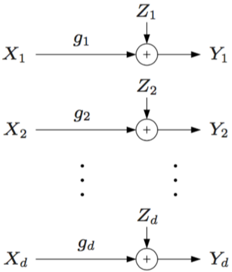
A Tight Upper Bound on the Second-Order Coding Rate of Parallel Gaussian Channels with Feedback

Vincent Y. F. Tan (NUS)

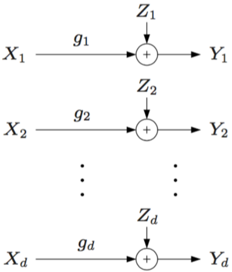
Joint work with Silas L. Fong (Toronto)

2017 Information Theory Workshop, Kaohsiung, Taiwan

Channel Model: The Parallel Gaussian Channel



Channel Model: The Parallel Gaussian Channel



Channel Law ($g_1 = g_2 = \ldots = g_d = 1$)

$$\mathbf{Y}_k = \mathbf{X}_k + \mathbf{Z}_k, \quad \begin{bmatrix} Y_{1,k} \\ Y_{2,k} \\ \vdots \\ Y_{d,k} \end{bmatrix} = \begin{bmatrix} X_{1,k} \\ X_{2,k} \\ \vdots \\ X_{d,k} \end{bmatrix} + \begin{bmatrix} Z_{1,k} \\ Z_{2,k} \\ \vdots \\ Z_{d,k} \end{bmatrix}, \quad k \in [1:n]$$

and $Z_{l,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, N_l)$ where $l \in [1:d]$.

Capacity of the Parallel Gaussian Channel

- Consider a peak power constraint

$$\Pr\left\{\frac{1}{n}\sum_{l=1}^{d}\sum_{k=1}^{n}X_{l,k}^2 \le P\right\} = 1$$

Capacity of the Parallel Gaussian Channel

- Consider a peak power constraint

$$\Pr\left\{\frac{1}{n}\sum_{l=1}^{d}\sum_{k=1}^{n}X_{l,k}^2 \le P\right\} = 1$$

- Define the capacity functions

$$\mathsf{C}(\mathbf{s}) = \sum_{l=1}^{d}\mathsf{C}\left(\frac{s_l}{N_l}\right), \quad \text{where} \quad \mathsf{C}(P) := \frac{1}{2}\log(1+P).$$

Capacity of the Parallel Gaussian Channel

- Consider a peak power constraint

$$\Pr\left\{\frac{1}{n}\sum_{l=1}^{d}\sum_{k=1}^{n}X_{l,k}^2 \le P\right\} = 1$$

- Define the capacity functions

$$\mathsf{C}(\mathbf{s}) = \sum_{l=1}^{d}\mathsf{C}\left(\frac{s_l}{N_l}\right), \quad \text{where} \quad \mathsf{C}(P) := \frac{1}{2}\log(1+P).$$

- "Vanishing-error" capacity is given by the water-filling solution

$$\mathsf{C}(\mathbf{P}^*), \quad \text{where} \quad \mathbf{P}^* = (P_1^*, P_2^*, \ldots, P_d^*)$$

and

$$P_l^* = (\lambda - N_l)^+ \quad \text{and } \lambda > 0 \text{ satisfies} \quad \sum_{l=1}^{d} P_l^* = P.$$

Capacity vs. Non-Asymptotic Fundamental Limits

- Define the non-asymptotic fundamental limit

$$M^*(n, \varepsilon, P) := \max\{M \in \mathbb{N} : \exists\, (n, M, \varepsilon, P)\text{-code}\}.$$

Maximum number of messages that can be supported by a length-n code with peak power P and average error prob. ε.

Capacity vs. Non-Asymptotic Fundamental Limits

- Define the non-asymptotic fundamental limit

$$M^*(n,\varepsilon,P) := \max\{M \in \mathbb{N} : \exists\,(n,M,\varepsilon,P)\text{-code}\}.$$

Maximum number of messages that can be supported by a length-n code with peak power P and average error prob. ε.

- Capacity result can be stated as

$$\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log M^*(n,\varepsilon,P) = \mathsf{C}(\mathbf{P}^*).$$

Capacity vs. Non-Asymptotic Fundamental Limits

- Define the non-asymptotic fundamental limit

$$M^*(n,\varepsilon,P) := \max\{M\in\mathbb{N} : \exists\, (n,M,\varepsilon,P)\text{-code}\}.$$

  Maximum number of messages that can be supported by a length-n code with peak power P and average error prob. ε.

- Capacity result can be stated as

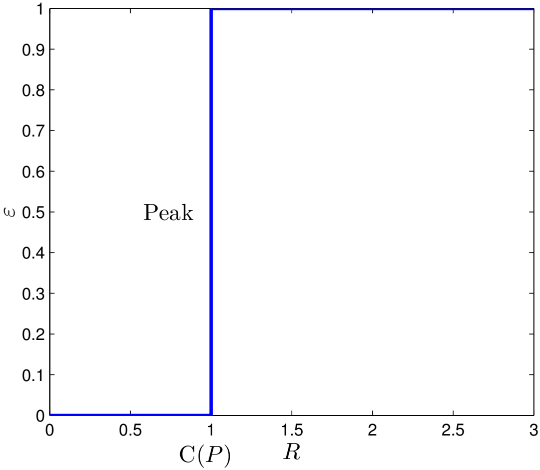
$$\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log M^*(n,\varepsilon,P) = \mathsf{C}(\mathbf{P}^*).$$

- In fact, the strong converse holds, and we have

$$\liminf_{n\to\infty}\frac{1}{n}\log M^*(n,\varepsilon,P) = \mathsf{C}(\mathbf{P}^*),\quad \forall\,\varepsilon\in(0,1).$$

Strong Converse

$$\varepsilon = \limsup_{n\to\infty} \mathsf{P}_\mathsf{e}^{(n)}, \qquad R = \liminf_{n\to\infty} \frac{1}{n} \log M^*(n, \varepsilon, P)$$



Second-Order Asymptotics

- Let
$$\mathsf{V}(P) := \frac{P(P+2)}{2(P+1)^2}$$
be the dispersion of the point-to-point AWGN channel.

Second-Order Asymptotics

- Let
$$\mathsf{V}(P) := \frac{P(P+2)}{2(P+1)^2}$$
be the dispersion of the point-to-point AWGN channel.
- Define the sum dispersion function to be
$$\mathsf{V}(\mathbf{s}) := \sum_{l=1}^{d} \mathsf{V}\left(\frac{s_l}{N_l}\right).$$

Second-Order Asymptotics

- Let

$$\mathsf{V}(P) := \frac{P(P+2)}{2(P+1)^2}$$

be the dispersion of the point-to-point AWGN channel.

- Define the sum dispersion function to be

$$\mathsf{V}(\mathbf{s}) := \sum_{l=1}^{d} \mathsf{V}\left(\frac{s_l}{N_l}\right).$$

- Then Polyanskiy (2010) and Tomamichel-T. (2015) showed that

$$\log M^*(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + O(\log n).$$

where $\mathbf{P}^* = (P_1^*, P_2^*, \ldots, P_d^*)$ is the optimal power allocation and

$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$$

Feedback

- If feedback is present, then the encoder at time k has access to message W and previous channel outputs $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{k-1})$, i.e.,

$$\mathbf{X}_k = \mathbf{X}_k(W, \mathbf{Y}^{k-1}), \quad \forall\, k \in [1:n].$$

Feedback

- If feedback is present, then the encoder at time k has access to message W and previous channel outputs $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{k-1})$, i.e.,

$$\mathbf{X}_k = \mathbf{X}_k(W, \mathbf{Y}^{k-1}), \quad \forall\, k \in [1:n].$$

- Define the non-asymptotic fundamental limit

$$M^*_{\mathsf{FB}}(n, \varepsilon, P) := \max\{M \in \mathbb{N} : \exists\, (n, M, \varepsilon, P)\text{-feedback code}\}.$$

Feedback

- If feedback is present, then the encoder at time k has access to message W and previous channel outputs $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{k-1})$, i.e.,

$$\mathbf{X}_k = \mathbf{X}_k(W, \mathbf{Y}^{k-1}), \quad \forall\, k \in [1:n].$$

- Define the non-asymptotic fundamental limit

$$M^*_{\mathsf{FB}}(n, \varepsilon, P) := \max\{M \in \mathbb{N} : \exists\,(n, M, \varepsilon, P)\text{-feedback code}\}.$$

- Since feedback does not increase capacity of point-to-point memoryless channels [Shannon (1956)],

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \to \infty} \frac{1}{n} \log M^*_{\mathsf{FB}}(n, \varepsilon, P) = \mathsf{C}(\mathbf{P}^*).$$

Feedback

- If feedback is present, then the encoder at time k has access to message W and previous channel outputs $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{k-1})$, i.e.,

$$\mathbf{X}_k = \mathbf{X}_k(W, \mathbf{Y}^{k-1}), \quad \forall\, k \in [1:n].$$

- Define the non-asymptotic fundamental limit

$$M^*_{\mathsf{FB}}(n, \varepsilon, P) := \max\{M \in \mathbb{N} : \exists\,(n, M, \varepsilon, P)\text{-feedback code}\}.$$

- Since feedback does not increase capacity of point-to-point memoryless channels [Shannon (1956)],

$$\lim_{\varepsilon \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log M^*_{\mathsf{FB}}(n, \varepsilon, P) = \mathsf{C}(\mathbf{P}^*).$$

- Natural question: What happens to the second-order terms?

Main Result

Theorem (Fong-T. (2017))

Feedback does not affect the second-order term for parallel Gaussian channels with feedback, i.e.,

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

Main Result

Theorem (Fong-T. (2017))

Feedback does not affect the second-order term for parallel Gaussian channels with feedback, i.e.,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Recall that without feedback

$$\log M^*(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + O(\log n).$$

Main Result

Theorem (Fong-T. (2017))

Feedback does not affect the second-order term for parallel Gaussian channels with feedback, i.e.,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Recall that without feedback

$$\log M^*(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + O(\log n).$$

- No guarantees on third-order term.

Main Result

Theorem (Fong-T. (2017))

Feedback does not affect the second-order term for parallel Gaussian channels with feedback, i.e.,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Recall that without feedback

$$\log M^*(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + O(\log n).$$

- No guarantees on third-order term.
- For achievability, encoder can ignore feedback.

Main Result

Theorem (Fong-T. (2017))

Feedback does not affect the second-order term for parallel Gaussian channels with feedback, i.e.,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Recall that without feedback

$$\log M^*(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + O(\log n).$$

- No guarantees on third-order term.
- For achievability, encoder can ignore feedback.
- Only need to prove converse part.

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p_{Y|X}^n(Y^n|x^n)}{p_{Y^*}^n(Y^n)}$$

has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p^n_{Y|X}(Y^n|x^n)}{p^n_{Y^*}(Y^n)} = -\frac{\|Y^n - x^n\|^2}{2N} + c$$

has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p^n_{Y|X}(Y^n|x^n)}{p^n_{Y^*}(Y^n)} = -\frac{\|Y^n - x^n\|^2}{2N} + c$$
$$= \frac{\langle x^n + Z^n, x^n \rangle}{N} + c$$

has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p^n_{Y|X}(Y^n|x^n)}{p^n_{Y^*}(Y^n)} = -\frac{\|Y^n - x^n\|^2}{2N} + c$$
$$= \frac{\langle x^n + Z^n, x^n\rangle}{N} + c$$
$$= \frac{1}{N}\sum_{k=1}^{n} Z_k x_k + c$$

has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p^n_{Y|X}(Y^n|x^n)}{p^n_{Y^*}(Y^n)} = -\frac{\|Y^n - x^n\|^2}{2N} + c$$
$$= \frac{\langle x^n + Z^n, x^n \rangle}{N} + c$$
$$= \frac{1}{N}\sum_{k=1}^{n} Z_k x_k + c$$

  has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

- This spherical symmetry argument [Polyanskiy (2011, 2013) and Fong-T. (2015)] yields

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) = n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1}(\varepsilon) + O(\log n).$$

Comparison to AWGN channel with feedback

- For the AWGN channel with feedback ($d = 1$), information density

$$\log \frac{p^n_{Y|X}(Y^n|x^n)}{p^n_{Y^*}(Y^n)} = -\frac{\|Y^n - x^n\|^2}{2N} + c$$
$$= \frac{\langle x^n + Z^n, x^n \rangle}{N} + c$$
$$= \frac{1}{N}\sum_{k=1}^{n} Z_k x_k + c$$

  has the same distribution for all x^n such that $\|x^n\|^2 = nP$.

- This spherical symmetry argument [Polyanskiy (2011, 2013) and Fong-T. (2015)] yields

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) = n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1}(\varepsilon) + O(\log n).$$

- Symmetry argument doesn't work for parallel Gaussians.

Elements of Converse Proof

Want to show

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathsf{FB}}(n,\varepsilon,P) - n\mathsf{C}(\mathbf{P}^*)}{\sqrt{n}} \le \sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon).$$

Elements of Converse Proof

Want to show

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathsf{FB}}(n,\varepsilon,P) - n\mathsf{C}(\mathbf{P}^*)}{\sqrt{n}} \le \sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon).$$

- Discretizing the power allocation: Turn power allocations into power types

Elements of Converse Proof

Want to show

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathsf{FB}}(n,\varepsilon,P) - n\mathsf{C}(\mathbf{P}^*)}{\sqrt{n}} \le \sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon).$$

- Discretizing the power allocation: Turn power allocations into power types
- Apply the meta-converse [Polyanskiy-Poor-Verdú (2010)]

Elements of Converse Proof

Want to show

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathsf{FB}}(n,\varepsilon,P) - n\mathsf{C}(\mathbf{P}^*)}{\sqrt{n}} \le \sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon).$$

- Discretizing the power allocation: Turn power allocations into power types
- Apply the meta-converse [Polyanskiy-Poor-Verdú (2010)]
- Apply Curtiss' theorem to the information spectrum term for close-to-optimal power types

Elements of Converse Proof

Want to show

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathsf{FB}}(n,\varepsilon,P) - n\mathsf{C}(\mathbf{P}^*)}{\sqrt{n}} \le \sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon).$$

- Discretizing the power allocation: Turn power allocations into power types
- Apply the meta-converse [Polyanskiy-Poor-Verdú (2010)]
- Apply Curtiss' theorem to the information spectrum term for close-to-optimal power types
- Apply large deviations bounds to far-from-optimal power types

Turn Power Allocation into Types: Part I

- Given a channel input $\mathbf{x}^n \in \mathbb{R}^{d\times n}$, define its power type to be

$$\begin{aligned}\phi(\mathbf{x}^n) &= \frac{1}{n}\left[\|x_1^n\|^2, \|x_2^n\|^2, \ldots, \|x_d^n\|^2\right] \\ &= \frac{1}{n}\left[\sum_{k=1}^n x_{1,k}^2, \sum_{k=1}^n x_{2,k}^2, \ldots, \sum_{k=1}^n x_{d,k}^2\right] \in \mathbb{R}_+^d.\end{aligned}$$

Turn Power Allocation into Types: Part I

- Given a channel input $\mathbf{x}^n \in \mathbb{R}^{d\times n}$, define its power type to be

$$\phi(\mathbf{x}^n) = \frac{1}{n}\left[\|x_1^n\|^2, \|x_2^n\|^2, \ldots, \|x_d^n\|^2\right]$$
$$= \frac{1}{n}\left[\sum_{k=1}^n x_{1,k}^2, \sum_{k=1}^n x_{2,k}^2, \ldots, \sum_{k=1}^n x_{d,k}^2\right] \in \mathbb{R}_+^d.$$

- Create a new code such that almost surely,

$$\sum_{k=1}^n X_{l,k}^2 = m_l P, \quad \text{for some } m_l \in [1:n] \text{ and } \forall\, l \in [1:d]$$

and

$$\sum_{l=1}^d \sum_{k=1}^n X_{l,k}^2 = nP, \quad \text{i.e.,} \quad \sum_{l=1}^d m_l = n.$$

Turn Power Allocation into Types: Part II

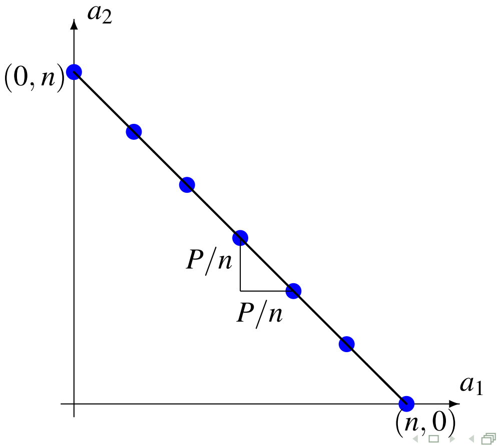
Power allocation vector set

$$\mathcal{S}^{(n)} := \left\{ \frac{P}{n} \cdot \mathbf{a} \,\middle|\, \mathbf{a} = (a_1, a_2, \ldots, a_d) \in \mathbb{Z}_+^d, \sum_{l=1}^{d} a_l = n \right\}$$

Turn Power Allocation into Types: Part II

Power allocation vector set

$$\mathcal{S}^{(n)} := \left\{ \frac{P}{n} \cdot \mathbf{a} \,\middle|\, \mathbf{a} = (a_1, a_2, \ldots, a_d) \in \mathbb{Z}_+^d, \sum_{l=1}^{d} a_l = n \right\}$$



Turn Power Allocation into Types: Part III

- Set of quantized power allocation vectors $\mathbf{s}$ with quantization level P/n that satisfy the equality power constraint

$$\sum_{l=1}^{d} s_l = P.$$

Turn Power Allocation into Types: Part III

- Set of quantized power allocation vectors $\mathbf{s}$ with quantization level P/n that satisfy the equality power constraint

$$\sum_{l=1}^{d} s_l = P.$$

- With the above construction, almost surely,

$$\phi(\mathbf{X}^n) \in \mathcal{S}^{(n)}.$$

Turn Power Allocation into Types: Part III

- Set of quantized power allocation vectors $\mathbf{s}$ with quantization level P/n that satisfy the equality power constraint

$$\sum_{l=1}^{d} s_l = P.$$

- With the above construction, almost surely,

$$\phi(\mathbf{X}^n) \in \mathcal{S}^{(n)}.$$

- So we can pretend that the codewords are discrete and leverage some existing theory from second-order analysis for DMCs.

Application of the Meta-Converse (PPV'10)

- By a relaxation of the meta-converse, for any $\xi > 0$,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) \le \log\xi - \log\left(1-\varepsilon-\Pr\left\{\log\frac{p^n_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}^n|\mathbf{X}^n)}{q_{\mathbf{Y}^n}(\mathbf{Y}^n)} \ge \log\xi\right\}\right)$$

Application of the Meta-Converse (PPV'10)

- By a relaxation of the meta-converse, for any $\xi > 0$,

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) \le \log \xi - \log \left(1 - \varepsilon - \Pr \left\{ \log \frac{p^n_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}^n|\mathbf{X}^n)}{q_{\mathbf{Y}^n}(\mathbf{Y}^n)} \ge \log \xi \right\} \right)$$

- Choose auxiliary output distribution $q_{\mathbf{Y}^n}$ carefully.

Application of the Meta-Converse (PPV'10)

- By a relaxation of the meta-converse, for any $\xi > 0$,

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) \le \log \xi - \log \left(1 - \varepsilon - \Pr\left\{ \log \frac{p^n_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}^n|\mathbf{X}^n)}{q_{\mathbf{Y}^n}(\mathbf{Y}^n)} \ge \log \xi \right\}\right)$$

- Choose auxiliary output distribution $q_{\mathbf{Y}^n}$ carefully.
- Inspired by Hayashi (2009), we choose

$$q_{\mathbf{Y}^n}(\mathbf{y}^n) = \frac{1}{2} \cdot \underbrace{\frac{1}{|\mathcal{S}^{(n)}|} \sum_{\mathbf{s} \in \mathcal{S}^{(n)}} \prod_{l,k} \mathcal{N}(y_{l,k}; 0, s_l + N_l)}_{q^{(1)}_{\mathbf{Y}^n}(\mathbf{y}^n)} + \frac{1}{2} \cdot \underbrace{\prod_{l,k} \mathcal{N}(y_{l,k}; 0, P_l + N_l)}_{q^{(2)}_{\mathbf{Y}^n}(\mathbf{y}^n)}$$

Application of the Meta-Converse (PPV'10)

- By a relaxation of the meta-converse, for any $\xi > 0$,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) \le \log\xi - \log\left(1-\varepsilon-\Pr\left\{\log\frac{p^n_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}^n|\mathbf{X}^n)}{q_{\mathbf{Y}^n}(\mathbf{Y}^n)} \ge \log\xi\right\}\right)$$

- Choose auxiliary output distribution $q_{\mathbf{Y}^n}$ carefully.
- Inspired by Hayashi (2009), we choose

$$q_{\mathbf{Y}^n}(\mathbf{y}^n) = \frac{1}{2}\cdot\underbrace{\frac{1}{|\mathcal{S}^{(n)}|}\sum_{\mathbf{s}\in\mathcal{S}^{(n)}}\prod_{l,k}\mathcal{N}(y_{l,k};0,s_l+N_l)}_{q^{(1)}_{\mathbf{Y}^n}(\mathbf{y}^n)} + \frac{1}{2}\cdot\underbrace{\prod_{l,k}\mathcal{N}(y_{l,k};0,P_l+N_l)}_{q^{(2)}_{\mathbf{Y}^n}(\mathbf{y}^n)}$$

- $q^{(1)}_{\mathbf{Y}^n}(\mathbf{y}^n)$: Unif. mixture of output dist. induced by input power types

Application of the Meta-Converse (PPV'10)

- By a relaxation of the meta-converse, for any $\xi > 0$,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) \le \log\xi - \log\left(1-\varepsilon-\Pr\left\{\log\frac{p^n_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}^n|\mathbf{X}^n)}{q_{\mathbf{Y}^n}(\mathbf{Y}^n)} \ge \log\xi\right\}\right)$$

- Choose auxiliary output distribution $q_{\mathbf{Y}^n}$ carefully.
- Inspired by Hayashi (2009), we choose

$$q_{\mathbf{Y}^n}(\mathbf{y}^n) = \frac{1}{2}\cdot\underbrace{\frac{1}{|\mathcal{S}^{(n)}|}\sum_{\mathbf{s}\in\mathcal{S}^{(n)}}\prod_{l,k}\mathcal{N}(y_{l,k};0,s_l+N_l)}_{q^{(1)}_{\mathbf{Y}^n}(\mathbf{y}^n)} + \frac{1}{2}\cdot\underbrace{\prod_{l,k}\mathcal{N}(y_{l,k};0,P_l+N_l)}_{q^{(2)}_{\mathbf{Y}^n}(\mathbf{y}^n)}$$

- $q^{(1)}_{\mathbf{Y}^n}(\mathbf{y}^n)$: Unif. mixture of output dist. induced by input power types
- $q^{(2)}_{\mathbf{Y}^n}(\mathbf{y}^n)$: Capacity-achieving output dist.

Almost-Optimal Types: Part I

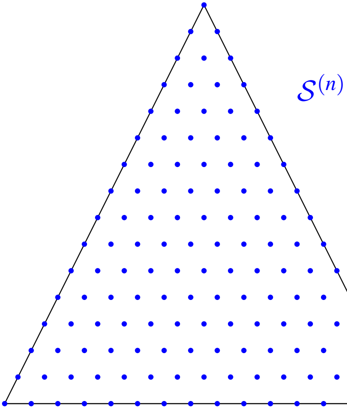
Define the set of almost-optimal power types

$$\Pi^{(n)} := \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \;\middle|\; \|\mathbf{s} - \mathbf{P}^*\|_2 \le \frac{1}{n^{1/6}} \right\}$$

Almost-Optimal Types: Part I

Define the set of almost-optimal power types

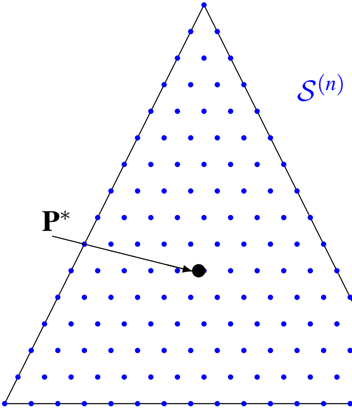
$$\Pi^{(n)} := \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 \le \frac{1}{n^{1/6}} \right\}$$



Almost-Optimal Types: Part I

Define the set of almost-optimal power types

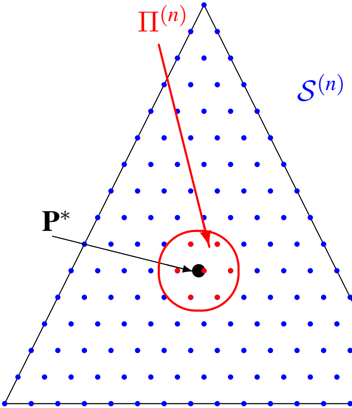
$$\Pi^{(n)} := \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 \le \frac{1}{n^{1/6}} \right\}$$



Almost-Optimal Types: Part I

Define the set of almost-optimal power types

$$\Pi^{(n)} := \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 \le \frac{1}{n^{1/6}} \right\}$$



Almost-Optimal Types: Part II

- Define threshold

$$\log \xi := n\mathsf{C}(\mathbf{P}^*) + \sqrt{n}\left(\sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon + \tau)\right) + \log\left(2|\mathcal{S}^{(n)}|\right)$$

Almost-Optimal Types: Part II

- Define threshold

$$\log \xi := n\mathsf{C}(\mathbf{P}^*) + \sqrt{n}\left(\sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon+\tau)\right) + \log\left(2|\mathcal{S}^{(n)}|\right)$$

- Contribution of information spectrum term on $\Pi^{(n)}$ is

$$\Pr\left\{\frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}}\sum_{k=1}^{n} U_k^{(\mathbf{P}^*)} \geq \Phi^{-1}(\varepsilon+\tau)\right\}$$

where

$$U_k^{(\mathbf{P}^*)} := \sum_{l=1}^{d} \frac{-\left(\frac{P_l^*}{N_l}\right)Z_{l,k}^2 + 2X_{l,k}Z_{l,k} + P_l^*}{2(P_l^* + N_l)} \qquad k \in [1:n]$$

and $\phi(\mathbf{X}^n) \in \Pi^{(n)}$.

Almost-Optimal Types: Part II

- Define threshold

$$\log \xi := n\mathsf{C}(\mathbf{P}^*) + \sqrt{n}\left(\sqrt{\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon+\tau)\right) + \log\left(2|\mathcal{S}^{(n)}|\right)$$

- Contribution of information spectrum term on $\Pi^{(n)}$ is

$$\Pr\left\{\frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}}\sum_{k=1}^{n} U_k^{(\mathbf{P}^*)} \geq \Phi^{-1}(\varepsilon+\tau)\right\}$$

where

$$U_k^{(\mathbf{P}^*)} := \sum_{l=1}^{d} \frac{-\left(\frac{P_l^*}{N_l}\right)Z_{l,k}^2 + 2X_{l,k}Z_{l,k} + P_l^*}{2(P_l^* + N_l)} \qquad k \in [1:n]$$

and $\phi(\mathbf{X}^n) \in \Pi^{(n)}$.

- Because of feedback, $U_k^{(\mathbf{P}^*)}$ is not independent across time!

Almost-Optimal Types: Part III

- Would like to approximate the nasty $\{U_k^{(\mathbf{P}^*)} \mid k \in [1:n]\}$ with the independent and identically distributed random variables

$$V_k^{(\mathbf{P}^*)} := \sum_{l=1}^{d} \frac{-\left(\frac{P_l^*}{N_l}\right) Z_{l,k}^2 + 2\sqrt{P_l} Z_{l,k} + P_l^*}{2(P_l^* + N_l)} \qquad k \in [1:n]$$

Almost-Optimal Types: Part III

- Would like to approximate the nasty $\left\{U_k^{(\mathbf{P}^*)} \mid k \in [1:n]\right\}$ with the independent and identically distributed random variables

$$V_k^{(\mathbf{P}^*)} := \sum_{l=1}^{d} \frac{-\left(\frac{P_l^*}{N_l}\right) Z_{l,k}^2 + 2\sqrt{P_l} Z_{l,k} + P_l^*}{2(P_l^* + N_l)} \qquad k \in [1:n]$$

- Due to our discretization procedure and the fact that $\phi(\mathbf{X}^n) \in \Pi^{(n)}$

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}} \sum_{k=1}^{n} U_k^{(\mathbf{P}^*)}\right)\right] = \lim_{n\to\infty} \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}} \sum_{k=1}^{n} V_k^{(\mathbf{P}^*)}\right)\right]$$

for all $t \in \mathbb{R}$.

Almost-Optimal Types: Part IV

- Curtiss' (aka Lévy's continuity) theorem:

$$\lim_{n\to\infty} \mathbb{E}[\exp(tX_n)] = \lim_{n\to\infty} \mathbb{E}[\exp(tY_n)], \quad \forall\, t \in \mathbb{R},$$

implies that

$$\lim_{n\to\infty} \Pr\{X_n \le a\} = \lim_{n\to\infty} \Pr\{Y_n \le a\}, \quad \forall\, a \in \mathbb{R}.$$

Almost-Optimal Types: Part IV

- Curtiss' (aka Lévy's continuity) theorem:

$$\lim_{n\to\infty} \mathbb{E}[\exp(tX_n)] = \lim_{n\to\infty} \mathbb{E}[\exp(tY_n)], \quad \forall\, t \in \mathbb{R},$$

  implies that

$$\lim_{n\to\infty} \Pr\{X_n \le a\} = \lim_{n\to\infty} \Pr\{Y_n \le a\}, \quad \forall\, a \in \mathbb{R}.$$

- Thus,

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}} \sum_{k=1}^{n} U_k^{(\mathbf{P}^*)} \ge \Phi^{-1}(\varepsilon+\tau) \right\}$$

Almost-Optimal Types: Part IV

- Curtiss' (aka Lévy's continuity) theorem:

$$\lim_{n\to\infty} \mathbb{E}[\exp(tX_n)] = \lim_{n\to\infty} \mathbb{E}[\exp(tY_n)], \quad \forall\, t \in \mathbb{R},$$

implies that

$$\lim_{n\to\infty} \Pr\{X_n \le a\} = \lim_{n\to\infty} \Pr\{Y_n \le a\}, \quad \forall\, a \in \mathbb{R}.$$

- Thus,

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}} \sum_{k=1}^{n} U_k^{(\mathbf{P}^*)} \ge \Phi^{-1}(\varepsilon+\tau) \right\}$$
$$= \lim_{n\to\infty} \Pr\left\{ \frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}} \sum_{k=1}^{n} V_k^{(\mathbf{P}^*)} \ge \Phi^{-1}(\varepsilon+\tau) \right\}$$

Almost-Optimal Types: Part IV

- Curtiss' (aka Lévy's continuity) theorem:

$$\lim_{n\to\infty} \mathbb{E}[\exp(tX_n)] = \lim_{n\to\infty} \mathbb{E}[\exp(tY_n)], \quad \forall\, t \in \mathbb{R},$$

implies that

$$\lim_{n\to\infty} \Pr\{X_n \le a\} = \lim_{n\to\infty} \Pr\{Y_n \le a\}, \quad \forall\, a \in \mathbb{R}.$$

- Thus,

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}} \sum_{k=1}^{n} U_k^{(\mathbf{P}^*)} \ge \Phi^{-1}(\varepsilon+\tau) \right\}$$

$$= \lim_{n\to\infty} \Pr\left\{ \frac{1}{\sqrt{n\mathsf{V}(\mathbf{P}^*)}} \sum_{k=1}^{n} V_k^{(\mathbf{P}^*)} \ge \Phi^{-1}(\varepsilon+\tau) \right\}$$

$$\stackrel{\text{CLT}}{=} 1-(\varepsilon+\tau).$$

Far-From-Optimal Types

- Contribution of information spectrum term on $\mathcal{S}^{(n)} \setminus \Pi^{(n)}$ is

$$\approx \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \geq \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\}, \quad \forall\, \mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}$$

Far-From-Optimal Types

- Contribution of information spectrum term on $\mathcal{S}^{(n)} \setminus \Pi^{(n)}$ is

$$\approx \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \geq \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\}, \quad \forall\, \mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}$$

- Recall

$$\mathcal{S}^{(n)} \setminus \Pi^{(n)} = \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 > \frac{1}{n^{1/6}} \right\}$$

Far-From-Optimal Types

- Contribution of information spectrum term on $\mathcal{S}^{(n)} \setminus \Pi^{(n)}$ is

$$\approx \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \geq \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\}, \quad \forall\, \mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}$$

- Recall

$$\mathcal{S}^{(n)} \setminus \Pi^{(n)} = \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 > \frac{1}{n^{1/6}} \right\}$$

- On this set,

$$\mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) = \Omega\left(\frac{1}{n^{1/3}} \right).$$

Far-From-Optimal Types

- Contribution of information spectrum term on $\mathcal{S}^{(n)} \setminus \Pi^{(n)}$ is

$$\approx \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \ge \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\}, \quad \forall\, \mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}$$

- Recall

$$\mathcal{S}^{(n)} \setminus \Pi^{(n)} = \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \,\middle|\, \|\mathbf{s} - \mathbf{P}^*\|_2 > \frac{1}{n^{1/6}} \right\}$$

- On this set,

$$\mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) = \Omega\left(\frac{1}{n^{1/3}} \right).$$

- Thus, by a standard exponential Markov inequality argument

$$\max_{\mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}} \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \ge \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\} \le \exp\left(-O(n^{1/3}) \right).$$

Far-From-Optimal Types

- Contribution of information spectrum term on $\mathcal{S}^{(n)} \setminus \Pi^{(n)}$ is

$$\approx \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \geq \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\}, \quad \forall\, \mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}$$

- Recall

$$\mathcal{S}^{(n)} \setminus \Pi^{(n)} = \left\{ \mathbf{s} \in \mathcal{S}^{(n)} \;\middle|\; \|\mathbf{s} - \mathbf{P}^*\|_2 > \frac{1}{n^{1/6}} \right\}$$

- On this set,

$$\mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) = \Omega\left(\frac{1}{n^{1/3}} \right).$$

- Thus, by a standard exponential Markov inequality argument

$$\max_{\mathbf{s} \in \mathcal{S}^{(n)} \setminus \Pi^{(n)}} \Pr\left\{ \frac{1}{n} \sum_{k=1}^{n} U_k^{(\mathbf{s})} \geq \mathsf{C}(\mathbf{P}^*) - \mathsf{C}(\mathbf{s}) \right\} \leq \exp\left(-O(n^{1/3}) \right).$$

- Polynomially many power types.

Wrap-Up and Future Work

- For parallel Gaussian channels with feedback,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

Wrap-Up and Future Work

- For parallel Gaussian channels with feedback,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Careful quantization of power allocation vector into power types

Wrap-Up and Future Work

- For parallel Gaussian channels with feedback,

$$\log M^*_{\mathsf{FB}}(n, \varepsilon, P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Careful quantization of power allocation vector into power types
- Bounding main information spectrum term using Curtiss' theorem

Wrap-Up and Future Work

- For parallel Gaussian channels with feedback,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Careful quantization of power allocation vector into power types
- Bounding main information spectrum term using Curtiss' theorem
- For third-order, need speed of convergence, i.e.,

$$\sup_{t\in\mathbb{R}} \left|\mathbb{E}[\exp(tX_n)] - \mathbb{E}[\exp(tY_n)]\right| \le f(n), \quad \text{for some } f(n) = o(1),$$

do we have

$$\sup_{a\in\mathbb{R}} \left|\Pr\{X_n \le a\} - \Pr\{Y_n \le a\}\right| \le g(n), \quad \text{for some desirable } g(n)?$$

Wrap-Up and Future Work

- For parallel Gaussian channels with feedback,

$$\log M^*_{\mathsf{FB}}(n,\varepsilon,P) = n\mathsf{C}(\mathbf{P}^*) + \sqrt{n\mathsf{V}(\mathbf{P}^*)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}).$$

- Careful quantization of power allocation vector into power types
- Bounding main information spectrum term using Curtiss' theorem
- For third-order, need speed of convergence, i.e.,

$$\sup_{t\in\mathbb{R}} |\mathbb{E}[\exp(tX_n)] - \mathbb{E}[\exp(tY_n)]| \le f(n), \quad \text{for some } f(n) = o(1),$$

do we have

$$\sup_{a\in\mathbb{R}} |\Pr\{X_n \le a\} - \Pr\{Y_n \le a\}| \le g(n), \quad \text{for some desirable } g(n)?$$

- Esséen smoothing lemma? Stein's method?

Full paper

Oct 2017 issue of IEEE Transactions on Information Theory

6474 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 63, NO. 10, OCTOBER 2017

A Tight Upper Bound on the Second-Order Coding Rate of the Parallel Gaussian Channel With Feedback

Silas L. Fong, *Member, IEEE*, and Vincent Y. F. Tan, *Senior Member, IEEE*

***Abstract*—This paper investigates the asymptotic expansion for the maximum rate of fixed-length codes over a parallel Gaussian channel with feedback under the following setting: a peak power constraint is imposed on every transmitted codeword, and the average error probabilities of decoding the transmitted message are non-vanishing as the blocklength increases. The main contribution of this paper is a self-contained proof of an upper bound on the first- and second-order asymptotics of the parallel Gaussian channel with feedback. The proof techniques involve developing an information spectrum bound followed by using Curtiss' theorem to show that a sum of dependent random variables associated with the information spectrum bound converges in distribution to a sum of independent random variables, thus facilitating the use of the usual central limit theorem. Combined with existing achievability results, our result implies that the presence of feedback does not improve the first- and second-order asymptotics.**

***Index Terms*—Curtiss' theorem, feedback, fixed-length codes, parallel Gaussian channel, second-order asymptotics.**

where $\{Z_{\ell,k}\}_{\ell\in\mathcal{L}}$ are independent Gaussian noises. For each $\ell \in \mathcal{L}$, the variance of the noise induced by the ℓ^{th} channel is assumed to be some positive number $N_\ell > 0$ for all channel uses, i.e., $\mathrm{Var}[Z_{\ell,k}] = N_\ell$ for all $k \in \mathbb{N}$. To keep notation compact, let $\boldsymbol{X}_k$, $\boldsymbol{Y}_k$ and $\boldsymbol{Z}_k$ denote the random column vectors $[X_{1,k}\ X_{2,k}\ \ldots\ X_{L,k}]^t$, $[Y_{1,k}\ Y_{2,k}\ \ldots\ Y_{L,k}]^t$ and $[Z_{1,k}\ Z_{2,k}\ \ldots\ Z_{L,k}]^t$ respectively. Then, the channel law (1) can be written as

$$\boldsymbol{Y}_k = \boldsymbol{X}_k + \boldsymbol{Z}_k. \tag{2}$$

Throughout this paper, we consider fixed-length codes over the parallel Gaussian channel, where the block length is denoted by n unless specified otherwise. Every codeword $\boldsymbol{X}^n$ transmitted by the source over n channel uses is subject to the following *peak power constraint* where $P > 0$ denotes the permissible power for $\boldsymbol{X}^n$:

Advertisement: Postdoc Positions in IT and ML at NUS

Postdoctoral positions available at the National University of Singapore

The Department of Electrical and Computer Engineering (ECE) at the National University of Singapore is offering positions for postdoctoral fellows who will work in information theory, machine learning and their intersection.

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

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

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

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