STANFORD

Lecture 7

Multiuser Channels

and the Capacity Region

April 24, 2023

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering

Instructor EE392AA – Spring 2023

Announcements & Agenda

- Announcements
  - PS 2 solutions at canvas
  - Problem Set #3 due Wed by midnight
  - Problem Set 4 due next week Tues 17:00
    - solutions night before exam
  - Read Section 2.6
  - stat-loading can have p_g over tones

- 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
  - Scheduling and Queuing

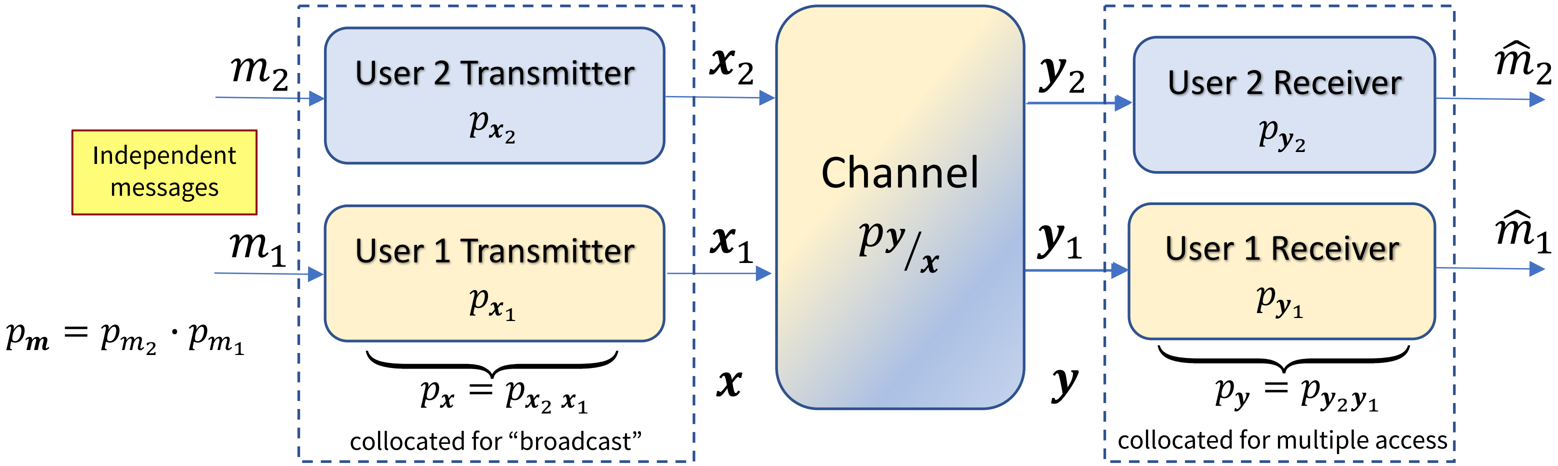
Multiuser (MU) Introduction

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)

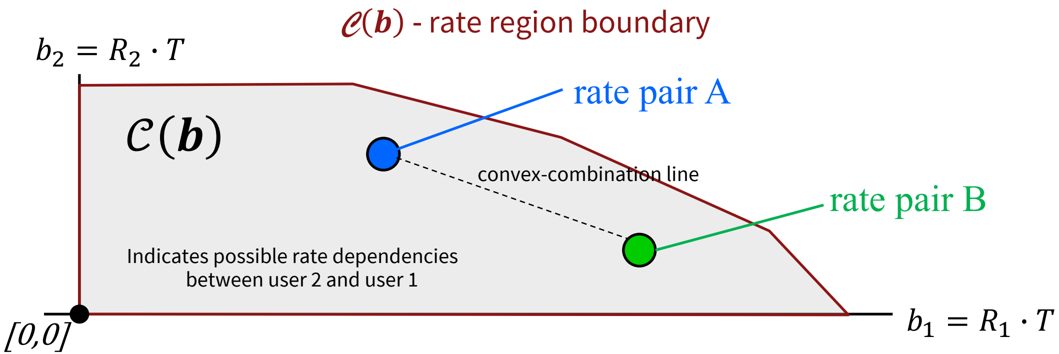
- There is a joint probability distribution $p_{\boldsymbol{xy}}$, from which come all marginals (e.g., input) and conditionals (channel)



- The data rates of user 1 and user 2 are mutually dependent (otherwise just two single-user channels)
- $b \rightarrow \boldsymbol{b} = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} = \boldsymbol{R} \cdot T = \begin{bmatrix} R_1 \cdot T \\ R_2 \cdot T \end{bmatrix}$; 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? rectangle (2), prism (3), Orthotope (U)
  "crosstalk free"

- The region is "convex hull" (union) of all 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.

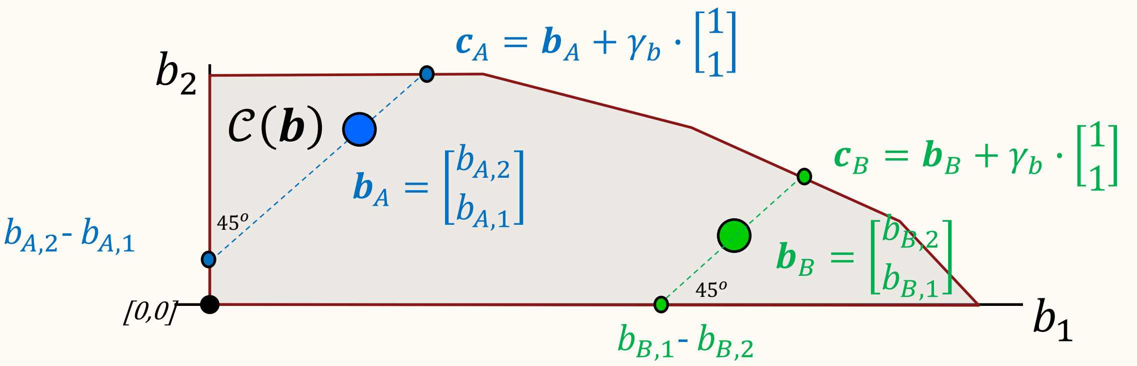
Multiuser Margin

- 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 distance to maximum value of $\bar{\mathcal{C}}$ $(\Gamma \cdot \gamma_m \cong 6 \cdot \bar{\gamma}_b$ dB , $\bar{\gamma}_b = 0$ if capacity-achieving code)

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



$$b_2 = b_1 + (b_{A,2} - b_{A,1})$$
$$b_1 = b_2 + (b_{B,1} - b_{B,2})$$

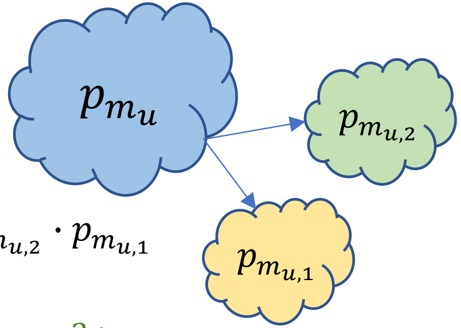
$$umax \triangleq \arg\left(\max_u b'_u\right)$$

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

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

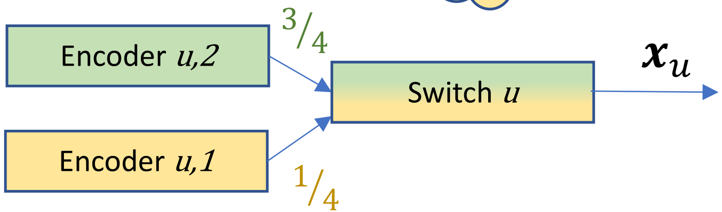
User Components (a.k.a. “time-sharing”)

- Two independent user components or subusers
  - Same transmitter and same receiver (“the user”)
- But the two subusers (codes used) can be separately encoded and decoded
- Bits per symbol: $b_u = b_{u,1} + b_{u,2}$

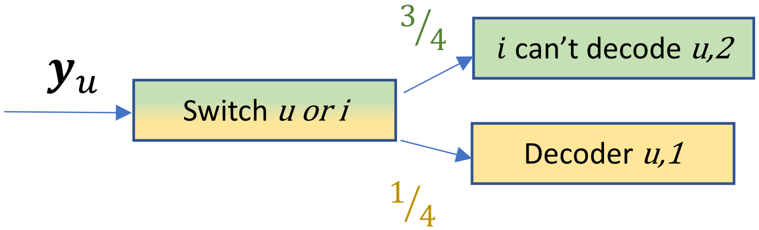


$$p_{m_u} = p_{m_{u,2}} \cdot p_{m_{u,1}}$$

- Another receiver $i \neq u$ may only be able to decode one of the components
  - (which it should do and remove if it can)
  - While the other remains as noise
- The two subusers can also simultaneously share dimensions where the fractions apportion information (or energy) 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 $\mathcal{C}(\boldsymbol{b})$ simply adds together the sub-users’ dimensional rates
  - and 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

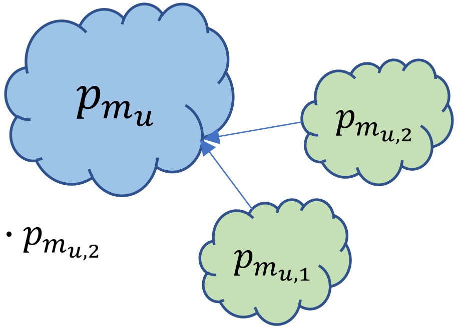
Macro Users

- Two users (or user components) that have identical impact/influence create a macro user
  - $p_{\dots x_u \dots x_i \, y} = p_{\dots x_i \dots x_u \, y}$ - 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 $y = x_1 + x_2 + n$ where both users 1 and 2 share the same energy
  - Same as a 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

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

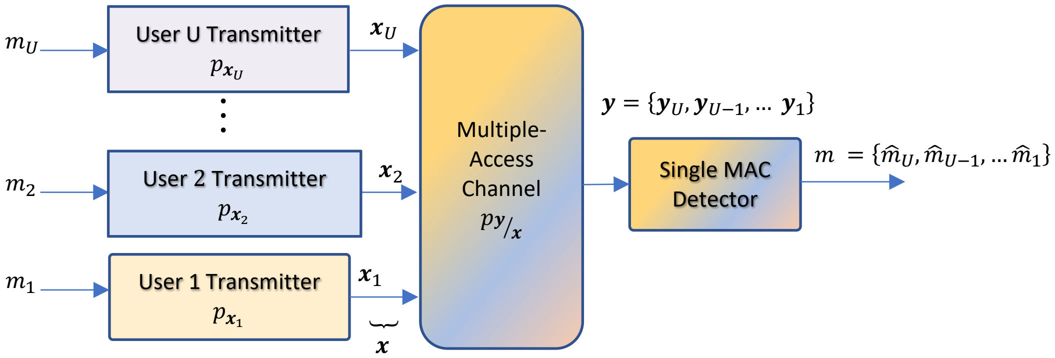


The 3 Basic MUs & Matrix AWGN

PS4.1 - 2.21 Multiuser Channel Types

Section 2.6.1

Multiple Access Channel (MAC)



Sec 2.6.1 and 2.7

- User transmitters in different locations (cannot coordinate to generate $\boldsymbol{x}$)

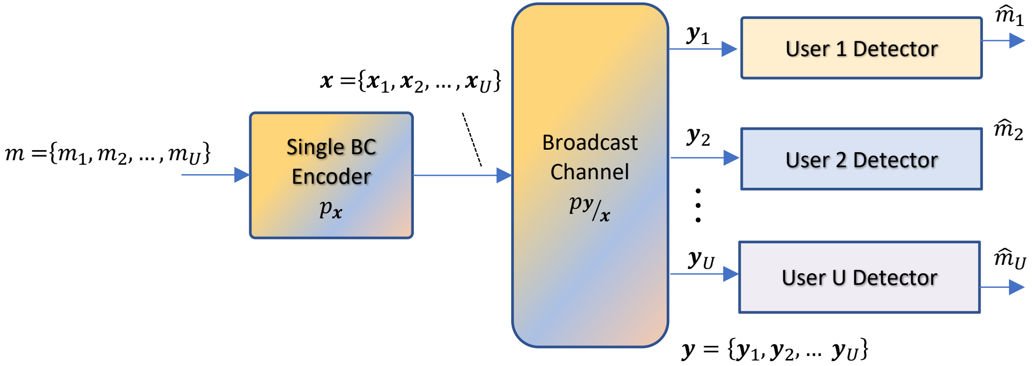
- Single receiver detects all users
  - Separates the users
  - Reliably decodes, $P_e \to 0$, by decoding and removing some (none or all) other users first
  - Suggests "user **order**" $\boldsymbol{\pi}$ (vector "priority") is important (decode $\boldsymbol{\pi}$'s 1st/bottom element first, … U … last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

- **Order is fundamental** to best MU design – the MAC has $U'!$ possible orders (each has $\boldsymbol{b}$)
  - and all potential convex combinations thereof
  - There is also a choice of input $p_{\boldsymbol{x}}$ (or the code), and all potential convex combinations thereof

Design

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

Broadcast Channel (BC)



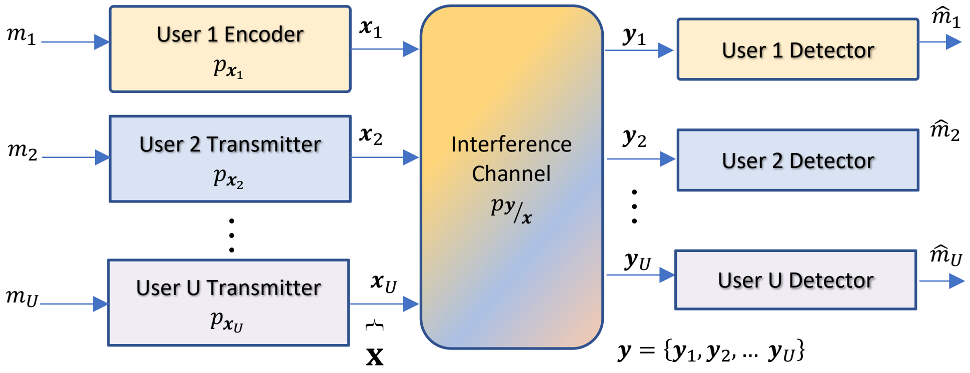
Sec 2.6.1 and 2.8

- "Dual" of MAC
- Receivers in different places – 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_x

Interference Channel



Sec 2.6.1 and 2.9

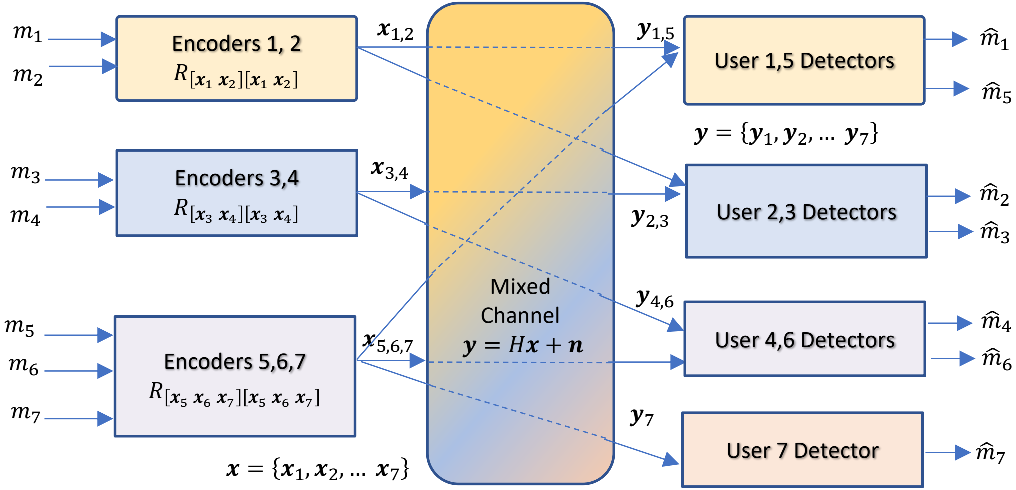
- Transmitters, and also receivers, are in different locations
  - No co-encoding of user messages nor coordinated reception of users
- Each receiver can use a decoding **order** to detect others first, if that is possible
  - Treat other users as noise if not possible to decode/remove first
  - Each receiver's order is column of matrix $\mathbf{\Pi}$.
- There are $(U'!)^U$ possible IC **orders:** … $U'!$ at each receiver, with $U' \leq U^2$
  - each user may have a subuser component for every user's receiver to detect

Design

1. $\mathbf{\Pi}$
2. p_x

Other Types / Combinations

- Mixed



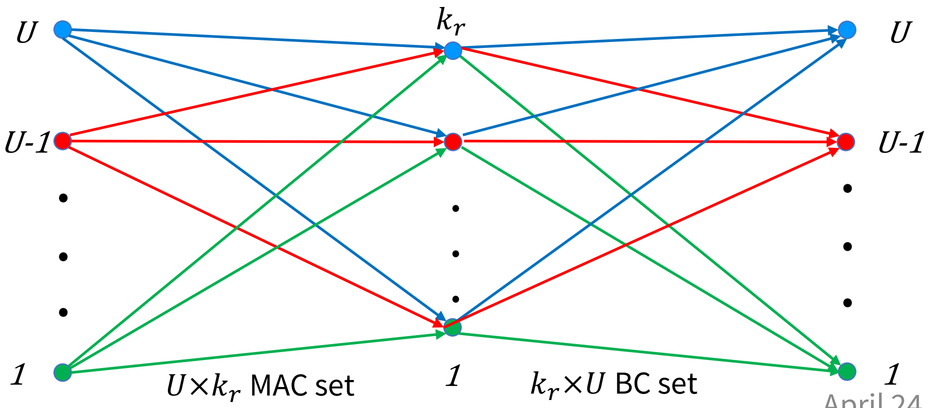
Different Macro views:

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}

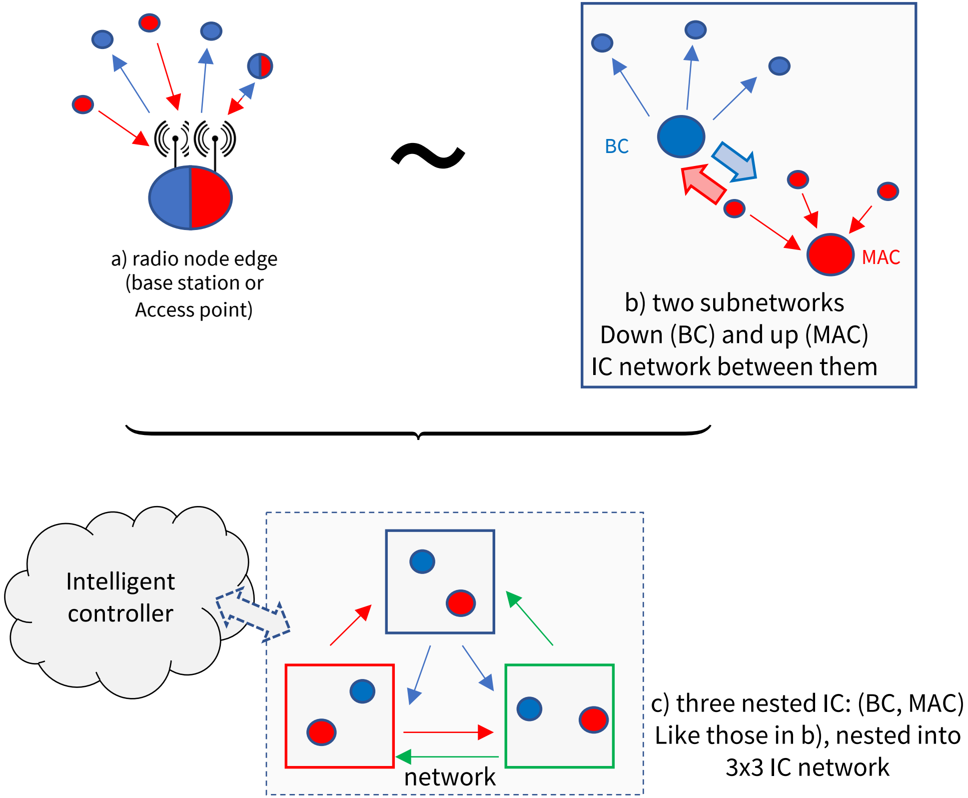
Macro design can shrink $\mathcal{C}(\boldsymbol{b})$

- Mesh/Relay
- Single User



Nested MU Channels

- Each nested MU channel → 1 macro user
- These macro users crosstalk into each other
  - Some users with macro group may decode
  - At given order in that group
  - Others are undecodable noise
- Treat as IC of macro users
  - Order the macro users
  - Decode all users within the local macro group



Rate Bounds & Detection

PS4.3 - 2.23 Mutual-Information Vector

Chain-Rule Reminder/Review

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

Lemma 2.3.4

- Think of the input components $\boldsymbol{x}_n$ as users, so $U \rightarrow N$ and $u \rightarrow n$ (may have U' replacing U in general)
- Any receiver output (or combination of them), $\boldsymbol{y}$, has chain-rule decomposition(s) and 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.

Simplify possible U' to just U in this section

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 Important?
  - The as-yet un-decoded users are "noise" (averaged to compute marginal dist'n p_{y_u/x_i} on which ML detector is based)
  - Are the other users reliably decodable? (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
  - Permutation vector has inverse

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

Prior-User Set & the $\mathcal{I}_{min}$ vector

- Enumerates reliably decodable users for any particular receiver, up to $U' \leq U^2$ in each column
  - Over-simplifed here to 1 subuser/user (channel externally has simplifying features presumably) to illustrate the algorithm
  - Otherwise, it would be 16 entries in each column and subuser components' rates add to the parent user's rate

- Prior-User Set is $\mathbb{P}_u(\boldsymbol{\pi}) = \{j \mid \boldsymbol{\pi}^{-1}(j) < \boldsymbol{\pi}^{-1}(u)\}$
  - That is "all the (sub) users before the desired user u in the given order $\boldsymbol{\pi}$.
  - Receiver u best decodes these "prior" (sub)users and removes them, while "post" (sub)users are noise
  - $\boldsymbol{\pi}$ can be any order in $\mathbb{P}_u(\boldsymbol{\pi})$, but the most interesting is usually $\boldsymbol{\pi}_u$ (receiver u's order)

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
$\mathbb{P}_u(\boldsymbol{\pi}_u)$	$\{1,2\}$	$\{2,4,1\}$	$\{1\}$	$\{4\}$

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

- List data-rate entries (mutual information bounds really) for those users who are decoded, and not 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

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

Decodable users and $\mathcal{I}_{min}$

- Mutual-information-like quantity
  - Relates to the prior-user set

$$\boldsymbol{\mathscr{I}}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \begin{bmatrix} \mathscr{I}_u\left(\boldsymbol{x}_{\pi_u(U)}; \boldsymbol{y}_u / \mathbb{P}_{\pi_u(U)}(\boldsymbol{\pi}_u)\right] \\ \vdots \\ \mathscr{I}_u\left(\boldsymbol{x}_{\pi_u(i)}; \boldsymbol{y}_u / \mathbb{P}_{\pi_u(i)}(\boldsymbol{\pi}_u)\right) \\ \vdots \\ \mathscr{I}_u\left(\boldsymbol{x}_{\pi_u(1)}; \boldsymbol{y}_u / \mathbb{P}_{\pi_u(1)}(\boldsymbol{\pi}_u)\right] \end{bmatrix}$$

- Decodable user set $\mathcal{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b})$
  - All users that receiver u can decode

- Best if $\mathcal{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b})$ = $\mathbb{P}_u(\boldsymbol{\pi}_u)$
  - But not necessarily so , and depends on $\boldsymbol{\Pi}$ and the attempted $\boldsymbol{b}$

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

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

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

Definition 2.6.2 **[Decodable Set and Minimum Mutual Information Vector]** *For a given* $\boldsymbol{\Pi}$, $p_{\boldsymbol{xy}}$, *and* $\boldsymbol{b}$, *each receiver* u *will be able to detect reliably (on average in the AEP sense) other* $(i \neq u)$ *subusers (user components) in the set*

$$i \in \mathscr{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b}) \qquad (2.242)$$

with $P_e \to 0$, *When receiver* u *can detect no other users reliably,* $\mathscr{D}_u(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}, \boldsymbol{b}) = \emptyset$, **with this order**.

Every multiuser channel has a minimum mutual-information vector with components

$$\mathcal{I}_{min,u}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \min_j \left\{ \mathscr{I}_j\left(\boldsymbol{x}_{\pi_j(u)}; \boldsymbol{y}_j / \mathbb{P}_{\pi_j(u)}(\boldsymbol{\pi}_j)\right) \right\} \ , \qquad (2.243)$$

and thus the **minimum mutual-information vector**, *or vertex, is*

$$\boldsymbol{\mathcal{I}}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) = \begin{bmatrix} \mathcal{I}_{min,U'}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \\ \vdots \\ \mathcal{I}_{min,u}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \\ \vdots \\ \mathcal{I}_{min,1}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \end{bmatrix} \ . \qquad (2.244)$$

In the most general case, the $\mathcal{I}_{min}$ *vector entries are sums of each user's subuser component rates that are minimally decodable everywhere; where more than one single user u's subuser components at any receiver i are decodable within the order, then the* $\mathcal{I}_{min}(i)$ *calculations should sum those reliably decodable components' subrates at receiver i before comparing the minima across all receivers.*

Example: sum of 3 users (MAC)

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

real subsymbols

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

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

- With one receiver, the π_u vectors are trivially scalars, so the $U!$ is 6, but the exponent simplifies to 1.
- 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 $b_3 = 2$.

Best Decodable Set

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

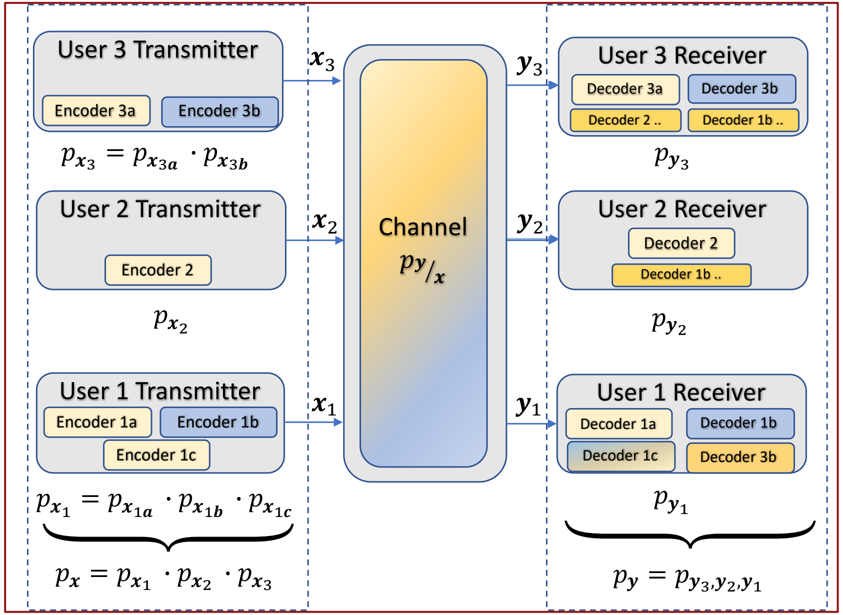
$$\boldsymbol{b} \preceq \mathcal{I}_{min}(\boldsymbol{\Pi}, p_{\boldsymbol{xy}}) \quad , \tag{2.234}$$

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 proof follows the example (on right) on Slide 20



The sub users can correspond to vertices within convex combinations

Optimum Detectors (2.6.3)

- Section 2.6.3 formalizes (general, including non-Gaussian, case) optimum detection
- There are various integrals/sums and definitions
- Formally what it means is each user's optimum detector for any given order $\boldsymbol{\pi}$ must
  - First detect all other users who are earlier in that order
  - Each such detector considers 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 subtracted in Gaussian case with no effect on further detection)
- The error-probability calculation then follows like single-user, simply with any "pre-users" no longer present and any "post-users" averaged (treated like noise)
- Thus, it is something you already know well – just complex notation for the multiuser case

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

- The general notation may be less helpful than the concepts of
  - The decodable users, $\mathfrak{D}_u(\boldsymbol{\Pi})$, are first detected and then "cancelled" – they contribute no "noise" (earlier in order)
  - Other users, $\overline{\mathfrak{D}}_u$, $(\boldsymbol{\Pi})\backslash u$ are not first detected and are "averaged" (treated as noise)

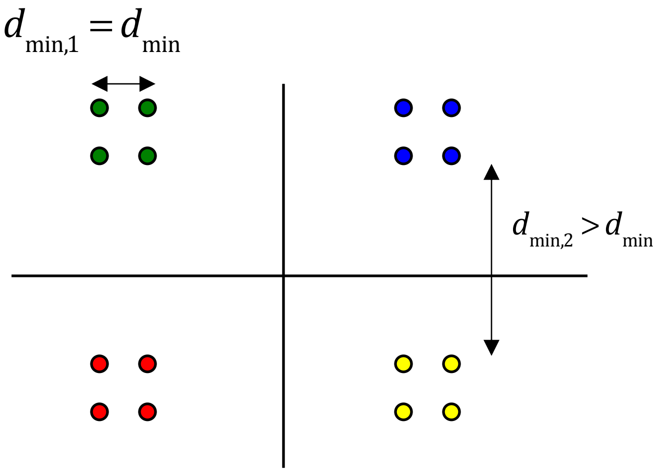
$$p_{\boldsymbol{x}_u/[\boldsymbol{y}\ \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}]}(\boldsymbol{\chi}_u, \boldsymbol{x}_{i\in\mathscr{D}_u(\boldsymbol{\Pi})}, \boldsymbol{y}) = \int_{\boldsymbol{\chi}\in\boldsymbol{x}_{\{\overline{\mathscr{D}}_u(\boldsymbol{\Pi})\setminus u\}}} p_{\boldsymbol{x}/[\boldsymbol{y}\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

integration/sum is 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 from channel prob

Simple Example



- The decoder should decode first red, green, blue, yellow; this treats the variation within each color as "noise"
- Then the decoder would re-center the constellation, and decide again which of the 4 same-color points
  - This effectively cancels the noise from the first step
- Yes, an overall decoder performs the same if all in one step if first decision is correct, but the basic concept expands

General MU Capacity Region and related optima

Section 2.6.4

Order-and-Distribution-Dependent Region

- Form a first convex hull of all $\mathcal{I}_{min}$ vectors FOR EACH GIVEN ORDER and input distribution

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

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

- Now maximize over the allowed input distributions (a 2nd convex hull operation, but now on distributions)

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

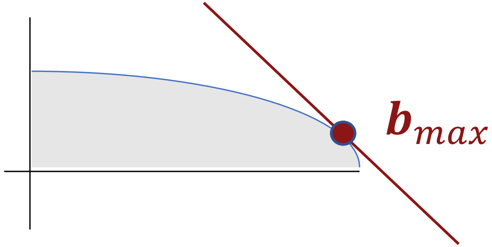
MU Capacity Region

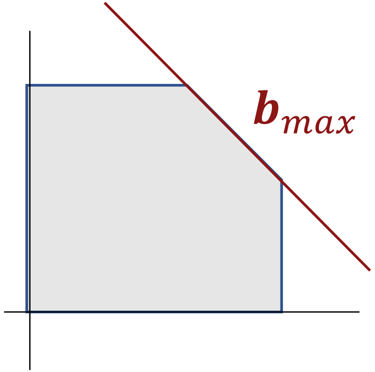
- the order search is "NP-hard"
- The distribution search can also be "NP-hard"
- Admissibility: Is $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{b})$? (often easier fortunately)

many cases simplify

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 $b_u = \mathcal{I}$ has a chain-rule decomposition
  - For any subset of output dimensions $\boldsymbol{y}$ and any subset of inputs $\boldsymbol{x}_{\boldsymbol{u}}$, $\mathcal{I}(\boldsymbol{x};\boldsymbol{y}) = \mathcal{I}\left(\boldsymbol{x}_{\boldsymbol{u}};\boldsymbol{y}\,/\boldsymbol{x}_{\boldsymbol{U}\setminus\boldsymbol{u}}\right) + \mathcal{I}\left(\boldsymbol{x}_{\boldsymbol{U}\setminus\boldsymbol{u}};\boldsymbol{y}\,\right)$; and with independent input messages, these are separable and can be separately maximized. The second term is a "single-user," $\boldsymbol{U}\setminus\boldsymbol{u}$, 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.**

Again with user components, treat $U \rightarrow U'$

Degraded-Matrix AWGN

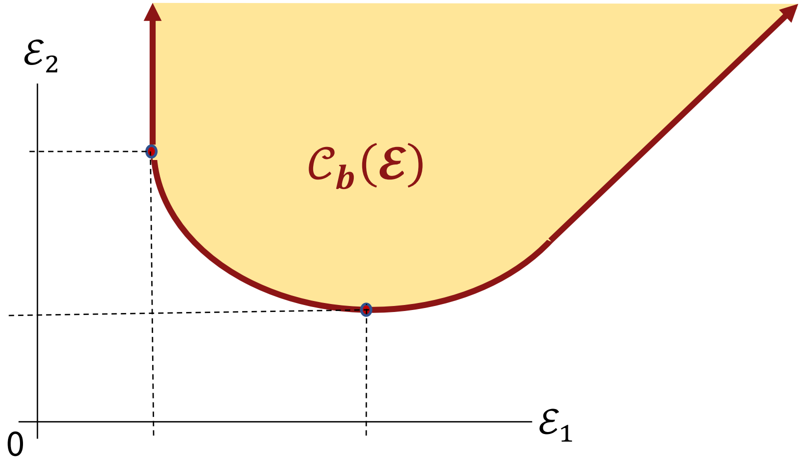
Definition 2.6.7 [(Subsymbol) Degraded multiuser Gaussian Channel] *A* **(subsymbol)-degraded AWGN multiuser channel** *has matrix ranks for H and/or $R_{\boldsymbol{xx}}$ that are ϱ_H and $\varrho_{R_{\boldsymbol{xx}}}$ respectively, such that*

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

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

- 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
  - Sometimes called 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

- Averages the capacity region over the variable-channel's parameter (joint if multiparameters) distribution
  - Assumes input is 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?
  - 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

Scheduling and Queuing

Section 2.6.7

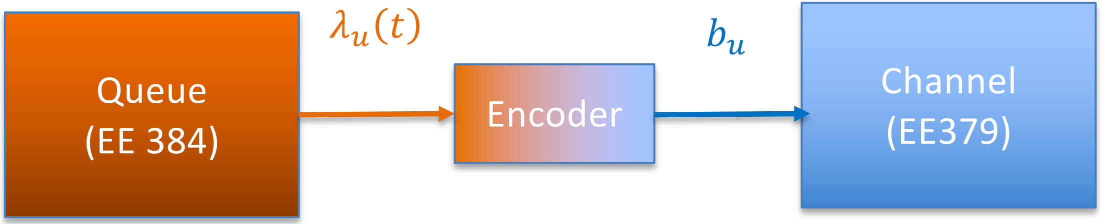
The real variation – the users' rates

Network Designer

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

Modem Designer

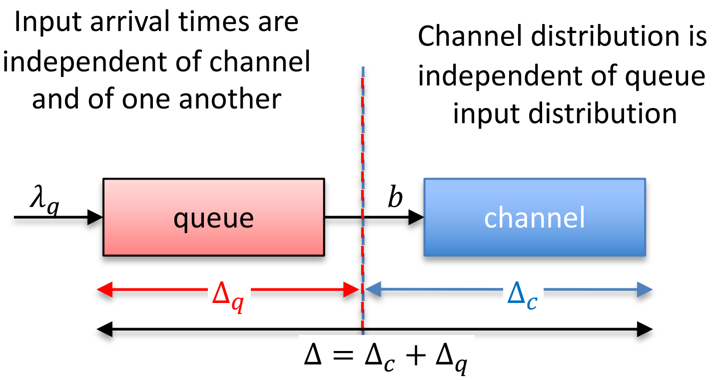
"source has a continuous
rate that is always on,
Each b_u is constant"



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

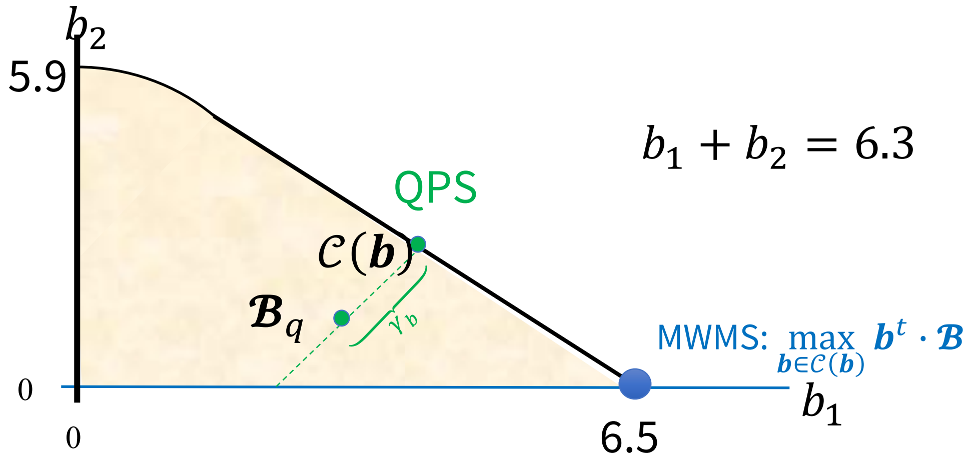
Queuing Basics

- Arrivals 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

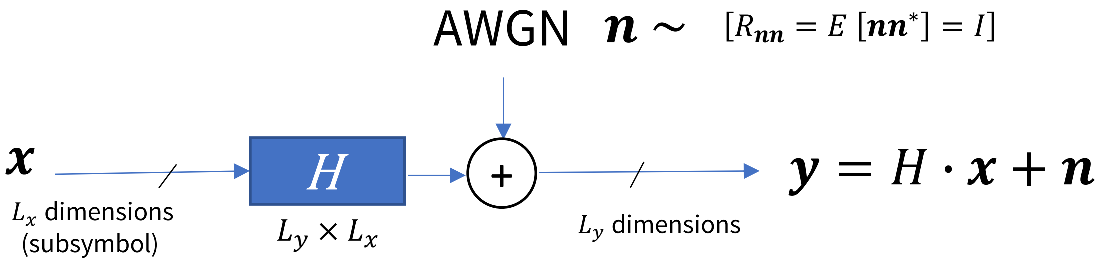


We'll learn later how to find if a point is admissible (the green QPS point on the boundary)

- 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)

End Lecture 7

Dimensionality Table & AWGN



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

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

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.