Short packet transmissions

Giuseppe Durisi

Chalmers, Sweden

October, 2017
Outline

1. Motivation: short packets and information theory
2. Foundations: nonasymptotic information theory
3. Application: short packets over MIMO fading channels
4. Application: minimization of energy per bit
Outline

1. **Motivation: short packets and information theory**
2. **Foundations: nonasymptotic information theory**
3. **Application: short packets over MIMO fading channels**
4. **Application: minimization of energy per bit**
Thanks to my collaborators

at Chalmers...

...and outside
Machine-type communications (MTC)

Key enabler of future autonomous systems

- **5G** ⇒ massive MTC; ultra-reliable, low-latency comm., narrow-band IoT
- **Low-power wireless-area networks** ⇒ LoRa-WAN, SigFox, ...
Unique characteristics of MTC traffic

- massive number of connected terminals
- transmitters are often idle
- short data packets
- low latency, high reliability
- high energy efficiency
- security and privacy
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

MTC and short packet transmission
The need for a nonasymptotic analysis
A motivating example: the AWGN channel

Example

Long-term evolution (4G)
- Long packets (500 bytes)
- Packet error probability of $10^{-1}$ at 5ms latency
- High reliability through retransmissions (HARQ)

MTC for factory automation
- Short packets: 100 bits of payload
- maximum delay of 100µs
- packet error probability in the range $[10^{-5}, 10^{-9}]$
Example

**Long-term evolution (4G)**
- Long packets (500 bytes)
- Packet error probability of $10^{-1}$ at 5ms latency
- High reliability through retransmissions (HARQ)

**MTC for factory automation**
- Short packets: 100 bits of payload
- Maximum delay of 100µs
- Packet error probability in the range $[10^{-5}, 10^{-9}]$

We need a fundamental paradigm shift in the design of wireless communication ⇒ new theoretical tools
Large-packet asymptotics

Claude E. Shannon (1916–2001)

The bit-pipe approximation

Claude E. Shannon (1916–2001)
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

MTC and short packet transmission
The need for a nonasymptotic analysis
A motivating example: the AWGN channel

Large-packet asymptotics

The bit-pipe approximation

Claude E. Shannon (1916–2001)

Used everywhere beyond PHY
- resource allocation & user scheduling
- delay analyses at the network level
Large-packet asymptotics

Claude E. Shannon (1916–2001)

The bit-pipe approximation

Used everywhere beyond PHY

- resource allocation & user scheduling
- delay analyses at the network level

If packet are shorts, bit-pipe approximation is not accurate!
This tutorial

The problem
Understand the tradeoff between latency, reliability and throughput when transmitting short packets

The tool
Finite-blocklength information theory

The outcome
Guidelines for the design of short-packet communication systems
What is latency?

A difficult quantity to define precisely
What is latency?

A difficult quantity to define precisely

End-to-end delay in access network

- encoding and decoding delays
- queuing delays
- transmission delays
What is latency?

A difficult quantity to define precisely

End-to-end delay in access network

- encoding and decoding delays
- queuing delays
- transmission delays
Packet size as a proxy for latency

- **propagation delay:**
  - distance 180 m $\Rightarrow$ delay of 30$\mu$s

- **packet duration:**
  - Smallest packet size in 4G: 14 OFDM symbols, 12 subcarriers
  - $n = 168$, duration = 1ms $\gg$ propagation delay
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Principles of coding theory

- More redundancy $\Rightarrow$ lower packet error probability $\epsilon$...
- ...but also lower transmission rate $R = k/n$

What is the tradeoff between $\epsilon$ and $R$ for fixed $n$?
A brief introduction to finite-blocklength IT

![Graph showing packet error probability $\epsilon$ versus rate $R$]

- Possible for $\epsilon^*(n, R)$
- Not possible

Beginning of last century
A brief introduction to finite-blocklength IT

1948: Shannon, channel capacity
A brief introduction to finite-blocklength IT

Vertical asymptotics \( \Rightarrow \) error exponent

(Gallager, \ldots )
A brief introduction to finite-blocklength IT

Horizontal asymptotics $\Rightarrow$ strong converse, fixed-error asymptotics
(Wolfowitz, Strassen,\ldots)
A brief introduction to finite-blocklength IT

Today: tight computationally-feasible bounds and accurate approximations
(Hayashi 2009, Polyanskiy, Poor and Verdú 2010, …)
The AWGN channel model

- $n$: blocklength (size of coded packet)
The AWGN channel model

- $n$: blocklength (size of coded packet)
- $\epsilon = \mathbb{P}[\widehat{W} \neq W]$: packet error probability

Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

MTC and short packet transmission
The need for a nonasymptotic analysis
A motivating example: the AWGN channel
The AWGN channel model

- $n$: blocklength (size of coded packet)
- $\epsilon = \mathbb{P}[\hat{W} \neq W]$: packet error probability
- $\|X^n\|^2 \leq n\rho$, where $\rho$ is the power constraint
The AWGN channel model

- $n$: blocklength (size of coded packet)
- $\epsilon = \mathbb{P}[\hat{W} \neq W]$: packet error probability
- $\|X^n\|^2 \leq n\rho$, where $\rho$ is the power constraint

Maximum coding rate $R^*(n, \epsilon)$

Largest rate among all codes with blocklength $n$ and packet error probability $\epsilon$
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

MTC and short packet transmission
The need for a nonasymptotic analysis
A motivating example: the AWGN channel

AWGN channel: $\rho = 0 \text{ dB}, \epsilon = 10^{-3}$

Channel capacity: $C_{\text{awgn}}(\rho) = \frac{1}{2} \log(1 + \rho)$
AWGN channel: \( \rho = 0 \text{ dB}, \ \epsilon = 10^{-3} \)
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

A motivating example: the AWGN channel

**AWGN channel:** $\rho = 0 \text{ dB}, \epsilon = 10^{-3}$

![Graph showing channel dispersion](image)

**Channel dispersion:**

$$V_{\text{awgn}}(\rho) = \frac{\rho(2 + \rho)}{2(1 + \rho)^2}$$
\[ R^*(n, \epsilon) \approx C_{\text{awgn}}(\rho) - \sqrt{\frac{V_{\text{awgn}}(\rho)}{n}} Q^{-1}(\epsilon) + \frac{1}{2} \frac{\log n}{n} \]
Probability of Error for Optimal Codes in a Gaussian Channel

By CLAUDE E. SHANNON

(Manuscript received October 17, 1958)

A study is made of coding and decoding systems for a continuous channel with an additive gaussian noise and subject to an average power limitation at the transmitter. Upper and lower bounds are found for the error probability in decoding with optimal codes and decoding systems. These bounds are close together for signaling rates near channel capacity and also for signaling rates near zero, but diverge between. Curves exhibiting these bounds are given.

Hard to compute!
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

So what is new?

Probability of Error for Optimal Codes
in a Gaussian Channel

By CLAUDE E. SHANNON
(Manuscript received October 17, 1955)

A study is made of coding and decoding systems for a continuous channel
with an additive gaussian noise and subject to an average power limitation
at the transmitter. Upper and lower bounds are found for the error prob-
ability in decoding with optimal codes and decoding systems. These bounds
are close together for signaling rates near channel capacity and also for sig-
naling rates near zero, but diverge between. Curves exhibiting these bounds
are given.

Hard to compute!

The “modern” bounds are

- computationally feasible
So what is new?

Probability of Error for Optimal Codes in a Gaussian Channel

By CLAUDE E. SHANNON
(Manuscript received October 17, 1958)

A study is made of coding and decoding systems for a continuous channel with an additive Gaussian noise and subject to an average power limitation at the transmitter. Upper and lower bounds are found for the error probability in decoding with optimal codes and decoding systems. These bounds are close together for signaling rates near channel capacity and also for signaling rates near zero, but diverge between. Curves exhibiting these bounds are given.

Hard to compute!

The “modern” bounds are

- computationally feasible
- general: bi-AWGN, fading channels, exponential-noise channels, etc...

```
function rate=metaconverse_rate(n,epsilon,P) %metaconverse bound
    gammatildeP=ncx2inv(1-epsilon,n,n/P);
    gammatildeQ=(gammatildeP-n*(1+1/P))/(1+P) + n/P;
    beta=ncx2cdf(gammatildeQ,n,n*(1+1/P));
    rate=-log(beta)/n*log(2);
end
```
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

So what is new?

Probability of Error for Optimal Codes in a Gaussian Channel
By CLAUDE E. SHANNON
(Manuscript received October 17, 1958)

A study is made of coding and decoding systems for a continuous channel with an additive gaussian noise and subject to an average power limitation at the transmitter. Upper and lower bounds are found for the error probability in decoding with optimal codes and decoding systems. These bounds are close together for signaling rates near channel capacity and also for signaling rates near zero, but diverge between. Curves exhibiting these bounds are given.

Hard to compute!

The “modern” bounds are
- computationally feasible
- general: bi-AWGN, fading channels, exponential-noise channels, etc...
- easy to approximate accurately

function rate=metaconverse_rate(n,epsilon,P) %metaconverse bound
gammatildeP=ncx2inv(1-epsilon,n,n/P);
gammatildeQ=(gammatildeP-n*(1+1/P))/(1+P) + n/P;
betatilde=ncx2cdf(gammatildeQ,n,n*(1+1/P));
rate=-log(betatilde)/(n*log(2));
end
Actual codes: bi-AWGN, $\epsilon = 10^{-4}$

![Graph showing the performance of various codes over different blocklengths.](image-url)
Outline

1. Motivation: short packets and information theory
2. Foundations: nonasymptotic information theory
3. Application: short packets over MIMO fading channels
4. Application: minimization of energy per bit
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Outline

- Review of information-theoretic asymptotic metrics...
- ... and of hypothesis testing
- The metaconverse theorem (upper bound) and its proof
- Achievability (lower) bounds
- The normal approximation
Shannon's coding theorem

The capacity of a stationary memoryless channel $P_{Y|X}$ is

$$C = \sup_{P_X} I(X; Y)$$
Mutual information and Shannon’s coding theorem

**Shannon’s coding theorem**

The capacity of a stationary memoryless channel $P_Y|X$ is

$$C = \sup_{P_X} I(X; Y)$$

**Mutual information**

$$I(X; Y) = \mathbb{E}_{X,Y} \left[ \log \frac{P_Y|X(Y|X)}{P_Y(Y)} \right], \quad X, Y \sim P_Y|X P_X$$

$$= D(P_Y|X P_X \parallel P_Y P_X)$$

Information density, relative entropy and the Golden formula
Binary-hypothesis testing
The metaconverse theorem
Achievability bounds and normal approximation
Upper bounds on $C$ through duality

Upper bound through the golden formula

$$I(X;Y) = D(P_Y |_{X} P_X \| P_Y P_X)$$

Upper bounds on $C$ through duality

Upper bound through the golden formula

\[
I(X;Y) = D(P_{Y|X}P_{X} \ || \ P_{Y}P_{X}) \\
= D(P_{Y|X}P_{X} \ || \ Q_{Y}P_{X}) - D(P_{Y} \ || \ Q_{Y}) \text{ for all } Q_{Y}
\]

Upper bounds on $C$ through duality

Upper bound through the golden formula

\[ I(X; Y) = D(P_Y | X P_X \parallel P_Y P_X) \]
\[ = D(P_Y | X P_X \parallel Q Y P_X) - D(P_Y \parallel Q Y) \text{ for all } Q_Y \]
\[ \leq D(P_Y | X P_X \parallel Q Y P_X) \]

Upper bounds on $C$ through duality

Upper bound through the golden formula

\[
I(X; Y) = D(P_Y | X P_X \| P_Y P_X) \\
= D(P_Y | X P_X \| Q_Y P_X) - D(P_Y \| Q_Y) \text{ for all } Q_Y \\
\leq D(P_Y | X P_X \| Q_Y P_X) \\
= \mathbb{E}_{X,Y} \left[ \log \frac{P_{Y|X}(Y|X)}{Q_Y(Y)} \right] \text{ linear in } P_X!
\]

Binary-hypothesis testing

\[
P_{X^n} \xrightarrow{?} X^n \xrightarrow{\text{test}} Z
\]

- \( Z \in \{0, 1\} \)
- \( Z = 0 \Rightarrow \text{test chooses } P_{X^n} \)

\[
Z = 0
\]

\[
Z = 1
\]
Optimal test

Neyman-Pearson $\beta$ function

Optimal test that minimizes error prob. under $Q_{X^n}$ given a constraint on the success prob. under $P_{X^n}$
Optimal test

\[ P_{X^n} \xrightarrow{?} X^n \xrightarrow{\text{test}} Z \]

**Neyman-Pearson \( \beta \) function**

- **Optimal test** that minimizes error prob. under \( Q_{X^n} \) given a constraint on the success prob. under \( P_{X^n} \)

\[ \beta_\alpha(P_{X^n}, Q_{X^n}) = \inf_{P_{Z|X^n} : P_{X^n}[Z=0] \geq \alpha} Q_{X^n}[Z = 0] \]
Neyman-Pearson & Stein Lemmas

Neyman-Pearson Lemma

- The optimal test involves thresholding the log-likelihood ratio

\[ \beta_\alpha(P_{X^n}, Q_{X^n}) = Q_{X^n} \left[ \log \frac{P_{X^n}}{Q_{X^n}} (X^n) \geq \gamma \right] \]

- where \( \gamma : P_{X^n} \left[ \log \frac{P_{X^n}}{Q_{X^n}} (X^n) \geq \gamma \right] = \alpha \)
Neyman-Pearson & Stein Lemmas

**Neyman-Pearson Lemma**
- The optimal test involves **thresholding** the log-likelihood ratio
- \( \beta_\alpha(P_{X^n}, Q_{X^n}) = Q_X^n \left[ \log \frac{P_{X^n}}{Q_{X^n}}(X^n) \geq \gamma \right] \)
- where \( \gamma : P_{X^n} \left[ \log \frac{P_{X^n}}{Q_{X^n}}(X^n) \geq \gamma \right] = \alpha \)

**Stein’s Lemma**
Assume that \( X^n \) has i.i.d. entries. Then \( \beta_\alpha(P_{X^n}, Q_{X^n}) \) decays to zero **exponentially fast** in \( n \)

\[ \log \beta_\alpha(P_{X^n}, Q_{X^n}) = -nD(P_X \| Q_X) + o(n) \]
The big picture

Capacity upper bound

\[ C \leq \sup_{P_X} D(P_X P_Y | X \| P_X Q_Y), \quad \text{for all } Q_Y \]
The big picture

**Capacity upper bound**

$$C \leq \sup_{P_X} D(P_X P_Y | X \| P_X Q_Y), \quad \text{for all } Q_Y$$

**A related hypothesis testing problem**

$$\log \beta_\alpha(P_{X^n} P_{Y^n} | X^n, P_{X^n} Q_{Y^n}) = -n D(P_X P_Y | X \| P_X Q_Y) + o(n)$$
The big picture

**Capacity upper bound**

\[ C \leq \sup_{P_X} D(P_X P_Y | X \parallel P_X Q_Y), \quad \text{for all } Q_Y \]

**A related hypothesis testing problem**

\[
\log \beta_\alpha(P_{X^n} P_{Y^n} | X^n, P_{X^n} Q_{Y^n}) = -nD(P_X P_Y | X \parallel P_X Q_Y) + o(n)
\]

**A nonasymptotic performance metric for short packets**

\[ \beta_\alpha(P_{X^n} P_{Y^n} | X^n, P_{X^n} Q_{Y^n}) \]
An \((n, M, \epsilon)\) code and the maximum coding rate

- **Message**: \(W \in \{1, 2, \ldots, M\}; \ k = \log M\)
- **Encoder**: maps \(W\) into \(X^n\)
- **Blocklength**: \(n\); number of inf. bits: \(\log_2 k\)
- **Decoder**: guesses \(W\) from \(Y^n\) with \(\epsilon = \Pr\{\hat{W} \neq W\}\)

\[ P_{Y^n|X^n} \]
An \((n, M, \epsilon)\) code and the maximum coding rate

- **Message:** \(W \in \{1, 2, \ldots, M\}\); \(k = \log M\)
- **Encoder:** maps \(W\) into \(X^n\)
- **Blocklength:** \(n\); number of inf. bits: \(\log_2 k\)
- **Decoder:** guesses \(W\) from \(Y^n\) with \(\epsilon = \mathbb{P}\{\hat{W} \neq W\}\)

**Maximum coding rate**

\[
R^*(n, \epsilon) = \max\{(\log M)/n : \exists (n, M, \epsilon)\text{-code}\}
\]
An \((n, M, \epsilon)\) code and the maximum coding rate

- **Message:** \(W \in \{1, 2, \ldots, M\}; k = \log M\)
- **Encoder:** maps \(W\) into \(X^n\)
- **Blocklength:** \(n\); number of inf. bits: \(\log_2 k\)
- **Decoder:** guesses \(W\) from \(Y^n\) with \(\epsilon = \mathbb{P}\{\hat{W} \neq W\}\)

**Maximum coding rate**

\[
R^*(n, \epsilon) = \max\{(\log M)/n : \exists(n, M, \epsilon)-\text{code}\}
\]

\[
C = \lim_{\epsilon \to 0} \lim_{n \to \infty} R^*(n, \epsilon)
\]
The metaconverse theorem

Theorem [Polyanskiy et al. ’10]

Every \((n, M, \epsilon)\)-code satisfies

\[
M \leq \sup_{P_{X^n}} \frac{1}{\beta_1 - \epsilon(P_{X^n}P_{Y^n} | X^n, P_{X^n}Q_{Y^n})}
\]

for all \(Q_{Y^n}\)
The metaconverse theorem

**Theorem [Polyanskiy et al. ’10]**

Every \((n, M, \epsilon)\)-code satisfies

\[
M \leq \sup_{P_{X^n}} \frac{1}{\beta_{1-\epsilon}(P_{X^n}P_{Y^n} | X^n, P_{X^n}Q_{Y^n})}
\]

for all \(Q_{Y^n}\)

**The metaconverse is tight [Vilar Vasquez et al. ’16]**

For a **given** \((n, M, \epsilon)\)-code with ML decoding

\[
M = \inf_{Q_{Y^n}} \frac{1}{\beta_{1-\epsilon}(P_{X^n}P_{Y^n} | X^n, P_{X^n}Q_{Y^n})}
\]

NP-hard problem!
Proof of metaconverse theorem (1)

- Fix $Q_{Y^n}$ and a $(n, M, \epsilon)$-code $\Rightarrow P_{X^n}$
Proof of metaconverse theorem (1)

- Fix $Q_{Y^n}$ and a $(n, M, \epsilon)$–code $\Rightarrow P_{X^n}$
- Use code as binary-hypothesis tester

$P_{Y^n|X^n} P_{X^n} \circ (X^n, Y^n) \Rightarrow P_{X^n} Q_{Y^n}$
Proof of metaconverse theorem (1)

- Fix $Q_{Y^n}$ and a $(n, M, \epsilon)$–code $\Rightarrow P_{X^n}$
- Use code as binary-hypothesis tester
  
  $P_{Y^n|X^n} P_{X^n} (X^n, Y^n) \quad ? \quad Z$

- Invert encoder to find $W$ from $X^n$
Proof of metaconverse theorem (1)

- Fix $Q_{Y^n}$ and a $(n, M, \epsilon)$–code $\Rightarrow P_{X^n}$
- Use code as binary-hypothesis tester

$$P_{Y^n|X^n} P_{X^n} \circ \ (X^n, Y^n)$$

- Invert encoder to find $W$ from $X^n$
- Apply decoder to $Y^n$ to find $\hat{W}$
Proof of metaconverse theorem (1)

- Fix $Q_{Y^n}$ and a $(n, M, \epsilon)$–code $\Rightarrow P_{X^n}$
- Use code as binary-hypothesis tester

$$P_{Y^n|X^n} P_{X^n} \circ (X^n, Y^n) \quad ? \quad P_{X^n} Q_{Y^n} \circ \quad \text{test} \quad Z$$

- Invert encoder to find $W$ from $X^n$
- Apply decoder to $Y^n$ to find $\hat{W}$
- Set $Z = 0$ if $\hat{W} = W$ and $Z = 1$ otherwise
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester

\[ P_{Y^n|X^n} P_{X^n} \] ? \( (X^n, Y^n) \) test \( Z \)

\[ P_{X^n} Q_{Y^n} \]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)

- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester

\[
P_{Y^n|X^n} P_{X^n} (X^n, Y^n) \xrightarrow{?} \text{test} \quad Z
\]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)

- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities

\[
P_{X^n} P_{Y^n} | X^n [Z = 0] \geq
\]
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester

\[
P_{Y^n|X^n} P_{X^n} \
\]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)
- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities

\[
P_{X^n} P_{Y^n} | X^n [Z = 0] \geq 1 - \epsilon
\]
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester
  
  \[ P_{Y^n|X^n}P_{X^n} \]  
  \[ (X^n, Y^n) \]  
  \[ \text{test} \]  
  \[ Z \]  
  
  \[ P_{X^n}Q_{Y^n} \]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)

- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities

  \[ P_{X^n}P_{Y^n} | X^n[Z = 0] \geq 1 - \epsilon \]

  \[ \leq P_{X^n}Q_{Y^n}[Z = 0] = \]
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester

\[ P_{Y^n | X^n} P_{X^n} \stackrel{\text{?}}{\rightarrow} (X^n, Y^n) \stackrel{\text{test}}{\rightarrow} Z \]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)

- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities

\[
P_{X^n P_{Y^n} | X^n}[Z = 0] \geq 1 - \epsilon
\]

\[
\beta_{1-\epsilon}(P_{X^n P_{Y^n} | X^n}, P_{X^n Q_{Y^n}}) \leq P_{X^n Q_{Y^n}}[Z = 0] =
\]
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester
  \[ P_{Y^n|X^n} P_{X^n} \]
  \[ ? \]
  \[ (X^n, Y^n) \]
  \[ test \]
  \[ Z \]
  \[ P_{X^n} Q_{Y^n} \]

- \( W \leftarrow (X^n, Y^n) \rightarrow \hat{W} \)

- Set \( Z = 0 \) if \( \hat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities
  \[ P_{X^n} P_{Y^n | X^n} [Z = 0] \geq 1 - \epsilon \]
  \[ \beta_{1-\epsilon}(P_{X^n} P_{Y^n | X^n}, P_{X^n} Q_{Y^n}) \leq P_{X^n} Q_{Y^n}[Z = 0] = \frac{1}{M} \]
Proof of metaconverse theorem (2)

- Use code as binary-hypothesis tester

\[ P_{Y^n|X^n} P_{X^n} \] 

\[ (X^n, Y^n) \]

\[ P_{X^n} Q_{Y^n} \]

\[ \text{test} \]

\[ Z \]

- \( W \leftarrow (X^n, Y^n) \rightarrow \widehat{W} \)

- Set \( Z = 0 \) if \( \widehat{W} = W \) and \( Z = 1 \) otherwise

- Compute probabilities

\[ P_{X^n} P_{Y^n} | X^n [Z = 0] \geq 1 - \epsilon \]

\[ \beta_{1-\epsilon}(P_{X^n} P_{Y^n} | X^n, P_{X^n} Q_{Y^n}) \leq P_{X^n} Q_{Y^n} [Z = 0] = \frac{1}{M} \]

- Optimize over \( P_{X^n} \)
Computing the metaconverse: the AWGN channel (1)

Theorem (Polyanskiy et al. ’10)

$$M \leq \sup_{P_{X^n}} \frac{1}{\beta_1 - \epsilon(P_{X^n}P_{Y^n}|X^n, P_{X^n}Q_{Y^n})}$$

- \(Y^n = X^n + Z^n\)
Computing the metaconverse: the AWGN channel (1)

Theorem (Polyanskiy et al. ’10)

\[ M \leq \sup_{P_{X^n}} \frac{1}{\beta_1 - \epsilon(P_{X^n}P_{Y^n} | X^n, P_{X^n}Q_{Y^n})} \]

- \( Y^n = X^n + Z^n \)
- \( C = \frac{1}{2} \log(1 + \rho); \ P_X^* = \mathcal{N}(0, \rho); \ P_Y^* = \mathcal{N}(0, 1 + \rho) \)
Computing the metaconverse: the AWGN channel (1)

Theorem (Polyanskiy et al. ’10)

\[
M \leq \sup_{P_{X^n}} \frac{1}{\beta_1 - \epsilon (P_{X^n}P_{Y^n} | X^n, P_{X^n}Q_{Y^n})}
\]

- \(Y^n = X^n + Z^n\)
- \(C = \frac{1}{2} \log(1 + \rho); \ P^*_X = \mathcal{N}(0, \rho); \ P^*_Y = \mathcal{N}(0, 1 + \rho)\)
- Set \(Q_{Y^n}(Y^n) = \prod_{k=1}^{n} P^*_Y(Y_k)\)
Computing the metaconverse: the AWGN channel (1)

**Theorem (Polyanskiy et al. ’10)**

\[
M \leq \sup_{P_{X^n}} \frac{1}{\beta_{1-\epsilon}(P_{X^n} P_{Y^n} \mid X^n, P_{X^n} Q_{Y^n})}
\]

- \( Y^n = X^n + Z^n \)
- \( C = \frac{1}{2} \log(1 + \rho); \ P^*_X = \mathcal{N}(0, \rho); \ P^*_Y = \mathcal{N}(0, 1 + \rho) \)
- Set \( Q_{Y^n}(Y^n) = \prod_{k=1}^{n} P^*_Y(Y_k) \)
- By symmetry,

\[
\beta_{1-\epsilon}(P_{X^n} P_{Y^n} \mid X^n, P_{X^n} Q_{Y^n}) = \beta_{1-\epsilon}(P_{Y^n} \mid X^n = \bar{x}, Q_{Y^n})
\]

where \( \bar{x} = [\sqrt{\rho} \ldots \sqrt{\rho}] \)
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Information density, relative entropy and the Golden formula
Binary-hypothesis testing
The metaconverse theorem
Achievability bounds and normal approximation

Computing the metaconverse: the AWGN channel (1)

Theorem (Polyanskiy et al. ’10)

\[ M \leq \frac{1}{\beta_{1-\epsilon}(P_{Y|X^n = \bar{x}, Q_{Y^n}})} \]

- \( Y^n = X^n + Z^n \)
- \( C = \frac{1}{2} \log(1 + \rho); P_X^* = \mathcal{N}(0, \rho); P_Y^* = \mathcal{N}(0, 1 + \rho) \)
- Set \( Q_{Y^n}(Y^n) = \prod_{k=1}^n P_Y^*(Y_k) \)
- By symmetry,

\[ \beta_{1-\epsilon}(P_{X^n P_{Y^n} | X^n, P_{X^n} Q_{Y^n}}) = \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}, Q_{Y^n}}) \]

where \( \bar{x} = [\sqrt{\rho} \ldots \sqrt{\rho}] \)
- Use Neyman-Pearson lemma
Computing the metaconverse: the AWGN channel (2)

- $P_{Y^n | X^n = \bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n$; $Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n$
- Neyman-Pearson lemma

$$\beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) = Q_{Y^n} \left[ \log \frac{P_{Y^n | X^n = \bar{x}}(Y^n)}{Q_{Y^n}(Y^n)} \geq \gamma \right]$$

where $\gamma$:

$$P_{Y^n | X^n = \bar{x}} \left[ \log \frac{P_{Y^n | X^n = \bar{x}}(Y^n)}{Q_{Y^n}(Y^n)} \geq \gamma \right] = 1 - \epsilon$$

function rate=metaconverse_rate(n,epsilon,\rho) %metaconverse bound
gammatildeP=ncx2inv(1-epsilon,n,n/\rho);
gammatildeQ=(gammatildeP-n*(1+1/\rho))/(1+\rho) + n/\rho;
betaf=ncx2cdf(gammatildeQ,n,n*(1+1/\rho));
rate=-log(betaf)/(n*log(2));
end
Computing the metaconverse: the AWGN channel (3)
The $\beta\beta$-bound (Yang et al. 2016)

**Theorem**

For all $0 < \epsilon < 1$, for all $P_{X^n}$ and $Q_{Y^n}$ there exists an $(n, M, \epsilon)$ code satisfying

$$\frac{M}{2} \geq \sup_{0 < \tau < \epsilon} \frac{\beta_\tau(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^n}P_{Y^n} \mid X^n, P_{X^n}Q_{Y^n})}$$

**Remarks**

- Recall golden formula!
  \[ I(X; Y) = D(P_{X^n}P_{Y^n} \mid X^n, P_{X^n}Q_{Y^n}) - D(P_{Y^n} \mid Q_{Y^n}) \]

- Closely related to the $\kappa\beta$ and the DT bounds in PPV '10

\[ \text{Motivation: short packets and information theory} \]
\[ \text{Foundations: nonasymptotic information theory} \]
\[ \text{Application: short packets over MIMO fading channels} \]
\[ \text{Application: minimization of energy per bit} \]

\[ \text{Information density, relative entropy and the Golden formula} \]
\[ \text{Binary-hypothesis testing} \]
\[ \text{The metaconverse theorem} \]
\[ \text{Achievability bounds and normal approximation} \]
The $\beta\beta$-bound: core idea

- **Codebook:** $\{c_1, c_2, \ldots, c_M\}$

\[
P_{Y^n|X^n}P_{X^n} \quad \triangleleft \quad (c_j, Y^n) \quad \triangleright \quad P_{X^n}Q_{Y^n} \quad Z_j
\]

- **Set** $\hat{W} = \hat{j}$ where...

- $\ldots\hat{j}$ is the first $j$ for which $Z_j = 0$.

---

The $\kappa_\beta$-bound

**Theorem**

For every $0 < \epsilon < 1$, there exists a $(n, M, \epsilon)$ code with codewords chosen from a set $\mathcal{F}$, satisfying

$$M \geq \sup_{0 < \tau < \epsilon} \sup_{Q_{Y^n}} \frac{\kappa_\tau(\mathcal{F}, Q_{Y^n})}{\sup_{x \in \mathcal{F}} \beta_{1-\epsilon}(P_{Y^n} | X^n = x, Q_{Y^n})}$$

where $\kappa_\tau$ is the minimum probability of error under $Q_{Y^n}$ when distinguishing between $Q_{Y^n}$ the collection $\{P_{Y^n} | X^n = x\}$

- $\epsilon$ is max probability of error (rather than average)
- Useful when $\beta_{1-\epsilon}(P_{Y^n} | X^n = x, Q_{Y^n})$ takes same value for all $x \in \mathcal{F}$ (AWGN)
Dependence-testing (DT) bound

**Theorem**

For every distribution $P_{X^n}$ there exists an $(n, M, \epsilon)$ code satisfying

$$\epsilon \leq \mathbb{P} \left[ i(X^n; Y^n) \leq \log \frac{M - 1}{2} \right] + \frac{M - 1}{2} \mathbb{P} \left[ i(X^n, \bar{Y}^n) > \log \frac{M - 1}{2} \right]$$

where $i(X^n; Y^n) = \log \frac{P_{Y^n|X^n}}{P_{Y^n}}$ and

$$(X^n, Y^n, \bar{Y}^n) \sim P_{X^n}(X^n)P_{Y^n|X^n}(Y^n|X^n)P_{Y^n}(\bar{Y}^n)$$
The DT-bound: core idea

- Codebook: \( \{c_1, c_2, \ldots, c_M\} \)

\[
\begin{align*}
(c_j, Y^n) & \Rightarrow \\
Z_j &= 0 \quad \text{if } i(c_j, Y^n) > \gamma \\
Z_j &= 1 \quad \text{otherwise}
\end{align*}
\]

- Set \( \hat{W} = \hat{j} \) where...

- \( \ldots \hat{j} \) is the first \( j \) for which \( Z_j = 0 \)

Useful when \( P_{Y^n} \) is “nice” for a good \( P^n_X (XAWGN) \)
Why is the DT bound "difficult to compute" for AWGN?

- Good input distribution \([Shannon '59]\): \(X^n\) uniform over the sphere

- \(Y^n = X^n + Z^n \Rightarrow P_{Y^n}\) is complicated
Normal approximation: the AWGN case (1)

**Metaconverse for AWGN**

\[ M \leq \frac{1}{\beta_{1-\epsilon}(P_{Y^n \mid X^n=\bar{x}}, Q_{Y^n})} \]

with

\[ P_{Y^n \mid X^n=\bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n \]

\[ Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n \]
Normal approximation: the AWGN case (1)

Metaconverse for AWGN

\[ \log M \leq -\log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) \]

with

\[ P_{Y^n | X^n = \bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n \quad Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n \]

\[ \log M \leq -\log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) + o(n) \]

\[ \log M \leq n C + o(n) \]
Normal approximation: the AWGN case (1)

Metaconverse for AWGN

$$R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n | X^n=\bar{x}, Q_{Y^n}})$$

with

$$P_{Y^n | X^n=\bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n \quad Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n$$
Normal approximation: the AWGN case (1)

**Metaconverse for AWGN**

\[
R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n})
\]

with

\[
P_{Y^n | X^n = \bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n \quad Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n
\]

**Stein’s lemma**

\[
-\log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) = nD(P_Y | X = \sqrt{\rho} \parallel Q_Y) + o(n)
\]
Normal approximation: the AWGN case (1)

Metaconverse for AWGN

\[ R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n | X^n=\bar{x}}, Q_{Y^n}) \]

with

\[ P_{Y^n | X^n=\bar{x}} = \mathcal{N}(\sqrt{\rho}, 1)^n \quad Q_{Y^n} = \mathcal{N}(0, 1 + \rho)^n \]

Stein’s lemma

\[-\log \beta_{1-\epsilon}(P_{Y^n | X^n=\bar{x}}, Q_{Y^n}) = nD(P_{Y | X=\sqrt{\rho}} \parallel Q_{Y}) + o(n)\]

\[= \frac{n}{2} \log(1 + \rho) + o(n)\]

\[\underbrace{nC}_{\text{normal approximation}}\]
Normal approximation: the AWGN case (2)

Metaconverse for AWGN

\[ R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n \mid X^n=\bar{x}}, Q_{Y^n}) \]

Stein’s lemma

\[ -\log \beta_{1-\epsilon}(P_{Y^n \mid X^n=\bar{x}}, Q_{Y^n}) = nC + o(n) \]

Approximation

\[ R^*(n, \epsilon) = C + o(1) \]
Normal approximation: the AWGN case (2)

Metaconverse for AWGN

\[ R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) \]

Stein’s lemma + Berry-Esseen CLT

\[ -\log \beta_{1-\epsilon}(P_{Y^n | X^n = \bar{x}}, Q_{Y^n}) = nC - \sqrt{nV} Q^{-1}(\epsilon) + o(\sqrt{n}) \]

\[ V = \text{Var} \left[ \log \frac{P_Y | X = \sqrt{\rho}}{Q_Y} \right] \frac{\rho(\rho + 2)}{2(1 + \rho)^2} \log^2 \epsilon \]

Approximation

\[ R^*(n, \epsilon) = C + o(1) \]
Normal approximation: the AWGN case (2)

Metaconverse for AWGN

$$R^*(n, \epsilon) \leq -\frac{1}{n} \log \beta_{1-\epsilon}(P_{Y^n | X^n=\bar{x}}, Q_{Y^n})$$

Stein’s lemma

$$-\log \beta_{1-\epsilon}(P_{Y^n | X^n=\bar{x}}, Q_{Y^n}) = nC - \sqrt{nV} Q^{-1}(\epsilon) + o(\sqrt{n})$$

$$V = \text{Var} \left[ \log \frac{P_Y}{Q_Y} \right] \frac{\rho(\rho + 2)}{2(1 + \rho)^2} \log^2 \epsilon$$

Approximation

$$R^*(n, \epsilon) = C - \sqrt{V/n} Q^{-1}(\epsilon) + o(1/\sqrt{n})$$
When is the normal approximation accurate? (1)

Normal approx. (Polyanskiy et al. ‘10, Tan & Tomamichel ‘15)

\[ R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{1}{2n} \log n + o\left(\frac{\log n}{n}\right) \]

\[ \epsilon = 10^{-4}, \rho = 0 \text{ dB} \]

Graph showing the capacity, metaconverse (PPV’10), normal approximation, and lower bound (Shannon ’59) with blocklength on the x-axis and rate [bit/ch. use] on the y-axis.
When is the normal approximation accurate? (2)

$R = 0.3, n = 300$

The graph shows the relationship between $\rho$ and $\epsilon$ for $R = 0.3$ and $n = 300$. The black line represents the metaconverse and the red line represents the normal approximation. The graph indicates the accuracy of the normal approximation for different values of $\rho$.
Application: ARQ on AWGN, $\rho = 0 \text{ dB}$ [PPV’10]

- Want to transmit $k$ bits
- $(n, 2^k, \epsilon)$-code + ARQ
- Long-term throughput
  \[ T(k) = \max_n \frac{k}{n} \left( 1 - \epsilon(k, n) \right) \]
- Optimize using normal approx.
Application: ARQ on AWGN, $\rho = 0$ dB \cite{PPV'10}

- Want to transmit $k$ bits
- $(n, 2^k, \epsilon)$-code + ARQ
- Long-term throughput
  \[ T(k) = \max_n \frac{k}{n} \left( 1 - \epsilon(k,n) \right) \]
- Optimize using normal approx.
Application: slotted Aloha

- $M$ users, $k$ bits per user
- frame of $n$ channel uses
- Probability of success:

\[
\frac{M}{S} \left(1 - \frac{1}{S}\right)^{M-1} \left(1 - \epsilon^*(k, nS)\right)
\]
Application: slotted Aloha

- $M$ users, $k$ bits per user
- frame of $n$ channel uses
- Probability of success:
  \[
  \frac{M}{S} \left(1 - \frac{1}{S}\right)^{M-1} \left(1 - \epsilon^*(k, nS)\right)
  \]
Outline

1. Motivation: short packets and information theory
2. Foundations: nonasymptotic information theory
3. Application: short packets over MIMO fading channels
4. Application: minimization of energy per bit
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Enter fading

AWGN channel with fluctuating SNR and multiple inputs/outputs
Enter fading

- AWGN channel with fluctuating SNR and multiple inputs/outputs
- Performance limits depend on:
  - How $\{H_k\}$ varies within the packet
  - Fading knowledge: noCSI, CSIR, CSIT, CSIRT
The memoryless block-fading model

\[ |h_k| \]

\[ n_c = 4 \]
\[ \ell = 5 \]
Relevance to 5G

- **Motivation:** short packets and information theory
- **Foundations:** nonasymptotic information theory
- **Application:** short packets over MIMO fading channels
- **Application:** minimization of energy per bit

**Model and nonasymptotic limits**
- FBL: quasi-static fading
- FBL: memoryless block fading

---

**Relevance to 5G**

- Time (T_c)
- Frequency (B_c)
- Larger than (B_c)
- Frequency (n_s)
- Time (n_o)
Two notions of capacity

**Outage capacity**
- $n_c \to \infty$, $\ell$ fixed
- Fading process stays “constant” over the packet
- $\times$ Does not capture the “cost” of learning the channel at the receiver

**Ergodic capacity**
- $\ell \to \infty$, $n_c$ fixed
- Fading process varies rapidly over the packet
- $\times$ Requires coding over many coherence intervals
- $\times$ Does not depend on $\epsilon$
Outage capacity

**SISO**

\[
C_{\text{out}, \epsilon} = \sup \{ R : P_{\text{out}}(R) \leq \epsilon \}
\]

where \( P_{\text{out}}(R) = \mathbb{P} \left[ \sum_{k=1}^{\ell} \log(1 + \rho |H_k|^2) \leq \ell R \right] \)
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Model and nonasymptotic limits
FBL: quasi-static fading
FBL: memoryless block fading

Outage capacity

**SISO**

\[ C_{out,\epsilon} = \sup\{ R : P_{out}(R) \leq \epsilon \} \]

where \( P_{out}(R) = \mathbb{P}\left[ \sum_{k=1}^{\ell} \log(1 + \rho |H_k|^2) \leq \ell R \right] \)

**SIMO**

\( C_{out,\epsilon} \) is the same for noCSI, CSIR, CSIT, CSITR

- The channel can be easily learnt at the receiver
- No power control possible
# Outage capacity

## SISO

$$C_{\text{out},\epsilon} = \sup\{R : P_{\text{out}}(R) \leq \epsilon\}$$

where

$$P_{\text{out}}(R) = \mathbb{P}\left[\sum_{k=1}^{\ell} \log(1 + \rho |H_k|^2) \leq \ell R\right]$$

## SIMO

$C_{\text{out},\epsilon}$ is **the same** for noCSI, CSIR, CSIT, CSITR

- The channel can be easily learnt at the receiver
- No power control possible

## MIMO

- CSIT: spatial waterfilling
- noCSIT: optimal input correlation matrix unknown (but Telatar conjecture)
Ergodic capacity

SISO

- CSIR: \( C_{\text{erg}}(\rho) = \mathbb{E} \left[ \log(1 + \rho |H|^2) \right] \)
- CSIRT: temporal waterfilling
- noCSI: \( C_{\text{erg}} \) not known in closed form, but high-SNR approx. [Marzetta & Hochwald ’00]:

\[
C_{\text{erg}}(\rho) = \left(1 - \frac{1}{n_c}\right) \log \rho + c + o_\rho(1)
\]

- All these results generalize to MIMO
Short-packets over fading channels: outline

- General quasi-static fading channel \((\ell = 1)\)
- Rayleigh memoryless block-fading channel
- Converse and achievability bounds at finite blocklength
- Engineering insights
Quasi-static fading channels: converse

Setup

Fading constant over the entire packet ($\ell = 1$)

Nonasymptotic converse for CSIRT

- **SIMO**: choose $Q_{Y^n}$ to be the capacity-achieving output distribution
- **MIMO**: choose a conditional distribution $Q_{Y^n|X^n}$ to allow for waterfilling (requires modified metaconverse)

Achievability: angle thresholding

$$y = h x$$

**Decoder**
- Declare $x^n(i)$ iff $\theta(x^n(i), Y^n) \leq \theta_\epsilon$
- Declare error otherwise
- Knowledge of $H$ or statistics of $H$ not required
- $\kappa_\beta$-bound to characterize performance
Numerical comparison: $1 \times 2$ Rician-fading channel

\begin{align*}
\epsilon &= 10^{-3} \\
\rho &= -1.55 \text{ dB} \\
K &= 20 \text{ dB}
\end{align*}

\begin{align*}
C_{out,\epsilon}^{\text{achievability CSIR}} &
\\
C_{out,\epsilon}^{\text{normal approx.}} &
\\
C_{out,\epsilon}^{\text{converse CSIR}} &
\\
\text{achievability noCSI} &
\\
\text{bit/channel use} &
\\
\text{blocklength } n_c &
\end{align*}
Numerical comparison: $1 \times 2$ Rician-fading channel

$\epsilon = 10^{-3}$
$\rho = -1.55 \text{ dB}$
$K = 20 \text{ dB}$

$90\% \ C_{out,\epsilon} : 120 - 320 \text{ CSIR}; 120 - 480 \text{ noCSI}; 1420 \text{ AWGN!}$
Fast convergence to $C_{\text{out},\epsilon}$: zero dispersion

**Theorem**

- Assume that the fading matrix has a smooth density $f_H(H)$, and
- $f_H(H) \to 0$ as $\|H\|_F \to \infty$ inverse polynomially (or faster).

Then,

$$R^*(n, \epsilon) = C_{\text{out},\epsilon} - \frac{1}{\sqrt{n}} + O\left(\frac{\log n}{n}\right)$$

for all four cases: noCSI, CSIR, CSIT, CSIRT.
Fast convergence to $C_{\text{out}, \epsilon}$: zero dispersion

**Theorem**

- Assume that the fading matrix has a smooth density $f_H(H)$, and
- $f_H(H) \to 0$ as $\|H\|_F \to \infty$ inverse polynomially (or faster).

Then,

$$R^*(n, \epsilon) = C_{\text{out}, \epsilon} - \frac{1}{\sqrt{n}} + \mathcal{O} \left( \frac{\log n}{n} \right)$$

for all four cases: noCSI, CSIR, CSIT, CSIRT

**Why?**

Main cause of error is deep fade and not additive noise!
MIMO Rayleigh block-fading model

- \( n_c \): coherence interval
- \( \ell \): number of time-frequency diversity branches
- \( m_t \times m_r \) MIMO channel
- noCSI
- \( R^*(n, \epsilon) \Rightarrow R^*(\ell, n_c, \epsilon) \)

A 5G design problem

- **Packet size:**
  168 symbols

- **14 OFDM symbols, 12 tones per symbol**

- **PEP:** $\epsilon = 10^{-5}$

- **SNR:** $\rho = 6$ dB

- **2 × 2 MIMO**
A 5G design problem

- packet size: 168 symbols
- 14 OFDM symbols, 12 tones per symbol
- PEP: $\epsilon = 10^{-5}$
- SNR: $\rho = 6$ dB
- $2 \times 2$ MIMO

![Diagram](https://via.placeholder.com/150)

$n = n_c \ell$

smaller resource blocks

time-frequency diversity branches $\ell$ (log scale)

coherence interval $n_c$ (log scale)
Asymptotic limit: outage capacity

\[ \lim_{n_c \to \infty} R^*(\ell, n_c, \epsilon) = C_{\text{out}, \epsilon} \]

\[ = \sup \{ R : \inf \mathbb{P}(P_{\text{out}}(R) \leq \epsilon) \} \]

\[ P_{\text{out}}(R) = \mathbb{P} \left\{ \sum_{k=1}^{\ell} \log \det (I_{m_r} + H_k^H Q_k H_k) \leq \ell R \right\} \]

- Same as for CSIR; CSI acquisition cost vanishes
- Fast convergence

\[ R^*(\ell, n_c, \epsilon) = C_{\text{out}, \epsilon} + O\left(\frac{\log n_c}{n_c}\right) \]
Outage capacity

![Graph showing outage capacity](image-url)
Asymptotic limit: Ergodic capacity

\[
\lim_{\ell \to \infty} R^*(\ell, n_c, \epsilon) = C_{\text{erg}} \\
= \frac{1}{n_c \ell} \sup I(X; Y)
\]

- It holds for all \(0 < \epsilon < 1\) (strong converse)
Asymptotic limit: Ergodic capacity

\[ n = n_c \ell \]

\[
\lim_{\ell \to \infty} R^*(\ell, n_c, \epsilon) = C_{\text{erg}} = \frac{1}{n_c \ell} \sup I(X; Y)
\]

- It holds for all \( 0 < \epsilon < 1 \) (strong converse)
- At high SNR [Zheng & Tse, ‘02; Yang et al., ‘12]

\[
C_{\text{erg}}(\rho) = m^* \left(1 - \frac{m^*}{n_c}\right) \log \rho + c + o(1)
\]

where \( m^* = \min\{m_t, m_r, \lfloor n_c/2 \rfloor\} \)

- CSI acquisition cost is explicit; some TX antennas should be switched off when \( n_c \) becomes small
Asymptotic limit: Ergodic capacity

\[
\lim_{\ell \to \infty} R^*(\ell, n_c, \epsilon) = C_{\text{erg}}
\]

\[
= \frac{1}{n_c \ell} \sup I(X; Y)
\]

- It holds for all \(0 < \epsilon < 1\) (strong converse)
- At high SNR [Zheng & Tse, ‘02; Yang et al., ‘12]

\[
C_{\text{erg}}(\rho) = m^* \left(1 - \frac{m^*}{n_c}\right) \log \rho + c + o(1)
\]

where \(m^* = \min\{m_t, m_r, \lfloor n_c/2 \rfloor\}\)

- CSI acquisition cost is explicit; some TX antennas should be switched off when \(n_c\) becomes small
- Tight bounds available [Alfano et al., ‘14, Devassy et al., ‘15]
The underlying geometry: \( m_t = m_r = m \)

\[
C_{\text{erg}} = m \left( 1 - \frac{m}{n_c} \right) \log(\rho) + o(\log(\rho))
\]
The underlying geometry: $m_t = m_r = m$ 

$$C_{\text{erg}} = m \left(1 - \frac{m}{n_c}\right) \log(\rho) + o(\log(\rho))$$
The underlying geometry: $m_t = m_r = m$

$$C_{\text{erg}} = m \left(1 - \frac{m}{n_c}\right) \log(\rho) + o(\log(\rho))$$
The underlying geometry: $m_t = m_r = m$

$C_{\text{erg}} = m \left(1 - \frac{m}{n_c}\right) \log(\rho) + o(\log(\rho))$

Communications on the Grassmannian manifold
Geometry suggests a signaling scheme

- **Uniform distribution** on the Grassmannian
  \[ X = \sqrt{n_c \rho} U \]
- \( U \): (truncated) **unitary** and **isotropically distributed**
- **Unitary space-time modulation (USTM)**
Ergodic capacity

Unitary space-time modulation (USTM)

Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Model and nonasymptotic limits
FBL: quasi-static fading
FBL: memoryless block fading

\[
C_{\text{erg}} \text{ (USTM lb)}
\]

\[
m_t^* = 1
\]

coherence interval \( n_c \) (log scale)

bit/channel use

time-frequency diversity branches \( \ell \) (log scale)
Asymptotic approximations on $R^*(\ell, n_c, \epsilon)$
Asymptotic approximations on $R^*(\ell, n_c, \epsilon)$

- Coherence interval $n_c$ (log scale)
- Bit/channel use
- $C_{\text{erg}}$ (USTM lb)
- $C_{\text{out},\epsilon}$
- $m_t^* = 1$

What is the nonasymptotic truth?
Nonasymptotic bounds: achievability

- **Dependency testing bound** \[\text{Polyanskiy et al., 2010}\]

- **USTM** over a **subset of TX antenna** used as input distribution

- **Induced output distribution** computed using **Itzykson-Zuber integral**

- **Optimize over number of TX antennas**
Nonasymptotic bounds: converse

- **Metaconverse theorem** [Polyanskiy et al., 2010]
- **USTM-induced** output distribution as $Q_{Y^{nc}}$
- **Optimize** over output distributions corresponding to a different number of active TX antennas
Nonasymptotic bounds on $R^*(\ell, n_c, \epsilon)$

- **Motivation:** short packets and information theory
- **Foundations:** nonasymptotic information theory
- **Application:** short packets over MIMO fading channels
- **Application:** minimization of energy per bit

- **Model and nonasymptotic limits**
- **FBL:** quasi-static fading
- **FBL:** memoryless block fading

### Graph
- **Coherence interval $n_c$ (log scale)**
- **Bit/channel use**
- **Time-frequency diversity branches $\ell$ (log scale)**
- **MC upper bound**
- **DT lower bound**
- **$R^*(n_c, l, \epsilon)$**
- **$C_{out, \epsilon}$**
- **$C_{\text{erg}}$ (USTM lb)**
- **$\ell^*$**
- **$m_t = 1$**

**Legend:**
- Blue: MC upper bound
- Red: DT lower bound
- Gray: $R^*(n_c, l, \epsilon)$
Diversity or multiplexing? $2 \times 2$ MIMO, $\epsilon = 10^{-5}$
Actual codes:

\[ n = 192, \ell = 8, k = 92, 1 \times 2 \text{ SIMO} \]
Outline

1. Motivation: short packets and information theory
2. Foundations: nonasymptotic information theory
3. Application: short packets over MIMO fading channels
4. Application: minimization of energy per bit
The plan

Minimum energy to transmit $k$ bits over AWGN and fading channels
The plan

Minimum energy to transmit $k$ bits over AWGN and fading channels

- Infinitely many bits: capacity per unit energy and the $-1.59$ dB limit
- Finitely many bits: bounds and approximation for AWGN...
- ... and for fading channels (CSI and noCSI)
Minimum energy per bit on AWGN channels

- Bandlimited AWGN channel \([Shannon '49]\)

\[
C = W \log_2 \left( 1 + \frac{P}{WN_0} \right) \quad \text{[bit/s]}
\]
Minimum energy per bit on AWGN channels

- Bandlimited AWGN channel [Shannon '49]

\[
C = W \log_2 \left( 1 + \frac{P}{WN_0} \right) \leq \frac{P}{N_0} \log_2 e \quad \text{[bit/s]}
\]
Minimum energy per bit on AWGN channels

- Bandlimited AWGN channel \([Shannon '49]\)

\[
C = W \log_2 \left( 1 + \frac{P}{WN_0} \right) \leq \frac{P}{N_0} \log e \quad \text{[bit/s]}
\]

- Energy per bit: \(E_b = P/C\)
Minimum energy per bit on AWGN channels

- Bandlimited AWGN channel \( [\text{Shannon } '49] \)
  \[
  C = W \log_2 \left( 1 + \frac{P}{WN_0} \right) \leq \frac{P}{N_0} \log_2 e \quad [\text{bit/s}]
  \]

- Energy per bit: \( E_b = P/C \)

- Minimum energy per bit:
  \[
  \left. \frac{E_b}{N_0} \right|_{\text{min}} = \frac{P/N_0}{P \log_2 e/N_0} = \log_e 2 = -1.59 \text{ dB}
  \]
Energy efficiency as derivative of spectral efficiency

In terms of SNR $\rho$:

$$\frac{E_b}{N_0} C = \frac{P}{N_0} \Rightarrow$$
Energy efficiency as derivative of spectral efficiency

In terms of SNR $\rho$:

\[
\frac{E_b}{N_0} C = \frac{P}{N_0} \Rightarrow \frac{E_b}{N_0} \cdot \frac{C}{W} = \frac{P}{N_0 W}
\]

\[= C(\rho) \Rightarrow = \rho\]
Energy efficiency as derivative of spectral efficiency

- In terms of SNR $\rho$:
  \[
  \frac{E_b}{N_0} C = \frac{P}{N_0} \Rightarrow \frac{E_b}{N_0} \cdot \frac{C}{W} = \frac{P}{N_0 W} = C(\rho) = \rho
  \]

- Hence,
  \[
  \frac{E_b}{N_0} C(\rho) = \rho \Rightarrow \frac{E_b}{N_0} = \frac{\rho}{C(\rho)}
  \]
Energy efficiency as derivative of spectral efficiency

- In terms of SNR $\rho$:
  \[
  \frac{E_b}{N_0} C = \frac{P}{N_0} \quad \Rightarrow \quad \frac{E_b}{N_0} \cdot \frac{C}{W} = \frac{P}{N_0W} \]
  
  Hence,
  \[
  \frac{E_b}{N_0} C(\rho) = \rho \quad \Rightarrow \quad \frac{E_b}{N_0} = \frac{\rho}{C(\rho)}
  \]

- Minimum energy per bit
  \[
  \left. \frac{E_b}{N_0} \right|_{\min} = \min_{\rho} \frac{\rho}{C(\rho)}
  \]
Energy efficiency as derivative of spectral efficiency

- In terms of SNR $\rho$:
  \[
  \frac{E_b}{N_0} C = \frac{P}{N_0} \quad \Rightarrow \quad \frac{E_b}{N_0} \cdot \frac{C}{W} = \frac{P}{N_0 W} = C(\rho) = \rho
  \]

- Hence,
  \[
  \frac{E_b}{N_0} C(\rho) = \rho \quad \Rightarrow \quad \frac{E_b}{N_0} = \frac{\rho}{C(\rho)}
  \]

- Minimum energy per bit
  \[
  \frac{E_b}{N_0} \bigg|_{\text{min}} = \min_{\rho} \frac{\rho}{C(\rho)} = \lim_{\rho \to 0} \frac{\rho}{C(\rho)} = \frac{1}{\dot{C}(0)}
  \]

- Achieved by Gaussian, but also BPSK, PPM, etc... [Golay '49]
Capacity per unit energy

- Stationary memoryless channel \((\mathcal{X}, P_Y|_{X,Y})\) with zero-energy symbol

Capacity per unit energy

- Stationary memoryless channel $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ with zero-energy symbol
- Capacity:

$$C(\rho) = \sup_{P_X : \mathbb{E}[|X|^2] \leq \rho} I(X; Y)$$

Capacity per unit energy

- **Stationary memoryless channel** \((\mathcal{X}, P_Y | X, \mathcal{Y})\) with zero-energy symbol
- **Capacity:**

\[
C(\rho) = \sup_{P_X : \mathbb{E}[|X|^2] \leq \rho} I(X; Y)
\]

- **Minimum energy per bit:** \(1/C(0)\)

---

Capacity per unit energy

- Stationary memoryless channel \((\mathcal{X}, P_{Y|X}, \mathcal{Y})\) with zero-energy symbol
- Capacity:
  \[
  C(\rho) = \sup_{P_X} \sup \left\{ I(X;Y) \right\} \quad \text{subject to} \quad \mathbb{E}[|X|^2] \leq \rho
  \]
- Minimum energy per bit: \(1/\hat{C}(0)\)
- The capacity per unit energy \(\hat{C}(0)\) can be computed directly
  \[
  \hat{C}(0) = \sup_{x \in \mathcal{X}} \frac{D(P_{Y|X=x} \| P_{Y|X=0})}{|x|^2}
  \]

---

Capacity per unit energy: achievability proof

- **Binary hypothesis test:**

  \[ P_{Y^n} \quad ? \quad Y_1 \cdots Y_n \quad \text{test} \quad Z \]

  \[ Q_{Y^n} \quad ? \quad Y_1 \cdots Y_n \quad \text{test} \quad Z \]

- \( Z = 0 \Rightarrow \text{choose } P_{Y^n}; \quad Z = 1 \Rightarrow \text{choose } Q_{Y^n} \)

- **Test:** conditional probability distribution \( P_Z | Y^n \)

- **Beta function:**

  \[ \beta_\alpha(P_{Y^n}, Q_{Y^n}) = \inf_{P_Z | Y^n : P_{Y^n}[Z=0] \geq \alpha} Q_{Y^n}[Z = 0] \]

- **Neyman-Pearson** optimal test: compare the LLR with threshold

  \[ \log \frac{P_{Y^n}}{Q_{Y^n}} \leq \gamma \]
Achievability proof: coding scheme

- $M$ codewords of length $MN$: $M$-PPM + length-$N$ repetition code
- **Decoder**: $M$ optimal binary hypothesis tests between $P_Y | X = c_m$ and $P_Y | X = 0$
Achievability proof: error probability analysis

\[ \mathbb{P}[\mathcal{E}] \leq \mathbb{P}[Z_1 = 1 \mid \mathbf{X} = \mathbf{c}_1] + \sum_{j=2}^{M} \mathbb{P}[Z_j = 0 \mid \mathbf{X} = \mathbf{c}_1] \]
Achievability proof: error probability analysis

\[
\mathbb{P}[\mathcal{E}] \leq \mathbb{P}[Z_1 = 1 \mid \mathbf{X} = \mathbf{c}_1] + \sum_{j=2}^{M} \mathbb{P}[Z_j = 0 \mid \mathbf{X} = \mathbf{c}_1]
\]

\[
= \mathbb{P}[Z_1 = 1 \mid \mathbf{X} = \mathbf{c}_1] + (M - 1) \mathbb{P}[Z_1 = 0 \mid \mathbf{X} = \mathbf{0}]
\]
Achievability proof: error probability analysis

\[ P[\mathcal{E}] \leq P[Z_1 = 1 \mid X = \mathbf{c}_1] + \sum_{j=2}^{M} P[Z_j = 0 \mid X = \mathbf{c}_1] \]

\[ = P[Z_1 = 1 \mid X = \mathbf{c}_1] + (M - 1) P[Z_1 = 0 \mid X = \mathbf{0}] \]

\[ = \frac{\epsilon}{2} + (M - 1) \beta_{1-\epsilon/2} \left( P_{Y^N \mid X^N=x_0}, P_{Y^N \mid X^N=0} \right) \]
Achievability proof: error probability analysis

\[ P[E] \leq P[Z_1 = 1 \mid X = c_1] + \sum_{j=2}^{M} P[Z_j = 0 \mid X = c_1] \]

\[ = P[Z_1 = 1 \mid X = c_1] + (M - 1) P[Z_1 = 0 \mid X = 0] \]

\[ = \frac{\epsilon}{2} + (M - 1) \beta_{1 - \epsilon/2} \left( P_{Y \mid X = x_0}^{N}, P_{Y \mid X = 0}^{N} \right) \]

\[ \leq \frac{\epsilon}{2} + \exp \left\{ -N \left[ D(P_{Y \mid X = x_0} \mid \mid P_{Y \mid X = 0}) - \frac{\log M}{N} + o(1) \right] \right\} \]

\[ \leq \frac{\epsilon}{2} + \exp \left\{ -N \left[ \frac{D(P_{Y \mid X = x_0} \mid \mid P_{Y \mid X = 0})}{\|x_0\|^2} - \frac{\log M}{N \|x_0\|^2} + o(1) \right] \right\} \]
Enter fading (again)

- AWGN channel with fluctuating SNR
- Channel gain known at Rx (CSIR) or unknown (noCSI)
- No CSIT; memoryless Rayleigh-fading channel with unit variance
Energy efficiency in fading channels

- For both CSIR and noCSI (!) [Jacobs '63, Pierce '66, Kennedy '69]

\[ \frac{E_b}{N_0} \bigg|_{\text{min}} = \log_2 e = -1.59 \text{ dB} \]

- CSIR: Gaussian, BPSK, PPM, . . .

- noCSI: flash signaling [Verdú '02]

---

Why flash signaling for noCSI?

- Input-output relation
  \[ Y = HX + Z \]

- Set \( N_0 = 1 \) \( \Rightarrow \) \( H, Z \sim \mathcal{CN}(0, 1) \)
Why flash signaling for noCSI?

- Input-output relation
  \[ Y = HX + Z \]

- Set \( N_0 = 1 \Rightarrow H, Z \sim \mathcal{CN}(0, 1) \)

- \( P_Y |_{X=x} = \mathcal{CN}(0, 1 + |x|^2); \quad P_Y |_{X=0} = \mathcal{CN}(0, 1) \)
Why flash signaling for noCSI?

- **Input-output relation**
  
  \[ Y = HX + Z \]

- **Set** \( N_0 = 1 \Rightarrow H, Z \sim \mathcal{CN}(0, 1) \)

- **Capacity per unit energy**

  \[
  \hat{C}(0) = \sup_{x \in \mathbb{C}} \frac{D(P_Y | X=x \| P_Y | X=0)}{|x|^2} = \sup_{x \in \mathbb{C}} \frac{|x|^2 \log_2 e - \log_2 (1 + |x|^2)}{|x|^2}
  \]
Why flash signaling for noCSI?

- Input-output relation
  \[ Y = HX + Z \]

- Set \( N_0 = 1 \) \( \Rightarrow \) \( H, Z \sim \mathcal{CN}(0, 1) \)

- \( P_Y | X=x = \mathcal{CN}(0, 1 + |x|^2) ; \quad P_Y | X=0 = \mathcal{CN}(0, 1) \)

- Capacity per unit energy
  \[
  \hat{C}(0) = \sup_{x \in \mathbb{C}} \frac{D(P_Y | X=x || P_Y | X=0)}{|x|^2} \\
  = \sup_{x \in \mathbb{C}} \frac{|x|^2 \log_2 e - \log_2 (1 + |x|^2)}{|x|^2} \\
  = \log_2 e
  \]
Leaving asymptotics: wideband slope [Verdú ’02]
Leaving asymptotics in another direction

- Finite number of bits
- Finite energy
- Nonzero error probability
Nonasymptotic framework

$k$ bits encoded onto $2^k$ finite-energy codewords (wideband limit)

$$\sum_{i=1}^{\infty} |x_i|^2 \leq E$$

Definition

$E_b^*(k, \epsilon)$: minimum energy per bit $E/k$ such that PEP $< \epsilon$
Nonasymptotic framework

$k$ bits encoded onto $2^k$ finite-energy codewords (wideband limit)

$$\sum_{i=1}^{\infty} |x_i|^2 \leq E$$

**Definition**

$E_b^*(k, \epsilon)$: minimum energy per bit $E/k$ such that PEP $< \epsilon$

$$\frac{E_b}{N_0} \bigg|_{\text{min}} = \lim_{\epsilon \to 0} \lim_{k \to \infty} \frac{E_b^*(k, \epsilon)}{N_0} = \log_e 2$$
The plan

**Tight bounds and approximations on** $E^*_b(k, \epsilon)$

- Two achievability bounds
- A converse bound
- Asymptotic approximations
- Applications: AWGN and fading channels

---


First achievability bound [Yang et al. ’16]

PPM + repetition coding + hypothesis testing decoder

For every input symbol $x_0$ and every $N \in \mathbb{N}$ such that $E = N |x_0|^2$

$$2^k - 1 \geq \sup_{0<\tau<\epsilon} \frac{\tau}{\beta_1-\epsilon+\tau}(P_{Y^N | X^N=x_0}, P_{Y^N | X^N=0})$$

- Similar to $\kappa/\beta$-bound
Second achievability bound [Yang et al. ’15]

PPM + repetition coding + maximum likelihood decoder

For every input symbol $x_0$ and every $N \in \mathbb{N}$ such that $E = N |x_0|^2$:

$$
\epsilon < \mathbb{E} \left[ \min \left\{ 1, (2^k - 1) P \left[ i(x_0, Y^N) \leq i(x_0, \bar{Y}^N) \mid Y^N \right] \right\} \right]
$$

where

$$
P_{Y^N, \bar{Y}^N}(y^N, \bar{y}^N) = P_{Y^N \mid X^N = x_0}(y^N) \cdot P_{Y^N \mid X^N = 0}(\bar{y}^N)
$$

$$
i(x; y^N) = \log \frac{P_{Y^N \mid X^N = x}}{P_{Y^N \mid X^N = 0}}
$$

● Similar to RCU-bound [Polyanskiy et al., 2010]
Converse bound

Metaconverse theorem

\[ 2^k \leq \sup_{\|x\|^2 \leq E} \frac{1}{\beta_1 - \epsilon} \left( \frac{P_Y | X=x, P_Y | X=0}{} \right) \]

Infinite-dimensional optimization required!
AWGN channel

Exploit symmetry; repetition not needed

**Achievability:** PPM + ML decoding

\[
\epsilon \leq \mathbb{E}\left[\min\left\{1, (2^k - 1)Q\left(\frac{2E}{N_0} + Z\right)\right\}\right]
\]

- \(Z \sim \mathcal{N}(0, 1)\);
- \(Q(\cdot)\): Gaussian Q function

**Converse**

\[
\frac{1}{M} \geq \frac{1}{\beta_{1-\epsilon}(P_Y|X=\sqrt{E}, P_Y|X=0)} = Q\left(\frac{2E}{N_0} + Q^{-1}(1 - \epsilon)\right)
\]
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

$E_b^*(k, \epsilon)$ vs. $k$ for AWGN channel ($\epsilon = 10^{-3}$)

$\frac{E_b^*(k, \epsilon)}{N_0} \quad [\text{dB}]$

- $\text{PPM + ML decoding}$
- $\text{meta-converse}$

$-1.59 \text{ dB}$

# information bits $k$

Giuseppe Durisi
short packets
$E_b^*(k, \epsilon)$ vs. $k$: for AWGN channel ($\epsilon = 10^{-3}$)

$$\frac{E_b^*(k, \epsilon)}{N_0} = \log_e 2 + \sqrt{\frac{2 \log e 2}{k}} Q^{-1}(\epsilon) + O\left(\frac{\ln k}{k}\right)$$

PPM + ML decoding

meta-converse

$-1.59$ dB

Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Minimum energy per bit in the asymptotic limits
Nonasymptotic bounds and approximations
Conclusions
The Rayleigh-fading case [Yang et al. ’15]

- **Computable** bounds on $E_b^*(k, \epsilon)$ as accurate as in the AWGN case
The Rayleigh-fading case [Yang et al. ’15]

- **Computed** bounds on $E^*_b(k, \epsilon)$ as accurate as in the AWGN case.

- **Asymptotics for CSIR** (same as AWGN):

\[
\frac{E^*_b(k, \epsilon)}{N_0} = \log_e 2 + \sqrt{\frac{\log_e 2}{k}} Q^{-1}(\epsilon) + O\left(\frac{\ln k}{k}\right)
\]
The Rayleigh-fading case \cite{Yang et al. '15}

- **Computable** bounds on $E_b^*(k, \epsilon)$ as accurate as in the AWGN case

- **Asymptotics for CSIR** (same as AWGN):

  $$\frac{E_b^*(k, \epsilon)}{N_0} = \log e \ 2 + \sqrt{\frac{\log e 2}{k}} Q^{-1}(\epsilon) + O\left(\frac{\ln k}{k}\right)$$

- **Asymptotics for noCSI**:

  $$\frac{E_b^*(k, \epsilon)}{N_0} = \log e \ 2 + c \sqrt[3]{\frac{\ln k}{k}} \left(3\sqrt[3]{Q^{-1}(\epsilon)}\right)^2 + o\left(\frac{1}{k^{1/3}}\right)$$

  $$c = \left(\frac{3}{\sqrt[3]{12}} + \frac{3}{\sqrt[3]{2/3}}\right) 3 \sqrt[3]{(\log_2 e)^2}$$
Minimum energy per bit: noCSI vs CSIR ($\epsilon = 10^{-3}$)

![Graph showing the comparison between noCSI and CSIR for minimum energy per bit.](image)

- **Converse**
- **Achievability**

- **noCSI**: $k \approx 6 \cdot 10^4$ (CSIR)
- **noCSI**: $k \approx 7 \cdot 10^7$ (noCSI)

- $-1.59$ dB
Minimum energy per bit: noCSI vs CSIR ($\epsilon = 10^{-3}$)

To achieve $E_b(k, \epsilon)/N_0 = -1.5$ dB:

- $k \approx 6 \cdot 10^4$ (CSIR)
- $k \approx 7 \cdot 10^7$ (noCSI)
The **CSIR case: achievability**

- **Transform** fading channel into **AWGN** by repetition coding

  **Tx:** repeat $N$ times each symbol $x$

  **Rx:** perform MRC $\Rightarrow \hat{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_i^* Y_i$

  As $N \to \infty$ we have by LLN and CLT:

  $\hat{Y} \to x \cdot \frac{1}{N} \sum_{i=1}^{N} |H_i|^2 + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_i^* Z_i \to x + Z$

  Fading is averaged out (irrespectively of fading distribution)
The CSIR case: achievability

- Transform fading channel into AWGN by repetition coding
  
  \[
  \text{Tx: repeat } N \text{ times each symbol } x \\
  \text{Rx: perform MRC } \Rightarrow \hat{Y} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_i^* Y_i
  \]

- As \( N \to \infty \) we have by LLN and CLT:
  
  \[
  \hat{Y} = x \cdot \frac{1}{N} \sum_{i=1}^{N} |H_i|^2 + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_i^* Z_i \to x + Z
  \]

- Fading is averaged out (irrespectively of fading distribution)
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Summary

- **AWGN and Rayleigh-fading channels with perfect CSIR:**
  \[
  \frac{E_b^*(k, \epsilon)}{N_0} \approx \log_e 2 + \sqrt{\frac{1}{k}} \text{const}
  \]

- **Rayleigh fading channels, noCSI:**
  \[
  \frac{E_b^*(k, \epsilon)}{N_0} \approx \log_e 2 + \sqrt{\frac{3 \ln k}{k}} \text{const}
  \]

- **Unchanged under MIMO and/or block-fading**
Further results I did not show you today

- Pilot transmission + mismatched nearest neighbor decoding
- CSIT + power control
- Variable-length codes with decision feedback (HARQ)
- Broadcasting a common message
- Packets through queues
- Massive random multiple access +FBL (many open problems!)
Conclusions

Finite-blocklength inf. theory

✅ Elegant theory
✅ Tight bounds for short-packet transmissions
✅ Many engineering insights for the design LP-WAN, 5G, and beyond
Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

SPECTRE: short-packet communication toolbox

- Collection of numerical routines in finite-blocklength information theory
- Available freely on GitHub
- Want to contribute? Contact us!

Motivation: short packets and information theory
Foundations: nonasymptotic information theory
Application: short packets over MIMO fading channels
Application: minimization of energy per bit

Bibliography (1)


Bibliography (2)


