)-code for the broadcast channel with a warden 
\(P_{Y_1, Y_2, Z|X}\) consists of

- Two message sets \(\mathcal{M}_j := \{1, \ldots, M_{jn}\}\) for \(j = 1, 2\);
- Two independent messages uniformly distributed over their respective message sets, i.e., 
  \(W_j \sim \text{Unif}(\mathcal{M}_j)\) for \(j = 1, 2\);
- One encoder \(f : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n\);
- Two decoders \(\varphi_j : \mathcal{Y}_j^n \rightarrow \mathcal{M}_j\) for \(j = 1, 2\);
Definition of a code

An \((n, M_1^n, M_2^n, \varepsilon, \delta)\)-code for the broadcast channel with a warden
\(P_{Y_1,Y_2,Z|X}\) consists of

- Two message sets \(\mathcal{M}_j := \{1, \ldots, M_j^n\}\) for \(j = 1, 2\);
- Two independent messages uniformly distributed over their respective message sets, i.e., \(W_j \sim \text{Unif}(\mathcal{M}_j)\) for \(j = 1, 2\);
- One encoder \(f : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow X^n\);
- Two decoders \(\varphi_j : Y^n_j \rightarrow \mathcal{M}_j\) for \(j = 1, 2\);

such that the following constraints hold:

 Reliability: \(
\Pr \left( \bigcup_{j=1}^{2} \{\hat{W}_j \neq W_j\} \right) \leq \varepsilon
\)

and

 Covertness: \(
D(\hat{Q}_{Z^n} \| Q_0^{x^n}) \leq \delta.
\)
Definition of Covert Capacity Region

- \((L_1, L_2) \in \mathbb{R}^2_+\) is \((\varepsilon, \delta)\)-achievable if there exists a sequence of \((n, M_{1n}, M_{2n}, \varepsilon_n, \delta)\)-codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n\delta}} \log M_{jn} \geq L_j, \quad j = 1, 2,
\]

\[
\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon.
\]
Definition of Covert Capacity Region

- \((L_1, L_2) \in \mathbb{R}_+^2\) is \((\varepsilon, \delta)\)-achievable if there exists a sequence of \((n, M_{1n}, M_{2n}, \varepsilon_n, \delta)\)-codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n\delta}} \log M_{jn} \geq L_j, \quad j = 1, 2,
\]

\[
\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon.
\]

- The \((\varepsilon, \delta)\)-covert capacity region \(\mathcal{L}_{\varepsilon, \delta} \subset \mathbb{R}_+^2\) is the closure of all \((\varepsilon, \delta)\)-achievable pairs of \((L_1, L_2)\).
Definition of Covert Capacity Region

- \((L_1, L_2) \in \mathbb{R}^2_+\) is \((\varepsilon, \delta)\)-achievable if there exists a sequence of \((n, M_{1n}, M_{2n}, \varepsilon_n, \delta)\)-codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n\delta}} \log M_{jn} \geq L_j, \quad j = 1, 2,
\]

\[
\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon.
\]

- The \((\varepsilon, \delta)\)-covert capacity region \(L_{\varepsilon, \delta} \subset \mathbb{R}^2_+\) is the closure of all \((\varepsilon, \delta)\)-achievable pairs of \((L_1, L_2)\).

- The \(\delta\)-covert capacity region

\[
L_\delta := \bigcap_{\varepsilon \in (0, 1)} L_{\varepsilon, \delta} = \lim_{\varepsilon \to 0} L_{\varepsilon, \delta}.
\]
Definition of Covert Capacity

For simplicity, assume binary-input channels, i.e., $\mathcal{X} = \{0, 1\}$.
Definition of Covert Capacity

- For simplicity, assume binary-input channels, i.e., $\mathcal{X} = \{0, 1\}$

- Given a DMC with a warden $P_{Y,Z|X}$, the *covert capacity* [Wang, Wornell, Zheng ’16 and Bloch ’16] is

$$L^*(P_{Y,Z|X}) := \sqrt{\frac{2D(W(\cdot|1)\|W(\cdot|0))^2}{\chi_2(Q_1\|Q_0)}}$$

where

$$W(\cdot|x) = P_{Y|X}(\cdot|x) \quad Q_x := P_{Z|X}(\cdot|x), \quad \forall x \in \mathcal{X},$$

and

$$\chi_2(Q_1\|Q_0) := \sum_z \frac{(Q_1(z) - Q_0(z))^2}{Q_0(z)}.$$
Assumption on the BC $P_{Y_1,Y_2,Z|X}$

**Assumption**

**Condition 1:** Fix a BC with a warden $P_{Y_1,Y_2,Z|X}$.

Let the covert capacities of $P_{Y_1,Z|X}$ and $P_{Y_2,Z|X}$ be $L_1^*$ and $L_2^*$ respectively.

If $L_1^* \geq L_2^*$ assume that

$$\max_{P_X} \frac{I(X; Y_1)}{I(X; Y_2)} \leq \frac{L_1^*}{L_2^*}.$$

Otherwise, if $L_1^* \leq L_2^*$ assume that

$$\max_{P_X} \frac{I(X; Y_2)}{I(X; Y_1)} \leq \frac{L_2^*}{L_1^*}.$$
Discussion of Condition 1

- Easy to check for binary-input BCs:

\[
\frac{D(W_1 \| W_0)}{D(V_1 \| V_0)} \leq \min_{\gamma \in [0,1]} \frac{D(W_\gamma \| W_0)}{D(V_\gamma \| V_0)}, \quad W_\gamma(y) = \sum_{x \in \mathcal{X}} P_\gamma(x) W(y|x).
\]
Discussion of Condition 1

- Easy to check for binary-input BCs:

\[
\frac{D(W_1 \| W_0)}{D(V_1 \| V_0)} \leq \min_{\gamma \in [0,1]} \frac{D(W_\gamma \| W_0)}{D(V_\gamma \| V_0)}, \quad W_\gamma(y) = \sum_{x \in X} P_\gamma(x) W(y|x).
\]

- Let \( W = P_{Y_1|X} = \text{BSC}(p) \) and \( V = P_{Y_2|X} = \begin{bmatrix} 1 - q_0 & q_0 \\ q_1 & 1 - q_1 \end{bmatrix} \).

\[
\text{Shaded area} \Rightarrow \text{Condition 1} \; \checkmark
\]
Main Result

Theorem (Tan-Lee (2018))

Assume Condition 1 holds for $P_{Y_1,Y_2,Z|X}$. For any $\delta > 0$ and $P_{Y_j|X}(\cdot | 1) \ll P_{Y_j|X}(\cdot | 0)$ for $j = 1, 2$,

$$\mathcal{L}_\delta = \left\{ (L_1, L_2) \in \mathbb{R}^2_+ : \frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \leq 1 \right\}.$$
Main Result

Theorem (Tan-Lee (2018))

Assume Condition 1 holds for $P_{Y_1, Y_2, Z|X}$. For any $\delta > 0$ and $P_{Y_j|X}(\cdot | 1) \ll P_{Y_j|X}(\cdot | 0)$ for $j = 1, 2$,

$$\mathcal{L}_\delta = \left\{ (L_1, L_2) \in \mathbb{R}^2_+ : \frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \leq 1 \right\}.$$
Theorem (Tan-Lee (2018))

Under Condition 1, the tuple \((L_1, L_2, L_{\text{key}})\) is achievable if and only if

\[
\frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \leq 1
\]

and

\[
L_{\text{key}} \geq \left( \frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \right) L_Z^* - L_1 - L_2,
\]

where

\[
L_Z^* = L^*(P_{Z,Z|X})
\]

is the self-covert capacity of the channel \(X \rightarrow Z\).
Effect of Key Size

Theorem (Tan-Lee (2018))

Under Condition 1, the tuple \((L_1, L_2, L_{\text{key}})\) is achievable if and only if

\[
\frac{L_1}{L^*_1} + \frac{L_2}{L^*_2} \leq 1
\]

and

\[
L_{\text{key}} \geq \left( \frac{L_1}{L^*_1} + \frac{L_2}{L^*_2} \right) L^*_Z - L_1 - L_2,
\]

where

\[
L^*_Z = L^*(P_{Z,Z|X})
\]

is the self-covert capacity of the channel \(X \rightarrow Z\).

If we operate on the boundary of the keyless covert capacity region,

\[
L^*_{\text{key}} = L^*_Z - L_1 - L_2.
\]
Use an optimal code for $X \rightarrow Y_1$ for $\rho n$ channel uses. If $\delta' < \delta$, 

$$
\log M_{1n} \approx \sqrt{\rho n \delta' L_1^*}
$$
Use an optimal code for $X \rightarrow Y_1$ for $\rho n$ channel uses. If $\delta' < \delta$,

$$
\log M_{1n} \approx \sqrt{\rho n \delta' L_1^*}
$$

Use another optimal code for $X \rightarrow Y_2$ over $(1 - \rho)n$ uses. Then,

$$
\log M_{2n} \approx \sqrt{(1 - \rho)n(\delta - \delta')L_2^*}
$$
Use an optimal code for $X \rightarrow Y_1$ for $\rho n$ channel uses. If $\delta' < \delta$,

$$\log M_{1n} \approx \sqrt{\rho n \delta' L_1^*}$$

Use another optimal code for $X \rightarrow Y_2$ over $(1 - \rho)n$ uses. Then,

$$\log M_{2n} \approx \sqrt{(1 - \rho)n(\delta - \delta') L_2^*}$$

Choose $\delta' = (1 - \rho)\delta$ achieves the point $(\rho L_1^*, (1 - \rho) L_2^*)$. 
Covert communication implies that $X^n$ must have low weight of order $\Theta\left(\frac{1}{\sqrt{n}}\right)$ [Wang, Wornell, Zheng ’16 and Bloch ’16], i.e.,

$$|\{i \in [n] : X_i = 1\}| = \Theta(\sqrt{n})$$
Covert communication implies that $X^n$ must have low weight of order $\Theta(\frac{1}{\sqrt{n}})$ [Wang, Wornell, Zheng ’16 and Bloch ’16], i.e.,

$$|\{i \in [n] : X_i = 1\}| = \Theta(\sqrt{n})$$

Hence throughput $\log M_n$ is of the order $\Theta(\sqrt{n})$
Covert communication implies that $X^n$ must have low weight of order $\Theta\left(\frac{1}{\sqrt{n}}\right)$ [Wang, Wornell, Zheng ’16 and Bloch ’16], i.e.,

$$|\{i \in [n] : X_i = 1\}| = \Theta(\sqrt{n})$$

Hence throughput $\log M_n$ is of the order $\Theta(\sqrt{n})$

For illustration purposes, consider a BS-BC $P_{Y_1,Y_2|X}$ is such that $P_{Y_j|X}$ for $j = 1, 2$ are BSCs.
Covert communication implies that $X^n$ must have low weight of order $\Theta\left(\frac{1}{\sqrt{n}}\right)$ [Wang, Wornell, Zheng ’16 and Bloch ’16], i.e.,

$$|\{i \in [n] : X_i = 1\}| = \Theta(\sqrt{n})$$

Hence throughput $\log M_n$ is of the order $\Theta(\sqrt{n})$.

For illustration purposes, consider a BS-BC $P_{Y_1,Y_2|X}$ is such that $P_{Y_j|X}$ for $j = 1, 2$ are BSCs.

Same intuition for Gaussian broadcast channels.

But use Entropy Power Inequality instead of Mrs. Gerber’s Lemma.
Superposition coding: Cloud center $u_2^n(w_2)$ carries message $w_2$; Satellite codeword $x^n(w_1, w_2) = u_1^n(w_1) \oplus u_2^n(w_2)$ carries $(w_1, w_2)$
**Superposition coding**: Cloud center $u^n_2(w_2)$ carries message $w_2$;
Satellite codeword $x^n(w_1, w_2) = u^n_1(w_1) \oplus u^n_2(w_2)$ carries $(w_1, w_2)$
Superposition coding: Cloud center $u_2^n(w_2)$ carries message $w_2$; Satellite codeword $x^n(w_1, w_2) = u_1^n(w_1) \oplus u_2^n(w_2)$ carries $(w_1, w_2)$

Since $x^n(w_1, w_2)$ has low weight (say $\alpha_n$) and $u_1^n(w_1)$ and $u_2^n(w_2)$ are randomly chosen, locations of 1’s in $u_1^n(w_1)$ and $u_2^n(w_2)$ are not likely to overlap.
Superposition coding: Cloud center $u_n^2(w_2)$ carries message $w_2$; Satellite codeword $x_n(w_1, w_2) = u_1^n(w_1) \oplus u_2^n(w_2)$ carries $(w_1, w_2)$

Since $x_n(w_1, w_2)$ has low weight (say $\alpha_n$) and $u_1^n(w_1)$ and $u_2^n(w_2)$ are randomly chosen, locations of 1’s in $u_1^n(w_1)$ and $u_2^n(w_2)$ are not likely to overlap.

Assume weight of $u_1^n(w_1)$ is $\rho \alpha_n$ and that of $u_2^n(w_2)$ is $(1 - \rho) \alpha_n$. 

\[
\begin{array}{c}
  u_1^n(w_1) \\
  \oplus \\
  u_2^n(w_2) \\
  \| \\
  x_n(w_1, w_2)
\end{array}
\]
Consider BSBCs

\[ Y_1 = X \oplus N_1, \quad Y_2 = X \oplus N_2, \quad N_j \sim \text{Bern}(p_j), \quad p_2 \geq p_1 \]
Consider BSBCs

\[ Y_1 = X \oplus N_1, \quad Y_2 = X \oplus N_2, \quad N_j \sim \text{Bern}(p_j), \quad p_2 \geq p_1 \]

Put

\[ \text{wt}(u_1^n(w_1)) = \rho \alpha_n n, \quad \text{and} \quad \text{wt}(u_2^n(w_2)) = (1 - \rho) \alpha_n n, \]

the superposition coding inner bound with \( X = U_1 \oplus U_2 \) reads

\((U_2 - X - Y_1 - Y_2)\)

\[ R_1 \leq I(X; Y_1|U_2) = I(U_1; Y_1 \oplus N_1) \approx \rho \alpha_n L_1^* \]

\[ R_2 \leq I(U_2; Y_2) = I(U_2; U_2 \oplus \tilde{N}_2) \approx (1 - \rho) \alpha_n L_2^* \]
Time-Division is Optimal for Some BCs: Why?

- Consider BSBCs
  \[ Y_1 = X \oplus N_1, \quad Y_2 = X \oplus N_2, \quad N_j \sim \text{Bern}(p_j), \quad p_2 \geq p_1 \]

- Put
  \[ \text{wt}(u_1^n(w_1)) = \rho \alpha_n n, \quad \text{and} \quad \text{wt}(u_2^n(w_2)) = (1 - \rho) \alpha_n n, \]

the superposition coding inner bound with \( X = U_1 \oplus U_2 \) reads

\[
(U_2 - X - Y_1 - Y_2)
\]

\[ R_1 \leq I(X; Y_1 | U_2) = I(U_1; Y_1 \oplus N_1) \approx \rho \alpha_n L_1^* \]
\[ R_2 \leq I(U_2; Y_2) = I(U_2; U_2 \oplus \tilde{N}_2) \approx (1 - \rho) \alpha_n L_2^* \]

- Hence, we can write
  \[
  \frac{R_1}{L_1^*} + \frac{R_2}{L_2^*} \approx \alpha_n
  \]
Lemma (El Gamal (1979))

Every \( (n, M_1^n, M_2^n, \varepsilon^n) \)-code for any BC satisfies

\[
\left( \log M_1^n \right) \left( 1 - \varepsilon^n \right) - 1 \leq \sum_{i=1}^{n} I(U_1^i; Y_1^i)
\]

\[
\left( \log M_2^n \right) \left( 1 - \varepsilon^n \right) - 1 \leq \sum_{i=1}^{n} I(U_2^i; Y_2^i)
\]

\[
\left( \log M_1^n + \log M_2^n \right) \left( 1 - \varepsilon^n \right) - 2 \leq \sum_{i=1}^{n} \left[ I(X_i; Y_1^i | U_2^i) + I(U_2^i; Y_2^i) \right]
\]

where \( U_1^i \) and \( U_2^i \) satisfy \((U_1^i, U_2^i) - X_i - (Y_1^i, Y_2^i)\).
Lemma (El Gamal (1979))

Every \((n, M_{1n}, M_{2n}, \varepsilon_n)\)-code for any BC satisfies

\[
\begin{align*}
(\log M_{1n})(1 - \varepsilon_n) - 1 & \leq \sum_{i=1}^{n} I(U_{1i}; Y_{1i}) \\
(\log M_{2n})(1 - \varepsilon_n) - 1 & \leq \sum_{i=1}^{n} I(U_{2i}; Y_{2i})
\end{align*}
\]
Lemma (El Gamal (1979))

Every \((n, M_{1n}, M_{2n}, \varepsilon_n)\)-code for any BC satisfies

\[
(\log M_{1n})(1 - \varepsilon_n) - 1 \leq \sum_{i=1}^{n} I(U_{1i}; Y_{1i})
\]

\[
(\log M_{2n})(1 - \varepsilon_n) - 1 \leq \sum_{i=1}^{n} I(U_{2i}; Y_{2i})
\]

\[
(\log M_{1n} + \log M_{2n})(1 - \varepsilon_n) - 2 \leq \sum_{i=1}^{n} \left[ I(X_i; Y_{1i}|U_{2i}) + I(U_{2i}; Y_{2i}) \right]
\]

\[
(\log M_{1n} + \log M_{2n})(1 - \varepsilon_n) - 2 \leq \sum_{i=1}^{n} \left[ I(U_{1i}; Y_{1i}) + I(X_i; Y_{2i}|U_{1i}) \right],
\]

where \(U_{1i}\) and \(U_{2i}\) satisfy \((U_{1i}, U_{2i}) - X_i - (Y_{1i}, Y_{2i})\).
Converse Proof: Upper Bound on $\lambda$-Sum Rate

- Assume $L_1^* \geq L_2^*$ (wlog) and let

$$\lambda = \frac{L_1^*}{L_2^*} \geq 1.$$
Converse Proof: Upper Bound on $\lambda$-Sum Rate

- Assume $L_1^* \geq L_2^*$ (wlog) and let

  $$\lambda = \frac{L_1^*}{L_2^*} \geq 1.$$  

- Combining previous inequalities and using a standard time-sharing random variable, we obtain

  $$\frac{1}{n} \left[ \log M_{1n} + \lambda \log M_{2n} - (1 + \lambda) \right] \leq \max_{U,X} I(X; Y_1|U) + \lambda I(U; Y_2)$$
Converse Proof: Upper Bound on $\lambda$-Sum Rate

- Assume $L_1^* \geq L_2^* \text{ (wlog)}$ and let
  \[
  \lambda = \frac{L_1^*}{L_2^*} \geq 1.
  \]

- Combining previous inequalities and using a standard
time-sharing random variable, we obtain
  \[
  \frac{1}{n} \left[ \log M_{1n} + \lambda \log M_{2n} - (1 + \lambda) \right] \leq \max_{U,X} I(X; Y_1|U) + \lambda I(U; Y_2)
  \]

- Problem: Maximization of $I(X; Y_1|U) + \lambda I(U; Y_2)$ over all $(U, X)$
  requires tools specific to the broadcast channel
Converse Proof: Upper Bound on $\lambda$-Sum Rate

- Assume $L_1^* \geq L_2^*$ (wlog) and let

\[ \lambda = \frac{L_1^*}{L_2^*} \geq 1. \]

- Combining previous inequalities and using a standard time-sharing random variable, we obtain

\[
\frac{1}{n} \left[ \log M_{1n} + \lambda \log M_{2n} - (1 + \lambda) \right] \leq \max_{U,X} I(X; Y_1|U) + \lambda I(U; Y_2)
\]

- Problem: Maximization of $I(X; Y_1|U) + \lambda I(U; Y_2)$ over all $(U, X)$ requires tools specific to the broadcast channel

- For the BS-BC, Mrs. Gerber’s Lemma [Wyner-Ziv (1973)] helps to simplify
Converse Proof: Concave Envelopes

- Remove $U$'s by exploiting tools from convex analysis

\[ \max P_{U,X} I(X;Y_1|U) + \lambda I(U;Y_2) = \max P_{U,X} I(X;Y_1|U) + \lambda [I(X;Y_2) - I(X;Y_2|U)] \]

Now, \[ \max P_{U|X} [I(X;Y_1|U) - \lambda I(X;Y_2|U)] = C[I(P_X, W) - \lambda I(P_X, V)] \]

where $W = P_{Y_1|X}$ and $V = P_{Y_2|X}$ and the concave envelope is defined as $C[f](x) := \inf \{ g(x) : g \geq f, g \text{ is concave} \}$
Converse Proof: Concave Envelopes

- Remove $U$’s by exploiting tools from convex analysis

- Note that $U - X - Y_2$ forms a Markov chain so

$$\max_{P_{U,X}} I(X; Y_1 | U) + \lambda I(U; Y_2)$$

$$= \max_{P_{U,X}} I(X; Y_1 | U) + \lambda [I(X; Y_2) - I(X; Y_2 | U)]$$

$$= \max_{P_X} \lambda I(X; Y_2) + \max_{P_{U|X}} [I(X; Y_1 | U) - \lambda I(X; Y_2 | U)]$$
Converse Proof: Concave Envelopes

- Remove $U$’s by exploiting tools from convex analysis

- Note that $U - X - Y_2$ forms a Markov chain so

  \[
  \max_{P_{U,X}} I(X; Y_1|U) + \lambda I(U; Y_2) \\
  = \max_{P_{U,X}} I(X; Y_1|U) + \lambda [I(X; Y_2) - I(X; Y_2|U)] \\
  = \max_{P_X} \lambda I(X; Y_2) + \max_{P_{U|X}} [I(X; Y_1|U) - \lambda I(X; Y_2|U)]
  \]

- Now,

  \[
  \max_{P_{U|X}} [I(X; Y_1|U) - \lambda I(X; Y_2|U)] = C[I(P_X, W) - \lambda I(P_X, V)]
  \]

  where $W = P_{Y_1|X}$ and $V = P_{Y_2|X}$ and the concave envelope is defined as

  \[
  C[f](x) := \inf\{g(x) : g \geq f, g \text{ is concave}\}
  \]
Converse Proof: Concave Envelopes

The usual superposition coding region is

\[ C = \bigcup_{P_X, P_{U|X}} \left\{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 \leq I(X; Y_1 | U), R_2 \leq I(U; Y_2) \right\} \]

Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \left\{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_\lambda(X) \right\} \]

where \( T_\lambda(X) := \mathbb{C}[I(X; Y_1) - \lambda I(X; Y_2)] \).
Converse Proof: Concave Envelopes

- The usual superposition coding region is

\[ C = \bigcup_{P_X, P_U|X} \{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2) \} \]

- Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_\lambda(X) \} \]

where \( T_\lambda(X) := C[I(X; Y_1) - \lambda I(X; Y_2)] \).
The usual superposition coding region is

$$C = \bigcup_{P_X, P_{U|X}} \{ (R_1, R_2) \in \mathbb{R}_+^2 \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2) \}$$

Using the concave envelope representation, we have

$$C = \bigcap_{\lambda \geq 1} \{ (R_1, R_2) \in \mathbb{R}_+^2 \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_{\lambda}(X) \}$$

where $T_{\lambda}(X) := \mathbb{C}[I(X; Y_1) - \lambda I(X; Y_2)]$. 
The usual superposition coding region is

\[ C = \bigcup_{P_X, P_U|X} \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2) \right\} \]

Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_\lambda(X) \right\} \]

where \( T_\lambda(X) := C[I(X; Y_1) - \lambda I(X; Y_2)] \).
Converse Proof: Concave Envelopes

- The usual superposition coding region is

\[ C = \bigcup_{P_X, P_{U|X}} \{(R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2)\} \]

- Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \{(R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T(\lambda)(X)\} \]

where \( T(\lambda)(X) := C[I(X; Y_1) - \lambda I(X; Y_2)] \).
The usual superposition coding region is

\[ C = \bigcup_{P_X, P_U|X} \left\{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2) \right\} \]

Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \left\{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_{\lambda}(X) \right\} \]

where \( T_{\lambda}(X) := C[I(X; Y_1) - \lambda I(X; Y_2)] \).
Converse Proof: Concave Envelopes

The usual superposition coding region is

\[ C = \bigcup_{P_X, P_U|X} \{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 \leq I(X; Y_1|U), R_2 \leq I(U; Y_2) \} \]

Using the concave envelope representation, we have

\[ C = \bigcap_{\lambda \geq 1} \{ (R_1, R_2) \in \mathbb{R}^2_+ \mid R_1 + \lambda R_2 \leq \max_{P_X} \lambda I(X; Y_2) + T_\lambda(X) \} \]

where \( T_\lambda(X) := C[I(X; Y_1) - \lambda I(X; Y_2)] \).
Converse Proof: A Tiny Bit of Analysis

Converse bound becomes

\[
\frac{\log M_{1n} + \lambda \log M_{2n}}{n} \leq \max_{P_X} \lambda \cdot I(P_X, V) + C[I(P_X, W) - \lambda \cdot I(P_X, V)]
\]
Converse Proof: A Tiny Bit of Analysis

- Converse bound becomes

\[
\frac{\log M_1 n + \lambda \log M_2 n}{n} \leq \max_{P_X} \lambda \cdot I(P_X, V) + \mathbb{C}[I(P_X, W) - \lambda \cdot I(P_X, V)]
\]

- Max over \( P_X = [1 - \alpha_n, \alpha_n] \) is over binary dist. with small mass at 1
Converse Proof: A Tiny Bit of Analysis

- Converse bound becomes
  \[
  \frac{\log M_1 n + \lambda \log M_2}{n} \lesssim \max_{P_X} \lambda \cdot I(P_X, V) + \mathbb{C}[I(P_X, W) - \lambda \cdot I(P_X, V)]
  \]

- Max over \( P_X = [1 - \alpha_n, \alpha_n] \) is over binary dist. with small mass at 1

- Using Condition 1, we have \( I(P_X, W) - \lambda \cdot I(P_X, V) \leq 0 \) for all \( P_X \)
Converse Proof: A Tiny Bit of Analysis

- Converse bound becomes
  \[
  \frac{\log M_{1n} + \lambda \log M_{2n}}{n} \leq \max_{P_X} \lambda \cdot I(P_X, V) + C[I(P_X, W) - \lambda \cdot I(P_X, V)]
  \]

- Max over \( P_X = [1 - \alpha_n, \alpha_n] \) is over binary dist. with small mass at 1

- Using Condition 1, we have \( I(P_X, W) - \lambda \cdot I(P_X, V) \leq 0 \) for all \( P_X \)

- \( C[I(P_X, W) - \lambda \cdot I(P_X, V)] \leq 0 \) for all \( P_X \).
Converse Proof: A Tiny Bit of Analysis

- Converse bound becomes
  \[
  \frac{\log M_1 n + \lambda \log M_2 n}{n} \leq \max_{P_X} \lambda \cdot I(P_X, V) + C \left[ I(P_X, W) - \lambda \cdot I(P_X, V) \right]
  \]

- Max over \( P_X = [1 - \alpha_n, \alpha_n] \) is over binary dist. with small mass at 1

- Using Condition 1, we have \( I(P_X, W) - \lambda \cdot I(P_X, V) \leq 0 \) for all \( P_X \)

- \( C \left[ I(P_X, W) - \lambda \cdot I(P_X, V) \right] \leq 0 \) for all \( P_X \).

- Left with \( I(P_X, V) \approx \alpha_n D(V(\cdot|1)||V(\cdot|0)) \), which is related to \( L_2^* \).
Converse Proof: A Tiny Bit of Analysis

- Converse bound becomes
  \[
  \frac{\log M_{1n} + \lambda \log M_{2n}}{n} \lesssim \max_{P_X} \lambda \cdot I(P_X, V) + \mathbb{C}[I(P_X, W) - \lambda \cdot I(P_X, V)]
  \]

- Max over \( P_X = [1 - \alpha_n, \alpha_n] \) is over binary dist. with small mass at 1

- Using Condition 1, we have \( I(P_X, W) - \lambda \cdot I(P_X, V) \leq 0 \) for all \( P_X \)

- \( \mathbb{C}[I(P_X, W) - \lambda \cdot I(P_X, V)] \leq 0 \) for all \( P_X \).

- Left with \( I(P_X, V) \approx \alpha_n D(V(\cdot|1)||V(\cdot|0)) \), which is related to \( L_2^* \).

- Finally, recalling that \( \lambda = \frac{L_1^*}{L_2^*} \),
  \[
  L_1 + \frac{L_1^*}{L_2^*} \cdot L_2 \lesssim \frac{L_1^*}{L_2^*} \cdot L_2 = L_1^*, \quad \Rightarrow \quad \frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \leq 1.
  \]
Conclusion and Open Problems

Concave envelope representation of bounds on capacity region with auxiliary RVs is very useful.

What can we say about BCs which don't satisfy Condition 1?

More than 2 legitimate receivers?
Concave envelope representation of bounds on capacity region with auxiliary RVs is very useful

- Concave envelope representation of bounds on capacity region with auxiliary RVs is very useful
Concave envelope representation of bounds on capacity region with auxiliary RVs is very useful.

What can we say about BCs which don’t satisfy Condition 1?
Concave envelope representation of bounds on capacity region with auxiliary RVs is very useful.

What can we say about BCs which don’t satisfy Condition 1?

More than 2 legitimate receivers?
Multiple Postdoc Positions in IT and ML at NUS

The Department of Electrical and Computer Engineering (ECE) at the National University of Singapore is offering positions for postdoctoral fellows who will work in information theory, machine learning and their intersection. The Department of Electrical and Computer Engineering (ECE) at the National University of Singapore (NUS) is offering positions for postdoctoral fellows who will work closely with Dr. Vincent Tan at the intersection of information theory, statistical signal processing, and machine learning. Some sample topics include:

- Fundamental performance limits (and algorithms) for dictionary learning (e.g., matrix factorization), ranking, and deep learning architectures;
- Learning in the presence of privacy constraints;
- Learning in the large alphabet regime;
- Learning of graphical models and other statistical models.

Working in traditional topics in Shannon’s information theory of interest to the PI will also be highly encouraged. Some sample topics include:

- Multi-user information theory;
- Strong converse and second-order asymptotics;
- Error exponent analysis and the method of types;
- Information-theoretic security;