Finite-Blocklength Wireless Communications:

From Nonasymptotic Bounds to Normal Approximations for MIMO Fading

Tobias Koch

Universidad Carlos III de Madrid

Workshop on Information Theory for Future Networks (March 3, 2026)

Collaborators

Giuseppe Durisi
Full Professor, Chalmers University of Technology

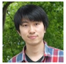
Alejandro Lancho
Cesar Nombela Research Fellow, UC3M

Johan Östman
Research Scientist, Zenseact / AI Sweden

Yury Polyanskiy
Professor, MIT

Chao Qi
Associate Professor, Xihua University

Gonzalo Vazquez-Vilar
Associate Professor, UC3M



Wei Yang
Senior Engineer, Qualcomm

Transmission of Short Packets

- Internet of Things
- Machine-to-machine communications
- tactile Internet
- ...

Short packets because...

→ reduced latency (latency $\propto$ packet length $\times$ baud rate)

→ efficient bandwidth usage

→ energy efficiency

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

IT Contributions to Wireless Communications

Traditionally, capacity / outage capacity of fading channels

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 vs. 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):

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

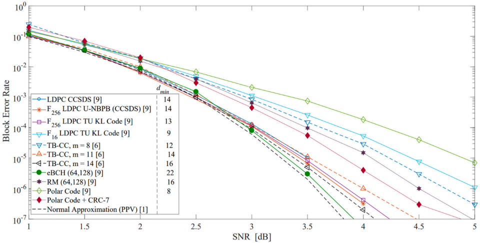
Y. Polyanskiy, H.V. Poor, S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, May 2010.

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)

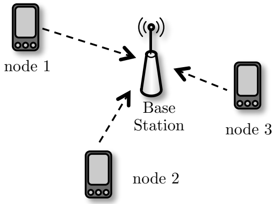
Note that coding techniques for URLLC need to be benchmarked against the PPV limit as the Shannon limit does not apply for such messages.

(Screenshot: `medium.com/5g-nr` $\rightarrow$ 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 $n_\mathsf{s} = n/\mathsf{s}$ 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



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

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

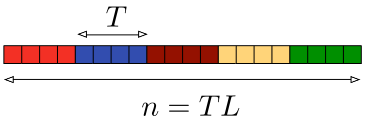
G. Durisi, T. Koch, P. Popovski, “Towards massive, ultrareliable, and low-latency wireless communication with short packets,” *Proc. IEEE*, Sep. 2016.

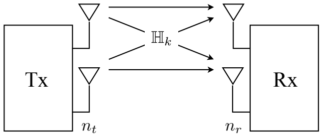
Finite-Blocklength Wireless Communications

- $R^*(n, \epsilon)$ 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**

Rayleigh Block-Fading Channel

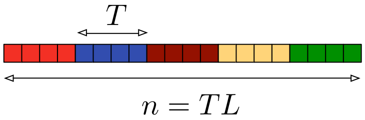


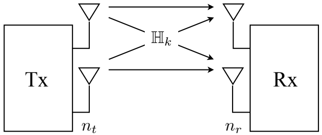


$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{W}_k, \quad k \in \mathbb{Z}$$

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

Channel State Information (CSI): Coherent vs. Noncoherent Settings





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

- **CSI@Tx:** $\{\mathbb{H}_k\}$ available at transmitter
  - requires feedback
  - may be unrealistic in high-mobility scenarios or time-critical applications
- **CSI@Rx:** $\{\mathbb{H}_k\}$ available at receiver
  - ignores the cost of obtaining CSI
  - lack of CSI@Rx does not preclude channel estimation
    ➔ pilot-aided channel estimation one possible coding scheme

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

Per-block power constraint:

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

Error probability:

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

or

$$\Pr\left(\widehat{B}^K \neq B^K\right) \leqslant \epsilon$$

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

Maximum coding rate:

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

Finite-Blocklength Wireless Communications

Nonasymptotic behavior of $R^*(L, T, \epsilon, \rho)$

Estimate $R^*(L, T, \epsilon, \rho)$ by means of bounds:

✔ very accurate

✘ must be evaluated numerically (time consuming, not in closed form)

Saddlepoint approximations of $R^*(L, T, \epsilon, \rho)$

✔ almost as accurate as bounds

✘ must be evaluated numerically (not in closed form)

✔ can be computed more efficiently

Asymptotic behavior of $R^*(L, T, \epsilon, \rho)$

Estimate $R^*(L, T, \epsilon, \rho)$ by means of asymptotic expansions (normal approximations,...):

✔ available in closed form

✘ only accurate in certain regimes

Bounds on $R^*(L, T, \epsilon, \rho)$ (1)

FBL bounds for noncoherent MIMO Rayleigh fading (DKOPY16):

- **DT lower bound**

$$R^*(L, T, \epsilon, \rho) \geqslant \max\left\{\frac{K}{TL}\colon \min_{1\leqslant \tilde{n}_t \leqslant n_t} \mathsf{E}\left[e^{-\left(\sum_{\ell=1}^{L} S_{\ell,\tilde{n}_t}(\mathbb{Z}_\ell) - \log(e^K - 1)\right)_+}\right] \leqslant \epsilon\right\}$$

- **MC upper bound**

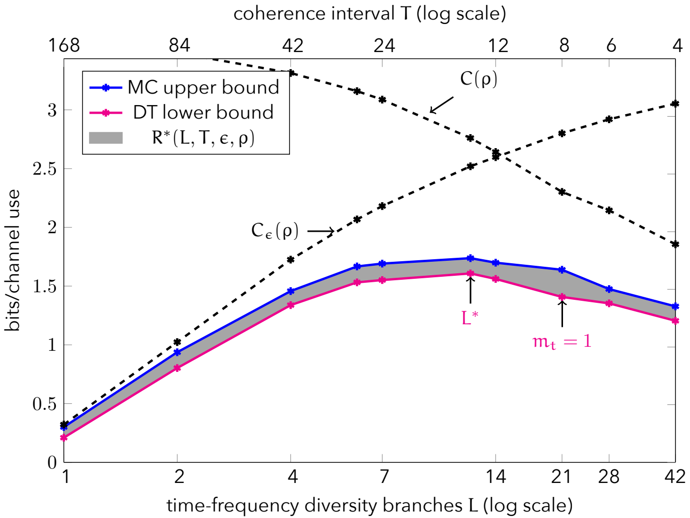
$$R^*(L, T, \epsilon, \rho) \leqslant \min_{1\leqslant \tilde{n}_t \leqslant n_t} \sup_{\Sigma_\ell} \frac{1}{n}\left(\gamma - \log\left(\Pr\left(\sum_{\ell=1}^{L} \bar{S}_{\ell,\tilde{n}_t}(\Sigma_\ell, \mathbb{Z}_\ell) \leqslant \gamma\right) - \epsilon\right)\right)$$

➔ $S_{\ell,\tilde{n}_t}(\mathbb{Z}_\ell)$, $\bar{S}_{\ell,\tilde{n}_t}(\Sigma_\ell, \mathbb{Z}_\ell)$: depend on $T \times n_r$ random matrix $\mathbb{Z}_\ell$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries

➔ difficult to analyze but can be computed using Monte Carlo integration

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

Bounds on $R^*(L, T, \epsilon, \rho)$ (2)



$n = TL = 168$

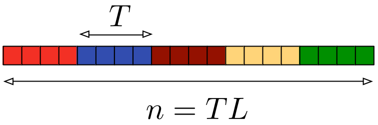
$\rho = 6$ dB

$n_t = n_r = 2$

$\epsilon = 10^{-5}$

Normal Approximation: Quasistatic Fading

Quasistatic fading: Fix L and let $T \to \infty$



Normal approximation for quasistatic fading (YDKP14):

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

➔ $C_\epsilon(\rho)$: outage capacity

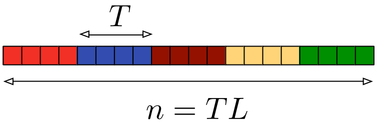
➔ holds irrespective of availability of CSI

➔ fast convergence to outage capacity

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

Normal Approximation: Ergodic Fading & CSI@Rx

Ergodic fading: Fix T and let $L \to \infty$



Normal approximation for ergodic fading & CSI@Rx (CP14):

$$R^*(L, T, \epsilon, \rho) = \mathsf{E}\left[\log\det\left(\mathsf{I} + \frac{\rho}{n_t}\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right] - \sqrt{\frac{V_c(\rho)}{L}}\, Q^{-1}(\epsilon) + o_L(L)$$

➔ $V_c(\rho)$: channel dispersion

➔ at high SNR: $V_c(\rho) \approx \frac{n_t}{T} + \mathsf{Var}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right]$

A. Collins and Y. Polyanskiy, "Coherent multiple-antenna block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, Jan. 2019.

Ergodic Fading: The Noncoherent Case

Without CSI@Tx & CSI@Rx, obtaining a normal approximation is difficult!

Standard strategy to obtain normal approximations:

1. Evaluate nonasymptotic achievability bound (RCU, DT, κ-β, ...) for capacity-achieving input/output distribution.
2. Evaluate nonasymptotic converse bound (meta converse) for capacity-achieving output distribution.
3. Perform asymptotic analysis as $n \to \infty$.

Problem: Capacity-achieving input/output distribution is unknown

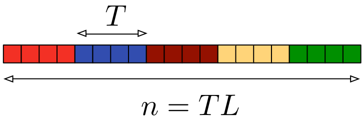
However: capacity at high-SNR is well understood

➔ unitary space-time modulation (USTM)

➔ high-SNR normal approximation feasible

Unitary Space-Time Modulation (USTM)

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



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

Theorem (HM00, ZT02):

When $T \geqslant n_t + n_r$, 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.

A High-SNR Normal Approximation for MIMO Fading

Ergodic fading: Fix T and let $L \to \infty$

High-SNR normal approximation for ergodic fading (QK26):

Assume that $T \geqslant n_r + n_t$ and $n_r \geqslant n_t$. Then,

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

where

$$\underline{R}_{\mathsf{USTM}}(\rho) = n_t\left(1 - \frac{n_t}{T}\right)\log\frac{T\rho}{n_t e} + \left(1 - \frac{n_t}{T}\right)\mathsf{E}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right] + \frac{1}{T}\log\frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)}$$

$$V = \frac{n_t}{T}\left(1 - \frac{n_t}{T}\right) + \left(1 - \frac{n_t}{T}\right)^2 \mathsf{Var}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right]$$

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

C. Qi and T. Koch, "On noncoherent multiple-antenna Rayleigh block-fading channels at finite blocklength," *IEEE Transactions on Information Theory*, March 2026.

Coherent vs. Noncoherent Fading Channels

High-SNR normal approximations for Rayleigh block-fading:

$$R^*(L, T, \epsilon, \rho) \approx C_{\{c,nc\}}(\rho) - \sqrt{\frac{V_{\{c,nc\}}(\rho)}{L}} Q^{-1}(\epsilon)$$

noncoherent setting (no CSI)

$$C_{nc}(\rho) \approx n_t \left(1 - \frac{n_t}{T}\right) \log \frac{\rho}{n_t} + \left(1 - \frac{n_t}{T}\right) \mathsf{E}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right] + \mathsf{K}(T, n_t)$$

$$V_{nc}(\rho) = \frac{n_t}{T}\left(1 - \frac{n_t}{T}\right) + \left(1 - \frac{n_t}{T}\right)^2 \mathsf{Var}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right]$$

coherent setting (CSI@Rx)

$$C_{c}(\rho) \approx n_t \log \frac{\rho}{n_t} + \mathsf{E}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right]$$

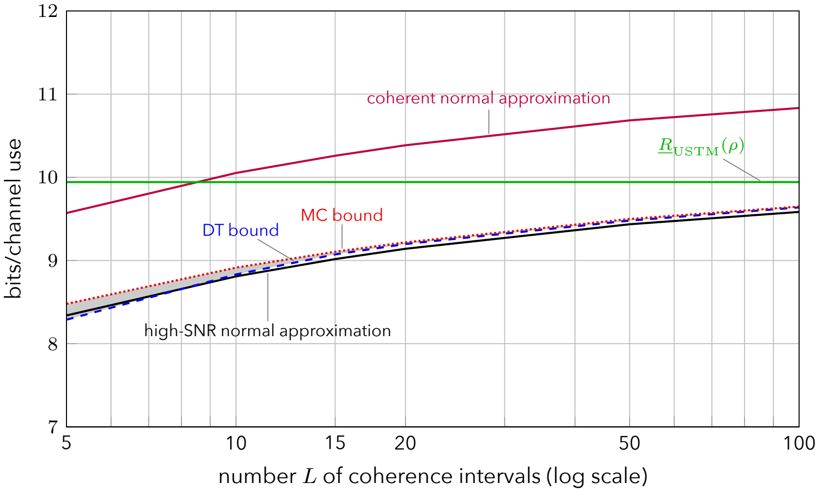
$$V_{c}(\rho) \approx \frac{n_t}{T} + \mathsf{Var}\left[\log\det\left(\mathbb{H}_\ell \mathbb{H}_\ell^{\mathsf{H}}\right)\right]$$

Heuristic:

→ **pre-log term of $C_{nc}(\rho)$ and channel dispersion $V_{nc}(\rho)$:**
- use coherent channel $(T - n_t)$ times per coherence interval

→ **transmit one pilot symbol per coherence interval and transmit antenna**

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



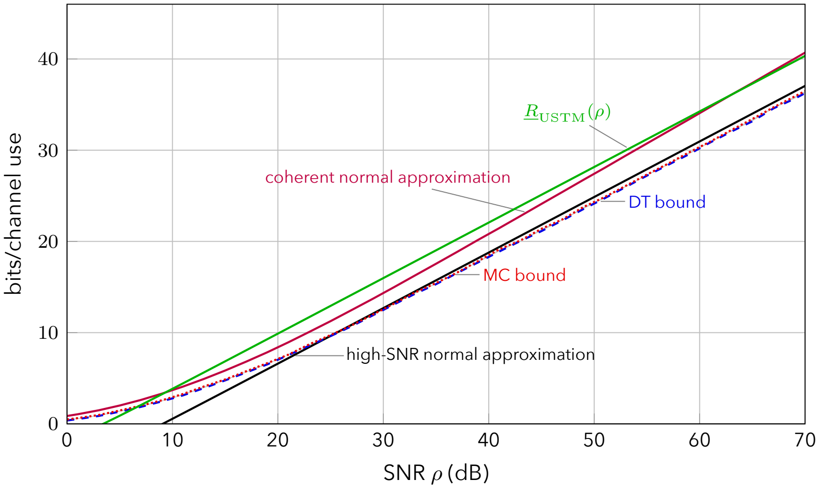
$T = 24$

$\rho = 15$ dB

$n_t = 2, n_r = 4$

$\epsilon = 10^{-3}$

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



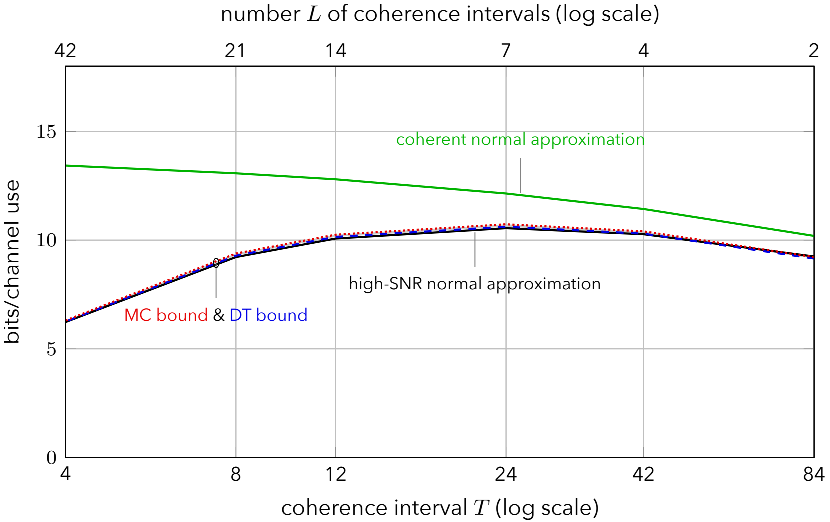
$T = 24$

$L = 7$

$n_t = n_r = 2$

$\epsilon = 10^{-5}$

Numerical Results: $n = LT$ Fixed



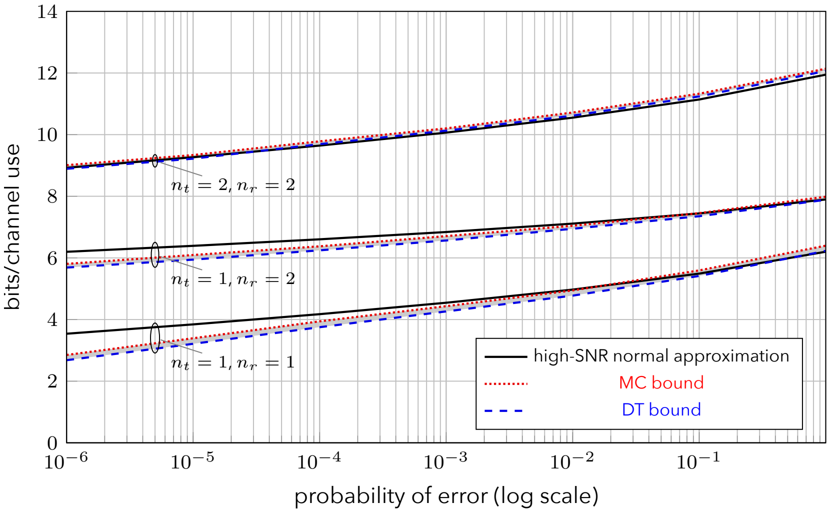
$n = LT = 168$

$\rho = 25$ dB

$n_t = n_r = 2$

$\epsilon = 10^{-3}$

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



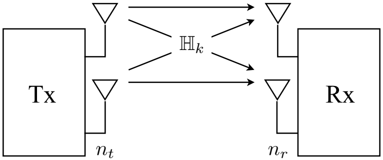
$T = 24$

$L = 7$

$\rho = 25$ dB

Diversity vs. Multiplexing at Finite Blocklength (1)

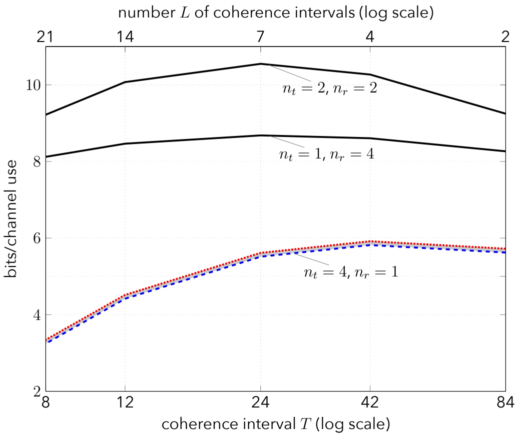
- **Spatial multiplexing:**
  - use antennas to increase rate for a given prob. of error
- **Spatial diversity:**
  - use antennas to decrease prob. of error for a given rate
- **"Modern systems use multiplexing only" (LJ10)**
  - abundant time- and frequency-selectivity available
  - spatial diversity is superfluous
  - based on outage capacity (relevant if $n \to \infty$)



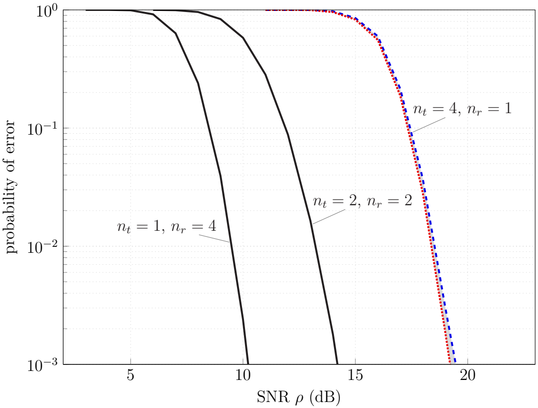
→ For short-packet transmissions, time- & frequency-selectivity is limited...

A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexity in modern MIMO systems," *IEEE Transactions on Communications*, January 2010.

Diversity vs. Multiplexing at Finite Blocklength (2)

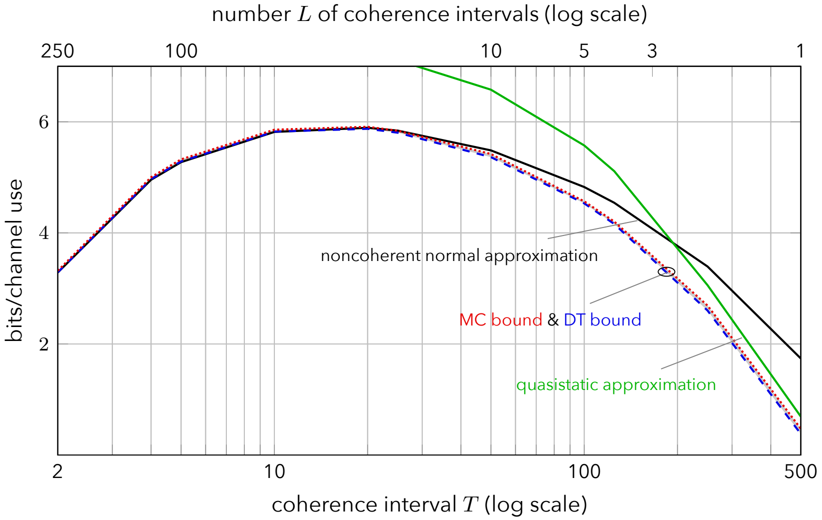


$n = TL = 168$, $\rho = 25$ dB, $\epsilon = 10^{-3}$



$R = 4$, $T = 24$, $L = 7$

Ergodic vs. Quasistatic Fading



$n = LT = 500$

$\rho = 25$ dB

$n_t = n_r = 1$

$\epsilon = 10^{-3}$

(High-SNR) Normal Approximations in a Nutshell

$$R^*(L, T, \epsilon, \rho) = \begin{cases} C_\epsilon(\rho) + o_T(T) o_T(T) & \text{quasi-static fading} \\ C(\rho) - \sqrt{\frac{V(\rho)}{L}} Q^{-1}(\epsilon) + o_L(L) o_L(L), & \text{ergodic fading} \end{cases}$$

✔ Available in closed-form

✔ Normal approximations for quasi-static and ergodic fading are complementary

- useful performance benchmark
- allow for analytical studies of tradeoffs in communication systems / performance of protocols

✘ Inaccurate for small ϵInaccurate for small ϵ

✘ High-SNR normal approximations are inaccurate for small SNR values

Finite-Blocklength Wireless Communications (again)

Nonasymptotic behavior of $R^*(L, T, \epsilon, \rho)$

Estimate $R^*(L, T, \epsilon, \rho)$ by means of bounds:

✔ very accurate

✘ must be evaluated numerically (time consuming, not in closed form)

Saddlepoint approximations of $R^*(L, T, \epsilon, \rho)$

✔ almost as accurate as bounds

✘ must be evaluated numerically (not in closed form)

✔ can be computed more efficiently

Asymptotic behavior of $R^*(L, T, \epsilon, \rho)$

Estimate $R^*(L, T, \epsilon, \rho)$ by means of asymptotic expansions (normal approximations,...):

✔ available in closed form

✘ only accurate in certain regimes

Asymptotic Expansions in Finite-Blocklength Information Theory

Finite-blocklength bounds: evaluate

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

for the i.i.d. random variables $Z_1, \ldots, Z_n$

Asymptotic expansions: perform asymptotic analysis of $\Pr\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_\ell \geqslant \gamma\right)$ as $n \to \infty$:

- Central limit theorem ➔ **normal approximation**
- Large deviations (exponential tilting + Chernoff bound) ➔ **error exponents**

- Exponential tilting + central limit theorem ➔ **saddlepoint expansion**

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:* $\psi(\zeta) = \log m(\zeta)$

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

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

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

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

where

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

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

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 $1/\sqrt{n}$

A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for short-packet wireless communications," IEEE Transactions on Wireless Communications, July 2020.

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

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

where

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

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

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

Lower Bound on Minimum Error Probability

Meta-converse bound (PPV10):

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

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

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

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

Numerical Results: Notation

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

- **Normal approximation "NA":**

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

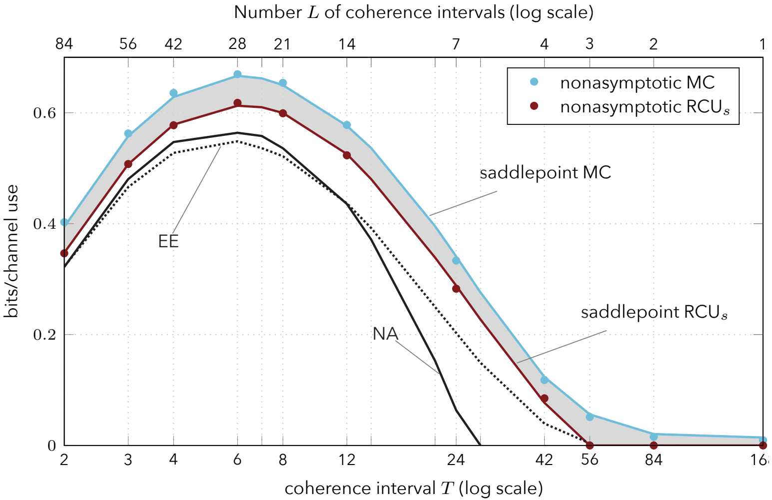
$C(\rho) = \mathsf{E}[i_{\ell,1}(\mathbf{X}_\ell;\mathbf{Y}_\ell)]$, $V = \mathsf{Var}[i_{\ell,1}(\mathbf{X}_\ell;\mathbf{Y}_\ell)]$

- **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: $n = LT$ Fixed



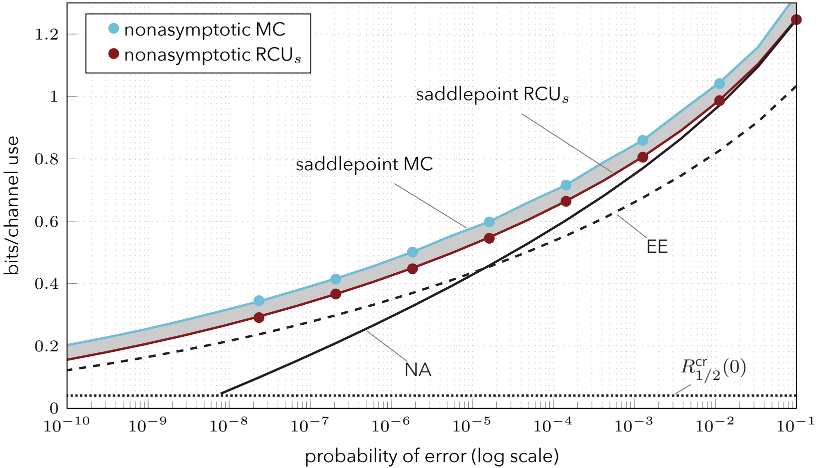
$n = LT = 168$

$\rho = 6$ dB

$n_t = n_r = 1$

$\epsilon = 10^{-5}$

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



$L = 14$

$T = 12$

$\rho = 6$ dB

$n_t = n_r = 1$

Saddlepoint Approximations in a Nutshell

Apply the saddlepoint approximation to finite-blocklength bounds:

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

Saddlepoint approximations versus nonasymptotic bounds

➔ **Complexity:**

- **Saddlepoint approximation:** compute $I_s(\rho)$, $\psi_{\rho,s}$, $\psi'_{\rho,s}$, $\psi''_{\rho,s}$, $\psi'''_{\rho,s}$
- **Nonasymptotic 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 approximations are indistinguishable from nonasymptotic bounds

Saddlepoint approximations are "easy-to-compute" alternatives to nonasymptotic bounds

Finite-Blocklength Wireless Communications

Nonasymptotic bounds

- ✔ Very accurate over entire range of parameters
- ✘ Must be computed numerically
- ✘ Computational cost high & grows with L

(High-SNR) normal approximations

- ✔ Available in closed-form
- ✔ Accurate for moderate SNR values and ϵ
- ✘ Inaccurate for small SNR values and ϵ

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

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

Saddlepoint approximations

- ✘ Must be computed numerically
- ✔ Computational cost low & independent of L
- ✔ Accurate over entire range of parameters

➔ **"easy-to-compute" alternatives to nonasymptotic bounds**

➔ **starting point for more refined approximations**