Throughput and Delay Optimal Resource Allocation in Multiple Access Fading Channels

DIMACS Network Information Theory Workshop
March 18, 2003

Edmund M. Yeh
Department of Electrical Engineering
Yale University

Joint work with Aaron Cohen, Brown University

Acknowledgments

Many thanks to Professor Robert Gallager and Professor Emre Telatar for their advice and encouragement.

Multiple Access Communications

- Multiple access (many to one): multiple senders transmit to one receiver (possibly) over fading channels.
- Ex: cellular telephony, satellite networks, local area networks.

Central Problems

- Contention/interference - resource sharing.
- Bursty sources $\Rightarrow$ random number of active senders.
- Network/MAC layer QOS issues - throughput, delay.
- Physical layer issues - channel modelling, coding, detection.

Need for Cross-Layer Approach

- Multiple access network theory (ALOHA, CSMA) - concentrates on source burstiness and delay; poor modelling of noise and interference.
- Multiple access information theory - concentrates on channel modelling and coding; ignores random arrival of messages and delay.
- Need more unified **cross-layer** framework:
  - Random packet arrivals affect resource sharing.
  - Choice of modulation and coding affects QOS issues.
  - Random fading affects resource allocation.
  - Gallager (85), Ephremides and Hajek (98).

New Approach

- Goal:
  - Combine information-theoretic limits with QOS issues.
  - Establish fundamental bounds on throughput/delay performance.
- Implementation:
  - Random arrivals, information-theoretic optimal coding.
  - Power control and rate allocation as function of fading and queue states to optimize throughput and delay

Previous Work

- Telatar and Gallager (95)
  - Achievable multiple access scheme with feedback.
  - Poisson arrivals; no queueing; single-user decoding; processor sharing system.
- Telatar (95)
  - Analogy between MAC and multi-processor queue.
  - Each user has **fixed** pool of bits to send.
  - Optimal processor assignment to minimize average packet delay.
- Yeh (01)
  - Poisson arrivals; queueing.
  - Optimal rate allocation from $\mathcal{C}$ to min. average packet delay.
  - Longer Queue Higher Rate (LQHR) policy strongly delay optimal.

Multiple Access Fading Channel

- Continuous-time M-user Gaussian multiple access fading channel with bandwidth W:

$$Y(t) = \sum_{i=1}^{M} \sqrt{H_i(t)} X_i(t) + Z(t).$$

- $\{Z(t)\}$: white Gaussian noise, density $N_0/2$.
- Slowly-varying and flat-fading (under-spread) channel.

Multiple Access Fading Channel

- Block fading model, block length $= T$.
- T large enough for reliable communication at a fixed fade.
- $\{\boldsymbol{H}(t) = (H_1(t), \ldots, H_M(t))\}$ modulated by finite-state ergodic Markov chain.
- Transmitter i has (long-term) average power constraint $\overline{P}_i$, and (short-term) peak power constraint $\hat{P}_i$.

Information-theoretic Capacity Region $\mathcal{C}(\boldsymbol{h}, \boldsymbol{p})$

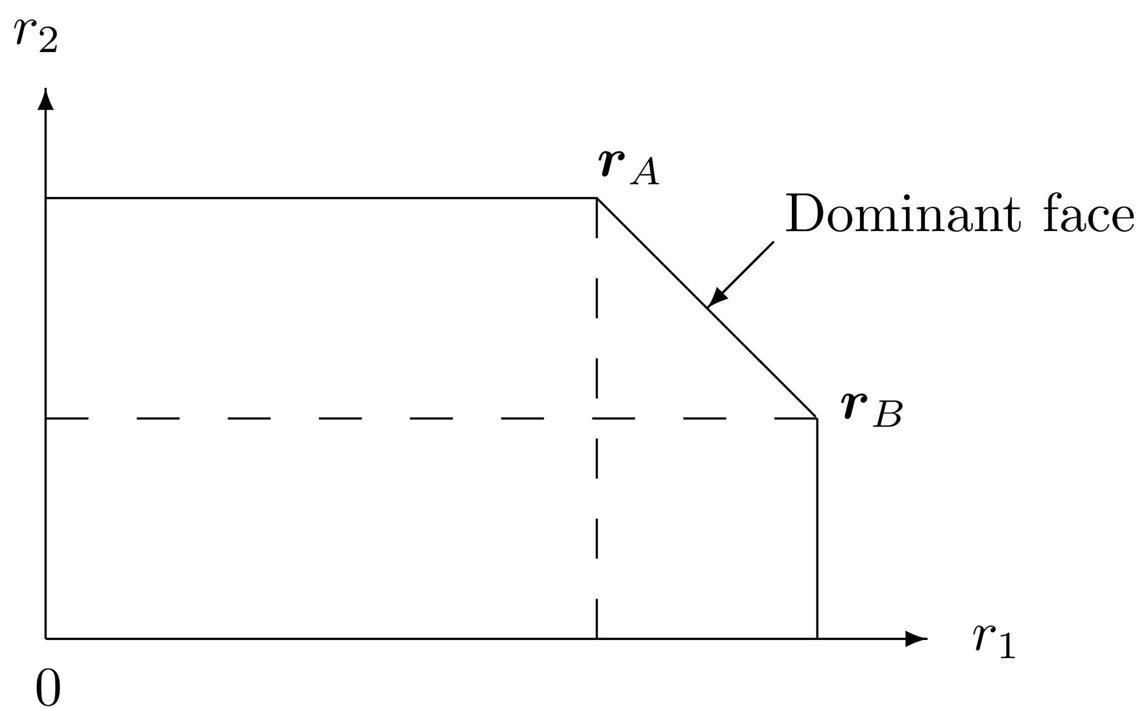
(Ahlswede, Liao, Cover, Wyner 1971-75)

- Fixed $\boldsymbol{h} = (h_1, \ldots, h_M)$ and $\boldsymbol{p} = (p_1, \ldots, p_M)$.
- $\mathcal{C}(\boldsymbol{h}, \boldsymbol{p})$ = set of $\boldsymbol{r} \in \mathbb{R}_+^M$ such that

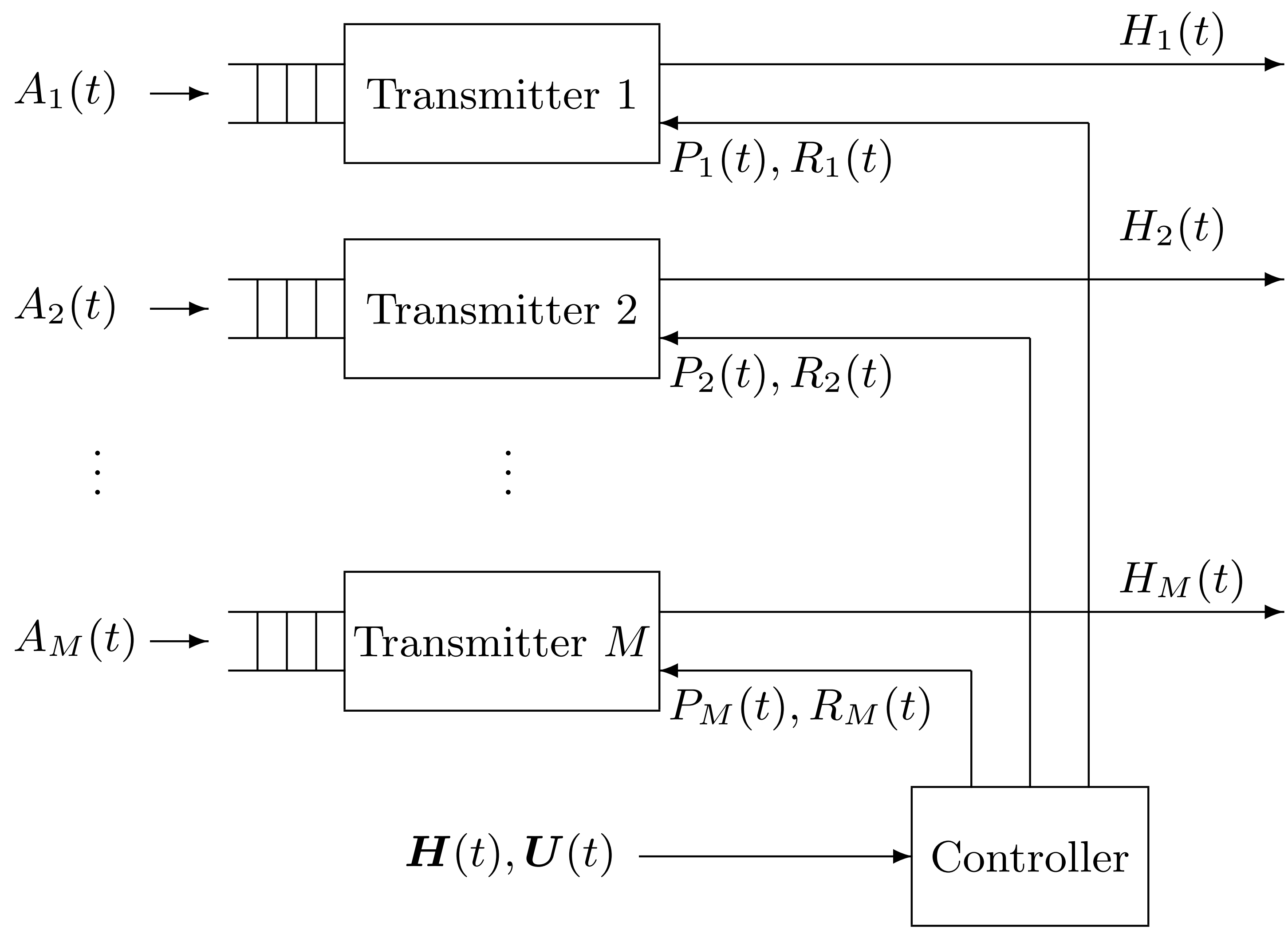
$$\sum_{i \in S} r_i \leq W \log \left(1 + \frac{\sum_{i \in S} h_i p_i}{N_0 W} \right), \ \ \forall S \subseteq \{1, \ldots, M\}.$$

- Reliable communication possible inside $\mathcal{C}(\boldsymbol{h}, \boldsymbol{p})$, impossible outside $\mathcal{C}(\boldsymbol{h}, \boldsymbol{p})$, for any coding and modulation scheme.
- **Polymatroid** structure (Tse and Hanly 98).

Two-User Capacity Region $\mathcal{C}(\boldsymbol{h}, \boldsymbol{p})$



Multiple Access Channel with Random Arrivals



Arrivals and Unfinished Work

- $\{A_i(t)\}$ = ergodic packet arrival process to transmitter i.
- User i packets i.i.d. $\sim F_{Z_i}(\cdot), \mathsf{E}[Z_i] < \infty$.
- $U_i(t)$ = number of untransmitted bits in queue i at time t.

Power Control and Rate Allocation

- Controller: $(\boldsymbol{H}(t), \boldsymbol{U}(t)) \mapsto (\boldsymbol{P}(t), \boldsymbol{R}(t))$.
- Two stages:

1. **Power control policy** $\mathcal{P}$:

$$\boldsymbol{p} = \mathcal{P}(\boldsymbol{h}, \boldsymbol{u})$$

   s.t. for all i, $\mathsf{E}[\mathcal{P}_i(\boldsymbol{H}, \boldsymbol{U})] \leq \bar{P}_i$, $\mathcal{P}_i(\boldsymbol{h}, \boldsymbol{u}) \leq \hat{P}_i$ for all $(\boldsymbol{h}, \boldsymbol{u})$.

2. **Rate allocation policy** $\mathcal{R}$:

$$\boldsymbol{r} = \mathcal{R}(\boldsymbol{h}, \boldsymbol{p}, \boldsymbol{u}) \in \mathcal{C}(\boldsymbol{h}, \boldsymbol{p}).$$

Main Results

- Stability region $\mathcal{S}$ of all bit arrival rates for which all queues can be kept finite.
- For given power control policy, find throughput optimal rate allocation policy.
- In symmetric scenario, find delay optimal rate allocation policy for any symmetric power control policy.

Stability Region $\mathcal{S}$

- $\lambda_i = \lim_{t\to\infty} A_i(t)/t$ = packet arrival rate to queue i.
- $\rho_i = \lambda_i \mathsf{E}[Z_i]$ = bit arrival rate to queue i.
- Define $f_i(\xi) = \limsup_{t\to\infty} \frac{1}{t}\int_0^t 1_{\{U_i(\tau)>\xi\}} d\tau$.
- System **stable** if $f_i(\xi) \to 0$ as $\xi \to \infty$ for all i.
- $\mathcal{S}$ = set of all $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_M)$ for which can stabilize system.

Stability Region $\mathcal{S}$

- Assume $\{A_i(t)\}$ modulated by finite-state ergodic Markov chain.

Theorem 1 $\mathcal{S} = \mathcal{C}(\overline{\boldsymbol{P}}, \hat{\boldsymbol{P}})$ = *information-theoretic capacity region under power control (Tse and Hanly 98).*

- $\mathcal{C}(\overline{\boldsymbol{P}}, \hat{\boldsymbol{P}}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \mathcal{C}(\mathcal{P})$.
- $\mathcal{F} = \{\mathcal{P} : \mathsf{E}[\mathcal{P}_i(\boldsymbol{H})] \leq \overline{P}_i, \forall i; \mathcal{P}_i(\boldsymbol{h}) \leq \hat{P}_i, \forall \boldsymbol{h}, \forall i\}$.
- $\mathcal{C}(\mathcal{P}) = \mathsf{E}[\mathcal{C}(\boldsymbol{H}, \mathcal{P}(\boldsymbol{H}))]$.

Stability Theorem

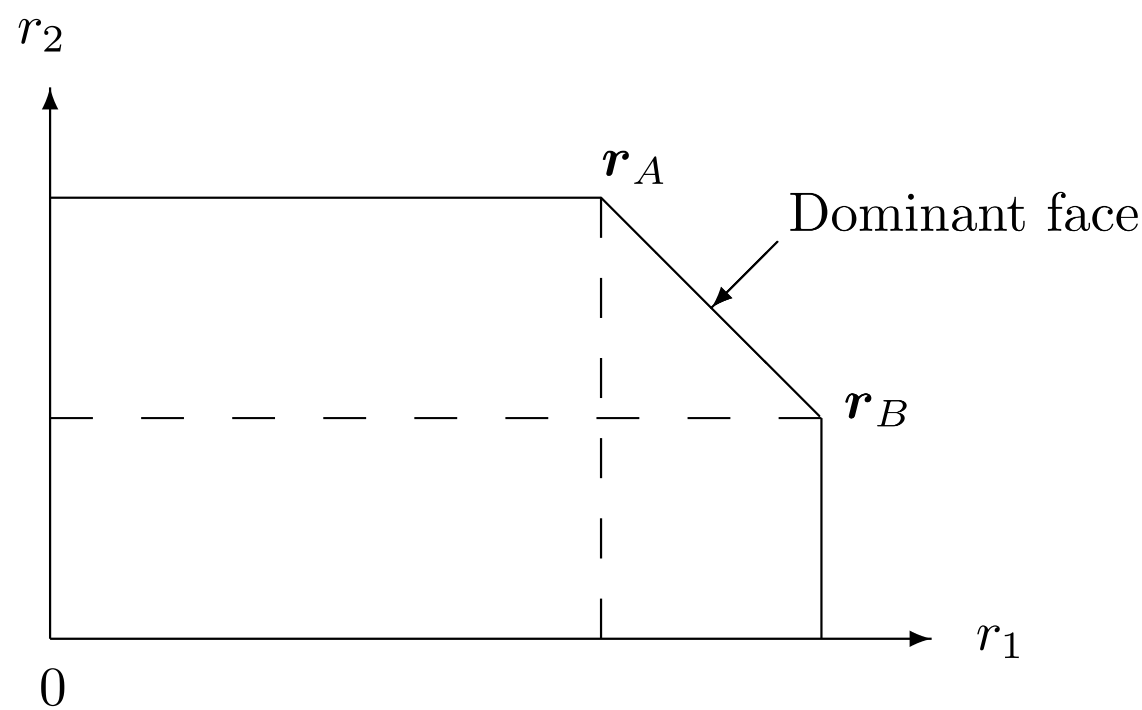
- **Achievability**: $\boldsymbol{\rho} \in \text{int}(\mathcal{S})$: knowing $\boldsymbol{\rho}$ and statistics of $\{\boldsymbol{H}(t)\}$, can stabilize system using stationary $\mathcal{P}, \mathcal{R}$ depending only on current channel state.
- **Converse:** $\boldsymbol{\rho} \notin \mathcal{S}$: cannot stabilize system, even with non-stationary policy with knowledge of queue state and/or knowledge of future events, so long as

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t p_i(\tau) d\tau \leq \overline{P}_i \ \forall i; \quad p_i(\tau) \leq \hat{P}_i, \forall \tau, \forall i.$$

Throughput Optimal Resource Allocation

- Find "universal" power/rate policy to stabilize system even if $\boldsymbol{\rho}$ not known, as long as $\boldsymbol{\rho} \in \mathrm{int}(\mathcal{S})$.
- Must use both $\boldsymbol{H}(t)$ and $\boldsymbol{U}(t)$.
- Suppose know $\boldsymbol{\rho} \in \mathcal{C}(\mathcal{P}) = \mathsf{E}[\mathcal{C}(\boldsymbol{H}, \mathcal{P}(\boldsymbol{H}))]$.
- Assume $\{H_i(kT)\}$ i.i.d. for each i, $\{A_i((k+1)T) - A_i(kT)\}$ i.i.d. for each i.
- Assume $\mathsf{E}[(A_i((k+1)T) - A_i(kT))^2] < \infty$.

No Work Conservation



Throughput Optimal Rate Allocation

Theorem 2 *Given $\mathcal{P} \in \mathcal{F}$, throughput optimal rate allocation policy is*

$$\boldsymbol{r}^* = \mathcal{R}^*(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}, \boldsymbol{u}), \boldsymbol{u}) = \arg \max_{\boldsymbol{r} \in \mathcal{C}(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}, \boldsymbol{u}))} \sum_{i=1}^{M} u_i r_i \qquad (1)$$

- Idea appeared in Tassiulas and Ephremides '92; McKeown, et al. '96; Tassiulas '97; Neely et al. '02.
- Here, motivated by delay optimality results.

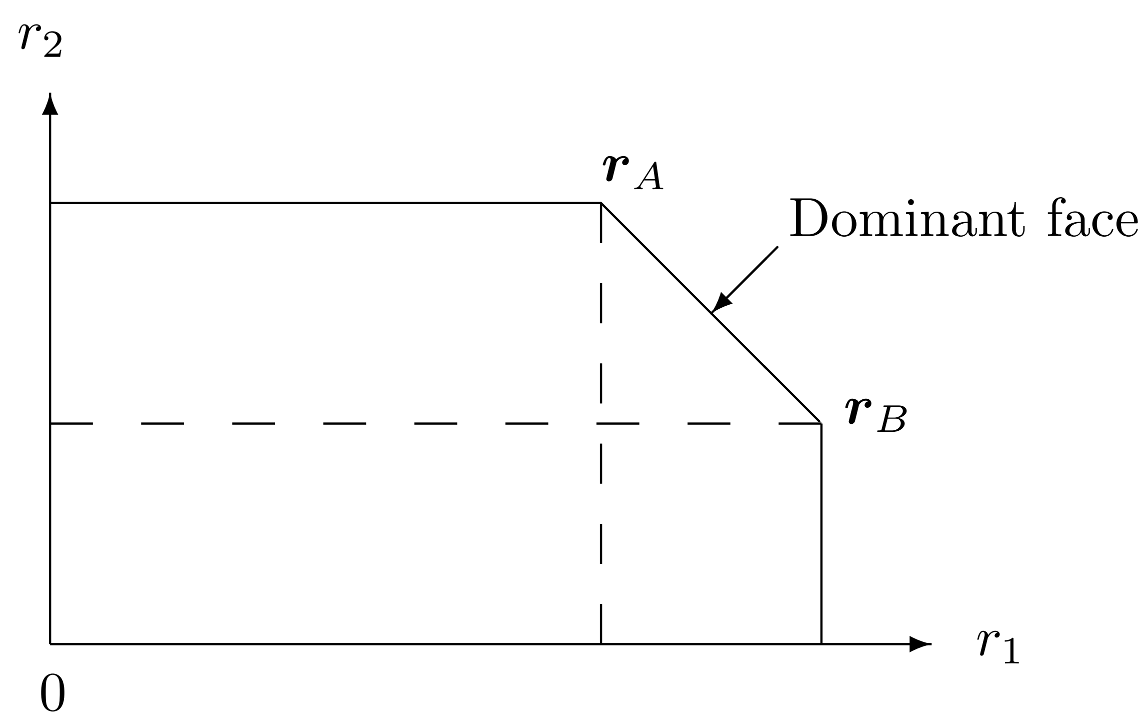
Longest Queue receives Highest Possible Rate (LQHPR)

- Due to **polymatroidal** nature of $\mathcal{C}(\boldsymbol{h}, \mathcal{P}(\boldsymbol{h}, \boldsymbol{u}))$, solution to (1) has special form.
- Order queues $u_{[1]} \geq u_{[2]} \geq \cdots \geq u_{[M]}$.

$$r^*_{[i]} = W \log \left(1 + \frac{h_{[i]} \mathcal{P}_{[i]}(\boldsymbol{h}, \boldsymbol{u})}{\sum_{j<i} h_{[j]}(t) \mathcal{P}_{[j]}(\boldsymbol{h}, \boldsymbol{u}) + N_0 W} \right)$$

- Longest Queue receives Highest Possible Rate (LQHPR).
- LQHPR $\Leftrightarrow$ **adaptive successive decoding**: $u_{[M]}$ decoded first, $u_{[1]}$ decoded last.

Two-User Rate Allocation



- $u_1 \geq u_2 : \boldsymbol{r}_B$ $\qquad$ $u_1 < u_2 : \boldsymbol{r}_A$

Proof of Stability Theorem

- Stability of Markov chains based on negative Lyapunov drift.
- $V(\boldsymbol{U}) = \sum_i U_i^2$.
- Show there exists compact set $\Gamma \subset \mathbb{R}^M$ s.t. for some $\epsilon > 0$,

$$\mathsf{E}[V(\boldsymbol{U}(t+T)) - V(\boldsymbol{U}(t))|\boldsymbol{U}(t)] \leq -\epsilon$$

whenever $\boldsymbol{U} \notin \Gamma$.

Delay Optimal Resource Allocation

- Beyond stabilization, keep queues as short as possible.
- Find feasible $\mathcal{P}$ and $\mathcal{R}$ to minimize $\lim_{t\to\infty} \mathsf{E}[\sum_{i=1}^{M} U_i(t)]$ (average bit delay) for $\boldsymbol{\rho} \in \mathrm{int}(\mathcal{S})$.

Delay Optimal Rate Allocation

- Focus on symmetric Poisson/exponential case.
- $\{A_i(t)\} = \text{Poisson}(\lambda)$ for each i.
- All packets i.i.d. $\sim \exp(\mu)$.
- Queue state $\boldsymbol{Q}(t) = (Q_1(t), \ldots, Q_M(t))$ - number of packets.
- For fixed $\mathcal{P}$, find $\mathcal{R}$ to minimize $\lim_{t\to\infty} \mathsf{E}\left[\sum_{i=1}^{M} Q_i(t)\right]$ (average packet delay).
- Yeh '01: non-faded symmetric MAC.

Delay Optimal Rate Allocation

- Symmetric fading process $\boldsymbol{H}(t)$:
  For any $\boldsymbol{a} = (a_1, \ldots, a_M)$, $\Pr(H_1(t) = a_1, \ldots, H_M(t) = a_M) = \Pr\left(H_1(t) = a_{\pi(1)}, \ldots, H_M(t) = a_{\pi(M)}\right)$ for any permutation π.
  e.g. for every t, $H_1(t), \ldots, H_M(t)$ i.i.d.
- Symmetric power control $\mathcal{P}(\boldsymbol{h}, \boldsymbol{q}) = \mathcal{P}(\boldsymbol{h})$:
  $\mathcal{P}_i(a_1, \ldots, a_M) = \mathcal{P}_{\pi^{-1}(i)}\left(a_{\pi(1)}, \ldots, a_{\pi(M)}\right)$ for all π.
  e.g. $M = 2$ and $a_1 > a_2$: $\mathcal{P}_1(a_1, a_2) = \mathcal{P}_2(a_2, a_1)$.
  e.g. Knopp and Humblet ('95).
- For this case, $\max \sum u_i r_i$ (LQHPR) policy is **delay optimal**.

Majorization and Weak Majorization

- Need to quantify load balancing.
- For $\boldsymbol{u} \in \mathbb{R}^M$, let

$$u_{[1]} \geq \cdots \geq u_{[M]}.$$

- For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^M$,

$$\boldsymbol{u} \prec_w \boldsymbol{v} \quad \text{if} \quad \sum_{i=1}^{k} u_{[i]} \leq \sum_{i=1}^{k} v_{[i]}, \; k = 1, \ldots, M.$$

  Say $\boldsymbol{u}$ **weakly majorized** by $\boldsymbol{v}$. If equality holds for $k = M$, say $\boldsymbol{u}$ **majorized** by $\boldsymbol{v}$: $\boldsymbol{u} \prec \boldsymbol{v}$.

- Ex: $(1\ 1) \prec_w (3\ 0)$, $(1\ 1) \prec (2\ 0)$.
- See Marshall and Olkin (79).

Stochastic Weak Majorization

- Use **stochastic coupling** to show weak majorization on queue vectors, in a **stochastic** sense.
- $\boldsymbol{U} = (U_1, \ldots, U_M)$, $\boldsymbol{V} = (V_1, \ldots, V_M)$ random vectors. $\boldsymbol{U}$ is said to be **stochastically weak-majorized** by $\boldsymbol{V}$, $\boldsymbol{U} \prec_w^{st} \boldsymbol{V}$, if there exist random vectors $\tilde{\boldsymbol{U}}$ and $\tilde{\boldsymbol{V}}$ such that

(a) $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ are identically distributed.

(b) $\boldsymbol{V}$ and $\tilde{\boldsymbol{V}}$ are identically distributed.

(c) $\tilde{\boldsymbol{U}} \prec_w \tilde{\boldsymbol{V}}$ a.s.

Strong Delay Optimality of LQHPR

Theorem 3 *Let* $\boldsymbol{q}_0$ *be initial queue state. Let* $\boldsymbol{Q}(t)$ *be queue evolution under* g_{LQHPR} *for* $t \geq 0$*. Let* $\boldsymbol{Q}'(t)$ *be corresponding quantity under any policy* $g \in G_{\mathcal{D}}$*. Then under all symmetric* $\mathcal{P}$*,*

$$\boldsymbol{Q}(t) \prec_w^{st} \boldsymbol{Q}'(t) \;\; \forall t \geq 0.$$

- Proof: generalize stochastic coupling argument for non-faded symmetric MAC.

Consequences

Corollary 1

$$\mathsf{E}\left[\varphi(\boldsymbol{Q}(t))\right] \leq \mathsf{E}\left[\varphi(\boldsymbol{Q}'(t))\right] \quad \forall t \geq 0$$

for all $\prec_w$-preserving $\varphi : \mathbb{R}^M \mapsto \mathbb{R}$ for which expectations exist.

- φ is $\prec_w$-**preserving** if $\boldsymbol{x} \prec_w \boldsymbol{y} \Rightarrow \varphi(\boldsymbol{x}) \leq \varphi(\boldsymbol{y})$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^M$.
- $\prec_w$-preserving $\Leftrightarrow$ Schur-convex, increasing.
- Includes all symmetric, convex and increasing real functions on $\mathbb{R}^M$.
- Examples:
  $\varphi(\boldsymbol{x}) = \max_{i_1 < i_2 < \cdots < i_k}(|x_{i_1}| + \cdots + |x_{i_k}|)$, $1 \leq k \leq M$;
  $\varphi(\boldsymbol{x}) = \sum_{i=1}^{M} |x_i|^r$ for $r \geq 1$ or $r \leq 0$;
  $\varphi(\boldsymbol{x}) = (\sum_{i=1}^{M} |x_i|^r)^{1/r}$ for $r \geq 1$.

Summary and Conclusions

- General framework for resource allocation in fading MAC with random arrivals.
- Stability region $\mathcal{S} = \mathcal{C}(\overline{\boldsymbol{P}}, \hat{\boldsymbol{P}})$.
- $\max \sum_i u_i r_i$ (LQPHR) policy throughput optimal for given $\mathcal{P}$.
- LQHPR minimizes average packet delay for any symmetric $\mathcal{P}$ in symmetric scenario.
- LQHPR implements adaptive successive decoding at physical layer.

Summary and Conclusions

- "Converse": LQHPR establishes **fundamental throughput/delay performance limit** for any multiple access coding scheme which meets any given required P_e (Fano).
- "Achievability": To approach rates in $\mathcal{D}$, need sufficiently large T and code over large number of bits.