Information-theoretic perspective on massive multiple-access

Yury Polyanskiy

Department of EECS
MIT

yp@mit.edu

SkolTech Mini Course, Jul. 2018

Legal notice: Some images in this presentation are borrowed from publicly available sources. The copyright on these images belongs to their original creators. For full copyright information please contact the author.
Lecture plan

1 Lecture 1: Short packets. Classical MAC.
   ▶ Motivation: Why work on MAC now? What is new?
   ▶ Finite blocklength IT: a few results
   ▶ Classical MAC IT

2 Lecture 2: Gaussian MAC. Modulation. CDMA.
   ▶ Orthogonal modulation (TDMA, FDMA, CDMA) and non-orthogonal (NOMA).
   ▶ Gaussian MAC. TIN. TIN+SIC. Rate-Splitting.
   ▶ Spectral efficiency and $E_b/N_0$.
   ▶ Randomly-spread CDMA. Effect of MUD.

3 Lecture 3: Massive MAC. Information theoretic analysis.
   ▶ Number of users scales with blocklength $K = \mu n$.
   ▶ Per-user probability of error (PUPE). Absence of strong converse.
   ▶ Gaussian-process achievability bound.

4 Lecture 4: Random-access
   ▶ Survey of attempts to formalize random-access.
   ▶ Our take: random-access = same-codebook.
   ▶ Achievability bound.
   ▶ Lattice-based coding scheme.
• Cell phone is powered on.
• Announces its presence on PRACH.
• Base station (periodically) gives permission to send.
• Summary:
  ▶ Random-Access is very low duty cycle.
  ▶ BS makes access **ORTHOGONAL** across users
  ▶ bulk of communication is over an interference-free single-user AWGN.
• What’s new in 5G?
Internet-of-Things

- Smart Agriculture
- Advanced Metering systems
- Fire alarms
- Home security and automation
- Oilfield and pipeline monitoring
- M-health
- Smart parking, intelligent traffic
- Waste and recycling
- Asset tracking and geo-location
- Animal tracking and livestock

Expected density: 100-500 devices per household/office
Soup of solutions

- Bluetooth LE
- Wi-Fi
- Wi-Gig
- HSDPA
- LTE
- LoRa
- Thread
- Sigfox
- Bluetooth
- Zigbee
- NFC
- UWB
- GPRS
- 868/915 ISM
- DECT
- Wi-Max
- HSPA
- Zwave
- NWave
- Long Range
- Long Battery Life
- Low Cost
- High Capacity
Two breeds of IoT

LPWAN

One base station covers 10 km
Q: What drains the battery? Examples (@ 3.3V):

<table>
<thead>
<tr>
<th>State</th>
<th>Arduino (w/o reg.)</th>
<th>XBee (Zigbee)</th>
<th>LP-WAN sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep</td>
<td>5 uA</td>
<td>1 uA</td>
<td>1-2 uA</td>
</tr>
<tr>
<td>CPU Running</td>
<td>50 uA</td>
<td>40 uA</td>
<td>60 uA</td>
</tr>
</tbody>
</table>

Duty-cycle of 1 sec / 20 min radio lasts 6-10 yr / AA bat. Caveat: Calculation assumes single-user.

Key problem: Energy usage will grow with # of sensors deployed. How much?

Sad: depends on technology? Happy: IT comes to rescue!
IoT is about battery life

Q: What drains the battery? Examples (@ 3.3V):

<table>
<thead>
<tr>
<th>Activity</th>
<th>Arduino (w/o reg.)</th>
<th>XBee (Zigbee)</th>
<th>LP-WAN sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep</td>
<td>5 uA</td>
<td>1 uA</td>
<td>1-2 uA</td>
</tr>
<tr>
<td>CPU Running</td>
<td>50 uA</td>
<td>40 uA</td>
<td>60 uA</td>
</tr>
<tr>
<td>Radio Xmit</td>
<td>40 mA</td>
<td>40 mA</td>
<td>20 mA</td>
</tr>
</tbody>
</table>

- Duty-cycle of 1 sec / 20 min radio lasts 6-10 yr / AA bat.
- Caveat: Calculation assumes single-user.
- Key problem: Energy usage will grow with # of sensors deployed. How much?
- Sad: depends on technology?
- Happy: IT comes to rescue!
**Q:** What drains the battery? Examples (@ 3.3V):

<table>
<thead>
<tr>
<th></th>
<th>Arduino (w/o reg.)</th>
<th>XBee (Zigbee)</th>
<th>LP-WAN sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep</td>
<td>5 uA</td>
<td>1 uA</td>
<td>1-2 uA</td>
</tr>
<tr>
<td>CPU Running</td>
<td>50 uA</td>
<td>40 uA</td>
<td>60 uA</td>
</tr>
<tr>
<td>Radio Xmit</td>
<td>40 mA</td>
<td>40 mA</td>
<td>20 mA</td>
</tr>
</tbody>
</table>

- Duty-cycle of 1 sec / 20 min radio lasts 6-10 yr / AA bat.
- **Caveat:** Calculation assumes single-user
- **Key problem:** Energy usage will grow with # of sensors deployed. How much?
- **Sad:** depends on technology? **Happy:** IT comes to rescue!
Envisioned solution:

- To save battery: sensors sleep all the time, except transmissions.
- ... uncoordinated transmissions.
- ... they wake up, blast the packet, go back to sleep.
- Focus on low-energy (low $E_b/N_0$)
- Focus on fundamental limits
- ... but with low-complexity solutions (single-user-only decoding).
Envisioned solution:

- To save battery: sensors sleep all the time, except transmissions.
- ... uncoordinated transmissions.
- ... they wake up, blast the packet, go back to sleep.
- Focus on low-energy \( (low \ E_b/N_0) \)
- Focus on fundamental limits
- ... but with low-complexity solutions (single-user-only decoding).

Issues we need to understand:

1. packets are short: finite-blocklength (FBL) info theory
2. multiple-access channel: Classical MAC
3. low-complexity MAC: modulation, CDMA, multi-user detection
4. massive random-access: many users, same-codebook codes (NEW)
Envisioned solution:

- To save battery: sensors sleep all the time, except transmissions.
- ... uncoordinated transmissions.
- ... they wake up, blast the packet, go back to sleep.
- Focus on low-energy ($E_b/N_0$)
- Focus on fundamental limits
- ... but with low-complexity solutions (single-user-only decoding).

Issues we need to understand:

1. packets are short: finite-blocklength (FBL) info theory
2. multiple-access channel: Classical MAC
3. low-complexity MAC: modulation, CDMA, multi-user detection
4. massive random-access: many users, same-codebook codes (NEW)

Supporting 10 users at 1Mbps is much easier than 1M users at 10bps.
FBL Info Theory: short intro
Case study: 1000-bit BSC

- Consider channel $BSC(n = 1000, \delta = 0.11)$
- How many data bits can we transmit with (block) $P_e \leq 10^{-3}$?
- Attempt 1: Repetition
  
  $$k = 47 \text{ bits via } [21,1,21]\text{-code}$$

- Attempt 2: Reed-Muller
  
  $$k = 112 \text{ bits via } [64,7,32]\text{-code}$$

- Shannon’s prediction: $C = 0.5 \text{ bit so}$
  
  $$k \approx 500 \text{ bit}$$
Case study: 1000-bit BSC

- Consider channel $BSC(n = 1000, \delta = 0.11)$
- How many data bits can we transmit with (block) $P_e \leq 10^{-3}$?
- Attempt 1: Repetition
  \[ k = 47 \text{ bits via } [21,1,21]\text{-code} \]
- Attempt 2: Reed-Muller
  \[ k = 112 \text{ bits via } [64,7,32]\text{-code} \]
- Shannon’s prediction: $C = 0.5$ bit so
  \[ k \approx 500 \text{ bit} \]
- Finite blocklength IT:
  \[ 414 \leq k \leq 416 \]
Abstract communication problem

Noisy channel

Goal: Decrease corruption of data caused by noise
Goal: Decrease corruption of data caused by noise

Solution: Code to diminish probability of error $P_e$.

Key metrics: Rate and $P_e$. 
Channel coding: principles

Data bits

Redundancy

Noisy channel

Pe

Reliability–Rate tradeoff

Possible

Impossible

Rate
Decreasing $P_e$ further:

1. More redundancy
   **Bad:** loses rate

2. Increase blocklength!
Decreasing $P_e$ further:

1. More redundancy
   **Bad:** loses rate

2. Increase blocklength!

\[ n = 100 \]
Decreasing $P_e$ further:

1. More redundancy
   - Bad: loses rate
2. Increase blocklength!
Decreasing $P_e$ further:

1. More redundancy
   - **Bad**: loses rate

2. Increase blocklength!

$n = 10^6$
Channel coding: Shannon capacity

Data bits

Redundancy

Noisy channel

Shannon: Fix $R < C$

$P_e \downarrow 0$ as $n \rightarrow \infty$

Reliability–Rate tradeoff

Channel capacity
Channel coding: Shannon capacity

Shannon: Fix $R < C$

$P_e \downarrow 0$ as $n \to \infty$

Question:
For what $n$ will $P_e < 10^{-3}$?
Channel coding: Gaussian approximation

Data bits \hspace{1cm} Redundancy

Noisy channel

Shannon: Fix $R < C$

$P_e \downarrow 0$ as $n \to \infty$

Question:
For what $n$ will $P_e < 10^{-3}$?

Channel dispersion

Channel capacity
Channel coding: Gaussian approximation

Noisy channel
Data bits
Redundancy

Shannon: Fix $R < C$

$P_e \downarrow 0$ as $n \to \infty$

Question:
For what $n$ will $P_e < 10^{-3}$?

Answer:

$n \gtrsim \text{const} \cdot \frac{V}{C^2}$
How to describe evolution of the boundary?

Classical results:

- **Vertical asymptotics**: fixed rate, reliability function
  Elias, Dobrushin, Fano, Shannon-Gallager-Berlekamp

- **Horizontal asymptotics**: fixed $\epsilon$, strong converse, $\sqrt{n}$ terms
  Wolfowitz, Weiss, Dobrushin, Strassen, Kemperman
XXI century:

- Tight non-asymptotic bounds
- Remarkable precision of normal approximation
- Extended results on *horizontal* asymptotics
  - AWGN, $O(\log n)$, cost constraints, feedback, *etc.*
Definition

\[ R^*(n, \epsilon) = \max \left\{ \frac{1}{n} \log M : \exists (n, M, \epsilon) \text{-code} \right\} \]

(max. achievable rate for blocklength \(n\) and prob. of error \(\epsilon\))

Rough summary: For ergodic channels

\[ R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n} Q^{-1}(\epsilon)}. \]
Connection to CLT

- Let $P_{Y^n|X^n} = P_{Y|X}^n$ be memoryless.

- Converse bounds (roughly):

$$R^*(n, \epsilon) \lesssim \epsilon\text{-th quantile of } \frac{1}{n} \log \frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$$

- Achievability bounds (roughly):

$$R^*(n, \epsilon) \gtrsim \epsilon\text{-th quantile of } \frac{1}{n} \log \frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$$
Connection to CLT

- Let $P_{Y^n|X^n} = P_{Y|X}^n$ be memoryless.
- Converse bounds (roughly):

  $$R^*(n, \epsilon) \lesssim \epsilon\text{-th quantile of } \frac{1}{n} \log \frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$$

- Achievability bounds (roughly):

  $$R^*(n, \epsilon) \gtrsim \epsilon\text{-th quantile of } \frac{1}{n} \log \frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$$

- Info-density $i(X^n; Y^n) = \log \frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$ is a sum of iid.

- Choice of $Q_{Y^n}$ is an art. Often c.a.o.d. works. Then, $\mathbb{E}[i(X^n; Y^n)] = nC$.

- So by CLT

  $$R^*(n, \epsilon) \approx \epsilon\text{-quantile of } \mathcal{N}(C, V/n)$$
FBL achievability bounds

- A random transformation $A \xrightarrow{P_{Y|X}} B$
- $(M, \epsilon)$ codes:
  \[ W \rightarrow A \rightarrow B \rightarrow \hat{W} \quad W \sim Unif\{1, \ldots, M\} \]
  \[ P[W \neq \hat{W}] \leq \epsilon \]
- For every $P_{XY} = P_X P_{Y|X}$ define information density:
  \[ \iota(x; y) \triangleq \log \frac{dP_{Y|X=x}}{dP_Y}(y) \]
  \[ \mathbb{E} [\iota(X; Y)] = I(X; Y) \]
  \[ \text{Var} [\iota(X; Y)|X] = V \]
  - Memoryless channels: $\iota(A^n; B^n) = \text{sum of iid.}$
  \[ \iota(A^n; B^n) \overset{d}{=} n I(A; B) + \sqrt{n} V Z, \quad Z \sim \mathcal{N}(0, 1) \]
- Goal: Prove FBL bounds.
  As by-product: $R^*(n, \epsilon) \gtrsim C - \sqrt{\frac{V}{n} Q^{-1}(\epsilon)}$
Theorem (Dependence Testing Bound)

For any $P_X$ there exists a code with $M$ codewords and

$$\epsilon \leq \mathbb{E} \left[ \exp \left\{ - \left| \nu_{X;Y}(X;Y) - \log \frac{M-1}{2} \right|^+ \right\} \right].$$

Highlights:

- Strictly stronger than Feinstein-Shannon
- ... and no optimization over $\gamma$!
- Easier to compute than RCU
- Easier asymptotics:
  $$\epsilon \leq \mathbb{E} \left[ e^{-n \frac{1}{n} \nu(X^n;Y^n) - R}^+ \right] \approx Q \left( \sqrt{\frac{n}{V}} \{ I(X;Y) - R \} \right)$$
- Has a form of $f$-divergence:
  $$1 - \epsilon \geq D_f(P_{XY} \parallel P_X P_Y)$$
DT bound: Proof

- Codebook: random $C_1, \ldots C_M \sim P_X$ iid
- Feinstein decoder:

$$\hat{W} = \text{smallest } j \text{ s.t. } \iota_{X;Y}(C_j;Y) > \gamma$$

- $j$-th codeword’s probability of error:

$$P[\text{error} \mid W = j] \leq \min \left( \underbrace{P[\iota_{X;Y}(X;Y) \leq \gamma]}_{(a)} + (j - 1) \underbrace{P[\iota_{X;Y}(\bar{X};Y) > \gamma]}_{(b)} \right)$$

  In $(a)$: $C_j$ too far from $Y$
  In $(b)$: $C_k$ with $k < j$ is too close to $Y$

- Average over $W$:

$$P[\text{error}] \leq P[\iota_{X;Y}(X;Y) \leq \gamma] + \frac{M-1}{2} P[\iota_{X;Y}(\bar{X};Y) > \gamma]$$
DT bound: Proof

- Recap: for every $\gamma$ there exists a code with

$$\epsilon \leq \mathbb{P}[\iota_{X;Y}(X;Y) \leq \gamma] + \frac{M-1}{2} \mathbb{P}[\iota_{X;Y}(\bar{X};Y) > \gamma] .$$

- Key step: closed-form optimization of $\gamma$.
- Introduce $\bar{X} \perp Y$: $\iota_{X;Y} = \log \frac{dP_{XY}}{dP_{\bar{X}Y}}$
- We have

$$P_{XY} \left[ \frac{dP_{XY}}{dP_{\bar{X}Y}} \leq e^\gamma \right] + \frac{M-1}{2} P_{\bar{X}Y} \left[ \frac{dP_{XY}}{dP_{\bar{X}Y}} > e^\gamma \right]$$

Bayesian dependence testing!

Optimum threshold: Ratio of priors $\Rightarrow \gamma^* = \log \frac{M-1}{2}$

- Change of measure argument:

$$P \left[ \frac{dP}{dQ} \leq \tau \right] + \tau Q \left[ \frac{dP}{dQ} > \tau \right] = \mathbb{E}_P \left[ \exp \left\{ - \left| \log \frac{dP}{dQ} - \log \tau \right|^+ \right\} \right] .$$
• Take a random transformation \( A \xrightarrow{P_{Y|X}} B \) (think \( A = \mathcal{A}^n \), \( B = \mathcal{B}^n \), \( P_{Y|X} = P_{Y^n|X^n} \))

• Input distribution \( P_X \) induces \( P_Y = P_{Y|X} \circ P_X \)
  \[
P_{XY} = P_X P_{Y|X}
\]

• Fix code:

\[
W \xrightarrow{encoder} X \rightarrow Y \xrightarrow{decoder} \hat{W}
\]

\( W \sim \text{Unif}[M] \) and \( M = \# \text{ of codewords} \)

Input distribution \( P_X \) associated to a code:

\[
P_X[.] \triangleq \frac{\# \text{ of codewords} \in (.)}{M}.
\]

• Goal: Upper bounds on \( \log M \) in terms of \( \epsilon \triangleq \mathbb{P}[\text{error}] \)

As by-product: \( R^*(n, \epsilon) \lesssim C - \sqrt{\frac{V}{n} Q^{-1}(\epsilon)} \)
Fano’s inequality

Theorem (Fano)

For any code

encoder \hspace{1cm} P_{Y|X} \hspace{1cm} decoder

\[ W \rightarrow X \rightarrow Y \rightarrow \hat{W} \]

with \( W \sim \text{Unif}\{1, \ldots, M\} \):

\[
\log M \leq \sup_{P_X} I(X; Y) + h(\epsilon) \cdot \frac{1}{1 - \epsilon}, \quad \epsilon = \mathbb{P}[W \neq \hat{W}]
\]

Implies weak converse:

\[
R^*(n, \epsilon) \leq \frac{C}{1 - \epsilon} + o(1).
\]

Proof: \( \epsilon\)-small \( \implies \) \( H(W|\hat{W})\)-small \( \implies \) \( I(X; Y) \approx H(W) = \log M \)
A (very long) proof of Fano via *channel substitution*

Consider two distributions on \((W, X, Y, \hat{W})\):

\[
P : \quad P_{WXY\hat{W}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}
\]

DAG: \hspace{1em} W \rightarrow X \rightarrow Y \rightarrow \hat{W}

\[
Q : \quad Q_{WXY\hat{W}} = P_W \times P_{X|W} \times Q_Y \times P_{\hat{W}|Y}
\]

DAG: \hspace{1em} W \rightarrow X \underline{\rightarrow} Y \rightarrow \hat{W}

Under \(Q\) the channel is useless:

\[
Q[W = \hat{W}] = \sum_{m=1}^{M} P_W(m)Q_{\hat{W}}(m) = \frac{1}{M} \sum_{m=1}^{M} Q_{\hat{W}}(m) = \frac{1}{M}
\]

Next step: data-processing for relative entropy \(D(\cdot||\cdot)\)
Data-processing for $D(\cdot \| \cdot)$

\[ D(P_A \| Q_A) \geq D(P_B \| Q_B) \]

Input distribution

\[ P_A \]

\[ Q_A \]

Output distribution

\[ P_B \]

\[ Q_B \]
Data-processing for $D(\cdot \| \cdot)$

\[ D(P_A \| Q_A) \geq D(P_B \| Q_B) \]

Apply to transform: $(W, X, Y, \hat{W}) \mapsto 1\{W \neq \hat{W}\}$:

\[ D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) \geq d(\mathbb{P}[W = \hat{W}] \| \mathbb{Q}[W = \hat{W}]) \]

\[ = d(1 - \epsilon \| \frac{1}{M}) \]

where $d(x \| y) = x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y}$. 

Yury Polyanskiy 
MAC tutorial
A proof of Fano via *channel substitution*

So far:

\[ D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) \geq d(1 - \epsilon \| \frac{1}{M}) \]

Lower-bound RHS:

\[ d(1 - \epsilon \| \frac{1}{M}) \geq (1 - \epsilon) \log M - h(\epsilon) \]

Analyze LHS:

\[ D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) = D(P_{XY} \| Q_{XY}) \]
\[ = D(P_X P_{Y|X} \| P_X Q_Y) \]
\[ = D(P_{Y|X} \| Q_Y | P_X) \]

(Recall: \( D(P_{Y|X} \| Q_Y | P_X) = \mathbb{E}_{x \sim P_X}[D(P_{Y|X=x} \| Q_Y)] \))
A proof of Fano via *channel substitution*: last step

Putting it all together:

\[
(1 - \epsilon) \log M \leq D(P_{Y|X} \parallel Q_Y | P_X) + h(\epsilon) \quad \forall Q_Y \quad \forall \text{code}
\]

Two methods:

1. Compute \( \sup_{P_X} \inf_{Q_Y} \) and recall

\[
\inf_{Q_Y} D(P_{Y|X} \parallel Q_Y | P_X) = I(X; Y)
\]

2. Take \( Q_Y = P_{Y}^* = \text{the caod} \) (capacity achieving output dist.) and recall

\[
D(P_{Y|X} \parallel P_{Y}^* | P_X) \leq \sup_{X} I(X; Y) \quad \forall P_X
\]

Conclude:

\[
(1 - \epsilon) \log M \leq \sup_{P_X} I(X; Y) + h(\epsilon)
\]

**Important:** Second method is particularly useful for FBL!
Tightening: from $D(\cdot \mid \cdot)$ to $\beta_\alpha(\cdot, \cdot)$

**Question:** How about replacing $D(\cdot \mid \cdot)$ with other divergences?

---

Note: Using $\beta_\alpha$ is aka meta-converse.
Tightening: from $D(\cdot \parallel \cdot)$ to $\beta_\alpha(\cdot, \cdot)$

**Question:** How about replacing $D(\cdot \parallel \cdot)$ with other divergences?

<table>
<thead>
<tr>
<th>$D(\cdot \parallel \cdot)$</th>
<th>relative entropy (KL divergence)</th>
<th>weak converse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_\lambda(\cdot \parallel \cdot)$</td>
<td>Rényi divergence</td>
<td>strong converse</td>
</tr>
<tr>
<td>$\beta_\alpha(\cdot, \cdot)$</td>
<td>Neyman-Pearson ROC curve</td>
<td>FBL bounds</td>
</tr>
</tbody>
</table>
**Tightening: from $D(\cdot||\cdot)$ to $\beta_\alpha(\cdot,\cdot)$**

**Question:** How about replacing $D(\cdot||\cdot)$ with other divergences?

<table>
<thead>
<tr>
<th>Divergence</th>
<th>Description</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D(\cdot</td>
<td></td>
<td>\cdot)$</td>
</tr>
<tr>
<td>$D_\lambda(\cdot</td>
<td></td>
<td>\cdot)$</td>
</tr>
<tr>
<td>$\beta_\alpha(\cdot,\cdot)$</td>
<td>Neyman-Pearson ROC curve</td>
<td>FBL bounds</td>
</tr>
</tbody>
</table>

**Note:** Using $\beta_\alpha$ is aka *meta-converse.*

... and leads to $R^*(n,\epsilon) \leq C - \sqrt[4]{n}Q^{-1}(\epsilon)$
General meta-converse principle

Steps:

• Select auxiliary channel $Q_{Y|X}$ (art)
  E.g.: $Q_{Y|X=x} = \text{center of a cluster of } x$

• Prove converse bound for channel $Q_{Y|X}$
  E.g.: $Q[W = \hat{W}] \lesssim \frac{\# \text{ of clusters}}{M}$

• Compute distance $D(\mathbb{P}||\mathbb{Q})$ between two spaces

\[
\mathbb{P} : P_{WXYW} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}
\]

vs.

\[
\mathbb{Q} : P_{WXYW} = P_W \times P_{X|W} \times Q_{Y|X} \times P_{\hat{W}|Y}
\]

• Apply data processing:

\[
D(P_{W,\hat{W}}||Q_{W,\hat{W}}) \leq D(P_{X,Y}||Q_{X,Y})
\]

• Key observation: This inequality connects $\mathbb{P}[\text{error}], \mathbb{Q}[\text{error}]$ and distance $D(\mathbb{P}||\mathbb{Q})$. 
FBL: summary

• All in all, these methods allow us to conclude:

\[ R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) \]

for a wide range of channels.

• Typically, \( V = \text{Var}[i(X; Y)|X] \) for cap.ach. distribution \( X \).
FBL: summary

• All in all, these methods allow us to conclude:

\[ R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) \]

for a wide range of channels.

• Typically, \( V = \text{Var}[i(X; Y) | X] \) for cap.ach. distribution \( X \).

• Example: The AWGN Channel

\[ Z \sim \mathcal{N}(0, \sigma^2) \]

\[ \downarrow \]

\[ X \quad \rightarrow \quad \oplus \rightarrow \quad Y \]

Codewords \( x^n \in \mathbb{R}^n \) satisfy power-constraint:

\[ \sum_{j=1}^{n} |x_j|^2 \leq nP \]

\[ C(P) = \frac{1}{2} \log(1 + P), \quad V(P) = \frac{\log^2 e}{2} \left( 1 - \frac{1}{(1 + P)^2} \right) \]

• Curious property of Gaussian noise: \( V(P) \leq \frac{\log^2 e}{2} \)
All in all, these methods allow us to conclude:

\[ R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) \]

for a wide range of channels.

Typically, \( V = \text{Var}[i(X; Y)|X] \) for cap.ach. distribution \( X \).

Example: The AWGN Channel

Below for Gaussian MAC we focus on m.i./capacity. By FBL there \( \exists \) codes within \( O\left(\frac{1}{\sqrt{n}}\right) \) uniformly in \( P \).

Codewords \( x^n \in \mathbb{R}^n \) satisfy power-constraint: \( \sum_{j=1}^{n} |x_j|^2 \leq nP \)

\[
C(P) = \frac{1}{2} \log(1 + P), \quad V(P) = \frac{\log^2 e}{2} \left(1 - \frac{1}{(1 + P)^2}\right)
\]

Curious property of Gaussian noise: \( V(P) \leq \frac{\log^2 e}{2} \)
Classical multiple-access IT
Core problem: many users, one channel
IT vs networks view on MAC

- Core problem: many users, one channel
- Networking folks:
  - ALOHA protocol (slotted) achieves:
    \[ \sum_i R_i \approx 0.37C \]
- Open problem: what max fraction \( \eta^* \) achievable?
  - State of the art [Tsybakov-Lihanov'87]: \( 0.476 \leq \eta^* \leq 0.568 \)
    (collision resolution codes)
IT vs networks view on MAC

- Core problem: many users, one channel
- Networking folks:
  - ALOHA protocol (slotted) achieves:
    \[ \sum_i R_i \approx 0.37C \]
  - Open problem: what max fraction \( \eta^* \) achievable?
    - State of the art [Tsybakov-Lihanov’87]: \( 0.476 \leq \eta^* \leq 0.568 \)
      (collision resolution codes)
  - IT: We want \( \sum_i R_i \gg C \)!
2-user MAC: IT formalism

- 2-input channel: $P_{Y|X_1,X_2}$ (memoryless)
- Random messages $W_1 \in [2^{nR_1}], W_2 \in [2^{nR_2}]
- Encoders: $X_1^n = f_1(W_1), X_2^n = f_2(W_2)$
- Joint decoder: $(\hat{W}_1, \hat{W}_2) = g(Y)$
- Joint probability of error:

$$\Pr[W_1 = \hat{W}_1, W_2 = \hat{W}_2] \geq 1 - \epsilon.$$
2-user MAC: IT formalism

- 2-input channel: $P_{Y|X_1,X_2}$ (memoryless)
- Random messages $W_1 \in \begin{bmatrix} 2^{nR_1} \end{bmatrix}, W_2 \in \begin{bmatrix} 2^{nR_2} \end{bmatrix}$
- Encoders: $X_1^n = f_1(W_1), X_2^n = f_2(W_2)$
- Joint decoder: $(\hat{W}_1, \hat{W}_2) = g(Y)$
- Joint probability of error: $\mathbb{P}[W_1 = \hat{W}_1, W_2 = \hat{W}_2] \geq 1 - \epsilon.$

- FBL fundamental limit (region): $R^*(n, \epsilon) = \{(R_1, R_2) : \exists (2^{nR_1}, 2^{nR_2}, \epsilon)-code\}$

- Asymptotics: $[\cdot] = \text{closure}$
  $$C_\epsilon = \left[ \liminf_{n \to \infty} R^*(n, \epsilon) \right], \quad C = \bigcap_{\epsilon > 0} C_\epsilon$$
Theorem (Ahlswede-Liao (capacity) + Dueck (Strong converse))

\[ C = C_\epsilon = \left[ \text{co} \left\{ \bigcup_{P_{X_1}, P_{X_2}} \text{Penta}(P_{X_1}, P_{X_2}) \right\} \right] \]

\[ \text{Penta}(P_{X_1}, P_{X_2}) \triangleq \left\{ (R_1, R_2) : \begin{array}{l} R_1 + R_2 \leq I(X_1, X_2; Y) \\ R_1 \leq I(X_1; Y|X_2) \\ R_2 \leq I(X_2; Y|X_1) \end{array} \right\} \]

- \text{co\{\cdot\}} – convex hull
- **Fun fact:** w/o syncronization \( C = \bigcup \text{Penta} \) but w/o \( \text{co\{\cdot\}} \) !
2-user MAC: capacity region

Theorem (Ahlswede-Liao (capacity) + Dueck (Strong converse))

\[
C = C_\epsilon = \left[ \operatorname{co} \left\{ \bigcup_{P_{X_1}, P_{X_2}} \text{Penta}(P_{X_1}, P_{X_2}) \right\} \right]
\]

\[
\text{Penta}(P_{X_1}, P_{X_2}) \triangleq \left\{ (R_1, R_2) : \begin{align*}
R_1 + R_2 &\leq I(X_1, X_2; Y) \\
R_1 &\leq I(X_1; Y|X_2) \\
R_2 &\leq I(X_2; Y|X_1)
\end{align*} \right\}
\]

- \(\operatorname{co}\{\cdot\}\) – convex hull
- **Fun fact:** w/o synchronization \(C = \bigcup \text{Penta}\) but w/o \(\operatorname{co}\{\cdot\}\)!
- Not true with cost constraints. In that case need time-sharing:

\[
C = \bigcup_{X_1, X_2, U} \left\{ (R_1, R_2) : \begin{align*}
R_1 + R_2 &\leq I(X_1, X_2; Y|U) \\
R_1 &\leq I(X_1; Y|X_2, U) \\
R_2 &\leq I(X_2; Y|X_1, U)
\end{align*} \right\}.
\]
Capacity = Union of pentagons

\[ \text{Penta}(P_{X_1}, P_{X_2}) \triangleq \left\{ (R_1, R_2) : \begin{align*}
R_1 + R_2 &\leq I(X_1, X_2; Y) \\
R_1 &\leq I(X_1; Y | X_2) \\
R_2 &\leq I(X_2; Y | X_1)
\end{align*} \right\} \]

Note: After taking \( \bigcup_{P_{X_1}, P_{X_2}} \) and convex-hull, resulting region may be curvilinear!
Theorem

\[ C = C_\epsilon = \left[ \text{co} \left\{ \bigcup_{P_{X_1}, P_{X_2}} \text{Penta}(P_{X_1}, P_{X_2}) \right\} \right] \]

Here is a standard proof

- Weak-converse:
  - sum-rate
    \[ R_1 + R_2 \lesssim \frac{1}{n} I(X_1^n, X_2^n; Y^n) \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i). \]
  - genie gives \( X_1^n \) to decoder
    \[ R_2 \lesssim \frac{1}{n} I(X_2^n; Y^n | X_1^n) \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{2i}; Y_i | X_{1i}). \]
  - Hence \( (R_1, R_2) \in \frac{1}{n} \sum_i \text{Penta}(P_{X_{1i}}, P_{X_{2i}}) \)
MAC theorem: standard proof (outline)

Theorem

\[ C = C_\epsilon = \co \left\{ \bigcup_{P_{X_1}, P_{X_2}} \text{Penta}(P_{X_1}, P_{X_2}) \right\} \]

Here is a standard proof

• Achievability:
  • Fix \( P_{X_1}, P_{X_2} \).
  • Generate codewords for user \( i \) from \((P_{X_1})^\otimes n\) iid
  • Decode via joint-typicality
  • Have \((M_1 - 1)(M_2 - 1)\) possibilities with both \( \hat{W}_1, \hat{W}_2 \) wrong (each w.p. \( \leq 2^{-nI(X_1, X_2; Y)} \))
  • Have \( M_i - 1 \) possibilities with \( \hat{W}_i \) wrong (each w.p. \( \leq 2^{-nI(X_i; Y|X_i)} \))
  • Hence, if \((R_1, R_2) \in \text{Penta}(P_{X_1}, P_{X_2})\) all three types of errors are small.
  • Let us understand this more carefully...
MAC achievability: details I

- Gen. $M_1 = 2^{nR_1}$ codewords $C_i^{iid} \sim (P_{X_1})^\otimes n$
- Gen. $M_2 = 2^{nR_2}$ codewords $D_i^{iid} \sim (P_{X_2})^\otimes n$
- True message $W_1 = i_0, W_2 = j_0$.
- Decoder sees $y^n$. How to decode?
MAC achievability: details I

- Gen. $M_1 = 2^{nR_1}$ codewords $C_i \sim iid (P_{X_1})^\otimes n$
- Gen. $M_2 = 2^{nR_2}$ codewords $D_i \sim iid (P_{X_2})^\otimes n$
- True message $W_1 = i_0, W_2 = j_0$.
- Decoder sees $y^n$. How to decode?
- Why is this not the same as decoding single-user $M_1 \times M_2$-size code?
MAC achievability: details I

- Gen. $M_1 = 2^{nR_1}$ codewords $C_i \overset{iid}{\sim} (P_{X_1})^\otimes n$
- Gen. $M_2 = 2^{nR_2}$ codewords $D_i \overset{iid}{\sim} (P_{X_2})^\otimes n$
- True message $W_1 = i_0, W_2 = j_0$.
- Decoder sees $y^n$. How to decode?
- Why is this not the same as decoding single-user $M_1 \times M_2$-size code?

Extra structure: $(C_{i_0}, D_{j}) \not\perp (C_{i_0}, D_{j_0})$
• Decoder sees $y^n$. How to decode?

• A good test for rejecting $(M_1 - 1)(M_2 - 1)$ codewords in $(P_{12})$:

$$(T_{12}) \quad i(c_i, d_j; y^n) \leq \gamma_{12} \implies \text{remove } (i, j) \text{ from consideration}$$

• $i(c, d; y^n) \triangleq \log \frac{P_{Y^n|X_1^n, X_2^n}(y^n|c, d)}{P_{Y^n}(y^n)}$

• Standard bound: $\forall i \neq i_0, j \neq j_0$:

$$\mathbb{P}[i(C_i, D_j; Y^n) > \gamma_{12}] \leq e^{-\gamma_{12}}$$

• Set $\gamma_{12} = \log(M_1 M_2) + \tau$ then test $(T_{12})$ filters all $(i, j) \in (P_{12})$
- Decoder sees $y^n$. How to decode?
- A good test for rejecting $(M_2 - 1)$ codewords in $(P_2)$:
  \[(T_2) \quad i(d_j; y^n|c_i) \leq \gamma_2 \Rightarrow \text{remove (i, j) from consideration}\]
- $i(d; y^n|c) \triangleq \log \frac{P_{Y^n|X^n_1,X^n_2}(y^n|c,d)}{P_{Y^n|X^n_1}(y^n|c)}$
- Standard bound: $\forall j \neq j_0$:
  \[
P[i(D_j; Y^n|C_{i_0}) > \gamma_2] \leq e^{-\gamma_2}\]
- Set $\gamma_2 = \log(M_2) + \tau$ then test $(T_2)$ filters all $(i_0, j) \in (P_2)$
• Decoder sees $y^n$. How to decode?

$$(T_{12}) \quad i(c_i, d_j; y^n) \leq n(R_1 + R_2) + \tau \quad \Rightarrow \text{remove } (i, j)$$

$$(T_1) \quad i(c_i; y^n|d_j) \leq nR_1 + \tau \quad \Rightarrow \text{remove } (i, j)$$

$$(T_2) \quad i(d_j; y^n|c_i) \leq nR_2 + \tau \quad \Rightarrow \text{remove } (i, j)$$

• This achieves:

$$\epsilon \leq 3e^{-\tau} + \mathbb{P} \left[ \{ i(X_1^n, X_2^n; Y^n) \leq n(R_1 + R_2) + \tau \} \cup \{ i(X_1^n; Y^n|X_2^n) \leq nR_1 + \tau \} \cup \{ i(X_2^n; Y^n|X_1^n) \leq nR_2 + \tau \} \right].$$

• By CLT a $(R_1, R_2)$ within $\frac{1}{\sqrt{n}}$ of the boundary of Penta is achievable.
• Decoder sees $y^n$. How to decode?

\[(T_{12}) \quad i(c_i, d_j; y^n) \leq n(R_1 + R_2) + \tau \quad \Rightarrow \text{remove } (i, j)\]

\[(T_1) \quad i(c_i; y^n|d_j) \leq nR_1 + \tau \quad \Rightarrow \text{remove } (i, j)\]

\[(T_2) \quad i(d_j; y^n|c_i) \leq nR_2 + \tau \quad \Rightarrow \text{remove } (i, j)\]

• This achieves:

$$\epsilon \leq 3e^{-\tau} + \mathbb{P}\left\{i(X_1^n, X_2^n; Y^n) \leq n(R_1 + R_2) + \tau\right\} \cup \left\{i(X_1^n; Y^n|X_2^n) \leq nR_1 + \tau\right\} \cup \left\{i(X_2^n; Y^n|X_1^n) \leq nR_2 + \tau\right\}.$$

• By CLT a $(R_1, R_2)$ within $\frac{1}{\sqrt{n}}$ of the boundary of Penta is achievable.

• Typical decoding
  ▶ Use $(T_{12})$ rule – this is like decoding single-user $M_1 \times M_2$-code (LDPC+LDGM structure!)
  ▶ After applying it, most often get only one (true) message left (!)
• Decoder sees $y^n$. How to decode?

\[
\begin{align*}
(T_{12}) & \quad i(c_i, d_j; y^n) \leq n(R_1 + R_2) + \tau \quad \Rightarrow \text{remove } (i, j) \\
(T_1) & \quad i(c_i; y^n | d_j) \leq nR_1 + \tau \quad \Rightarrow \text{remove } (i, j) \\
(T_2) & \quad i(d_j; y^n | c_i) \leq nR_2 + \tau \quad \Rightarrow \text{remove } (i, j)
\end{align*}
\]

• This achieves:

\[
\epsilon \leq 3e^{-\tau} + \mathbb{P}\left[ \left\{ i(X_1^n, X_2^n; Y^n) \leq n(R_1 + R_2) + \tau \right\} \cup \left\{ i(X_1^n; Y^n | X_2^n) \leq nR_1 + \tau \right\} \cup \left\{ i(X_2^n; Y^n | X_1^n) \leq nR_2 + \tau \right\} \right].
\]

• By CLT a $(R_1, R_2)$ within $\frac{1}{\sqrt{n}}$ of the boundary of Penta is achievable.

• Typical decoding
  ▶ Use $(T_{12})$ rule – this is like decoding single-user $M_1 \times M_2$-code (LDPC+LDGM structure!)
  ▶ After applying it, most often get only one (true) message left (!)
  ▶ Unless $R_1 = I(X_1; Y | X_2) + O\left(\frac{1}{\sqrt{n}}\right)$.
  ▶ In this case, many $(i, j)$’s remain. But they are all in one column!
  ▶ Hence decode $W_2$. Conditioned on $X_2$ – decode $M_1$-code.
Example: Binary Adder Channel (BAC)

\[ Y = X_1 + X_2 \quad X_i \in \{0, 1\}, Y \in \{0, 1, 2\} \]

- Maximal sum-rate:

\[ C_{sum} = \max_{A,B} I(A, B; Y) = \max H(A + B) = \frac{3}{2} \log 2 \]

- Each user can send 1 bit/ch.use. But together \( \frac{3}{2} \) bit/ch.use. How?
Example: Binary Adder Channel (BAC)

- Take $R_1 = 1$. Then $X_2 \rightarrow Y$ sees channel:
Example: Binary Adder Channel (BAC)

- Take $R_1 = 1$. Then $X_2 \rightarrow Y$ sees channel:

\[
\begin{array}{c}
0 \\
1 \\
\end{array} \xrightarrow{\frac{1}{2}} \begin{array}{c}
\frac{1}{2} \\
\frac{1}{2} \\
2 \\
\end{array} \xrightarrow{\frac{1}{2}} \begin{array}{c}
0 \\
1 \\
\end{array} = \text{BEC}(1/2)
Example: Binary Adder Channel (BAC)

- Take $R_1 = 1$. Then $X_2 \to Y$ sees channel:

\[
\begin{array}{c c}
0 & \frac{1}{2} \\
\frac{1}{2} & 0 \\
1 & \frac{1}{2} \\
\frac{1}{2} & 2
\end{array}
\]

$= \text{BEC}(1/2)$

- successive interference cancellation (SIC):

\[
\begin{array}{c}
A^n \quad Y^n \quad \text{Dec} \quad \hat{A}^n \\
B^n \quad \hat{B}^n
\end{array}
\]
Example: Binary Adder Channel (BAC)

\[ Y = X_1 + X_2 \quad X_i \in \{0, 1\}, Y \in \{0, 1, 2\} \]

- Analyzing FBL achievability we can show: (maximal sumrate)

\[ R^*_\text{sum}(n, \epsilon) \geq \frac{3}{2} - \sqrt{\frac{1}{4n}} Q^{-1}(\epsilon) + O(\log n) . \]

- Open problem: Prove \( R^*_\text{sum}(n, \epsilon) \leq \frac{3}{2} + \sqrt{\frac{1}{n}} K_\epsilon \)
Example: Binary Adder Channel (BAC)

\[ Y = X_1 + X_2 \quad \text{where} \quad X_i \in \{0, 1\}, \quad Y \in \{0, 1, 2\} \]

- Analyzing FBL achievability we can show: (maximal sum rate)

\[ R^*_{sum}(n, \epsilon) \geq \frac{3}{2} - \sqrt{\frac{1}{4n}} Q^{-1}(\epsilon) + O(\log n) . \]

- **Open problem:** Prove \( R^*_{sum}(n, \epsilon) \leq \frac{3}{2} + \sqrt{\frac{1}{n}} K_\epsilon \)
- ... not even asking for \( K_\epsilon < 0 \)
- ... So far best-known result (Ahslwede): \( R^*_{sum} \leq \frac{3}{2} + c\sqrt{\frac{1}{n}} \log n \)
Example: Binary Adder Channel (BAC)

\[ Y = X_1 + X_2 \quad X_i \in \{0, 1\}, Y \in \{0, 1, 2\} \]

- Analyzing FBL achievability we can show: (maximal sum rate)

\[ R^*_{\text{sum}}(n, \epsilon) \geq \frac{3}{2} - \sqrt{\frac{1}{4n}} Q^{-1}(\epsilon) + O(\log n). \]

- **Open problem:** Prove \( R^*_{\text{sum}}(n, \epsilon) \leq \frac{3}{2} + \sqrt{\frac{1}{n}} K\epsilon \)
- ... not even asking for \( K\epsilon < 0 \)
- ... So far best-known result (Ahslwede): \( R^*_{\text{sum}} \leq \frac{3}{2} + c \sqrt{\frac{1}{n}} \log n \)
- The state is so bad that even for \( \epsilon = 0 \) we only know (Fano):

\[ R^*_{\text{sum}}(n, \epsilon = 0) \leq \frac{3}{2} \]

- **Open problem:** Prove \( \lim_{n \to \infty} R^*_{\text{sum}}(n, \epsilon = 0) < \frac{3}{2} \).
Example: Binary Adder Channel (BAC)

\[ Y = X_1 + X_2 \quad X_i \in \{0, 1\}, Y \in \{0, 1, 2\} \]

- Analyzing FBL achievability we can show: (maximal sumrate)

\[
R^*_{sum}(n, \epsilon) \geq \frac{3}{2} - \sqrt{\frac{1}{4n}Q^{-1}(\epsilon)} + O(\log n).
\]

- Open problem: Prove \( R^*_{sum}(n, \epsilon) \leq \frac{3}{2} + \sqrt{\frac{1}{n}K_{\epsilon}} \)

- Conjecture: [Ajjanagadde-P.'15] for all \(0 < \alpha < 1\)

\[
\max_{A^n \perp \perp B^n} H_\alpha(A^n + B^n) = nH_\alpha\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)
\]

where \(H_\alpha(\cdot)\) is Renyi entropy.

- If true implies Open problem. How?
MAC: revisit weak-converse (genie)

\[ \mathbb{P} : \begin{align*}
W_1 &\rightarrow X_1 \quad W_1 \\
W_2 &\rightarrow X_2 \quad W_2
\end{align*} \]

\[ \mathbb{Q} : \begin{align*}
\hat{W}_1 &\rightarrow X_1 \quad \hat{W}_1 \\
\hat{W}_2 &\rightarrow X_2 \quad \hat{W}_2
\end{align*} \]

\[ \mathbb{P}[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \]

\[ \mathbb{Q}[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1} \]

... apply data processing of \( D(\cdot \| \cdot) \) ...

\[ \downarrow \]

\[ d(1 - \epsilon \parallel \frac{1}{M_1}) \leq D(P_{Y|X_1X_2} \| Q_{Y|X_1} | P_{X_1} P_{X_2}) \]

Optimizing \( Q_{Y|X_1} \):

\[ \log M_1 \leq \frac{I(X_1; Y| X_2) + h(\epsilon)}{1 - \epsilon} \]
MAC: revisit weak-converse (genie)

\[ P[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \]

\[ Q[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1 M_2} \]

\[ d(1 - \epsilon \| \frac{1}{M_1}) \leq D(P_{Y|X_1 X_2} \| Q_Y | P_{X_1} P_{X_2}) \]

Optimizing \( Q_Y \):

\[ \log M_1 M_2 \leq \frac{I(X_1, X_2; Y) + h(\epsilon)}{1 - \epsilon} \]

Together with previous: full (pentagon) weak converse
MAC: towards strong-converse

\[ P[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \quad \text{and} \quad Q[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1 M_2} \]

\[ \ldots \text{use Renyi } D_{\lambda}(\cdot \| \cdot) \ldots \]

\[ D_{\lambda}(P_{X_1 X_2 Y} \| P_{X_1} P_{X_2} Q_Y) \geq d_{\lambda}(1 - \epsilon \| \frac{1}{M_1 M_2}) \]

Selecting \( \lambda = 1 + \frac{1}{\sqrt{n}} \) yields (for BAC)

\[ \log M_1, M_2 \leq \sup_{A^n \perp \perp B^n} H_{\alpha_n}(A^n + B^n) + K \sqrt{n} \]

with \( \alpha_n = 1 - \frac{1}{\sqrt{n}} \).
Classical MAC: summary

- Trivially generalizes to $K$-user MAC:
  \[
Penta = \left\{ (R_1, \ldots, R_K) : \sum_{i \in S} R_i \leq I(X_S; Y | X_{Sc}) \forall S \subset [K] \right\}
  \]

- Classic IT: Fix $K$ let $n \to \infty$.
- Use joint probability of error:
  \[
  \mathbb{P}[W_1 = \hat{W}_1, \ldots, W_K = \hat{W}_k] \geq 1 - \epsilon.
  \]

- New FBL issue: for $K = 100$ need $2^{100}$ tests in achievability.
Classical MAC: summary

- Trivially generalizes to $K$-user MAC:

$\text{Penta} = \{(R_1, \ldots, R_K) : \sum_{i \in S} R_i \leq I(X_S; Y | X_{Sc}) \forall S \subset [K]\}$

- Classic IT: Fix $K$ let $n \to \infty$.

- Use joint probability of error:

$\mathbb{P}[W_1 = \hat{W}_1, \ldots, W_K = \hat{W}_K] \geq 1 - \epsilon$.

- New FBL issue: for $K = 100$ need $2^{100}$ tests in achievability.

- What is new today?
  - Many-user scaling [D. Guo et al]: $K = \mu n$, $n \to \infty$
  - New probability of error [P.'17]: $\frac{1}{K} \sum_i \mathbb{P}[W_i \neq \hat{W}_i] \leq \epsilon$
  - Same-codebook coding [P.'17]: $X_i \in \mathcal{C}$ for all $i$. 
Gaussian MAC. Modulation

Let’s put on our engineering boots.
The classical model: $K$-user multiple-access channel

\[ Y(t) = X_1(t) + \cdots + X_K(t) + Z(t) \]
The classical model: $K$-user multiple-access channel

$Y(t) = X_1(t) + \cdots + X_K(t) + Z(t)$

- Users send coded waveforms $X_j(t)$
- Additive Gaussian noise $Z(t)$
- Base station’s job: estimate $X_j$ from the knowledge of $Y(t)$
How to avoid inter-user interference?

These are called orthogonal schemes. Key problem: resources divided among active and inactive users (or need costly resource ack/grant protocol). In IoT most are inactive ⇒ huge waste of bandwidth.
How to avoid inter-user interference?

These are called **orthogonal schemes**

**Key problem:** resources divided among active and inactive (!) users

(or need costly resource ack/grant protocol)

in IoT most are inactive ⇒ huge waste of bandwidth
This “pie-slicing” philosophy comes from:

- **Given**: $W$ Hz bandwidth and duration $T$ sec:
- **By XYZ Theorem**: d.o.f. $n = 2WT$
  $$XYZ \in \{\text{Kotelnikov, Nyquist, Shannon, Slepian, ...}\}$$
- **TDMA, FDMA, CDMA**: just different bases in $\mathbb{R}^{2WT}$.
  (Fine print: CDMA = Orthogonal CDMA here).

... cheating: user $K$’s power is $2^2K$ larger than user 1’s.

**Challenge**: users only allowed to send $\pm 1$, can we have $K \gg n$?
Orthogonal and non-orthogonal multiple access (NOMA)

This “pie-slicing” philosophy comes from:

- Given: $W$ Hz bandwidth and duration $T$ sec:
- By XYZ Theorem: d.o.f. $n = 2WT$
  
  $XYZ \in \{ \text{Kotelnikov, Nyquist, Shannon, Slepian, ...} \}$
- TDMA, FDMA, CDMA: just different bases in $\mathbb{R}^{2WT}$.
  (Fine print: CDMA = Orthogonal CDMA here).
- Is there value in having $K > n$? (non-orthogonal signalling)
- Is it even possible to have $K \gg n$ or even $K \gg n$?
Orthogonal and non-orthogonal multiple access (NOMA)

This “pie-slicing” philosophy comes from:

- Given: $W \text{ Hz bandwidth and duration } T \text{ sec:}$
- By XYZ Theorem: d.o.f. $n = 2WT$
  
  $XYZ \in \{ \text{Kotelnikov, Nyquist, Shannon, Slepian, ...} \}$
- TDMA, FDMA, CDMA: just different bases in $\mathbb{R}^{2WT}$.
  (Fine print: CDMA = Orthogonal CDMA here).
- Is there value in having $K > n$? (non-orthogonal signalling)
- Is it even possible to have $K > n$ or even $K \gg n$?
- Silly: Take $n = 1$ and let user $j$ send a bit via $\{0, 2^j\}$. 
This “pie-slicing” philosophy comes from:

- Given: \( W \) Hz bandwidth and duration \( T \) sec:
- By XYZ Theorem: d.o.f. \( n = 2WT \)
  
  \[ XYZ \in \{ \text{Kotelnikov, Nyquist, Shannon, Slepian, \ldots} \} \]
- TDMA, FDMA, CDMA: just different bases in \( \mathbb{R}^{2WT} \).
  (Fine print: CDMA = Orthogonal CDMA here).
- Is there value in having \( K > n \)? (non-orthogonal signalling)
- Is it even possible to have \( K > n \) or even \( K \gg n \)?
- Silly: Take \( n = 1 \) and let user \( j \) send a bit via \( \{0, 2^j\} \).
- ... cheating: user \( K \)’s power is \( 2^{2K} \) larger than user 1’s.
Orthogonal and non-orthogonal multiple access (NOMA)

This “pie-slicing” philosophy comes from:

- Given: \( W \) Hz bandwidth and duration \( T \) sec:
- By XYZ Theorem: d.o.f. \( n = 2WT \)
  \( XYZ \in \{ \text{Kotelnikov, Nyquist, Shannon, Slepian, ...} \} \)
- TDMA, FDMA, CDMA: just different bases in \( \mathbb{R}^{2WT} \).
  (Fine print: CDMA = Orthogonal CDMA here).
- Is there value in having \( K > n \)? (non-orthogonal signalling)
- Is it even possible to have \( K > n \) or even \( K \gg n \)?
- Silly: Take \( n = 1 \) and let user \( j \) send a bit via \( \{0, 2^j\} \).
- ... cheating: user \( K \)'s power is \( 2^{2K} \) larger than user 1’s.
- Challenge: users only allowed to send \( \pm 1 \), can we have \( K \gg n \)?
Achieving capacity of $K$-user BAC with zero-error

\[ Y = \sum_{j=1}^{K} X_j \quad X_i \in \{\pm 1\} \]

- Known: $C_{\text{sum}}(K) = H(\text{Bin}(K, 1/2)) \approx \frac{1}{2} \log K$.
- IOW, for sending 1-bit (each) the frame-length $n \approx \frac{2K}{\log_2 K} \ll K$.

How can $K > n$ users signal in $n$ dimensions simultaneously?
Achieving capacity of $K$-user BAC with zero-error

\[ Y = \sum_{j=1}^{K} X_j \quad X_i \in \{\pm 1\} \]

- Known: \( C_{\text{sum}}(K) = H(\text{Bin}(K, 1/2)) \approx \frac{1}{2} \log K \).
- IOW, for sending 1-bit (each) the frame-length \( n \approx \frac{2K}{\log_2 K} \ll K \).

**How can \( K > n \) users signal in \( n \) dimensions simultaneously?**

- Lindström, Cantor-Mills, Khachatrian-Martirossian: even with zero-error!

First, recall a particularly nice orthogonal basis:

\[
H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad H_{m+1} = \begin{bmatrix} H_m & H_m \\ H_m & -H_m \end{bmatrix}
\]

(each user is modulating his row)

- K.-M. noticed you can add more rows!
Recursive construction (Cantor-Mills, Khachatrian-Martirossian)

How can $K > n$ users signal in $n$ dimensions simultaneously?

- Walsh-Hadamard basis:
  
  $H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$  
  
  $H_2 = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \end{bmatrix}$  
  
  $H_{m+1} = \begin{bmatrix} H_m & H_m \\ H_m & -H_m \end{bmatrix}$

- K.-M. signals:
  
  $A_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$  
  
  $A_2 = \begin{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix} \end{bmatrix}$

- Key property: $x \mapsto x A_m$ is injective on $\{\pm 1\}^{K_m}$, $K_m = \frac{m}{2} 2^m + 1$
How can $K > n$ users signal in $n$ dimensions simultaneously?

- Walsh-Hadamard basis:

\[
H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \quad H_{m+1} = \begin{bmatrix} H_m & H_m \\ H_m & -H_m \end{bmatrix}
\]

- K.-M. signals:

\[
A_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad \tilde{A}_{m+1} = \begin{bmatrix} A_m & A_m \\ A_m & -A_m \end{bmatrix}
\]

- Key property: $x \mapsto xA_m$ is injective on $\{\pm1\}^{K_m}$, $K_m = \frac{m}{2}2^m + 1$

- Number of users at dimension $n$: $K \approx \frac{1}{2}n \log_2 n$ (optimal!)
Recursive construction (Cantor-Mills, Khachatrian-Martirossian)

How can $K > n$ users signal in $n$ dimensions simultaneously?

- **Walsh-Hadamard basis:**
  
  $$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$
  
  $$H_{m+1} = \begin{bmatrix} H_m & H_m \\ H_m & -H_m \end{bmatrix}$$

- **K.-M. signals:**
  
  $$A_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$
  
  $$A_{m+1} = \begin{bmatrix} A_m & A_m \\ A_m & -A_m \end{bmatrix}$$

- **Key property:** $x \mapsto x A_m$ is injective on $\{\pm 1\}^{K_m}$, $K_m = \frac{m}{2} 2^m + 1$

- **Number of users at dimension $n$:** $K \approx \frac{1}{2} n \log_2 n$ (optimal!)

- **Idea:** $(\pm 1)^{2^m} \cdot H_m$ has many “holes”; add $\pm 1$-vectors there.
• Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\}^m$ and $v2K_{m-1}+1 = 0$.

• Equivalently: $v \in \{0, 1\}^m$ (just use $v \mapsto \frac{1+v}{2}$)
\begin{align*}
\tilde{A}_{m+1} &= \begin{bmatrix}
A_m & A_m \\
A_m & -A_m
\end{bmatrix} \\
A_{m+1} &= \begin{bmatrix}
A_m & A_m \\
A_m & -A_m
\end{bmatrix}
\end{align*}

- Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\} \otimes K_m$ and $v_{2K_m-1+1} = 0$.
- Equivalently: $v \in \{0, 1\} \otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)
- Let $v = [x \ y \ z]$ and

$$[x \ y \ z]\tilde{A}_m = [g \ h]$$
• Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm1\}^\otimes K_m$ and $v2K_{m-1}+1 = 0$.
• Equivalently: $v \in \{0, 1\}^\otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)
• Let $v = [x \ y \ z]$ and

$$[x \ y \ z]\tilde{A}_m = [g \ h] \Rightarrow g - h = [x \ y \ z] \begin{pmatrix} 0 \\ 2A_{m-1} \\ 2I_{2m-1} \end{pmatrix}$$
Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\} \otimes K_m$ and $v2K_{m-1}+1 = 0$.

Equivalently: $v \in \{0, 1\} \otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)

Let $v = [x \ y \ z]$ and

$$[x \ y \ z] \tilde{A}_m = [g \ h] \Rightarrow g - h = [x \ y \ z] \begin{pmatrix} 0 \\ 2A_{m-1} \\ 2I_{2^{m-1}} \end{pmatrix}$$

$z_1 = 0$, so by adding $(g - h)_1$ to $(g - h)_\ell$ we get:

$$2z_\ell = (g - h)_1 + (g - h)_\ell - 2y \cdot v_\ell \quad \ell = 2, \ldots, 2^{m-1}$$

where $v_\ell$ is sum of 1-st and $\ell$-th column of $A_{m-1}$.
• Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\} \otimes K_m$ and $v2K_{m-1} + 1 = 0$.

• Equivalently: $v \in \{0, 1\} \otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)

• Let $v = [x\ y\ z]$ and

$$[x\ y\ z]\tilde{A}_m = [g\ h] \Rightarrow g - h = [x\ y\ z] \begin{pmatrix} 0 \\ 2A_{m-1} \\ 2I_{2m-1} \end{pmatrix}$$

• $z_1 = 0$, so by adding $(g - h)_1$ to $(g - h)_\ell$ we get:

$$2z_\ell = (g - h)_1 + (g - h)_\ell - 2y \cdot v_\ell \quad \ell = 2, \ldots, 2^{m-1}$$

where $v_\ell$ is sum of 1-st and $\ell$-th column of $A_{m-1}$

• Key: $v_\ell$’s entries are $\{0, 2\}$. Take mod 4 of (*) and decode $z_\ell$’s!

• Subtracting $z_\ell$’s we get system:

$$[x\ y] \begin{pmatrix} A_{m-1} & A_{m-1} \\ A_{m-1} & -A_{m-1} \end{pmatrix} = [g'\ h']$$
• Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\} \otimes K_m$ and $v_{2K_m-1+1} = 0$.

• Equivalently: $v \in \{0, 1\} \otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)

• Let $v = [x \ y \ z]$ and

$$[x \ y \ z] \tilde{A}_m = [g \ h] \Rightarrow g - h = [x \ y \ z] \begin{pmatrix} 0 \\ 2A_{m-1} \\ 2I_{2m-1} \end{pmatrix}$$

• $z_1 = 0$, so by adding $(g - h)_1$ to $(g - h)_\ell$ we get:

$$2z_\ell = (g - h)_1 + (g - h)_\ell - 2y \cdot v_\ell \quad \ell = 2, \ldots, 2^{m-1}$$

where $v_\ell$ is sum of 1-st and $\ell$-th column of $A_{m-1}$

• Key: $v_\ell$’s entries are $\{0, 2\}$. Take mod 4 of $(*)$ and decode $z_\ell$’s!

• Subtracting $z_\ell$’s we get system:

$$[x \ y] \begin{pmatrix} A_{m-1} & A_{m-1} \\ A_{m-1} & -A_{m-1} \end{pmatrix} = [g' \ h'] \quad \Rightarrow \quad xA_{m-1} = \frac{g' + h'}{2}$$
• Want to show: $v$ is decodable from $v\tilde{A}_m$ for any $v \in \{\pm 1\} \otimes K_m$ and $v_{2K_{m-1}+1} = 0$.
• Equivalently: $v \in \{0, 1\} \otimes K_m$ (just use $v \mapsto \frac{1+v}{2}$)
• Let $v = [x \ y \ z]$ and

$$[x \ y \ z]\tilde{A}_m = [g \ h] \Rightarrow g - h = [x \ y \ z] \begin{pmatrix} 0 \\ 2A_{m-1} \\ 2I_{2m-1} \end{pmatrix}$$

• $z_1 = 0$, so by adding $(g - h)_1$ to $(g - h)_\ell$ we get:

\[
\begin{align*}
(\ast) & \quad 2z_\ell = (g - h)_1 + (g - h)_\ell - 2y \cdot v_\ell \quad \ell = 2, \ldots, 2^{m-1} \\
\end{align*}
\]

where $v_\ell$ is sum of 1-st and $\ell$-th column of $A_{m-1}$
• Key: $v_\ell$’s entries are $\{0, 2\}$. Take mod 4 of $(\ast)$ and decode $z_\ell$’s!
• Subtracting $z_\ell$’s we get system:

$$[x \ y] \begin{pmatrix} A_{m-1} & A_{m-1} \\ A_{m-1} & -A_{m-1} \end{pmatrix} = [g' \ h'] \Rightarrow xA_{m-1} = \frac{g' + h'}{2} \Rightarrow \text{induct}$$
• When user inputs are constrained (to $\pm 1$), can have $K \gg n$ and still recover inputs.
• Total information grows with $K$: $H(X_1 + \cdots + X_K) \sim \frac{1}{2} \log K$. (This is similar to $\frac{1}{2} \log (1 + KP)$ in GMAC.)
• Lots of smart ideas in MAC codes.
Reflections

- When user inputs are constrained (to $\pm 1$), can have $K \gg n$ and still recover inputs.
- Total information grows with $K$: $H(X_1 + \cdots + X_K) \sim \frac{1}{2} \log K$.
  (This is similar to $\frac{1}{2} \log(1 + KP)$ in GMAC.)
- Lots of smart ideas in MAC codes.
- Information theory structures it all into:

\[
C = \bigcup_{X_1, \ldots, X_K, U} \{(R_1, \ldots, R_K) : R_S \leq I(X_S; Y|X_{Sc}, U)\}
\]
• When user inputs are constrained (to ±1), can have $K \gg n$ and still recover inputs.
• Total information grows with $K$: $H(X_1 + \cdots + X_K) \sim \frac{1}{2} \log K$. (This is similar to $\frac{1}{2} \log(1 + KP)$ in GMAC.)
• Lots of smart ideas in MAC codes.
• Information theory structures it all into:

$$C = \bigcup_{X_1, \ldots, X_K, U} \{(R_1, \ldots, R_K) : R_S \leq I(X_S; Y | X_{Sc}, U)\}$$

• Similar to how all the smarts (Reed-Muller, BCH, LDPC, Polar, ...) are hidden behind

$$C = \max_X I(X; Y)$$
• When user inputs are constrained (to $\pm 1$), can have $K \gg n$ and still recover inputs.
• Total information grows with $K$: $H(X_1 + \cdots + X_K) \sim \frac{1}{2} \log K$.
  (This is similar to $\frac{1}{2} \log(1 + KP)$ in GMAC.)
• Lots of smart ideas in MAC codes.
• Information theory structures it all into:

$$C = \bigcup_{X_1, \ldots, X_K, U} \{(R_1, \ldots, R_K) : R_S \leq I(X_S; Y | X_{Sc}, U)\}$$

• Similar to how all the smarts (Reed-Muller, BCH, LDPC, Polar, ...) are hidden behind

$$C = \max_X I(X; Y)$$

• We understand that “pie-slicing” point of view of radio-MAC is wrong. What is right?
2-user Gaussian MAC

\[ Y = X_1 + X_2 + Z \]
\[ Z \sim \text{iid } \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]
2-user Gaussian MAC

\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Evaluating capacity region:

\[ R_1 + R_2 \leq I(X_1, X_2; Y) \leq \frac{1}{2} \log(1 + P_1 + P_2) \]

\[ R_i \leq I(X_i; Y | X_i) = I(X_i; X_i + Z) \leq \frac{1}{2} \log(1 + P_i) \]
2-user Gaussian MAC

\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Evaluating capacity region:

\[ R_1 + R_2 \leq I(X_1, X_2; Y) \leq \frac{1}{2} \log(1 + P_1 + P_2) \]

\[ R_i \leq I(X_i; Y|X_i) = I(X_i; X_i + Z) \leq \frac{1}{2} \log(1 + P_i) \]
2-GMAC rates for TDMA

\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Here is a TDMA:
  - Partition block: \( n = \lambda n + (1 - \lambda)n \)
  - User 1 sends in \( \lambda n \): \( R_1 = \frac{1}{2} \log(1 + P_1) \)
  - User 2 sends in \( \bar{\lambda} n \): \( R_2 = \frac{1}{2} \log(1 + P_2) \)

Note: low-complexity decoder – two users are decoded separately.
2-GMAC rates for TDMA

\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Here is a TDMA:
  - Partition block: \( n = \lambda n + (1 - \lambda) n \)
  - User 1 sends in \( \lambda n \): \( R_1 = \frac{1}{2} \log(1 + P_1) \)
  - User 2 sends in \( \bar{\lambda} n \): \( R_2 = \frac{1}{2} \log(1 + P_2) \)
2-GMAC rates for TDMA

\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Here is a TDMA:
  - Partition block: \( n = \lambda n + (1 - \lambda)n \)
  - User 1 sends in \( \lambda n \): \( R_1 = \frac{1}{2} \log(1 + P_1) \)
  - User 2 sends in \( \bar{\lambda} n \): \( R_2 = \frac{1}{2} \log(1 + P_2) \)

- Note: low-complexity decoder – two users are decoded separately.
2-GMAC rates for FDMA

\[ Y = X_1 + X_2 + Z \]
\[ Z \sim iid \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Here is a FDMA:
  - Use **Fourier** transform to change \( n=\)time to \( n=\)frequency.
  - Partition block: \( n = \lambda n + (1 - \lambda) n \)
  - User 1 sends in \( \lambda n \):
    \[ R_1 = \frac{\lambda}{2} \log(1 + \frac{P_1}{\lambda}) \]
  - User 2 sends in \( \bar{\lambda} n \):
    \[ R_2 = \frac{\bar{\lambda}}{2} \log(1 + \frac{P_2}{\bar{\lambda}}) \]
2-GMAC rates for FDMA

\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

Here is a FDMA:

- Use **Fourier** transform to change \( n \)=time to \( n \)=frequency.
- Partition block: \( n = \lambda n + (1 - \lambda)n \)
- User 1 sends in \( \lambda n \):
  \[ R_1 = \frac{\lambda}{2} \log(1 + \frac{P_1}{\lambda}) \]
- User 2 sends in \( \bar{\lambda}n \):
  \[ R_2 = \frac{\bar{\lambda}}{2} \log(1 + \frac{P_2}{\bar{\lambda}}) \]
2-GMAC rates for FDMA

\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

• Here is a FDMA:
  ▶ Use Fourier transform to change \( n = \text{time} \) to \( n = \text{frequency} \).
  ▶ Partition block: \( n = \lambda n + (1 - \lambda)n \)
  ▶ User 1 sends in \( \lambda n \):
    \[ R_1 = \frac{\lambda}{2} \log(1 + \frac{P_1}{\lambda}) \]
  ▶ User 2 sends in \( \bar{\lambda}n \):
    \[ R_2 = \frac{\bar{\lambda}}{2} \log(1 + \frac{P_2}{\bar{\lambda}}) \]

\[ \lambda^* = \frac{P_1}{P_1 + P_2} \]
achieves optimal sumrate
2-GMAC rates for TIN

\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Treat-interference-as-noise (TIN):
  - Each user treats the other as noise (single-user decoders)
  - Random coding ensures noise is Gaussian.
  - Rates: \[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1+P_2}), \quad R_2 = \frac{1}{2} \log(1 + \frac{P_2}{1+P_1}) \]
2-GMAC rates for TIN

\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Treat-interference-as-noise (TIN):
  - Each user treats the other as noise (single-user decoders)
  - Random coding ensures noise is Gaussian.
  - Rates: \[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1+P_2}), R_2 = \frac{1}{2} \log(1 + \frac{P_2}{1+P_1}) \]

- TIN point can be inside/outside TDMA.
\[ Y = X_1 + X_2 + Z \]

\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]

\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Consider a corner point:

\[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1 + P_2}), \quad R_2 = \frac{1}{2} \log(1 + P_2). \]
\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Consider a corner point:

\[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1 + P_2}), \quad R_2 = \frac{1}{2} \log(1 + P_2). \]

- User 1 can be decoded by TIN. But then can subtract it out!
\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Consider a corner point:

\[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1 + P_2}), \quad R_2 = \frac{1}{2} \log(1 + P_2). \]

- User 1 can be decoded by TIN. But then can subtract it out!
\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Consider a corner point:
  \[ R_1 = \frac{1}{2} \log(1 + \frac{P_1}{1 + P_2}) \], \[ R_2 = \frac{1}{2} \log(1 + P_2) \].

- User 1 can be decoded by TIN. But then can subtract it out!

- So far: achieved three optimal points via SU-decoding. Any more?
Rate-splitting

\[ Y = X_1 + X_2 + Z \]
\[ Z \overset{iid}{\sim} \mathcal{N}(0, 1) \]
\[ \mathbb{E}[(X_1)^2] \leq P_1, \mathbb{E}[(X_2)^2] \leq P_2 \]

- Split user 1 into two virtual users 1A and 1B:

\[ R_1 = R_{1A} + R_{1B}, \quad P_1 = P_{1A} + P_{1B} \]

- A funny order of decoding:
  - Decode \( X_{1A} \) via TIN: \( R_{1A} = \frac{1}{2} \log(1 + \frac{P_{1A}}{1 + P_{1B} + P_2}) \)
  - Subtract \( X_{1A} \), decode \( X_2 \): \( R_2 = \frac{1}{2} \log(1 + \frac{P_2}{1 + P_{1B}}) \)
  - Subtract \( X_2 \), decode \( X_{1B} \): \( R_{1B} = \frac{1}{2} \log(1 + P_{1B}) \)

- Simple check:

\[ R_{1A} + R_{1B} + R_2 = \frac{1}{2} \log(1 + P_1 + P_2) \quad \text{sumrate optimal} \]

by varying \( P_{1A} + P_{1B} = P_1 \) can achieve any point!!
\[ Y(t) = X_1(t) + \cdots + X_K(t) + Z(t) \]

- Assume equal-power setting \( P_i = P \). Capacity region (sumrate):

\[
\sum_{i=1}^{K} R_i \leq \frac{1}{2} \log(1 + KP)
\]
$Y(t) = X_1(t) + \cdots + X_K(t) + Z(t)$

- **single-user** decoders achieve:
  - FDMA optimal at symmetric point: $R_i = \frac{1}{2K} \log(1 + KP)$
  - TIN+SIC achieves all vertices.
  - Rate-Splitting all points of optimal sumrate.

- Is that it? Let us see...
K-user GMAC: Reflections

- So total capacity:

\[ C_{\text{sum}} = \frac{1}{2} \log_2(1 + K P) \text{ bit/rdof} \]

Growing to \( \infty \) as \( K \to \infty \).
• So total capacity:

\[ C_{\text{sum}} = \frac{1}{2} \log_2(1 + KP) \text{ bit/rdof} \]

growing to \( \infty \) as \( K \to \infty \).

• But at the same time, per-user rate:

\[ C_{\text{sym}} = \frac{1}{2K} \log_2(1 + KP) \to 0. \]

• The crucial performance metric: HRH energy-per-bit

\[ \frac{E_b}{N_0} \triangleq \frac{\text{total energy spent}}{2 \times \text{total \# bits}} = \frac{nKP}{2nC_{\text{sum}}} \]
• So total capacity:

\[ C_{\text{sum}} = \frac{1}{2} \log_2(1 + KP) \ \text{bit/rdof} \]

growing to \( \infty \) as \( K \to \infty \).

• But at the same time, per-user rate:

\[ C_{\text{sym}} = \frac{1}{2K} \log_2(1 + KP) \to 0. \]

• The crucial performance metric: HRH energy-per-bit

\[ \frac{E_b}{N_0} \triangleq \frac{\text{total energy spent}}{2 \times \text{total \#\ bits}} = \frac{nKP}{2nC_{\text{sum}}} \]

• As \( K \to \infty \):

\[ \frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)} \to \infty \quad !!! \]

• Capacity \( \nearrow \), but each user works harder and moves fewer bits/sec!
So total capacity:

\[ C_{sum} = \frac{1}{2} \log_2(1 + KP) \text{ bit/rdof} \]

growing to \( \infty \) as \( K \to \infty \).

But at the same time, per-user rate:

\[ C_{sym} = \frac{1}{2K} \log_2(1 + KP) \to 0. \]

The crucial performance metric: HRH energy-per-bit

\[
\frac{E_b}{N_0} \triangleq \frac{\text{total energy spent}}{2 \times \text{total \# bits}} = \frac{nKP}{2nC_{sum}}
\]

As \( K \to \infty \):

\[
\frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)} \to \infty \quad !!!
\]

Capacity \( \nearrow \), but each user works harder and moves fewer bits/sec!

Correct scaling: \( P_{tot} = KP \) should be fixed!
- Studying this tradeoff is the favorite pastime of ComSoc
- Sp.eff. $\rho \triangleq \frac{\text{total \ # \ of \ data \ bits}}{\text{total real d.o.f.}}$
- We have:
  \[ \rho = \frac{1}{2} \log(1 + KP), \quad \frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)} \]
- Regardless of $K$: (and any sumrate-optimal arch)
  \[ \frac{E_b}{N_0} = \frac{2^{2\rho} - 1}{2\rho} \geq -1.59 \text{ dB} \]
• Studying this tradeoff is the favorite pastime of ComSoc.

• Sp.eff. \( \rho \) ≜ total # of data bits / total real d.o.f.

• We have:

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

  \[ E_b N_0 = K^2 P \log(1 + KP) \]

  \[ \rho = \frac{1}{2} \log(1 + KP) \geq -0.59 \text{ dB} \]

  \[ E_b N_0 = 2^{\frac{1}{2} \rho - \frac{1}{2} \rho} \geq -1.59 \text{ dB} \]

• Compare to TIN:

  \[ \rho = K^2 \log_2(1 + \frac{P_1 + (K-1)P}{K}) \]

  \[ P_\text{tot} \rightarrow \infty \rightarrow \frac{1}{2} \ln 2 \]

  \[ E_b N_0 = (1 + P_\text{tot}) \ln 2 \]

• IMPORTANT:

  \[ \rho \leq \frac{1}{2} \ln 2 = 0.72 \text{ bit/rdof} \]

  \[ \rho \leq 0.72 \text{ bit/rdof} \]

• IMPORTANT:

  Essentially optimal for low sp.eff.
Spectral efficiency vs. $\frac{E_b}{N_0}$

- Studying this tradeoff is the favorite pastime of ComSoc
- Sp.eff. $\rho \equiv \frac{\text{total # of data bits}}{\text{total real d.o.f.}}$
- We have:

$$\rho = \frac{1}{2} \log(1 + KP), \quad \frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)}$$

- regardless of $K$ : (and any sumrate-optimal arch)

$$\frac{E_b}{N_0} = \frac{2^{2\rho} - 1}{2\rho} \geq -1.59 \text{ dB}$$

- Compare to TIN: $\rho = \frac{K}{2} \log_2\left(1 + \frac{P}{1+(K-1)P}\right)$ $\xrightarrow{K \to \infty} \frac{1}{2 \ln 2} \frac{P_{tot}}{1+P_{tot}}$
• Studying this tradeoff is the favorite pastime of ComSoc

• Sp.eff. $\rho \triangleq \frac{\text{total # of data bits}}{\text{total real d.o.f.}}$

• We have:

$$\rho = \frac{1}{2} \log(1 + KP), \quad \frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)}$$

• regardless of $K$: (and any sumrate-optimal arch)

$$\frac{E_b}{N_0} = \frac{2^{2\rho} - 1}{2\rho} \geq -1.59 \text{ dB}$$

• Compare to TIN: $\rho = \frac{K}{2} \log_2(1 + \frac{P}{1+(K-1)P}) \xrightarrow{K \to \infty} \frac{1}{2 \ln 2} \frac{P_{tot}}{1+P_{tot}}$

$$\rho = \frac{1}{2 \ln 2} \frac{P_{tot}}{1 + P_{tot}}, \quad \frac{E_b}{N_0} = (1 + P_{tot}) \ln 2$$
• Studying this tradeoff is the favorite pastime of ComSoc.
• Speff. $\rho \equiv \frac{\text{total # of data bits}}{\text{total real d.o.f.}}$
• We have:
  $$\rho = \frac{1}{2} \log(1 + KP)$$
  $$\frac{E_b}{N_0} = KP \log(1 + KP)$$
• regardless of $K$ (and any sumrate-optimal arch)
• $\rho \geq -0.59 \text{dB}$
• Compare to TIN:
  $$\rho = \frac{K}{2} \log_2(1 + P_{\text{tot}})$$
  $$\frac{E_b}{N_0} = (1 + P_{\text{tot}}) \ln 2$$

Spectral efficiency vs. $\frac{E_b}{N_0}$
- Studying this tradeoff is the favorite pastime of ComSoc
- Sp.eff. $\rho \triangleq \frac{\text{total # of data bits}}{\text{total real d.o.f.}}$
- We have:

$$\rho = \frac{1}{2} \log(1 + KP), \quad \frac{E_b}{N_0} = \frac{KP}{\log(1 + KP)}$$

- regardless of $K$ (and any sumrate-optimal arch)

$$\frac{E_b}{N_0} = \frac{2^{2\rho} - 1}{2\rho} \geq -1.59 \text{ dB}$$

- Compare to TIN: $\rho = \frac{K}{2} \log_2(1 + \frac{P}{1+(K-1)P}) \xrightarrow{K \to \infty} \frac{1}{2 \ln 2} \frac{P_{tot}}{1+P_{tot}}$

$$\rho = \frac{1}{2 \ln 2} \frac{P_{tot}}{1+P_{tot}}, \quad \frac{E_b}{N_0} = (1 + P_{tot}) \ln 2$$

- IMPORTANT: $\rho \leq \frac{1}{2 \ln 2} = 0.72 \text{ bit/rdof}$
- IMPORTANT: Essentially optimal for low sp.eff.
• Given that TIN is not bad for low sp.eff., let us try to achieve it.
• **Problem:** Per-user rate $= \frac{\rho}{K}$ and is **very small** for large $K$. 
• Given that TIN is not bad for low sp.eff., let us try to achieve it.
• **Problem:** Per-user rate $= \frac{\rho}{K}$ and is very small for large $K$. Aside:
  ▶ For IT Soc: Channel with $C = 0.5$ and channel with $C = 0.001$ are not fundamentally different.
  ▶ For ComSoc: First channel is OK (turbo/LDPC/polar), second is a nightmare.
  ▶ **Why?** First, SNR needs to be brought up to a reasonable level.
  ▶ This is the idea of modulation.
• Given that TIN is not bad for low sp.eff., let us try to achieve it.
• **Problem:** Per-user rate $= \frac{\rho}{K}$ and is very small for large $K$.
  Aside:
  ▶ For IT Soc: Channel with $C = 0.5$ and channel with $C = 0.001$ are not fundamentally different.
  ▶ For ComSoc: First channel is OK (turbo/LDPC/polar), second is a nightmare.
  ▶ Why? First, SNR needs to be brought **up** to a reasonable level.
  ▶ This is the idea of modulation.
  ▶ Another issue: how do you do TIN practically? A code with $\pm 1$ entries will create a very non-Gaussian interference!
• Given that TIN is not bad for low sp.eff., let us try to achieve it.

• **Problem:** Per-user rate $= \frac{\rho}{K}$ and is **very small** for large $K$.

• **Solution:** each user modulates some $N$-signature $s_i \in \mathbb{R}^N$ 

<table>
<thead>
<tr>
<th>N</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>pink</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

• Think of $N$-blocks as new super-symbols. Effective channel:

$$Y^N = s_1 B_1 + s_2 B_2 + \cdots s_K B_k + Z^N, \quad \|s_i\| = 1$$

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{tot}}{\beta}$
- new rate: $\frac{\rho N}{K} = \frac{\rho}{\beta}$ in bits / one $B$-symbol.
- with proper choice should have $\frac{\rho}{\beta} \sim 1$ as ComSoc likes.
- $N$-blocks are new super-symbols. Effective channel:
  \[ Y^N = s_1 B_1 + s_2 B_2 + \cdots s_K B_k + Z^N, \quad \|s_i\| = 1 \]

  - Set $\beta = \frac{K}{N}$
  - new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{tot}}{\beta}$.

- **Side observation:**
  - If $s_i$'s are chosen orthogonally and $K = N$, this is FDMA (hence optimal).
  - But incurs FBL loss – important when $K \sim n$. Ignore for now.
  - So why not do so? Many reasons:
    - $K$ may vary, but $N$ should be constant.
    - Requires central distribution of signatures among ACTIVE users.
    - Asynchrony kills orthogonality
  - Early Qualcomm: random-like $s_i$'s resolve all issues, and are good enough for TIN!
• $N$-blocks are new super-symbols. Effective channel:

$$Y^N = s_1 B_1 + s_2 B_2 + \cdots + s_K B_k + Z^N, \quad \|s_i\| = 1$$

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{\text{tot}}}{\beta}$.

• Idea 1: Decode via matched-filter + SU decoders:

$$\hat{B}_i = \langle s_i, Y^N \rangle = B_i + \tilde{Z}_i, \quad \text{Var}[\tilde{Z}_i] = 1 + NP \sum_{j \neq i} |\langle s_i, s_j \rangle|^2$$

• Idea 2: Select $s_i$ randomly. (attractive sys. arch.)
• When $s_i$'s are random and $N$ large:

$$|\langle s_i, s_j \rangle| \approx \frac{1}{\sqrt{N}} \quad \text{w.h.p.}$$

• So SU-decoder sees effective SNR:

$$\text{SNR} = \frac{NP}{1+(K-1)P} = \frac{P_{\text{tot}}}{1+P_{\text{tot}} \beta}$$
• $N$-blocks are new super-symbols. Effective channel:

$$Y^N = s_1 B_1 + s_2 B_2 + \cdots s_K B_k + Z^N, \quad \|s_i\| = 1$$

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{\text{tot}}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder

• End result:

$$\rho_{\text{CDMA}} = \frac{\beta}{2} \log_2 \left( 1 + \frac{P_{\text{tot}}}{1 + P_{\text{tot}} \frac{1}{\beta}} \right) \quad \frac{E_b}{N_0} = \frac{P_{\text{tot}}}{2\rho_{\text{CDMA}}}$$

- As $\beta \to \infty$ we approach TIN.
- So classical CDMA folks (Viterbi...) were only trying to achieve TIN.
• $N$-b

• End


- As $\beta \to \infty$ we approach TIN.
- So classical CDMA folks (Viterbi...) were only trying to achieve TIN.
CDMA: going beyond TIN

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{\text{tot}}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder

$$\rho_{\text{CDMA}} = \frac{\beta}{2} \log_2 \left( 1 + \frac{P_{\text{tot}}}{1 + P_{\text{tot}} \beta} \right) \quad \frac{E_b}{N_0} = \frac{P_{\text{tot}}}{2\rho_{\text{CDMA}}}$$

- So far we considered matched-filter arch.:

$$\hat{B}_1 = \langle s_1, Y^N \rangle$$

$$\vdots$$

$$\hat{B}_K = \langle s_K, Y^N \rangle$$

- Can we do better?
CDMA: going beyond TIN

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{tot}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder
  
  $$\rho_{CDMA} = \frac{\beta}{2} \log_2\left(1 + \frac{P_{tot}}{1 + P_{tot} \beta}\right) \quad \frac{E_b}{N_0} = \frac{P_{tot}}{2\rho_{CDMA}}$$

- So far we considered matched-filter arch.: 
  
  $$\hat{B}_1 = \langle s_1, Y^N \rangle$$
  $$\ldots$$
  $$\hat{B}_K = \langle s_K, Y^N \rangle$$

- Can we do better? Yes! via multi-user detection (MUD).
- In one of two ways:
  - Signal-processing: Estimate $\hat{B}_K$ via MMSE or decorrelator. Note: does not leverage knowledge of distribution of $B_i$
  - Coding: Use joint-decoding of $\hat{B}_K$ also leveraging knowledge that (e.g.) $B_i = \pm 1$
CDMA+MUD vs OFDM

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{tot}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder

$$\rho_{CDMA} = \frac{\beta}{2} \log_2(1 + \frac{P_{tot}}{1 + P_{tot} \beta}) \quad \frac{E_b}{N_0} = \frac{P_{tot}}{2\rho_{CDMA}}$$

- multi-user detectors (MUD) improve performance of random-CDMA.
- E.g. MMSE detector yields (Tse-Hanly/Verdú-Shamai formula)

$$\rho_{MMSE} = \frac{\beta}{2} \log_2(1 + P_1 - \frac{1}{4} \mathcal{F}), \quad P_1 = \frac{P_{tot}}{\beta}$$

where $\mathcal{F} = (\sqrt{P_1(1 + \sqrt{\beta})^2 + 1} - \sqrt{P_1(1 - \sqrt{\beta})^2 + 1})^2$
CDMA+MUD vs OFDM

- Allows to beat TIN's $\rho \leq 0.72$ bit/rdof bottleneck.
- Still, industry converged to OFDM: spectrum is too precious.
- IoT: centralized orthogonalization impossible! Comeback of MUD?

\[ \text{Optimal TIN CDMA-MMSE: best } \beta \]

\[ \rho_{\text{MMSE}} = \beta^2 \log_2(1 + \frac{P_1}{P_{\text{tot}}} - \frac{1}{4} F) \]

\[ P_1 = P_{\text{tot}} \beta \]

\[ F = \left( \sqrt{P_1} \sqrt{1 + \sqrt{\beta}} + 1 \right) - \sqrt{P_1} \sqrt{1 - \sqrt{\beta}} + 1 \]
CDMA+MUD vs OFDM

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{\text{tot}}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder

$$\rho_{\text{CDMA}} = \frac{\beta}{2} \log_2 \left(1 + \frac{P_{\text{tot}}}{1 + P_{\text{tot}} \beta} \right) \quad \frac{E_b}{N_0} = \frac{P_{\text{tot}}}{2\rho_{\text{CDMA}}}$$

- multi-user detectors (MUD) improve performance of random-CDMA.
- E.g. MMSE detector yields (Tse-Hanly/Verdú-Shamai formula)

$$\rho_{\text{MMSE}} = \frac{\beta}{2} \log_2 \left(1 + P_1 - \frac{1}{4} \mathcal{F} \right), \quad P_1 = \frac{P_{\text{tot}}}{\beta}$$

where $\mathcal{F} = \left(\sqrt{P_1(1+\sqrt{\beta})^2} + 1 - \sqrt{P_1(1-\sqrt{\beta})^2} + 1\right)^2$

- Allows to beat TIN’s $\rho \leq 0.72$ bit/rdof bottleneck.
- Still, industry converged to **OFDM**: spectrum is too precious.
CDMA+MUD vs OFDM

- Set $\beta = \frac{K}{N}$
- new power-constraint: $\mathbb{E}[B_i^2] \leq NP = \frac{P_{tot}}{\beta}$.
- random (non-orthogonal) signatures
- matched-filter + SU-decoder

$$\rho_{CDMA} = \frac{\beta}{2} \log_2 \left(1 + \frac{P_{tot}}{1 + \frac{P_{tot}}{\beta}} \right) \quad \frac{E_b}{N_0} = \frac{P_{tot}}{2\rho_{CDMA}}$$

- multi-user detectors (MUD) improve performance of random-CDMA.
- E.g. MMSE detector yields (Tse-Hanly/Verdú-Shamai formula)

$$\rho_{MMSE} = \frac{\beta}{2} \log_2 \left(1 + P_1 - \frac{1}{4} \mathcal{F} \right), \quad P_1 = \frac{P_{tot}}{\beta}$$

where $\mathcal{F} = \left(\sqrt{P_1(1 + \sqrt{\beta})^2 + 1} - \sqrt{P_1(1 - \sqrt{\beta})^2 + 1}\right)^2$

- Allows to beat TIN’s $\rho \leq 0.72$ bit/rdof bottleneck.
- Still, industry converged to OFDM: spectrum is too precious.
- IoT: centralized orthogonalization impossible! Comeback of MUD?
New problems: many users with short packets
The classical model: K-user multiple-access channel

\[ Y(t) = X_1(t) + \cdots + X_K(t) + Z(t) \]
The classical model: K-user multiple-access channel

\[ Y(t) = X_1(t) + \cdots + X_K(t) + Z(t) \]

- Before: Fix \( K \), let \( n \to \infty \). Few users. Large payloads.
- **Now**: Huge \( K \). Small payload.
- **Random-access**: User activity – random, uncoordinated
On number of sensors (user density)

• **Key metric:** $\mu$ in users/rdof

\[
\mu = \frac{\text{# of active users per frame}}{\text{size of frame}}
\]

• $K_{tot}$ sensors sending with period $T_{per}$ (sec) in band $B$ (Hz)

\[
\mu = \frac{K_{tot}}{2BT_{per}}
\]

• **Futuristic example:**
  - City of $10^6$.
  - Each house has $10^2$ devices.
  - Each dev sends every 10 min, $T_{per} = 600$ s.
  - sub-GHz bandwidth is scarce: ISM $B = 20$ MHz.
  - $\mu \approx 4 \cdot 10^{-3}$.
On number of sensors (user density)

- **Key metric:** $\mu$ in users/rdof
  
  $$\mu = \frac{\text{# of active users per frame}}{\text{size of frame}}$$

- $K_{tot}$ sensors sending with period $T_{per}$ (sec) in band $B$ (Hz)
  
  $$\mu = \frac{K_{tot}}{2BT_{per}}$$

- **Futuristic example:**
  - City of $10^6$.
  - Each house has $10^2$ devices.
  - Each dev sends every 10 min, $T_{per} = 600$ s.
  - sub-GHz bandwidth is scarce: ISM $B = 20$ MHz.
  - $\mu \approx 4 \cdot 10^{-3}$.

- **Another point of view:**
  - Traditional comm: focus on sp.eff. $\rho$ vs $\frac{E_b}{N_0}$. Why?
  - $\frac{\rho B}{K}$ = per-user speed?
  - or is it $\frac{\rho B}{\text{speed}}$ = number of happy users?
New twists compared to classic MAC

Problem 1 large $K \to \infty$, fixed payload $\log_2 M$

Relevant asymptotics: $K, n \to \infty$ with $\frac{K}{n} = \mu$.

Problem 2 “user-centric” probability of error

$$P_e \triangleq \frac{1}{K} \sum_j \mathbb{P}[\hat{X}_j \neq X_j]$$

Problem 3 “random-access”

indistinguishable users (same-codebook), non-asymptotics.
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M \approx 10^2$ bits.
- Figures of merit: energy-per-bit and user density

\[
\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2 \log_2 M} \quad \mu \triangleq \frac{K}{n}
\]
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M \approx 10^2$ bits.
- Figures of merit: energy-per-bit and user density

$$\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2 \log_2 M}$$

$$\mu \triangleq \frac{K}{n}$$

Problem 1: “massive” number of users

- Number of users $K = \mu n$ scales linearly with blocklength!
- Q: Ok, but what $\mu$ should we look at?
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M \approx 10^2$ bits.
- Figures of merit: energy-per-bit and user density

$$\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2\log_2 M}$$

$\mu \triangleq \frac{K}{n}$

Problem 1: “massive” number of users

- Number of users $K = \mu n$ scales linearly with blocklength!
- Q: Why scale linearly? A: # of devices waking up $\propto$ time.
- Q: Ok, but what $\mu$ should we look at?
  A: $\mu \sim 10^{-3}$. Here is why:
    - City of $10^6$.
    - Each house has $10^2$ devices.
    - Each dev sends 1-10 times/hour.
    - sub-GHz bandwidth is scarce, unlikely to ever get $> 20$ MHz.
    - $\Rightarrow \frac{K}{n} \approx 10^{-3} \ldots 10^{-2}$. This relation is unlikely to change soon.
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M$ bits.
- Figures of merit: energy-per-bit and user density

$$E_b \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2 \log_2 M}$$
$$\mu \triangleq \frac{K}{n}$$

Problem 1: “massive” number of users

- Number of users $K = \mu n$ scales linearly with blocklength!
- [Chen-Chen-Guo’17]: Fix per-user power to $P$ (i.e. codeword $\|c\|^2 \leq nP$), then

$$\log M^*_{user}(K = \mu n, n, P) \approx \frac{1}{2\mu} \log(1 + \mu nP)$$

- Note: this corresponds to $\frac{E_b}{N_0} \to \infty$.
- Our work: What about finite $\frac{E_b}{N_0}$?
New twists compared to classic MAC

Problem 1 large $K \to \infty$, fixed payload $\log_2 M$

Relevant asymptotics: $K, n \to \infty$ with $\frac{K}{n} = \mu$.

Problem 2 “user-centric” probability of error

$$P_e \triangleq \frac{1}{K} \sum_j \mathbb{P}[\hat{X}_j \neq X_j]$$

Problem 3 “random-access”

indistinguishable users (same-codebook), non-asymptotics.
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M$ bits.
- Figures of merit: energy-per-bit and user density

\[
\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2 \log_2 M}
\]

\[
\mu \triangleq \frac{K}{n}
\]

- Regime: $K = \mu n$, $n \to \infty$.

Problem 2: “user-centric” prob. of error

- For finite $\frac{E_b}{N_0}$ we have ( Why? See next...)

\[
P[W_1 = \hat{W}_1, \ldots W_K = \hat{W}_K] \to 0 \quad \text{as } n \to \infty
\]

- ⇒ NEED to switch to per-user $P_e$, PUPE :

\[
P_e = \frac{1}{K} \sum_{i=1}^{K} P[W_i \neq \hat{W}_i]
\]
Theorem

Suppose $K$ users send one bit each with finite energy $\mathcal{E}$ over the GMAC (with arbitrary $n$): $Y^n = \sum_{i=1}^{K} X_i + Z^n$. Then we have

$$\mathbb{P}[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{\mathcal{E} \log e}{2} + \log 2 \frac{\mathcal{E} \log e}{\log K}.$$ 

And, thus, classical probability of error $\rightarrow 1$ as $K \rightarrow \infty$. 

$Eb/N_0 \rightarrow \infty$ for classical probability of error
\( Eb/N_0 \rightarrow \infty \) for classical probability of error

**Theorem**

Suppose \( K \) users send one bit each with finite energy \( \mathcal{E} \) over the GMAC (with arbitrary \( n \)): \( Y^n = \sum_{i=1}^{K} X_i + Z^n \). Then we have

\[
\mathbb{P}[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{\mathcal{E} \log e}{2} + \log 2 \frac{1}{\log K}.
\]

And, thus, classical probability of error \( \rightarrow 1 \) as \( K \rightarrow \infty \).

**Proof:**

- **WLOG** can assume: \( Y = \sum c_i W_i + Z \), where \( c_i \in \mathbb{R}^n \) and \( W_i \sim \text{Ber}(1/2) \).
- **Genie**: Reveal vector of \( W_i \)'s to within Hamming-distance 1.
- **New problem**: See \( Y = c_U + Z \), \( U \sim [K] \). **Goal**: find \( U \).
$E b / N_0 \rightarrow \infty$ for classical probability of error

**Theorem**

Suppose $K$ users send one bit each with finite energy $E$ over the GMAC (with arbitrary $n$): $Y^n = \sum_{i=1}^{K} X_i + Z^n$. Then we have

$$P[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{E \log e}{2} + \frac{\log 2}{\log K}.$$ 

And, thus, classical probability of error $\rightarrow 1$ as $K \rightarrow \infty$.

**Proof:**

- **WLOG** can assume: $Y = \sum c_i W_i + Z$, where $c_i \in \mathbb{R}^n$ and $W_i \sim \text{Ber}(1/2)$.
- **Genie**: Reveal vector of $W_i$’s to within Hamming-distance 1.
- **New problem**: See $Y = cU + Z$, $U \sim [K]$. **Goal**: find $U$.
- **Fano + Capacity calculation**:

$$P[U = \hat{U}] \log K - \log 2 \leq I(c_U; Y)$$
Theorem

Suppose $K$ users send one bit each with finite energy $\mathcal{E}$ over the GMAC (with arbitrary $n$): $Y^n = \sum_{i=1}^{K} X_i + Z^n$. Then we have

$$\mathbb{P}[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{\mathcal{E} \log e}{2} + \log 2 \frac{\log K}{\log K}.$$ 

And, thus, classical probability of error $\to 1$ as $K \to \infty$.

Proof:

- **WLOG** can assume: $Y = \sum c_i W_i + Z$, where $c_i \in \mathbb{R}^n$ and $W_i \sim \text{Ber}(1/2)$.
- **Genie**: Reveal vector of $W_i$’s to within Hamming-distance 1.
- Fano + Capacity calculation:

$$\mathbb{P}[U = \hat{U}] \log K - \log 2 \leq I(cU; Y) \leq \frac{n}{2} \log \left(1 + \frac{\mathcal{E}}{n}\right) \leq \frac{\log e}{2} \mathcal{E}$$
Theorem (AWGN)

Suppose $K$ users send one bit each with finite energy $\mathcal{E}$ over the GMAC (with arbitrary $n$): $Y^n = \sum_{i=1}^{K} X_i + Z^n$. Then we have

$$\Pr[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{\mathcal{E} \log e}{2} + \log 2 \frac{\log K}{\log K}.$$ 

Same proof:

Theorem (BSC)

Let $G$ be a $K \times n$ generating matrix with $\leq \mathcal{E}$ ones per row. Then over $BSC(\delta)$ and all $n$:

$$1 - \Pr[\text{block error}] \leq \frac{d(\delta || \bar{\delta}) \mathcal{E} + \log 2}{\log K}$$
$Eb/N_0 \rightarrow \infty$ for classical probability of error

**Theorem (AWGN)**

Suppose $K$ users send one bit each with finite energy $\mathcal{E}$ over the GMAC (with arbitrary $n$): $Y^n = \sum_{i=1}^{K} X_i + Z^n$. Then we have

$$
\mathbb{P}[X_1 = \hat{X}_1, \ldots, X_K = \hat{X}_K] \leq \frac{\mathcal{E} \log e}{2} + \log 2 \frac{\log K}{\log K}.
$$

**Same proof:**

**Theorem (BSC)**

Let $G$ be a $K \times n$ generating matrix with $\leq \mathcal{E}$ ones per row. Then over $BSC(\delta)$ and all $n$:

$$
1 - \mathbb{P}[\text{block error}] \leq \frac{d(\delta||\bar{\delta})\mathcal{E} + \log 2}{\log K}.
$$

**Puzzle:** Genie + Fano method fails for BEC! (Proof by induction works.)
$K$-user GMAC under PUPE: surprise

- Per-user probability of error as

$$P_e = \frac{1}{K} \sum_{i=1}^{K} \mathbb{P}[W_i \neq \hat{W}_i].$$

- Let’s forget about $K = \mu n$ and consider ...

- **Classical regime:** $K$-fixed, power $P$ fixed, $n \to \infty$. Symmetric capacity

$$C_{sym}(K) = \frac{1}{2K} \log(1 + KP).$$

- But no strong converse (!)

$$C_{sym,\epsilon}(K) > C_{sym}(K - 1) \quad \forall \epsilon \geq \frac{1 + \log e K}{K}$$

- **Lesson:** When PUPE above $\frac{\log K}{K}$, far from usual GMAC+JPE.
K-user GMAC under PUPE: no strong converse

- Let $C_{sym,\epsilon}(K)$ be the max achievable symmetric rate ($K$-fixed, \( n \to \infty \)) under PUPE

$$
\frac{1}{K} \sum_{i=1}^{K} \mathbb{P}[W_i \neq \hat{W}_i] \leq \epsilon.
$$

Note that sequence:

$$\frac{1}{2} K \log(1 + KP)$$

is monotonically decreasing.

First part: by union bound PUPE $\leq \epsilon$ implies JPE $\leq K \epsilon$ + strong-converse for GMAC.

Second part: Choose codebooks for symmetric-rate point of $(K-1)$-GMAC

- Each user sends 0 w.p. $\epsilon$. Then w.p. $1 - (1 - \epsilon)^K$ only $(K-1)$ are active.
Let \( C_{\text{sym}, \epsilon}(K) \) be the max achievable symmetric rate (\( K \)-fixed, \( n \to \infty \)) under PUPE

\[
\frac{1}{K} \sum_{i=1}^{K} \mathbb{P}[W_i \neq \hat{W}_i] \leq \epsilon.
\]

**Theorem (P.-Telatar’16)**

We have:

\[
C_{\text{sym}, \epsilon}(K, \epsilon) = \begin{cases} 
\frac{1}{2K} \log(1 + KP), & \epsilon < 1/K \\
\geq \frac{1}{2(K-1)} \log(1 + (K - 1)P), & \epsilon \geq \frac{1 + \log_e K}{K}
\end{cases}
\]

- Note that sequence: \( \frac{1}{2K} \log(1 + KP) \) is monotonically decreasing.
- First part: by union bound PUPE \( \leq \epsilon \) implies JPE \( \leq K\epsilon \) + strong-converse for GMAC.
$K$-user GMAC under PUPE: no strong converse

- Let $C_{sym, \epsilon}(K)$ be the max achievable symmetric rate ($K$-fixed, $n \to \infty$) under PUPE

$$
\frac{1}{K} \sum_{i=1}^{K} \mathbb{P}[W_i \neq \hat{W}_i] \leq \epsilon.
$$

**Theorem (P.-Telatar’16)**

We have: $C_{sym, \epsilon}(K, \epsilon) = \begin{cases} 
\frac{1}{2K} \log(1 + KP), & \epsilon < 1/K \\
\geq \frac{1}{2(K-1)} \log(1 + (K - 1)P), & \epsilon \geq \frac{1+\log_e K}{K}
\end{cases}$

- Note that sequence: $\frac{1}{2K} \log(1 + KP)$ is monotonically decreasing.
- First part: by union bound PUPE $\leq \epsilon$ implies JPE $\leq K\epsilon$ + strong-converse for GMAC.
- Second part: Choose codebooks for symmetric-rate point of $(K - 1)$-GMAC
- Each user sends 0 w.p. $\epsilon$. Then w.p. $1 - (1 - \epsilon)^K$ only $(K - 1)$ are active.
New twists compared to classic MAC

Problem 1 large $K \to \infty$, fixed payload $\log_2 M$

Relevant asymptotics: $K, n \to \infty$ with $\frac{K}{n} = \mu$.

Problem 2 “user-centric” probability of error

$$P_e \triangleq \frac{1}{K} \sum_j \mathbb{P}[\hat{X}_j \neq X_j]$$

Problem 3 “random-access”

indistinguishable users (same-codebook), non-asymptotics.
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M$ bits.
- Figures of merit: energy-per-bit and user density

\[
\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2 \log_2 M}
\]
\[
\mu \triangleq \frac{K}{n}
\]

- Regime: $K = \mu n$, $n \to \infty$.
- PUPE definition: $P_e \triangleq \frac{1}{K} \sum_{j=1}^{K} \mathbb{P}[X_j \neq \hat{X}_j]$.

Next: new results
Recap: MAC setting and performance metrics

- Perfectly synchronized $K$-user Gaussian MAC with blocklength $n$
- Each user transmits $\log_2 M$ bits.
- Figures of merit: energy-per-bit and user density

\[
\frac{E_b}{N_0} \triangleq \frac{\mathbb{E}[\|X^n\|^2]}{2\log_2 M}
\]

\[
\mu \triangleq \frac{K}{n}
\]

- Regime: $K = \mu n$, $n \to \infty$.
- PUPE definition: $P_e \triangleq \frac{1}{K} \sum_{j=1}^{K} \mathbb{P}[X_j \neq \hat{X}_j]$.

Next: new results

- Converse bound (via reduction to known problems)
- Achievability bound (via Gaussian process theory)
Communication with \((\mu, M, \epsilon)\) is asymptotically \((n \to \infty)\) feasible only if both of these hold:

\[
(1 - \epsilon)\mu \log_2 M \leq \frac{1}{2} \log_2 (1 + \mu P_{\text{tot}}) + \mu h(\epsilon)
\]

\[
\frac{1}{M} \geq Q \left( \sqrt{\frac{P_{\text{tot}}}{\mu}} + Q^{-1}(1 - \epsilon) \right).
\]

where \(P_{\text{tot}} = 2\mu \log_2 M \cdot \frac{E_b}{N_0}\) is the total received power.

- First bound: A working code recovers \(W \in [M]^K\) with Hamming distortion \(\leq \epsilon\). Comparing sum-capacity with rate-distortion function we get the bound.
Theorem

Communication with \((\mu, M, \epsilon)\) is asymptotically \((n \to \infty)\) feasible only if both of these hold:

\[
(1 - \epsilon)\mu \log_2 M \leq \frac{1}{2} \log_2 (1 + \mu P_{\text{tot}}) + \mu h(\epsilon)
\]

\[
\frac{1}{M} \geq Q \left( \sqrt{\frac{P_{\text{tot}}}{\mu}} + Q^{-1}(1 - \epsilon) \right).
\]

where \(P_{\text{tot}} = 2\mu \log_2 M \cdot \frac{E_b}{N_0}\) is the total received power.

- Second bound: To get small \(\frac{E_b}{N_0}\) one necessarily needs to code over large payloads (i.e. \(\log_2 M \gg 1\)) – this is [PPV’11].

- Namely, we use the genie argument. At least one of \(K\) users should have \(P_e \leq \epsilon\).

- Even if that user communicated alone over a \(n = \infty\) AWGN channel, he’d need large total energy-per-bit if \(M\) is small.
Theorem

Communication with \((\mu, M, \epsilon)\) is asymptotically \((n \to \infty)\) feasible only if both of these hold:

\[
(1 - \epsilon) \mu \log_2 M \leq \frac{1}{2} \log_2 (1 + \mu P_{\text{tot}}) + \mu h(\epsilon)
\]

\[
M \geq Q(\sqrt{P_{\text{tot}} \mu} + Q - 1 (1 - \epsilon))
\]

where \(P_{\text{tot}} = 2 \mu \log_2 M \cdot E_b N_0\) is the total received power.

- Second bound: To get small \(E_b N_0\) one necessarily needs to code over large payloads (i.e. \(\log_2 M \gg 1\)) — this is \([P.-Poor-Verdú’11]\).
- Namely, we use the genie argument. At least one of \(K\) users should have \(P_e \leq \epsilon\).
- Even if that user communicated alone over an \(n = \infty\) AWGN channel, he’d need large total energy-per-bit if \(M\) is small.

\[\text{Yury Polyanskiy} \quad \text{MAC tutorial} \quad 90\]
Theorem (Thrampoulidis-Zadik-P.’18)

For each $\beta > 0$ there exists codes with $\frac{E_b}{N_0} = \frac{\beta^2}{2\log_2 M}$ and PUPE $\epsilon$ provided that

$$\theta \mu \log M + \mu h(\theta) < \frac{1}{2} \log(1 + \beta^2 \theta \mu) + \frac{\log e}{2} \left( \frac{\psi(\beta, \theta, \mu)}{1 + \beta^2 \theta \mu} - 1 \right)$$

for all $\theta \in [\epsilon, 1]$ where

$$\psi(\beta, \theta, \mu) = \sqrt{1 + \beta^2 \theta \mu} - \frac{\beta \mu}{\sqrt{2\pi}} e^{-\frac{1}{2}(Q^{-1}(\theta))^2}$$
Theorem (Thrampoulidis-Zadik-P.’18)

For each $\beta > 0$ there exists codes with $\frac{E_b}{N_0} = \frac{\beta^2}{2 \log_2 M}$ and PUPE $\epsilon$ provided that

$$\theta \mu \log M + \mu h(\theta) < \frac{1}{2} \log(1 + \beta^2 \theta \mu) + \frac{\log e}{2} \left( \frac{\psi(\beta, \theta, \mu)}{1 + \beta^2 \theta \mu} - 1 \right)$$

for all $\theta \in [\epsilon, 1]$ where

$$\psi(\beta, \theta, \mu) = \sqrt{1 + \beta^2 \theta \mu} - \frac{\beta \mu}{\sqrt{2 \pi}} e^{-\frac{1}{2}(Q^{-1}(\theta))^2}$$

Proof outline:

- Use random gaussian codebooks
- Use maximum likelihood decoder (not optimal!): $\min \|Y - \sum_i c_i\|_2$
- Use information-density thresholding trick
- Use Gaussian process theory (Gordon’s lemma) to evaluate the bound
• Generate codewords $c_m^{(j)} \sim \mathcal{N}(0, P I_n)$, $j \in [K]$, $m \in [M]$, where $P = \frac{\beta^2}{n}$
• Use ML decoder (suboptimal!):

$$
\hat{W} = \arg\min_{w_1, \ldots, w_K} \| Y - (c_{w_1}^{(1)} + \cdots + c_{w_K}^{(K)}) \|_2^2.
$$

• Define

$$
F(S_0) = \{ \exists (m_j)_{j \in S_0} : \| Y - (c(S_0^c) + \sum_{j \in S_0} c_m^{(j)}) \|_2 \leq \| Y - c([K]) \|_2, m_j \neq W_j \forall j \}
$$

• We have:

$$
\mathbb{P}[d_H(W, \hat{W}) = t] \leq \mathbb{P} \left[ \bigcup_{S_0 : |S_0| = t} F(S_0) \right]
$$

• Main goal: Show $\mathbb{P} \left[ \bigcup_{S_0 : |S_0| = t} F(S_0) \right] \to 0$ for all $t = \theta n$, $\theta \in [\epsilon, 1]$.
• Intermediate step: Bound $\mathbb{P}[F(S_0) | c_{[K]}, Y, W_{[K]}]$
• Define information density

\[ i_t(u; y | v) = \frac{n}{2} \log(1 + Pt) + \frac{\log e}{2} \left( \frac{\|y - v\|_2^2}{1 + P't} - \|y - u - v\|_2^2 \right), \]

• Define \( c(T) = \sum_{j \in T} c^{(j)}_{W_j}, \) \( c' = \sum_{j \in S_0} c^{(j)}_{m_j} \) for some \( m_j \neq W_j. \) Then:

\[ \{ \| Y - (c(S_0^c) + c') \|_2 \leq \| Y - c([K]) \|_2 \} = \{ i_t(c'; Y | c(S_0^c)) \geq i_t(c(S_0); Y | c(S_0)) \}. \]

• Let \( A_1, \ldots, A_K \stackrel{iid}{\sim} \mathcal{N}(0, PI_n) \) and \( B = \sum_i A_i + Z. \) For any \( S_0 \in \binom{[K]}{t} \):

\[ \log \frac{dP_{A_{S_0}}|A_{S_0^c}, B}{dP_{A_{S_0^c}}} = i_t(u; y | v), \]

where \( u = \sum_{j \in S_0} A_j, \) \( v = \sum_{j \in S_0^c} A_j \) and \( y = B. \)

• And thus we get:

\[ \mathbb{P} \left[ i_t(c'; Y | c(S_0^c)) > \gamma | Y, c[K], W[K] \right] \leq e^{-\gamma} \]
• We have shown (via union bound):
\[ P[F(S_0)|c[K], Y, W[K]] \leq M^t \exp\{-i_t(c(S_0); Y|c(S_0^c))\}. \]

• So we now use a smart union bound:
\[ P[\bigcup_{S_0} F(S_0)] \leq M^t \binom{K}{t} \exp\{-\gamma\} + P[I_t \leq \gamma], \]

where \( I_t = \min_{S_0} i_t(c(S_0); Y|c(S_0^c)) \)

• Left to study the extrema of Gaussian matrix \( G \in \mathbb{R}^{n \times \mu n} \) with 
\( \text{iid} \sim \mathcal{N}(0, 1) \)
\[ \Phi \triangleq \frac{1}{n} \min \left\{ \left\| \frac{\beta}{\sqrt{n}} Gx + Z \right\|_2 : x \in \{0, 1\}^{\mu n}, \|x\|_0 = \theta \mu n \right\} \]

• After dualizing norm, we get a problem:
\[ \mathbb{P}[\min_u \max_v A_{u,v} \leq c] \leq? \]
• We have shown (via union bound):
  \[ P[F(S_0) \mid c[K], Y, W[K]] \leq M^t \exp\{-i_t(c(S_0); Y \mid c(S_0^c))\}. \]

• So we now use a smart union bound:
  \[ P[\bigcup S_0 F(S_0)] \leq M^t \binom{K}{t} \exp\{-\gamma\} + P[I_t \leq \gamma], \]
  where \( I_t = \min_{S_0} i_t(c(S_0); Y \mid c(S_0^c)) \)

• Left to study the extrema of Gaussian matrix \( G \in \mathbb{R}^{n \times \mu n} \) with \( \text{iid} \sim \mathcal{N}(0, 1) \)
  \[ \Phi \triangleq \frac{1}{n} \min \left\{ \left\| \frac{\beta}{\sqrt{n}} G x + Z \right\|_2 : x \in \{0, 1\}^{\mu n}, \|x\|_0 = \theta \mu n \right\} \]

• After dualizing norm, we get a problem:
  \[ P[\min_u \max_v A_{u,v} \leq c] \leq P[\min_u \max_v B_{u,v} \leq c] \]

• Gaussian comparison method: Bound extrema of \( A \) via extrema of a simpler process \( B \)
**Theorem (Slepian)**

Let \( \{A_v\}_{v \in \mathcal{V}} \) and \( \{B_v\}_{v \in \mathcal{V}} \) be zero-mean Gaussian processes, s.t. \( \text{Cov}(A) \leq \text{Cov}(B) \) and \( \text{Var}[A_v] = \text{Var}[B_v] \) for all \( v \) then

\[
\mathbb{E}[\max_v A_v] \geq \mathbb{E}[\max_v B_v]
\]
Slepian’s lemma (1962) and Gordon’s lemma (1985)

**Theorem (Slepian)**

Let \( \{A_v\}_{v \in V} \) and \( \{B_v\}_{v \in V} \) be zero-mean Gaussian processes, s.t. \( \text{Cov}(A) \leq \text{Cov}(B) \) and \( \text{Var}[A_v] = \text{Var}[B_v] \) for all \( v \) then

\[
\max_v A_v \succeq \max_v B_v \quad \text{(stoch. domination)}
\]
Slepian’s lemma (1962) and Gordon’s lemma (1985)

**Theorem (Slepian)**

Let \( \{A_v\}_{v \in V} \) and \( \{B_v\}_{v \in V} \) be zero-mean Gaussian processes, s.t. \( \text{Cov}(A) \leq \text{Cov}(B) \) and \( \text{Var}[A_v] = \text{Var}[B_v] \) for all \( v \) then

\[
\max_v A_v \succeq \max_v B_v \quad \text{(stoch. domination)}
\]

**Theorem (Gordon)**

Let \( \{A_{u,v}\} \) and \( \{B_{u,v}\} \) be zero-mean Gaussian processes, s.t.

1. \( \text{Var}[A_{u,v}] = \text{Var}[B_{u,v}] \)
2. \( \mathbb{E}[A_{u,v}A_{u,v}'] \leq \mathbb{E}[B_{u,v}B_{u,v}'] \) for all \( u, v, v' \)
3. \( \mathbb{E}[A_{u,v}A_{u',v}'] \geq \mathbb{E}[B_{u,v}B_{u',v}'] \) for all \( u \neq u', v, v' \). Then:

\[
\min_u \max_v A_{u,v} \preceq \min_u \max_v B_{u,v}
\]
Slepian’s lemma (1962) and Gordon’s lemma (1985)

Theorem (Slepian)

Let \( \{A_v\}_{v \in \mathcal{V}} \) and \( \{B_v\}_{v \in \mathcal{V}} \) be zero-mean Gaussian processes, s.t. \( \text{Cov}(A) \leq \text{Cov}(B) \) and \( \text{Var}[A_v] = \text{Var}[B_v] \) for all \( v \) then

\[
\max_v A_v \succeq \max_v B_v \quad \text{(stoch. domination)}
\]

Theorem (Gordon)

Let \( \{A_{u,v}\} \) and \( \{B_{u,v}\} \) be zero-mean Gaussian processes, s.t.

1. \( \text{Var}[A_{u,v}] = \text{Var}[B_{u,v}] \)
2. \( \mathbb{E}[A_{u,v}A_{u,v}'] \leq \mathbb{E}[B_{u,v}B_{u,v}'] \) for all \( u, v, v' \)
3. \( \mathbb{E}[A_{u,v}A_{u',v'}] \geq \mathbb{E}[B_{u,v}B_{u',v'}] \) for all \( u \neq u', v, v' \).

Then:

\[
\min_u \max_v A_{u,v} \succeq \min_u \max_v B_{u,v}
\]

Remark: 2) implies \( A^*_u = \max_v A_{u,v} \succeq B^*_u = \max_v B_{u,v} \).

3) implies \( \{A^*_u\} \) is “more-correlated” than \( \{B^*_u\} \).
User density vs. Energy-per-bit: best bounds
User density vs. Energy-per-bit: CDMA (w/o MUD)

- Converse Achievability (Gaussian Process)
- Achievability: TIN (aka CDMA w/o MUD)
User density vs. Energy-per-bit: TDMA

- Converse Achievability (Gaussian Process)
- Achievability: TIN (aka CDMA w/o MUD)
- Achievability: TDMA
User density vs. Energy-per-bit: higher reliability

- Converse
- Achievability (Gaussian Process)
- Achievability: TIN (aka CDMA w/o MUD)
- Achievability: TDMA
Problem 3: Information theory of random-access
Prior work on MAC/random-access

It’s a mess...
Prior work on MAC/random-access

It’s a mess...

- Channel model: collision vs. additive
- Noise model: noiseless, stochastic or worst-case
- Coding with or without feedback (as in CSMA)
- Probability of error: zero, vanishing or fixed $> 0$.
- Probability of error: per-user vs all-users
- User activity: always-on vs sporadic
- Finite blocklength vs $n \to \infty$
- Various asymptotics: $K = \text{const}, n \to \infty$ vs both $K, n \to \infty$
Classification by user activity

Identifiable users
individual codebooks
$K_{tot} < \infty$

Non-identifiable users
one (same) codebook
$K_{tot} = \infty$
Classification by user activity

Identifiable users
- individual codebooks
  \[ K_{tot} < \infty \]
- All active
  \[ K_a = K_{tot} \]

Non-identifiable users
- one (same) codebook
  \[ K_{tot} = \infty \]
- Some active
  \[ K_a < K_{tot} \]
Classification by user activity

- **MAC**
  - **Identifiable users**
    - Individual codebooks
    - $K_{tot} < \infty$
  - **Non-identifiable users**
    - One (same) codebook
    - $K_{tot} = \infty$
  - **All active**
    - $K_a = K_{tot}$
    - Active set known
  - **Some active**
    - $K_a < K_{tot}$
    - Active set unknown
Sample of prior work

Identifiable users
individual codebooks
$K_{tot} < \infty$

Non-identifiable users
one (same) codebook
$K_{tot} = \infty$

• Classical IT
  [Liao’72], [Ahlswede’73]
• Orthogonal schemes TDMA/FDMA
• Rate splitting [Rimoldi-Urbanke’99]
• Finite blocklength [MolavianJazi-Laneman’14-16]
• Many-user [Chen-Guo’14]
Sample of prior work

- Non-orthogonal CDMA, MUD
- Randomly-spread CDMA
  - [Tse-Hanly’99], [Verdù-Shamai’99]
- [Mathys’90]
- LDS, SCMA
Sample of prior work

**MAC**

**Identifiable users**
- individual codebooks
- $K_{tot} < \infty$

- **All active**
  - $K_a = K_{tot}$
  - Active set known

- **Some active**
  - $K_a < K_{tot}$
  - Active set unknown

**Non-identifiable users**
- one (same) codebook
- $K_{tot} = \infty$

- **Active set**
  - Active set known
  - Active set unknown

- **Non-orthogonal CDMA, MUD**
- **Randomly-spread CDMA**
  - [Tse-Hanly'99], [Verdú-Shamai'99]
  - [Mathys'90]
- **LDS, SCMA**

- **Many-access** [Chen-Chen-Guo'17]
- **Blind-detection for CDMA**
  - [BarDavid-Plotnik-Rom'93]
- **conflict-avoiding codes**
  - [Bassalygo-Pinsker'83], B. Tsybakov
Sample of prior work

MAC

Identifiable users
individual codebooks
\( K_{tot} < \infty \)

Non-identifiable users
one (same) codebook
\( K_{tot} = \infty \)

- ALOHA [Abramson’70]
- [Massey-Mathys’85]
- Collision-resolution protocols
  [Capetanakis’79]
- Superimposed codes
  [Ericson-Gyorfi’88]
  [Furedi-Ruszinkó’99]
- \( B_r \)-codes [Dyachkov-Rykov’81]
- Coded Slotted ALOHA
  [Casini et al’07],[Liva’11]
- Compressed sensing
  [Jin-Kim-Rao’11]
Key definition: random-access code

\[ f : [M] \rightarrow \mathbb{R}^n \] is a random-access code for \( K_a \) users if \( \exists \) list-\( K_a \) decoder \( g \) s.t.

\[
\mathbb{P}[W_j \notin g(f(W_1) + \cdots + f(W_{K_a}) + Z)] \leq \epsilon \quad \forall j \in [K_a]
\]

where \( W_i \overset{iid}{\sim} \text{Unif}[M] \).

For \( \epsilon = 0 \) this was studied:

- Noiseless channels: \( B_r \)-codes [Dyackhov-Rykov’81]
- Worst-case noise: superimposed codes [Ericson-Gyorfi’88, Furedi-Ruszinkó’99]
Definition (P.’17)

\( f : [M] \rightarrow \mathbb{R}^n \) is a **random-access code** for \( K_a \) users if \( \exists \) list-\( K_a \) decoder \( g \) s.t.

\[ \mathbb{P}[W_j \not\in g(f(W_1) + \cdots + f(W_{K_a}) + Z)] \leq \epsilon \quad \forall j \in [K_a] \]

where \( W_i \overset{iid}{\sim} \text{Unif}[M] \).

For \( \epsilon > 0 \) this is:

- Just compressed sensing: \( Y = X \beta + Z \), \( X \) is \( K_a \)-out-of-\( M \) sparse.
- \( \Rightarrow \) studied by many, but not w.r.t. \( \frac{E_b}{N_0} \) and not with \( M = 2^{\Theta(n)} \).
Same-codebook codes = compressed sensing

- random-access = all users share same codebook
- ... obviously decoding is up to permutation of users
- **New problems:** capacity/error-exponent/zero-error capacity
- Equivalent to compressed-sensing [Jin-Kim-Rao’11]
Same-codebook codes = compressed sensing

- random-access = all users share same codebook
- ... obviously decoding is upto permutation of users
- **New problems**: capacity/error-exponent/zero-error capacity

**Equivalent to compressed-sensing** \([\text{Jin-Kim-Rao'11}]\)

- Let same-codebook (column) vectors be \(c_1, \ldots c_j\).

\[
X = \begin{pmatrix} c_1 & \cdots & c_M \end{pmatrix}
\]

- Let \(\beta \in \{0, 1\}^M\) with \(\beta_j = 1\) if codeword \(j\) was transmitted

- Then the problem is:

\[
Y = X\beta + Z, \quad \text{Goal: } \mathbb{E}[\|\beta - \hat{\beta}(Y)\|] \to \min
\]

(linear regression with sparsity \(\|\beta\|_0 = K_a\) aka comp.sensing).
Same-codebook codes = compressed sensing

- random-access = all users share same codebook
- ... obviously decoding is upto permutation of users
- **New problems:** capacity/error-exponent/zero-error capacity
- Equivalent to compressed-sensing [Jin-Kim-Rao’11]
- Let same-codebook (column) vectors be $c_1, \ldots, c_j$.

$$X = \begin{pmatrix} c_1 & \cdots & c_M \end{pmatrix}$$

- Let $\beta \in \{0, 1\}^M$ with $\beta_j = 1$ if codeword $j$ was transmitted
- Then the problem is:

$$Y = X\beta + Z, \quad \text{Goal: } \mathbb{E}[\|\beta - \hat{\beta}(Y)\|] \rightarrow \min$$

(linear regression with sparsity $\|\beta\|_0 = K_a$ aka comp.sensing).
- The famous $n \sim 2K_a \log_e M$ is just **TIN**:

$$\log_e M \approx \frac{n}{2} \log_e (1 + \frac{P}{1 + (K_a - 1)P}) \approx \frac{n}{2K_a}$$

So all the $L_1$ (LASSO) frenzy is just to achieve TIN (hehe...)
Key definition: random-access code

Definition (P.'17)

\( f: [M] \rightarrow \mathbb{R}^n \) is a random-access code for \( K_a \) users if \( \exists \) list-\( K_a \) decoder \( g \) s.t.

\[
\mathbb{P}[W_j \notin g(f(W_1) + \cdots + f(W_{K_a}) + Z)] \leq \epsilon \quad \forall j \in [K_a]
\]

where \( W_i \overset{iid}{\sim} \text{Unif}[M] \).

This definition is answer to many prayers, but . . .

Bad news: Asymptotics of \( K_a = \mu n, n \rightarrow \infty \) is nonsense.
Prototypical random-access code: ALOHA

- $n$-frame is partitioned into $L = \frac{n}{n_1}$ subframes of length $n_1$
- Each of $K_a$ users places his $n_1$-codeword into a random subframe.
- Per-user error: $P_e \approx \Pr[Bino(K_a - 1, \frac{1}{L}) > 0] \approx \frac{K_a}{L} e^{-\frac{K_a}{L}}$
II. RANDOM CODING BOUND

**Theorem 1.** Fix $P' < P$. There exists an $(M,n,\epsilon)$ random-access code for $K_a$-user GMAC satisfying power-constraint $P$ and

$$\epsilon \leq \sum_{t=1}^{K_a} \frac{t}{K_a} \min(p_t, q_t) + p_0,$$

where

$$p_0 = \frac{(K_a)}{M} + K_a \mathbb{P}\left[\frac{1}{n} \sum_{j=1}^{n} Z_j^2 > \frac{P}{P'}\right],$$

$$p_t = e^{-nE(t)},$$

$$E(t) = \max_{0 \leq \rho, \rho_1 \leq 1} -\rho \rho_1 t R_1 - \rho_1 R_2 + E_0(\rho, \rho_1)$$

$$E_0 = \rho_1 a + \frac{1}{2} \log(1 - 2b\rho_1)$$

$$a = \frac{\rho}{2} \log(1 + 2P't\lambda) + \frac{1}{2} \log(1 + 2P't\mu)$$

$$b = \rho \lambda - \frac{\mu}{1 + 2P't\mu}, \quad \mu = \frac{\rho \lambda}{1 + 2P't\lambda}$$

$$\lambda = \frac{P't - 1 + \sqrt{D}}{4(1 + \rho_1 \rho) P't},$$

$$D = (P't - 1)^2 + 4P't \frac{1 + \rho \rho_1}{1 + \rho}$$

$$R_1 = \frac{1}{n} \log M - \frac{1}{n} \log(t!),$$

$$R_2 = \frac{1}{n} \log\left(\frac{K_a}{t}\right)$$

$$q_t = \inf_{\gamma} \mathbb{P}[I_t \leq \gamma] + \exp\{n(R_1 + R_2) - \gamma\}$$

**Remark:** For classical regime $K_a$-fixed, $n \to \infty$ and $\epsilon \to 0$

$$C_{\text{random-access}}(K_a) = \frac{1}{2K_a} \log(1 + K_a P).$$
Random-coding achievability bound

- Generate $M$ codewords: $c_i \sim \mathcal{N}(0, P)^{\otimes n}$.
- WLOG, users send $c_1, c_2, \ldots, c_{K_a}$.
- Decoder sees
  \[ Y = c_1 + \cdots + c_{K_a} + Z \]
- Define sum-codewords $c(S) \triangleq \sum_{i \in S} c_i$
- ML-decoder (not optimal!)
  \[ \hat{S} = \arg \min_S \|c(S) - Y\| \]
- Error-analysis:
  \[
P_e \leq \sum_{t=1}^{K_a} \frac{t}{K_a} \mathbb{P}[t\text{-misguessed}]
  \]
  \[
  \mathbb{P}[t\text{-misguessed}] \leq \mathbb{P}\left[ \bigcup_{S \in \binom{[K_a]}{t}} \bigcup_{S' \in \binom{M-K_a}{t}} \|c(S) - c(S') + Z\| \leq \|Z\| \right]
  \]
Random-coding achievability bound

- Generate \( M \) codewords: \( c_i \sim N(0,P)^{\otimes n} \)

Analysis I:
- Condition on \( Z, c_1, \ldots, c_{K_a} \)
- Use Chernoff + Gallager \( \rho \)-trick for \( \mathbb{P}[\cup S' \cdots | c_{1}^{K_a}, Z] \)
- Use another Gallager \( \rho \)-trick for \( \mathbb{P}[\cup S \cdots | Z] \)
- Finally take expectation over \( Z \)

ML-decoder (not optimal!)

\[
\hat{S} = \arg \min_S \| c(S) - Y \|
\]

Error-analysis:

\[
P_e \leq \sum_{t=1}^{K_a} \frac{t}{K_a} \mathbb{P}[t\text{-misguessed}]
\]

\[
\mathbb{P}[t\text{-misguessed}] \leq \mathbb{P}
\left[
\bigcup_{S \in \binom{K_a}{t}} \bigcup_{S' \in \binom{M-K_a}{t}} \| c(S) - c(S') + Z \| \leq \| Z \|
\right]
\]
Random-coding achievability bound

- Generate $M$ codewords: $c_i \sim N(0, P^\otimes n)$.

**Analysis I:**
- Condition on $Z, c_1, \ldots, c_{K_a}$
- Use Chernoff + Gallager $\rho$-trick for $\mathbb{P}[\bigcup S', \ldots | c_1^{K_a}, Z]$
- Use another Gallager $\rho$-trick for $\mathbb{P}[\bigcup S \cdots | Z]$
- Finally take expectation over $Z$

- ML-decoder (not optimal!)

**Analysis II:**
- Define information density appropriately
- Use Feinstein’s trick to bound

$$\mathbb{P}[\bigcup S \cup S' \cdots] \leq \mathbb{P}[i_{\text{min}}(X_1^{K_a}; Y) < \gamma] + \binom{K_a}{t} \binom{M}{t} e^{-\gamma}$$

$$i_{\text{min}} = \min_S i_t(c(S); Y | c(S^c))$$

- $i_{\text{min}} \approx \max$ of Gaussian process indexed by $t$-subsets of $[K_a]$

$$\mathbb{P}[t\text{-misguessed}] \leq \mathbb{P} \left[ \bigcup_{S \in \binom{[K_a]}{t}} \bigcup_{S' \in \binom{[M-K_a]}{t}} \| c(S) - c(S') + Z \| \leq \| Z \| \right]$$
Random-coding achievability bound

- Generate $M$ codewords: $c_i \sim N(0, P) \otimes n$

**Analysis I:**
- Condition on $Z, c_1, \ldots, c_{K_a}$
- Use Chernoff + Gallager $\rho$-trick for $\mathbb{P}[\bigcup S' \cdots | c_1^{K_a}, Z]$
- Use another Gallager $\rho$-trick for $\mathbb{P}[\bigcup S \cdots | Z]$
- Finally take expectation over $Z$

**IML-decoder (not optimal!)**

**Analysis II:**
- Define information density appropriately
- Use Feinstein’s trick to bound
  \[
  \mathbb{P}[\bigcup S \cup S' \cdots ] \leq \mathbb{P}[\min_i (X_1^{K_a}; Y) < \gamma] + \binom{K_a}{t} \binom{M}{t} e^{-\gamma}
  \]
- $\min_i = \min S i_t(c(S); Y | c(S^c))$
- $\min \approx \max$ of Gaussian process indexed by $t$-subsets of $[K_a]$

**Classical IT:** term $S$ goes $\to 0$ if $I(X_S; Y | X_{S^c}) > \sum_{i \in S} R_i$
Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

For $K_a \lessapprox 50$ dominant term $t \leq 3$

For $K_a \gg 150$ dominant term $t = K_a$
Energy−per−bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

For $K_a \lesssim 50$ dominant term $t \leq 3$
For $K_a \gtrsim 150$ dominant term $t = K_a$
Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- **NOMA: random–coding achievability**
- **Lower bound**
- **ALOHA**
Energy–per–bit vs. number of users. Payload \( k = 100 \text{ bit} \), frame \( n = 30000 \text{ rdof} \), \( P_e = 0.1 \)

- **ALOHA**
- **DT–TIN bound**
- **NOMA: random–coding achievability**
- **Lower bound**

**Notes:**
- \( \text{Eb/N0, dB} \)
- \# active users
- Yury Polyanskiy
- MAC tutorial 114
Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- ALOHA
- NOMA: random–coding achievability
- Lower bound
- Coded ALOHA (irreg., rep. rate = 3.6)
- Coded ALOHA (2–regular)
... and randomly-spread CDMA w/ optimal MUD

Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- ALOHA
- NOMA: random–coding achievability
- Lower bound
- Coded ALOHA (irreg., rep. rate = 3.6)
- Coded ALOHA (2–regular)
- Random CDMA, BPSK, optimal MUD; $K_a/N=1$
New twists compared to classic MAC

Problem 1 large $K \rightarrow \infty$, fixed payload $\log_2 M$

- Relevant asymptotics: $K, n \rightarrow \infty$ with $\frac{K}{n} = \mu$.

Problem 2 “user-centric” probability of error

- $P_e \triangleq \frac{1}{K} \sum_j \mathbb{P}[\hat{X}_j \neq X_j]$

Problem 3 “random-access”

- indistinguishable users (same-codebook), non-asymptotics.
Low-complexity random-access over GMAC
Key challenge:

Providing multiple-access to massive number of **UNCOORDINATED**
and infrequently communicating devices
Key challenge:
Providing multiple-access to massive number of UNCOORDINATED and infrequently communicating devices

Typical scenario:
- Huge # of users $K_{\text{tot}} \approx 10^6 - 10^7$
- Still large # of active users $K_a \approx 1 - 500$
- Small data payload, e.g. $k = 100$ bits
- Blocklength $n \sim 10^4$
- $\frac{k}{n} \ll 1$, but system spectral efficiency $\rho = \frac{K_a \cdot k}{n} \sim 1$
Key challenge:

Providing multiple-access to massive number of UNCOORDINATED
and infrequently communicating devices

Typical scenario:

- Huge \# of users $K_{\text{tot}} \approx 10^6 - 10^7$
- Still large \# of active users $K_a \approx 1 - 500$
- Small data payload, e.g. $k = 100$ bits
- Blocklength $n \sim 10^4$
- $\frac{k}{n} \ll 1$, but system spectral efficiency $\rho = \frac{K_a \cdot k}{n} \sim 1$

The goal is to communicate with the smallest possible energy-per-bit
Simple scheme I: Treat interference as noise (TIN)

**Theorem (DT-TIN bound)**

There exists $\mathcal{C} \subset B(0, \sqrt{nP})$ of size $M$ such that

$$
\Pr[X_1 \notin \{\text{top-}K_a \text{ closest c/w to } Y\}] \leq \mathbb{E} \left[ e^{-|i(X;X+Z)-\log M|} \right]
$$

where $Y = X_1 + \cdots + X_{K_a} + Z$, $X_i$ - uniform on $\mathcal{C}$, $X \sim \mathcal{N}(0, P) \otimes n$ and $Z \sim \mathcal{N}(0, 1) \otimes n$.

**Remarks:**

- Decoder searches for top-$K_a$ closest codewords
- Achieves about $\log M \approx nC_{TIN}(P) - \sqrt{nV_{TIN}(P)}Q^{-1}(\epsilon)$

$$
C_{TIN}(P) = \frac{1}{2} \log \left(1 + \frac{P}{1+(K_a-1)P}\right), \quad V_{TIN}(P) = \frac{P\log^2 e}{1+(K_a-1)P}.
$$

- Spectral efficiency as $K_a \to \infty$ is bounded by $\frac{\log_2 e}{2} \approx 0.72 \text{ bit.}$
Simple scheme I: Treat interference as noise (TIN)

Energy-per-bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- ALOHA
- DT-TIN bound
- NOMA: random-coding achievability
- Lower bound
Simple scheme II: $T$-fold ALOHA

- Each user places his $n_1$-codeword into one of $L$ subframes.
- Assume any $T$-fold collision is resolvable.
- Per-user error: $P_e \approx P[Bino(K_a - 1, \frac{1}{L}) > T] \approx \left(\frac{K_a}{L}\right)^T e^{-\frac{K_a}{L}}$
Simple scheme II: $T$-fold ALOHA

Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- NOMA: random–coding achievability
- Lower bound
- ALOHA
- DT–TIN bound
- 5–fold ALOHA

Yury Polyanskiy
MAC tutorial
Simple scheme II: $T$-fold ALOHA

Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

Want $T$-MAC codes for $T \sim 3-10$
Our scheme: high-level idea

- Send lattice points
- **Decode sum of codewords** via single-user decoder [Nazer-Gastpar’11]
- Use a subset of points forming a Sidon set (all sums $c_1 + c_2$ distinct)
- Single-lattice (no MMSE scaling): $R \approx \frac{1}{2K} \log^+ P$
- Nested-lattice (with MMSE scaling): $R \approx \frac{1}{2K} \log^+ \left( \frac{1}{K} + P \right)$
  
  Warning: issues with same-dither
- Lose power-factor compared to $\frac{1}{2K} \log(1 + KP)$
Sample performance of new scheme

![Sample performance of new scheme](image-url)
Many ideas appeared separately:

- Compute-and-forward [Nazer-Gastpar’11]
- Explicit codes for the modulo-2 binary adder channel [Lindström’69, Bar-David et al.’93]
- 2-user codes for $\mathbb{F}_q$-adder MAC [Dumer-Zinoviev’78, Dumer’95]
- Concatenation of codes with good minimum distance and codes for the BAC [Ericson-Levenshtein’94]
- Concatenation of CoF inner codes with syndrome decoding for compressed sensing [Lee-Hong’16]
Details of our scheme

Three phases:
- Sidon set: $\{0, 1\}^k \rightarrow \mathbb{F}_p^n$
- Compute-and-forward: $\mathbb{F}_p^n \rightarrow \mathbb{R}^{n_1}$
- $T$-fold ALOHA: Place $n_1$-codeword in a random subframe
Inner code (CoF):
Convert $T$-user GMAC into a mod-$p$ (noiseless) adder MAC.

$w_1, \ldots, w_T$ are vectors in $\mathbb{Z}_p$

$\mathcal{C}_{lin}$ is linear code over $\mathbb{Z}_p$
Inner code (CoF):
Convert $T$-user GMAC into a mod-$p$ (noiseless) adder MAC.

$w_1, \ldots, w_T$ are vectors in $\mathbb{Z}_p$
$C_{\text{lin}}$ is linear code over $\mathbb{Z}_p$

\[ y_{\text{BAC}} = \left[ \sum_{i=1}^{T} w_i \right] \mod p \]
Concatenation scheme

**Inner code (CoF):**
Convert $T$-user GMAC into a mod-$p$ (noiseless) adder MAC.

**Outer code (BAC):**
$C_{\text{BAC}}$ code for mod-$p$ adder $T$-MAC  
Here: only $p = 2$

$y_{\text{BAC}} = \left[ \sum_{i=1}^{T} w_i \right] \mod p$
Concatenation scheme

Inner code (CoF):
Convert $T$-user GMAC into a mod-$p$ (noiseless) adder MAC.

Outer code (BAC):
$C_{BAC}$ code for mod-$p$ adder $T$-MAC  Here: only $p = 2$

\[ \sum_{i=1}^{T} w_i \mod p \]

$w_1, \ldots, w_T$ are vectors in $\mathbb{Z}_p$

$C_{\text{lin}}$ is linear code over $\mathbb{Z}_p$
• $C_{\text{lin}} \subset \{0, 1\}^n$ is a binary linear code (shifted to $\pm \sqrt{P}$)

• Receive $y = \sum_{i=1}^{T} x_i + z$, shift, rescale, take mod-2, get

$$y_{\text{CoF}} = [x + z] \mod 2$$

where $x = [\sum_i x_i] \mod 2 \in C_{\text{lin}} \subset \{0, 1\}^n$

• The channel from $x$ to $y_{\text{CoF}}$ is a **BMS with folded Gsn noise**

$$\implies \text{Designing } C_{\text{lin}} \text{ is a standard coding task}$$

**Normal approximation:** $\log |C_{\text{lin}}| \approx nC - \sqrt{nV} Q^{-1}(\epsilon_{\text{code}})$
More on the CoF phase

- $C_{\text{lin}} \subset \{0,1\}^n$ is a binary linear code (shifted to $\pm \sqrt{P}$)
- Receive $y = \sum_{i=1}^{T} x_i + z$, shift, rescale, take mod-2, get
  $$y_{\text{CoF}} = [x + z] \mod 2$$
  where $x = [\sum_i x_i] \mod 2 \in C_{\text{lin}} \subset \{0,1\}^n$
- The channel from $x$ to $y_{\text{CoF}}$ is a BMS with folded Gsn noise
  $\implies$ Designing $C_{\text{lin}}$ is a standard coding task
  Normal approximation: $\log |C_{\text{lin}}| \approx nC - \sqrt{nVQ^{-1}(\epsilon_{\text{code}})}$

What is lost in the conversion $y \mapsto y_{\text{CoF}}$?

Sum-capacity of $y$ grows like $\log(T \cdot P)$
Capacity of $y_{\text{CoF}}$ only grows like $\log(P)$
More on the CoF phase

- $C_{\text{lin}} \subset \{0, 1\}^n$ is a binary linear code (shifted to $\pm \sqrt{P}$)
- Receive $y = \sum_{i=1}^{T} x_i + z$, shift, rescale, take mod-2, get

$$y_{\text{CoF}} = [x + z] \mod 2$$

where $x = [\sum_i x_i] \mod 2 \in C_{\text{lin}} \subset \{0, 1\}^n$

- The channel from $x$ to $y_{\text{CoF}}$ is a BMS with folded Gsn noise
  $\implies$ Designing $C_{\text{lin}}$ is a standard coding task
  Normal approximation: $\log |C_{\text{lin}}| \approx nC - \sqrt{nVQ^{-1}}(\epsilon_{\text{code}})$

What is lost in the conversion $y \mapsto y_{\text{CoF}}$?

Sum-capacity of $y$ grows like $\log(T \cdot P)$
Capacity of $y_{\text{CoF}}$ only grows like $\log(P)$

$T$-fold ALOHA reduces “power-loss” to $1/T$ instead of $1/K_a$
More on the BAC Phase

\[ y_{\text{BAC}} = \left[ \sum_{i=1}^{T} w_i \right] \mod 2, \quad w_1, \ldots, w_T \in C_{\text{BAC}} \]

Need to decode a list \( \{w_1, \ldots, w_T\} \)

Symmetric-capacity: \( C_{\text{sym}} = \frac{1}{T} \)
More on the BAC Phase

\[ y_{\text{BAC}} = \left[ \sum_{i=1}^{T} w_i \right] \mod 2, \quad w_1, \ldots, w_T \in \mathcal{C}_{\text{BAC}} \]

Need to decode a list \( \{w_1, \ldots, w_T\} \)

Symmetric-capacity: \( C_{\text{sym}} = \frac{1}{T} \)

How to construct explicit codes?

- Let \( H = [h_1 | \cdots | h_N] \) be the parity-check matrix of a \( T \)-error correcting code
- \( \Rightarrow \) all \( T \)-sums of columns are distinct
- Set \( \mathcal{C}_{\text{BAC}} = \{h_1, \ldots, h_N\} \)
- BCH parity check matrix: \( R_{\text{BAC}} = \frac{1}{T} \) (optimal!)
- Encoding: easy (just compute \( \alpha, \alpha^3, \cdots, \alpha^{2T-1} \))
More on the BAC Phase

\[
y_{\text{BAC}} = \left[ \sum_{i=1}^{T} w_i \right] \mod 2, \quad w_1, \ldots, w_T \in \mathcal{C}_{\text{BAC}}
\]

Need to decode a list \( \{w_1, \ldots, w_T\} \)

Symmetric-capacity: \( C_{\text{sym}} = \frac{1}{T} \)

How to construct explicit codes?

- Let \( H = [h_1 | \cdots | h_N] \) be the parity-check matrix of a \( T \)-error correcting code
- \( \Rightarrow \) all \( T \)-sums of columns are distinct
- Set \( \mathcal{C}_{\text{BAC}} = \{h_1, \ldots, h_N\} \)
- BCH parity check matrix: \( R_{\text{BAC}} = \frac{1}{T} \) (optimal!)
- Encoding: easy (just compute \( \alpha, \alpha^3, \cdots, \alpha^{2T-1} \))

Problem: decoding complexity of BCH linear in \( n = 2^k - 1 \)
Decoding:

- $\alpha_1, \ldots, \alpha_T \in \mathbb{F}_{2^k}$ are messages
- $y_{\text{BAC}} = H e' - \text{syndrome (!)} \implies$ we know $\sum_i (\alpha_i)^s$, $s \leq 2T$
- **Error locator**: Berlekamp-Massey yields coeffs of
  \[
  \sigma(z) = \prod_{i=1}^{T} (1 + \alpha_i z)
  \]
- **Find roots of** $\sigma(\cdot)$ e.g. via [Rabin’80]
- **Invert roots**: using the identity $\alpha^{-1} = \alpha^{2^k} - 1$

Total complexity: $O(kT^2 \log^2(T) \log \log(T))$ operations in $\mathbb{F}_{2^k}$
The spectral efficiency $\rho = \frac{K_a \cdot k}{n}$ of our scheme is at most $R_{\text{lin}}$.

What if $\rho > 1$?

Solution: - work with $p > 2$

- CoF phase requires good linear codes over $\mathbb{F}_p$
- BAC phase can be implemented using $H = [h_1 | \cdots | h_n]$ of a $[n = p^s - 1, n - k = 2T]$ Reed-Solomon code over $\mathbb{F}_{p^s}$ with

$$C_{\text{BAC}} = \{ \alpha h_i : \alpha \in \mathbb{F}_{p^s} \setminus \{0\}, i = 1, \ldots, p^s - 1 \}$$

- Can use nested lattice to achieve the 1.53dB shaping gain
- **Drawback**: hard to analyze finite blocklength
Approximate performance

Asymptotic optimum: \( \left( \frac{E_b}{N_0} \right)^* = \frac{2^{2\rho - 1}}{2\rho} \), with \( \rho = \frac{K_a \cdot k}{n} \).

Let \( L = \frac{K_a}{\alpha T} \) for \( \alpha \in (0, 1] \) be number of subframes.

\[ P_e \approx \mathbb{P}[T\text{-collision}] = \Pr \left( \text{Binomial} \left( K_a - 1, \frac{\alpha T}{K_a} \right) \geq T \right) \]

Linear code rate \( R_{\text{lin}} = \frac{\rho}{\alpha} \)

\[
\Delta = \left( \frac{E_b}{N_0} \right) \text{dB} - \left( \frac{E_b}{N_0} \right)^* \text{dB}
\approx 6\rho \frac{1 - \alpha}{\alpha} + 10 \log_{10}(\alpha)
\]

T-Collision avoidance loss due to a \( 1/\alpha \) increase in spectral efficiency
Approximate performance

Asymptotic optimum: \( \left( \frac{E_b}{N_0} \right)^* = \frac{2^{2\rho-1}}{2\rho}, \) with \( \rho = \frac{K_a \cdot k}{n} \).

Let \( L = \frac{K_a}{\alpha T} \) for \( \alpha \in (0, 1] \) be number of subframes

\[
P_e \approx \mathbb{P}[T\text{-collision}] = \Pr \left( \text{Binomial} \left( K_a - 1, \frac{\alpha T}{K_a} \right) \geq T \right)
\]

Linear code rate \( R_{\text{lin}} = \frac{\rho}{\alpha} \)

\[
\Delta = \left( \frac{E_b}{N_0} \right) \text{dB} - \left( \frac{E_b}{N_0} \right)^* \text{dB}
\approx 6 \rho \frac{1 - \alpha}{\alpha} + 10 \log_{10}(\alpha) + 10 \log_{10}(T)
\]

CoF loss from the reduction \( y \mapsto y_{\text{CoF}} \)
Approximate performance

Asymptotic optimum: \( \left( \frac{E_b}{N_0} \right)^* = \frac{2^{2\rho-1}}{2\rho}, \) with \( \rho = \frac{K_a \cdot k}{n}. \)

Let \( L = \frac{K_a}{\alpha T} \) for \( \alpha \in (0, 1] \) be number of subframes

\[ P_e \approx \mathbb{P}[\text{T-collision}] = \Pr \left( \text{Binomial} \left( K_a - 1, \frac{\alpha T}{K_a} \right) \geq T \right) \]

Linear code rate \( R_{\text{lin}} = \frac{\rho}{\alpha} \)

\[
\Delta = \left( \frac{E_b}{N_0} \right) \text{dB} - \left( \frac{E_b}{N_0} \right)^* \text{dB}
\]

\[
\approx 6\rho \frac{1-\alpha}{\alpha} + 10 \log_{10}(\alpha) + 10 \log_{10}(T) - 10 \log_{10}(1 - 2^{-2\rho})
\]

Loss of +1 in computation rate
Approximate performance

Asymptotic optimum: \( \left( \frac{E_b}{N_0} \right)^* = \frac{2^{2\rho - 1}}{2\rho} \), with \( \rho = \frac{K_a \cdot k}{n} \).

Let \( L = \frac{K_a}{\alpha T} \) for \( \alpha \in (0, 1] \) be number of subframes.

\[
Pe \approx \mathbb{P}[T\text{-collision}] = \Pr \left( \text{Binomial} \left( K_a - 1, \frac{\alpha T}{K_a} \right) \geq T \right)
\]

Linear code rate \( R_{\text{lin}} = \frac{\rho}{\alpha} \)

\[
\Delta = \left( \frac{E_b}{N_0} \right) \text{dB} - \left( \frac{E_b}{N_0} \right)^* \text{dB}
\]

\[
\approx 6\rho \frac{1 - \alpha}{\alpha} + 10 \log_{10}(\alpha) + 10 \log_{10}(T) - 10 \log_{10}(1 - 2^{-2\rho}) + 1.53
\]

Shaping loss
Low-complexity schemes: summary
MAC with random path-loss
\[ Y(t) = H_1 X_1(t) + \cdots + H_K X_K(t) + Z(t) \]

- More realistic model: waveforms added with random gains
- Standard work-around: use pilots
- Impossible without coordination!
New multi-access protocol (2018): idea

- **Step 1:** Partition entire frame into subframes of length $N$

![Diagram showing partitioning of a frame into subframes]
New multi-access protocol (2018): idea

- **Step 1:** Partition entire frame into subframes of length $N$.

- **Step 2:** Each user randomly selects a subframe for communication. *Important:* $K$ and $\frac{n}{N}$ are chosen so that $> T$-fold collisions are improbable.
New multi-access protocol (2018): idea

- **Step 1:** Partition entire frame into subframes of length \( N \)

- **Step 2:** Each user randomly selects a subframe for communication.

- **Step 3:** Users encode their data via sparse-graph (LDPC) codes

- **Step 4:** Decoder uses joint Tanner graph (LDPC+LDGM structure) to iteratively decode data and learn the channel gains.
New multi-access protocol (2018): idea

- **Step 1**: Partition entire frame into subframes of length $N$

  ![Partitioned Frame Diagram](image)

- **Step 2**: Each user randomly selects a subframe for communication.
- **Step 3**: Users encode their data via sparse-graph (LDPC) codes
- **Step 4**: Decoder uses joint Tanner graph (LDPC+LDGM structure) to iteratively decode data and learn the channel gains!

![Joint Tanner Graph Diagram](image)
New multi-access protocol (2018): results

Fading MAC: $E_b/N_0$ (dB) vs $K_a$ for $n=30000$, $k=100$ bits, $P_e=0.1$

- 1-ALOHA using LDPC scheme
- 2-ALOHA using LDPC scheme
- 3-ALOHA using LDPC scheme
- 4-ALOHA using LDPC scheme
- 2-ALOHA using FBL bound
- 3-ALOHA using FBL bound
- 4-ALOHA using FBL bound
- 1-ALOHA using FBL bound
- Shamai-Bettesh asymptotic bound
- Converse

ALOHA breaks down at about $\sim 20$ users
NEW scheme at about $\sim 250$ users
New multi-access protocol (2018): results

Fading MAC: $E_b/N_0$ (dB) vs $K_a$ for $n=30000$, $k=100$ bits, $P_e=0.1$

ALOHA breaks down at about $\sim 20$ users
NEW scheme at about $\sim 250$ users
Other ideas for low-complexity schemes

- Work in progress by several groups
  - Narayanan-Chamberland
  - P.-Frolov
  - Durisi-Dalai
  - Popovski-Liva
  - ... (sorry to those I forgot)

- Methods we did not cover:
  - Coded Slotted ALOHA
  - ... including with MPR capability
  - iterative decoding same-codebook LDPCs
  - super-imposed codes

- Problem is even more interesting with fading
  - Random channel gains $H_j$ help distinguish users.
  - With many users, order statistics of $H_j$'s becomes deterministic.
Envisioned solution:

- To save battery: sensors sleep all the time, except transmissions.
- ... uncoordinated transmissions.
- ... they wake up, blast the packet, go back to sleep.
- Focus on low-energy \((\text{low } E_b/N_0)\)
- Focus on fundamental limits
- ... but with low-complexity solutions (single-user-only decoding).
Envisioned solution:

- To save battery: sensors sleep all the time, except transmissions.
- ... uncoordinated transmissions.
- ... they wake up, blast the packet, go back to sleep.
- Focus on low-energy (low $E_b/N_0$)
- Focus on fundamental limits
- ... but with low-complexity solutions (single-user-only decoding).

Issues we need to understand:

1. packets are short: finite-blocklength (FBL) info theory
2. multiple-access channel: Classical MAC
3. low-complexity MAC: modulation, CDMA, multi-user detection
4. massive random-access: many users, same-codebook codes (NEW)

Supporting 10 users at 1Mbps is much easier than 1M users at 10bps.
Thank you!
Extra: More plots
AlOHa + codes repairing 5-fold collisions

Energy–per–bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- NOMA: random–coding achievability
- Lower bound
- ALOHA
- DT–TIN bound
- 5–fold ALOHA

Yury Polyanskiy
MAC tutorial
Energy-per-bit vs. number of users. Payload $k = 100$ bit, frame $n = 30000$ rdof, $P_e = 0.1$

- ALOHA
- ALOHA + 5MAC
- NOMA: Treat interference as noise (TIN)
- NOMA: random-coding achievability
- Lower bound
- Coded ALOHA (irreg., rep. rate = 3.6)
- Coded ALOHA (2-regular)
- Random CDMA, BPSK, optimal MUD; $K_a/N=1$