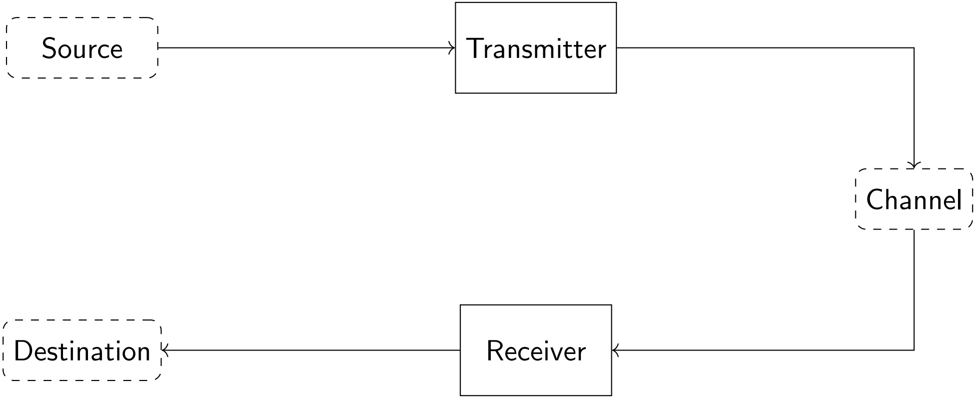
New Paradigms in Source Coding and Channel Coding

Adeel Mahmood

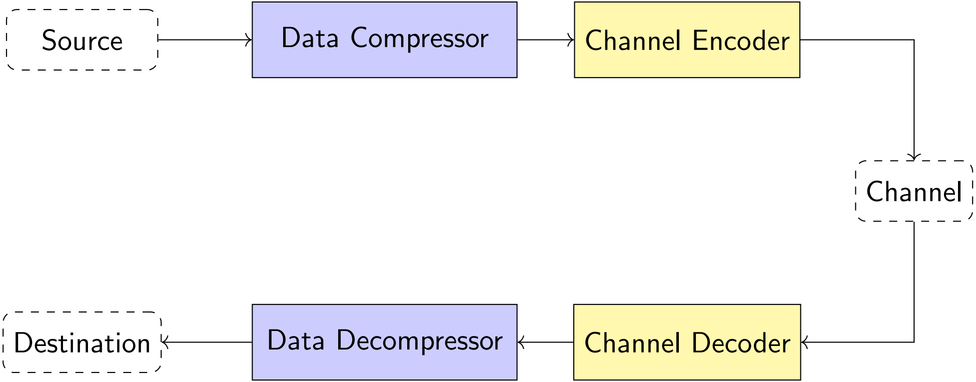
School of Electrical and Computer Engineering
Cornell University

B Exam

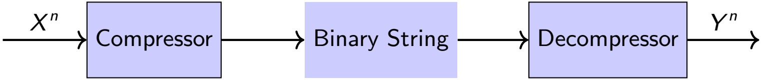
Shannon's Source-Channel Separation Theorem



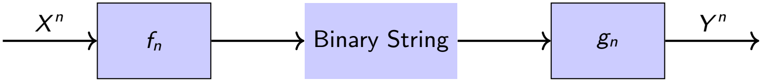
Shannon's Source-Channel Separation Theorem



Source Coding

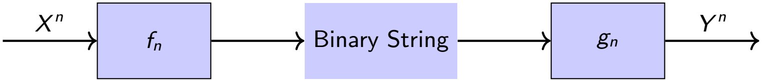


Source Coding



$$C_n = (f_n, g_n)$$

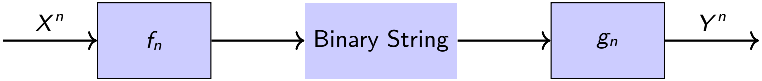
Source Coding



$$C_n = (f_n, g_n)$$

Define the compression rate

$$R(C_n, p) := \frac{\mathbb{E}\left[\mathrm{len}(f_n(X^n))\right]}{n}$$

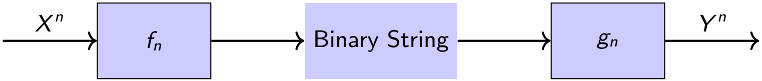


$$C_n = (f_n, g_n)$$

Define the compression rate

$$R(C_n, p) := \frac{\mathbb{E}\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{P_{Y|X}:\mathbb{E}[\rho(X,Y)]\leq d} I(X;Y)$$

Source Coding



$$C_n = (f_n, g_n)$$

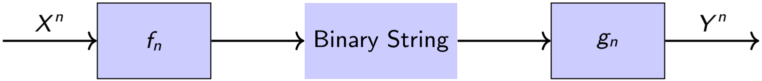
Define the compression rate

$$R(C_n, p) := \frac{\mathbb{E}\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{P_{Y|X} : \mathbb{E}[\rho(X,Y)] \leq d} I(X; Y)$$

A compression scheme C_n is ...

- ... <u>non-universal</u> if C_n depends on p and $\lim_{n \to \infty} R(C_n, p) - R(p, d) = 0$

Source Coding



$$C_n = (f_n, g_n)$$

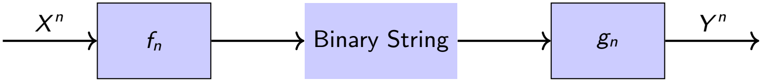
Define the compression rate

$$R(C_n, p) := \frac{\mathbb{E}\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{P_{Y|X}:\mathbb{E}[\rho(X,Y)]\leq d} I(X;Y)$$

A compression scheme C_n is ...

- ... non-universal if C_n depends on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$
- ... weakly universal if C_n does not depend on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$ for all p

Source Coding



$$C_n = (f_n, g_n)$$

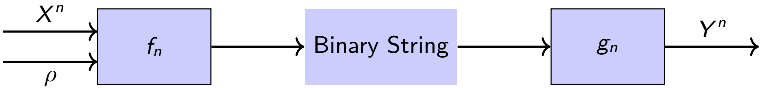
Define the compression rate

$$R(C_n, p) := \frac{\mathbb{E}\left[\mathrm{len}(f_n(X^n))\right]}{n} \geq R(p, d) := \min_{P_{Y|X} : \mathbb{E}[\rho(X,Y)] \leq d} I(X; Y)$$

A compression scheme C_n is ...

- ... <u>non-universal</u> if C_n depends on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$
- ... <u>weakly</u> universal if C_n does not depend on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$ for all p
- ... <u>strongly</u> universal if C_n does not depend on p and $\lim_{n\to\infty} \sup_p \left[R(C_n, p) - R(p, d)\right] = 0$

Source Coding



$$C_n = (f_n, g_n)$$

Define the compression rate

$$R(C_n, p) \coloneqq \frac{\mathbb{E}\left[\operatorname{len}(f_n(X^n))\right]}{n} \geq R(p, d) \coloneqq \min_{P_{Y|X}:\mathbb{E}[\rho(X,Y)]\leq d} I(X;Y)$$

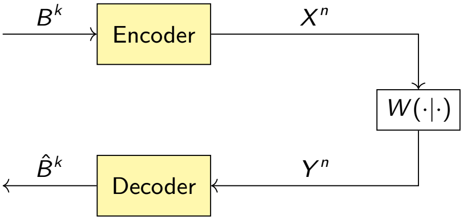
A compression scheme C_n is . . .

- . . . <u>non-universal</u> if C_n depends on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$
- . . . <u>weakly</u> universal if C_n does not depend on p and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$ for all p
- . . . <u>strongly</u> universal if C_n does not depend on p and $\lim_{n\to\infty} \sup_p \left[R(C_n, p) - R(p, d)\right] = 0$
- . . . <u>universal distortion</u> if C_n does not depend on p and ρ and $\lim_{n\to\infty} R(C_n, p) - R(p, d) = 0$

References

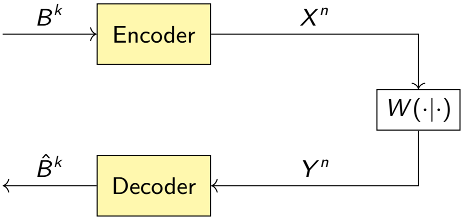
- A. Mahmood and A. B. Wagner, "Minimax Rate-Distortion," in IEEE Transactions on Information Theory, 2023
- A. Mahmood and A. B. Wagner, "Lossy Compression With Universal Distortion," in IEEE Transactions on Information Theory, 2023

Channel Coding



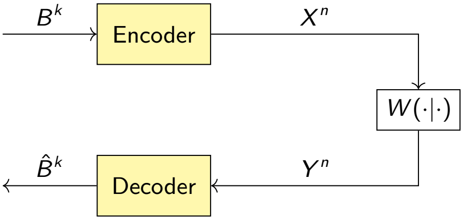
- Input alphabet $\mathcal{X}$; output alphabet $\mathcal{Y}$
- $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$

Channel Coding



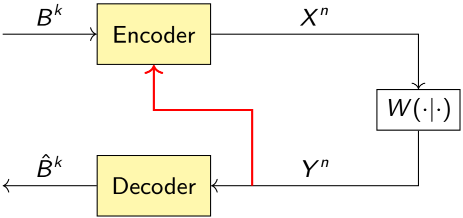
- Input alphabet $\mathcal{X}$; output alphabet $\mathcal{Y}$
- $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$
- Assume a memoryless channel: $W(Y^n|X^n) = \prod_{i=1}^{n} W(Y_i|X_i)$

Channel Coding



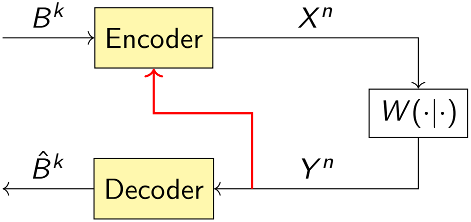
- Input alphabet $\mathcal{X}$; output alphabet $\mathcal{Y}$
- $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$
- Assume a memoryless channel: $W(Y^n|X^n) = \prod_{i=1}^{n} W(Y_i|X_i)$
- Two types of channels:
  - DMC: $\mathcal{X}$ and $\mathcal{Y}$ are finite
  - AWGN: $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $W(\cdot|x) = \mathcal{N}(x, 1)$

Channel Coding



- Input alphabet $\mathcal{X}$; output alphabet $\mathcal{Y}$
- $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$
- Assume a memoryless channel: $W(Y^n|X^n) = \prod_{i=1}^n W(Y_i|X_i)$
- Two types of channels:
  - DMC: $\mathcal{X}$ and $\mathcal{Y}$ are finite
  - AWGN: $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $W(\cdot|x) = \mathcal{N}(x, 1)$
- With feedback, $\text{Enc}(B^k, Y^{m-1}) = X_m$ for all $m = 1, \ldots, n$

The Channel Coding Problem



- Maximize the rate $R = k/n$
- Minimize the probability of error $\Pr(B^k \neq \hat{B}^k)$

Known Results

Definition: $P_e(n, R) := \min\{\epsilon : \exists \text{ an } (n, R) \text{ code with avg. error prob.} \leq \epsilon\}$

Known Results

Definition: $P_e(n,R) := \min\{\epsilon : \exists \text{ an } (n,R) \text{ code with avg. error prob. } \leq \epsilon\}$

Theorem (for DMC):

$$\lim_{n\to\infty} P_e(n,R) = \begin{cases} 0 & \text{if } R < C \quad \text{[Shannon '48]} \\ 1 & \text{if } R > C \quad \text{[Wolfowitz '57]} \end{cases}$$

$$\text{where } C = \max_{P_X} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

Known Results

Definition: $P_e(n,R) := \min\{\epsilon : \exists \text{ an } (n,R) \text{ code with avg. error prob. } \leq \epsilon\}$

Theorem (for DMC):

$$\lim_{n\to\infty} P_e(n,R) = \begin{cases} 0 & \text{if } R < C \quad \text{[Shannon '48]} \\ 1 & \text{if } R > C \quad \text{[Wolfowitz '57]} \end{cases}$$

$$\text{Let } R = C + \frac{\beta}{\sqrt{n}}$$

$$\text{where } C = \max_{P_X} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

Known Results

Definition: $P_e(n,R) := \min\{\epsilon : \exists \text{ an } (n,R) \text{ code with avg. error prob.} \leq \epsilon\}$

Theorem (for DMC):

$$\lim_{n\to\infty} P_e(n,R) = \begin{cases} 0 & \text{if } R < C \quad \text{[Shannon '48]} \\ 1 & \text{if } R > C \quad \text{[Wolfowitz '57]} \end{cases}$$

$$\lim_{n\to\infty} P_e\left(n, C + \frac{\beta}{\sqrt{n}}\right) =$$

$$\text{where } C = \max_{P_X} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

Known Results

Definition: $P_e(n,R) := \min\{\epsilon : \exists \text{ an } (n,R) \text{ code with avg. error prob.} \le \epsilon\}$

Theorem (for DMC):

$$\lim_{n\to\infty} P_e(n,R) = \begin{cases} 0 & \text{if } R < C \quad \text{[Shannon '48]} \\ 1 & \text{if } R > C \quad \text{[Wolfowitz '57]} \end{cases}$$

$$\lim_{n\to\infty} P_e\left(n, C + \frac{\beta}{\sqrt{n}}\right) = \Phi\left(\frac{\beta}{\sqrt{V}}\right) \qquad \text{[cf. Strassen '62]}$$

$$\text{where } C = \max_{P_X} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

$$V = \mathbb{E}\left[\mathrm{Var}\left(\log \frac{W(Y|X)}{P^* W(Y)} \middle| X\right)\right]$$

Known Results

Definition: $P_e(n,R) := \min\{\epsilon : \exists \text{ an } (n,R) \text{ code with avg. error prob.} \le \epsilon\}$

Theorem (for DMC):

$$\lim_{n\to\infty} P_e(n,R) = \begin{cases} 0 & \text{if } R < C \quad \text{[Shannon '48]} \\ 1 & \text{if } R > C \quad \text{[Wolfowitz '57]} \end{cases}$$

$$\lim_{n\to\infty} P_e\left(n, C + \frac{\beta}{\sqrt{n}}\right) = \Phi\left(\frac{\beta}{\sqrt{V}}\right) \qquad \text{[cf. Strassen '62]}$$

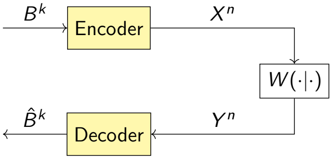
Theorem (for AWGN):

$$\lim_{n\to\infty} P_e(n,R) = 0 \text{ for all } R$$

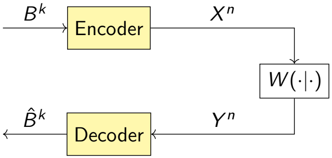
$$\text{where } C = \max_{P_X} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

$$V = \mathbb{E}\left[\mathrm{Var}\left(\log \frac{W(Y|X)}{P^* W(Y)}\middle| X\right)\right]$$

Channel Coding with Cost Constraint

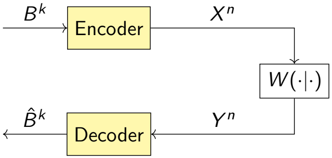


Channel Coding with Cost Constraint



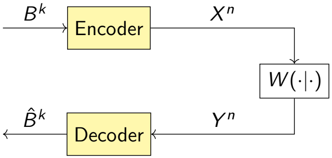
- $c(\cdot) : \mathcal{X} \mapsto [0, \infty)$

Channel Coding with Cost Constraint



- $c(\cdot) : \mathcal{X} \mapsto [0, \infty)$
- cost threshold: Γ

Channel Coding with Cost Constraint



- $c(\cdot) : \mathcal{X} \mapsto [0, \infty)$
- cost threshold: Γ
- **Def** (Almost sure cost constraint): An (n, R, Γ) code is an (n, R) code satisfying

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \quad \text{a.s.}$$

[Han '98; Hayashi '09; Csiszár and Körner '11; Kostina and Verdú '15]

Channel Coding with Cost Constraint

Definition: $P_e(n, R, \Gamma) := \min\{\epsilon : \exists \text{ an } (n, R, \Gamma) \text{ code with avg. error prob.} \leq \epsilon\}$

Channel Coding with Cost Constraint

Definition: $P_e(n, R, \Gamma) := \min\{\epsilon : \exists \text{ an } (n, R, \Gamma) \text{ code with avg. error prob.} \le \epsilon\}$

Theorem:

$$\lim_{n\to\infty} P_e(n, R, \Gamma) = \begin{cases} 0 & \text{if } R < C(\Gamma) \\ 1 & \text{if } R > C(\Gamma) \end{cases}$$

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right) = \Phi\left(\frac{\beta}{\sqrt{V(\Gamma)}}\right) \quad \text{[Hayashi '09]}$$

$$\text{where } C(\Gamma) = \max_{P_X : \mathbb{E}[c(X)] \le \Gamma} I(X;Y) = \mathbb{E}\left[\log \frac{W(Y|X)}{P^* W(Y)}\right]$$

$$V(\Gamma) = \mathbb{E}\left[\mathrm{Var}\left(\log \frac{W(Y|X)}{P^* W(Y)} \middle| X\right)\right]$$

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.} \quad \text{prohibits codewords generated i.i.d. } P^*,$$

where $P^* = \text{argmax}_{P_X:\mathbb{E}[c(X)]\leq\Gamma} I(X;Y)$.

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.} \quad \text{prohibits codewords generated i.i.d. } P^*,$$

where $P^* = \mathrm{argmax}_{P_X : \mathbb{E}[c(X)] \leq \Gamma} I(X;Y)$.

> **Theorem 7.7.1** (*Channel coding theorem*) *For a discrete memoryless channel, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.*
>
> *Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.*
>
> **Proof:** We prove that rates $R < C$ are achievable and postpone proof of the converse to Section 7.9.
>
> *Achievability*: Fix $p(x)$. Generate a $(2^{nR}, n)$ code at random according to the distribution $p(x)$. Specifically, we generate 2^{nR} codewords independently according to the distribution
>
> $$p(x^n) = \prod_{i=1}^{n} p(x_i). \qquad (7.60)$$

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.} \quad \text{prohibits codewords generated i.i.d. } P^*,$$

where $P^* = \operatorname{argmax}_{P_X:\mathbb{E}[c(X)]\leq\Gamma} I(X;Y)$.

Theorem 7.7.1 (*Channel coding theorem*) *For a discrete memoryless channel, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.*

Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

Proof: We prove that rates $R < C$ are achievable and postpone proof of the converse to Section 7.9.

Achievability: Fix $p(x)$. Generate a $(2^{nR}, n)$ code at random according to the distribution $p(x)$. Specifically, we generate 2^{nR} codewords independently according to the distribution

$$p(x^n) = \prod_{i=1}^{n} p(x_i). \tag{7.60}$$

In fact, when $\boldsymbol{P}$ is the i.i.d. of P_{M+} or P_{M-}, $I(\epsilon, C \,|\, \boldsymbol{P}, \boldsymbol{W})$ is equal to $\sqrt{V_W^+ F^{-1}(\epsilon)}$ or $\sqrt{V_W^- F^{-1}(\epsilon)}$. Thus, (66) holds. Therefore, the achievability part (the direct part) of Theorem 2 holds. Therefore, it is sufficient to prove the converse part (67).

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.} \quad \text{prohibits codewords generated i.i.d. } P^*,$$

where $P^* = \text{argmax}_{P_X : \mathbb{E}[c(X)] \leq \Gamma} I(X;Y)$.

Theorem 7.7.1 (*Channel coding theorem*) *For a discrete memoryless channel, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.*

Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

Proof: We prove that rates $R < C$ are achievable and postpone proof of the converse to Section 7.9.

Achievability: Fix $p(x)$. Generate a $(2^{nR}, n)$ code at random according to the distribution $p(x)$. Specifically, we generate 2^{nR} codewords independently according to the distribution

$$p(x^n) = \prod_{i=1}^{n} p(x_i). \tag{7.60}$$

Capacity of wireless channels

one picks the codewords randomly and independently, with the components of each codeword i.i.d. $\mathcal{N}(0, P)$, then with very high probability the randomly chosen code will do the job at any rate $R < C$. This is the so-called *i.i.d. Gaussian code*. A sketch of this random coding argument can be found in

In fact, when $\boldsymbol{P}$ is the i.i.d. of P_{M+} or P_{M-}, $I(\epsilon, C \,|\, \boldsymbol{P}, \boldsymbol{W})$ is equal to $\sqrt{V_W^+ F^{-1}(\epsilon)}$ or $\sqrt{V_W^- F^{-1}(\epsilon)}$. Thus, (66) holds. Therefore, the achievability part (the direct part) of Theorem 2 holds. Therefore, it is sufficient to prove the converse part (67).

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.} \quad \text{prohibits codewords generated i.i.d. } P^*,$$

where $P^* = \text{argmax}_{P_X : \mathbb{E}[c(X)] \leq \Gamma} I(X;Y)$.

Theorem 7.7.1 (*Channel coding theorem*) *For a discrete memoryless channel, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.*

Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

Proof: We prove that rates $R < C$ are achievable and postpone proof of the converse to Section 7.9.

Achievability: Fix $p(x)$. Generate a $(2^{nR}, n)$ code at random according to the distribution $p(x)$. Specifically, we generate 2^{nR} codewords independently according to the distribution

$$p(x^n) = \prod_{i=1}^{n} p(x_i). \tag{7.60}$$

Capacity of wireless channels

one picks the codewords randomly and independently, with the components of each codeword i.i.d. $\mathcal{N}(0, P)$, then with very high probability the randomly chosen code will do the job at any rate $R < C$. This is the so-called *i.i.d. Gaussian code*. A sketch of this random coding argument can be found in

Code construction: We define the common randomness (revealed to the encoder and decoder before the transmission starts) to be a random variable U as follows:

$$P_U \triangleq \underbrace{P_{X^\infty} \times \ldots \times P_{X^\infty}}_{|\mathcal{W}|}, \tag{19}$$

where X^∞ consists of i.i.d. copies drawn from (any) capacity-achieving input distribution. A realization of U defines $|\mathcal{W}|$

In fact, when $\boldsymbol{P}$ is the i.i.d. of P_{M+} or P_{M-}, $I(\epsilon, C \,|\, \boldsymbol{P}, \boldsymbol{W})$ is equal to $\sqrt{V_W^+}F^{-1}(\epsilon)$ or $\sqrt{V_W^-}F^{-1}(\epsilon)$. Thus, (66) holds. Therefore, the achievability part (the direct part) of Theorem 2 holds. Therefore, it is sufficient to prove the converse part (67).

Limitations of Almost Sure Cost Constraint

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}}$$

1. Prohibits codewords generated i.i.d. P^*

Limitations of Almost Sure Cost Constraint

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}}$$

1. Prohibits codewords generated i.i.d. P^*
2. Does not allow random encoders which produce noise-like channel input

Limitations of Almost Sure Cost Constraint

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}}$$

1. Prohibits codewords generated i.i.d. P^*
2. Does not allow random encoders which produce noise-like channel input
   2.1 A feedback scheme using i.i.d. and c.c. codewords leads to second-order improvement for DMCs [Mahmood and Wagner '23]

Constant-composition (c.c.) codewords: a codebook in which every codeword contains exactly the same number of each symbol, i.e., each codeword is just a different permutation of one fixed symbol-count pattern.

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}$$

1. Prohibits codewords generated i.i.d. P^*
2. Does not allow random encoders which produce noise-like channel input
   2.1 A feedback scheme using i.i.d. and c.c. codewords leads to second-order improvement for DMCs [Mahmood and Wagner '23]
3. Does not allow second-order improvement with feedback in case of unique P^*

Constant-composition (c.c.) codewords: a codebook in which every codeword contains exactly the same number of each symbol, i.e., each codeword is just a different permutation of one fixed symbol-count pattern.

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}$$

1. Prohibits codewords generated i.i.d. P^*
2. Does not allow random encoders which produce noise-like channel input
   2.1 A feedback scheme using i.i.d. and c.c. codewords leads to second-order improvement for DMCs [Mahmood and Wagner '23]
3. Does not allow second-order improvement with feedback in case of unique P^*
   3.1 Even third-order improvement impossible for AWGN channel [Fong and Tan '15]

Constant-composition (c.c.) codewords: a codebook in which every codeword contains exactly the same number of each symbol, i.e., each codeword is just a different permutation of one fixed symbol-count pattern.

Limitations of Almost Sure Cost Constraint

$$\frac{1}{n}\sum_{i=1}^{n} c(X_i) \leq \Gamma \text{ a.s.}$$

1. Prohibits codewords generated i.i.d. P^*
2. Does not allow random encoders which produce noise-like channel input
   2.1 A feedback scheme using i.i.d. and c.c. codewords leads to second-order improvement for DMCs [Mahmood and Wagner '23]
3. Does not allow second-order improvement with feedback in case of unique P^*
   3.1 Even third-order improvement impossible for AWGN channel [Fong and Tan '15]
4. Too strict $\equiv$ No cost variability

Constant-composition (c.c.) codewords: a codebook in which every codeword contains exactly the same number of each symbol, i.e., each codeword is just a different permutation of one fixed symbol-count pattern.

A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$

A Relaxed Cost Constraint

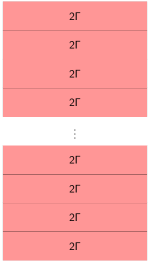
$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$

- Strong converse fails due to non-ergodic power consumption:

  [Polyanskiy '10; Kostina and Verdú '15]

A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$



- Strong converse fails due to non-ergodic power consumption:

  [Polyanskiy '10; Kostina and Verdú '15]

  - Consider a rate-$C(2\Gamma)$ codebook where $C(2\Gamma) > C(\Gamma)$

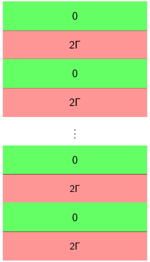
A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$

- Strong converse fails due to non-ergodic power consumption:

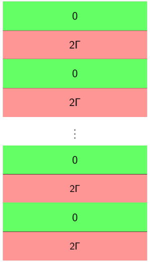
  [Polyanskiy '10; Kostina and Verdú '15]

  - Consider a rate-$C(2\Gamma)$ codebook where $C(2\Gamma) > C(\Gamma)$
  - Zero out half the codewords (assuming $\min_x c(x) = 0$)



A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$



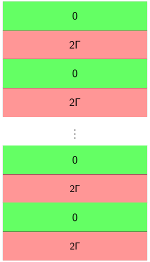
- Strong converse fails due to non-ergodic power consumption:

  [Polyanskiy '10; Kostina and Verdú '15]

  - Consider a rate-$C(2\Gamma)$ codebook where $C(2\Gamma) > C(\Gamma)$
  - Zero out half the codewords (assuming $\min_x c(x) = 0$)
  - $P_e \approx 1/2$

A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma$$

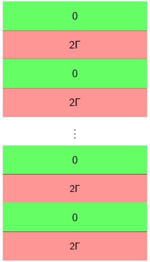


- Strong converse fails due to non-ergodic power consumption:

  [Polyanskiy '10; Kostina and Verdú '15]
  - Consider a rate-$C(2\Gamma)$ codebook where $C(2\Gamma) > C(\Gamma)$
  - Zero out half the codewords (assuming $\min_x c(x) = 0$)
  - $P_e \approx 1/2$
  - Average cost: Γ

A Relaxed Cost Constraint

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \le \Gamma$$



- Strong converse fails due to non-ergodic power consumption:

  [Polyanskiy '10; Kostina and Verdú '15]

  - Consider a rate-$C(2\Gamma)$ codebook where $C(2\Gamma) > C(\Gamma)$
  - Zero out half the codewords (assuming $\min_x c(x) = 0$)
  - $P_e \approx 1/2$
  - Average cost: Γ
- Second-order rate is also infinite:

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right) = 0 \quad \text{for all } \beta.$$

A Goldilocks Cost Constraint

A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint

A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint

A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

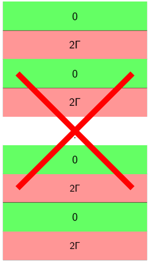
1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint
3. Ensures that the cost concentrates around its mean

A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint
3. Ensures that the cost concentrates around its mean

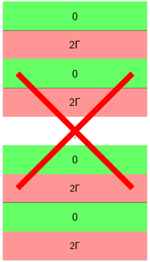


A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint
3. Ensures that the cost concentrates around its mean
4. Does not allow blasting power on errors in the feedback case
5. Yet, feedback improvement is possible

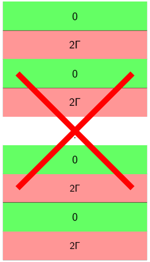


A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint
3. Ensures that the cost concentrates around its mean
4. Does not allow blasting power on errors in the feedback case
5. Yet, feedback improvement is possible
6. Allows for i.i.d. P^* codewords (unless $\Delta < \mathrm{Var}_{P^*}(c(X))$)

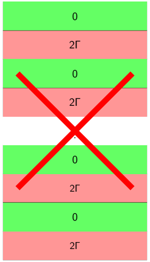


A Goldilocks Cost Constraint

Def: An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

1. Taking $\Delta \to \infty$ recovers the expected cost constraint
2. Taking $\Delta = 0$ is similar to the a.s. cost constraint
3. Ensures that the cost concentrates around its mean
4. Does not allow blasting power on errors in the feedback case
5. Yet, feedback improvement is possible
6. Allows for i.i.d. P^* codewords (unless $\Delta < \mathrm{Var}_{P^*}(c(X))$)
7. Accounts for $O(1/n)$ fluctuations for random encoding



A Goldilocks Cost Constraint

- **Def:** An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

A Goldilocks Cost Constraint

- **Def:** An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

- **Def:** $P_e(n, R, \Gamma, \Delta) \coloneqq \min\{\epsilon : \exists$ an (n, R, Γ, Δ) code with avg. error prob. $\leq \epsilon\}$

A Goldilocks Cost Constraint

- **Def:** An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

- **Def:** $P_e(n, R, \Gamma, \Delta) := \min\{\epsilon : \exists$ an (n, R, Γ, Δ) code with avg. error prob. $\leq \epsilon\}$

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e\left(n, R, \Gamma, \Delta\right) = \begin{cases} 0 & \text{if } R < C(\Gamma) \\ 1 & \text{if } R > C(\Gamma) \end{cases}$$

A Goldilocks Cost Constraint

- **Def:** An (n, R, Γ, Δ) code is an (n, R) code satisfying

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \le \Gamma \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \le \frac{\Delta}{n}$$

- **Def:** $P_e(n, R, \Gamma, \Delta) := \min\{\epsilon : \exists$ an (n, R, Γ, Δ) code with avg. error prob. $\le \epsilon\}$

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e(n, R, \Gamma, \Delta) = \begin{cases} 0 & \text{if } R < C(\Gamma) \\ 1 & \text{if } R > C(\Gamma) \end{cases}$$

$$\begin{aligned} \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) = \quad & \min_{U} && \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right] \\ & \text{subject to} && \mathbb{E}[U] \ge \beta \\ & && \mathrm{Var}(U) \le C'(\Gamma)^2 \Delta \\ & && |\mathrm{supp}(U)| \le 3 \end{aligned}$$

A Goldilocks Cost Constraint

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) = \begin{array}{rl} \min\limits_{U} & \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right] \\ \text{subject to} & \mathbb{E}[U] \geq \beta \\ & \mathrm{Var}(U) \leq C'(\Gamma)^2\Delta \\ & |\mathrm{supp}(U)| \leq 3 \end{array}$$

A Goldilocks Cost Constraint

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) = \min_{U} \quad \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

$$\text{subject to} \quad \mathbb{E}[U] \geq \beta$$

$$\mathrm{Var}(U) \leq C'(\Gamma)^2 \Delta$$

$$|\mathrm{supp}(U)| \leq 3$$

- As $\Delta \to 0$, RHS $\to \Phi\left(\frac{\beta}{\sqrt{V(\Gamma)}}\right)$ [as in a.s. constraint]

A Goldilocks Cost Constraint

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) = \begin{array}{rl} \displaystyle\min_{U} & \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right] \\ \text{subject to} & \mathbb{E}[U] \geq \beta \\ & \mathrm{Var}(U) \leq C'(\Gamma)^2\Delta \\ & |\mathrm{supp}(U)| \leq 3 \end{array}$$

- As $\Delta \to 0$, RHS $\to \Phi\left(\frac{\beta}{\sqrt{V(\Gamma)}}\right)$ [as in a.s. constraint]
- As $\Delta \to \infty$, RHS $\to 0$ [as in expected constraint]

A Goldilocks Cost Constraint

Theorem 1 (Mahmood and Wagner '24, '25)

For any channel with a unique P^*,

$$\lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) = \begin{array}{rl} \displaystyle\min_{U} & \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right] \\ \text{subject to} & \mathbb{E}[U] \geq \beta \\ & \mathrm{Var}(U) \leq C'(\Gamma)^2 \Delta \\ & |\mathrm{supp}(U)| \leq 3 \end{array}$$

- As $\Delta \to 0$, RHS $\to \Phi\left(\frac{\beta}{\sqrt{V(\Gamma)}}\right)$ [as in a.s. constraint]
- As $\Delta \to \infty$, RHS $\to 0$ [as in expected constraint]

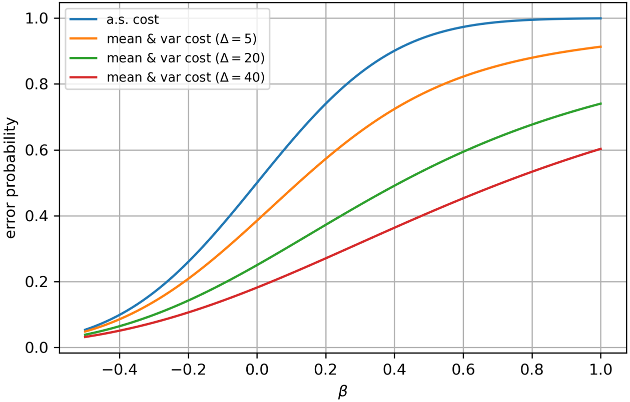
Theorem 2 (Mahmood and Wagner '24)

For any $\Delta > 0$,

$$0 < \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) < \Phi\left(\frac{\beta}{\sqrt{V(\Gamma)}}\right)$$

Numerical Example

BSC with crossover prob. $p = 0.3$, $c(x) = x$, $\Gamma = 0.2$, $\epsilon = 0.1$:



Does Feedback Improve Second-Order Performance?

Theorem 3 (Mahmood and Wagner '24)

For any $\Delta > 0$ and a channel (DMC only) with a unique P^*,

$$\limsup_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) < \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right)$$

Does Feedback Improve Second-Order Performance?

Theorem 3 (Mahmood and Wagner '24)

For any $\Delta > 0$ and a channel (DMC only) with a unique P^*,

$$\limsup_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) < \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right)$$

Theorem 4 (Mahmood and Wagner '24, Fong and Tan '15)

For the a.s. cost constraint and any channel with a unique P^*,

$$\lim_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right) = \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right)$$

Does Feedback Improve Second-Order Performance?

Theorem 3 (Mahmood and Wagner '24)

For any $\Delta > 0$ and a channel (DMC only) with a unique P^*,

$$\limsup_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) < \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right)$$

Theorem 4 (Mahmood and Wagner '24, Fong and Tan '15)

For the a.s. cost constraint and any channel with a unique P^*,

$$\lim_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right) = \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right)$$

- For compound-dispersion DMCs, feedback improves second-order performance [Wagner, Shende, Altuğ '20]

} channels without cost constraints

Does Feedback Improve Second-Order Performance?

Theorem 3 (Mahmood and Wagner '24)

For any $\Delta > 0$ and a channel (DMC only) with a unique P^*,

$$\limsup_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right) < \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma, \Delta\right)$$

Theorem 4 (Mahmood and Wagner '24, Fong and Tan '15)

For the a.s. cost constraint and any channel with a unique P^*,

$$\lim_{n\to\infty} P_{e,fb}\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right) = \lim_{n\to\infty} P_e\left(n, C(\Gamma) + \frac{\beta}{\sqrt{n}}, \Gamma\right)$$

- For compound-dispersion DMCs, feedback improves second-order performance [Wagner, Shende, Altuğ '20]
- For unique P^*, feedback does not improve second-order performance [Wagner, Shende, Altuğ '20]

} channels without cost constraints

Proof: Main Ideas

Lemma (cf. Feinstein '54; Shannon '57; Verdú and Han '94)

Let $(X^n, Y^n) \sim P \circ W$.

Proof: Main Ideas

Lemma (cf. Feinstein '54; Shannon '57; Verdú and Han '94)

Let $(X^n, Y^n) \sim P \circ W$.

1. *(Converse) Every* (n, R) *code induces a distribution* P *on the channel input* X^n *and*

$$P_e \gtrapprox \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

Proof: Main Ideas

Lemma (cf. Feinstein '54; Shannon '57; Verdú and Han '94)

Let $(X^n, Y^n) \sim P \circ W$.

1. *(Converse) Every* (n, R) *code induces a distribution* P *on the channel input* X^n *and*

$$P_e \gtrapprox \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR\right)$$

2. *(Achievability) For a channel input* $X^n \sim P$ *and rate* R*, there is an* (n, R) *code with*

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \le nR\right)$$

Proof: Main Ideas

Lemma (cf. Feinstein '54; Shannon '57; Verdú and Han '94)

Let $(X^n, Y^n) \sim P \circ W$.

1. *(Converse) Every* (n, R) *code induces a distribution* P *on the channel input* X^n *and*

$$P_e \gtrsim \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

2. *(Achievability) For a channel input* $X^n \sim P$ *and rate* R*, there is an* (n, R) *code with*

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Designing a code $\equiv$ **Designing the distribution** P

Proof: Optimal Code Design

Let $(X^n, Y^n) \sim P \circ W$.

1. (Converse)

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

2. (Achievability)

$$P_e \lessapprox \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Converse Proof

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \right)$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

$$\geq \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \middle| X^n = X^n\right)$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

$$\geq \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \middle| X^n = X^n\right)$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$
\begin{aligned}
P_e &\gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right) \\
&\geq \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \middle| X^n = X^n\right) \\
&\geq \inf_P \int_{S^*} dP(x^n)\mathbb{P}\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \leq nR \middle| X^n = x^n\right)
\end{aligned}
$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR\right)$$

$$\ge \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR \middle| X^n = X^n\right)$$

$$\ge \inf_P \int_{S^*} dP(x^n)\mathbb{P}\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \le nR \middle| X^n = x^n\right) \implies P_e \gtrapprox \inf_U \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

$$\text{s.t.} \quad \mathbb{E}[U] \ge \beta$$

$$\mathrm{Var}(U) \le C'(\Gamma)^2\Delta$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$P_e \gtrsim \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

$$\geq \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \middle| X^n = X^n\right)$$

$$\geq \inf_P \int_{S^*} dP(x^n)\mathbb{P}\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \leq nR \middle| X^n = x^n\right) \quad \Longrightarrow \quad P_e \gtrsim \inf_U \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

$$\text{s.t.} \quad \mathbb{E}[U] \geq \beta$$

$$\mathrm{Var}(U) \leq C'(\Gamma)^2\Delta$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n c(X_i)\right] \leq \Gamma \quad \textbf{and} \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n c(X_i)\right) \leq \frac{\Delta}{n}$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$
\begin{aligned}
P_e &\gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right) \\
&\geq \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR \Big| X^n = X^n\right) \\
&\geq \inf_P \int_{S^*} dP(x^n)\mathbb{P}\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \leq nR \Big| X^n = x^n\right)
\end{aligned}
$$

$\Longrightarrow$

$$
P_e \gtrapprox \inf_U \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]
$$

s.t. $\mathbb{E}[U] \geq \beta$

$\mathrm{Var}(U) \leq C'(\Gamma)^2\Delta$

$|\mathrm{supp}(U)| \leq 3$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n c(X_i)\right] \leq \Gamma \quad \textbf{and} \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n c(X_i)\right) \leq \frac{\Delta}{n}$$

Converse Proof

- Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
- Restrict analysis to $S^* = \left\{x^n : \frac{1}{n}\sum_{i=1}^n c(x_i) \approx \Gamma\right\}$
- Choose Q to be capacity-cost achieving output distribution, denoted by Q^*

$$P_e \gtrsim \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR\right)$$

$$\ge \inf_P \sup_Q \int_{S^*} dP(X^n)\mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR \middle| X^n = X^n\right)$$

$$\ge \inf_P \int_{S^*} dP(x^n)\mathbb{P}\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \le nR \middle| X^n = x^n\right) \implies$$

$$P_e \gtrsim \min_U \ \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

s.t. $\mathbb{E}[U] \ge \beta$

$\mathrm{Var}(U) \le C'(\Gamma)^2\Delta$

$|\mathrm{supp}(U)| \le 3$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n c(X_i)\right] \le \Gamma \quad \text{and} \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n c(X_i)\right) \le \frac{\Delta}{n}$$

Achievability Proof

$$P_e \lessapprox \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

↑

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$

$$P_e \lessapprox \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose

$$P = \begin{cases} p_1 \text{Unif}(T^n(t_1)) + p_2 \text{Unif}(T^n(t_2)) + p_3 \text{Unif}(T^n(t_3)) & \text{(for DMC)} \\ p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1}) & \text{(for AWGN)} \end{cases}$$

$$P_e \lessapprox \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

$$P_e \lesssim \inf_{P} \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1})$,

$$P_e \lesssim \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1})$,

   where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

$$P_e \lessapprox \inf_P \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR \right)$$

Infimum is over all distributions P such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right] \leq \Gamma \qquad \textbf{and} \qquad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} c(X_i)\right) \leq \frac{\Delta}{n}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1})$,

   where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \text{Unif}(S^{n-1}_{R_1}) + p_2 \text{Unif}(S^{n-1}_{R_2}) + p_3 \text{Unif}(S^{n-1}_{R_3})$,
   where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \le nR\right)$$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \le nR\right)$$

↓

$$P_e \gtrapprox \min_U \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

s.t. $\mathbb{E}[U] \ge \beta$

$\text{Var}(U) \le C'(\Gamma)^2 \Delta$

$|\text{supp}(U)| \le 3$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1})$,

   where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

$$\begin{aligned} P_e \gtrapprox \min_U \ & \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right] \\ \text{s.t.} \quad & \mathbb{E}[U] \geq \beta \\ & \mathrm{Var}(U) \leq C'(\Gamma)^2 \Delta \\ & |\mathrm{supp}(U)| \leq 3 \end{aligned}$$

Achievability Proof

1. Let $R = C(\Gamma) + \frac{\beta}{\sqrt{n}}$
2. Choose $P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1})$,

   where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

$$P_e \lessapprox \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right) \neq \mathbb{P}\left(\sum_{i=1}^{n} \log \frac{W(Y_i|X_i)}{PW(Y_i)} \leq nR\right)$$

$$P_e \gtrapprox \inf_P \sup_Q \mathbb{P}\left(\log \frac{W(Y^n|X^n)}{Q(Y^n)} \leq nR\right)$$

$$P_e \gtrapprox \min_U \mathbb{E}\left[\Phi\left(\frac{U}{\sqrt{V(\Gamma)}}\right)\right]$$

s.t. $\mathbb{E}[U] \geq \beta$

$\mathrm{Var}(U) \leq C'(\Gamma)^2 \Delta$

$|\mathrm{supp}(U)| \leq 3$

Mixture of Uniform Distributions on $(n-1)$ spheres

- Let $X^n \sim P$, where $$P = p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1}),$$ where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$

Mixture of Uniform Distributions on $(n-1)$ spheres

- Let $X^n \sim P$, where $$P = p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1}),$$ where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$
- Let $Y^n \sim PW$, where $PW = p_1 Q_{R_1}^{\text{unif}} + p_2 Q_{R_2}^{\text{unif}} + p_3 Q_{R_3}^{\text{unif}}$

Mixture of Uniform Distributions on $(n-1)$ spheres

- Let $X^n \sim P$, where $$P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1}),$$ where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$
- Let $Y^n \sim PW$, where $PW = p_1 Q_{R_1}^{\text{unif}} + p_2 Q_{R_2}^{\text{unif}} + p_3 Q_{R_3}^{\text{unif}}$
- Would prefer $Q_{R_i}^{\text{i.i.d.}} \longleftarrow$ induced output distribution when the input is $\mathcal{N}\left(\mathbf{0}, \frac{R_i}{\sqrt{n}} I_n\right)$

Mixture of Uniform Distributions on $(n-1)$ spheres

- Let $X^n \sim P$, where $$P = p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1}),$$ where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$
- Let $Y^n \sim PW$, where $PW = p_1 Q_{R_1}^{\text{unif}} + p_2 Q_{R_2}^{\text{unif}} + p_3 Q_{R_3}^{\text{unif}}$
- Would prefer $Q_{R_i}^{\text{i.i.d.}} \longleftarrow$ induced output distribution when the input is $\mathcal{N}\left(\mathbf{0}, \frac{R_i}{\sqrt{n}} I_n\right)$

Lemma 1 (Mahmood and Wagner '25)

Whether $Y^n \sim Q_{R_i}^{\text{unif}}$ or $Y^n \sim Q_{R_i}^{\text{i.i.d.}}$, for some constant $C > 0$,

$$\mathbb{P}\left(\left|\log \frac{Q_{R_i}^{\text{unif}}(Y^n)}{Q_{R_i}^{\text{i.i.d.}}(Y^n)}\right| \leq C \log n\right) \to 1 \text{ as } n \to \infty.$$

Mixture of Uniform Distributions on $(n-1)$ spheres

- Let $X^n \sim P$, where
$$P = p_1 \mathrm{Unif}(S_{R_1}^{n-1}) + p_2 \mathrm{Unif}(S_{R_2}^{n-1}) + p_3 \mathrm{Unif}(S_{R_3}^{n-1}),$$
where each $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$ for $i \neq j$
- Let $Y^n \sim PW$, where $PW = p_1 Q_{R_1}^{\mathrm{unif}} + p_2 Q_{R_2}^{\mathrm{unif}} + p_3 Q_{R_3}^{\mathrm{unif}}$
- Would prefer $Q_{R_i}^{\mathrm{i.i.d.}} \longleftarrow$ induced output distribution when the input is $\mathcal{N}\left(\mathbf{0}, \frac{R_i}{\sqrt{n}} I_n\right)$

Lemma 1 (Mahmood and Wagner '25)

Whether $Y^n \sim Q_{R_i}^{\mathrm{unif}}$ or $Y^n \sim Q_{R_i}^{\mathrm{i.i.d.}}$, for some constant $C > 0$,
$$\mathbb{P}\left(\left| \log \frac{Q_{R_i}^{\mathrm{unif}}(Y^n)}{Q_{R_i}^{\mathrm{i.i.d.}}(Y^n)} \right| \leq C \log n \right) \to 1 \text{ as } n \to \infty.$$

Lemma 2 (Mahmood and Wagner '25)

Whether $Y^n \sim Q_{R_i}^{\mathrm{unif}}$ or $Y^n \sim Q_{R_j}^{\mathrm{unif}}$, for some constant $C > 0$,
$$\mathbb{P}\left(\left| \log \frac{Q_{R_i}^{\mathrm{unif}}(Y^n)}{Q_{R_j}^{\mathrm{unif}}(Y^n)} \right| \leq C \log n \right) \to 1 \text{ as } n \to \infty.$$

Choosing High vs. Low Variance to Minimize $\Pr(X + F(X) \leq C)$

Decision freedom: After observing $X = x$, set $F(x)$ equal to either $Y \sim \mathcal{N}(1, 1)$ or $Z \sim \mathcal{N}(1, 2)$

Choosing High vs. Low Variance to Minimize $\Pr(X + F(X) \leq C)$

Decision freedom: After observing $X = x$, set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$

1. $$\min_F \Pr\left(X + F(X) \leq C\right)$$

Choosing High vs. Low Variance to Minimize $\Pr(X + F(X) \leq C)$

Decision freedom: After observing $X = x$, set $F(x)$ equal to either $Y \sim \mathcal{N}(1, 1)$ or $Z \sim \mathcal{N}(1, 2)$

1. $$\min_F \Pr\left(X + F(X) \leq C\right) = \int dx\, p(x) \min_F \Pr(F(x) \leq C - x)$$

Choosing High vs. Low Variance to Minimize $\Pr(X + F(X) \leq C)$

Decision freedom: After observing $X = x$, set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$

1. $$\min_F \Pr\left(X + F(X) \leq C\right) = \int dx\, p(x) \min_F \Pr(F(x) \leq C - x)$$

2. $$\min_F \Pr(F(x) \leq C - x) = \min\left\{F_Y(C - x), F_Z(C - x)\right\}$$

Choosing High vs. Low Variance to Minimize $\Pr(X + F(X) \leq C)$

Decision freedom: After observing $X = x$, set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$

1. $$\min_F \Pr\left(X + F(X) \leq C\right) = \int dx\, p(x) \min_F \Pr(F(x) \leq C - x)$$

2. $$\min_F \Pr(F(x) \leq C - x) = \min\left\{F_Y(C - x), F_Z(C - x)\right\}$$
$$= \begin{cases} F_Y(C - x) & \text{if } x \geq x_{\text{th}} \qquad \text{(low variance)} \\ F_Z(C - x) & \text{if } x < x_{\text{th}} \qquad \text{(high variance)} \end{cases}$$

Application to Feedback

- Minimize $\Pr(X + F(X) \leq C)$
- Set $F(x)$ equal to either $Y \sim \mathcal{N}(1, 1)$ or $Z \sim \mathcal{N}(1, 2)$
- $$\min_F \Pr(x + F(x) \leq C) = \begin{cases} F_Y(C - x) & \text{if } x \geq x_{\text{th}} \\ F_Z(C - x) & \text{if } x < x_{\text{th}} \end{cases}$$ (low variance) (high variance)

$$P_e \approx \Pr\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$

Application to Feedback

- Minimize $\Pr(X + F(X) \leq C)$
- Set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$
- $$\min_F \Pr(x + F(x) \leq C) = \begin{cases} F_Y(C - x) & \text{if } x \geq x_{\text{th}} \quad \text{(low variance)} \\ F_Z(C - x) & \text{if } x < x_{\text{th}} \quad \text{(high variance)} \end{cases}$$

$$\begin{aligned} P_e &\approx \Pr\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right) \\ &\approx \Pr\left(\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} + \log \frac{W(Y^{n/2+1:n}|X^{n/2+1:n})}{P_F W(Y^{n/2+1:n}|Y^{n/2})} \leq nR\right) \end{aligned}$$

- Let P_F be the feedback-adjusted distribution for the second half

Application to Feedback

- Minimize $\Pr(X + F(X) \leq C)$
- Set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$
- $$\min_F \Pr(x + F(x) \leq C) = \begin{cases} F_Y(C - x) & \text{if } x \geq x_{\text{th}} \quad \text{(low variance)} \\ F_Z(C - x) & \text{if } x < x_{\text{th}} \quad \text{(high variance)} \end{cases}$$

$$P_e \approx \Pr\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$
$$\approx \Pr\left(\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} + \log \frac{W(Y^{n/2+1:n}|X^{n/2+1:n})}{P_F W(Y^{n/2+1:n}|Y^{n/2})} \leq nR\right)$$

- Let P_F be the feedback-adjusted distribution for the second half
- If $\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} \leq$ threshold, then choose $P_F = p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1})$

Application to Feedback

- Minimize $\Pr(X + F(X) \leq C)$
- Set $F(x)$ equal to either $Y \sim \mathcal{N}(1,1)$ or $Z \sim \mathcal{N}(1,2)$
- $$\min_F \Pr(x + F(x) \leq C) = \begin{cases} F_Y(C-x) & \text{if } x \geq x_{\text{th}} \\ F_Z(C-x) & \text{if } x < x_{\text{th}} \end{cases} \qquad \begin{matrix} \text{(low variance)} \\ \text{(high variance)} \end{matrix}$$

$$P_e \approx \Pr\left(\log \frac{W(Y^n|X^n)}{PW(Y^n)} \leq nR\right)$$
$$\approx \Pr\left(\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} + \log \frac{W(Y^{n/2+1:n}|X^{n/2+1:n})}{P_F W(Y^{n/2+1:n}|Y^{n/2})} \leq nR\right)$$

- Let P_F be the feedback-adjusted distribution for the second half
- If $\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} \leq$ threshold, then choose $P_F = p_1 \text{Unif}(S_{R_1}^{n-1}) + p_2 \text{Unif}(S_{R_2}^{n-1}) + p_3 \text{Unif}(S_{R_3}^{n-1})$
- If $\log \frac{W(Y^{n/2}|X^{n/2})}{PW(Y^{n/2})} >$ threshold, then choose $P_F = \text{Unif}(S_R^{n-1})$

References

- A. Mahmood and A. B. Wagner, “Channel coding with mean and variance cost constraints,” IEEE Transactions on Information Theory, 2025
- A. Mahmood and A. B. Wagner, “Channel coding for Gaussian channels with mean and variance constraints,” 2025. [Online]. Available: arXiv
- A. Mahmood and A. B. Wagner, ”Improved Channel Coding Performance Through Cost Variability,” 2024. [Online]. Available: arXiv
- A. Mahmood and A. B. Wagner, “Timid/bold coding for channels with cost constraints,” in 2023 IEEE International Symposium on Information Theory (ISIT), 2023