Information Theory for Low-Latency Wireless Communications

LOLITA

Tobias Koch

Universidad Carlos III de Madrid

Information Theory and Tapas Workshop (January 27, 2023)

Collaborators

Giuseppe Durisi
Full Professor, Chalmers University of Technology

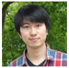
Alejandro Lancho
Marie Curie Research Fellow, MIT

Johan Östman
Research Scientist, Zenseact / AI Sweden

Yury Polyanskiy
Professor, MIT

Chao Qi
Postdoctoral Researcher, UC3M

Gonzalo Vazquez-Vilar
Associate Professor, UC3M



Wei Yang
Senior Engineer, Qualcomm

Low-latency wireless communications

- Internet of Things
- Machine-to-machine communications
- tactile Internet
  - → **low latency and high reliability**
  - → **transmission of short packets with low probability of error**

Example: Ultra-reliable and low-latency communications

Services supported by 5G:

- enhanced mobile broadband (eMBB)
- massive machine-type communications (mMTC)
- **ultra-reliable and low-latency communications (URLLC)**

URLLC

- Total latency less than 1ms
  - low-latency + limited bandwidth = short packets
- Less than 1 packet loss in 10^5 packets
  - requires powerful forward error correction (channel coding)

Fundamental limits

Traditionally, fundamental limits means capacity / outage capacity

Capacity of coherent MIMO fading channels

E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Transactions on Emerging Telecommunications Technologies*, November 1999.

Capacity of noncoherent MIMO fading channels

T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Transactions on Information Theory*, January 1999.

L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channels,” *IEEE Transactions on Information Theory*, February 2002.

Capacity versus outage

L. Ozarow, S. Shamai, and A. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Transactions on Vehicular Technology*, May 1994.

G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels,” *IEEE Transactions on Information Theory*, July 1999.

⋮

Capacity versus short packets

Channel capacity C**:** largest rate R such that $P_e \to 0$ as $n \to \infty$

➔ requires the transmission of long packets

➔ not necessarily a good benchmark for short-packet communications

Maximum coding rate $R^*(n, \epsilon)$**:** Largest rate R for which there exists a channel code of blocklength n such that $P_e \leqslant \epsilon$

➔ $R^*(n, \epsilon) \to C$ as $n \to \infty$

➔ behavior of $n \mapsto R^*(n, \epsilon)$ relevant for short-packet communications

Finite-blocklength information theory

Nonasymptotic behavior of $R^*(n, \epsilon)$

Estimate $R^*(n, \epsilon)$ by means of bounds:

- *Lower bounds:* dependence-testing bound (PPV10), RCU_s bound (MGiF11)
- *Upper bound:* meta-converse bound (PPV10)

Asymptotic behavior of $R^*(n, \epsilon)$

- *Error exponents:*

$$P_e^*(n, R) = e^{-nE_r(R)+o(n)} \qquad (E_r(R)\text{: reliability function})$$

- *Normal approximation* (Strassen'62, Hayashi'09, PPV10):

$$R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \qquad (V\text{: channel dispersion})$$

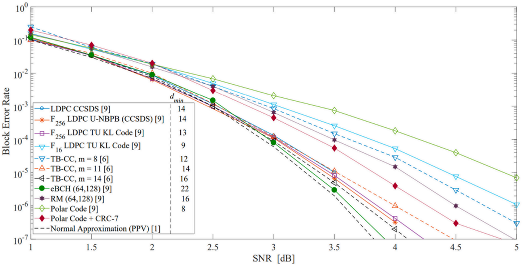
Y. Polyanskiy, H.V. Poor, S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, May 2010.

A. Martinez, A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *Proc. Inf. Theory and Appl. Workshop (ITA)*, Feb. 2011.

Normal approximation as benchmark

Coding for short blocks

URLLC will require sending of very short messages in the 10 to 100-bit range. Channel coding schemes like LDPC and Polar codes need to be fine-tuned short messages. Candidate coding schemes for URLLC are benchmarked in the following figure.

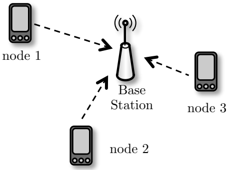


Channel coding candidates for URLLC (credit: Short Block Length Codes for URLLC)

Note that coding techniques for URLLC need to be benchmarked against the PPV limit as the Shannon limit does not apply for such messages.

(Screenshot: `medium.com/5g-nr` → Ultra-Reliable Low-Latency Communication (URLLC))

NA in the analysis of communication protocols



Example: Framed ALOHA protocol

- d devices, each sending k bits to base station
- n channel uses divided into s slots of $n_s = n/s$ channel uses
- each device picks randomly a slot to send its packet
  - if $\geqslant 2$ devices pick the same slot, then all packets are lost
  - if only one device picks a slot, then packet is lost with probability

$$\epsilon^*(k, n_s) \approx Q\left(\frac{n_s C - k + (\log n_s)/2}{\sqrt{n_s V}}\right)$$

How to choose s to maximize prob. of successful transmission?

Example: Framed ALOHA protocol

Numerical example:

- $\mathsf{d} = 10$, $\mathsf{k} = 192$ bits, $n = 800$
- AWGN channel with SNR $\rho = 10$ dB

Probability of successful transmission:

$$P_S = \frac{\mathsf{d}}{\mathsf{s}}\left(1 - \frac{1}{\mathsf{s}}\right)^{\mathrm{d}-1}\left(1 - \epsilon^*(\mathsf{k}, n_{\mathsf{s}})\right)$$

➔ *optimal number of slots* s*:* $\mathsf{s}^* = 6$

➔ *classic framed-ALOHA protocol (*$\epsilon^*(\mathsf{k}, n_{\mathsf{s}}) = 0$*):* $\mathsf{s}^* = 10$

G. Durisi, T. Koch, P. Popovski, "Towards massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, Sep. 2016.

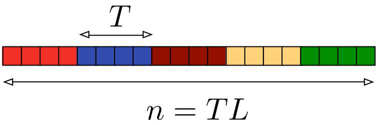
FBL information theory for wireless communications

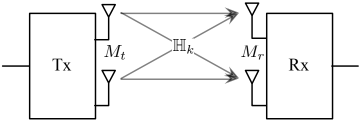
- $R^*(n, \epsilon)$ depends critically on assumed channel model
- most results derived for AWGN channel or DMCs
- these channels do not capture:
  - coherence time/bandwidth
  - channel estimation overhead
  - number of transmit/receive antennas
  - tradeoff between diversity and multiplexing

→ **need FBL information theory for wireless communication channels**

→ **LOLITA**

Rayleigh block-fading channel

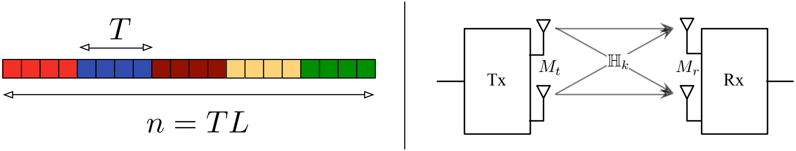




$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{W}_k, \quad k \in \mathbb{Z}$$

- $\{\mathbb{H}_k\}$ blockwise IID, $\mathbb{H}_k \sim \mathcal{N}_{\mathbb{C}}(0, \mathsf{I}_{M_r \times M_t})$ (Rayleigh fading)
- $\{\mathbf{W}_k\} \sim$ IID $\mathcal{N}_{\mathbb{C}}(0, \mathsf{I}_{M_r})$
- M_t transmit antennas, M_r receive antennas
- T: coherence interval
- L: number of time-frequency branches

No *a priori* CSI - noncoherent setting



- $\{\mathbb{H}_k\}$ **unknown at transmitter (no CSI@Tx):**
  - high-mobility scenarios or time-critical applications
  - avoid need for feedback link
- $\{\mathbb{H}_k\}$ **unknown at receiver (no CSI@Rx):**
  - characterizing cost of obtaining CSI
  - pilot-aided channel estimation one possible coding scheme

Maximum coding rate $R^*(L, T, \epsilon, \rho)$

Per-block power constraint:

$$\sum_{\ell=1}^{T} \|\mathbf{X}_{jT+\ell}\|^2 \leqslant T\rho, \quad \forall j$$

Maximum error probability:

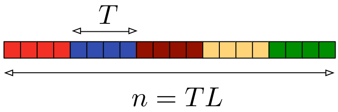
$$\max_{b^K \in \{0,1\}^K} \Pr\left(\widehat{B}^K \neq B^K \mid B^K = b^K\right) \leqslant \epsilon$$

Rate: $R \triangleq \frac{K}{LT}$

Maximum coding rate:

$$R^*(L, T, \epsilon, \rho) = \left\{ \begin{array}{c} \text{largest rate } R \text{ for which there exists} \\ \text{an encoder and decoder satisfying} \\ \text{the power and error prob. constraints} \end{array} \right\}$$

Quasistatic fading channels (fix L and let $T \to \infty$)



Normal approximation (YDKP14)

$$R^*(L, T, \epsilon, \rho) = C_\epsilon(\rho) + \mathcal{O}_T\left(\frac{\log T}{T}\right)$$

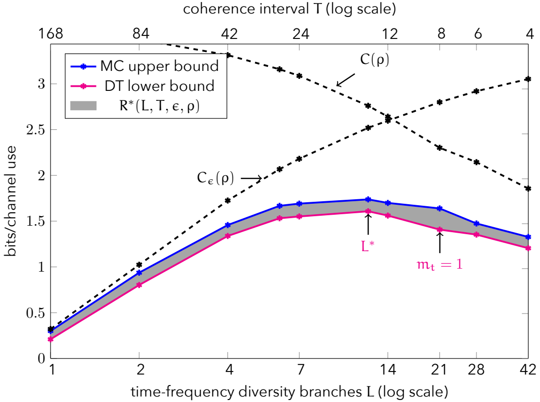
- ➔ channel dispersion V is zero
- ➔ holds irrespective of availability of CSI
- ➔ holds irrespective of number of transmit and receive antennas

W. Yang, G. Durisi,T. Koch, Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, Jul. 2014.

Ergodic fading channels (fix T and let $L \to \infty$)

No normal approximation but nonasymptotic bounds available:

- must be evaluated numerically
- mathematical analysis of $R^*(L, T, \epsilon, \rho)$ difficult



$n = TL = 168$

$\rho = 6$ dB

$M_t = M_r = 2$

$\epsilon = 10^{-5}$

G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, Feb. 2016.

The high-SNR normal approximation

- Single-antenna Rayleigh block-fading channel
- **Ergodic case:** Fix T and let $L \to \infty$

High-SNR normal approximation

$$R^*(L, T, \epsilon, \rho) = \underline{R}_{\mathsf{USTM}}(\rho) + o_\rho(1) - \sqrt{\frac{V + o_\rho(1)}{LT}} Q^{-1}(\epsilon) + \mathcal{O}_L\left(\frac{\log L}{L}\right)$$

where

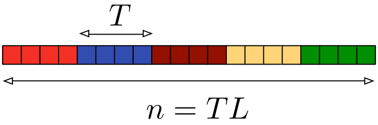
$$\underline{R}_{\mathsf{USTM}}(\rho) = \left(1 - \frac{1}{T}\right) \log(T\rho) - \frac{\log \Gamma(T)}{T} + \frac{1}{T} {}_2F_1\left(1, T-1; T; \frac{T\rho}{1+T\rho}\right)$$
$$- \left(1 - \frac{1}{T}\right) \left[\log(1+T\rho) + \frac{T\rho}{1+T\rho} - \psi(T-1)\right]$$

$$V = T\left(1 - \frac{1}{T}\right)^2 \frac{\pi^2}{6} + 1 - \frac{1}{T}$$

$o_\rho(1)$: terms that vanish as $\rho \to \infty$, $\mathcal{O}_L(\log L/L)$: terms of order $\log L/L$

Unitary Space-Time Modulation (USTM)

- $\mathbb{X}_j = \begin{pmatrix} \leftarrow & \mathbf{X}_{(j-1)T+1} & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{X}_{(j-1)T+T} & \rightarrow \end{pmatrix}$



- $\mathbb{X}_1, \ldots, \mathbb{X}_L$ are IID
- $\mathbb{X}_j = \sqrt{T\rho}\,\mathbb{U}_j \quad \rightarrow \mathbb{U}_j$: isotropically-distributed unitary matrix

Theorem (HM00, ZT02):

When $T \geqslant M_t + M_r$, the rate R_{USTM} achievable with USTM satisfies

$$\lim_{\rho \to \infty} \left\{ C(\rho) - R_{\text{USTM}}(\rho) \right\} = 0.$$

B. M. Hochwald and T. L. Marzetta, “Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading,” *IEEE Transactions on Information Theory*, March 2000.

Features of the proof: Achievability

Dependence testing bound (PPV10)

For every $P_{\mathbf{X}^L}$ and $\gamma(\cdot)$, there exists an encoder and decoder such that

$$P_e \leqslant \Pr\big(i(\mathbf{X}^L;\mathbf{Y}^L) \leqslant \log\gamma(\mathbf{X}^L)\big) + (2^K-1)\sup_{\mathbf{x}^L}\Pr\big(i(\mathbf{x}^L;\mathbf{Y}^L) > \log\gamma(\mathbf{x}^L)\big)$$

where $i(\mathbf{x}^L;\mathbf{y}^L) \triangleq \log\frac{f_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{y}^L|\mathbf{x}^L)}{f_{\mathbf{Y}^L}(\mathbf{y}^L)}$.

- DT bound with USTM channel inputs
- Output PDF induced by USTM inputs:

$$f_{\mathbf{Y}}^{(U)}(\mathbf{y}) = \frac{e^{-\frac{\|\mathbf{y}\|^2}{1+T\rho}}\|\mathbf{y}\|^{2(1-T)}\Gamma(T)}{\pi^T(1+T\rho)}\tilde{\gamma}\left(T-1,\frac{T\rho\|\mathbf{y}\|^2}{1+T\rho}\right)\left(1+\frac{1}{T\rho}\right)^{T-1}$$

- Perform central-limit-type analysis of $i(\mathbf{X}^L;\mathbf{Y}^L) = \sum_{\ell=1}^{L} i(\mathbf{X}_\ell;\mathbf{Y}_\ell)$

Features of the proof: Converse

Weakened meta-converse bound (Verdú-Han bound) (PPV10)

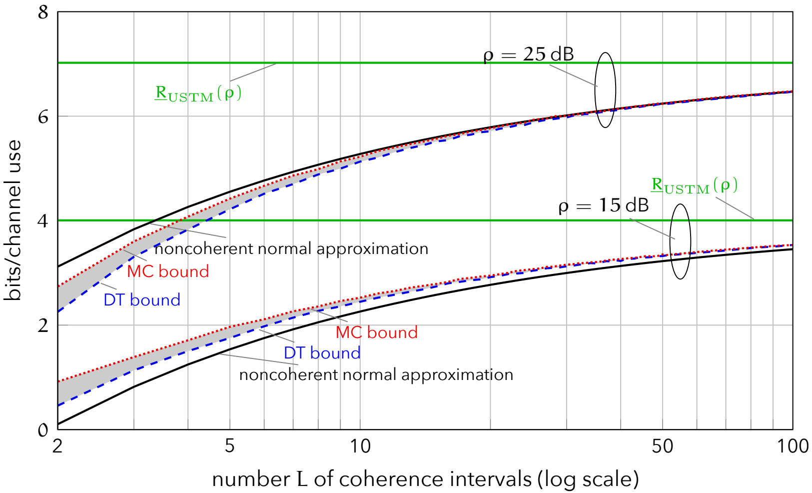
For every $\xi(\cdot)$, the maximum coding rate is upper-bounded by

$$R^*(L,T,\epsilon,\rho) \leqslant \sup_{\mathbf{x}^L} \left\{ \frac{\log \xi(\mathbf{x}^L)}{LT} - \frac{\log\left(1-\epsilon-\Pr\left(j(\mathbf{x}^L;\mathbf{Y}^L) \geqslant \log \xi(\mathbf{x}^L)\right)\right)}{LT} \right\}$$

where $j(\mathbf{x}^L;\mathbf{y}^L) \triangleq \log \frac{f_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{y}^L|\mathbf{x}^L)}{q_{\mathbf{Y}^L}(\mathbf{y}^L)}$ and $q_{\mathbf{Y}^L}$ is an auxiliary pdf.

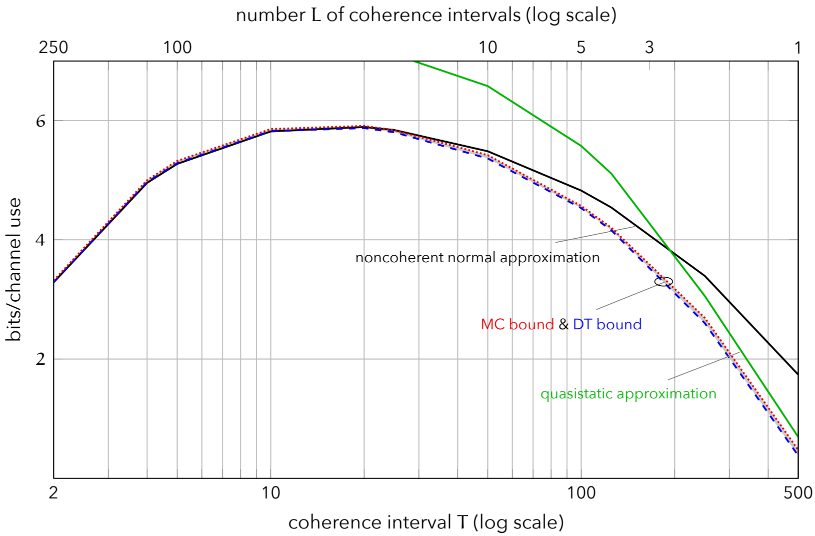
- Auxiliary pdf $q_{\mathbf{Y}^L}$: choose pdf $f^{(U)}_{\mathbf{Y}^L}$ induced by USTM inputs
- Perform central-limit-type analysis of $j(\mathbf{x}^L;\mathbf{Y}^L) = \sum_{\ell=1}^{L} j(\mathbf{x}_\ell;\mathbf{Y}_\ell)$
- Optimize over $\mathbf{x}^L$

Numerical results: $L \mapsto R^*(L, T, \epsilon, \rho)$



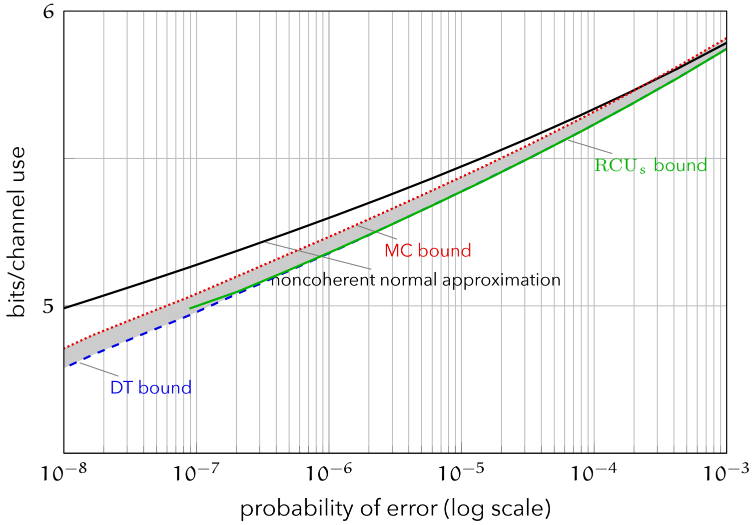
$T = 20$, $\rho = \{15 \text{ dB}, 25 \text{ dB}\}$, $M_t = M_r = 1$, $\epsilon = 10^{-3}$

Numerical results: LT fixed



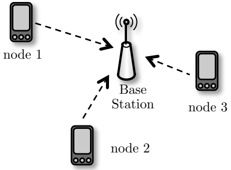
$LT = 500$, $\rho = 25$ dB, $M_t = M_r = 1$, $\epsilon = 10^{-3}$

Numerical results: $\epsilon \mapsto R^*(L, T, \epsilon, \rho)$



$L = 25$, $T = 20$, $\rho = 25$ dB, $M_t = M_r = 1$

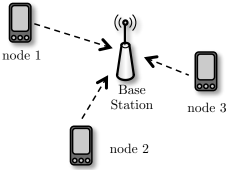
Framed ALOHA protocol revisited (1)



- $\mathsf{d} = 12$ devices, each sending $\mathsf{k} = 256$ bits to base station
- $n = LT = 480$ channel uses divided into s slots of $n_{\mathsf{s}} = n/\mathsf{s}$ channel uses
- each device picks randomly a slot to send its packet
  - if $\geqslant 2$ devices pick the same slot, then all packets are lost
  - if only one device picks a slot, then packet is lost with probability

$$\epsilon^*(\mathsf{k}, n_{\mathsf{s}}, \rho) \approx Q\left(\frac{n_{\mathsf{s}} C_\ell(\rho) - \mathsf{k}}{\sqrt{n_{\mathsf{s}} V_\ell(\rho)}}\right), \quad \ell \in \{\text{AWGN}, \text{block-fading}\}$$

Framed ALOHA protocol revisited (2)



SNR	T	optimal number of slots s			
		noncoherent block-fading channel	coherent block-fading channel	AWGN channel	classic ALOHA protocol
$\rho = 15$ dB	5	$s = 4$	$s = 6$	$s = 8$	$s = 12$
	20	$s = 6$	$s = 6$	$s = 8$	$s = 12$
$\rho = 25$ dB	5	$s = 8$	$s = 12$	$s = 12$	$s = 12$
	20	$s = 8$	$s = 8$	$s = 12$	$s = 12$

➔ **less favorable channel $\Leftrightarrow$ fewer slots (larger blocklength per slot)**

Discussion: High-SNR normal approximation

✔ Available in closed-form

✔ Accurate for SNR $\geqslant 15$ dB, $L \geqslant 10$, and moderate ϵ

- useful performance benchmark
- can be used to analyze the performance of communication protocols

✘ Inaccurate for SNR < 15 dB, $L < 10$, and small ϵ

URLLC

- Total latency less than 1ms
  - low-latency + limited bandwidth = short packets
- Less than 1 packet loss in 10^5 packets
  - requires power forward error correction (channel coding)

Saddlepoint approximation

$Z_1, \ldots, Z_n$: sequence of i.i.d., zero-mean, random variables with pdf f_Z

- *moment generating function:* $m(\zeta) = \mathsf{E}\left[e^{\zeta Z_\ell}\right]$
- *cumulant generating function:* $\psi(\zeta) = \log m(\zeta)$

$$\Pr\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_\ell \geqslant \gamma\right)$$

$$= \int_{\gamma}^{\infty} \frac{n}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} e^{n(\psi(\zeta)-\zeta z)} \mathrm{d}\zeta \mathrm{d}z \qquad \text{inv. Laplace transform}$$

$$= \frac{n}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{n\zeta} e^{n(\psi(\zeta)-\zeta\gamma)} \mathrm{d}\zeta \qquad \text{solve integral over } z$$

$$\approx \frac{1}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{\zeta} e^{n(\psi(\tau)+\psi'(\tau)(\zeta-\tau)+\frac{1}{2}\psi''(\tau)(\zeta-\tau)^2-\zeta\gamma)} \mathrm{d}\zeta \qquad \text{Taylor series of } \psi(\cdot)$$

$$= e^{n(\psi(\tau)-\tau\gamma)} \frac{1}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{\zeta} e^{\frac{n}{2}\psi''(\tau)(\zeta-\tau)^2} \mathrm{d}\zeta \qquad \text{choose } \tau \text{ s.t. } \psi'(\tau) = \gamma$$

$$= e^{n(\psi(\tau)-\tau\gamma)} Q\left(\sqrt{n\psi''(\tau)\tau^2}\right) e^{\frac{n}{2}\psi''(\tau)\tau^2} \qquad \text{solve integral over } \zeta$$

Saddlepoint approximation

$Z_1, \ldots, Z_n$: sequence of i.i.d., zero-mean, random variables with pdf f_Z

- *moment generating function:* $m(\zeta) = \mathsf{E}\left[e^{\zeta Z_\ell}\right]$
- *cumulant generating function:* $\psi(\zeta) = \log m(\zeta)$

$$
\begin{aligned}
&\Pr\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_\ell \geqslant \gamma\right) \\
&= \int_\gamma^\infty \frac{n}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} e^{n(\psi(\zeta)-\zeta z)} \mathrm{d}\zeta \mathrm{d}z && \text{inv. Laplace transform} \\
&= \frac{n}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{n\zeta} e^{n(\psi(\zeta)-\zeta\gamma)} \mathrm{d}\zeta && \text{solve integral over } z \\
&\approx \frac{1}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{\zeta} e^{n(\psi(\tau)+\psi'(\tau)(\zeta-\tau)+\frac{1}{2}\psi''(\tau)(\zeta-\tau)^2-\zeta\gamma)} \mathrm{d}\zeta && \text{Taylor series of } \psi(\cdot) \\
&= e^{n(\psi(\tau)-\tau\gamma)} \frac{1}{2\pi \mathrm{i}} \int_{\tau-\mathrm{i}\infty}^{\tau+\mathrm{i}\infty} \frac{1}{\zeta} e^{\frac{n}{2}\psi''(\tau)(\zeta-\tau)^2} \mathrm{d}\zeta && \text{choose } \tau \text{ s.t. } \psi'(\tau) = \gamma \\
&= e^{n(\psi(\tau)-\tau\gamma)} Q\left(\sqrt{n\psi''(\tau)\tau^2}\right) e^{\frac{n}{2}\psi''(\tau)\tau^2} && \text{solve integral over } \zeta
\end{aligned}
$$

Saddlepoint expansion

$Z_1, \ldots, Z_n$: sequence of i.i.d., zero-mean, random variables

- *moment generating function:* $m(\zeta) = \mathsf{E}\left[e^{\zeta Z_\ell}\right]$
- *cumulant generating function:* $\psi(\zeta) = \log m(\zeta)$
- *characteristic function:* $\varphi(\zeta) = \mathsf{E}\left[e^{\mathrm{i}\zeta Z_\ell}\right]$

Z_ℓ is *lattice* if it is supported on $b, b \pm h, b \pm 2h, \ldots$ (for some b and h)

➔ Z_ℓ is *nonlattice* if it is not lattice

➔ Z_ℓ is nonlattice iff $|\varphi(\zeta)| < 1$ for every $\zeta \neq 0$

Saddlepoint expansion (Daniels'54, Feller'71, Jensen'95,...)
Let $Z_1, \ldots, Z_n$ be i.i.d. nonlattice random variables of positive variance. Assume that $m(\zeta) < \infty$ on an open interval around $\zeta = 0$. Then

$$\Pr\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_\ell \geqslant \gamma\right) = e^{n(\psi(\tau)-\tau\gamma)}\left[Q\left(\sqrt{n\psi''(\tau)\tau^2}\right)e^{\frac{n}{2}\psi''(\tau)\tau^2} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right]$$

Saddlepoint expansions of bounds on $R^*(L, T, \epsilon, \rho)$

Weakened meta-converse bound (PPV10)

$$R^*(L, T, \epsilon, \rho) \leqslant \sup_{\mathbf{x}^L} \left\{ \frac{\log \xi(\mathbf{x}^L)}{LT} - \frac{\log\left(1 - \epsilon - \Pr\left(\sum_{\ell=1}^{L} j(\mathbf{x}_\ell; \mathbf{Y}_\ell) \geqslant \log \xi(\mathbf{x}^L)\right)\right)}{LT} \right\}$$

- $Z_\ell \leftrightarrow j(\mathbf{x}_\ell; \mathbf{Y}_\ell)$
- $\gamma \leftrightarrow \log \xi(\mathbf{x}^L)$

However:

- $j(\mathbf{x}_\ell; \mathbf{Y}_\ell)$ depends on system parameters such as
  - SNR
  - number of transmit and receive antennas
  - ...
- We wish the error terms $\mathcal{O}(1/\sqrt{n})$ to be uniform in these parameters

Families of distributions

$Z_{1,\theta}, \ldots, Z_{n,\theta}$: sequence of i.i.d., zero-mean, random variables

- depends on parameter $\theta \in \Theta$
- $m(\zeta) \to m_\theta(\zeta)$, $\psi(\zeta) \to \psi_\theta(\zeta)$, $\varphi(\zeta) \to \varphi_\theta(\zeta)$

Family of random variables $Z_{\ell,\theta}$ is *nonlattice* if

$$\sup_{\theta \in \Theta} |\varphi_\theta(\zeta)| < 1, \quad \text{for every } \zeta \neq 0$$

We assume that (for some $\zeta_0 > 0$)

- $\sup_{\theta \in \Theta, |\zeta| < \zeta_0} m_\theta^{(4)}(\zeta) < \infty$
- $\inf_{\theta \in \Theta, |\zeta| < \zeta_0} \psi_\theta''(\zeta) > 0$

Saddlepoint expansion for families of distributions

Saddlepoint expansion for families of distributions

Let $Z_{1,\theta}, \ldots, Z_{n,\theta}$ be a family of i.i.d. nonlattice random variables. Then

$$\Pr\left(\sum_{\ell=1}^{n} Z_{\ell,\theta} \geqslant \gamma\right) = e^{n[\psi_\theta(\tau) - \tau\psi'_\theta(\tau)]}\left[\Psi_\theta(\tau, n) + \frac{\mathsf{K}_\theta(\tau, n)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right]$$

where

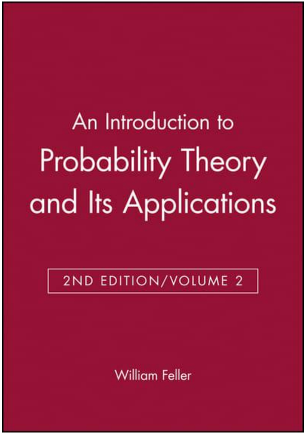
$$\Psi_\theta(\tau, n) = Q\left(\sqrt{n\psi''_\theta(\tau)\tau^2}\right) e^{\frac{n}{2}\psi''_\theta(\tau)\tau^2}$$

$$\mathsf{K}_\theta(\tau, n) = \frac{\psi'''_\theta(\tau)}{6\psi''_\theta(\tau)^{3/2}}\left(-\frac{1}{\sqrt{2\pi}} + \frac{\tau^2 n\psi''_\theta(\tau)}{\sqrt{2\pi}} - \tau^3\psi''_\theta(\tau)^{3/2} n^{3/2}\Psi_\theta(\tau, n)\right)$$

and τ is the solution to $n\psi'_\theta(\tau) = \gamma$.

$o(1/\sqrt{n})$: term that is uniform in τ and θ and vanishes faster than $1/\sqrt{n}$

Features of the proof



- $F_\theta^{\star n}$: distribution of $\sum_{\ell=1}^{n}(Z_{\ell,\theta} - \gamma/n)$
- $\vartheta_{\theta,\tau}$: tilted distribution

$$\vartheta_{\theta,\tau}^{\star n}(x) = e^{-n\psi_\theta(\tau)+\tau\gamma} \int_{-\infty}^{x} e^{\tau t} \mathrm{d}F_\theta^{\star n}(t)$$

- Tail distribution of $\sum_{\ell=1}^{n} Z_{\ell,\theta}$:

$$\Pr\left(\sum_{\ell=1}^{n} Z_{\ell,\theta} \geqslant \gamma\right) = e^{n\psi_\theta(\tau)-\tau\gamma} \int_{0}^{\infty} e^{-\tau y} \mathrm{d}\vartheta_{\theta,\tau}^{\star n}(y)$$

- Perform expansion of $\vartheta_{\theta,\tau}^{\star n}$ around zero-mean Gaussian distribution with variance $n\psi_\theta''(\tau)$

Upper bound on minimum error probability

Minimum error probability $\epsilon^*(L, T, R, \rho)$**:** smallest error probability P_e for which there exists a channel code of blocklength n and rate R

RCU$_s$ **bound** (MGiF11)

For every $s > 0$, there exists an encoder and decoder such that

$$\epsilon^*(L, T, R, \rho) \leqslant \Pr\left(\sum_{\ell=1}^{L}\big(I_s(\rho) - i_{\ell,s}(\mathbf{X}_\ell; \mathbf{Y}_\ell)\big) \geqslant L I_s(\rho) + \log U - LTR\right)$$

where

$$i_{\ell,s}(\mathbf{x}_\ell; \mathbf{y}_\ell) \triangleq \log \frac{f_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\mathbf{x}_\ell)^s}{\int f_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\tilde{\mathbf{x}})^s \mathrm{d}P_{\mathbf{X}_\ell}(\tilde{\mathbf{x}})}$$

$I_s(\rho) \triangleq \mathrm{E}[i_{\ell,s}(\mathbf{X}_\ell; \mathbf{Y}_\ell)]$, and $U \sim \mathcal{U}([0,1])$.

➔ for $s = 1$, RCU$_s$ bound = DT bound

Lower bound on minimum error probability

Weakened meta-converse bound (Verdú-Han bound) (PPV10)

For every $\xi > 0$ and $s > 0$,

$$\epsilon^*(L, T, R, \rho) \geqslant \Pr\left(\sum_{\ell=1}^{L}\big(J_s(\rho) - j_{\ell,s}(\mathbf{X}_\ell; \mathbf{Y}_\ell)\big) \geqslant s(LJ_s(\rho) - \log \xi)\right) - e^{\log \xi - LTR}$$

where

$$j_{\ell,s}(\mathbf{x}_\ell; \mathbf{y}_\ell) \triangleq \log \frac{f_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\mathbf{x}_\ell)}{q^s_{\mathbf{Y}_\ell}(\mathbf{y}_\ell)}$$

$$q^s_{\mathbf{Y}_\ell}(\mathbf{y}_\ell) \triangleq \frac{1}{\mu(s)}\left(\int f_{\mathbf{Y}_\ell|\mathbf{X}_\ell}(\mathbf{y}_\ell|\tilde{\mathbf{x}})^s \mathrm{d}P_{\mathbf{X}_\ell}(\tilde{\mathbf{x}})\right)^{1/s}$$

and $J_s(\rho) \triangleq \mathrm{E}[j_{\ell,s}(\mathbf{X}_\ell; \mathbf{Y}_\ell)]$.

➔ meta-converse bound with auxiliary distribution $q^s_{\mathbf{Y}_\ell}$

Saddlepoint expansions of RCU_s and MC bounds

Slightly more restrictive per-block power constraint:

$$\sum_{\ell=1}^{T} \|\mathbf{X}_{jT+\ell}\|^2 \leqslant T\rho, \quad \forall j$$

Apply saddlepoint expansion with

- $\theta = (\rho, s)$
- $m_{\rho,s}(\tau) \triangleq \mathsf{E}\left[e^{\tau[I_s(\rho) - i_{\ell,s}(\mathbf{X}_\ell;\mathbf{Y}_\ell)]}\right]$
- $\psi_{\rho,s}(\tau) \triangleq \log m_{\rho,s}(\tau)$

For meta-converse bound, use that

$$j_{\ell,s}(\mathbf{x}_\ell;\mathbf{y}_\ell) = \log \mu(s) + \frac{1}{s} i_{\ell,s}(\mathbf{x}_\ell;\mathbf{y}_\ell)$$

Saddlepoint expansions of RCU_s and MC bounds

Slightly more restrictive per-block power constraint:

$$\sum_{\ell=1}^{T} \|\mathbf{X}_{jT+\ell}\|^2 = T\rho, \quad \forall j$$

Apply saddlepoint expansion with

- $\theta = (\rho, s)$
- $m_{\rho,s}(\tau) \triangleq \mathsf{E}\left[e^{\tau[I_s(\rho) - i_{\ell,s}(\mathbf{X}_\ell;\mathbf{Y}_\ell)]}\right]$
- $\psi_{\rho,s}(\tau) \triangleq \log m_{\rho,s}(\tau)$

For meta-converse bound, use that

$$j_{\ell,s}(\mathbf{x}_\ell;\mathbf{y}_\ell) = \log \mu(s) + \frac{1}{s} i_{\ell,s}(\mathbf{x}_\ell;\mathbf{y}_\ell)$$

Numerical results: Notation

- **Saddlepoint approximations:**
  "saddlepoint MC" and "saddlepoint RCU$_s$"

- **Normal approximation "NA":**

$$R^*(L, T, \epsilon, \rho) \approx \frac{C(\rho)}{T} - \sqrt{\frac{V}{LT^2}} Q^{-1}(\epsilon)$$

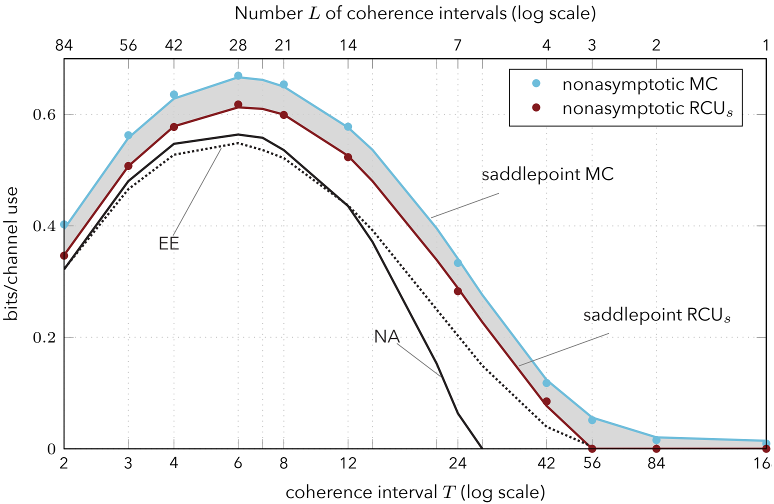
$C(\rho) = \mathsf{E}[i_{\ell,1}(\mathbf{X}_\ell; \mathbf{Y}_\ell)]$, $V = \mathsf{Var}[i_{\ell,1}(\mathbf{X}_\ell; \mathbf{Y}_\ell)]$

- **Error-exponent approximation "EE":**

$$\text{Solve} \quad \epsilon^*(L, T, R, \rho) = e^{-L[\tau\psi'_{\rho,1/(1+\tau)} - \psi_{\rho,1/(1+\tau)}]} \quad \text{for } R$$

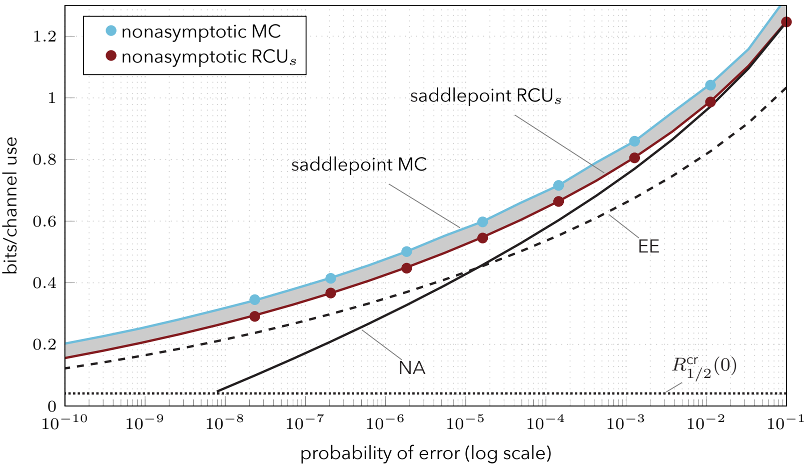
where τ is such that $\frac{1}{T}\left(I_{1/(1+\tau)}(\rho) - \psi'_{\rho,1/(1+\tau)}\right) = R$

Numerical results: LT fixed



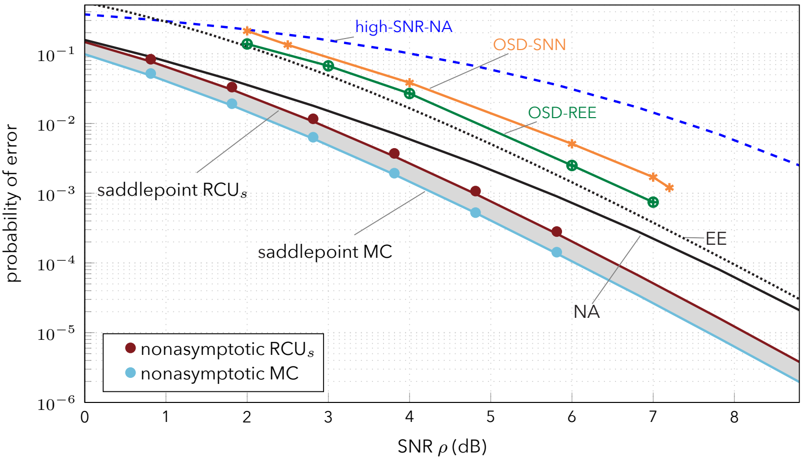
$LT = 168$, $\rho = 6$ dB, $\epsilon = 10^{-5}$, $M_t = M_r = 1$

Numerical results: $\epsilon \mapsto R^*(L, T, \epsilon, \rho)$



$L = 14$, $T = 12$, $\rho = 6$ dB, $M_t = M_r = 1$

Numerical results: $\rho \mapsto \epsilon^*(L, T, \epsilon, \rho)$



$$L = 7,\ T = 24,\ R = 0.48,\ M_t = M_r = 1$$

J. Östman *et al.*, “Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?” *IEEE Trans. Commun.*, Feb. 2019.

Discussion: Saddlepoint approximations

Saddlepoint approximations versus nonasymptotic bounds

- Complexity:
  - **Saddlepoint approx.:** compute $I_s(\rho)$, $\psi_{\rho,s}$, $\psi'_{\rho,s}$, $\psi''_{\rho,s}$, $\psi'''_{\rho,s}$
  - **Nonasymp. bound:** compute $I_s(\rho)$ and Pr $\left(\sum_{\ell=1}^{L} i_{\ell,s}(\mathbf{X}_\ell;\mathbf{Y}_\ell) \geqslant \gamma\right)$
- Accuracy:
  - Saddlepoint approx. are indistinguishable from nonasymp. bounds

Saddlepoint approx. are easy-to-compute alternatives to nonasymp. bounds

Saddlepoint approximations versus other asymptotics expansions

- **Normal approx.:** go-to choice for suff. large SNR and error probabilities
  - *For example:* $\rho = 6$ dB and $\epsilon \geqslant 10^{-3}$
- **Error exponents:** go-to choice for suff. small SNR and error probabilities
  - *For example:* $\rho = 6$ dB and $\epsilon \leqslant 10^{-7}$
- **Saddlepoint approx.:** accurate over entire range of system parameters

Saddlepoint approximations for non-i.i.d. codebooks

- Standard saddlepoint approximation applies to sum of i.i.d. RVs

- $j_{1,s}(\mathbf{X}_1; \mathbf{Y}_1), \ldots, j_{L,s}(\mathbf{X}_L; \mathbf{Y}_L)$ are only i.i.d. if $\mathbf{X}_1, \ldots, \mathbf{X}_L$ are
  (we sidestepped this issue by changing the power constraint)

- Saddlepoint approx. can be generalized to independent RVs:
  - ➔ moment generating function: $m_{k,\theta}(\zeta) = \mathsf{E}\left[e^{\zeta Z_{k,\theta}}\right]$
  - ➔ cumulant generating function: $\psi_{k,\theta}(\zeta) = \log m_{k,\theta}(\zeta)$

$$\rightarrow \quad \bar{\psi}_{n,\theta}(\zeta) = \frac{1}{n}\sum_{k=1}^{n} \psi_{k,\theta}(\zeta)$$

  - ➔ replace in saddlepoint approximation $\psi_\theta(\tau)$ by $\bar{\psi}_{n,\theta}(\tau)$

Information Theory for Low-Latency Wireless Communications

High-SNR normal approximation

- ✔ Available in closed-form
- ✔ Accurate for SNR $\geqslant 15$ dB, $L \geqslant 10$, and moderate ϵ
- ✘ Inaccurate for SNR < 15 dB, $L < 10$, and small ϵ

➔ **useful performance benchmark (where accurate)**

➔ **proxy for $R^*(L, T, \epsilon, \rho)$**

Saddlepoint approximation

- ✘ Must be computed numerically
- ✔ Computational cost low and independent of L
- ✔ Very accurate over entire range of system parameters

➔ **easy-to-compute alternative to nonasymptotic bounds**

➔ **starting point for more refined approximations**