STANFORD

Supplementary Lecture 7A

General Capacity Region via Atoms

April 20, 2026

JOHN M. CIOFFI

Hitachi Professor Emeritus of Engineering

Instructor EE379B – Spring 2026

Announcements & Agenda

- Agenda
  - Atomization and Completeness
  - Capacity Region and Converse
  - Gauss Everywhere

It helps to view this after understanding MAC/L7

Atomization and Completeness

Section 2.6.4.1-2

Subuser Atoms

Atomization decomposes users/subuser to their smallest components, "atoms."

- **Subuser subsets:** Let $S \subseteq \{[1{:}\,U] \backslash \emptyset\}$ for receiver indices and $T \subseteq \{[1{:}\,U] \backslash \emptyset\}$ for transmitter indices.
- For an S, atom $a = (i, S)$ contains all receivers $i \in S$ that reliably decode the atomic subuser rate:
  - $b_{i,S} \equiv b_a$.
  - For each of the $(2^U - 1)$ possible S subsets, a indexes the common decodable (subuser) information for user i.
  - Across all U users, there are $U \cdot (2^U - 1)$ atoms $a \in \mathcal{A}$.
- A receiver r has decodable atoms in set $\mathcal{A}_r$ and all the atoms in $\mathcal{A} \backslash \mathcal{A}_r$ are noise.
  - Removed through integration/summing to marginal distributions of remaining atoms in $\mathcal{A}_r$.

While large $|\mathcal{A}|$ is finite, it completely represents all possible solutions.

Fano's Inequality (2.3.5)

- For any decoder with error probability $P_e^{(N)}$, the input signal's conditional entropy satisfies

$$\textbf{FANO:}\ \ \bar{\mathcal{H}}_{\boldsymbol{x}/\boldsymbol{y}_r} \le \mathcal{H}\left(P_e^{(N)}\right) + P_e^{(N)} \cdot \bar{b}$$

$$\mathcal{H}(p) \triangleq p \cdot \log_2 p + (1-p) \cdot \log_2(1-p)$$
$$\mathcal{H}(0) = 0$$

- So, for a good code with error probability $P_e^{(N)} \to 0$, FANO says the input is known with certainty.

- (Single-user) Converse if $\bar{b} = \bar{\mathcal{C}} + \delta$, then rearranging FANO yields

$$P_e \ge \frac{N \cdot \delta - 1}{\bar{\mathcal{C}} + \bar{b}}$$

- Bad news, flip coins as N gets bigger.

- But MU Capacity has an additional multiuser Fano use.

- MU use shows that **any/all** good-code set with all users' messages reliably decodable:
  - will differ from the a good code set based on the $2^U - 1$ possible S choices by an *asymptotically vanishing (zeroing)* rate-vector.
  - Thus, the atoms are complete to specify multiuser information.

MU Fano Use

- Subuser message m_i at receiver r, with $a = (i, S)$ and $r \in S$ has Fano IE:

$$\frac{1}{N} \cdot \mathcal{H}_{m_i | \boldsymbol{y}_r} \leq \frac{1}{N} \cdot \left[1 + P_{e,i \to r}^{(N)} \cdot \log_2 |M_i|\right] \to \underbrace{0}_{N \to \infty}$$

So, this is "$o(1)$" → 0 linearly

- Choose $\delta_N = {}^{1}/_{\sqrt{N}}$ so that $N \cdot \delta_N \to \infty$ and reduce $b_i' = b_i - \delta_N$ by deleting codewords from this code.
  - New-code receivers still work reliably, just with lower rate as fewer codewords, the deletions leave $M_i' = 2^{N \cdot b_i'}$.
- The $\mathcal{H}_{m_i | \boldsymbol{y}_r}$ for the original code goes to zero more quickly (via Fano above) than the $\delta_N \to 0$.
- Map original code's messages m_{i,S_i} to new b_i' code, when $i \in S$, but to constant (1) for other $i \notin S$ (need not decode because above is only for a receiver $r \in S$.
  - Reduces number of codewords; original number of codewords was $M_i = 2^{N \cdot b_i}$, and new is $M_i' = 2^{N \cdot b_i - N \cdot \delta_N}$.
    - So, decoding remains reliable (same code, just some codewords no longer being transmitted).
  - But the amount remaining is more than enough to cover the original codes approach to $\mathcal{H}_{m_i | \boldsymbol{y}_r}$, ($\delta_N > \frac{1}{N}$).
  - This atomic construction covers any optimum code set's codewords, all of them.
- Any good code set differs by vanishing small data rate from codes based on $2^U - 1$ atomic set based on S.
  - Even smaller subset is complete for special cases like MAC/BC (U) and IC (U^2).

User-group rate bounds

- Sum over a receiver r atomic subset group cannot exceed rate for perfect cancellation of others not in group.
- Atomic increment:

$$\Delta_r(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}}) \triangleq \mathcal{I}\left(\boldsymbol{x}_a; \boldsymbol{y}_r / \boldsymbol{x}_{\mathbb{P}_{\boldsymbol{\pi}_r}(a)}\right) \text{ and } \mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}}) = \min_{r \in S(a)} \Delta_r(a, \boldsymbol{\pi}_r, p_{\boldsymbol{x}})$$

- User i's data rate is:

$$b_i = \sum_{S \ni i} b_{i,S} \quad ; \quad m_i = \{m_{i,S} \mid \emptyset \neq S \subseteq [1{:}U]\}$$

- Transmitter subsets:

$$\boldsymbol{x}_T \triangleq \{\boldsymbol{x}_i : i \in T\}, \qquad \boldsymbol{x}_{T^c} \triangleq \{\boldsymbol{x}_i : i \in ([1:U] \setminus T)\}.$$

- Result above formally:

$$\boxed{\sum_{i \in T} \sum_{S \ni r} b_{i,S} \leq \mathcal{I}(\boldsymbol{x}_T; \boldsymbol{y}_r / \boldsymbol{x}_{T^c})}$$

Capacity Region, & Converse

Section 2.6

Achievability of Imin approach

- Any atom that can is reliably decoded $\forall\, r \in S$, then the atom's rate must satisfy:

$$b_a \triangleq \lim_{n\to\infty} \frac{1}{n} \cdot \mathcal{H}_{m_{i,S}} \cdot$$

$$b_a \leq \min_{r\in S} \liminf_{n\to\infty} \frac{1}{n} \cdot \mathcal{I}\left(m_{i,S}; \boldsymbol{y}_r^n / M_{\mathrm{prev}_{\pi_r}(a)}\right)$$

- The rate b_a 's upper bound $\mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$ is reliably decodable by Lecture 7's GDFE/successive decoding, just applied to the atom
  - And anything ahead of it in the order is given/removed (chain rule).

- Via dimension sharing (L7 also), any convex combination over orders (and input probabilities) is also achievable

Capacity Region Expression

$$\mathcal{C}(\boldsymbol{b}) = \underbrace{\sum_{\substack{S \ni i \\ i \in [1:U]}}}_{\textit{collapse to U dimensions}} \left\{ \bigcup_{p_{\boldsymbol{x}} \in \mathcal{P};\boldsymbol{\Pi}}^{conv} \{ b_{i,S} \mid 0 \leq b_{i,S} \leq \mathcal{I}_{min}[(i,S),\boldsymbol{\Pi},p_{\boldsymbol{x}}] \} \right\}$$

- The sum in front just sums for each user all its subuser atoms' rates, as $\mathcal{C}(\boldsymbol{b})$ is U-dimensional.
- The convex hull is over all the possible orders and input distributions (this often simplifies).
- The inner set are just the bounding planes for the atoms
  - These are similar to those in L7's MAC/chain rule, but now over an expanded rate-dimenstionality (as much as $2^U - 1$).
- Why not earlier? This approach uses the finite atomic structure (which differs from any optimal solution by vanishing small rate) and chain rule rather than try to randomly characterize multiuser auxiliary variables.
  - The completeness of the finite-enumerable atomic structure is the new concept that allows this progress.

Converse

- The converse is relatively easy at this point.

- The atomization has created a set of achievable regions, but the converse for those MAC-like regions just apply directly here also to them all. A point outside their convex hull has nonzero/bad P_e .
  - Single user Fano applies also if one prefers.

- The hard part was the completeness of the $2^U - 1$ being vanishingly small rate difference from any good IC code set. Once that is established, the results just reuse MAC-like results along with the Imin concept.

Gauss Everywhere!

Section 2.6

This is intuitively appealing, but a little harder than the capacity region.

Gaussian inputs are MU Optimum (with AGN)

- Any multiuser additive Gaussian noise channel best uses Gaussian input distributions:
  - **UNDER AUTOCORRELATION (or trace thereof) ENERGY CONSTRAINTS.**

- This is true, even if you hear "worst-case noise distribution" or other reservations, maybe from someone claiming to know better.

- So, just like single user (and basically because the atomization decomposes almost into many, but finite, single-user-like channels), all multiuser Gaussian channels have all Gaussian inputs as best.

  - This was known previously for MAC and BC but often stated as "possible but unknown in general."
  - **It's true in general as this S7A and Section 2.6.5.4**.

- The proof leverages the earlier completeness, but also needs several a few other information-theory results, particularly:
  - Support functions,
  - Abel Transforms,
  - Fischer information, and
  - De Brujin's identity.

Support-Function Argument

- Restate the $\mathcal{I}_{min}$ in terms of atoms: $0 \le b_a \le \mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$.

- User rate is summed over these atoms: $b_i = \sum_{S \ni i} b_{i,S}$.

- The **support** looks at non-negative weighted rate sums $\boldsymbol{\mu} \succcurlyeq \mathbf{0}$; $\boldsymbol{w} \succcurlyeq \mathbf{0}$.

$$\sum_{i=1}^{U} \mu_i \cdot b_i = \sum_{a \epsilon \mathcal{A}} w_a \cdot b_a$$

- The outer atomic bounds then have view as function of the input distribution $p_{\boldsymbol{x}}$.

$$\max_{0 \le b_a \le \mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})} \sum_{a \epsilon \mathcal{A}} w_a \cdot b_a = \sum_{a \epsilon \mathcal{A}} w_a \cdot \mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}}) \triangleq \Phi_{\boldsymbol{\Pi}, \boldsymbol{w}}(p_{\boldsymbol{x}})$$

Eliminate (Gauss) min

- Reintroduce the receiver index r :

$$\Delta_r(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}}) \triangleq \mathcal{I}\left(\boldsymbol{x}_a; \boldsymbol{y}_r / \boldsymbol{x}_{\mathbb{P}_{\boldsymbol{\pi}_r}(a)}\right)$$

$$\mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}}) = \min_{r \in S(a)} \Delta_r(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$$

- For a specific r that corresponds to where the minimum $\mathcal{I}_{min}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$ occurs

$$r_a^* \in \arg\min_{r \in S(a)} \Delta_r(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$$

- The support function for this receiver becomes

$$\Phi_{\boldsymbol{\Pi}, \boldsymbol{w}}(p_{\boldsymbol{x}}) = \sum_{a \epsilon \mathcal{A}} w_a \cdot \Delta_{r_a^*}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$$

- So must show that the following needs to be maximum for any $r_a \in \{r_a\}$ for Gaussian $p_{\boldsymbol{x}}$.

$$\Psi_{\boldsymbol{\Pi}, \boldsymbol{w}}(p_{\boldsymbol{x}}; \{r_a\}) \triangleq \sum_{a \epsilon \mathcal{A}} w_a \cdot \Delta_{r_a}(a, \boldsymbol{\Pi}, p_{\boldsymbol{x}})$$

Receiver-wise Decompostion and Abel

- Decompose over receivers: $$\Psi_{\mathbf{\Pi},\boldsymbol{w}}(p_{\boldsymbol{x}};\{r_a\}) = \sum_{r=1}^{U}\sum_{\{a|r_a=r\}} w_a \cdot \Delta_{r_a}(a,\boldsymbol{\Pi},p_{\boldsymbol{x}})$$

- Recognize $\Delta_r(a,\boldsymbol{\Pi},p_{\boldsymbol{x}})$ as differences of differential entropies: $$\Delta_{r_a}(a,\boldsymbol{\Pi},p_{\boldsymbol{x}}) = \mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{\mathbb{P}(a)}} - \mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{[\mathbb{P}(a),\boldsymbol{x}_a]}}$$

- Recognize support sum in terms of differences ("Abel Trans"):

$$\sum_{a\epsilon\mathcal{A}} w_a \cdot \Delta_{r_a}(a,\boldsymbol{\Pi},p_{\boldsymbol{x}}) = \sum_{k=1}^{K_r} \alpha_{r,k} \cdot \mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{T_{r,k-1}}} - \beta_r \cdot \mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{T_{(r,K_r)}}}$$

$$T_{r,0} = \emptyset, \qquad T_{r,0} \subseteq T_{r,1} \subseteq \cdots \subseteq T_{r,K_r}.$$

$$\alpha_{r,k} \triangleq w_{r,k} - w_{r,k-1}, \qquad \beta_r \triangleq w_{r,K_r},$$

$$\sum_{k=1}^{K_r} \alpha_{r,k} = \beta_r \ .$$

- Several algebra steps and a reordering (priors only) to make alphas nonnegative where necessary leads to

$$\sum_{a\epsilon\mathcal{A}} w_a \cdot \Delta_{r_a}(a,\boldsymbol{\Pi},p_{\boldsymbol{x}}) = \sum_{k=1}^{K_r} \alpha_{r,k} \cdot \left(\mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{1:k-1}} - \mathscr{H}_{\boldsymbol{y}_r|\boldsymbol{x}_{1:K_r}}\right).$$

Gaussian Smoothing

- The differential entropy $\mathscr{H}_{\boldsymbol{y}_{r|\mathbb{P}}(a)}$ is maximized by Gaussian, even if there is uncancelled other atoms.

- Proving this starts with Gaussian smoothing.
  - The smoothing adds a variable (energy t) Gaussian Z_k to random variables U_k think of these as users/receivers sums of user and crosstalk.

$$U_k \triangleq \sum_i A_{k,i} \cdot \boldsymbol{x}_i + \boldsymbol{n}_k, \qquad U_k^G \triangleq \sum_i A_{k,i} \cdot \boldsymbol{x}_i^G + \boldsymbol{n}_k; \qquad U_k(t) \triangleq U_k + \sqrt{t} \cdot \boldsymbol{z}_k, \qquad U_k^G(t) \triangleq U_k^G + \sqrt{t} \cdot \boldsymbol{z}_k.$$

- Take non-neg weighted differences of smoothed variables D.E.:

$$\Delta(t) \triangleq \sum_k c_k \cdot \left[\mathscr{H}_{\left(U_k^G(t)\right) - h\left(U_k(t)\right)}\right]$$

- Need Fisher information, which is defined (as trace of below when MIMO)

For a random vector U with density p_U, the Fisher information is defined as

$$J(U) \triangleq \mathbb{E}\left[\|\nabla \log p_U(U)\|^2\right].$$

De Bruijn's Inequality

- Take derivative with respect to the smooth-scaling variable t

$$\frac{d}{dt}\Delta(t) = \frac{1}{2}\sum_k c_k \cdot \left[J(U_k^G(t)) - J(U_k(t))\right],$$

- Gaussian minimizes Fischer Information over all distributions with the same variance.

- Thus, the derivative above is always ≤ 0 and for $t \to \infty$, the Gaussian $\boldsymbol{z}$ dominates, so derivative is 0.
- $\Delta(t)$ always decreases from $t = 0$ to $t \to \infty$, and therefore $\Delta(0) > 0$, which by its definition then makes the Gaussian differential entropy for the sum (even when it averages uncancelled other users) the maximizing distribution.

$$\sum_k c_k \cdot h\left(\sum_i A_{k,i} \cdot X_i^G + N_k\right) \geq \sum_k c_k \cdot h\left(\sum_i A_{k,i} \cdot X_i + N_k\right) \qquad \Psi(p;\{r_a\}) \leq \Psi(p_G;\{r_a\})$$

$$\Phi_{\pi,w}(p) \leq \Phi_{\pi,w}(p_G).$$

Q E D.

- Communication engineers can now use canonical design for all MU Matrix AWGN channels.
  - It is optimum, and the chain-rule, successive-decoding DFE structures are applicable to Gaussian designs,
  - This allows much simplification as we will see. No further theoretical trepidations need hinder progress!

End Supplementary Lecture 7A