Energy efficiency based resource allocation

Philippe Ciblat

Joint work with C. Le Martret and X. Leturc

Outline

- Energy efficiency criterion
- Application to Hybrid ARQ (HARQ) based system
  - *Review on Fractional Programming*
  - *Short introduction to HARQ*
  - *Application to a practical scheme*

Generic resource allocation optimization problem

Energy-efficiency based problem

$$\min_{\boldsymbol{\theta}} f(\{\mathcal{E}_k(\boldsymbol{\theta}_k)\}_{k=1,\dots,K})$$

$$\text{s.t.} \quad \begin{aligned} &\mathbf{QoS}_k(\boldsymbol{\theta}_k) \geq \mathbf{QoS}_k^{(0)}, \ \forall k \in \{1,\dots,K\} \\ &\mathbf{C}(\{\boldsymbol{\theta}_k\}_{k=1,\dots,K}) \geq 0 \\ &\mathbf{C}_k(\boldsymbol{\theta}_k) \geq 0, \ \forall k \in \{1,\dots,K\} \end{aligned}$$

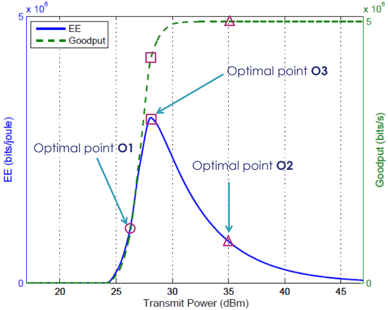
with the following energy efficiency

$$\mathcal{E}_k = \frac{\text{\# total amount of data correctly delivered by link } k}{\text{\# total consumed energy on link } k}.$$

QoS constraints: FER, delay, data rate

Why Energy Efficiency? example 1

- **O1**: minimum power with data rate constraint ($\geq$ 1Mbits/s)
- **O2**: maximum data rate with power constraint ($\leq$ 35dBm)
- **O3**: maximum energy efficiency

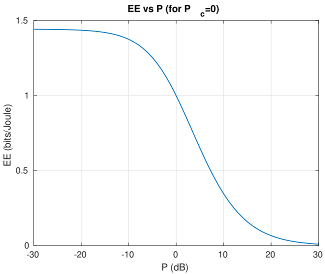


Remarks:

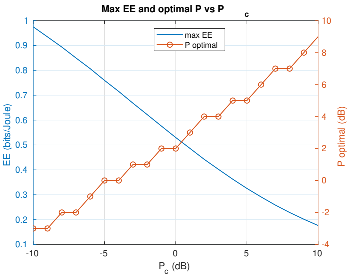
- we do not control the operating point
- consumed energy $=$ transmit power $+$ circuitry power (P_c)

Why Energy Efficiency? example 2

- Data rate: $\log_2(1 + P)$ with transmit power P
- Consumed power: $P + P_c$



EE vs P (for $P_c = 0$)



Max EE (left) and best P (right) vs P_c

Remarks:

- EE makes sense iff $P_c \neq 0$
- EE operating point strongly depends on P_c

Why Energy Efficiency? example 3

- Q_r remaining battery (%),
- T_t time to transmit the messages (s),
- N_p number of transmitted messages,
- and Data rate (Mbits/s)

		Q_r	T_t (s)	N_p	Data rate
10^7 sent messages	EE	96	297	10^7	4.3
	MTO	85	256	10^7	5
	MPO	89	1 280	10^7	1
Full battery use	EE	0	8 327	2.8×10^8	4.3
	MTO	0	1 800	7×10^7	5
	MPO	0	12 180	9.5×10^7	1

Capacity based Energy efficiency [Zappone2015]

- K users
- Downlink communications (EE computed at BTS)
- FDMA (with equal bandwidth for each user)

$$\max_{\{P_k\}} \frac{\sum_{k=1}^{K} \overbrace{\log_2(1+G_kP_k)}^{R_k}}{\sum_{k=1}^{K} P_k + P_c}$$

s.t.

$$\begin{aligned} \sum_{k=1}^{K} P_k &\leq P_{\max} \\ P_k &\geq 0, \ \forall k \end{aligned}$$

Remarks:

- Warning: the power constraint is not necessary saturated.
- Ratio between concave function and convex function
- Linear constraints

$\Rightarrow$ Resorting to **Fractional programming (FP)**

Review on Fractional Programming - 1

$$\max_{\mathbf{x}\in\mathcal{C}} \frac{f(\mathbf{x})}{g(\mathbf{x})}$$

with concave function f, and positive convex function g.

- Let q^* be the maximum value (assumed non-negative).
- Let $\mathbf{x}^*$ be the argmax value.

Lemma 1

$\mathbf{x}^*$ is achieved iff

$$\max_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) - q^*.g(\mathbf{x}) = 0$$

Consequence:

- If q^* is known in advance, just solve

$$\max_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) - q^*.g(\mathbf{x})$$

- As f concave and g convex, $f - q^*.g$ is concave

Convex optimization

Review on Fractional Programming - 2

Lemma 2

Let

$$F(q) := \max_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) - q.g(\mathbf{x})$$

is a strictly decreasing function in $q \in \mathbb{R}_+$

Consequence:

- $q \mapsto F(q)$ is continuous
- $\lim_{q\to 0} F(q) = \max_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) > 0$ (as q^* is non-negatve)
- $\lim_{q\to\infty} F(q) = -\infty$ (as g is non-negative)
- q^* is the unique root of F
- Any root-finding algorithm works!
- but each computation of F requires a convex optimization

Practical algorithm

Lemma 3 : Dinkelbach algorithm [1967]

Start with $q_0 = 0$, select an arbitrary small ε.
Iterate over n

1. Given q_n, find
$$\mathbf{x}_n^* = \arg\max_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) - q_n.g(\mathbf{x})$$
2. Then
$$q_{n+1} = \frac{f(\mathbf{x}_n^*)}{g(\mathbf{x}_n^*)}$$
3. Stop when $F(q_{n+1}) < \varepsilon$

Result: this algorithm converges to $(q^*, \mathbf{x}^*)$ up to ε.

HARQ based energy efficiency

Only channel statistics known at the transmitter

- fast-varying Rayleigh/Rice fading channel
- costly to report instantaneous channel realizations
- cheaper to report statistics due to its coherence time

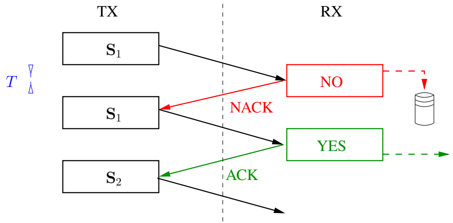
HARQ to handle unknown channel variation

k-th link characterization

- OFDMA
- Subcarriers are statistically equivalent
  - γ_k: bandwidth proportion assigned to link k
  - Q_k: energy used by link k in one OFDM symbol
    - independent of subcarrier
    - $E_k = Q_k/\gamma_k$: energy of link k in entire bandwidth
- Rice fading channel

Type-I HARQ

Type-I HARQ: packet **S** is composed by coded symbols s_n



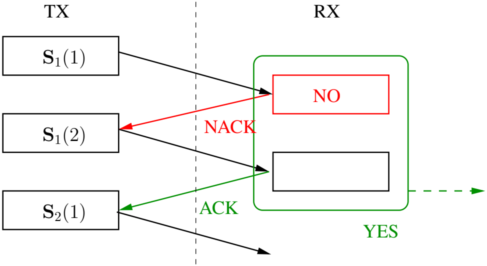
- first packet is more protected
- there is less retransmission
- transmission delay is reduced
- Efficiency is upper-bounded by the code rate

Drawbacks

- Each received packet is treated independently
- Mis-decoded packet is thrown in the trash

Type-II HARQ

Memory at RX side is considered ⇒ Type-II HARQ



Main examples:

- *Chase Combining* (CC)
- *Incremental Redundancy* (IR)

$$\begin{aligned} Y_1 &= S_1(1) + N_1 \\ Y_2 &= S_1(2) + N_2 \end{aligned}$$

then joint detection on

$$Y = [Y_1, Y_2]$$

Performance metrics

- **Packet Error Rate** (PER):

$$\text{PER} = \text{Prob(message is not decoded)}$$

- **Efficiency** (*Throughput*/*Goodput*/etc):

$$\eta = \frac{\text{information bits received without error}}{\text{transmitted bits}}$$

- **(Mean) delay**:

$$d = \#\ \text{transmitted packets when message is correctly received}$$

- **Jitter**:

$$\sigma_d = \text{delay standard deviation}$$

Quality of Service (QoS)

- Data: PER and efficiency
- Voice on IP: delay
- Video Streaming: efficiency and jitter

Our practical optimization problem

Type-I HARQ with Rice channel and minimum efficiency constraints

$$\max_{\gamma, \mathbf{E}} \sum_{k=1}^{K} \frac{m_k R_k \gamma_k (1 - \pi_k(G_k E_k))}{\kappa_{1,k} \gamma_k E_k + \kappa_{2,k}}$$

s.t. $m_k R_k \gamma_k (1 - \pi_k(G_k E_k)) \geq \eta_k^{(0)}$, $\sum_{k=1}^{K} \gamma_k \leq 1$, $\gamma_k \geq 0$, $E_k \geq 0$
with

- $G_k = |A_k|^2 + \varsigma_k^2$
- π_k probability that one frame in error

$$\pi_k(G_k E_k) \approx a_k \left(b_k \sum_{\ell=1}^{4} c_\ell \frac{e^{-\frac{|A_k|^2 G_k E_k \theta_\ell d_k}{1+\varsigma_k^2 G_k E_k \theta_\ell d_k}}}{1 + \varsigma_k^2 G_k E_k \theta_\ell d_k} \right)^{\delta_k}$$

Remark: More difficult than just information-theoretic metric

How to solve it?

- $f_k : x \mapsto 1 - \pi_k(G_k x)$ concave
- change of variables $(\gamma_k, E_k) \mapsto (\gamma_k, Q_k)$, then

$$(\gamma_k, Q_k) \mapsto \gamma_k(1 - \pi_k(G_k Q_k/\gamma_k)) = \gamma_k f_k(Q_k/\gamma_k)$$

  is concave as perspective of f_k

  Consequently, in (γ_k, Q_k)
  - Numerator: concave
  - Denominator: convex (as linear)
  - Constraints set: convex set

Results (fractional programming tool)

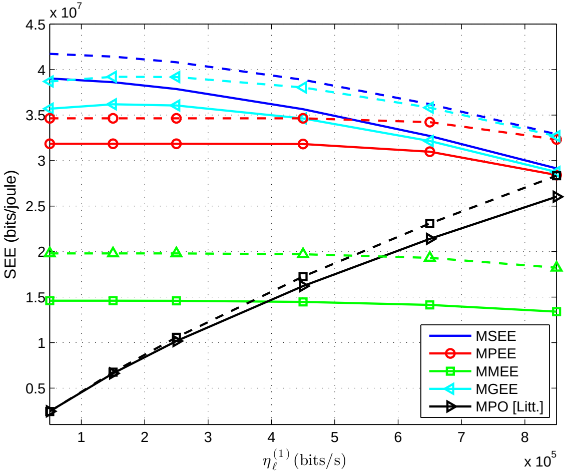
- Jong's algorithm: solve at iteration i (with the above constraints)

$$\max_{\gamma, \mathbf{Q}} \sum_{k=1}^{K} u_k^{(i)} m_k R_k \gamma_k (1 - \pi_k(G_k Q_k/\gamma_k)) - v_k^{(i)} \kappa_{1,k} Q_k$$

  and update $u_k^{(i)}$ and $v_k^{(i)}$ according to well-defined equations
- KKT can be written in closed-form

Numerical results

- $K = 10$ links, Bandwidth $W = 5$ MHz
- QPSK, convolutional code of rate $1/2$
- Rician factor: 10 (dashed line), 0 (solid line)



Future works

- Model for P_c :
  - How to take into account the manufacturing
  - How to take into account the decoding power consumption

$$= a - b \times \log(C - R)$$

    with C the Shannon capacity and R the current data rate
  - Extension to massive MIMO done but without two previous items and also depends on the hybrid RF.
- More philosophical thoughts: what do we want to do with a network?
  - If P_c is large (not efficient), then EE leads to high P, and more Green House Gas (GHG)
  - If P_c is low (very efficient), then EE leads to low P, and to lowtech (low rate, for instance)

Sketch of proof for Lemma 1

- Let $\mathbf{x}^*$ be the optimal solution of RHS of Lemma 1. It means

$$f(\mathbf{x}^*) - q^* g(\mathbf{x}^*) = 0 \Rightarrow \frac{f(\mathbf{x}^*)}{g(\mathbf{x}^*)} = q^*$$

  Moreover $\forall \mathbf{x} \neq \mathbf{x}^*$,

$$f(\mathbf{x}) - q^* g(\mathbf{x}) \leq 0 \Rightarrow \frac{f(\mathbf{x})}{g(\mathbf{x})} \leq q^*$$

  which proves that $\mathbf{x}^*$ is the optimal solution of FP.
- Let $\mathbf{x}^*$ be the optimal solution of FP. As $q^* = f(\mathbf{x}^*)/g(\mathbf{x}^*)$, we get

$$\frac{f(\mathbf{x})}{g(\mathbf{x})} < q^* \Rightarrow f(\mathbf{x}) - q^* g(\mathbf{x}) \leq 0$$

  for any $\mathbf{x} \neq \mathbf{x}^*$.

Sketch of proof for Lemma 2

- Assume $q_1 > q_2$
- $\mathbf{x}_1^*$ the argmax with q_1, and $\mathbf{x}_2^*$ the argmax with q_2

$$F(q_1) = f(\mathbf{x}_1^*) - q_1 g(\mathbf{x}_1^*) \overset{(a)}{<} f(\mathbf{x}_1^*) - q_2 g(\mathbf{x}_1^*)$$
$$\overset{(b)}{<} f(\mathbf{x}_2^*) - q_2 g(\mathbf{x}_2^*) = F(q_2)$$

(a) $q_1 > q_2$ and g is a positive function. Strict inequality.

(b) $\mathbf{x}_2^*$ is the argmax for q_2

Sketch of proof for Lemma 3

Step 1: sequence $\{q_n\}_n$ is strictly increasing.

- Assuming $F(q_n) = f(\mathbf{x}_n^*) - q_n g(\mathbf{x}_n^*) > 0$ (True for $F(q_0)$)
- $f(\mathbf{x}_n^*) - q_{n+1} g(\mathbf{x}_n^*) = 0$

$$q_{n+1} - q_n = \frac{F(q_n)}{g(\mathbf{x}_n^*)} > \frac{F(q_n)}{g_{\max}} > 0 \quad (1)$$

with $g_{\max} = \max_{\mathbf{x} \in \mathcal{C}} g(\mathbf{x})$ (it exists if $\mathcal{C}$ compact)

Step 2: convergence to q^*

- Due to stopping criterion, bounded increasing sequence, and so $\lim_{n \to \infty} q_n = \overline{q}$
- Assuming that $\lim_{n \to \infty} F(q_n) = F(\overline{q}) > \varepsilon$, i.e, $\overline{q} < q^* - \delta$ with $F(q^* - \delta) = \varepsilon$.
- but as q_n converges, $(q_{n+1} - q_n)$ converges to 0, and Eq. (1) implies

$$F(q_n) \to 0 \Rightarrow q_n \to q^*$$

which leads to a contradiction.