Massive multiple-access: A very short overview

William Weijia Zheng

Feb. 02, 2024

Some *past* news from the market side…

- 2020 data: "average time spent on Mobile phone in USA is ≈ 4 hours per day"

"Global System for Mobile Communication Association"

- 2023 data (GSMA): "***Mobile penetration is approaching saturation*** in most markets around the world, especially among adult and urban populations." Back to 2021, the total number of mobile devices exceeded the world population. (Some people may have ≥ 2 phones.)

- Screen quality will soon improve to a level that people can barely distinguish. *(I can barely distinguish 4K and 8K)*

- Lessons we learnt:
  - Almost all people who want a phone have one, or more.
  - People already spent a lot of time on content-rich Apps, with good QoE.
  - ***Maybe it is time for us to diversify the wireless services to machines*** **:)**

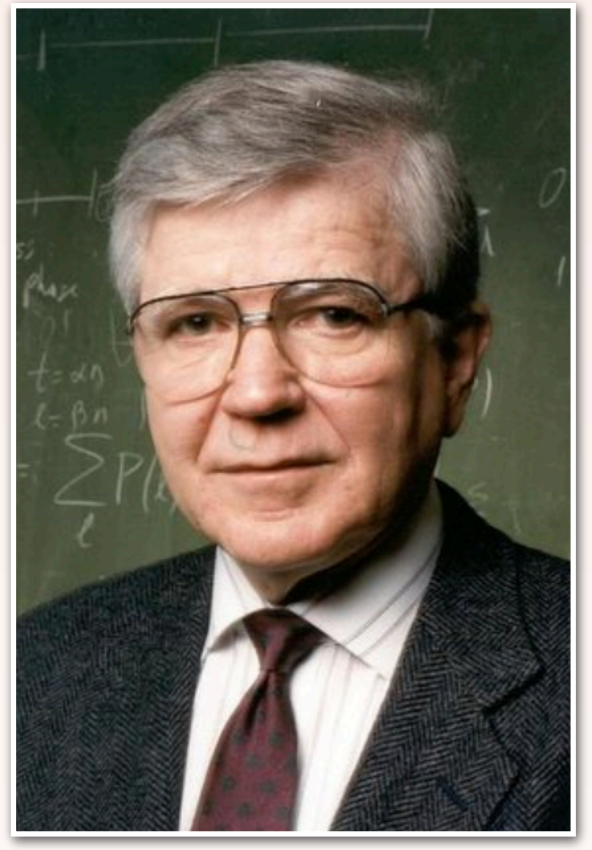
Machines are unlike human

- ***Timescales tailored for human:***
  - YouTube counts a view if you watch a video $\geq 30s$.
  - When visiting a website, most people want to wait up to $\leq 3s$.
- ***Timescales for machines:***
  - Microcontrollers work in μs $(= 10^{-6}s)$
  - LTE schedule: in ms $(= 10^{-3}s)$
  - ….
- Given the small timescale of machines, they do not (have the resource to) transmit large payloads. Hence a grant-free manner of communication is preferable. *(When the payload is* 100*bits, then to spend extra resource to acquire grant is not a good deal.)*
- *"Device transmissions are often sporadic with very short payloads"*

Motivation of massive MAC

- "A challenge … is to provide co-existence over the same band of a massive (from hundreds to thousands) number of infrequently communicating devices. …"

- **[NOT a brand new problem]** In 1985, Robert Gallager (who proposed LDPC in 1960) called for "a coding technology that is applicable for a large set of transmitters of which a small, but variable subset simultaneously use the channel."



R. Gallager (1931-)

- ***Sounds interesting. But…*** why not referring back to classical MAC questions? Or, is there any old solution that can work well in this new setting?

Envisioned features of massive MAC

- Comparing with conventional MAC, the differences of massive MAC are:
  - *Finite block length effect* (as mentioned, sensors/microcontrollers do not transmit long payloads as human) E.g., 100 – 200bits per connection.
  - $K_{tot} \sim 10^3$ to 10^4 devices around the access point (AP), but only an ***unknown*** and ***time-varying*** subset of devices with size $K_a \sim 1$ to 300 are active at any time *(random access)*.
  - Devices may carry the same chipset *(because it is cheap)*, that is, they all share the same codebook, and they operate one the same (congested) band.
  - ***Given all discussed, in the following, we will only focus on Gaussian MAC.***

(Classical) 2-sender Gaussian MAC

- $Y = g_1 X_1 + g_2 X_2 + Z$. *Where $Y, X_1, X_2, Z \in \mathbb{R}^n$, with each entry of $Z \sim N(0,1)$. $g_1, g_2 \in \mathbb{R}$ are (known) channel gains.*
- In transmission time $i \in [1,n]$, the channel output is $Y_i = g_1 X_{1i} + g_2 X_{2i} + Z_i$.
- We require $\frac{1}{n}\sum_{i=1}^{n} x_{ji}^2(m_j) \le P$ *(avg. transmission power constraints)*
- We define the received powers (SNRs) as $S_j = g_j^2 P, j \in \{1,2\}$.

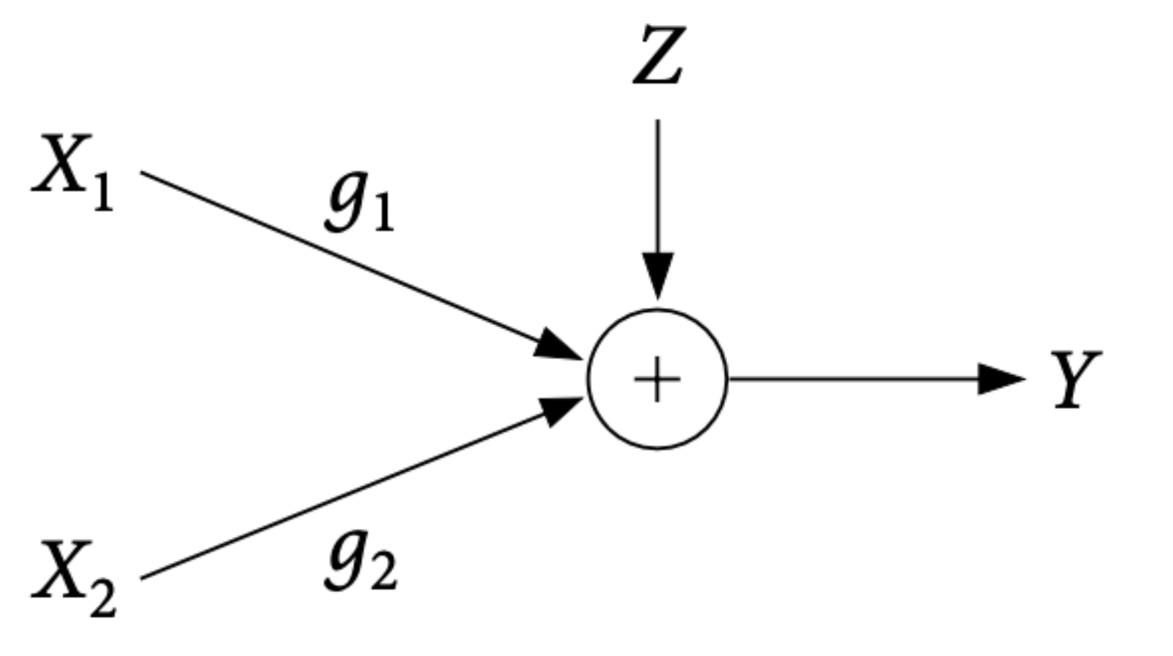


Figure 4.6. Gaussian multiple access channel.

Theorem 4.4. The capacity region of the Gaussian MAC is the set of rate pairs (R_1, R_2) such that

$$R_1 \le \mathsf{C}(S_1),$$
$$R_2 \le \mathsf{C}(S_2),$$
$$R_1 + R_2 \le \mathsf{C}(S_1 + S_2),$$

where $\mathsf{C}(x)$ is the Gaussian capacity function. **Where** $C(x) = \frac{1}{2}\log(1+x), x \ge 0.$

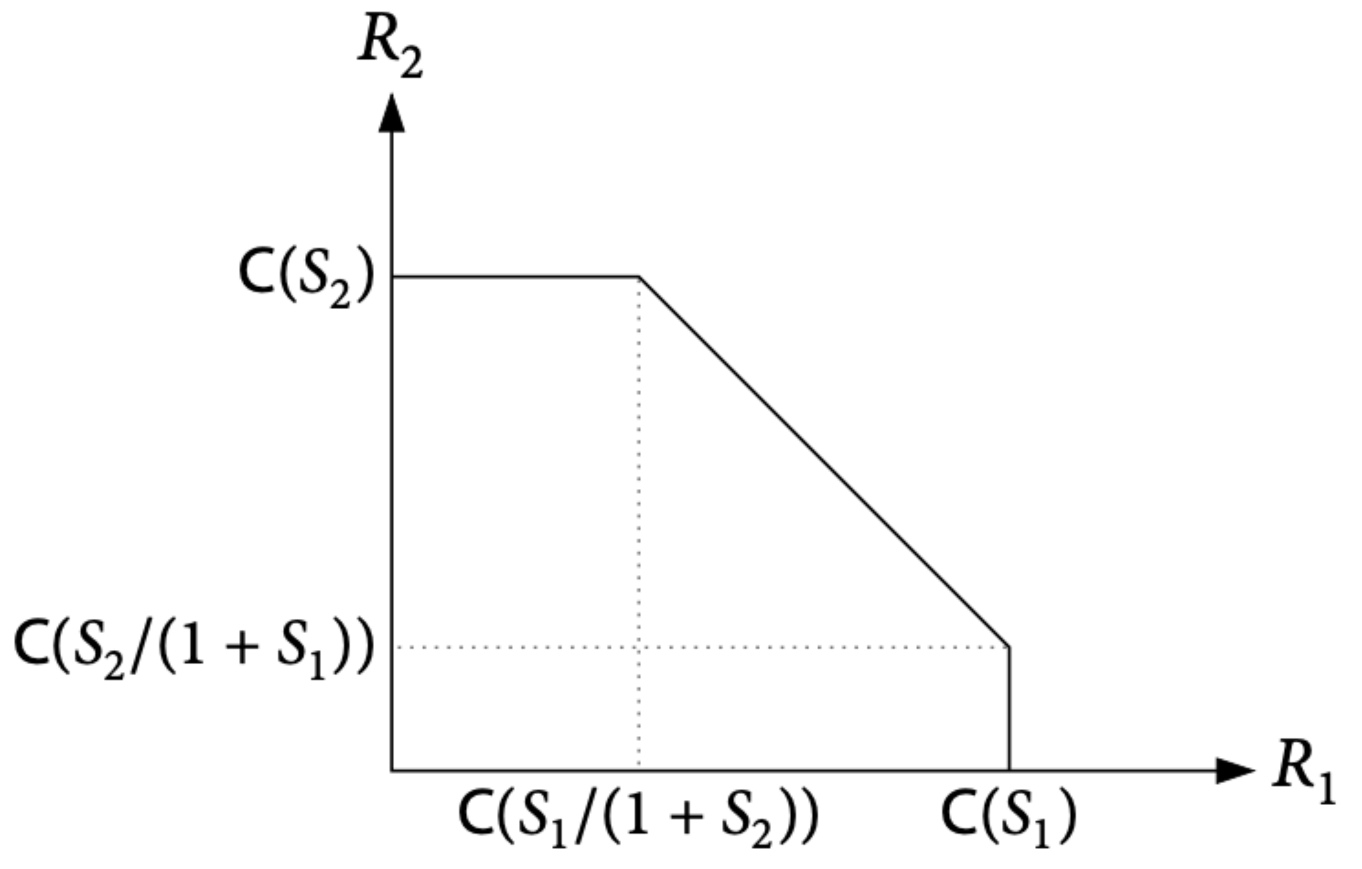


Figure 4.7. Capacity region of the Gaussian MAC.

(Classical) 2-sender Gaussian MAC

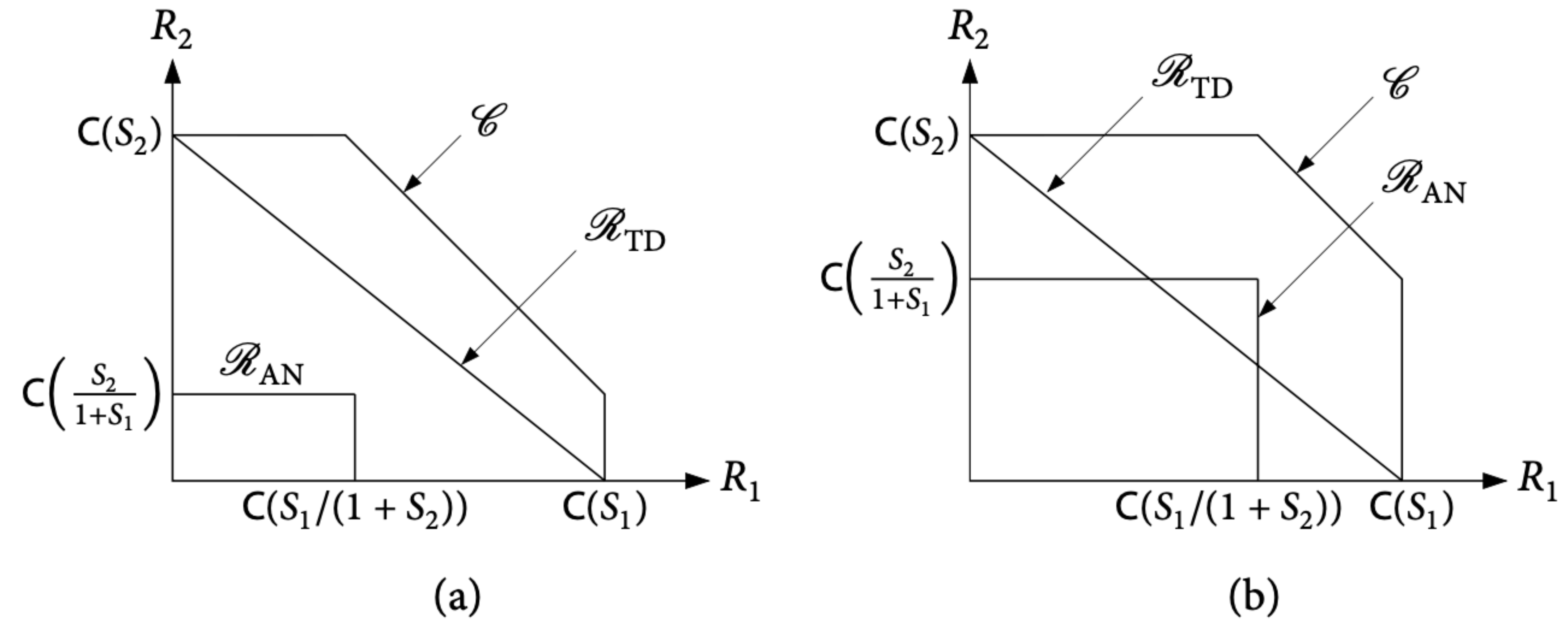


Figure 4.8. Comparison between time division (region $\mathscr{R}_{TD}$) and treating the other codeword as noise (region $\mathscr{R}_{AN}$): (a) high SNR, (b) low SNR.

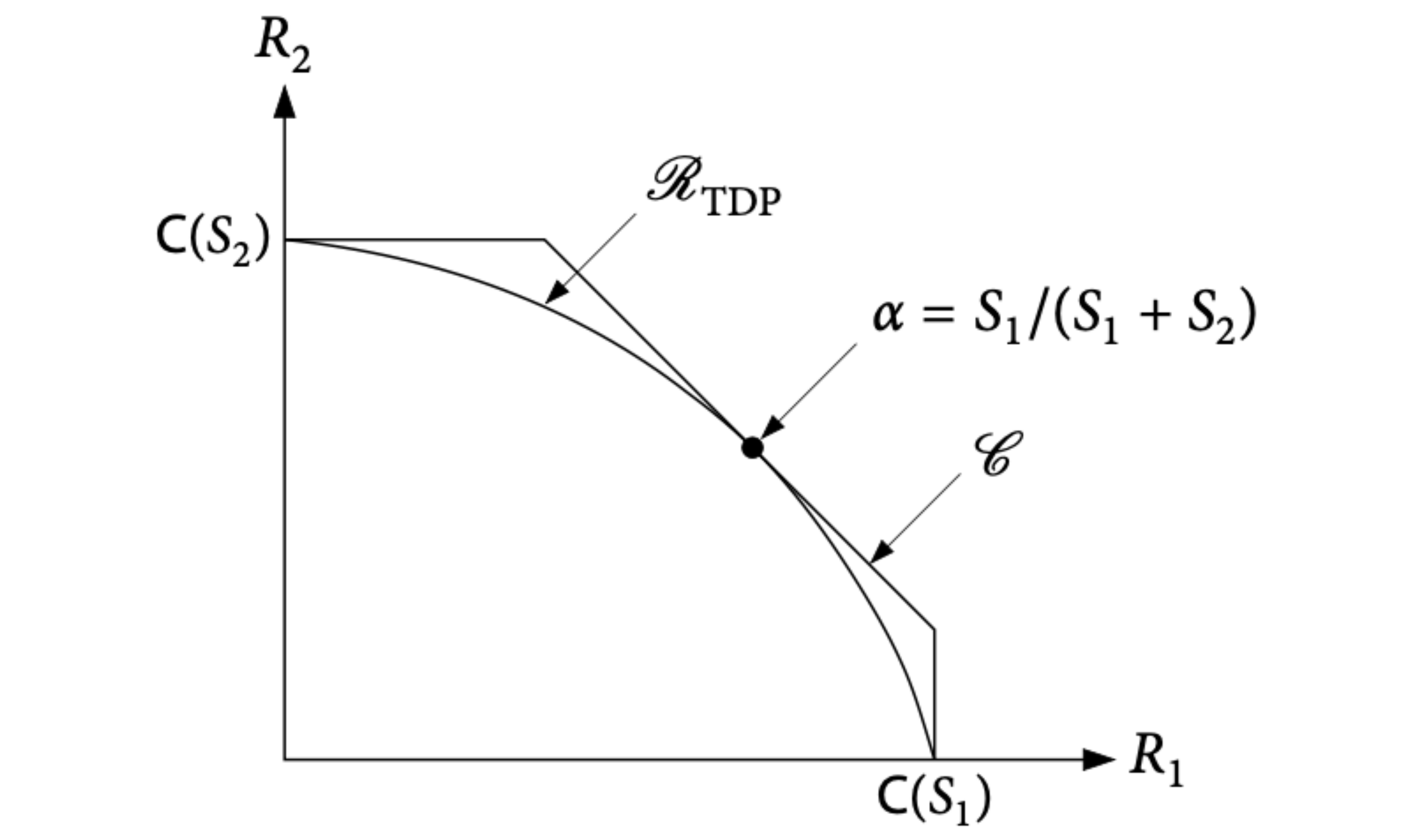
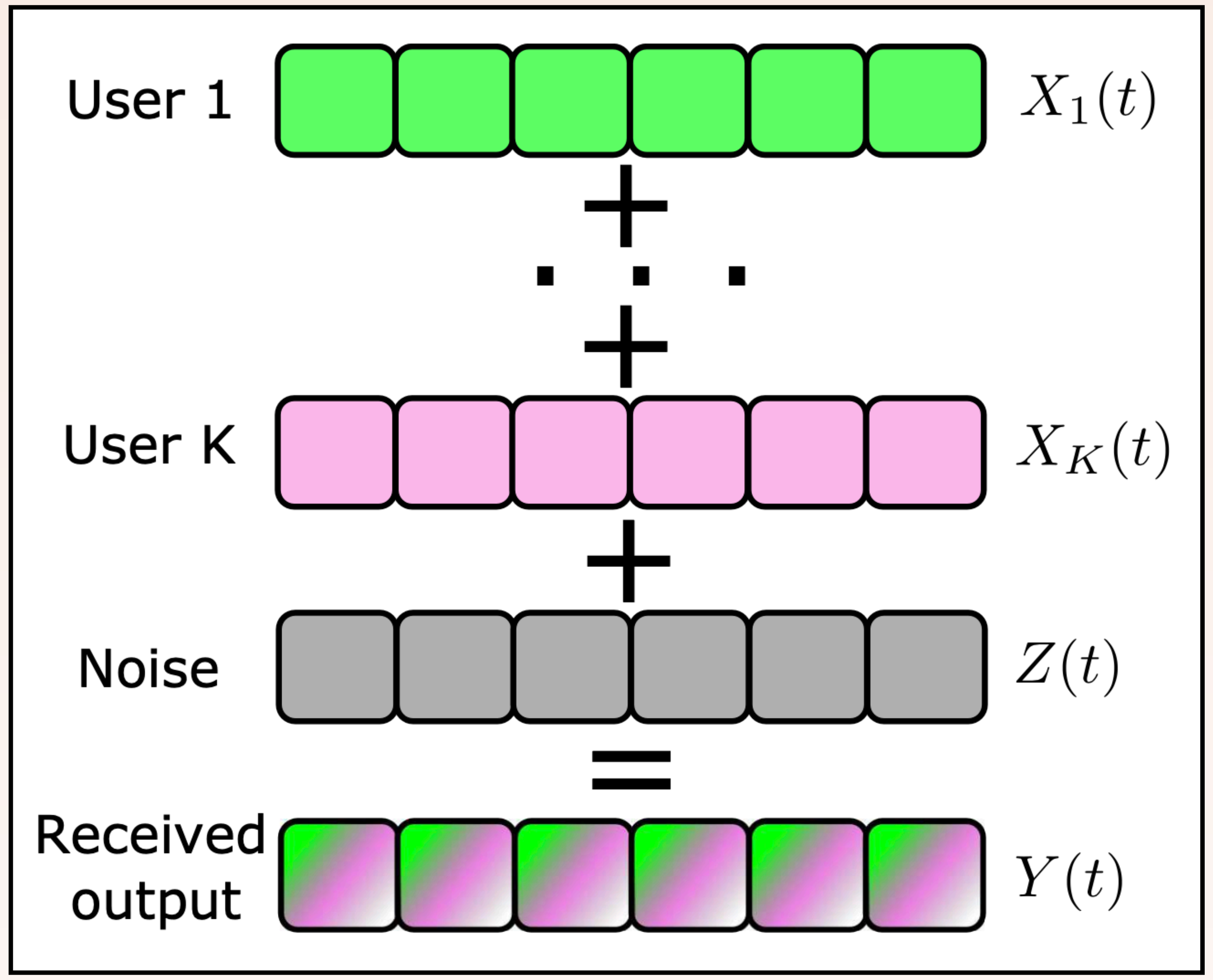


Figure 4.9. Time division with power control (region $\mathscr{R}_{TDP}$).

- ***Takeaways***:
  - To attain (rate-sum) capacity, you need orthogonal schemes (to avoid interference). [See, figure 4.9]
  - However, orthogonal schemes require some form of coordination.
  - *However, massive MAC is envisioned <u>NOT to do coordination</u>, <u>nor to orthogonalize the devices</u>.*

System model (UGMAC code)

- **Users** : $K_a \ll K_{tot}$ active users. Each user $i \in [K_a]$ selects message $W_i \overset{iid}{\sim} \text{Unif}([M])$
- **Encoder $f : [M] \to \mathbb{R}^n$** : maps W_i to a codeword $f(W_i) \in \mathbb{R}^n$. *All users use the same encoder f.*
- **Gaussian multiple access channel:** $Y = \sum_{i=1}^{K_a} f(W_i) + Z$
- **Decoder g:** studies Y to produces a list, namely $g(Y) \subset [M]$, with the hope of $g(Y) \approx \{W_1, \ldots, W_{K_a}\}$



- We say (f, g) is an (n, M, K_a, P, ϵ) UGMAC code if both conditions are satisfied:
  - (Energy constraint) $\forall w \in [M]$: $\|f(w)\|^2 \leq nP$
  - (Per-user probability of error) $\forall i \in [K_a] : \mathbb{P}(W_i \notin g(Y) \text{ or } \boxed{E_{coll}}) \leq \epsilon$
- ***Remarks:***
  - $\boxed{E_{coll}}$ stands for the event of collision. Which has negligible probability $\mathbb{P}(E_{coll}) \leq \binom{K_a}{2}/M$.
  - Actually, one can derive the exact $\mathbb{P}(E_{coll}) = 1 - \frac{\binom{M}{K_a} K_a!}{M^{K_a}}$, which is a birthday problem.

* A side note on the names of such system model

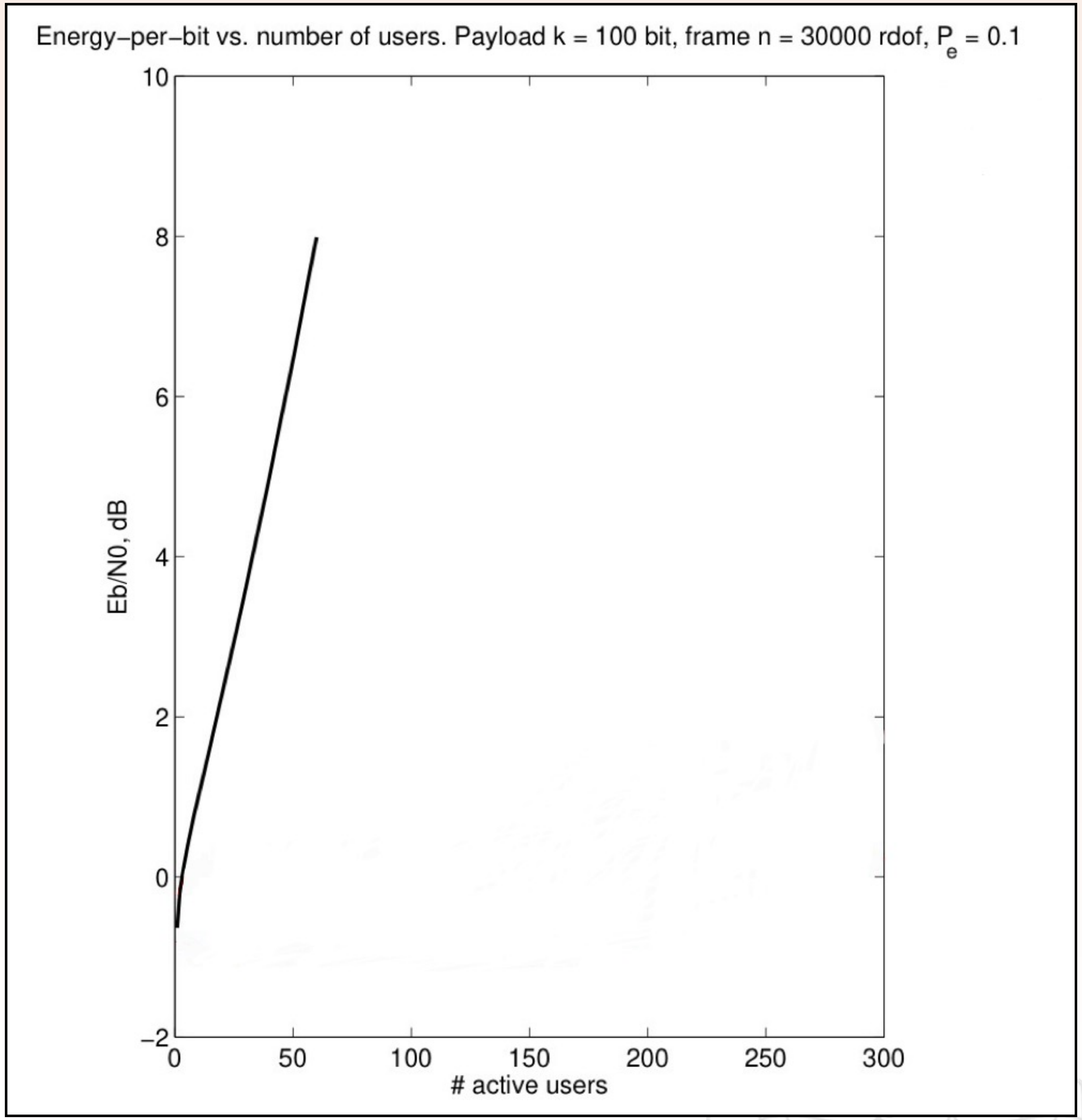
- **Users** : $K_a \ll K_{tot}$ active users. Each user $i \in [K_a]$ selects message $W_i \overset{iid}{\sim} \mathrm{Unif}([M])$
- **Encoder $f : [M] \to \mathbb{R}^n$** : maps W_i to a codeword $f(W_i) \in \mathbb{R}^n$. *All users use the same encoder f.*
- **(Not necessary Gaussian) multiple access channel:** $Y = \mathrm{someChannel}\{f(W_1), \ldots, f(W_{K_a})\}$
- **Decoder g:** studies Y and produces a list, namely $g(Y) \subset [M]$. Hope $g(Y) \approx \{W_1, \ldots, W_{K_a}\}$

- ***What differs from traditional model is: decoder g is not required to know who sends what.***
- In the literature, there are many names for the above system model:
  - "Unsourced random/multiple access" ***Pro:*** Very commonly used name. And it stresses that there is no identity problem; ***Con:*** it is totally fine if the codewords themselves contain identity information.
  - "Massive random/multiple access" ***Pro:*** used by the initial paper; ***Con:*** if $K_a = 10$ it seems not very "massive", but we don't want to exclude those "non-massive" cases.
  - "Uncoordinated random/multiple access" ***Few paper use this name.***

Any candidate scheme? … ALOHA

- x-axis: K_a, number of active users
- y-axis: minimal E_b/N_0 (dB) such that at least 90% of messages are recovered.

- By "ALOHA", what we really mean is an (optimistic) slotted ALOHA.
- Each frame is of length n. We partition the frames into m subframes.
  - When ≥ 2 devices select the same subframe, both signals are lost.
- When number of active users increases, ALOHA curve shots up to sky.
- ***Subframe-collision is a big headache.***



An achievability bound (Polyanskiy 2017)

- For any (n, M, K_a, P, ϵ) and any $P' < P$, there exists a UGMAC code with
- $$\mathrm{PUPE} \le p_0 + \boxed{\sum_{t=1}^{K_a} \frac{t}{K_a} e^{-nE(t)}}$$
- where $p_0 = \dfrac{\binom{K_a}{2}}{M} + K_a \mathbb{P}\left(\sum_{j=1}^{n} Z_j^2 > \dfrac{nP}{P'}\right)$. And $E(t)$ is some complicated expression.

$Z_j \overset{iid}{\sim} N(0,1)$

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

$$\epsilon \le \sum_{t=1}^{K_a} \frac{t}{K_a} \min(p_t, q_t) + p_0\,, \tag{3}$$

where

$$p_0 = \frac{\binom{K_a}{2}}{M} + K_a \mathbb{P}\left[\frac{1}{n}\sum_{j=1}^{n} Z_j^2 > \frac{P}{P'}\right], \tag{4}$$

$$p_t = e^{-nE(t)}, \tag{5}$$

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

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

$$a = \frac{\rho}{2}\log(1 + 2P't\lambda) + \frac{1}{2}\log(1 + 2P't\mu) \tag{6}$$

$$b = \rho\lambda - \frac{\mu}{1 + 2P't\mu}, \quad \mu = \frac{\rho\lambda}{1 + 2P't\lambda} \tag{7}$$

$$\lambda = \frac{P't - 1 + \sqrt{D}}{4(1 + \rho_1\rho)P't}, \tag{8}$$

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

$$R_1 = \frac{1}{n}\log M - \frac{1}{nt}\log(t!) \tag{9}$$

$$R_2 = \frac{1}{n}\log\binom{K_a}{t} \tag{10}$$

$$q_t = \inf_{\gamma} \mathbb{P}[I_t \le \gamma] + \exp\{n(tR_1 + R_2) - \gamma\}$$

Original theorem from the paper :(

Proof (sketch)

- Randomly generate the codebook: $\forall i = 1,2,\ldots,M$, $c_i \sim N(0,P')^{\otimes n}$. Since we assumed $P' < P$. We have high probability such that $\|c_i\|^2 \le nP$.
- Inspect p_0, the first term is an upper bound of $\mathbb{P}(E_{coll})$.
- The second term corresponds to $\mathbb{P}(\exists i \in [K_a] \text{ s.t. } \|c_{W_i}\|^2 > nP)$
- So, for the summation term, we are safe to assume all chosen codewords are <u>distinct</u> and <u>satisfying the power constraint.</u>

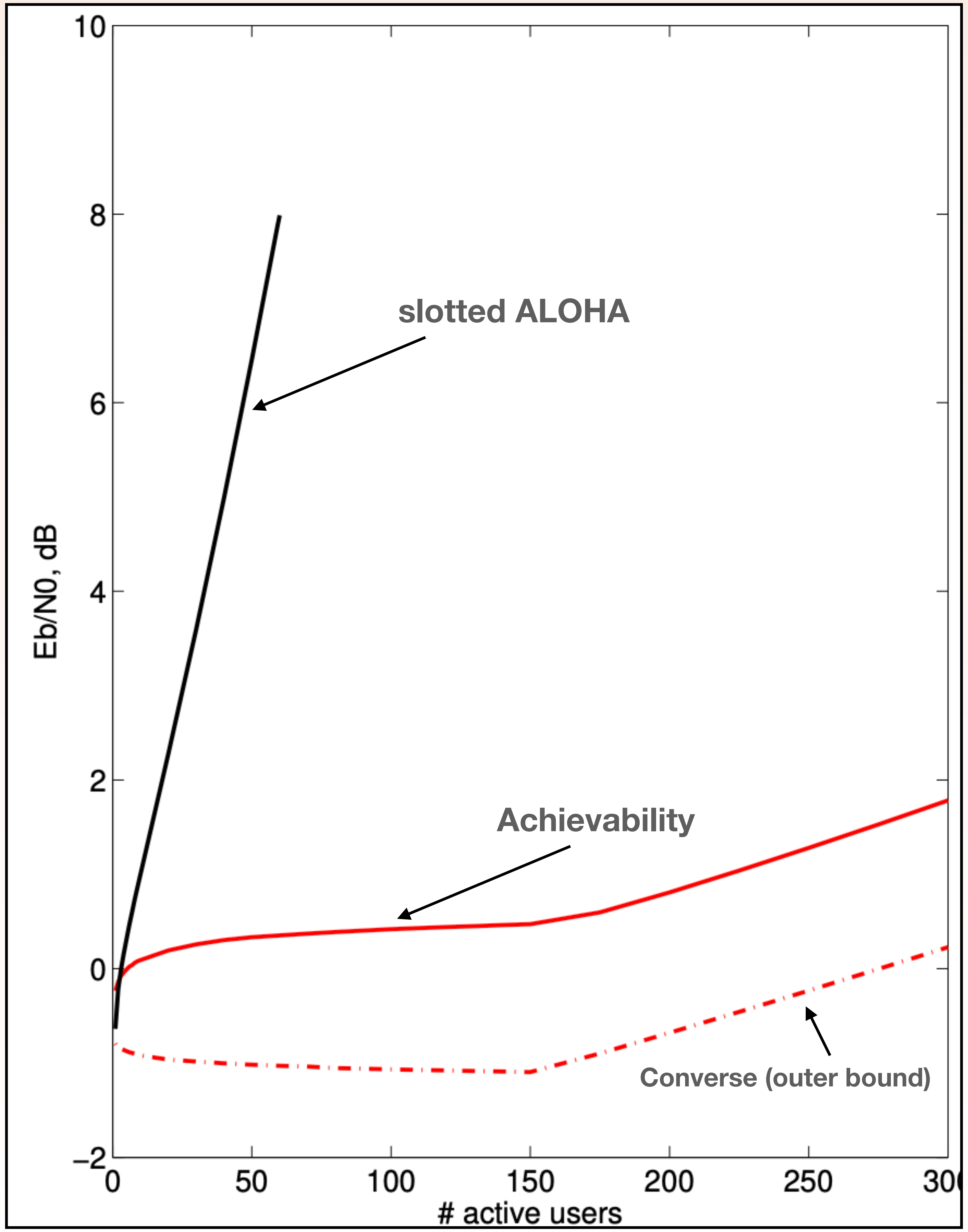
An achievability bound (Polyanskiy 2017)

- By symmetry, we can assume $c_1, c_2, \ldots, c_{K_a}$ are chosen and transmitted.
- Decoder receives $Y \in \mathbb{R}^n$, where $Y = c_1 + c_2 + \ldots + c_{K_a} + Z$.
- The decoder want to recover a subset of codewords $\hat{S} \subset [M]$ with $|\hat{S}| = K_a$. The decoder hopes that $\hat{S} \approx \{c_1, c_2, \ldots, c_{K_a}\} = [K_a]$.
- Define $c(A) = \sum_{i \in A} c_i$, where $A \subset [M]$.
- **(Decoding rule) Decoder estimates $\hat{S} = \arg \min_{S:|S|=K_a} \|c(S) - Y\|$.**
- Once we select the set of suspected messages $\hat{S}$, we can calculate PUPE $= \frac{1}{K_a} |[K_a] \backslash \hat{S}|$
- Define event "$t -$ false" as $\{ |[K_a] \backslash \hat{S}| = t \}$.
- Observe that if $t -$ fasle happens, then PUPE $= t/K_a$

An achievability bound (Polyanskiy 2017)

- Recall that $t-\text{false} \iff \{|[K_a]\backslash\hat{S}| = t\}$.
- Then, $\widehat{\text{PUPE}} \le \sum_{t=1}^{K_a} \frac{t}{K_a}\mathbb{P}(t-\text{false})$. We then only need to bound $\mathbb{P}(t-\text{false}) \le e^{-nE(t)}$
- Note that $t-\text{false} \iff \exists S_0 \subset [K_a]$ (a set of transmitted messages) was erroneously replaced with $S_0' \subset \{K_a+1,\ldots,M\}$ and $|S_0| = |S_0'| = t$.
- ***[Lemma of Gaussian ball]*** If $Z \sim N(0,aI_n)$, then $\mathbb{P}(\|Z+u\| \le v) \le e^{-nE_{ball}(u,v,a)}$, where $E_{ball}(u,v,a)$ is complicated (which is itself an optimisation problem)
- Define event $F(S_0,S_0') = \{\|c(S_0) - c(S_0') + Z\| \le \|Z\|\}$
- $\mathbb{P}(t-\text{false}) \le \mathbb{P}\{\bigcup_{S_0 \in \binom{K_a}{t}} \bigcup_{S_0' \in \binom{M-K_a}{t}} F(S_0,S_0')\}$. Define event $F(S_0) = \cup_{S_0'} F(S_0,S_0')$.
- The idea is to use: $\mathbb{P}(F(S_0,S_0')\,|\,c(S_0),Z) \le e^{-nE_{ball}}$.
- *Marginalising steps …… need to use Gallager's ρ-trick twice and some optimisations, straightforward but not trivial*
- After two-layer tedious marginalising, we arrive at $\mathbb{P}(t-\text{false}) = \mathbb{P}(\cup_{S_0}F(S_0)) \le e^{-nE(t)}$. And we're done.

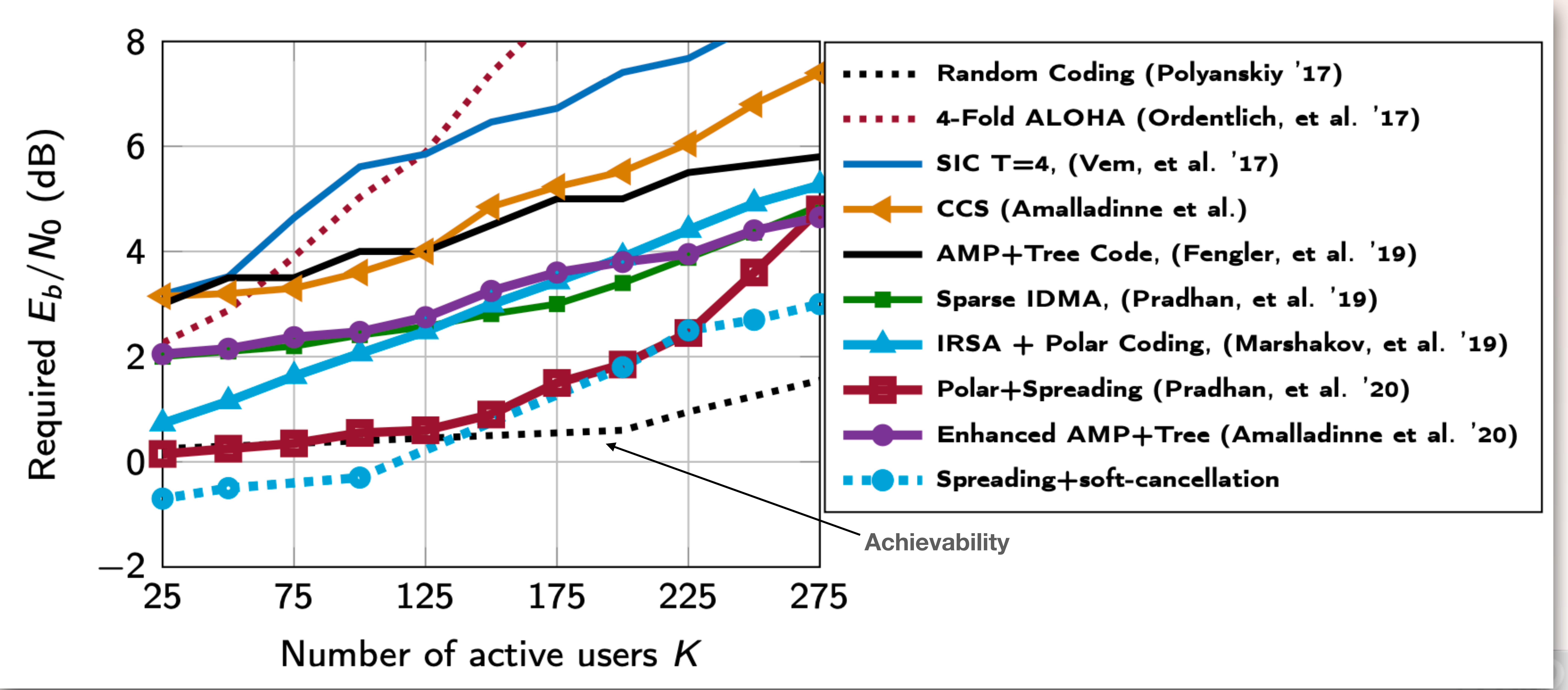
Plotting the achievability bound (for GMAC)



- Payload length (message bits) 100bits $\iff$ $M = 2^{100}$
- Frame length $n = 30000$. Recall $f(W_i) \in \mathbb{R}^n$.
- PUPE tolerance $= 0.1 = 10\,\%$ *(In practice impossible, for theory ok)*

- Remarks:
  - Converse bound: it is impossible for any code to perform below it. *Proof is hard to read, hence omitted.*
  - All three curves are calculated, instead of simulated.
  - The converse bound is actually decreasing from $K_a \in [1,150]$ *(Any intuition?)*
  - Both the achievability and converse indicate that $K_a = 150$ is crucial. Cost goes up later.

Some more curves



- People work on this problem since 2017.
- Amongst the shown curves, some beat the achievability bound we just introduced.

References

- Yury Polyanskiy, "*A perspective on massive random-access,*" 2017 ISIT, Aachen, Germany, 2017, pp. 2523-2527 *[this is the paper I refer to as the basis and the achievability bound part is from this paper]*

- Abbas El Gamal and Young-Han Kim. "*Network Information Theory*" Cambridge University Press, USA, 2012. *[Most screenshots on 2-sender Gaussian MAC are from Chap. 4]*

- K. R. Narayana, J. -F. Chamberland, Y. Polyanskiy, *"Unsourced Multiple Access (UMAC): Information Theory And Coding", 2021 ISIT. [a 3-hour tutorial of ISIT 2021, I covered part of its Part 1 only]*

- Lin Dai, "*Lecture 5, Multiple Access*" (lecture slides) , https://www.ee.cityu.edu.hk/~lindai/6603_Lecture5.pdf *[A very informative and concise powerpoint on Gaussian multiple access channels, as well as Gaussian broadcast channel]*

- V. K. Amalladinne, J. -F. Chamberland and K. R. Narayanan, *"A Coded Compressed Sensing Scheme for Unsourced Multiple Access,"* in *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6509-6533, Oct. 2020 *[Paper for the "CCS" curve in last page. Surprisingly, it has a tutorial video series: https://youtube.com/playlist?list=PLUd5FtcfdZflIMvB9X0gBIyEFG27uWOdt&si=pW3ED4Qgd3TDlBj]*

- GSMA, "*The Mobile Economy 2023*", *[where I found the business data at beginning]*

- ……