Cross-Layer Design for Wireless Communications

Philippe Mary

IETR UMR 6164/ INSA Rennes
20 av. des Buttes de Coësmes, CS 70839, F-35708 Rennes.

Contact

- IETR : Bâtiment 6
- Bureau : 006
- Email : philippe.mary@insa-rennes.fr
- website : http ://pmary.perso.insa-rennes.fr

Schedule

- Lecture class : 10h
- Lecture slot (exercises included) : 2h
- Evaluation :
  - 1 Test duration 2h scheduled on the 02/02/2015 at 10h00

Outline

- Part I : Layered architectures
- Part II : PHY/MAC resource allocation and system models
- Part III : A cautionary perspective of cross-layer design

Bibliography

- These lecture notes have been produced with the following materials
  - Ana I. Pérez-Neira, Marc Realp Campalans, "Cross-Layer Resource Allocation in Wireless Communications", *Elsevier*,
  - Marvin K. Simon, Mohamed Slim-Alouini, "Digital Communication Over Fading Channels", *Wiley*,
  - Jean-Marie Gorce, "Les technologies 802.11 pour les réseaux sans-fil" *INSA de Lyon*.
  - David Tse, Pramod Viswanath "Fundamentals of Wireless Communications"
  - several research articles

Aim of the course

- New doors on wireless system designs
- Focus to radio systems
- Cross-layer => Interaction between several layers (OSI)
  - PHY Layer (physical layer)
  - MAC Layer (medium access control layer)
  - Network Layer
- What we need to know
  - Difference between a cross-layer architecture and a layered architecture
  - Wireless systems characteristics => multi-users, interference
    - PHY : noise, fading, interference,
    - MAC : channel access, congestion management (interference)
    - Network : routing, high level interference (Network architecture dependent)

Introduction

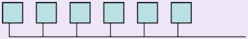
- Definition : A communication network is a set of electronic devices (nodes) connected to each other following some rules (routing)
- Two important characteristics for a network
  - Transmission technology
    - Broadcasting : TV num, Radio
    - Point to point network (communication networks : Cellular, sensor networks, ...)
  - Scale
    - BAN (Body Area Networks)
    - PAN (Personal Area Networks) - Office
    - LAN (Local Area Networks) - building, campus
    - MAN (Metropolitan Area Networks) - Town
    - WAN (Wide Area Networks) - Country, continent
    - Internet - Planet

Communication network

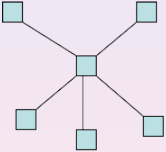
- Every packet has an unique or a finite number of destinations
- Direct link (1 hop) : A radio link exists (?) between two nodes ;
- Multi-hop communications : the source cannot reach the destination in one hop
  - A specific route is needed (Which criteria ?)
  - Depends on the cost function (distance, energy, latency, loss probability, ...)
- Routing algorithm
  - pro-active
  - reactive

Topology

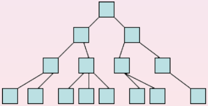
- Bus : Wired network of computer



- Star : Cellular, WLAN, etc.

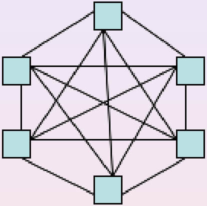


- Tree : hierarchical architecture

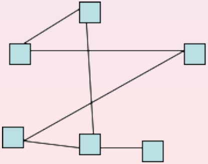


Topology (cc'ed)

- Ad-hoc network : connected nodes without an infrastructure
  - Fully connected



  - Partially connected

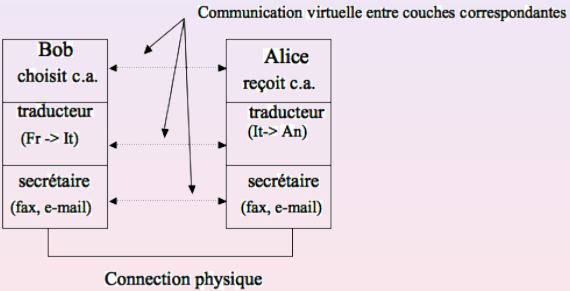


Layered Protocol Architecture

- The networks are built according layers above to each other
- "Divide and Conquer" approach
  - Separate the functionalities => reduced complexity
- Aim of each layer : provides services to the upper layers whilst hiding implementation issues for the upper layers.
- Definition : the set of layers and protocols is called network architecture

The birthday card

- Bob (French) wishes sending a birthday card to Alice (American), and uses an agency for the translation



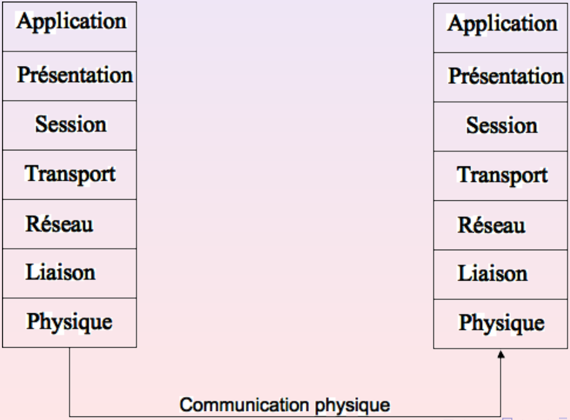
- The layer *n* communicates with the associated layer by a set of rules called layer *n* protocol

Functionalities of the birthday card

- layer 1 protocol : fax
  - Dealt with the same layer at the destination
  - Can be changed to e-mail
- layer 2 protocol : intermediate translate language
  - italian can be switched in chinese or spanish
- Each layer adds informations which can be read only by the same layer at the destination
  - addition of header and/or tailer to packets
- Important properties for a layered architecture :
  - Each layer performs a set of functions
  - The information flow between layers must be minimized
  - How many layers should we design ?
    - few => too many functionalities per layer
    - a lot => complex architecture

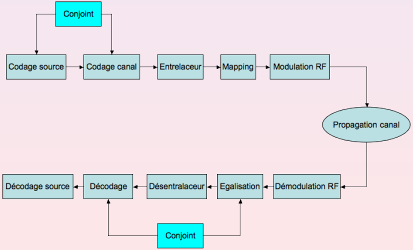
OSI model

- OSI : Open Systems Interconnection
  - Alternate models e.g. TCP/IP
- 7 different layers



OSI layers : PHY layer

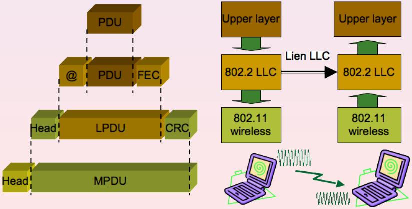
- Function : information transmission (bits) in a noisy channel
- Main design functions :
  - representation of "0" and "1"
  - bit duration
  - transmission type (simplex, duplex)
  - how to start/finish a connection
  - source/channel coding, etc.



OSI layers : Data link layer

Can be separated in two sub-layer, i.e. Logical Link Layer (LLC) and Medium Access Control (MAC)

- First error-free layer
- LLC
  - Interface between the MAC and the network layer (3)
  - Data exchange between users in a LAN
  - Data link layer error control : retransmission of packets (ARQ)
  - The above layer is waiting for an error-free transmission service (i.e. TCP/IP)



OSI layers : Data link layer

The MAC layer manages the medium access between users

- Interface between the LLC and the PHY layer
- Frame addressing (physical address)
- Allow to be associated to a network
- Medium access manager
  - Avoid and detect collisions (CSMA/CD, ex : Ethernet)
  - Wireless channel : no collision detection => CSMA/CA (WLAN)
  - ALOHA protocol
  - Resource allocation in centralized networks (TDMA, FDMA, CDMA, etc.)

OSI layers : Network layer

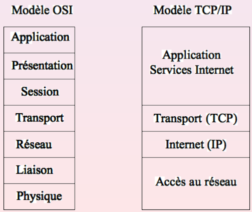
- Routing packets from a source to a destination
- Universal addressing, e.g. IPv4
- Congestion control (too many request on a particular node)
- Connected mode : a link should be established before sending a packet between two nodes. Packet order saved
- Non connected mode : "postal card" mode => every packet carries the destination address ; independent routing
- Commutation (connected) : one route is computed once for all
- Routing (non connected) : a route is established for each packet

OSI Layers : Transport Layer

- The last layer before the applications layers
- Reliable end-to-end information delivery
- Main function : rearrange in small packets the data from the session layer and ensure that they are delivered at the destination
  - Control Mechanism of error and flow
- Provides services for connected and non-connected mode
- Connected mode => Packets can be rearranged (TCP/IP)
- TCP can be included in the Transport Layer

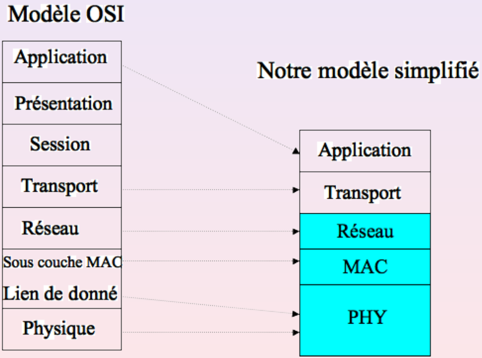
OSI model : Applications Layers

- Session Layer (5) : Advanced services, dialogues management. Turns on, manages and finishes the remote connections
- Presentation Layer (6) : Convert in bytes the syntax and semantic of the application layer for the other layers of the OSI stack
- Application Layer (7) : The closest to the user. Interacts with the application software



Simplified OSI model

- Only the first three layers will be considered



Advantages of layered architectures

Modularity

- Simplicity
- Easy debugging
- Standardization
- Development time reduced for the new protocols

Is there any inconvenient with the layered architecture ?

- Assumption : the layers can be optimized independently from each other
  - Limit in the wireless case ?
  - Is it efficient ?
  - What is the other solutions ?
- Do not forget : a wireless network does not include the links !

Cross-layer design

- Advantages
  - Exploit the interaction between the layers
  - In wireless : Dependency between layers and their functionalities
- Disadvantages
  - The interactions between protocols from several layers are difficult to model
  - Probably lead to complex algorithms
  - Is the modularity conserved ?
- X-layer design : understanding and exploitation of the interactions between layers

Cross-Layer vs Layered

- Should the OSI model be rejected if we talk about x-layer ?
- Do we need of a network architecture ?
- Is the x-layer design suitable for any type of networks and all kind of application ?

The layered architecture cannot be rejected and all layers cannot be optimized together without any frontiers

- "spaghetti code"
- No way to make it sustainable

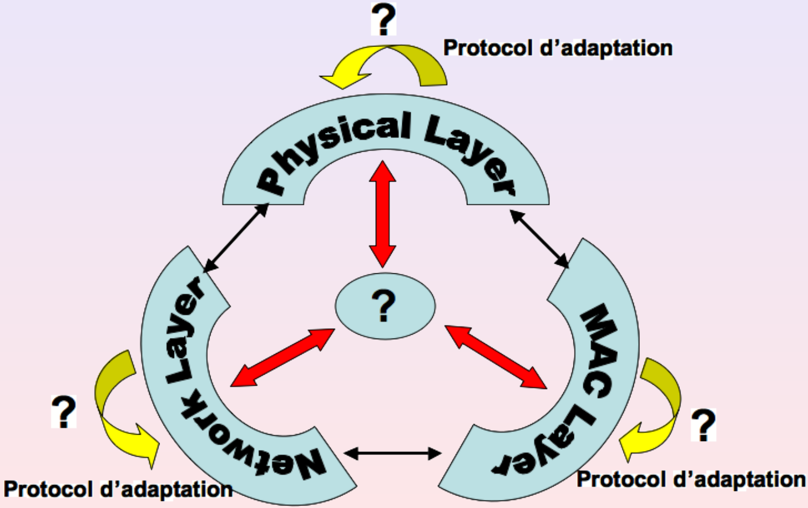
Solution => Holistic view of wireless networks

Cross-layer design

Let us focus on the first three layers

- Derive an abstract model for each layer and QoS criteria
- Design adaptive protocols at each level
- What are the needed exchanges between each pair of layers
- Obtain a tradeoff between performances, complexity and scaling

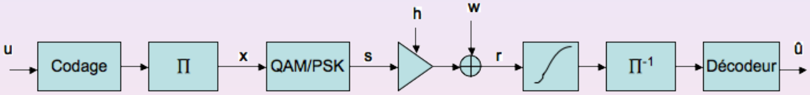
Cross-layer design



Design criteria

- Several QoS criteria according to the layer
- PHY
  - BER, capacity/throughput, PER, SINR
- MAC
  - Channel access delay, throughput without error
- Network
  - Delay, throughput, rejection probability
- Other criteria
  - Energy consumption (Devices / network life duration)
  - Utility of a user
  - => Every layer concerned

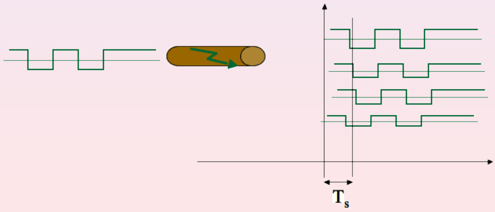
Capacity



- The output signal is : $r = h \cdot s + w$ et $w \sim \mathcal{CN}\left(0, \sigma^2\right)$
- The capacity of the channel is $C\left(\gamma\right) = \log_2\left(1+\gamma\right)$ bits/s/Hz

Wireless link characteristics

- QoS PHY criteria : BER
- BER depends on the received power (fading, selective channels, ...), coding, components, etc.
- Channel : Fading + shadowing + pathloss.
  $P_r = P_t \cdot \mathrm{PL} \cdot \alpha_{\mathrm{shad}} \cdot |h|^2$
- Multi-user $\Longrightarrow$ interference
- Flat Fading : $\tau_{\max} < T_s$
  - Random summation of many paths



Bit Error Rate (BER) in AWGN

- Important criterium for a non-coded modulation scheme (M-QAM, M-PSK)
- Exact expression for M-QAM and M-PSK quite hard to obtain => approximations

$$BER_{MQAM}(\gamma) \approx 4\left(1-\frac{1}{\sqrt{M}}\right)\frac{1}{\log_2 M}\sum_{i=0}^{\sqrt{M}/2-1} Q\left((2i+1)\sqrt{\frac{3\gamma}{M-1}}\right)$$

$$BER_{MPSK}(\gamma) \approx \frac{2}{\max\left(\log_2 M, 2\right)}\sum_{i=1}^{\max(M/4,1)} Q\left(\sin\frac{(2i-1)\pi}{M}\sqrt{2\gamma}\right)$$

Instantaneous and average SNR

- Base band and flat fading channel : $y(t) = h(t)x(t) + n(t)$
  - $h(t)$ is a single complex coefficient (gaussian for instance) and constant over one symbol : $h(t) = \alpha(t)e^{j\theta(t)}$
- instantaneous SNR (over one symbol) :

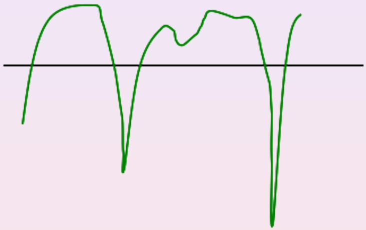
$$\gamma(t) = \frac{E\left[|h(t)x(t)|^2\right]}{E\left[|n(t)|^2\right]} = \alpha(t)^2 \frac{E_s}{N_0}$$

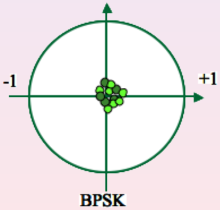
- Average SNR :

$$\overline{\gamma}(t) = E\left[\gamma(t)\right] = E\left[\alpha(t)^2\right] \frac{E_s}{N_0} = \Omega \frac{E_s}{N_0}$$

Rayleigh channel

- NLOS channel : summation of random variables with random phase and module

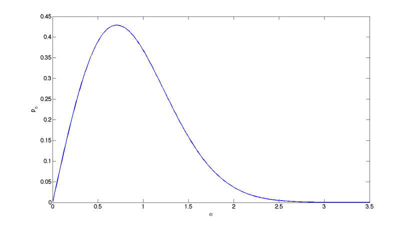




Rayleigh Fading

- The channel magnitude modelized by a Rayleigh law

$$p_\alpha(\alpha) = \begin{cases} \frac{2\alpha}{\Omega}\exp\left(-\frac{\alpha^2}{\Omega}\right) & \alpha \geq 0 \\ 0 & \textit{elsewhere} \end{cases}$$

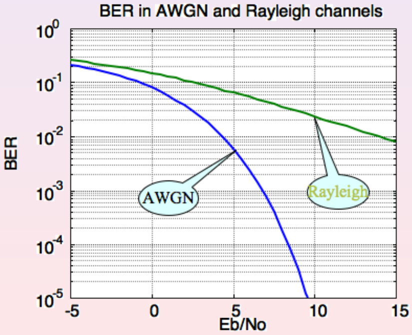


- The law of the SNR is $p_\gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}}$

Fading effect on the BER

- The instantaneous SNR is time varying=> error probability also :

$$P_b(E) = \int_0^\infty P_b(E|\gamma)\, p_\gamma(\gamma)\, d\gamma = \frac{1}{2}\left[1 - \sqrt{\frac{\overline{\gamma}}{1+\overline{\gamma}}}\right]$$



Packet Error Rate (PER)

- Real systems : mapping + modulation and coding scheme (MCS)
- Effective transmission rate (information) : $R^{(m)} = b^{(m)} c^{(m)}$
- Decoded bits are passed to the MAC/Link
    - Received packet at PHY with a non zero PER. Depends on m (MCS) and $\gamma \Rightarrow PER^{(m)}(\gamma)$
    - Transmission probability : $PSR^{(m)}(\gamma) = 1 - PER^{(m)}(\gamma)$

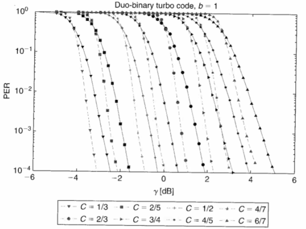


Figure 2.3 *PER curves for BPSK. Continuous lines: k = 288, dashed lines: k = 1152.*

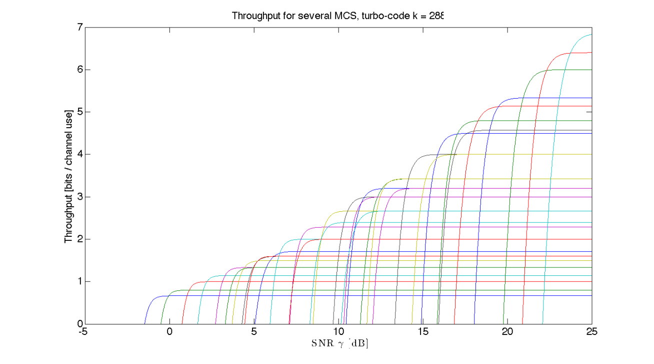
MAC/Link Throughput

- Shannon capacity : Optimal MCS without errors
- Transmission errors => $PER \neq 0$
- Amount of data without errors delivered from MAC/Link to the above layers at t :

$$\eta^{(m)}(\gamma) = R^{(m)}\mathbf{1}\{Z_i = 1\} \text{ bits / channel use}$$

  - Instantaneous value varying in time (even though γ constant !)
- Throughput : $\overline{\eta}^{(m)}(\gamma) = E_t\left[\eta^{(m)}(\gamma)\right] = R^{(m)}PSR^{(m)}(\gamma)$

Throughput vs MCS



- Not all the MCS are interesting !

Link adaptation

- System abilities to match MCS, power to the radio link quality
  - Radio link quality : SNR (SINR with interference)
- MCS adaptation to optimize a cost function (e.g. BER, throughput, etc.)
- Cross-layer => throughput maximization MAC/Link (e.g. HSDPA)
- CSI (Channel State Information) is needed at the transmitter
  - TDD system => Symmetry on the channel

Link Adaptation (cc'ed)

- Ideal feedback channel + SNR known at Tx

$$\overline{\eta}^{(m)}(\gamma) = R^{(m^*)} PSR^{(m^*)}(\gamma) \quad (1)$$

- The optimal MCS is given by :

$$m^* = \arg\max_m R^{(m)} PSR^{(m)}(\gamma)$$

- Assumption : $R^{(m-1)} < R^{(m)} < R^{(m+1)}$, $\exists\ \gamma_{th}^{(m)}$ such as :

$$R^{(m-1)} PSR^{(m-1)}\left(\gamma_{th}^{(m)}\right) = R^{(m)} PSR^{(m)}\left(\gamma_{th}^{(m)}\right) \quad (2)$$

- Optimal MCS belongs to : $\gamma \in \left[\gamma_{th}^{(m)}, \gamma_{th}^{(m+1)}\right)$

Link Adaptation (cc'ed)

- The SNR threshold is the level where the throughput curves cross, eq. (2)

Table 2.2 *SNR thresholds for throughput maximization.*

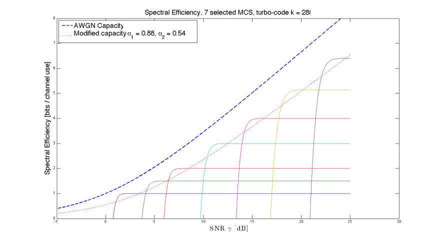
m	1	2	3	4	5	6	7	8
MCS$(b^{(m)}, c^{(m)})$	$\left(1, \frac{1}{3}\right)$	$\left(2, \frac{1}{2}\right)$	$\left(2, \frac{3}{4}\right)$	$\left(4, \frac{1}{2}\right)$	$\left(4, \frac{3}{4}\right)$	$\left(6, \frac{2}{3}\right)$	$\left(6, \frac{6}{7}\right)$	$\left(8, \frac{4}{5}\right)$
Rate $R^{(m)}$	0.33	1	1.5	2	3	4	5.14	6.4
$k = 288, \gamma_{\text{th}}^{(m)}$		0.7 dB	4.2 dB	6.6 dB	10.3 dB	14.1 dB	17.9 dB	22 dB
$k = 1152, \gamma_{\text{th}}^{(m)}$		0.9 dB	4.1 dB	6.5 dB	10.2 dB	13.9 dB	17.7 dB	21.8 dB

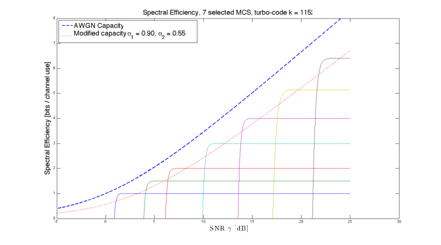
Analytical approximation of the throughput envelope

- Equation (1) is the throughput envelope
- It can be approached empirically :

$$\overline{\eta}\left(\gamma\right) = \alpha_1 \log_2\left(1 + \alpha_2\gamma\right)$$

- α_1, α_2 depends on the MCS number





Link Adaptation with QoS constraint

- VoIP, streaming => transmission with QoS certified
- PHY, MAC/Link => PSR minimum, minimum throughput, délai max
- Taking into account the QoS, eq. (1) becomes :

$$\overline{\eta}^{(m)}(\gamma) = R^{(m^*)} PSR^{(m^*)}(\gamma)$$

- with : $m^* = \arg\max_m R^{(m)} PSR^{(m)}(\gamma)$
- s.t. $PSR^{(m)}(\gamma) \geq PSR_{QoS} \Longleftrightarrow \gamma \geq \gamma_{QoS}^{(m)}$

Ergodic Spectral Efficiency

- Time varying radio channel
- If the fading is ergodic => ergodic capacity

$$\overline{C} = E_\gamma \left[\log_2 (1 + \gamma)\right]$$

- Sufficiently long code words in order to encompass all fading states (true if fast fading)
- Slow fading channel => outage probability
- Similar definition for the average throughput

$$\overline{\eta} = E_\gamma \left[R^{(m^*)} PSR^{(m^*)} (\gamma)\right]$$

Average Throughput in Rayleigh fading

- Fading => smooth throughput compared to the gaussian case

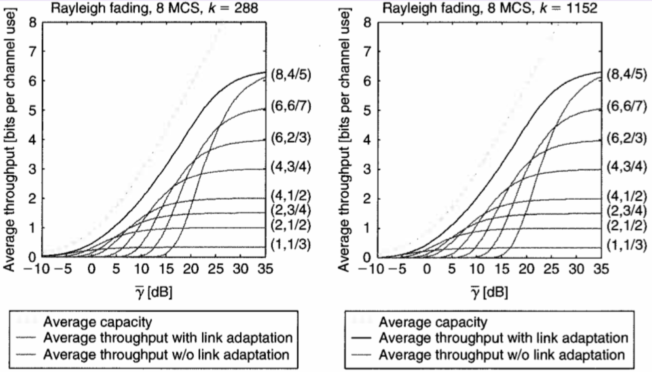


Figure 2.9 *Average throughput in Rayleigh fading channel with link adaptation (8 MCS) and without link adaptation.*

problem statement

- Spectral efficiency region : set of achievable data rate
- Tradeoff between the individual spectral efficiency of each user in competition for the channel access
- Resource allocation : scheduling strategies, prioritizing policy, physical resource management between users
- What physical resources can be shared between users ?
    - Nature gives us four degrees of freedom : time, frequency, power and space... is that all ? (to be cc'ed)

Signal model for multi-user SIMO multiple access channel

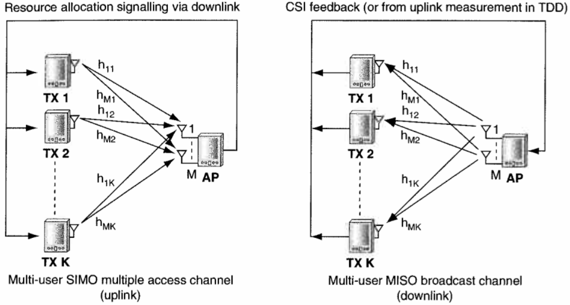


Figure 3.1 *Multi-user SIMO multiple access (left) and MISO broadcast (right) channels.*

- Aim : optimize the achievable rate region under constraints

$$r = Hs + w$$

Signal model (cc'ed)

- AP : processing by $V \in \mathbb{C}^{M \times K}$ et $y = V^t Hs + z$
- the $k-$th entry of the received signal is :

$$y_k = v_k^t h_k s_k + \sum_{k' \neq k} v_k^t h_{k'} s_{k'} + z_k$$

- The SINR of the user k is :

$$\gamma_k = \frac{\left| v_k^t h_k \right|^2 p_k}{\sum_{\substack{k' \in \mathbb{K} \\ k' \neq k}} \left| v_k^t h_{k'} \right|^2 p_{k'} + \sigma^2}$$

- $\mathbb{K} = \{k \mid k \in \{1, \cdots, K\} ; p_k > 0\}$

Successive Interference Cancelation (SIC)

- If all users are decoded in the same time
  - *near-far effect* problem
  - Interferer like noise => suboptimal
- SIC receiver : decode each entry of **s** by removing all decoded entries from the received signal
  - Iterative receiver
  - Decode users by decreasing power order
- SIC decoding order : $\pi = \{\pi_1, \cdots, \pi_K\}$. $\pi_i = k$ => the user k is the $i-$th to be decoded

SIC receiver at the AP

$$\begin{aligned} y_k^{SIC} &= v_k^t h_k s_k + \sum_{\substack{k' \in \mathbb{K} \\ k' \neq k}} v_k^t h_{k'} s_{k'} - \sum_{\substack{k' \in \mathbb{K} \\ k' \in \{\pi_1, \cdots, \pi_{i-1}\}}} v_k^t h_{k'} \hat{s}_{k'}^{SIC} + z_k, \\ &= v_k^t h_k s_k + \sum_{\substack{k' \in \mathbb{K} \\ k' \in \{\pi_{i+1}, \cdots, \pi_K\}}} v_k^t h_{k'} s_{k'} + z_k \end{aligned}$$

- $\hat{s}_{k'}^{SIC}$ current symbol guess with $y_{k'}^{SIC}$
- Inconvenient
  - may be error propagation

Downlink

- Multi-user broadcast channel MISO
- Precoding strategies => User's CSI feedback to the AP
- Broadcast : signal s_k and interferer $s_{k'}$ experience the same channel
- The user k receives :

$$y_k = h_k^{(l)} v_k s_k + \sum_{\substack{k' \in \mathbb{K} \\ k' \neq k}} h_k^{(l)} v_{k'} s_{k'} + z_k$$

- Precoding matrix $\mathbf{V}$ very important (e.g. Dirty Paper Coding)

Power resource allocation policy

- The SINR of each user depends on p_k and H
- Block fading (constant channel during T seconds)
- Resource allocation policy : each channel realization => power allocation $p\left(H\right) = \{p_1, \cdots, p_k\}$
- The transmitting power is limited
  - Downlink : $\sum_{k=1}^{K} p_k \leq P_t$
  - Uplink : $p_k \leq P \;\; \forall k \in \{1, \cdots, K\}$
- The power resource allocation is $P = \left\{p\left(H\right) : H \in \mathbb{C}^{M \times K}\right\}$

Resource Allocation and Spectral Efficiency

- Let put $R_k(\mathbf{H}, \mathbf{p}(\mathbf{H}))$: shannon capacity or user k throughput
- Average Spectral Efficiency Region : Set of the individual average spectral efficiency which can be obtained via a resource allocation policy
- Power resource allocation policy **P**, the average spectral efficiency vector is :

$$\overline{\mathbf{R}} = \left\{ E\left[R_1(\mathbf{H}, \mathbf{p}(\mathbf{H}))\right], \cdots, E\left[R_K(\mathbf{H}, \mathbf{p}(\mathbf{H}))\right] \right\}$$

Achievable data rate region

- Set of the achievable spectral efficiency

$$S = \bigcup_{\mathbf{P}\in\mathbb{P}} \bigcup_{\pi\in\Pi} \left\{\overline{\mathbf{R}} : \overline{R}_k = E\left[R_k\left(\mathbf{H}, \mathbf{p}\left(\mathbf{H}\right)\right)\right], k \in \{1, \cdots, K\}\right\}$$

  - Π is the set of permutations of $\{1, \cdots, K\}$ (decoding order)
- The convex hull S defined the spectral efficiency region :

$$\Omega = \left\{\sum_{k=1}^{K} \theta_k \overline{R}_k \mid 0 \leq \theta_k \leq 1, \sum_{k=1}^{K} \theta_k = 1, \overline{R}_k \in S, k \in \{1, \cdots, K\}\right\}$$

Optimization and region boundaries

- The boundary of Ω can be achieved by a vector $\boldsymbol{\theta} = \{\theta_1, \cdots, \theta_K\}$

$$p^*_{\boldsymbol{\theta}}(\mathbf{H}) = \arg\max_{\mathbf{p}(\mathbf{H})} \left(\max_{\pi} \sum_{k=1}^{K} \theta_k R_k(\mathbf{p}(\mathbf{H}), \mathbf{H}) \right)$$

- Let $\overline{\mathbf{R}}^*_\theta = \left\{ \overline{R}^*_1, \cdots, \overline{R}^*_K \right\}$ the optimal spectral efficiencies for $\boldsymbol{\theta}$
- No general relationship between $\boldsymbol{\theta}$ et $\overline{\mathbf{R}}^*_\theta$ can be derived
- The hyperplane $\sum_{k=1}^{K} \theta_k \overline{R}_k = cst$ is tangent to the boundary of Ω at $\left\{ \overline{R}^*_1, \cdots, \overline{R}^*_K \right\}$

Spectral efficiency region for two users

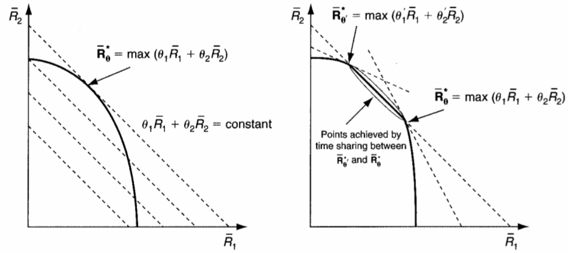


Figure 3.2 *Two-user average spectral efficiency region and maximum weighted sum rate.*

- For a certain constant value, the bound is achieved
- A same straight line can be tangent to several points

Some interesting functioning points