STANFORD

Lecture 7

Multiuser Channels

and the Capacity Region

April 23, 2024

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering

Instructor EE379B – Spring 2024

Announcements & Agenda

- Announcements
  - Mid term May 2, in class.
  - Final leaning towards 24-hour take home
    - Send email when you start
    - Send completed test 24 hours later, roughly June 7-10 range
  - PS3 due tomorrow
  - PS4 due Tuesday 4/30 (so solutions can be distributed)

- Problem Set 4 = PS4 (due May 2)
  1. 2.21 Multiuser Channel Types
  2. 2.22 Multiuser Detector Margin
  3. 2.23 Mutual-Information Vector
  4. 2.24 Time-Division Multiplexing region
  5. 2.25 MAC regions

- Agenda
  - Multi-User (MU) Introduction
    - Where used?; What is a multi-user data rate?; order & decodability
  - The 3 basic MU types and the matrix AWGN
  - Rate Bounds and Detection
  - General MU Capacity Region and other optima
  - Back-Ups – not presented
    - Scheduling and Queuing
    - Some useful slides on AWGN labelling

Multiuser (MU) Introduction

(definitions and fundamentals)

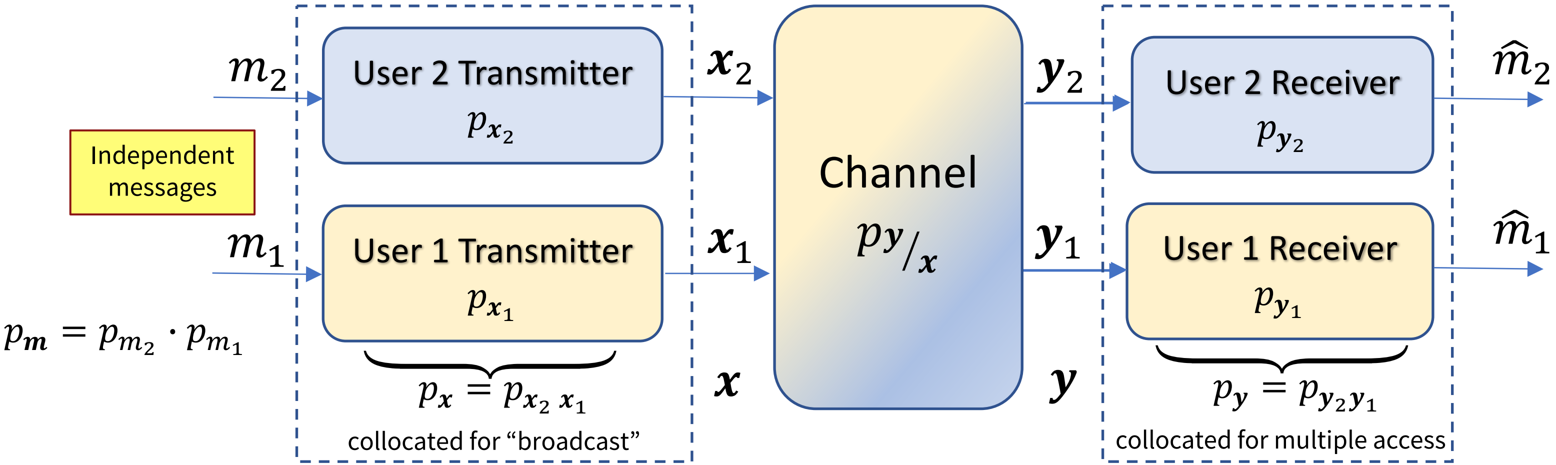
Section 2.6 intro

U>1 users

- Downlink/stream – one to many (“broadcast”)
- Uplink/stream – many to one (“multiple access)
- Relay signals (“mesh”)
- Overlapping combinations (Wi-Fi, or cell, or really all) – “interference”

MU Mathematical Model (Section 2.6)

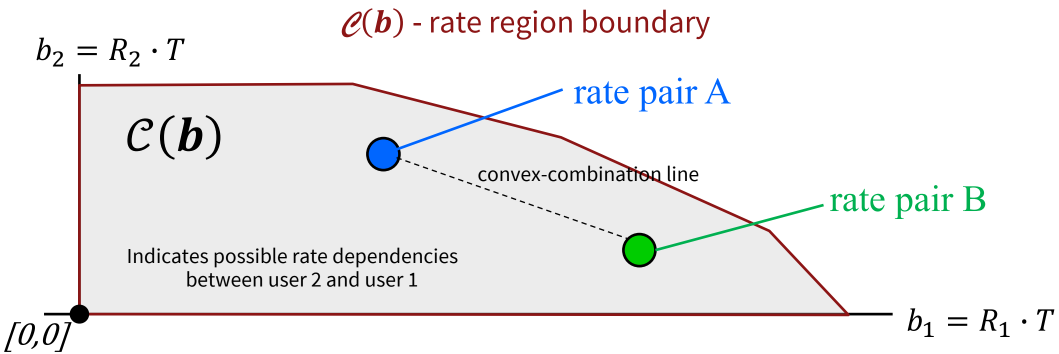
- There is a joint probability distribution $p_{\boldsymbol{xy}}$ that determines all marginals (e.g., input) and conditionals (channel), $p_{\boldsymbol{y}/\boldsymbol{x}} = \frac{p_{\boldsymbol{xy}}}{p_{\boldsymbol{x}}}$.



- User 1 & 2's data rates are mutually dependent (otherwise just two single-user channels).
- $b \rightarrow \boldsymbol{b} = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} = \boldsymbol{R} \cdot T = \begin{bmatrix} R_2 \cdot T \\ R_1 \cdot T \end{bmatrix}$; the bits/sub-symbol becomes a U −dimensional vector, $u = 1, \ldots, U$.
- Single-user is a (degenerate) subset of multiuser.

The Rate Region

- "Reliably decodable" set of users' bits/subsymbol vectors that can be achieved $P_e \to 0$ (AEP).



- All "convex combinations" (on the line connecting points) must trivially be achievable too.
- What is $\mathcal{C}(\boldsymbol{b})$ if two independent single-user channels? "crosstalk free"

rectangle (2), prism (3), Orthotope (U)

- The region is "convex hull" (union) of achievable points over all "allowed" $p_{\boldsymbol{xy}}$, or really over $p_{\boldsymbol{x}}$,
  - because $p_{\boldsymbol{y}/\boldsymbol{x}}$ (the general MU channel description) is given (fixed).

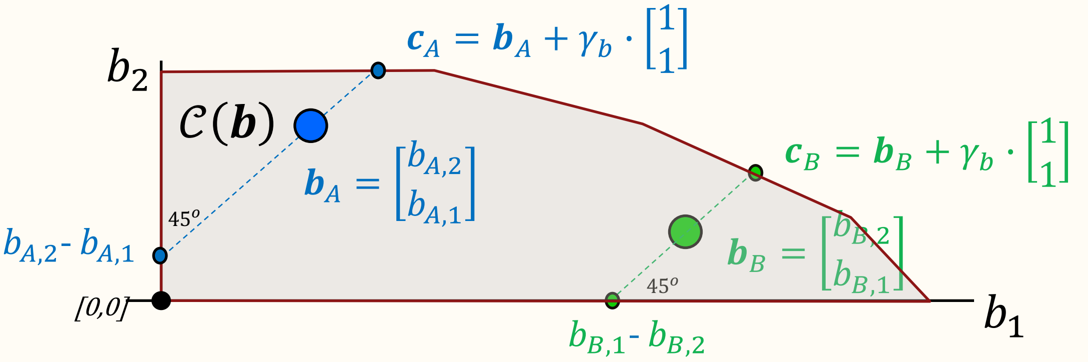
Multiuser Margin

- Single-user (energy) margin $\bar{b} = \frac{1}{2} \cdot \log_2\left(1 + \frac{SNR}{\Gamma \cdot \gamma_m}\right)$ measures safety for $\bar{b}$ if SNR changes.

$$\gamma_m = \frac{1}{\Gamma} \cdot \frac{SNR}{2^{2\bar{b}} - 1}$$

- The **bit gap** is $\gamma_b = \mathcal{C} - b$ where $\bar{\mathcal{C}} = \frac{1}{2} \cdot \log_2(1 + SNR) = \bar{b} + \bar{\gamma}_b$ - so measures rate gap to $\bar{\mathcal{C}}$.
  - $\Gamma \cdot \gamma_m \cong 6 \cdot \bar{\gamma}_b$ dB , $\bar{\gamma}_b = 0$ if the code achieves capacity (6 dB/bit-dimension).

- **Multiuser bit gap** measures to $\boldsymbol{c}_{\boldsymbol{b}'} \in \mathcal{C}(\boldsymbol{b})$, the rate region boundary, so $\gamma_{\boldsymbol{b}'} \cdot \mathbf{1} = \boldsymbol{c}_{\boldsymbol{b}'} - \boldsymbol{b}'$



$$\begin{aligned} b_2 &= b_1 + (b_{A,2} - b_{A,1}) \\ b_1 &= b_2 + (b_{B,1} - b_{B,2}) \end{aligned}$$

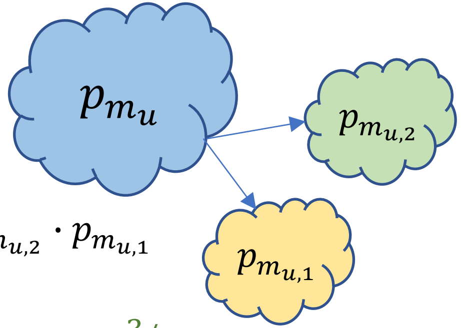
$$umax \triangleq \arg\left(\max_u b'_u\right)$$

$$b_{umax} = \left(\sum_{i \neq umax} b_i\right) + b'_{umax} - \left(\sum_{i \neq umax} b'_i\right)$$

- Multiuser (energy) margin still is then same as single-user margin. $\Gamma \cdot \gamma_m \cong 6 \cdot \bar{\gamma}_b$ dB

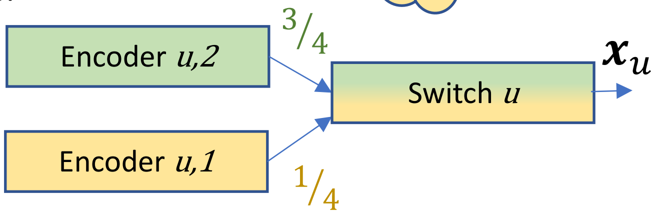
User Components (a.k.a. “time/dimension-sharing”)

- Two independent **user components or subusers** have the
  - same transmitter and same receiver (“different components of same user”).
- These two subusers (codes used) can be separately encoded and decoded.

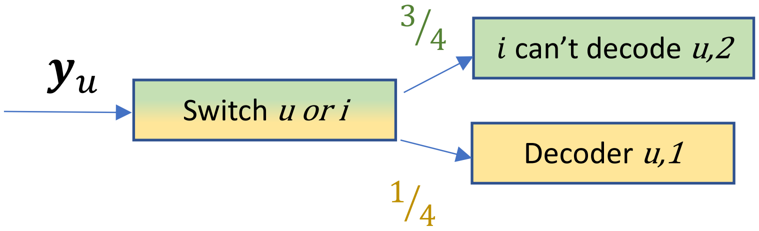


$$p_{m_u} = p_{m_{u,2}} \cdot p_{m_{u,1}}$$

- Bits per symbol is $b_u = b_{u,1} + b_{u,2}$.
- Other users’ receivers $i \neq u$ may decode all, none, or some of these components:
  - which they should do and remove if possible, or
  - otherwise they are averaged in marginal (remains as noise when Gaussian).



- The two subusers may simultaneously share dimensions, apportioning fractional information (or energy when Gaussian) to each.
- U can increase to $U + 1$, or more generally to $U \leq U' \leq U^2$ components.
- $\mathcal{C}(\boldsymbol{b})$, and b, can also expand to U' dimensions:
  - Original $\mathcal{C}(\boldsymbol{b})$ adds together the sub-users’ dimensional rates,
  - and thus decreases its dimensionality.



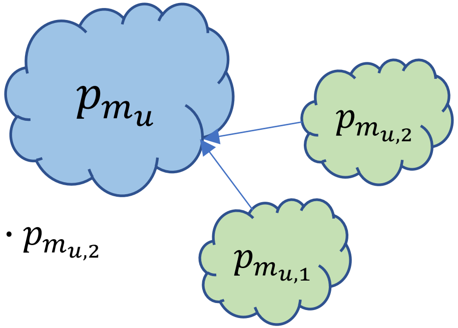
- Some information theorists call this “**time-sharing**,”
  - but user components is more accurate and general, and extrapolates to all types of dimensions and combinations.

Macro Users

- Two users (or user components) that have identical impact/influence create **a macro user.**
  - $p_{\ldots x_u \ldots . x_i \; y} = p_{\ldots x_i \ldots . x_u \; y}$ - interchange of the users does not change the joint probability distribution.
  - These two could be considered one macro user, where any partition of this macro user's rate to the two original users is feasible.

- Simple example is $y = x_1 + x_2 + n$, where both users 1 and 2 share the same energy.
  - This is ~ single-user channel with macro user $x = x_1 + x_2$, for which any division of $b = b_1 + b_2$ is possible.

- This can simplify some capacity-region construction.



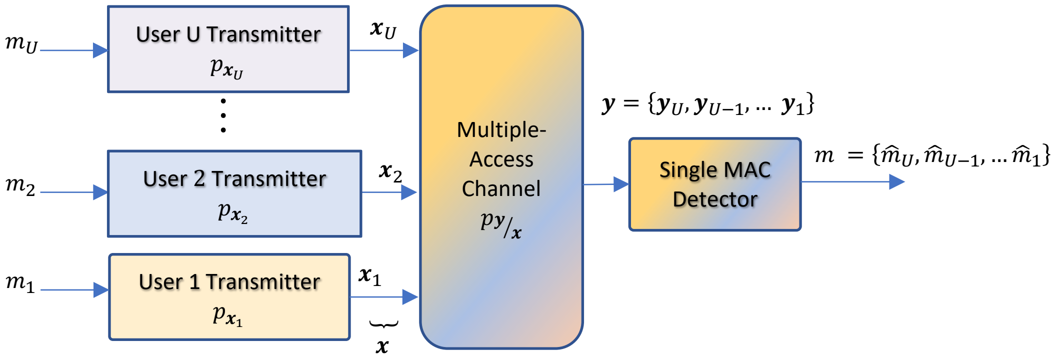
$$p_{m_u} = p_{m_{u,1}} \cdot p_{m_{u,2}}$$

The 3 Basic MUs & Matrix AWGN

PS4.1 - 2.21 Multiuser Channel Types

Section 2.6.1

Multiple Access Channel (MAC)



Sec 2.6.1 and 2.7

- User transmitters are in different locations (cannot coordinate to encode/modulate $\boldsymbol{x}$).
  - All use a good single-user code (see 379A, Chapter 8).

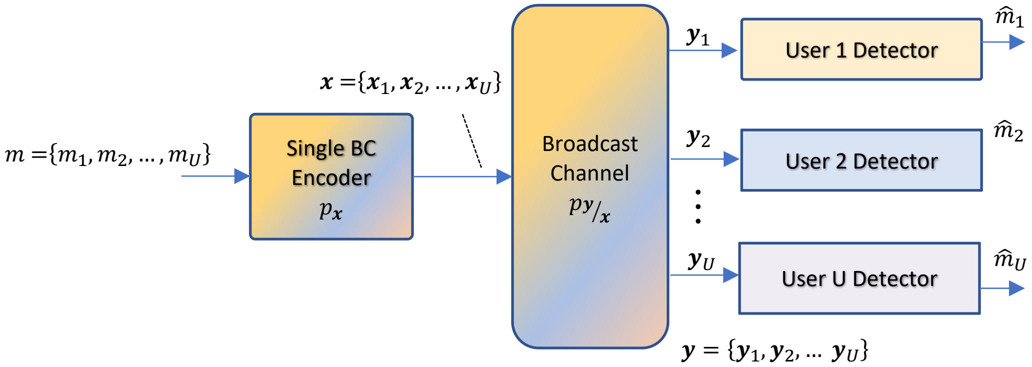
- Single receiver detects all users and:
  - separates the users,
  - reliably decodes, $P_e \to 0$, by decoding and removing some (none or all) other users first, which
  - suggests "user **order**" $\boldsymbol{\pi}$ (vector "priority") is important (decode $\boldsymbol{\pi}$'s 1st/bottom element first, … U … last at top).
  - If subusers, then up to U' subusers might be decoded, where again $U \leq U' \leq U^2$.

- **Order is fundamental** to best MU design – the MAC has $U'!$ possible orders (each has a $\boldsymbol{b}$)
  - and all potential convex combinations thereof.
  - There is also an input $p_{\boldsymbol{x}}$ choice (or code choice), and all potential convex combinations thereof.
  - For MAC, there will be ways to simplify so that $U' = U$.

Design

1. $\boldsymbol{\pi}$
2. $p_{\boldsymbol{x}}$

Broadcast Channel (BC)



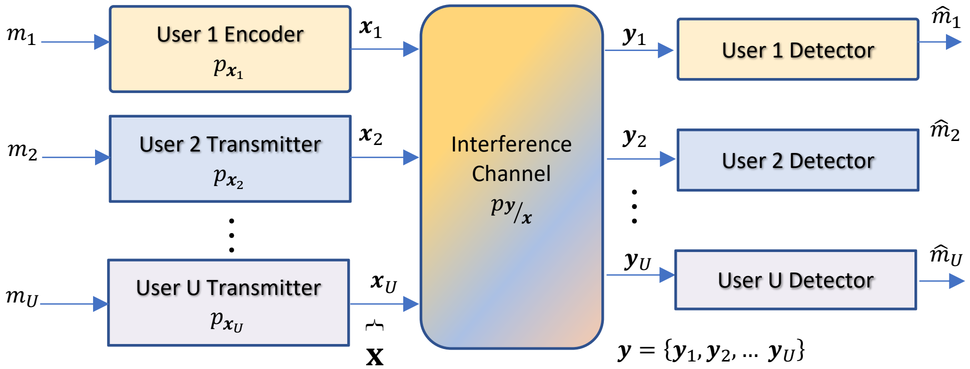
Sec 2.6.1 and 2.8

- The BC is the **"dual"** of special type of MAC.
  - This eventually allows common design method.
- Receivers are in different places, so they cannot "co-process" $\boldsymbol{y}$'s user outputs.
- Transmitter can co-encode/generate $\boldsymbol{x}$, although input messages remain independent.
  - Who encodes first? (may be at disadvantage)
  - Who encodes last? (knowing other users' signals is an advantage)
  - What then is the **order**?

Design

1. $\boldsymbol{\pi}$
2. $p_{\boldsymbol{x}}$

Interference Channel (IC)



Sec 2.6.1 and 2.9

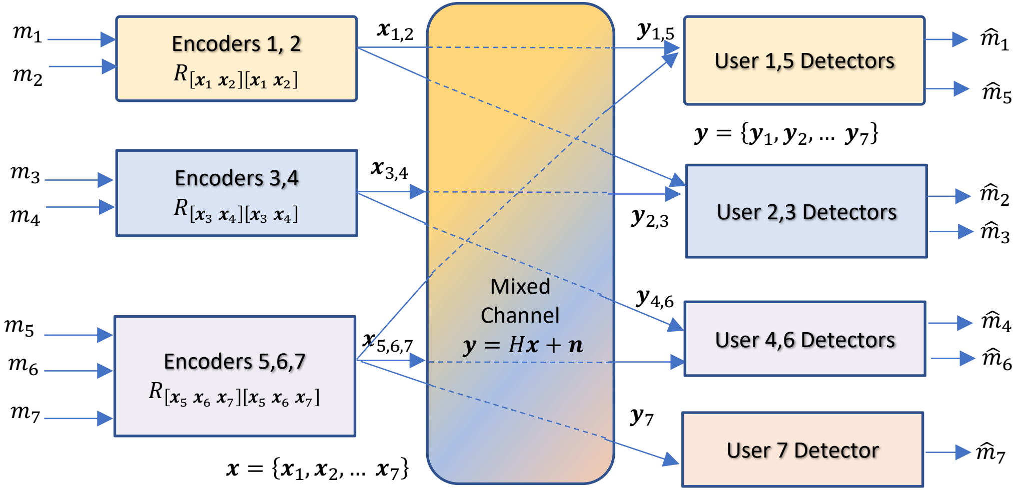
- Users' transmitters, and also receivers, are in different locations.
  - No co-encoding of user messages nor coordinated reception is possible.
  - Views: A set of MACs with same inputs or a set of BCs with same outputs.
- Each receiver can use a decoding **order** to detect others first, if that is possible.
  - The rcvr treats other users as noise if not possible to decode/remove first.
  - Each receiver's order is column of **matrix order Π**.
- There are $(U'!)^U$ possible IC **orders:** ... $U'!$ at each receiver, with $U' \le U^2$.
  - Each user may have a subuser component for every user's receiver to detect, large **FINITE.**

Design

1. Π
2. p_x

Other MU Types / Combinations

- Mixed



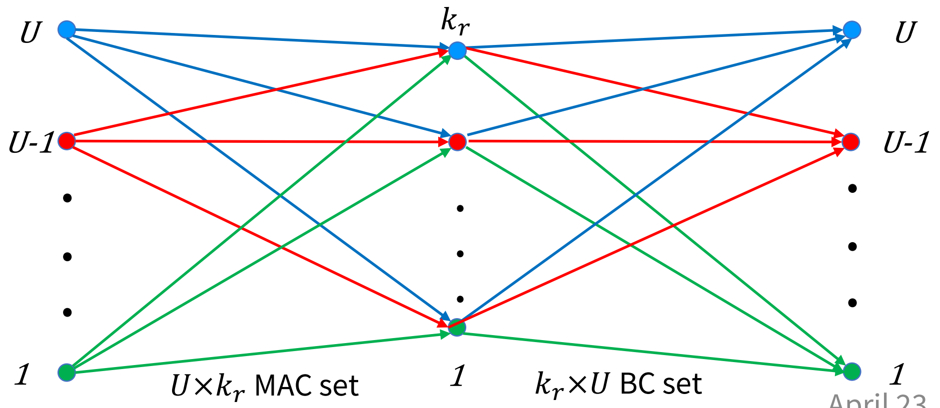
Different Macro views:

IC of 4 MAC macros:
{1,5}, {2,3}, {4,5}
{7} is single user.

IC of 3 BC macros:
{1,2}, {3,4}, {5,6,7}.

Macro design can shrink $\mathcal{C}(\boldsymbol{b})$ dimensionality.

- Mesh/Relay



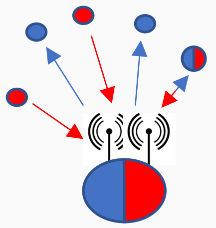
k_r **relays**

- Single User

Nested MU Channel Examples

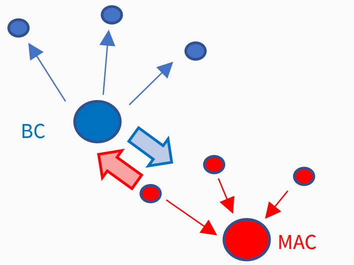
- Each **nested MU** channel → 1 macro user.

- These macro users crosstalk into each other.
  - Some users with macro group may decode
  - with any given order in that group.
  - Those not decoded are undecodable "noise."



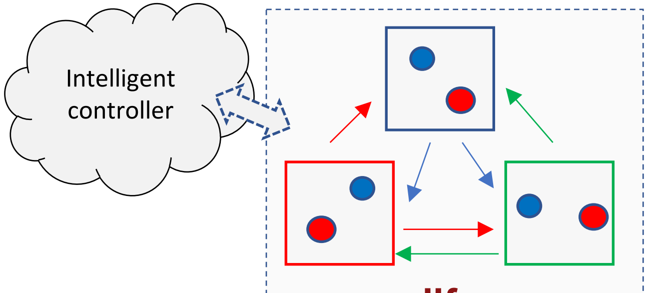
a) radio node edge (**cellular base station** or **Wi-Fi Access point**)

~



b) two nested MUs: **down** (BC) and **up** (MAC) with IC network between them

- Design treats as IC of macro users:
  - orders the macro's subusers.
  - Rcvrs decode all subusers within the local macro group
  - or use "multi-level" waterfilling (end of 379B).



c) three nested IC: (BC, MAC) like those in b), nested into 3x3 IC network

Rate Bounds & Detection

PS4.3 - 2.23 Mutual-Information Vector

Chain-Rule Reminder/Review

$$\mathbb{I}(\boldsymbol{x};\boldsymbol{y}) = \sum_{n=1}^{N} \mathbb{I}(\boldsymbol{x}_n;\boldsymbol{y}\,/[\boldsymbol{x}_{n-1}\ \ \cdots\ \ \boldsymbol{x}_1])$$

Lemma 2.3.4

- Think of the input components $\boldsymbol{x}_n$ as users, so $U \to N$ and $u \to n$ (may replace U with U' in general).
- Any receiver output (or combination of them), $\boldsymbol{y}$, has chain-rule decomposition(s); for the given $p_{\boldsymbol{xy}}$, this $\mathbb{I}(\boldsymbol{x};\boldsymbol{y})$ represents a maximum (sum-user) data rate by AEP.
- Each sum term has similar interpretation, given the “previously decoded” (given) other users.
- The capacity region points must correspond to chain-rule $\mathbb{I}$ terms in $\mathcal{C}(\boldsymbol{b})$ for each user receiver in that point’s construction.
- User **decoding order** characterizes the different “chain-rule” compositions.

Some data rate bounds

- **Sum-Rate bound:** $b = \sum_{u=1}^{U} b_u \leq \mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$ - full transmit/receiver coordination is vector coding.

- **Average User u bound:** $\mathcal{I}_u(x_u; \boldsymbol{y}) \leq \mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$- this does NOT bound b_u (could remove other user(s) first)
  - $\mathcal{I}_u(x_u; \boldsymbol{y})$ treats **all** other users as ``noise.''

- The **conditional mutual information** ~first decodes the conditioning users' messages correctly (reliably) and then removes them from the detection process.
  - For AWGN, this operation corresponds to remodulating, filtering by the known channel, and subtracting the result from receiver u 's signal.
  - There are ways to simplify this prior-user removal process.

- All chain-rule bounds apply also if $U \rightarrow U' = U^2$.

Fundamental: User Priority → "order"

- Receiver u decodes who first?, last? $\boldsymbol{\pi}_u = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ or …. (U'! Choices)
- Why is this important?
  - The as-yet un-decoded users are "noise" (averaged to compute marginal dist'n $p_{\boldsymbol{y}_u/\boldsymbol{x}_i}$, upon which ML detector is based).
  - Are the other users reliably decodable? (They must be treated as "noise" if not.)

If others decoded first, then it is successive decoding or "generalized decision feedback," for some order π_u.

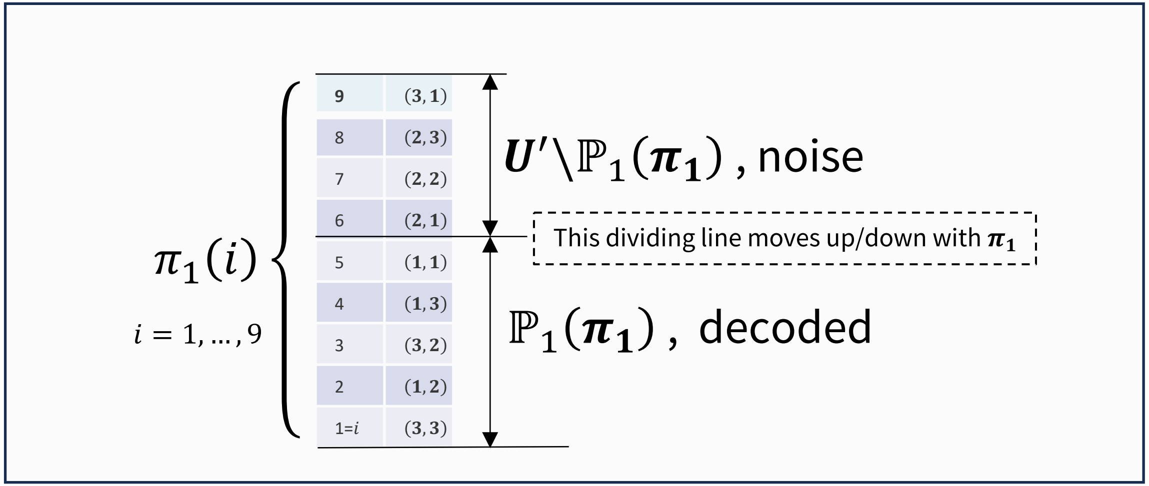
- Same for all users – so there is a "global" order possibility: $\boldsymbol{\Pi} = [\boldsymbol{\pi}_U \quad \cdots \quad \boldsymbol{\pi}_1]$ with $(U'!)^U$ choices
- The designer might check all $\boldsymbol{\Pi}$, and then take convex combinations for each and every allowed $p_{\boldsymbol{x}}$.
  - It simplifies in many situations (including MAC, BC, and sometimes IC).
- **Order vector** and **inverse:**
  - Any permutation vector has inverse.
  - Same as 379A interleave, different use.

$$\boldsymbol{\pi}_u = \begin{bmatrix} \pi(U') \\ \vdots \\ \pi(1) \end{bmatrix} \quad \boldsymbol{\pi}_u^{-1} = \begin{bmatrix} U' \\ \vdots \\ 1 \end{bmatrix} \qquad j = \pi(i) \rightarrow i = \pi^{-1}(j)$$

Prior-User Set

- U^2 subusers subdivide into ordered pairs (u, u') where $u = 1, \ldots, U$ and $u' = 1, \ldots, U$.
  - Receiver u has components from users $u' = 1, \ldots, U$.

- In order $\pi(u, u')$, when receiver u has decoded all $u' = 1, \ldots, U$ (all its subuser components), it is done.
  - Any higher-in-order user components are noise (marginals).
  - Can reindex order so that $\pi(u, u') = \pi(i), i = 1, \ldots, U^2$ where i counts from bottom up.
  - The prior-user set for this order $\mathbb{P}_u(\boldsymbol{\pi})$ occurs for smallest i that contains $\{(u,1),(u,2),\ldots,(u,U)\} \subseteq \mathbb{P}_u(\boldsymbol{\pi})$.
- Example for $U = 3$:



More Formal Prior-User Set

- Define the receiver-indexed sets
  - $S_1 \triangleq \{(1,1), \dots, (1,U)\}$
  - $S_2 \triangleq \{(2,1), \dots, (2,U)\}$
  - $\vdots$
  - $S_U \triangleq \{(U,1), \dots, (U,U)\}$
  - $S = S_1 \cup S_2 \cup \dots \cup S_U$.

$$i_u^*(\boldsymbol{\pi}_u) \triangleq \arg\min_j [\pi_u(S_u) \subseteq \{1{:}j\}]$$

- Prior-User Set is $P_u\left(\pi_{i_u^*}\right) \triangleq \{(u,i) | \pi(u,i) \leq i_u^*(\boldsymbol{\pi}_u)\}$.
- For given order and input p_x, find each user's worst data rate.
- Example – let's say this particular order $\boldsymbol{\Pi}$ just happens to have the 4 subusers/user aligned successively at each receiver,
  - which simplifies illustration from 16 to 4.

rcvr/ User i	$\boldsymbol{\pi}_4(i)$	$\boldsymbol{\pi}_3(i)$	$\boldsymbol{\pi}_2(i)$	$\boldsymbol{\pi}_1(i)$
$i = 4$	3	3	4	3
$i = 3$	4	2	3	2
$i = 2$	1	4	2	1
$i = 1$	2	1	1	4
$\mathbb{P}_u(\boldsymbol{\pi}_u)$	$\{1,2\}$	$\{2,4,1\}$	$\{1\}$	$\{4\}$

$$\boldsymbol{\Pi} = \begin{bmatrix} 3 & 3 & 4 & 3 \\ 4 & 2 & 3 & 2 \\ 1 & 4 & 2 & 1 \\ 2 & 1 & 1 & 4 \end{bmatrix}$$

- Designer lists data-rate entries (mutual information bounds really) for those users who are decoded, and ∞ for those treated as noise.

$\mathfrak{I}$	$\mathfrak{I}_4$	$\mathfrak{I}_3$	$\mathfrak{I}_2$	$\mathfrak{I}_1$
top	∞	$\mathcal{I}_3(3/1,2,4)$ 20	∞	∞
	$\mathcal{I}_4(4/1,2)$ 10	$\mathcal{I}_3(2/1,4)$ 9	∞	∞
	$\mathcal{I}_4(1/2)$ 5	$\mathcal{I}_3(4/1)$ 4	$\mathcal{I}_2(2/1)$ 4	$\mathcal{I}_1(1/4)$ 2
bottom	$\mathcal{I}_4(2)$ 1	$\mathcal{I}_3(1)$ 2	$\mathcal{I}_2(1)$ 2	$\mathcal{I}_1(4)$ 5

$$\mathcal{I}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \begin{bmatrix} 4 \\ 20 \\ 1 \\ 2 \end{bmatrix}$$

The minimum Mutual – Info Vector $\mathcal{I}_{min}$

- **Mutual-information-like** quantity
  - follows the prior-user set.

$$\mathscr{I}_u\left(\boldsymbol{x}_{\pi_u(i)};\boldsymbol{y}_u/\mathbb{P}_{\pi_u(i)}(\boldsymbol{\pi}_u)\right) \triangleq \begin{cases} \infty & i > \pi_u^{-1}(u) \\ \mathcal{I}_u\left(\boldsymbol{x}_{\pi_u(i)};\boldsymbol{y}_u/\mathbb{P}_{\pi_u(i)}(\boldsymbol{\pi}_u)\right) & i \le \pi_u^{-1}(u) \end{cases}$$

- A worst rate for each and every user $\mathcal{I}_{min,u}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}})$ to compare to the user's implemented rate b_u.

$$\mathcal{I}_{min,u}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \min_{i\in\{1,\ldots,U\}} \left\{\mathscr{I}_i\left(\boldsymbol{x}_u;\boldsymbol{y}_i/\mathbb{P}_u(\boldsymbol{\pi}_i)\right)\right\}$$

- This tacitly implies sum over each user's subuser components.

Lemma 2.6.1 [Best Decodable Set] *When good codes (with $\Gamma = 0$ dB), given $\boldsymbol{\Pi}$ and $p_{\boldsymbol{xy}}$, and with*

$$\boldsymbol{b} \preceq \mathcal{I}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \quad , \tag{2.234}$$

then

$$\mathbb{P}_u(\boldsymbol{\pi}_u) \subseteq \mathscr{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b}) \tag{2.235}$$

and receiver u reliably achieves the data rate $b = \mathcal{I}_u(\boldsymbol{x}_u;\boldsymbol{y}_u/\mathbb{P}_u(\boldsymbol{\pi}_u))$ with order $\boldsymbol{\pi}_u$.

The $\mathcal{I}_{min}$ vector

$$\mathcal{I}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \begin{bmatrix} \mathcal{I}_{min,U}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \\ \vdots \\ \mathcal{I}_{min,u}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \\ \vdots \\ \mathcal{I}_{min,1}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \end{bmatrix}$$

$$\boldsymbol{b} \preceq \mathcal{I}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}})$$

Example: sum of 3 users (MAC)

$$y = x_1 + x_2 + x_3 + n$$

real subsymbols

Order $\boldsymbol{\Pi}$	b_1	b_2	b_3
$[1\ 2\ 3]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\mathcal{E}_1+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\mathcal{E}_1+\mathcal{E}_2+\sigma^2}\right)}{2}$
$[1\ 3\ 2]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\mathcal{E}_1+\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\mathcal{E}_1+\sigma^2}\right)}{2}$
$[3\ 1\ 2]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\mathcal{E}_1+\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\sigma^2}\right)}{2}$
$[2\ 3\ 1]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\mathcal{E}_2+\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\mathcal{E}_2+\sigma^2}\right)}{2}$
$[2\ 1\ 3]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\mathcal{E}_2+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\mathcal{E}_1+\mathcal{E}_2+\sigma^2}\right)}{2}$
$[3\ 2\ 1]^*$	$\frac{\log_2\left(1+\frac{\mathcal{E}_1}{\mathcal{E}_2+\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_2}{\mathcal{E}_3+\sigma^2}\right)}{2}$	$\frac{\log_2\left(1+\frac{\mathcal{E}_3}{\mathcal{E}_2+\mathcal{E}_3+\sigma^2}\right)}{2}$

Position in order determines whether other signals are noise or pre-decoded and then pre-subtracted.

- With the MAC's one receiver:
  - the $\boldsymbol{\pi}_u \equiv \boldsymbol{\pi}$ vectors,
  - $U' = U$,
  - $U! = 6$ for this example, so there are only 6 orders to consider.
- There are many other situations that simplify also.

If energies are $\sigma_n^2 = .001$, $\mathcal{E}_1 = 3.072$, $\mathcal{E}_2 = 1.008$, and $\mathcal{E}_3 = .015$, then with Gaussian codes ($p_{\boldsymbol{x}}$ Gaussian) the order $[123]^*$ corresponds to $b_1 = 1$, $b_2 = 3$, and $b_3 = 2$.

Optimum Detectors (2.6.3)

- Section 2.6.3 formalizes (general, including non-Gaussian, case) optimum detection.
- There are various integrals/sums and definitions.
- More simply, each user's optimum detector -- for any given order $\boldsymbol{\pi}$ – must:
  - first (asymptotically/reliably) detect all other "earlier" users (MMSE→MAP, chain rule) reliably ("no errors"), and
  - then consider all "later" users as noise (this generalizes to integration over margin distribution on non AWGNs).
- Each such detector considers all earlier users as given (which means they can be cancelled in Gaussian case with no further detection effect).
- The error-probability calculation follows single-user, simply with any "pre-users" no longer present and any "post-users" averaged (treated like noise).
- Thus, these are the same 379A decoders – just more complex notation for the multiuser case,
  - with some modulation-level preprocessing.

Multi-User Detection (MUD) – 2.6.2

- Optimum remains $\max_{\hat{x}_u} \left\{ p_{x_u/y} \right\}$ where the y is the receiver input for detection (MAP detection).

$$P_e(u) = 1 - P_c(u) = 1 - \sum_{i=1}^{M_u} P_{c/i}(u) \cdot p_i(u)$$

$P_{c/i}$ is the probability for message i
averaged over all the possible y's for which i is selected
(Decision Region).

- But the receiver now might estimate another user earlier (order), so P_e becomes order dependent.

- More generally for any given $p_{\boldsymbol{xy}}$:
  - All **decodable users**, $\mathfrak{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b})$, are first detected and then "cancelled" – they contribute no "noise" (earlier in order). Abbreviate here $\mathfrak{D}_u(\boldsymbol{\Pi})$, but remains a function of all 3 $(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b})$.
  - Other users, $\overline{\mathfrak{D}}_{u'}(\boldsymbol{\Pi})\backslash u$ are not first detected and are "averaged" (treated as noise).

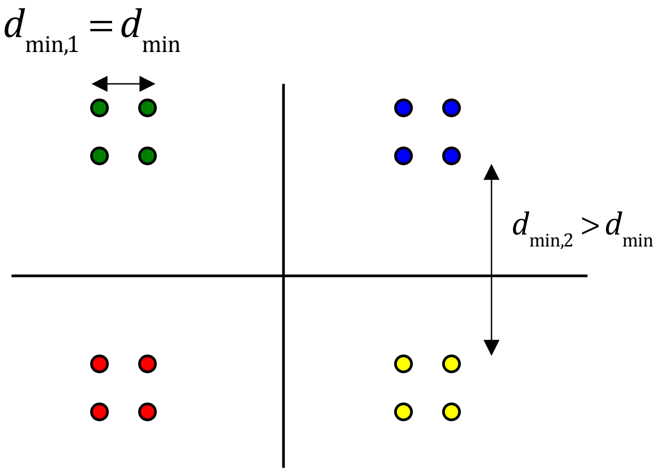
$$p_{\boldsymbol{x}_u/[\boldsymbol{y}\ \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}]}(\boldsymbol{\chi}_u, \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y}) = \int_{\boldsymbol{\chi}\in\boldsymbol{x}_{\{\overline{\mathscr{D}}_u(\boldsymbol{\Pi})\backslash u\}}} p_{\boldsymbol{x}/[\boldsymbol{y}\ \boldsymbol{x}_{\{i\in\mathscr{D}_u(\boldsymbol{\Pi})\}}]}(\boldsymbol{\chi}, \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y}) \cdot d\boldsymbol{\chi}$$

Integration/sum
is over the noise ave

$$p_{\boldsymbol{x}/[\boldsymbol{y}\ \boldsymbol{x}_{i\in\{\mathscr{D}_u(\boldsymbol{\Pi})\}}]}(\boldsymbol{\chi}_u, \boldsymbol{\chi}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y}) = \frac{p_{\boldsymbol{y}/\boldsymbol{x}}(\boldsymbol{\chi}_{i\in\overline{\mathscr{D}}_u(\boldsymbol{\Pi})}, \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y}) \cdot p_{\boldsymbol{x}}(\boldsymbol{\chi}_{i\in\overline{\mathscr{D}}_u(\boldsymbol{\Pi})})}{p_{\boldsymbol{y}/\boldsymbol{x}_{\{i\in\mathscr{D}_u(\boldsymbol{\Pi})\}}}(\boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y})}$$

Term inside integral
Derives from $p_{\boldsymbol{xy}}$.

Simple Example



Design can (optimally) reuse all the single-user good codes !!!

- The decoder should decode first red, green, blue, yellow; this treats the variation within each color as "noise."
- Then the decoder re-centers the constellation and decides further which of the 4 same-color points.
  - This effectively cancels the noise from the first step.
- Yes, an overall decoder performs the same if earlier decisions are correct, but the basic concept expands.
  - Again, MMSE (which is chain rule) is optimum detector if previous users (asymptotically reliable – no errors) are correct.

General MU Capacity Region and related optima

Section 2.6.4

3 General Search Steps

- Search 1: Find $\mathcal{I}_{\boldsymbol{min}}$ for given $\boldsymbol{\Pi}$ and $p_{\boldsymbol{xy}}$
- Search 2: Generate these $\mathcal{I}_{\boldsymbol{min}}$'s convex hull over all orders $\boldsymbol{\Pi}$ for the achievable region $\mathcal{A}(\boldsymbol{b}\ , p_{\boldsymbol{xy}})$
- Search 3: Generate a 2nd Convex hull over all probability distributions $p_{\boldsymbol{x}}$ for $\mathcal{C}(\boldsymbol{b}\ )$
- These searches can be complex for general case, but do simplify for Gaussian MAC, BC, and IC.

Order-and-Distribution-Dependent Region

- **Order Step** forms a first convex hull of all $\mathcal{I}_{min}$ vectors FOR EACH GIVEN ORDER and input distribution.

$$\mathcal{A}(\boldsymbol{b}\,, p_{xy}) = \bigcup_{\Pi}^{conv} \mathcal{I}_{min}(\boldsymbol{\Pi}, p_{xy})$$

Achievable Region

- Any point outside $\mathcal{A}(\boldsymbol{b}, p_x)$ will in the AEP sense have large error probability for at least one receiver.
  - The orders are "dimension shared" across different designs (the convex hull / union) operation .... sub users.
  - Every order and all convex combinations thereof have been considered, so it it could have been decoded it was inside $\mathcal{A}(\boldsymbol{b}, p_x)$.

- **Distribution Step** forms hull over the allowed input distributions (a 2nd convex hull operation).

$$\mathcal{C}(\boldsymbol{b}\,) = \bigcup_{p_x}^{conv} \mathcal{A}(\boldsymbol{b}\,, p_{xy})$$

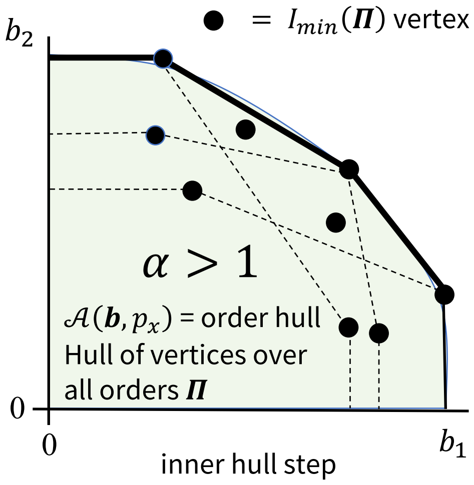
MU Capacity Region

- The order search is "NP-hard."
- The distribution search can also be "NP-hard."
- **Admissibility:** Is $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{b})$? (often easier fortunately)

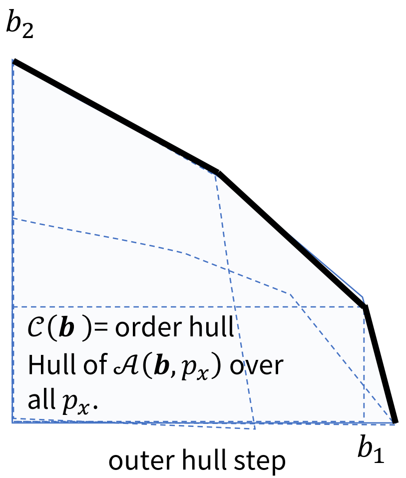
many cases simplify

The two convex-hull steps

- The **order-vertices'** hull

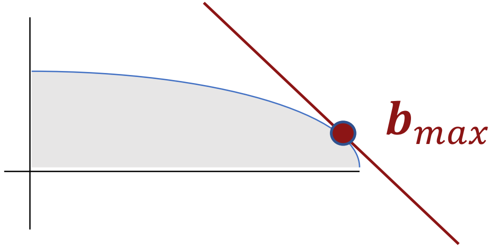


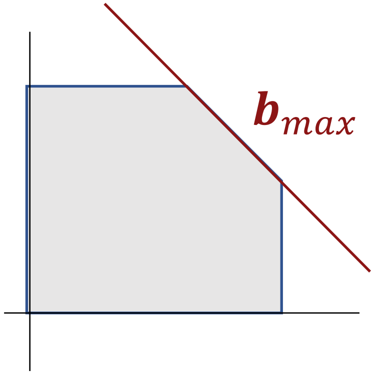
- The **input-distributions'** hull



Maximum Rate Sum

- The **rate sum** is $\mathbf{1}^* \boldsymbol{b}$, or simply the sum of the user bits/symbol.
- This is a hyperplane in U-space.
- This plane with normal vector $\mathbf{1}$ will be tangent to $\mathcal{C}(\boldsymbol{b})$ at $\boldsymbol{b}_{max}$, where $\mathbf{1}^* \boldsymbol{b}_{max}$= b_{max} , the maximum sum rate.





MU Matrix AWGN Channels

- $\mathcal{C}(\boldsymbol{b})$ for a multi-user AWGN channel $\boldsymbol{y} = H \cdot \boldsymbol{x} + \boldsymbol{n}$ will have all users input distributions as Gaussian at the region's (non-zero) boundary, $\mathscr{C}(\boldsymbol{b})$.

  - Each of these points is a mutual information that for each receiver/user $b_u = \mathbb{I}$ has a chain-rule decomposition.

  - For any subset of output dimensions $\boldsymbol{y}$ and any subset of inputs $\boldsymbol{x_u}$, $\mathbb{I}(\boldsymbol{x};\boldsymbol{y}) = \mathbb{I}\left(\boldsymbol{x_u};\boldsymbol{y}\ /\boldsymbol{x}_{\boldsymbol{U}\backslash\boldsymbol{u}}\right) + \mathbb{I}\left(\boldsymbol{x}_{\boldsymbol{U}\backslash\boldsymbol{u}};\boldsymbol{y}\ \right)$.
    - With independent input messages, these are separable and can be separately maximized.
    - The second term is a "single-user," $\boldsymbol{U}\backslash\boldsymbol{u}$, channel, and this channel thus has optimum Gaussian input.
    - The uncancelled users' crosstalk may contribute in MMSE sense to noise, which then is sum of Gaussians that is also Gaussian.

  - (Proof by induction: last user is single-user channel, which has Gaussian; then next to last has Gaussian xtalk and noise, so it also is Gaussian …), the optimum $\boldsymbol{u}$ is also Gaussian. This also works for any user subset $\boldsymbol{u}$. **QED.**

In general, with user components, treat $\boldsymbol{U} \rightarrow \boldsymbol{U}'$.

Degraded-Matrix AWGN

Definition 2.6.7 [(Subsymbol) Degraded multiuser Gaussian Channel] *A* **(subsymbol)-degraded AWGN multiuser channel** *has matrix ranks for H and/or $R_{\boldsymbol{xx}}$ that are ϱ_H and $\varrho_{R_{\boldsymbol{xx}}}$ respectively, such that*

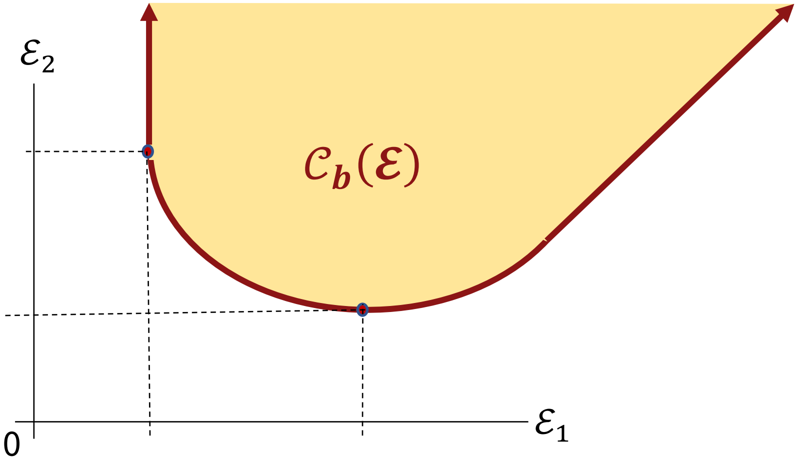
$$\min\left\{\varrho_{R_{\boldsymbol{xx}}}, \varrho_H\right\} < U \quad . \tag{2.284}$$

Otherwise, the channel is **non-degraded**. *The literature often omits the word "subsymbol," but it is tacit in degraded-channel definitions.*

This degraded definition depends on channel AND input.

- What "degraded" means physically is that there are not enough dimensions to carry all users independently.
  - There are other chain-rule conditional-probability definitions, but they appear equivalent.
- If all users energize, some must co-exist on the available (subsymbol) dimensions.
  - A name is NOMA (new name for old subject) – Non-Orthogonal Multiple Access (associated with IoT where U can be very large).
- Non-degraded channels (Massive MIMO is an example) have a surplus of dimensions (less likely to be degraded).
- $R_{\boldsymbol{nn}}$ is never singular on real channels, so noise whitening should not reduce the rank.
  - however, we will see a special case where design will assume a fictitious singular noise, so we'll need care on this when used.

Capacity-Energy Region (AWGN only)



- Essentially redraws the capacity regions for different energy vectors with fixed $\boldsymbol{b}$.
  - Trivially, any point within is reliably achievable, while points outside have insufficient energy.

- If a given $\boldsymbol{\mathcal{E}}_x \in \mathcal{C}_{\boldsymbol{b}}(\boldsymbol{\mathcal{E}})$, then $\boldsymbol{b}$ is **admissible** when also $\boldsymbol{b}_{\boldsymbol{\mathcal{E}}_x} \in \mathcal{C}(\boldsymbol{b})$.

Ergodic Capacity Region

- Design averages the capacity region over the variable-channel's parameter (joint if multiparameters) distribution.
  - This assumes messages are independent of parameters.

- Example: The **ergodic capacity region** is $\langle \mathcal{C}(\boldsymbol{b}) \rangle = \mathbb{E}_{\boldsymbol{H}}[\mathcal{C}(\boldsymbol{b})]$ for the matrix AWGN:

  - *interesting result* – The distribution $p_{\boldsymbol{x}}$ that maximizes the ergodic capacity when H is Raleigh (any user) fading is a discrete distribution (so then not Gaussian); extends well-known result for single user.

  - The AEP results don't hold because they assume the INPUT distribution is ergodic – and that is not necessarily true if the channel is varying (the reversal of input/channel limits for large blocklength may not hold and Rayleigh is example).

  - This presumably extends to multiuser case; however most channel variation for wideband (e.g. modern wireless) have codeword lengths/delays for good codes that are less than the coherence time, so Gaussian good codes remain in wide use. Thus, might as well go with Gaussian/known-good-codes for "quasi-stationary" assumption.

- **Outage Capacity Region?**
  - There is some work on "zero-outage" capacity region (depending on definition may not be same as $\langle \mathcal{C}(\boldsymbol{b}) \rangle$).

  - Not necessarily just $(1 - P_{out}) \cdot \langle \mathcal{C}(\boldsymbol{b}) \rangle$), like single-user case because of "which user outage?" question, although it probably is a decent measure anyway.

  - Probably more important to look at user input-rate variation (and contention for which point in $\mathcal{C}(\boldsymbol{b})$) and layer 2/3 buffer overflow outages, etc.

End Lecture 7

(Back-Up slides FYI)

Scheduling and Queuing

Section 2.6.7

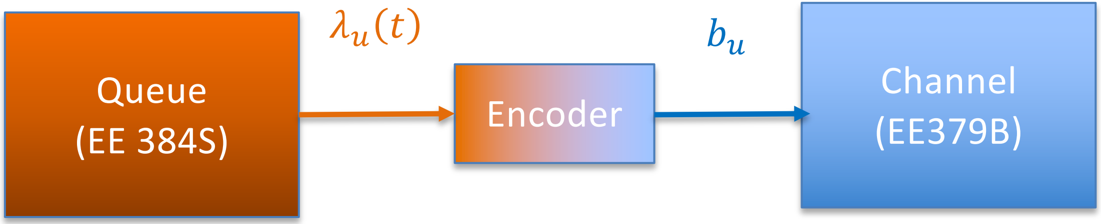
The real variation – the users' rates

Network Designer

says the "channel has a continuous rate that apportions dynamically as needed to any user, $\sum_u \lambda_u(t) = constant$."

Modem Designer

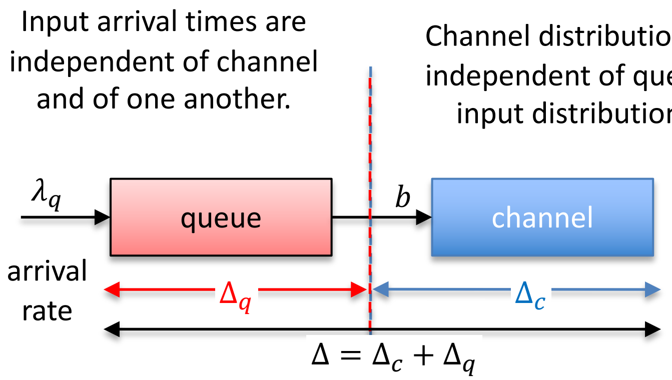
says the "source has a continuous rate that is always on, each b_u is constant."



- Neither of these two design perspectives is (always) correct.
  - See also Appendix A's queuing theory basics.

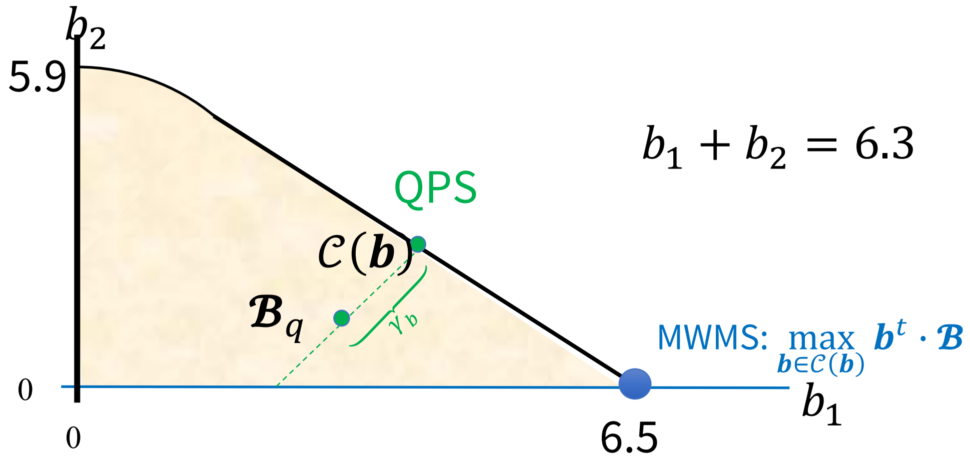
Queuing Basics

- Arrivals are independent of channel variation.
- $\mathcal{B} = \lambda_q \cdot \Delta \longrightarrow \mathbb{E}[\mathcal{B}] = \mathbb{E}[\lambda_q] \cdot \mathbb{E}[\Delta]$ = number of bits in system (Little's Theorem).
- $\mathbb{E}[\lambda_q] \leq \mathbb{E}[b]$ for stable operation.
- Multiuser Form
  - $\mathbb{E}[\boldsymbol{\mathcal{B}}] = \mathbb{E}[\boldsymbol{\lambda}_q] \odot \mathbb{E}[\boldsymbol{\Delta}]$.



Solution: Queue Proportional Scheduling

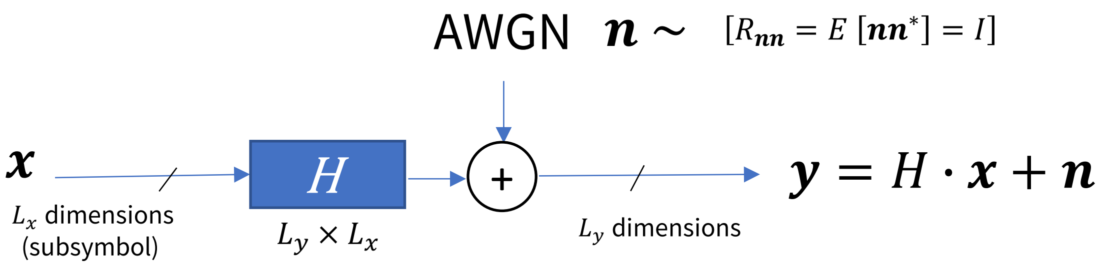
- Send data rate in capacity region that has user rate vector as scaled version of user queue depths.



We'll learn later how to find if a point is admissible (the green QPS point on the boundary).

- The design point is proportional to users relative queue depths, and has margin γ_b.
- QPS (Queue Proportional Scheduling) has lowest average delay of all scheduling methods.
- Less jitter than MWMS, fair among users (QPS empties the queues faster).

Dimensionality Table & AWGN



Type	$\boldsymbol{x}$ Number of inputs	$\boldsymbol{y}$ Number of outputs	H
multiple access	$U \cdot L_x$	L_y	$[H_U \; ... \; H_2 \; H_1]$
broadcast	L_x	$U \cdot L_y$	$\begin{bmatrix} H_1 \\ \vdots \\ H_{U-1} \\ H_U \end{bmatrix}$
interference	$U \cdot L_x$	$U \cdot L_y$	$\begin{bmatrix} H_{UU} & ... & H_{U1} \\ \vdots & \ddots & \vdots \\ H_{2U} & ... & H_{21} \\ H_{1U} & ... & H_{11} \end{bmatrix}$

Table 2.2: Table of dimensionality for the multi-user Gaussian channel $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n}$.

Best Decodable Set

Lemma 2.6.1 **[Best Decodable Set]** *When good codes (with* $\Gamma = 0$ *dB), given* $\mathbf{\Pi}$ *and* $p_{\boldsymbol{xy}}$*, and with*

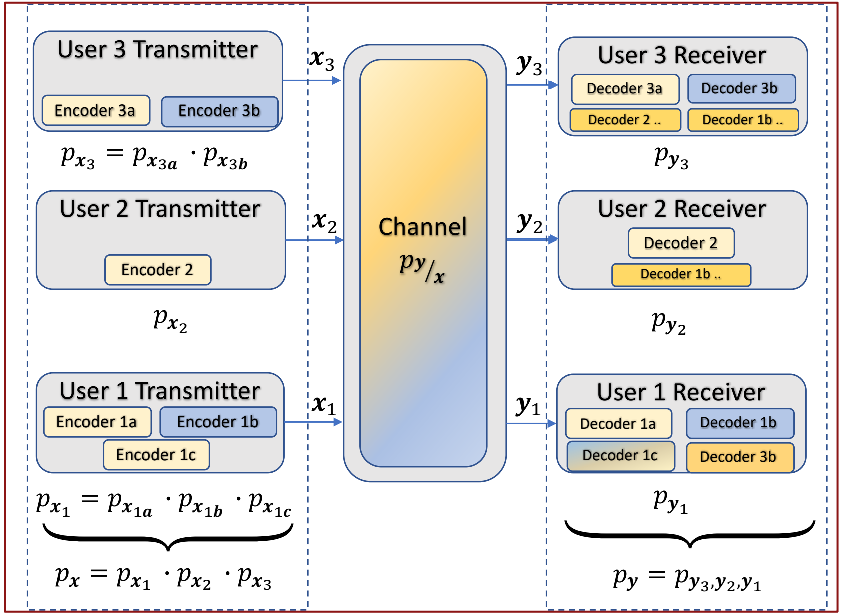
$$\boldsymbol{b} \preceq \boldsymbol{\mathcal{I}}_{min}(\mathbf{\Pi}, p_{\boldsymbol{xy}}) \quad , \tag{2.234}$$

then

$$\mathbb{P}_u(\boldsymbol{\pi}_u) \subseteq \mathscr{D}_u(\mathbf{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b}) \tag{2.235}$$

and receiver u *reliably achieves the data rate* $b = \mathcal{I}_u(\boldsymbol{x}_u; \boldsymbol{y}_u / \mathbb{P}_u(\boldsymbol{\pi}_u))$ *with order* $\boldsymbol{\pi}_u$*.*

- The proof follows the example (on right) on Slide L7:21



The sub users can correspond to vertices within convex combinations