Third-Order Asymptotics: Old and New

Vincent Y. F. Tan (National University of Singapore; NUS)

Joint works with M. Tomamichel, Y. Sakai and M. Kovačević

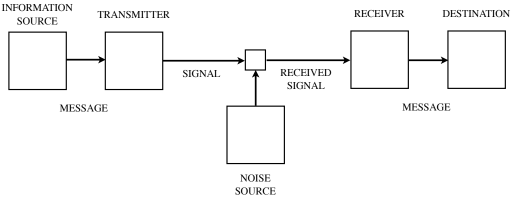
Workshop on Probability and Information Theory
(held at the University of Hong Kong, on 20th August 2019)

Outline

1. Introduction
2. Old Contribution
3. New Contribution

Outline

Introduction: Transmission of Information



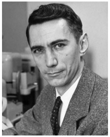
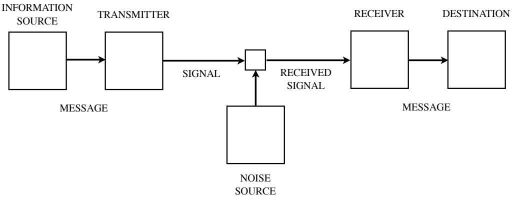


Figure: Shannon's Figure 1

- Information theory $\equiv$ Finding fundamental limits for reliable information transmission

Introduction: Transmission of Information



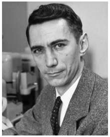
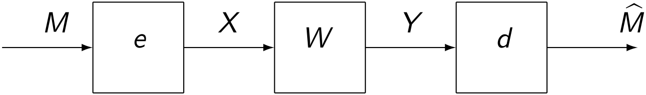


Figure: Shannon's Figure 1

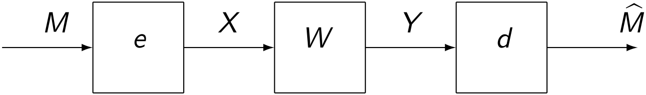
- Information theory $\equiv$ Finding fundamental limits for reliable information transmission
- Channel coding: Concerned with the maximum rate of communication in bits/channel use

Channel Coding (One-Shot)



- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set

Channel Coding (One-Shot)

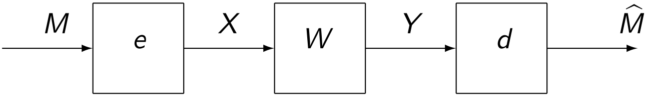


- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set
- The average error probability $p_{\rm err}(\mathcal{C})$ is

$$p_{\rm err}(\mathcal{C}) \ := \ \Pr\left[\widehat{M} \neq M\right]$$

where M is uniform on $\mathcal{M}$

Channel Coding (One-Shot)



- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set
- The average error probability $p_{\mathrm{err}}(\mathcal{C})$ is

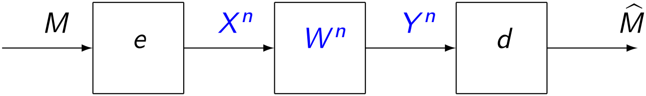
$$p_{\mathrm{err}}(\mathcal{C}) \;:=\; \Pr\left[\widehat{M} \neq M\right]$$

  where M is uniform on $\mathcal{M}$

- Maximum code size at ε-error is

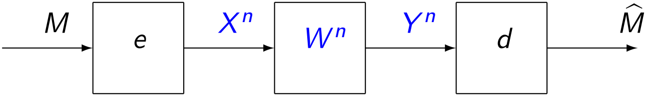
$$M^*(W, \varepsilon) := \sup\left\{m \,\middle|\, \exists \mathcal{C} \;\text{ s.t. }\; m = |\mathcal{M}|,\, p_{\mathrm{err}}(\mathcal{C}) \leq \varepsilon\right\}$$

Channel Coding (n-Shot)



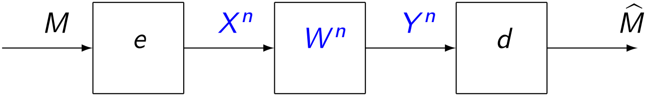
- Consider n independent uses of a channel

Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel

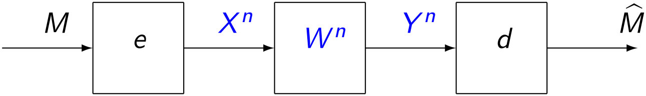
Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel
- For vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

Channel Coding (n-Shot)



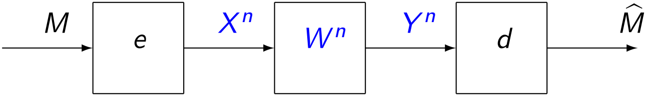
- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel
- For vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

- Maximum code size at average error ε and blocklength n is

$$M^*(W^n, \varepsilon)$$

Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel
- For vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

- Maximum code size at average error ε and blocklength n is

$$M^*(W^n, \varepsilon)$$

- Consider both discrete- and continuous-time channels.

Outline

Old Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion
- Going beyond the normal approx terms

M. Tomamichel

Old Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion
- Going beyond the normal approx terms

M. Tomamichel

Theorem (Tomamichel-Tan (2013))

For all DMCs with positive ε-dispersion V_ε,

$$\log M^*(W^n, \varepsilon) \le nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

where $\Phi(a) := \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x$

Old Contribution: Remarks

- Our bound

$$\log M^*(W^n, \varepsilon) \le nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

Old Contribution: Remarks

- Our bound

$$\log M^*(W^n, \varepsilon) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- Best upper bound till date:

$$\log M^*(W^n, \varepsilon) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \left(|\mathcal{X}| - \frac{1}{2}\right)\log n + O(1)$$

V. Strassen (1964)

Polyanskiy-Poor-Verdú (2010)

Old Contribution: Remarks

- Our bound

$$\log M^*(W^n, \varepsilon) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- Best upper bound till date:

$$\log M^*(W^n, \varepsilon) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \left(|\mathcal{X}| - \frac{1}{2}\right)\log n + O(1)$$

V. Strassen (1964)

Polyanskiy-Poor-Verdú (2010)

- Requires new converse techniques

Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) \right]$$

Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon)\right]$$

- $\rho_n = O(\log n)$ if channel is non-exotic

Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon)\right]$$

- $\rho_n = O(\log n)$ if channel is non-exotic
- ρ_n may be important at very short blocklengths

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV}\Phi^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV}\Phi^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- For the BEC [PPV10]

$$\rho_n = O(1)$$

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC + \sqrt{nV}\Phi^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- For the BEC [PPV10]

$$\rho_n = O(1)$$

- For the AWGN under maximum (or peak) power constraints [PPV10, Tan-Tomamichel (2015)]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ are positive and $C > 0$. Then,

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ are positive and $C > 0$. Then,

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

- This is an achievability result but BEC doesn't satisfy assumptions

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ are positive and $C > 0$. Then,

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

- This is an achievability result but BEC doesn't satisfy assumptions
- Assumption may be relaxed to

$$\exists\, P \in \Pi \quad \text{s.t.} \quad V^{\mathrm{r}}(P, W) := V\left(PW, \frac{P \times W}{PW}\right) > 0$$

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ are positive and $C > 0$. Then,

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

- This is an achievability result but BEC doesn't satisfy assumptions
- Assumption may be relaxed to

$$\exists\, P \in \Pi \quad \text{s.t.} \quad V^{\mathrm{r}}(P, W) := V\left(PW, \frac{P \times W}{PW}\right) > 0$$

- Based on the concentration bound [Polyanskiy's thesis]

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^{n} X_i\right) \mathbb{I}\left\{\sum_{i=1}^{n} X_i \geq \gamma\right\}\right] \leq 2\left(\frac{\log 2}{\sqrt{2\pi}} + \frac{12\,T}{\sigma}\right) \frac{\exp(-\gamma)}{\sigma\sqrt{n}}.$$

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2}\log n + O(1)$$

- This is a converse result

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric
- The set of weakly input-symmetric channels is very thin

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2}\log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric
- The set of weakly input-symmetric channels is very thin
- We dispense of this symmetry assumption

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved
- For BSC

$$\rho_n = \frac{1}{2} \log n + O(1)$$

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved
- For BSC
$$\rho_n = \frac{1}{2} \log n + O(1)$$
- We can dispense of the positive ε-dispersion assumption

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved
- For BSC
$$\rho_n = \frac{1}{2} \log n + O(1)$$
- We can dispense of the positive ε-dispersion assumption
- No need for unique CAID

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

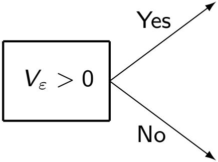
If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved
- For BSC
$$\rho_n = \frac{1}{2} \log n + O(1)$$
- We can dispense of the positive ε-dispersion assumption
- No need for unique CAID
- "A Tight Upper Bound for the Third-Order Asymptotics for Most DMCs" M. Tomamichel and V. Y. F. Tan, IEEE T-IT, Nov 2013

Main Result: Tight Third-Order Term

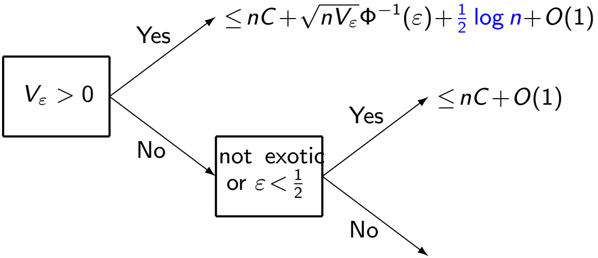
All cases are covered



$$\leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

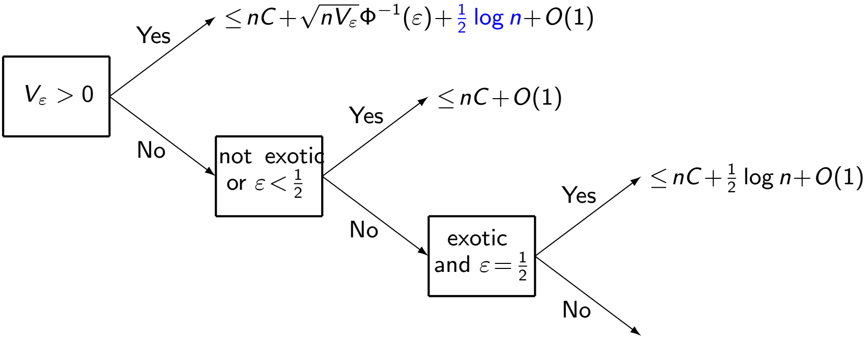
Main Result: Tight Third-Order Term

All cases are covered



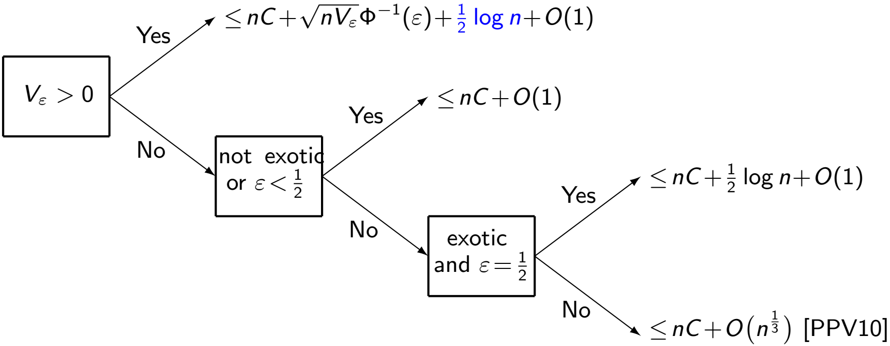
Main Result: Tight Third-Order Term

All cases are covered



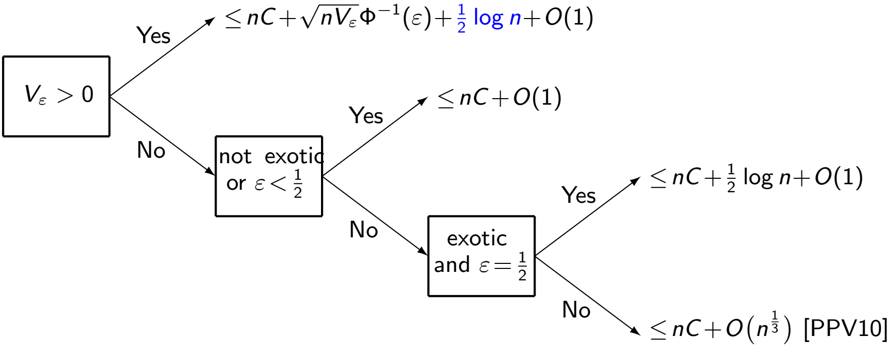
Main Result: Tight Third-Order Term

All cases are covered



Main Result: Tight Third-Order Term

All cases are covered



W is exotic if $V_{\max}(W) = 0$ and $\exists\, x_0 \in \mathcal{X}$ such that

$$D(W(\cdot|x_0)\|Q^*) = C, \quad \text{and} \quad V(W(\cdot|x_0)\|Q^*) > 0.$$

Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2}\log n + O(1)$

Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2} \log n + O(1)$
- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight

Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2} \log n + O(1)$

- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight

- We need a convenient converse bound for general DMCs

Proof Technique for Tight Third-Order Term

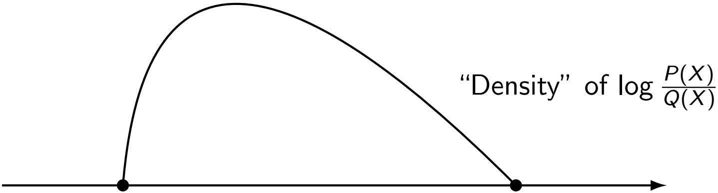
- For the regular case, $\rho_n \le \frac{1}{2}\log n + O(1)$
- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight
- We need a convenient converse bound for general DMCs
- Information spectrum divergence

$$D_s^{\varepsilon}(P\|Q) := \sup\left\{R \,:\, P\left(\log\frac{P(X)}{Q(X)} \le R\right) \le \varepsilon\right\}$$

"Information Spectrum Methods in Information Theory"
by T. S. Han (2003)

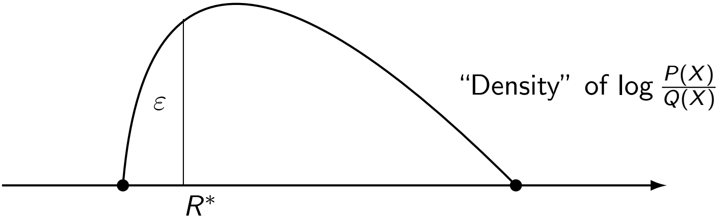
Proof Technique: Information Spectrum Divergence

$$D_s^{\varepsilon}(P\|Q) := \sup\left\{R \,\middle|\, P\left(\log\frac{P(X)}{Q(X)} \leq R\right) \leq \varepsilon\right\}$$



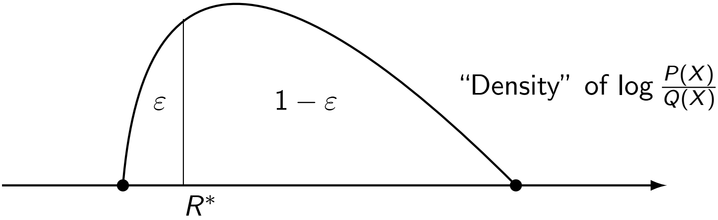
Proof Technique: Information Spectrum Divergence

$$D_s^\varepsilon(P\|Q) := \sup\left\{R \,\middle|\, P\left(\log\frac{P(X)}{Q(X)} \le R\right) \le \varepsilon\right\}$$



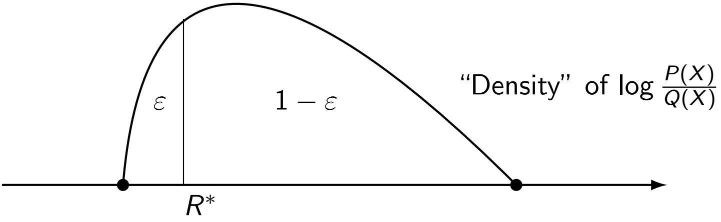
Proof Technique: Information Spectrum Divergence

$$D_s^{\varepsilon}(P\|Q) := \sup\left\{ R \,\middle|\, P\left(\log \frac{P(X)}{Q(X)} \le R\right) \le \varepsilon \right\}$$



Proof Technique: Information Spectrum Divergence

$$D_s^\varepsilon(P\|Q) := \sup\left\{R \,\middle|\, P\left(\log\frac{P(X)}{Q(X)} \le R\right) \le \varepsilon\right\}$$



If X^n is i.i.d. P, the Berry-Esseen theorem yields

$$D_s^\varepsilon(P^n\|Q^n) = nD(P\|Q) + \sqrt{nV(P\|Q)}\Phi^{-1}(\varepsilon) + O(1)$$

Proof Technique: Symbol-Wise Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

Proof Technique: Symbol-Wise Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \left(\max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) \right) + \log \frac{1}{\delta}$$

Proof Technique: Symbol-Wise Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \left(\max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) \right) + \log \frac{1}{\delta}$$

- Choose $\delta = n^{-\frac{1}{2}}$ so $\log \frac{1}{\delta} = \frac{1}{2} \log n$

Proof Technique: Symbol-Wise Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \left(\max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) \right) + \log \frac{1}{\delta}$$

- Choose $\delta = n^{-\frac{1}{2}}$ so $\log \frac{1}{\delta} = \frac{1}{2} \log n$

- Since all $\mathbf{x}$ within a type class result in the same $D_s^{\varepsilon+\delta}$ (if $Q^{(n)}$ is permutation invariant), it's really a max over types $P_{\mathbf{x}} \in \mathcal{P}_n(\mathcal{X})$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \leq \max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log \frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x})\|Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- First term: $Q_{\mathbf{k}}$'s and $\lambda(\mathbf{k})$'s designed to form an $n^{-\frac{1}{2}}$-cover of $\mathcal{P}(\mathcal{Y})$:

$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k} \in \mathcal{K} \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \le n^{-\frac{1}{2}}.$$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \leq \max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log \frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2} \sum_{\mathbf{k} \in \mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} (PW)^n(\mathbf{y})$$

- First term: $Q_{\mathbf{k}}$'s and $\lambda(\mathbf{k})$'s designed to form an $n^{-\frac{1}{2}}$-cover of $\mathcal{P}(\mathcal{Y})$:

$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k} \in \mathcal{K} \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \leq n^{-\frac{1}{2}}.$$

- Second term: Uniform mixture over output distributions induced by input types [Hayashi (2009)]

Proof Technique: Novel Choice of Output Distribution

- First term is

$$\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) \quad \text{where} \quad \lambda(\mathbf{k}) = \frac{\exp\left(-\gamma\|\mathbf{k}\|_2^2\right)}{F}$$

  and **k** indexes distance to the capacity-achieving output distribution (CAOD). Can be shown that $F < \infty$.

Proof Technique: Novel Choice of Output Distribution

- First term is

$$\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) \quad \text{where} \quad \lambda(\mathbf{k}) = \frac{\exp\left(-\gamma\|\mathbf{k}\|_2^2\right)}{F}$$

  and **k** indexes distance to the capacity-achieving output distribution (CAOD). Can be shown that $F < \infty$.

- Choose each $Q_{\mathbf{k}}$ as follows:

$$Q_{\mathbf{k}}(y) := Q^*(y) + \frac{k_y}{\sqrt{n\zeta}},$$

  where $\mathcal{K} := \left\{\mathbf{k} \in \mathbb{Z}^{|\mathcal{Y}|} : \sum_y k_y = 0, k_y \geq -Q^*(y)\sqrt{n\zeta}\right\}$

Proof Technique: Novel Choice of Output Distribution

- First term is

$$\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) \quad \text{where} \quad \lambda(\mathbf{k}) = \frac{\exp\big(-\gamma\|\mathbf{k}\|_2^2\big)}{F}$$

  and $\mathbf{k}$ indexes distance to the capacity-achieving output distribution (CAOD). Can be shown that $F < \infty$.

- Choose each $Q_{\mathbf{k}}$ as follows:

$$Q_{\mathbf{k}}(y) := Q^*(y) + \frac{k_y}{\sqrt{n\zeta}},$$

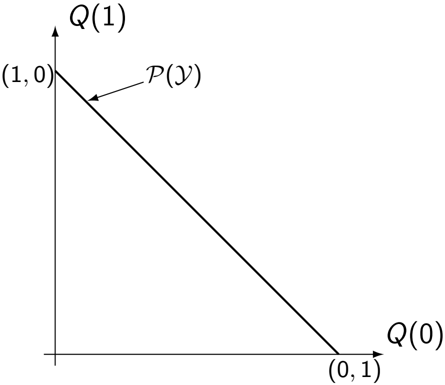
  where $\mathcal{K} := \big\{\mathbf{k} \in \mathbb{Z}^{|\mathcal{Y}|} : \sum_y k_y = 0, k_y \geq -Q^*(y)\sqrt{n\zeta}\big\}$

- By construction, ensures that

$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k} \in \mathcal{K}, \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \leq \frac{1}{\sqrt{n}}.$$

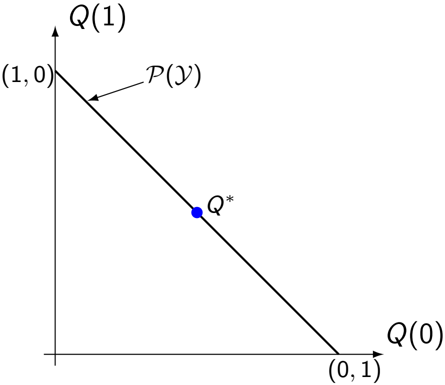
Proof Technique: Novel Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



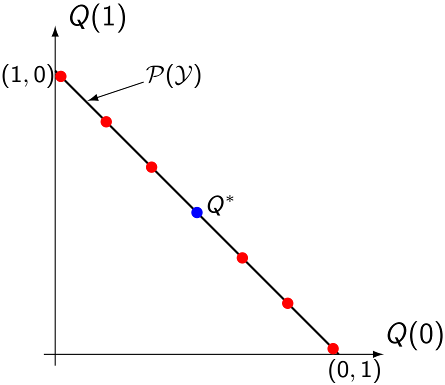
Proof Technique: Novel Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



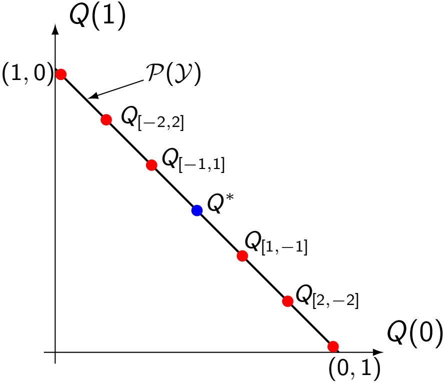
Proof Technique: Novel Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



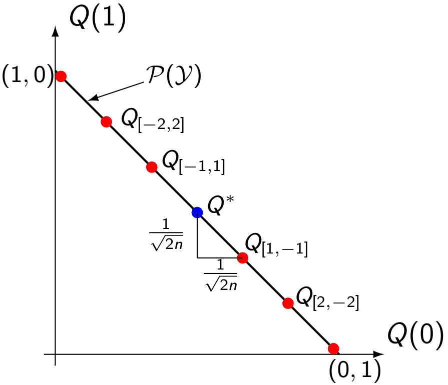
Proof Technique: Novel Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

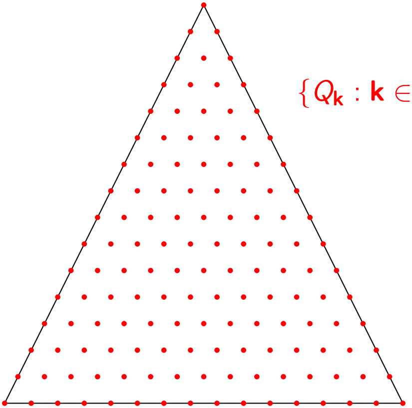


Proof Technique: Novel Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

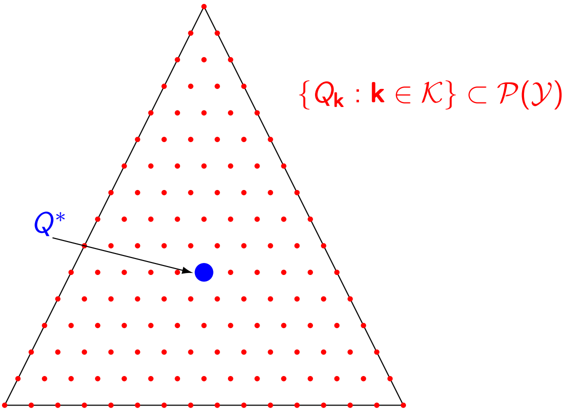


Proof Technique: Novel Choice of Output Distribution

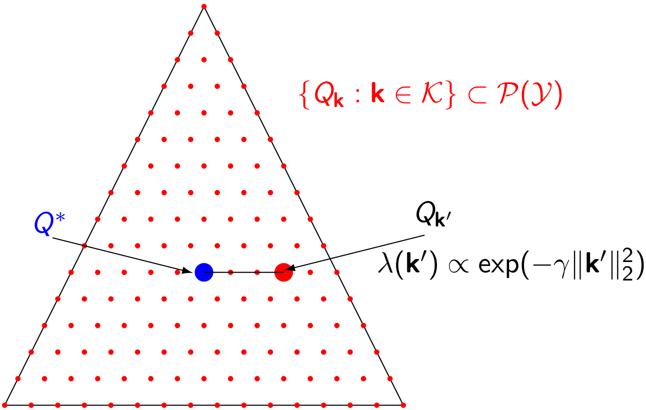


$\{Q_{\mathbf{k}} : \mathbf{k} \in \mathcal{K}\} \subset \mathcal{P}(\mathcal{Y})$

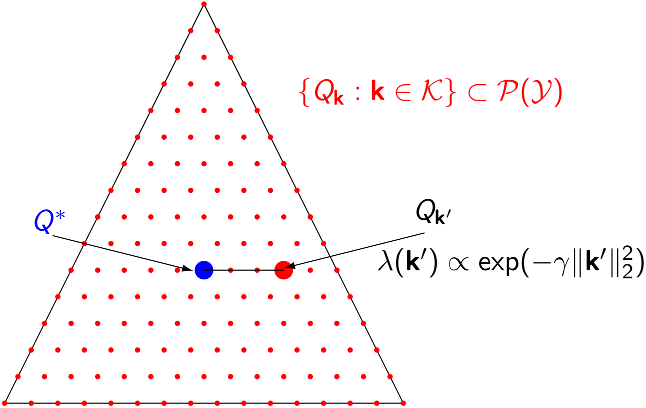
Proof Technique: Novel Choice of Output Distribution



Proof Technique: Novel Choice of Output Distribution



Proof Technique: Novel Choice of Output Distribution



$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k} \in \mathcal{K}, \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \leq \frac{1}{\sqrt{n}}.$$

Proof Technique: Standard Choice of Output Distn.

- Recall the output distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

Proof Technique: Standard Choice of Output Distn.

- Recall the output distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- Second term: Uniform mixture over output distributions induced by input types [Hayashi (2009)]

$$\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y}).$$

Proof Technique: Standard Choice of Output Distn.

- Recall the output distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- Second term: Uniform mixture over output distributions induced by input types [Hayashi (2009)]

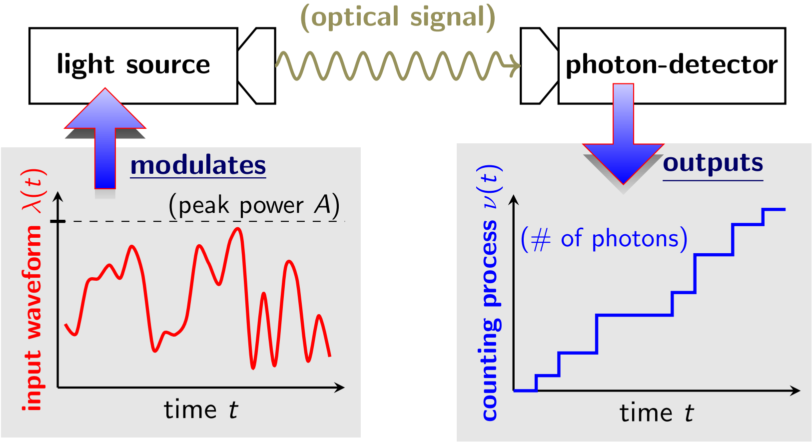
$$\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y}).$$

- Serves to take care of "bad input types" (i.e., types $P \in \mathcal{P}_n(\mathcal{X})$ such that PW is far from Q^*)

Outline

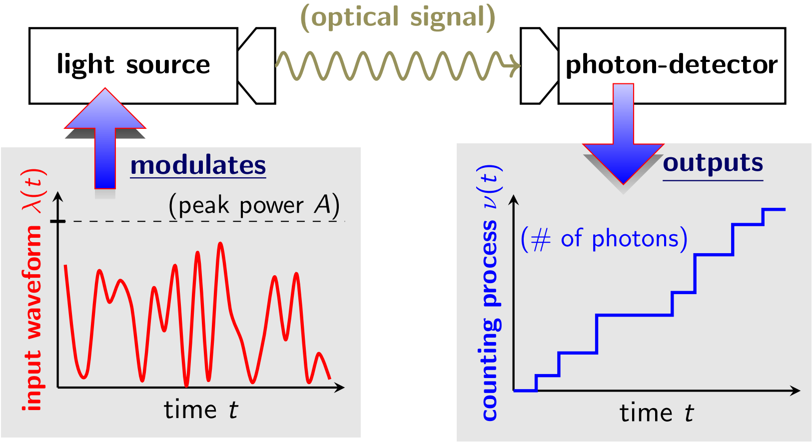
Mathematical Model of Poisson Channel (1/3)

Consider the following optical communication:



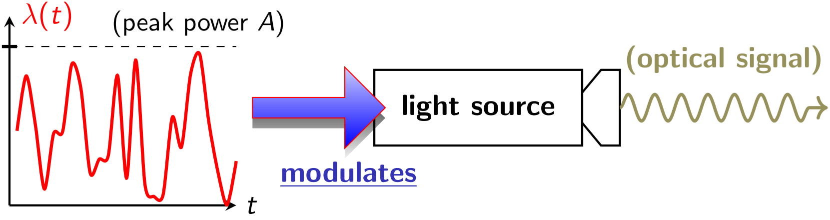
Mathematical Model of Poisson Channel (1/3)

Consider the following optical communication:

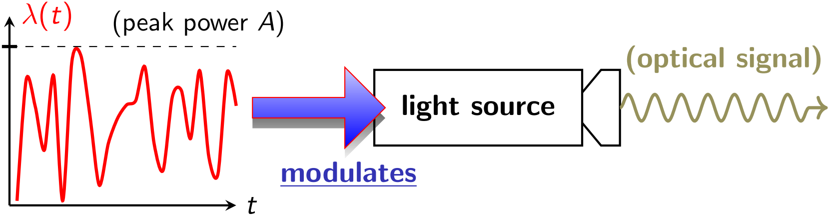


Remark: This is a **continuous-time** channel ($0 \le t < T$).

Mathematical Model of Poisson Channel (2/3)



Mathematical Model of Poisson Channel (2/3)



Optical Signal is Modulated by Input Waveform $\lambda(t)$

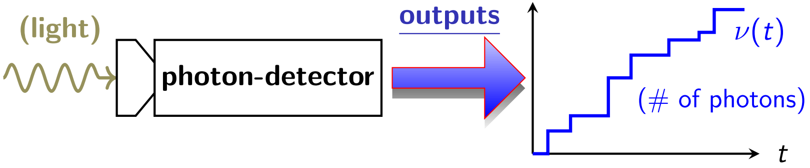
- an integrable function $\lambda(\cdot)$ defined on the time block $[0, T)$;
- with **peak power constraint** **(**$A > 0$**)**:

$$0 \leq \lambda(t) \leq A \qquad \forall t \in [0, T);$$

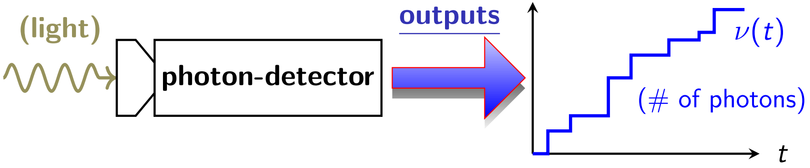
- with **average power constraint** **(**$0 \leq \sigma \leq 1$**)**:

$$\frac{1}{T} \int_0^T \lambda(t)\,\mathrm{d}t \leq \sigma A.$$

Mathematical Model of Poisson Channel (3/3)



Mathematical Model of Poisson Channel (3/3)



Output is Poisson counting process $\{\nu(t)\}_{0 \le t < T}$

$$\nu(0) = 0 \quad \text{a.s.} \qquad \text{and} \qquad \mathbb{P}\{\nu(t+\tau) - \nu(t) = k\} = \frac{e^{\Lambda}\,\Lambda^k}{k!}$$

for each $t, \tau \in \mathbb{R}_{\ge 0}$ and $k \in \mathbb{Z}_{\ge 0}$, where Λ is given by

$$\Lambda \overset{\text{def}}{=} \int_t^{t+\tau} \big(\lambda(u) + \lambda_0 \big) \, \mathrm{d}u.$$

- **input waveform** (intensity of light) $\lambda : [0, T) \to [0, A]$
- **dark current** (background noise level) $0 \le \lambda_0 < \infty$

Block Coding Scheme for Poisson Channel

- **input alphabet** is the set of waveforms $\lambda(\cdot)$

$$\mathcal{W}(T, A, \sigma) \stackrel{\text{def}}{=} \left\{ \lambda : [0, T) \to [0, A] \;\middle|\; \frac{1}{T} \int_0^T \lambda(t)\, \mathrm{d}t \leq \sigma A \right\},$$

where A (resp. σ) is the peak (resp. average) power constraint.

Block Coding Scheme for Poisson Channel

- **input alphabet** is the set of waveforms $\lambda(\cdot)$

$$\mathcal{W}(T, A, \sigma) \stackrel{\text{def}}{=} \left\{ \lambda : [0, T) \to [0, A] \;\middle|\; \frac{1}{T} \int_0^T \lambda(t)\,\mathrm{d}t \le \sigma A \right\},$$

where A (resp. σ) is the peak (resp. average) power constraint.

- **output alphabet** is the set of possible counting processes $\nu(\cdot)$

$$\mathcal{S}(T) \stackrel{\text{def}}{=} \{ g : [0, T) \to \mathbb{Z}_{\ge 0} \mid g(0) = 0 \text{ and } g(t_1) \ge g(t_2), t_1 < t_2 \}$$

Block Coding Scheme for Poisson Channel

- **input alphabet** is the set of waveforms $\lambda(\cdot)$

$$\mathcal{W}(T, A, \sigma) \stackrel{\text{def}}{=} \left\{ \lambda : [0, T) \to [0, A] \;\middle|\; \frac{1}{T} \int_0^T \lambda(t) \, \mathrm{d}t \le \sigma A \right\},$$

where A (resp. σ) is the peak (resp. average) power constraint.

- **output alphabet** is the set of possible counting processes $\nu(\cdot)$

$$\mathcal{S}(T) \stackrel{\text{def}}{=} \{ g : [0, T) \to \mathbb{Z}_{\ge 0} \mid g(0) = 0 \text{ and } g(t_1) \ge g(t_2), t_1 < t_2 \}$$

A (T, M, A, σ)-code (ϕ, ψ) for Poisson channel

- encoder $\phi : \{1, 2, \ldots, M\} \to \mathcal{W}(T, A, \sigma)$
- decoder $\psi : \mathcal{S}(T) \to \{1, 2, \ldots, M\}$

Block Coding Scheme for Poisson Channel (Cont'd)

Block Coding Scheme for Poisson Channel (Cont'd)

A (T, M, A, σ)-code (ϕ, ψ) for Poisson channel

- encoder $\phi : \{1, 2, \ldots, M\} \to \mathcal{W}(T, A, \sigma)$
- decoder $\psi : \mathcal{S}(T) \to \{1, 2, \ldots, M\}$

Block Coding Scheme for Poisson Channel (Cont'd)

A (T, M, A, σ)-code (ϕ, ψ) for Poisson channel

- encoder $\phi : \{1, 2, \ldots, M\} \to \mathcal{W}(T, A, \sigma)$
- decoder $\psi : \mathcal{S}(T) \to \{1, 2, \ldots, M\}$

A $(T, M, A, \sigma, \varepsilon)_{\text{avg}}$-code (ϕ, ψ) for Poisson channel

A (T, M, A, σ)-code (ϕ, ψ) is called a $(T, M, A, \sigma, \varepsilon)_{\text{avg}}$-code if

$$\frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\{\psi(\nu) = m \mid \lambda = \phi(m)\} \geq 1 - \varepsilon.$$

Here, λ is the r.v. induced by the encoder ϕ with uniform messages.

Poisson Channel Capacity (1st-Order Asymptotics)

Denote by M^* the max. M s.t. $\exists$ a $(T, M, A, \sigma, \varepsilon)_{\text{avg}}$-code.

Theorem (Kabanov'78; Davis'80; Wyner'88)

$$\log M^* = T\,C^* + \mathrm{o}(T) \qquad (\text{as } T \to \infty),$$

where

$$\begin{cases} C^* \overset{\text{def}}{=} A\left((1-p^*)\,s\log\dfrac{s}{p^*+s} + p^*\,(1+s)\log\dfrac{1+s}{p^*+s}\right), \\ s \overset{\text{def}}{=} \dfrac{\lambda_0}{A} \qquad (\text{ratio of } \textbf{dark current } \lambda_0 \text{ to } \textbf{PPC } A), \\ p^* \overset{\text{def}}{=} \min\{\sigma, p_0\} \qquad (\text{role of } \textbf{CAID}, \text{ where } \sigma \text{ is } \textbf{APC}), \\ p_0 \overset{\text{def}}{=} \dfrac{(1+s)^{1+s}}{s^s\,\mathrm{e}} - s. \end{cases}$$

Poisson Channel Dispersion (2nd-Order Asymptotics)

- Denote by M^* the max s.t. $\exists$ a $(T, M, A, \sigma, \varepsilon)_{\mathsf{avg}}$-code.

Poisson Channel Dispersion (2nd-Order Asymptotics)

- Denote by M^* the max s.t. $\exists$ a $(T, M, A, \sigma, \varepsilon)_{\mathsf{avg}}$-code.
- We seek second- and third-order terms

$$\log M^* = T\,C^* + \sqrt{T}\,L + \rho_T, \qquad T \to \infty.$$

Poisson Channel Dispersion (2nd-Order Asymptotics)

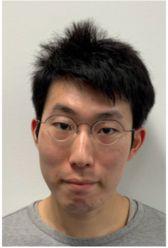
- Denote by M^* the max s.t. $\exists$ a $(T, M, A, \sigma, \varepsilon)_{\mathsf{avg}}$-code.
- We seek second- and third-order terms
$$\log M^* = T\, C^* + \sqrt{T}\, L + \rho_T, \qquad T \to \infty.$$
- Many works since 2013 on multi-terminal channels and sources
- First work on higher-order asymptotics for continuous-time channels

Poisson Channel Dispersion (2nd-Order Asymptotics)

- Denote by M^* the max s.t. $\exists$ a $(T, M, A, \sigma, \varepsilon)_{\mathrm{avg}}$-code.
- We seek second- and third-order terms

$$\log M^* = T\,C^* + \sqrt{T}\,L + \rho_T, \qquad T \to \infty.$$

- Many works since 2013 on multi-terminal channels and sources
- First work on higher-order asymptotics for continuous-time channels



Yuta Sakai

Mladen Kovačević

Poisson Channel Dispersion (2nd-Order Asymptotics)

Theorem (Sakai–Tan–Kovačević'19: `arXiv:1903.10438`)

$$\log M^* = T\,C^* + \sqrt{T\,V^*}\,\Phi^{-1}(\varepsilon) + \rho_T,$$

where the Poisson channel dispersion V^* is given by

$$V^* \overset{\text{def}}{=} A\left((1-p^*)\,s\log^2\frac{s}{p^*+s} + p^*\,(1+s)\log^2\frac{1+s}{p^*+s}\right),$$

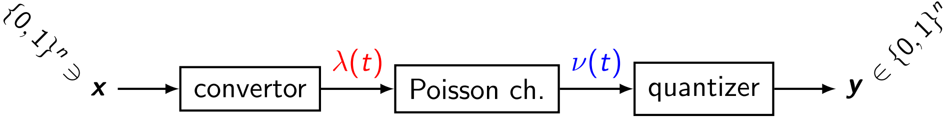
and the third-order term ρ_T satisfies

$$\frac{1}{2}\log T + \mathrm{O}(1) \le \rho_T \le \log T + \mathrm{O}(1) \qquad (\text{as } T \to \infty).$$

Result: 2nd-order term $\sqrt{V^*}\,\Phi^{-1}(\varepsilon)$ and bounds on 3rd-order term ρ_T

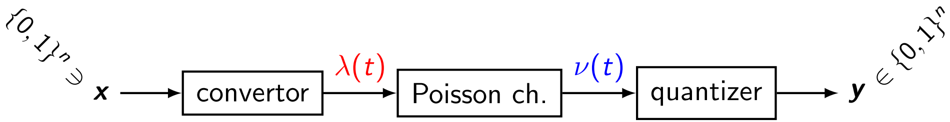
Proof Ideas of Second- and Third-Order Asymptotics

In both converse and achievability parts, we shall employ **Wyner's discretization argument** (Wyner'88):



Proof Ideas of Second- and Third-Order Asymptotics

In both converse and achievability parts, we shall employ **Wyner's discretization argument** (Wyner'88):



Converse Part

- symbol-wise meta converse bound (Tomamichel–Tan'13)
- **novel choice of output distribution** (projected ϵ-net)

Proof Ideas of Second- and Third-Order Asymptotics

In both converse and achievability parts, we shall employ **Wyner's discretization argument** (Wyner'88):

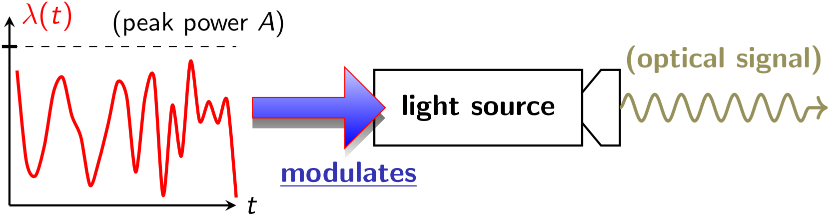
Converse Part

- symbol-wise meta converse bound (Tomamichel–Tan'13)
- **novel choice of output distribution** (projected ϵ-net)

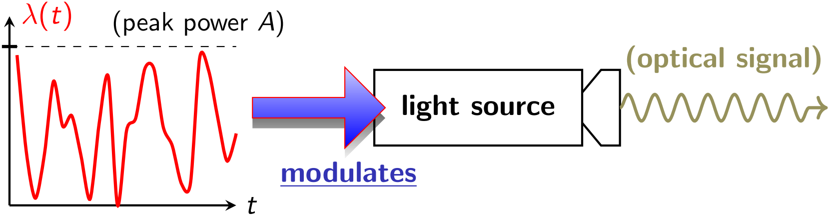
Achievability Part

- random coding union bound (PPV'10) with cost constraint
- some other techniques to handle the continuous nature (e.g., logarithmic Sobolev inequality)

Wyner's Discretization Part I: Input Restriction

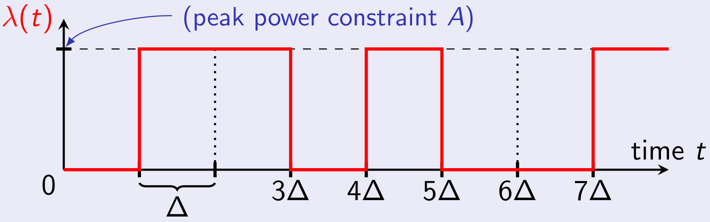


Wyner's Discretization Part I: Input Restriction



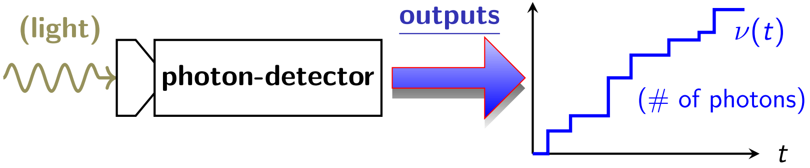
Discretization of $\{\lambda(t)\}_{0\le t<T}$ into n Blocks (here, $\Delta = T/n$)

input waveform $\lambda(t)$ is restricted to be **square**, e.g.,

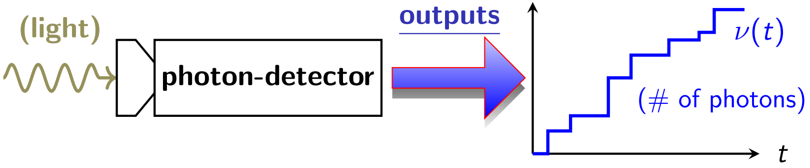


That is, we may think of $\lambda(t)$ as **a binary sequence** $\{x_k\}_{k=1}^{n}$.

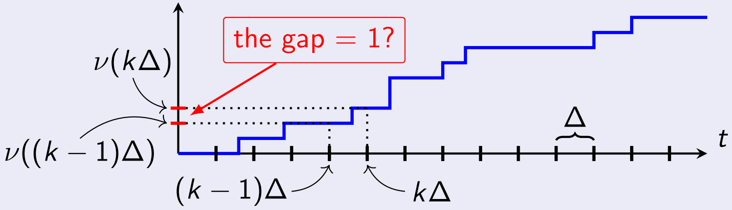
Wyner's Discretization Part II: Output Quantization



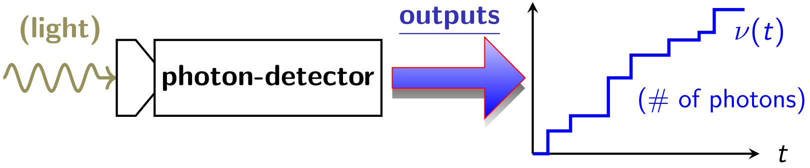
Wyner's Discretization Part II: Output Quantization



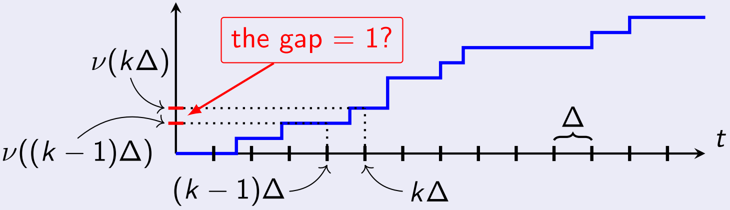
Discretization of $\{\nu(t)\}_{0\leq t<T}$ into n Blocks (here, $\Delta = T/n$)



Wyner's Discretization Part II: Output Quantization



Discretization of $\{\nu(t)\}_{0\le t<T}$ into n Blocks (here, $\Delta = T/n$)



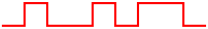
Poisson counting process $\nu(t)$ is quantized as $\{y_k\}_{k=1}^n$:

$$y_k \overset{\text{def}}{=} \begin{cases} 0 & \text{if } \nu(k\Delta) - \nu((k-1)\Delta) \neq 1, \\ 1 & \text{if } \nu(k\Delta) - \nu((k-1)\Delta) = 1. \end{cases}$$

Overall Diagram of Wyner's Discretization

- input sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$
  (which is converted to a square wave $\lambda(t)$: )
- output sequence $\boldsymbol{y} = (y_1, \ldots, y_n) \in \{0, 1\}^n$
  (which is obtained by quantizing the counting process $\nu(t)$)

Overall Diagram of Wyner's Discretization

- input sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$
  (which is converted to a square wave $\lambda(t)$: )
- output sequence $\boldsymbol{y} = (y_1, \ldots, y_n) \in \{0,1\}^n$
  (which is obtained by quantizing the counting process $\nu(t)$)

Discretized channel $W_n^n : \{0,1\}^n \to \{0,1\}^n$

$$W_n^n(\boldsymbol{y} \mid \boldsymbol{x}) \stackrel{\text{def}}{=} \prod_{i=1}^{n} W_n(y_i \mid x_i),$$

where the single-letter channel $W_n : \{0,1\} \to \{0,1\}$ **depends** on n.

Remark: the discretization error is negligible as $n \to \infty$ (next slide).

Wyner's Discretization Well-Approximates Poisson Channel

Denote by

- $M^*_{\text{Poisson}}(\varepsilon)$: fundamental limit of Poisson channel
- $M^*(W_n^n, \varepsilon)$: fundamental limit of discretized channel W_n^n

Lemma (Wyner'88)

There exist a sequence $\epsilon_n = \mathrm{o}(1)$ and a subsequence $\{n_k\}_{k=1}^{\infty}$ s.t.

$$M^*_{\text{Poisson}}(\varepsilon) = M^*(W_{n_k}^{n_k}, \varepsilon + \epsilon_{n_k}) \qquad (\forall k \geq 1).$$

Wyner's Discretization Well-Approximates Poisson Channel

Denote by

- $M^*_{\text{Poisson}}(\varepsilon)$: fundamental limit of Poisson channel
- $M^*(W_n^n, \varepsilon)$: fundamental limit of discretized channel W_n^n

Lemma (Wyner'88)

There exist a sequence $\epsilon_n = \mathrm{o}(1)$ and a subsequence $\{n_k\}_{k=1}^{\infty}$ s.t.

$$M^*_{\text{Poisson}}(\varepsilon) = M^*(W_{n_k}^{n_k}, \varepsilon + \epsilon_{n_k}) \qquad (\forall k \geq 1).$$

Therefore, we observe that

$$\log M^*_{\text{Poisson}}(\varepsilon) \leq \limsup_{n \to \infty} \log M^*(W_n^n, \varepsilon + \epsilon_n),$$

implying that it suffices to examine **the RHS** in the converse part.

Meta Converse Bound and Output Distribution

Apply the symbol-wise meta converse (Tomamichel–Tan'13):

$$\log M^*(W_n^n, \varepsilon + \epsilon_n) \leq \max_{\boldsymbol{x} \in \{0,1\}^n} D_{\mathrm{s}}^{\varepsilon+\epsilon_n+\eta}(\underbrace{W_n^n(\cdot \mid \boldsymbol{x})}_{\text{discretized Poisson channel}} \| Q^{(n)}) + \log \frac{1}{\eta}$$

Meta Converse Bound and Output Distribution

Apply the symbol-wise meta converse (Tomamichel–Tan'13):

$$\log M^*(W_n^n, \varepsilon + \epsilon_n) \leq \max_{\boldsymbol{x} \in \{0,1\}^n} D_{\mathrm{s}}^{\varepsilon+\epsilon_n+\eta}(\underbrace{W_n^n(\cdot \mid \boldsymbol{x})}_{\text{discretized Poisson channel}} \,\|\, Q^{(n)}) + \log \frac{1}{\eta}$$

Since $Q^{(n)} \in \mathcal{P}(\{0,1\}^n)$ is arbitrary, we substitute

$$Q^{(n)}(\boldsymbol{y}) = \frac{1}{3} \prod_{i=1}^{n} P^*_{[-\kappa]} W_n(y_i) + \frac{1}{3} \prod_{i=1}^{n} P^*_{[\kappa]} W_n(y_i) + \frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \leq p^* + m/T \leq 1}}^{\infty} \mathrm{e}^{-\gamma m^2/T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)$$

where $\kappa = \frac{1}{2} \min\{\sigma, 1/\mathrm{e}\} > 0$ and $P^*_{[u]}(1) = p^* + u$.

Meta Converse Bound and Output Distribution

Apply the symbol-wise meta converse (Tomamichel–Tan'13):

$$\log M^*(W_n^n, \varepsilon + \epsilon_n) \le \max_{\boldsymbol{x} \in \{0,1\}^n} D_s^{\varepsilon+\epsilon_n+\eta}\big(\underbrace{W_n^n(\cdot \mid \boldsymbol{x})}_{\text{discretized Poisson channel}} \,\|\, Q^{(n)} \big) + \log \frac{1}{\eta}$$

Since $Q^{(n)} \in \mathcal{P}(\{0,1\}^n)$ is arbitrary, we substitute

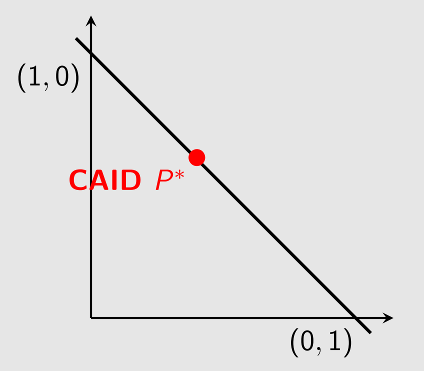
$$Q^{(n)}(\boldsymbol{y}) = \frac{1}{3} \prod_{i=1}^{n} P^*_{[-\kappa]} W_n(y_i) + \frac{1}{3} \prod_{i=1}^{n} P^*_{[\kappa]} W_n(y_i) + \frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \le p^* + m/T \le 1}}^{\infty} \mathrm{e}^{-\gamma m^2 / T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)$$

where $\kappa = \frac{1}{2} \min\{\sigma, 1/\mathrm{e}\} > 0$ and $P^*_{[u]}(1) = p^* + u$.

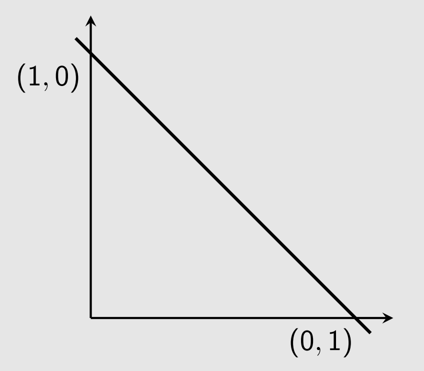
- **third term** is the main part of **our novel construction**
- **first** and **second** terms are to apply Lipschitz properties

ϵ-Net Argument: Tomamichel–Tan's Original Choice

Consider a binary-input binary-output channel $W : \{0,1\} \to \{0,1\}$.



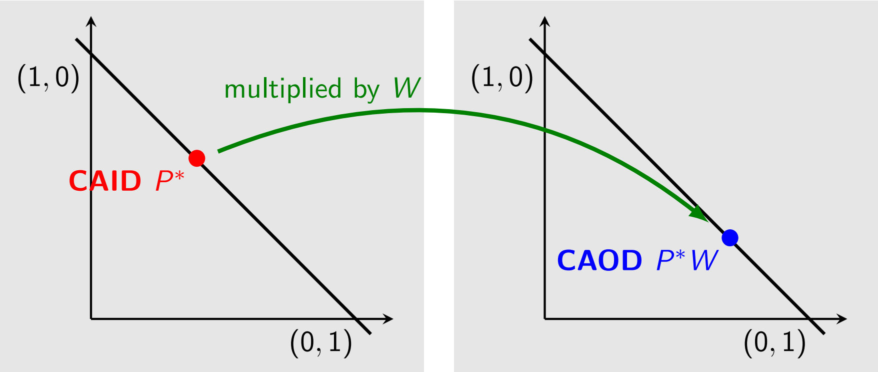
[input probab. simplex]



[output probab. simplex]

ϵ-Net Argument: Tomamichel–Tan's Original Choice

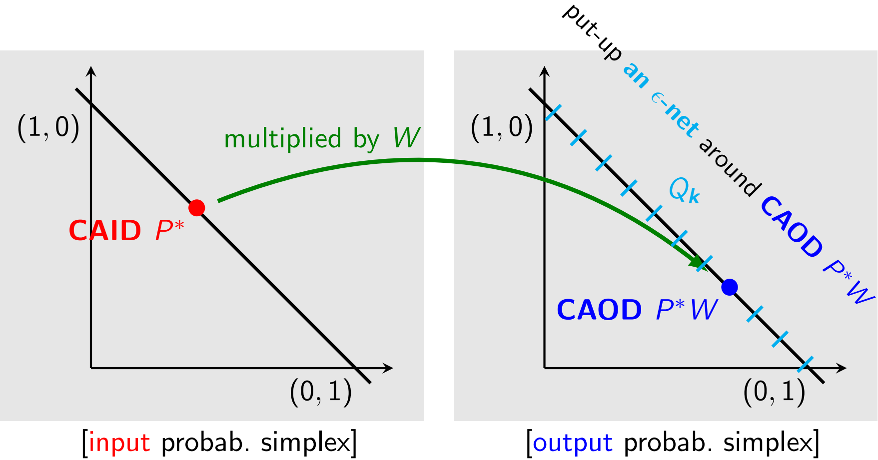
Consider a binary-input binary-output channel $W : \{0,1\} \to \{0,1\}$.



[input probab. simplex] [output probab. simplex]

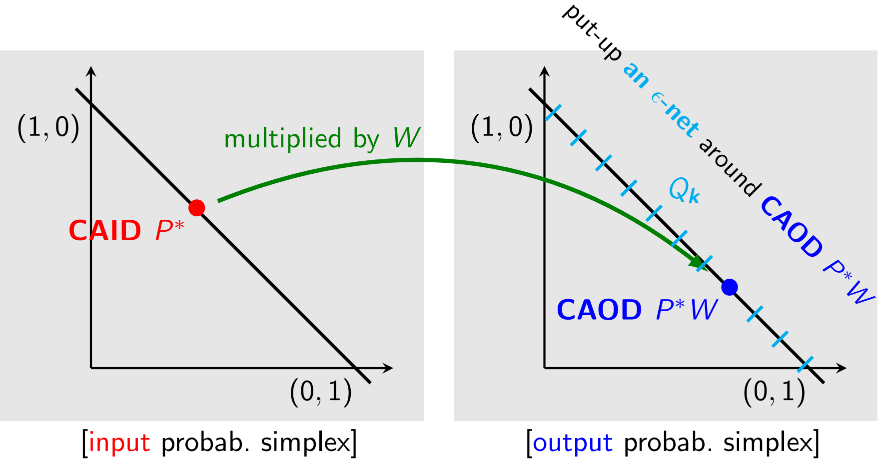
ϵ-Net Argument: Tomamichel–Tan's Original Choice

Consider a binary-input binary-output channel $W : \{0,1\} \to \{0,1\}$.



ϵ-Net Argument: Tomamichel–Tan's Original Choice

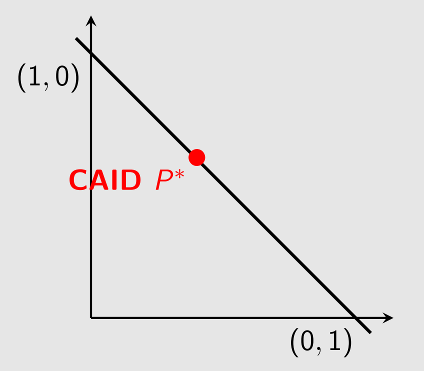
Consider a binary-input binary-output channel $W : \{0,1\} \to \{0,1\}$.



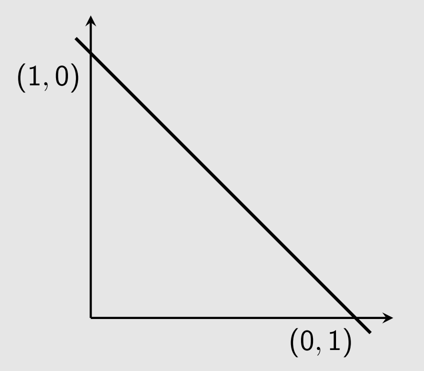
Use a convex combination of **the ϵ-net**: $\sum_{\boldsymbol{k}} \mu(\boldsymbol{k}) Q_{\boldsymbol{k}}^n$.

ϵ-Net Argument: For Discretized Poisson Channels

Consider a (single-letter) discretized channel $W_n : \{0,1\} \to \{0,1\}$.



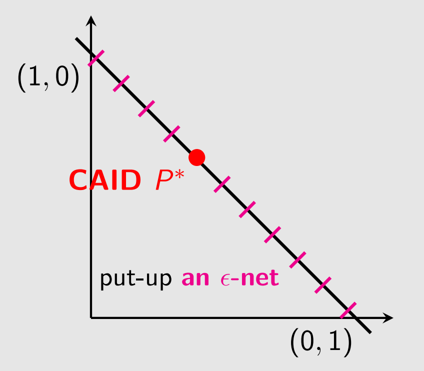
[input probab. simplex]



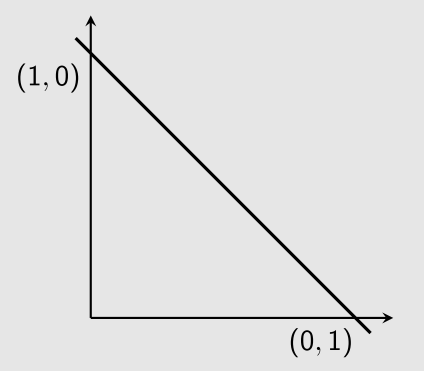
[output probab. simplex]

ϵ-Net Argument: For Discretized Poisson Channels

Consider a (single-letter) discretized channel $W_n : \{0,1\} \to \{0,1\}$.



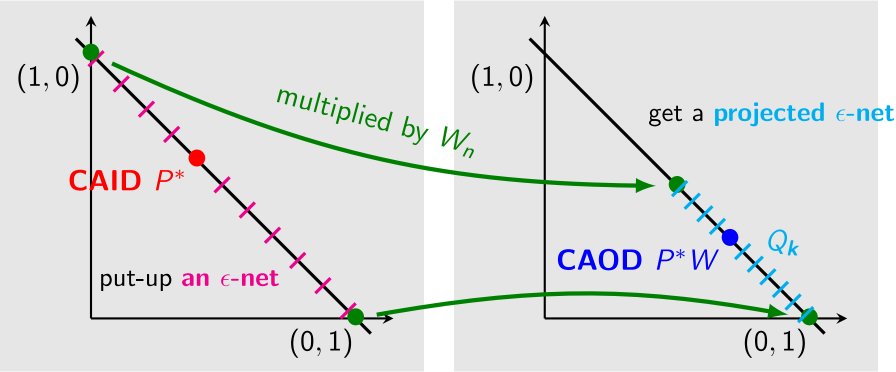
[input probab. simplex]



[output probab. simplex]

ϵ-Net Argument: For Discretized Poisson Channels

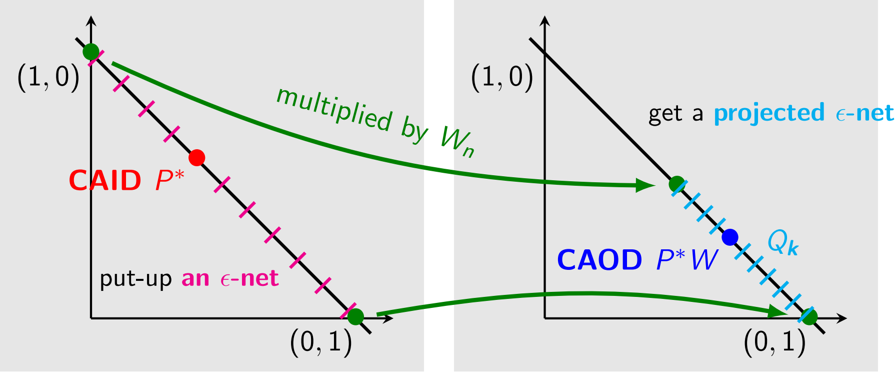
Consider a (single-letter) discretized channel $W_n : \{0,1\} \to \{0,1\}$.



[input probab. simplex] [output probab. simplex]

ϵ-Net Argument: For Discretized Poisson Channels

Consider a (single-letter) discretized channel $W_n : \{0,1\} \to \{0,1\}$.



[input probab. simplex] [output probab. simplex]

Use a convex combination of **the projected ϵ-net**: $\sum_{\boldsymbol{k}} \mu(\boldsymbol{k}) Q_{\boldsymbol{k}}^n$

Why This Choice of Output Distn. and not TT13?

- Recall that we chose

$$\text{Third Term of } Q^{(n)}(\boldsymbol{y}) = \frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \le p^* + m/T \le 1}}^{\infty} \mathrm{e}^{-\gamma m^2/T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)$$

Why This Choice of Output Distn. and not TT13?

- Recall that we chose

$$\text{Third Term of } Q^{(n)}(\boldsymbol{y}) = \frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \le p^* + m/T \le 1}}^{\infty} \mathrm{e}^{-\gamma m^2/T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)$$

- Need to control normalization constant F.
- By the sifting property of D_{s}, appears as $\log F$ bound on $\log M^*$.

Why This Choice of Output Distn. and not TT13?

- Recall that we chose

$$\text{Third Term of } Q^{(n)}(\boldsymbol{y}) = \frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \le p^* + m/T \le 1}}^{\infty} \mathrm{e}^{-\gamma m^2/T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)$$

- Need to control normalization constant F.
- By the sifting property of D_{s}, appears as $\log F$ bound on $\log M^*$.
- By direct calculation

$$F < \sum_{m=-\infty}^{\infty} \mathrm{e}^{-\gamma m^2/T} < 1 + \int_{-\infty}^{\infty} \mathrm{e}^{-\gamma m^2/T} \,\mathrm{d}m = 1 + \sqrt{\frac{\pi T}{\gamma}} = \mathrm{O}(\sqrt{T})$$

Why This Choice of Output Distn. and not TT13?

- Recall that we chose

$$\text{Third Term of } Q^{(n)}(\boldsymbol{y}) = \boxed{\frac{1}{3F} \sum_{\substack{m=-\infty: \\ 0 \le p^* + m/T \le 1}}^{\infty} \mathrm{e}^{-\gamma m^2/T} \prod_{i=1}^{n} P^*_{[m/T]} W_n(y_i)}$$

- Need to control normalization constant F.
- By the sifting property of D_{s}, appears as $\log F$ bound on $\log M^*$.
- By direct calculation

$$F < \sum_{m=-\infty}^{\infty} \mathrm{e}^{-\gamma m^2/T} < 1 + \int_{-\infty}^{\infty} \mathrm{e}^{-\gamma m^2/T} \, \mathrm{d}m = 1 + \sqrt{\frac{\pi T}{\gamma}} = \mathrm{O}(\sqrt{T})$$

- Tomamichel-Tan's construction in the output distn. space cannot handle the non-stationary W_n^n.

Concluding Remarks

- Full understanding of third-order asymptotics for DMCs

Concluding Remarks

- Full understanding of third-order asymptotics for DMCs
- Second- and third-order asymptotics for the Poisson channel

$$\log M^*(T, A\sigma, \varepsilon) = T\, C^* + \sqrt{T\, V^*}\, \Phi^{-1}(\varepsilon) + \rho_T,$$

where

$$\frac{1}{2}\log T + \mathrm{O}(1) \le \rho_T \le \log T + \mathrm{O}(1)$$

Concluding Remarks

- Full understanding of third-order asymptotics for DMCs
- Second- and third-order asymptotics for the Poisson channel

$$\log M^*(T, A\sigma, \varepsilon) = T\,C^* + \sqrt{T\,V^*}\,\Phi^{-1}(\varepsilon) + \rho_T,$$

  where

$$\frac{1}{2}\log T + \mathrm{O}(1) \le \rho_T \le \log T + \mathrm{O}(1)$$

- Different choices of output distributions.

Concluding Remarks

- Full understanding of third-order asymptotics for DMCs
- Second- and third-order asymptotics for the Poisson channel

$$\log M^*(T, A\sigma, \varepsilon) = T\,C^* + \sqrt{T\,V^*}\,\Phi^{-1}(\varepsilon) + \rho_T,$$

  where

$$\frac{1}{2}\log T + \mathrm{O}(1) \leq \rho_T \leq \log T + \mathrm{O}(1)$$

- Different choices of output distributions.
- Check out `arXiv:1903.10438`.