# Energy efficiency optimization in wireless networks

Vineeth Satheeskumar Varma Satheeskumar Varma

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir®

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir®

# Université Paris-Sud

Ecole Doctorale : STITS
Laboratoire des signaux et systèmes

Discipline : Physique

## Thèse de doctorat

Soutenue le 20/06/2014 par

**Vineeth Satheeskumar Varma**

# Optimisation de l'efficacité énergétique des réseaux de communication

| | | |
|---|---|---|
| **Directeur de thèse :** | Samson Lasaulce | Prof. (CNRS/LSS) |
| **Co-directeurs de thèse :** | Salah Eddine Elayoubi | Senior Researcher (Orange Labs) |
| | Merouane Debbah | Prof. (Supélec) |
| **Composition du jury :** | | |
| Président du jury : | Rachid Elazouzi | Prof. (Uni. of Avignon) |
| Rapporteurs : | Aris Moustakas | Asst. Prof. (NCU of Athens) |
| | Tijani Chahed | Prof. (Telecom SudParis) |
| Examinateurs : | Marceau Coupechoux | Asst. Prof (Telecom ParisTech) |
| | Luca Sanguinetti | Asst. Prof (Uni. of Pisa) |
| Invités : | ..... | ..... (.....) |

# Table des matières

**Résumé**

Le principal objectif de la thèse était d'établir un cadre d'étude des communications efficace énergétiquement en définissant et en justifiant de nouvelles mesures d'efficacité énergétique pour divers systèmes sans fils. En général, le rendement énergétique est défini comme le rapport entre le débit total et la puissance totale consommée par l'émetteur. Cette définition implique que, lorsqu'une re-transmission est autorisée (dans le cas d'une erreur sur les paquets), la maximisation de l'efficacité énergétique peut conduire directement à minimiser l'énergie dépensée lors de la transmission d'une unité d'information.

La technique d'utilisation de plusieurs antennes à la fois sur l'émetteur et le récepteur (MIMO ) , est maintenant bien établie dans le domaine de la communication sans fil et de la théorie de l'information. Dans ce travail, l'efficacité énergétique d'un système MIMO point-à-point est étudiée à la fois dans le cas d'information imparfaite sur l'état du canal (CSI) est disponible uniquement à l'émetteur et dans le cas d'information imparfaite sur l'état du canal (CSI) est disponible uniquement au niveau du récepteur. Les résultats de l'analyse indiquent que, dans les deux cas, il existe une unique puissance d'émission optimale et que celle-ci permettra de maximiser l'efficacité énergétique.

D'autres résultats montrent également que, lorsque l'information sur l'état du canal est imparfaite, et l'estimation de canal est nécessaire, utiliser toutes les antennes d'émission disponibles n'est pas optimal pour la communication à faible consommation d'énergie lorsque le temps de cohérence du canal est fini. L'efficacité énergétique d'un réseau de petites cellules centralisé utilisant un MIMO virtuel pour une meilleure efficacité spectrale a également été analysée dans ce travail. , Il a ainsi été montré que le paramètre ayant l'effet maximal sur l'efficacité énergétique est le schéma du mode de sommeil des petites cellules. Dans la plupart des systèmes en pratique, les émetteurs (dans la couche physique) ont un tampon de mémoire dans lequel des paquets de données arrivent à partir des couches supérieures selon un processus stochastique. L'efficacité énergétique inter-couche de l'émetteur, en tenant compte de la nature sporadique de ces arrivées de paquets a été étudiée dans la section suivante de la thèse. Une nouvelle métrique d'efficacité énergétique a été conçue pour un tel système et il a été montré que cette nouvelle métrique est quasi- concave par rapport à la puissance d'émission sous certaines hypothèses sur le taux d'arrivée des paquets. Une formulation de ce problème par la théorie des jeux, où chaque émetteur agit en tant que décideur, a également été proposée dans un réseau d'interférence et il a été montré que le jeu de commande de puissance résultant avait un équilibre de Nash unique. En outre, un algorithme de meilleure réponse qui permettrait

aux décideurs indépendants d'atteindre cet équilibre de manière distribuée a également été proposé.

En conclusion, ce travail développe un cadre pour l'efficacité énergétique pour les cas généraux de contrôle de puissance dans la couche physique avec MIMO et inter-couche avec une arrivée de paquets sporadique. Des algorithmes centralisés et décentralisés pour atteindre un point de fonctionnement économe en énergie pour les systèmes à l'étude ont été proposés. Les résultats mettent en évidence le compromis entre la consommation d'énergie et les taux de date en ce qui concerne l'efficacité énergétique des réseaux sans fil.Le principal objectif de la thèse était d'établir un cadre d'étude des communications efficaceénergétiquement en définissant et en justifiant de nouvelles mesures d'efficacité énergétique pour divers systèmes sans fils.

**Abstract**

The main objective of the thesis was to establish a framework for energy-efficient communication by defining and justifying novel energy-efficiency metrics for various wireless systems and settings. In general, the energy-efficiency is defined as the ratio of the total data rate to the total power consumed at the transmitter. This definition implies that, when re-transmission is allowed (in the case of outage), maximizing energy-efficiency can directly lead to minimizing the energy spent in transmitting a unit of information.

The technique of using multiple antennas both at the transmitter and receiver (MIMO), is well established in the field of wireless communication and information theory. In this work, the energy-efficiency of a MIMO point-to-point link is studied both when imperfect channel state information (CSI) is available at the transmitter and when imperfect CSI is available only at the receiver. The analytical results indicate that in both cases, a unique optimal transmit power exists which will maximize the energy-efficiency. Additional results also show that when the CSI is imperfect, and channel estimation is required, using all the available transmit antennas is not optimal for energy-efficient communication when the coherence time of the channel is finite. The energy-efficiency of a centralized small cell network using virtual MIMO for a higher spectrum efficiency was also analyzed in this work. In this, the parameter of interest with the maximum impact on the energy efficiency was found to be the sleep mode scheme of the small cells.

In most practical systems, the transmitters (in the physical layer) have a memory buffer into which data packets arrive from the upper layers through a stochastic process. The cross-layer energy efficiency of the transmitter, taking into account the sporadic nature of these packet arrivals was studied in the next section of the thesis. A novel energy-efficiency metric was designed for such a system and it was shown that this new metric is quasi-concave with respect to the transmit power under certain assumptions on the packet arrival rate. A game theoretical formulation of this problem, with each transmitter acting as a decision maker, was also studied in an interference network and it was shown that the resulting power control game had a unique Nash equilibrium. Furthermore, a best response algorithm that would allow independent decision makers to reach this equilibrium in a distributed manner was also found.

In conclusion, this work develops a framework for energy-efficiency for the general cases of power control in the physical layer with MIMO and cross-layer with a sporadic packet arrival. Both centralized and decentralized algorithms for achieving an energy-efficient working point for the systems under consideration were proposed. The results highlight the trade-off

between power consumption and date rates in energy efficient wireless networks.

# Chapter 1

# Introduction

This manuscript is focused on the study and design of energy-efficient wireless systems. The ideas discussed can be applied for both large scale energy management while operating radio towers and base stations, as well as micro-management of mobile terminals.

## 1.1 Background and motivation

Over the past two decades, designing energy-efficient communication terminals has become an important issue. This is not surprising for terminals which have to be autonomous as far as energy is concerned, such as cellular phones, unplugged laptops, wireless sensors, and mobile robots. More surprisingly, energy consumption has also become a critical issue for the fixed infrastructure of wireless networks. For instance, Vodafone's global energy consumption for 2007-2008 was about 3000 GWh [1], which corresponds to emitting 1.45 million tons of CO2 and represents a monetary cost of a few hundred million Euros. The Information and Communication Technology industry currently accounts for 2% of worldwide carbon emissions. With exponential increases in information and communication traffic, the global carbon footprint is expected to double over the next 10 years. While incremental approaches to reducing energy consumption are critical, ultimately they will fail to keep pace, thus requiring significant advancements in the energy-efficiency of the communication sector. This context explains, in part, why concepts like "green communications" have emerged as seen from [2, 3] and [4]. Using large multiple antennas, white spaces, virtual multiple input multiple output (MIMO) systems, and small cells with sleep mode are envisioned to be some of the ways of reducing energy consumption drastically.

Designing green wireless networks [5, 2, 6] has become increasingly important for modern wireless networks, in particular, to manage operating costs. A challenge for modern (beyond 4G and 5G) cellular networks is not

only to respond to the explosion of data rates, but also to manage network energy consumption. The concept of small cell networks appears as a good candidate solution to raise such a challenge (see e.g., [7]). As small cell networks will be distributed to large extent and subject to high inter-cell interference, designing distributed energy-efficient interference management schemes appears as a natural need. For being able to design green networks, find a suitable energy-efficiency metric is crucial. In [8], the EE of a communication between a transmitter and a receiver is defined as the ratio of the net data rate to the radiated power; the corresponding quantity is a measure of the average number of bits successfully received per joule consumed at the transmitter. Quite recently, there has been a resurgence of interest in this performance metric. There are several reasons for this and only a few are provided. First, the EE as defined in [8], mathematically translates in a simple manner the trade-off between the benefit of increasing the transmit power in terms of data rate, and the induced cost in terms of consumed energy or amount of interference generated. Second, as motivated in [9], there are applications in which the allowable delay is not tightly constrained. Therefore, the data rate is a less relevant measure than the energy needed to transmit the information and EE naturally appears as a metric to be optimized.

On [11] the authors bridge the gap between the pioneering work by Verdú on the capacity per unit cost for static channels [12] and the more pragmatic definition of energy-efficiency proposed by [8] for quasi-static single input single output (SISO) channels. Indeed, in [11], energy-efficiency is defined as the ratio of the probability that the channel mutual information is greater than a given threshold to the used transmit power. Assuming perfect channel state information at the receiver (CSIR) and the knowledge of the channel distribution at the transmitter, the pre-coding matrix is then optimized for several special cases. While [11] provides interesting insights into how to allocate and control power at the transmitter, a critical issue is left unanswered; to what extent do the conclusions of [11] hold in more practical scenarios such as those involving imperfect CSI? Answering this question was one of the motivations for the work summarized in Chapter 2.

As most practical communication systems work with some sort of sporadic arrival for data packets behind the physical layer, the energy efficiency of the system would not be the same if packets arrived constantly. In most of the literature on energy-efficiency, this is not considered and Chapter 3 explores this issue. A distributed algorithm that maximizes individual energy-efficiency of a transmitter is also proposed. The case where there is a non-static number of users connected to a network is studied in Chapter 4. Finally, a technique of exchanging CSI information between interfering and distributed transmitter-receiver pairs through power level coding is shown in Chapter 4.

## 1.2 Basic definition of the energy efficiency metric

In the proposed approach, the goal pursued is to maximize the number of information bits transmitted successfully per Joule consumed at the transmitter. This is different from the most conventional approach which consists in minimizing the transmit power under a transmission rate constraint: [13] perfectly represents this body of literature. In the latter and related works, efficiency is not the main motivation. [14] provides a good motivation as to how energy-efficiency can be more relevant than minimizing power under a rate constraint. Indeed, in a communication system without delay constraints, rate constraints are generally irrelevant whereas the way energy is used to transmit the (sporadic) packets is of prime interest. Rather, the approach discussed in this manuscript follows the original works on energy-efficiency which includes [15, 8, 16, 17, 18]. The current state of the art indicates that, since [11], there have been no works where the MIMO case is treated by exploiting the cumulative distribution of the channel mutual information (i.e., the outage probability) at the numerator of the performance metric. As explained below, the analysis goes much further than [11] by considering effects such as channel estimation error effects. Several works address the issue of power allocation for outage probability minimization [19, 20, 21] under imperfect channel state information.

In [8], EE is defined as the ratio between the average net data transmission rate and the power consumed for sending a given packet. When the radiated power is considered as the transmission cost, this ratio merely equals $\frac{Rf(\gamma_i(\underline{p}))}{p_i}$ where the quantity $R$ is the constant gross data rate (determined by the coding and modulation schemes) on the radio interface, $\underline{p}$ is the vector of power levels and $\gamma_i$ the SINR of user $i$. Each packet transmitted on the channel is received without any errors with a probability which depends on the quality of the communication link, the interference, and transmit power levels. The corresponding block or packet success rate (also called efficiency function) is precisely the function $f(\gamma_i(\underline{p}))$ above. The function $f : [0, +\infty) \rightarrow [0, 1]$ is a sigmoidal [1] or S-shaped function verifying $f(0) = 0$ and $\lim_{x\rightarrow\infty} f(x) = 1$ (see [37] for more details). Common examples for $f$ are $f(x) = (1 - e^{-x})^M$, $f(x) = e^{-\frac{c}{x}}$ [11], where $M \geq 1$ is the packet length and $c > 0$ is some constant related to spectral efficiency. Energy-efficiency is particularly relevant when packet re-transmission is allowed. When there is no re-transmission, the energy [2] consumed to send $V$ bits while transmitting at the power level $p_i$ is $p_i \frac{V}{R}$. Minimizing energy amounts to minimizing $p_i$ in the absence of re-transmissions. However, when re-transmission is al-

1. A sigmoidal function is a function which is initially convex for $\gamma \in [0, \gamma_+]$ and eventually concave for $\gamma \in [\gamma_+, \infty)$.

2. Here, the energy under consideration is the energy associated with the radiated signal.

lowed (typically by using an automatic repeat request -ARQ- protocol, that is used at the physical layer independently of the architecture at the upper layer), the average duration to send a packet equals $\frac{V}{Rf(\gamma_i(\underline{p}))}$ and the energy consumed becomes $p_i \frac{V}{Rf(\gamma_i(\underline{p}))}$. Clearly, minimizing energy amounts to maximizing EE. This means that, at least in presence of re-transmissions, the classical approach which involves minimizing $p_i$ (subject to some quality of service constraints) induces a loss in terms of minimizing the energy consumption. To be precise, minimizing the power subject to a QoS constraint does not minimize the energy consumed by the system when the system has a certain amount of data to be transferred. This will be illustrated through some numerical results in this manuscript.

Of course, when the system under consideration has just a single transmitter, the energy efficiency can be easily defined to be the ratio of it's average data rate to power consumption. However, in a multi-agent system, there are several ways of defining the energy-efficiency and the best choice is not so straight-forward or clear. For example, in a small cell cluster with two base stations serving two users, the energy efficiency can be defined in two ways. As either the sum of the individual energy efficiencies of each base station or as the sum of the total rate in the network divided by the total power. The second choice will certainly lead to a minimization of the total energy spent by the system but may result in an optimal strategy that is partial to one agent. In the small cell example, the optimal solution might be to never serve the user further away from the base stations. In distributed systems, each agent either optimizes their individual energy efficiencies or if possible, the sum. When mobile users optimize their energy efficiency by tuning the uplink transmit power, the most natural assumption is that they maximize their individual energy efficiency as discussed in [8]. When the base station optimizes it's energy efficiency, the second choice is opted as discussed in some sections of Chapter 2 and 4.

## 1.3 Structure of the manuscript and publications

In comparison to existing literature, the main contributions of the manuscript can be summarized as follows:

- One of the scenarios under investigation concerns the case of single-user MIMO energy-efficiency both with and without imperfect CSI available at the transmitter.
- Energy efficiency of a virtual MIMO system formed in a cluster of small cells and some users.
- The energy efficiency of a cross-layer system, i.e taking into account the sporadic nature of data packet arrival: defining the appropriate metric and it's properties.
- Both centralized and de-centralized solutions for many of the above

scenarios.
- The expected energy-efficiency definition and optimization considering the dynamic nature of users connected to a network.
- A method of exchanging CSI through power level coding.
- In all the above cases, instead of considering the radiated power only for the cost of transmitting, the total power consumed by the transmitter is accounted for.

The manuscript is therefore structured as follows. Chapter. II is a study of energy efficiency of MIMO systems. This chapter is based around three publications (publications are indexed by **J**. for journal publications, **C**. for conference proceedings, **B**. for book chapters and **P**. for patents) that are provided in the appendix, namely:

**C**.**1** V.S Varma, S. Lasaulce, M. Debbah and S.E. Elayoubi, "Impact of Mobility on Wireless Green Networks", European Signal Processing Conference (EUSIPCO) 2011.

**J**.**1** V.S Varma, S. Lasaulce, M. Debbah and S.E. Elayoubi, "An Energy Efficient Framework for the Analysis of MIMO Slow Fading Channels", IEEE Trans. Signal Proc., Vol 61, 10, pp: 2647-2659.

**C**.**2** V.S Varma, S.E. Elayoubi, M. Debbah and S. Lasaulce, "On the Energy Efficiency of Virtual MIMO Systems", IEEE Int. Symposium on personal, indoor and mobile communications (PIMRC 2013).

The work on this chapter has also led to some interesting patents that have been submitted for approval:

**P**.**1** V.S. Varma, S.E. Elayoubi, M. Debbah and S. Lasaulce, "Virtual MIMO optimal antenna selection and sleep mode implementation", No: 200113-FR (filed).

**P**.**2** V.S. Varma, S.E. Elayoubi and M. Debbah, "An efficient scheme for beam-forming in home base stations", No: 200313-FR (filed).

Chapter 3 deals with the impact of the often ignored queuing process of packets, that are present before the physical layer in typical communication systems, on energy efficiency. This chapter is based around the following publications:

**C**.**3** V.S Varma, S. Lasaulce, Y. Hayel, S. Eddine Elayoubi and Merouane Debbah, "Cross-Layer Design for Green Power Control",IEEE Int. Conference on Communications (ICC 2012).

**J**.**2** Vineeth S Varma, Samson Lasaulce, Yezekael Hayel, Salah Eddine Elayoubi and Merouane Debbah, "A Cross-Layer Approach for Energy-Efficient Distributed Power Control", IEEE Trans. on Vehicular Tech. (revised)

Chapter 4 is a summary of techniques and methods that can improve the energy-efficiency of both centralized and decentralized systems. It is partly composed of original ideas that are yet to be published, and partly based

on the following publications:

**C.4** V.S Varma, S.E. Elayoubi, S. Lasaulce and M. Debbah, "A Flow Level Perspective on Base Station Power Allocation in Green Networks", ACM International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2012, Best student paper award).

**C.5** M. Mhiri, V.S Varma, M.L. Truest, S. Lasaulce and A. Samet, "On the benefits of repeated game models for green cross-layer power control in small cell", BlackSeaComm 2013

The work associated with this chapter has also inspired the following patents:

**P.3** S. Lasaulce, V.S. Varma and R. Visoz, "Coding information through power levels", Ref No: 200288FR01-PJ. (filed)

**P.4** V.S. Varma, S. Lasaulce and S.E. Elayoubi, "Sleep mode for network resources assisted by information from transportation system traffic detectors", 200047FR0-DM. (filed)

Other contributions which will not be discussed in this manuscript have been obtained and/or published in one book chapter, three conference papers and one other patent:

**B.1** V.S. Varma, E.V. Belmega, S. Lasaulce and M. Debbah," Energy Efficient Communications in MIMO Wireless Channels: Information Theoretical Limits", CRC Press book on green communications, 2012.

**C.6** F. Meriaux, V.S Varma, S. Lasaulce, "Mean Field Energy Games in Wireless Networks", IEEE Asilomar 2012. (invited paper)

**C.7** V.S Varma, S. Lasaulce, M. Debbah and S.E. Elayoubi, "Green Power Control for large MIMO systems", Colloque Gretsi 2013.

**C.8** B. Perabathini, M. Debbah, M. Kountouris and A. Conte, "Physical Limits of point-to-point communication systems", IEEE WiOpt 2014 - PhysCommNet Workshop. (invited paper)

**P.5** "Scheme for broadcasting and receiving accompanying data for TV or radio services", V.S. Varma and S.E. Elayoubi. Ref No: 1H518040/655.SF. (filed)

# Chapter 2

# Energy efficiency analysis of MIMO systems

The focus of this chapter is on the energy-efficiency of MIMO systems. Two important cases of interest are studied here; the case of single user or point-to-point MIMO and that of multi-user virtual MIMO. These cases are arguably two of most practical applications of MIMO technology. In the first case, the MIMO system is formed between several antennas localized at one point (transmitter) and several antennas localized at another point (receiver). Even if many users are connected to the same base station in a network, but are served on multiple channels, the results from the first case apply as all the MIMO channels are in parallel. However, if all the users are connected through the same spectral band at the same time, the situation belongs to the second case of virtual MIMO. This chapter discusses and studies the energy efficiency of both these cases in detail. Note that in this chapter and other chapters in which the transmit power is controlled by a single entity, the power will be represented by $P$ whereas in distributed power control like in Chapter 3 (and the introduction), $p_i$ will be used for individual powers and $\underline{p}$ for the vector of individual powers. This choice was made for improved understanding while reading and to avoid confusion.

## 2.1 Single user or point-to-point MIMO systems

### 2.1.1 System model

A point-to-point multiple input and multiple output communication unit is studied in this section. The dimensionality of the input and output is given by the numbers of antennas but the analysis holds for other scenarios such as virtual MIMO systems [26]. If the total transmit power is given as $P$, the average SNR is given by :

$$\gamma = \frac{P}{\sigma^2} \tag{2.1}$$

where $\sigma^2$ is the reception noise variance.The signal at the receiver is modeled by :

$$\underline{y} = \sqrt{\frac{\gamma}{M}}\mathbf{H}\underline{s} + \underline{z} \tag{2.2}$$

where $\mathbf{H}$ is the $N \times M$ channel transfer matrix and $M$ (resp. $N$) the number of transmit (resp. receive) antennas. The entries of $\mathbf{H}$ are i.i.d. zero-mean unit-variance complex Gaussian random variables. The vector $\underline{s}$ is the $M$-dimensional column vector of transmitted symbols follows a complex normal distribution, and $\underline{z}$ is an $N$-dimensional complex white Gaussian noise distributed as $\mathcal{N}(\underline{0}, \mathbf{I})$. Denoting by $\mathbf{Q} = \mathbb{E}[\underline{s}\underline{s}^H]$ the input covariance matrix (called the pre-coding matrix), which satisfies

$$\frac{1}{M}\mathrm{Tr}(\mathbf{Q}) = 1 \tag{2.3}$$

where Tr stands for the trace operator. The power constraint is expressed as :

$$P \leq P_{\max} \tag{2.4}$$

where $P_{\max}$ is the maximum available power at the transmitter.

The channel matrix $\mathbf{H}$ is assumed to evolve in a quasi-static manner : the channel is constant for some time interval, after which it changes to an independent value that it holds for the next interval [22]. This model is appropriate for the slow-fading case where the time with which $\mathbf{H}$ changes is much larger than the symbol duration.

When the same (imperfect) CSI is available at the transmitter and receiver, by estimating the channel for $t$ time, and sending the information to the transmitter within $t_f$ time, the energy-efficiency $\eta_T$ is defined as:

$$\eta_T(P, \mathbf{Q}, \hat{\mathbf{H}}) = \frac{R\left(1 - \frac{t+t_f}{T}\right) F_L\left[I_{\mathrm{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}}) - \frac{R}{R_0}\right]}{aP + b} \tag{2.5}$$

where $R$ is the transmission rate in bit/s, $T$ is the block duration in s, $R_0$ is a parameter which has unit Hz (e.g., the system bandwidth), and $a > 0$, $b \geq 0$ are parameters to relate the transmitter radiated power to its total consumed power ;define $\xi = \frac{R}{R_0}$ as the spectral efficiency. $I_{\mathrm{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}})$ denotes the mutual information with imperfect CSITR (the receiver also has the exact same CSI as the transmitter). This form of the energy-efficiency is inspired from early definitions provided in works like [8], and studies the gain in data rate with respect to the cost which is the power consumed. The numerator represents the benefit associated with transmitting namely, the net transmission rate (called the goodput in [27]) of the communication and is measured in bit/s. The goodput comprises a term $1-\frac{t+t_f}{T}$ which represents the loss in terms of information rate due to the presence of a training and feedback mechanism (for duration $t$ seconds and $t_f$ seconds resp. in a $T$s

long block) [1]. The denominator of (2.5) represents the cost of transmission in terms of power. The proposed form for the denominator of (2.5) is inspired from [25] where the authors propose to relate the average power consumption of a transmitter (base stations in their case),to the average radiated or radio-frequency power by an affine model.

The term $F_L(.)$ represents the transmission success probability. The details on this function are given inA.2. The bounds on $F_L$ can be expressed as $F_L(I_{\text{ICSITR}}(0,0,\mathbf{H}) - \xi) = 0$ (no reliable communication when transmit power is zero) and as $F_L \to 1$ when $P \to \infty$. Therefore, in the presence of CSI at the transmitter, outage occurs even when the mutual information is more than the targeted rate due to the noise and finite code-lengths. In this scenario, the energy-efficiency is maximized when the parameters $\mathbf{Q}$ and $P$ are optimized.

In the absence of CSI at the transmitter, the earlier definition of energy efficiency is not suitable since $\mathbf{H}$ is random, $\eta_T$ is also a random quantity. Additionally, in this case, it is impossible to know if the data transmission rate is lower than the instantaneous channel capacity as the channel varies from block to block. Therefore, in this case, the source of outage is primarily the variation of the channel [32], and using (2.5) directly is not suitable. As the channel information is unavailable at the transmitter, define $\mathbf{Q} = \frac{\mathbf{I}_M}{M}$, meaning that the transmit power is allocated uniformly over the transmit antennas. Under this assumption, the average energy-efficiency can be calculated as the expectation of the instantaneous energy-efficiency over all possible channel realizations. For large $L$, it has been shown in [32] (and later used in other works like [11]) that the expression can be rewritten and approximated to :

$$\eta_R(P,t) = \frac{R\left(1-\frac{t}{T}\right)\Pr_{\mathbf{H}}\left[I_{\text{ICSIR}}(P,t,\hat{\mathbf{H}}) \geq \xi\right]}{aP+b} \tag{2.6}$$

where $\Pr_{\mathbf{H}}$ represents the probability evaluated over the realizations of the random variable $\mathbf{H}$. Here, $I_{ICSIR}$ represents the mutual information of the channel with imperfect CSI at the receiver. Let us comment on this definition of energy efficiency. This definition is similar to the earlier definition in all most ways. Here the parameter $t$, represents the length of the training sequence used to learn the channel at the receiver [2]. The major difference here is that the expression for the success rate is the probability that the associated mutual information is above a certain threshold. This definition

1. In this case, its assumed that the feedback mechanism is sufficient to result in perfect knowledge of $\hat{\mathbf{H}}$ at the transmitter. This is done because, assuming a different imperfect CSI at the transmitter from the receiver creates too much complexity and this problem is beyond the scope of this manuscript.

2. In this case, the optimization is done over $P$ and $t$ assuming imperfect CSI at the receiver. A parameter here not explicitly stated, but indicated nevertheless, is $M$ due to the number of transmit antennas affecting the effectiveness of training

of the outage is shown to be appropriate and compatible with the earlier definition when only statistical knowledge of the channel is available [32].

Each transmitted block of data is assumed to comprise a training sequence in order for the receiver to be able to estimate the channel; the training sequence length in symbols is denoted by $t_s$ and the block length in symbols by $T_s$. Continuous counterparts of the latter quantities are defined by $t = t_s S_d$ and $T = T_s S_d$, where $S_d$ is the symbol duration in seconds. In the training phase, all $M$ transmitting antennas broadcast orthogonal sequences of known pilot/training symbols of equal power on all antennas. The receiver estimates the channel, based on the observation of the training sequence, as $\widehat{\mathbf{H}}$ and the error in estimation is given as $\Delta\mathbf{H} = \mathbf{H} - \widehat{\mathbf{H}}$. Concerning the number of observations needed to estimate the channel, note that typical channel estimators generally require at least as many measurements as unknowns [23], that is to say $t_s \geq M$. The channel estimate normalized to unit variance is denoted by $\widetilde{\mathbf{H}}$. From [23] its known that the mutual information is the lowest when the estimation noise is Gaussian. Taking the worst case noise, it has been shown in [22] that the following observation equation

$$\widetilde{\underline{y}} = \sqrt{\frac{\gamma_{\text{eff}}(\gamma, t)}{M}}\widetilde{\mathbf{H}}\underline{s} + \widetilde{\underline{z}} \tag{2.7}$$

perfectly translates the loss in terms of mutual information [3] due to channel estimation provided that the effective SNR $\gamma_{\text{eff}}(\gamma, t)$ and equivalent observation noise $\widetilde{\underline{z}}$ are defined properly namely,

$$\begin{cases} \widetilde{\underline{z}} & = & \sqrt{\frac{\gamma}{M}}\Delta\mathbf{H}\underline{s} + \underline{z} \\ \gamma_{\text{eff}}(\gamma, t) & = & \frac{\frac{t}{MS_d}\gamma^2}{1+\gamma+\gamma\frac{t}{MS_d}} \end{cases} . \tag{2.8}$$

As the worst case scenario for the estimation noise is assumed, all formulas derived in the following sections give lower bounds on the mutual information and success rates. Note that the lower bound is tight (in fact, the lower bound is equal to the actual mutual information) when the estimation noise is Gaussian which is true in practical cases of channel estimation.

Now that the energy efficiency metric has been defined for the single user MIMO case both with and without CSI at the transmitter, the next step forward is to study the properties of this metric in order to easily optimize it.

### 2.1.2 Imperfect CSITR available

When perfect CSITR or CSIR is available, the mutual information of a MIMO system, with a pre-coding scheme $\mathbf{Q}$ and channel matrix $\mathbf{H}$ can be

3. It is implicitly assumed that the mutual information is taken between the system input and output; this quantity is known to be very relevant to characterize the transmission quality of a communication system (see e.g. [34] for a definition).

expressed as:

$$I_{\text{CSITR}}(P, \mathbf{Q}, \mathbf{H}) = \log \left| \mathbf{I}_M + \frac{P}{M\sigma^2} \mathbf{H}\mathbf{Q}\mathbf{H}^H \right| \tag{2.9}$$

The notation $|\mathbf{A}|$ denotes the determinant of the (square) matrix $\mathbf{A}$. With imperfect CSIT, which is exactly the same as the CSIR (i.e., both the transmitter and the receiver have the same channel estimate $\hat{\mathbf{H}}$), a lower bound on the mutual information can be found from several works like [19, 21] etc. This lower bound for $I_{\text{ICSITR}}$ is used, which is expressed as:

$$I_{\text{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}}) = \log \left| \mathbf{I}_M + \hat{\mathbf{H}} \frac{P}{M\sigma^2(1+\gamma\sigma_E^2)} \mathbf{Q}\hat{\mathbf{H}}^H \right| \tag{2.10}$$

where $\hat{\mathbf{H}}$ is the estimated channel and $1-\sigma_E^2$ is the variance of $\hat{\mathbf{H}}$. Considering the block fading channel model, from [19] and [23], its concluded that $\sigma_E^2 = \frac{1}{1+\gamma\frac{t}{M}}$. Simplifying :

$$I_{\text{ICSITR}}(P, \mathbf{Q}, \widehat{\mathbf{H}}) = \log \left| \mathbf{I}_M + \frac{\gamma_{\text{eff}}}{M} \widehat{\mathbf{H}}\mathbf{Q}\widehat{\mathbf{H}}^H \right|. \tag{2.11}$$

Having defined the mutual information to be used for (2.5), the optimization of $\eta_T$ is pursued.

Studying (2.5) and (2.11), its seen that varying the power allocation (or the corresponding pre-coding matrix) $\mathbf{Q}$, affects only the success rate $F_L(.)$ and the total power $P$ is the only term that is present outside $F_L(.)$. As $F_L(.)$ is known to be an increasing function, if the total power is a constant, optimizing the energy efficiency $\eta_T$ amounts to simply maximizing the mutual information $I_{\text{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}})$. This is a well documented problem and it gives a "water-filling" type of solution [36]. Rewriting (2.9) as

$$I_{\text{ICSITR}}(P, \mathbf{Q}, \widehat{\mathbf{H}}) = \log \left| \mathbf{I}_M + \frac{\gamma_{\text{eff}}}{M} \mathbf{D}\mathbf{S}\mathbf{D}^H \right| \tag{2.12}$$

where the optimal covariance matrix $\mathbf{Q} = \mathbf{V}\mathbf{S}\mathbf{V}^H$ is achieved through the singular value decomposition of the channel matrix $\widehat{\mathbf{H}} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ and an optimal diagonal covariance matrix $\mathbf{S} = \text{diag}[s_1, \ldots, s_{\min(M,N)}, 0, \ldots, 0]$. The water-filling algorithm can be performed by solving:

$$s_i = \left( \mu - \frac{1}{\gamma \|d_i\|^2} \right)^+, \text{ for } i = 1, 2, \cdots, \min(M, N) \tag{2.13}$$

where $d_i$ are the diagonal elements of $\mathbf{D}$ and $\mu$ is selected such that $\Sigma_{i=1}^{\min(M,N)} s_i = M$. Here $(x)^+ = \max(0, x)$, this implies that $s_i$ can never be negative. The actual number of non-vanishing entries in $\mathbf{S}$ depends on the values of $d_i$ as well $\gamma$ (and thus $P$). Examining (2.13), its seen that when $\gamma \to 0$, the water-filling algorithm will lead to choosing $s_j = M$ and $s_i = 0$ for all $i \neq j$, where

$j$ is chosen such that $d_j = \max(d_i)$ (beamforming). Similarly for $\gamma \to \infty$, $s_i = \frac{M}{\min(M,N)}$ (uniform power allocation).

From (2.5), its seen that the parameters that can be optimized in order to maximize the energy efficiency are $\mathbf{Q}$ and $P$. Note that for every different $P$, the optimal power allocation $\mathbf{Q}$ changes according to (2.13) as $\gamma$ is directly proportional to $P$. Therefore optimizing this parameter is not a trivial exercise. Practically, $P$ represents the total radio power, that is, the total power transmitted by the antennas. This power determines the total consumed power $b + aP$, of base stations or mobile terminals and so, optimizing this power is of great importance.

In this section, a theorem on the properties of $\eta_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ is provided, where $\mathbf{Q}_{WF(P)}$ is the power allocation obtained by using the water-filling algorithm and iteratively solving (2.13) with power $P$. This procedure is said to be "iterative" because, after solving equation 2.13, if any $s_j < 0$, then set $s_j = 0$ and the equation is resolved until the all solutions are positive. For optimization, desirable properties on $\eta_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ are differentiability, quasi-concavity and the existence of a maximum. The following theorem states that these properties are in fact satisfied by $\eta_T$.

**Theorem 2.1.1** *The energy-efficiency function $\eta_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ is quasi-concave with respect to $P$ and has a unique maximum $\eta_T(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}})$, where $P^*$ satisfies the following equation :*

$$\begin{aligned} &\tfrac{\partial F_L[I_{\text{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi]}{\partial P} \left(P^* + \tfrac{b}{a}\right) \\ &-F_L[I_{\text{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi] \quad = 0 \end{aligned} \tag{2.14}$$

*where $\frac{\partial}{\partial P}$ is the partial derivative.*

The proof of this theorem can be found in Appendix A.2. Thus, the optimal transmit power for imperfect CSITR depends on several factors like

- the channel estimate $\widehat{\mathbf{H}}$,
- the target spectral efficiency $\xi$,
- the ratio of the constant power consumption to the radio-frequency (RF) power efficiency $\frac{b}{a}$,
- the channel training time $t$ and
- the noise level $\sigma^2$.

Note that in this model, its assumed that the CSI at the transmitter is exactly identical to CSI at the receiver. Because of this, consider the feedback mechanism to be perfect and take a constant time $t_f$. Although in practice, $t_f$ plays a role in determining the efficiency and the optimal power, in the model $t_f$ is a constant and does not appear in the equation for $P^*$.

### 2.1.3 Imperfect CSIR and no CSIT

As discussed in the introduction, when re-transmission is possible, maximizing the energy efficiency is preferred over power minimization. For the case without re-transmission, power minimization with a quality of service constraint can be chosen as shown in A.1. This can lead to some interesting results on the choice of number of transmit antennas in a MIMO system when under imperfect CSIR. The problem with re-transmissions, has already been well analyzed in [11] when perfect CSI is available at the receiver and $b = 0$. So, in this section consider the case when imperfect CSI is available and is obtained through channel training. For $I_{ICSIR}(P, t, \mathbf{H})$, use a lower bound on the mutual information obtained from the equivalent observation equation (2.7), derived in [23]:

$$I_{ICSIR}(P, t, \hat{\mathbf{H}}) = \log \left| \mathbf{I}_M + \frac{1}{M} \gamma_{\text{eff}} \left( \frac{LP}{\sigma^2}, t \right) \hat{\mathbf{H}} \hat{\mathbf{H}}^H \right| \tag{2.15}$$

Note that here, $Q = \frac{\mathbf{I}_M}{M}$ is used and has been shown to be optimal in [11]. In this section the focus is to generalize [11] to a more realistic scenario where the total power consumed by the transmitter (instead of the radiated power only) and imperfect channel knowledge are accounted for.

By inspecting (2.6) and (2.15) its seen that using all the available transmit power can be suboptimal. For instance, if the available power is large and all of it is used, then $\eta_R(P, t)$ tends to zero. Since $\eta_R(P, t)$ also tends to zero when $P$ goes to zero (see [11]), there must be at least one maximum at which energy-efficiency is maximized, showing the importance of using the optimal fraction of the available power in certain regimes. The objective of this section is to study those aspects namely, to show that $\eta_R$ has a unique maximum for a fixed training time length and provide the equation determining the optimum value of the transmit power.

From [37] its known that a sufficient condition for the function $\frac{f(x)}{x}$ to have a unique maximum is that the function $f(x)$ be sigmoidal. To apply this result in the context, one can define the function $f$ by

$$f(\gamma_{\text{eff}}) = \Pr \left[ \log \left| \mathbf{I}_M + \frac{1}{M} \gamma_{\text{eff}} \mathbf{H} \mathbf{H}^H \right| \geq \xi \right]. \tag{2.16}$$

For the SISO case, for a channel with $h$ following a complex normal distribution, it can be derived that $f(\gamma) = \exp\left(-\frac{2^\xi - 1}{\gamma}\right)$ which is sigmoidal. It turns out that proving that $f$ is sigmoidal in the general case of MIMO is a non-trivial problem, as advocated by the current state of relevant literature [11, 38, 39]. In [11], $\eta_R(P)$ under perfect CSIR, was conjectured to be quasi-concave for general MIMO, and proven to be quasi-concave for the follwing special cases:

**(a)** $M \geq 1$, $N = 1$;

**(b)** $M \to +\infty$, $N < +\infty$, $\lim_{M\to\infty} \frac{N}{M} = 0$;
**(c)** $M < +\infty$, $N \to +\infty$, $\lim_{N\to\infty} \frac{M}{N} = 0$;
**(d)** $M \to +\infty$, $N \to +\infty$, $\lim\limits_{M\to+\infty, N\to+\infty} \dfrac{M}{N} = \ell < +\infty$;
**(e)** $\sigma^2 \to 0$;
**(f)** $\sigma^2 \to +\infty$;

In the following proposition, a sufficient condition to ensure that $\eta_R(P,t)$ is quasi-concave w.r.t $P$ is given.

**Proposition 2.1.2 (Optimization of $\eta_R(P,t)$ w.r.t $P$)** *If $\eta_R(P)$ with perfect CSIR is quasi-concave w.r.t $P$, then $\eta_R(P,t)$ is a quasi-concave function with respect to $P$, and has a unique maximum.*

This proposition is proved in Appendix A.2.

Quasi-concavity is an attractive property for the energy-efficiency as quasi-concave functions can be easily optimized numerically. Additionally, this property can also be used in multi-user scenarios for optimization and for proving the existence of a Nash Equilibrium in energy-efficient power control games [8, 40, 41].

The expression of $\eta_R(P,t)$ shows that only the numerator depends on the fraction of training time. Choosing $t = 0$ maximizes $1 - \frac{t}{T}$ but the block success rate vanishes. Choosing $t = T$ maximizes the latter but makes the former term go to zero. Again, there is an optimal trade-off to be found. Interestingly, it is possible to show that the function $\eta_R(P^*,t)$ is strictly concave w.r.t. $t$ for any MIMO channels in terms of $(M,N)$, where $P^*$ is a maximum of $\eta_R$ w.r.t $P$. This property can be useful when performing a joint optimization of $\eta_R$ with respect to both $P$ and $t$ simultaneously. This is what the following proposition states.

**Proposition 2.1.3 (Maximization of $\eta(P^*(t),t)$ w.r.t $t$)** *The energy-efficiency function $\eta_R(P^*(t),t)$ is a strictly concave function with respect to $t$ for any $P^*(t)$ satisfying $\frac{\partial\eta_R}{\partial P}(P^*,t) = 0$ and $\frac{\partial^2\eta_R}{\partial P^2}(P^*,t) < 0$, i.e, at the maximum of $\eta_R$ w.r.t. $P$.*

The proof of this proposition is provided in Appendix A.2. The parameter space of $\eta_R$ is two dimensional and continuous as both $P$ and $t$ are continuous and thus the set $\eta(P^*(t),t)$ is also continuous and the proposition is mathematically sound. The proposition assures that the energy-efficiency can been maximized w.r.t. the transmit power and the training time jointly, provided $\eta_R(P,t)$ is quasi-concave w.r.t $P$ for all $t$.

Note that the energy-efficiency function is shown to be concave only when it has already been optimized w.r.t $P$. The optimization problem studied here is basically, a joint-optimization problem, and its shown that once $\eta(P,t)$ is maximized w.r.t $P$ for all $t$, then, $\eta(P^*(t),t)$ is concave w.r.t $t$.

### 2.1.4 Numerical results and interpretations

Several simulations that support the proposed conjectures as well as expand on the analytical results are presented here. All simulations are performed using Monte-Carlo simulations as there is no expression available for the outage of a general MIMO system.

The $F_L$ used here is based on the results in [28], $F_L = Q_{func}(\frac{\xi - I_{\text{ICSITR}}(P,\mathbf{Q}_{WF},\mathbf{H})}{\sqrt{\frac{2\gamma}{(1+\gamma)L}}})$, $L$ being the code-length. This is the Gaussian approximation that is very accurate for $L$ large enough and from simulations observe that for $L \geq 10$ the approximation is quite valid.

First of all, numerical results are presented that support and present the analytical results through figures. The first two figures shown assume imperfect CSITR obtained through training and use a $2 \times 2$ MIMO system. The quasi-concavity of the the energy-efficiency function w.r.t the transmit power is shown in Figure 2.1 for $\xi = 1$ and $\xi = 4$, and $t_s = 2$ and $t_s = 10$. This figure shows that for a higher target rate, a longer training time yields a better energy-efficiency.

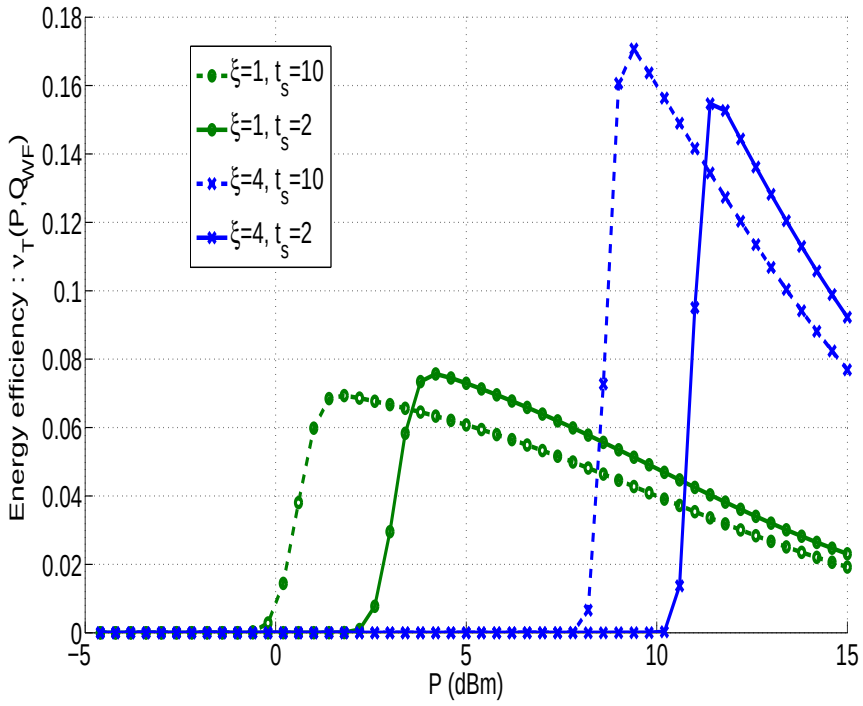


Figure 2.1: CSIT: Energy efficiency ($\eta_T$) in bits/J v.s transmit power (P) in dBm for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $\frac{b}{a} = 10$ mW for certain values of $\xi$ and $t_s$.

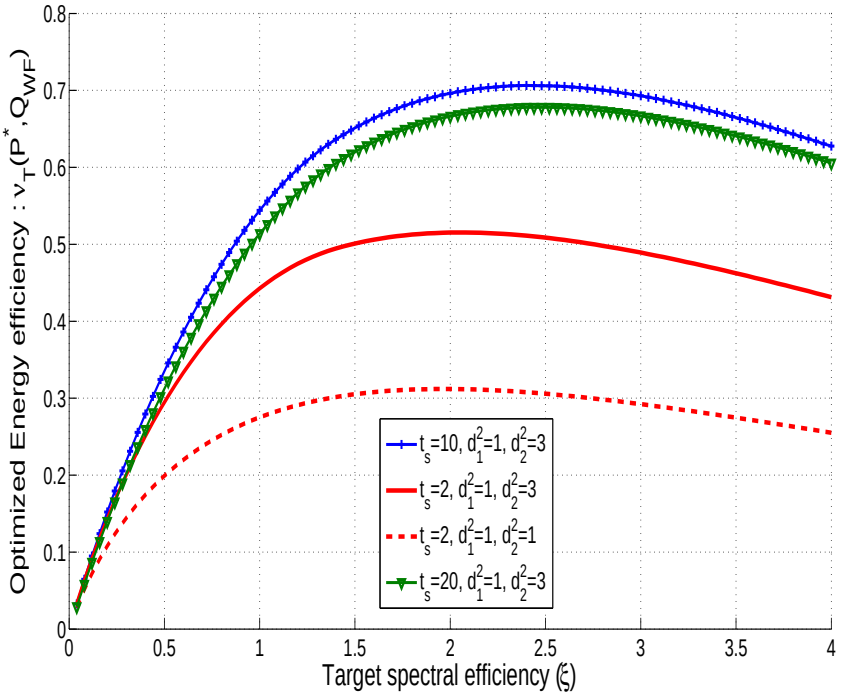


Figure 2.2: CSIT: Optimal energy-efficiency ($\eta_T(P^*, \mathbf{Q}_{WF})$) in bits/J v.s spectral efficiency ($\xi$) for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $L = 100$ and $\frac{b}{a} = 1$ mW.

It can be observed that the plots are quasi-concave and so there is an optimal target rate to use for each channel condition and code-length. In Figure 2.2, $\eta_T$ is always optimized over $P$ and $\mathbf{Q}$. Observe that $\eta_T^*(\xi)$ is also quasi-concave and has a unique maximum for each value of $d_i$ and $t_s$ (representing the channel Eigen-values as from equations (2.12), (2.13) and training time lengths). $d_i$ is ordered in an ascending order, i.e. in this case, with $d_1^2 \leq d_2^2$. The parameters used are: $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $L = 100$ and $\frac{b}{a} = 1$ mW with $t_s = 2, 10$ and 20 for $d_1^2 = 1$, $d_2^2 = 3$, and $t_s = 2$ for $d_1^2 = d_2^2 = 1$.This figure also implies that the training time and target rate can be optimized to yield the maximum energy-efficiency for a given coherence-time and channel fading.

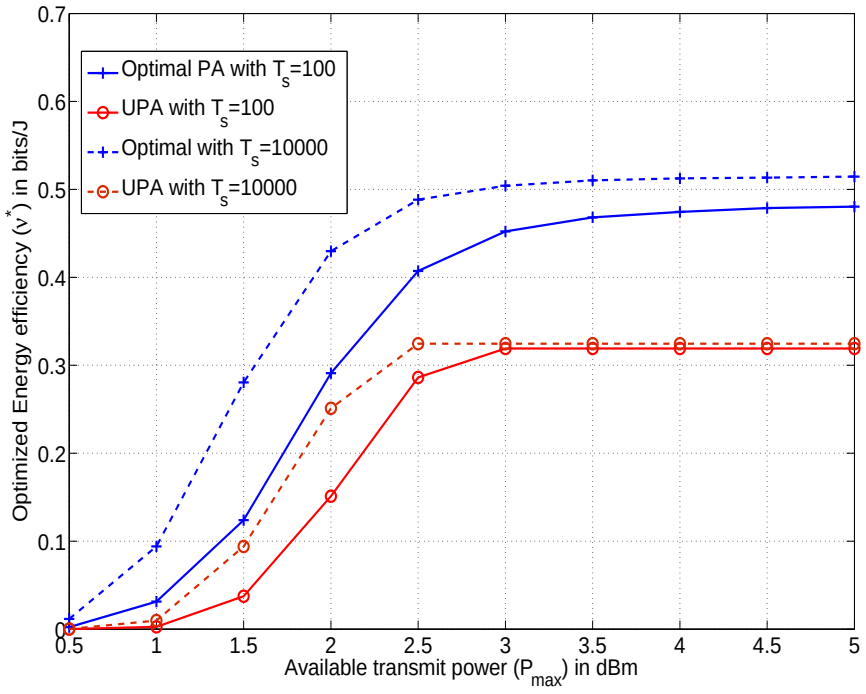


Figure 2.3: Optimal energy-efficiency ($\eta_R(P^*)$) in bits/J v.s available transmit power ($\sup(P)$) for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $R = 1$bps and $\frac{b}{a} = 1$ mW.

In Figure 2.3, the energy efficiency function that uses optimized power allocation is compared to uniform power allocation. In both cases, the training time and the transmit power is optimized and plot the optimized energy efficiency v.s $P_{\max}$. Note that the optimized PA always yields a better performance when compared to UPA and at low power, UPA has almost zero efficiency while the optimal PA yields a finite efficiency. The gain observed can be considered as the major justification in using non-uniform power allocation and sending the channel state information to the transmitter. However, when the block length is small, imperfect CSIT results in a smaller gain as seen from the relatively larger gap between $T_s = 100$ and $T_s = 10000$ when compared to the size of the gap in UPA.

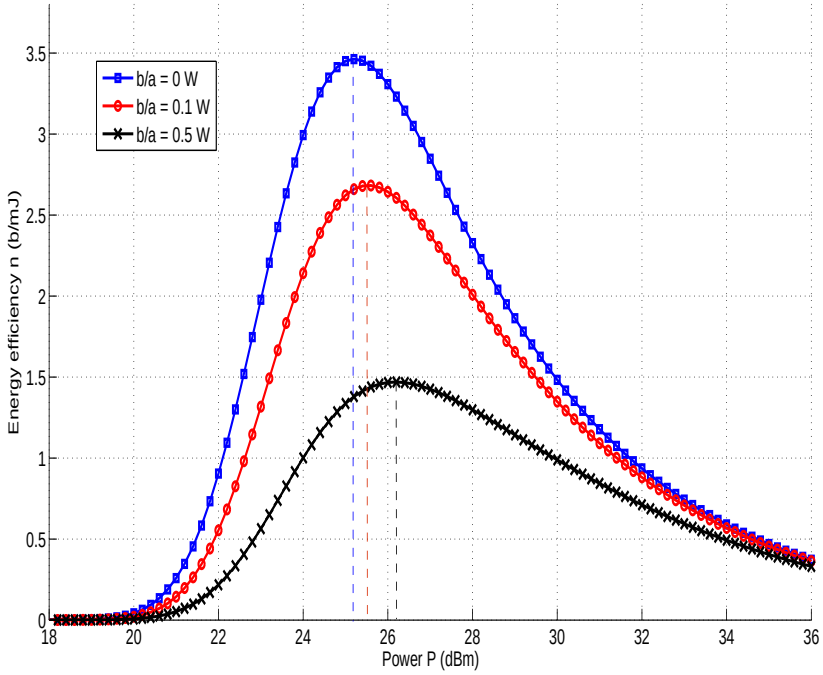


Figure 2.4: No CSIT: Energy efficiency ($\eta_R$) v.s transmit power (P) with $t_s = M = N = 4$,$R = 1600$bps, $\xi = \frac{R}{R_0}$=16 and $T_s = 55$ symbols.

In the following plots, the case where no CSIT is available is analyzed. $\frac{\sigma^2}{L} = 1mW$ and so the power $P$ can be expressed in $dBm$ easily. Also note that $\frac{b}{a}$ has the unit of power and is expressed in Watts (W). $S_d = 15$ $\mu$s from LTE standards [42]. Figure 2.4 studies the energy efficiency as a function of the transmit power ($P$) for different values of $\frac{b}{a}$ and illustrates the quasi-concavity of the energy efficiency function w.r.t $P$. The parameters used are $R = 1600$, $\xi = \frac{R}{R_0} = 16$, $T_s = 55$ and $M = N = t = 4$.

## 2.2 Multi-user MIMO energy-efficiency

When many users collectively use MIMO, it can either be done through combining point-to-point MIMO with OFDM or through virtual MIMO. As the first case is simply a trivial extension of the previous section; the goal of this section is to provide insights on how to design green radio access networks, especially in the framework of virtual MIMO systems. Indeed, classical network architectures are focused on integrated, macro base stations, where each cell covers a pre-determined area, and inter-cell interference is

reduced by the means of fixed frequency reuse patterns [46]. Heterogeneous Networks (HetNets) introduced a new notion of small cells where pico or femto base stations are deployed within the coverage area of the macro base stations [47]. Virtual MIMO is a step forward in this context that allows distributed systems of base stations/antennas that cover a common area and cooperate in order to increase the overall spectral efficiency [48]. This section focuses on these latter solutions and aims at addressing the problem from an energy efficiency point of view.

While using many non-localized antennas, a major issue considering the energy efficiency is the added energy cost of deploying several antenna. Sleep mode mechanisms have thus been regarded as a solution for this issue; they consist in deactivating network resources that have low traffic load, eliminating thus both the variable and constant parts of the energy consumption [45]. This mechanism has been applied to macro networks [45], as well as to heterogeneous networks with macro and small cells [47]. This section aims to extend this concept to virtual MIMO networks, where an antenna that is not significantly contributing to the network capacity (for a given configuration of user positions and radio channels) is put into sleep mode.

### 2.2.1 System model

The wireless system under consideration is the downlink in a virtual MIMO system within a small cell cluster. To be precise, each of the small cell base stations are connected to a central processor and so they act as antennas for the virtual MIMO as shown in Fig 2.5.

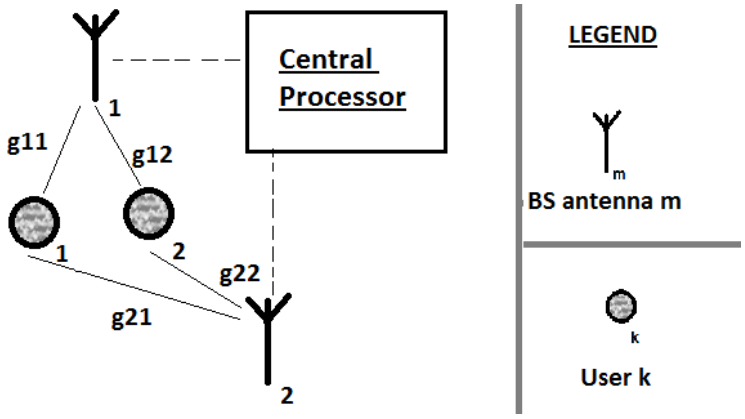


Figure 2.5: An example illustration of a $2 \times 2$ virtual MIMO with $g_{i,j}$ representing the channel between BS antenna $i$ and user $j$.

Refer to the set of these base stations as the "cluster". Each user is equipped with a single receive antenna. In order to eliminate interference

zero-forcing is implemented. Consider a block-fading channel model where the channel fading stays is assumed to stay constant for the duration of the block and changes from block to block. The base stations require the channel state information available at the user end in order to implement the zero-forcing technique. Therefore, in each block channel a training and feedback mechanism happens, after which data is transmitted. Also assume that every base station is capable of entering into a "sleep-mode". In this mode, the base station does not send any pilot signals and therefore does not perform the training or feedback actions consuming a lesser quantity of power compared to the active base stations. Let there be $M$ base stations in the cluster and $K$ users. Define $\mathcal{K} = \{1, 2, \ldots, K\}$ and $\mathcal{M} = \{1, 2, \ldots, M\}$ the sets of users and base station antennas.

As the transmit antennas are not co-located, each of them have an individual power budget. When a base station is active, it consumes a constant power of $b$ due to the power amplifier design and training or feedback costs. Additionally, it consumes a power $P_m\|x_m\|^2$ proportional to the radiated power, where $P_m \leq P_{\max}$ and $\|x_m\| \leq 1$ is the signal transmitted and $P_{\max}$ is the power constraint [45][49]. When it is placed on sleep mode, it is assumed that it only consumes power $c$ where $c < b$. Denote by $\mathbf{s}$ the sleep mode state vector of the cluster with elements $s_m \in \{0, 1\}$. The base station $m$ is in sleep mode when $s_m = 1$ and active when $s_m = 0$. Thus the power consumption of the m-th base station is $cs_i + (1 - s_i)(b + P_m\|x_m\|^2)$. The total power consumption of the cluster is given by:

$$P_{tot} = \sum_{m=1}^{M} cs_m + (1 - s_m)(b + P_m\|x_m\|^2) \tag{2.17}$$

For any given state of the cluster, define $\omega(\mathbf{s})$ as the total number of base stations that are active. This value can be calculated as $\omega(\mathbf{s}) = M - \sum_m s_m$. If $M < K$, zero-forcing can not be used. However, if $M > K$, and $\omega(\mathbf{s}) \geq K$, then the zero-forcing technique can be implemented by choosing $K$ base stations to transmit the data signals after all $\omega(\mathbf{s})$ active base stations train and obtain feedback on their channels. The communication between the active base stations and the users is done using zero-forcing pre-coding and the details on this method are given in A.3.

### 2.2.2 Energy efficiency optimization

This work aims at minimizing the energy consumption by base stations. If each user in the network is connected to download some data, then the total energy consumed by the network is the total power consumed multiplied by the total duration for which the user stays connected. Energy efficiency (EE) is a metric that is often used to measure this, and maximizing the energy efficiency leads to minimizing the total energy consumed.

Before defining the EE, first calculate the total power consumption of the network. From (2.17) and the work detailed in A.3, the total power consumed is given by:

$$P_{tot}(P_0, \tilde{\mathbf{H}}, \beta) = \sum_{m=1}^{M} cs_m + (1 - s_m) \times \left( b + P_0 \left\| \frac{(\mathbf{H}^{-1}(\tilde{\mathbf{H}}, \beta)\mathbf{u})_{\beta^{-1}(m)}}{\alpha(\tilde{\mathbf{H}}, \beta)} \right\|^2 \right) \tag{2.18}$$

Here define:

$$\forall m \in \mathcal{M}; \beta^{-1}(m) = \begin{cases} j & \text{if } j \in \mathcal{N} \text{ exists } s.t\, \beta(j) = m \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

and $()_j$ is the $j-th$ element if $j \neq 0$ and is 0 if $j = 0$. In this scenario, define the instantaneous energy efficiency as:

$$\eta(P_0, \tilde{\mathbf{H}}, \beta) = \frac{\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}, \beta))}{P_{tot}(P_0, \tilde{\mathbf{H}}, \beta)} \tag{2.20}$$

where $f()$ gives the effective throughput as a function of the SINR. $f(\gamma_k) = \log(1 + \gamma_k)$ for example. However the base station energy efficiency for a longer duration is studied, the effects of fast fading in $\tilde{\mathbf{H}}$ gets averaged and in this case a more reasonable definition for the EE is:

$$\bar{\eta}(P_0, \mathbf{G}, \beta) = \frac{\mathbb{E}_{\bar{\mathbf{H}}}[\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}}), \beta))]}{\mathbb{E}_{\bar{\mathbf{H}}}[P_{tot}(P_0, \tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}}), \beta)]} \tag{2.21}$$

Some properties of the energy efficiency metric w.r.t $P_0$ are discussed below. If the goal of a system is to be energy efficient using power control, then one important question arises: Is there a unique power for which the energy efficiency is maximized ? The answer to this question lies in the following proposition:

**Proposition 2.2.1** *Given a certain path loss matrix* $\mathbf{G}$ *and a selection of transmitting base stations* $\beta$ *in the virtual MIMO system, the average EE* $\bar{\eta}(P_0, \mathbf{G}, \beta)$ *is maximized for a unique* $P_0^*$ *and is quasi-concave in* $P_0$*.*

The proof and further discussion can be found from A.3.

Given a certain sleep mode state $\mathbf{s}$, there are $\omega(\mathbf{s})$ base stations active that train and obtain feedback. From this set $\zeta$, $K$ base stations have to be picked for zero-forcing. This choice is mathematically expressed by $\beta$. The $\beta$ that optimizes the energy efficiency depends on the channel state $\tilde{\mathbf{H}}$. The following proposition details the method of choosing the $\beta$ that optimizes EE.

**Proposition 2.2.2** *When* $\frac{P_0}{b} \to 0$*, the* $\beta^*$ *that maximizes* $\bar{\eta}(P_0, \mathbf{G}, \beta)$ *is obtained by:*

$$\beta^* = \arg\min[\alpha(\tilde{\mathbf{H}}, \beta);\ \beta \in \{\mathcal{N} \to \mathcal{K}\}] \tag{2.22}$$

The proof and further discussion can be found from A.3.

### 2.2.3 Numerical results

In this section, simulations and numerical calculations are used to study the effectiveness of the proposal as well as the advantages offered. The common parameters are:

1. $c = \frac{b}{10}$ W
2. $P_{\max} = 2$ W
3. $f(\gamma) = B\log(1+\gamma)$
4. $\sigma^2 = 1$ mW

Where $B = 10^6$ hz is the bandwidth.

The fast fading co-efficient consider is $\bar{h}_{i,j} = o(\pi_{m,k})\Omega + 0.1\xi$. Where $\xi \in CN(0,1)$, a is the direct line of sight factor which plays a dominant role in most small cell networks, $o_{\pi_{m,k}} \in 0, 1$ is the shadow factor and $o(pi_{m,k}) = 1$ with probability $\pi_{m,k}$. Here $\pi_{m,k}$ is the probability that the receiver $k$ has line of sight with the BS antenna $m$. Take $\pi_{m,k} = 0.5\,\forall(k,m)$.

The presented results study the case of two users $K = 2$ served by a small cell cluster of three base stations, i.e $M = 3$. In addition to zero-forcing, when there are two users a single base station could also alternately use Orthogonal Frequency-Division Multiple Access (OFDMA) to serve the two users and keep the other two BS in sleep mode. Two main regimes of interest are plotted:

1. $b = 1$W : This regime represents the futuristic case where power amplifier efficiencies are quite high and the constant power consumed is lower than the maximum RF output power.
2. $b = 10$W : This regime represents the more current state of the art w.r.t power amplifier efficiency where in small cell antennas, a large portion of the power is lost as a fixed cost.

Two possible values of $\Omega$, the line of sight factor, are considered. The case $\Omega = 10$ is representative of pico-cells that are deployed externally, whereas the case $\Omega = 0$ represents femto-cells deployed internally and no line of sight communication is possible.

The deployment of antennas and the user locations are shown in Fig 2.6. In this setting take $g_{1,1} = g_{2,1} = 4$, $g_{3,1} = g_{1,2} = g_{2,2} = 0.1$ and $g_{3,2} = 10$.

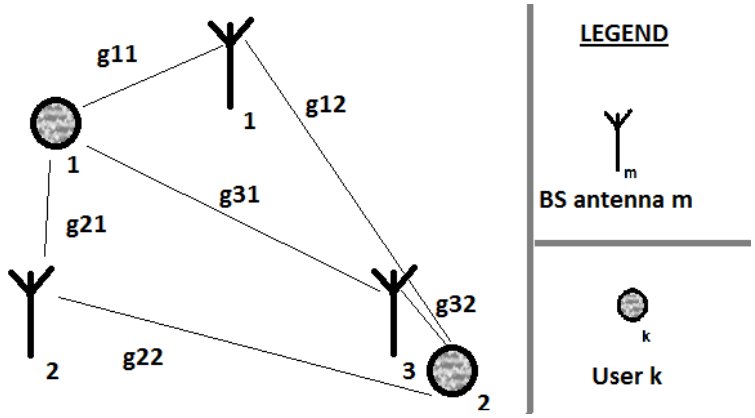


Figure 2.6: Setting B schematic

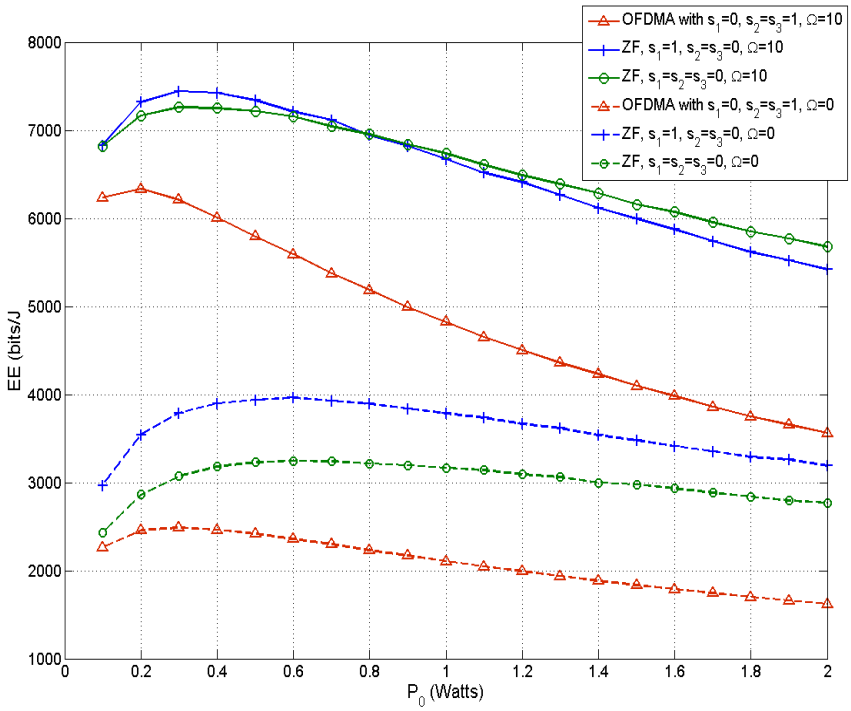


Figure 2.7: Setting B: EE v.s $P_0$ for $b = 1$ W

In Fig 2.7, similarly to what was done in the previous setting, study the EE of a VMIMO system with a very efficient power amplifier. In this figure, for both $\Omega = 0$ and $\Omega = 1$ its seen that having to use 2 BS antennas and put

one on sleep mode is the most efficient. In this setting, the configuration of BS and users are asymmetric and the BS to be put in sleep mode has to be chosen carefully. BS 1 and 2 are symmetric and are close to user 1, but 3 is closer to user 2. In this case choosing $s_1 = 1$ or $s_2 = 1$ is efficient, but $s_3 = 1$ is highly inefficient.

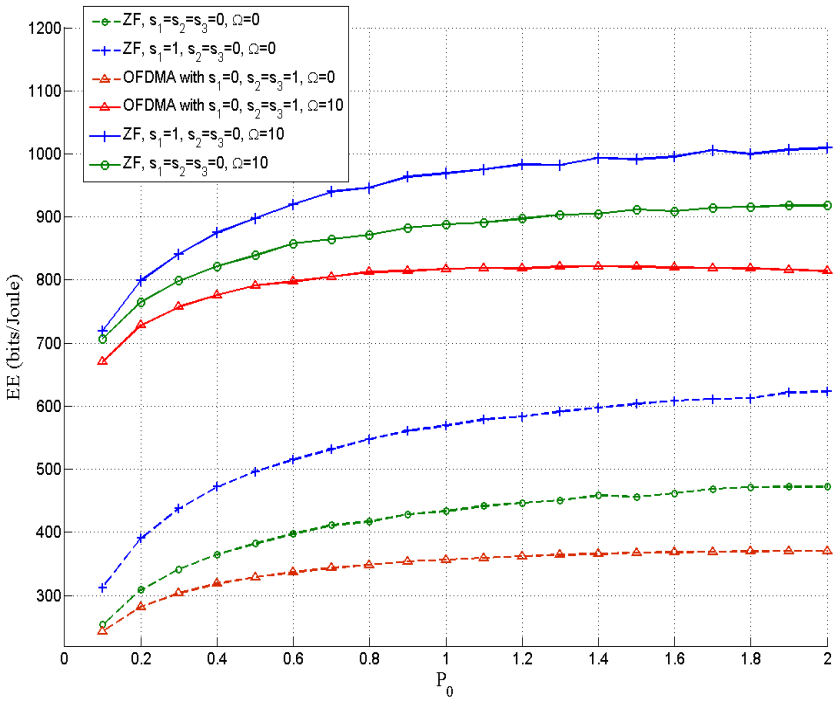


Figure 2.8: Setting B: EE v.s $P_0$ for $b = 10$ W

In this setting, its seen from Fig 2.8 that using OFDMA to divide resources between the two users is not as efficient as ZF due to the higher SNR when served by nearby BS antennas. Like in Fig 2.7, choosing $s_1 = 1$ or $s_2 = 1$ and zero-forcing is always the most efficient solution.

Additional numerical results where OFDM with sleep mode can be more efficient over ZF, can be found from A.3.

# Chapter 3

# A cross-layer approach to energy efficiency

In this chapter, as described in the introduction, the practical case of random packet arrivals onto the packet layer queue are accounted for while formulating the energy efficiency metric. This leads to a very interesting set of problems and questions about the properties of this "new" metric. One of the main questions answered is, "Does the work in [8] hold true even with this more practical metric under consideration?". If so, what sort of algorithm or process can be used for decentralized power control?

## 3.1 System model

The purpose of this section is to describe the communication model considered for cross-layer energy-efficient power control, which consists in expressing the SINR and packet arrival rate for a given user. A general interference network is considered with $N$ transmitter-receiver pairs, in which each transmitter communicates with its respective receiver, while under interference from the other transmitters [50]. Let $\mathcal{N} = \{1, 2, ..., N\}$ be the set of transmitters. Transmitter $i \in \mathcal{N}$ transmits with power level $p_i \in [0, P_{\max}]$, where $P_{\max} > 0$ is the maximum possible transmit power, which is identical for all transmitters (the analysis does not lose its generality with this assumption). The vector $\underline{p} = (p_1, p_2, \ldots, p_N)$ will be referred to as the power or action profile on the current data block or packet. Denote by $\underline{p}_{-i}$, the $(N-1)$ dimensional vector obtained by removing the $i^{\text{th}}$ component from $\underline{p}$. For notational simplicity, sometimes $\underline{p}$ is represented as $(p_i, \underline{p}_{-i})$, when the dependence of certain functions on $p_i$ has to be shown explicitly. By transmitting at $p_i$, each user $i$ has a resulting SINR $\gamma_i$ at his receiver of

interest which is a function of $\underline{p}$, and is assumed to be given by:

$$\gamma_i(\underline{p}) = \frac{p_i g_{ii}}{\sigma_i^2 + \sum_{j=1, j\neq i}^{N} p_j g_{ji}} \tag{3.1}$$

where $g_{ji}$ represents the quasi-static or block fading channel gain of the link between transmitter $j$ and receiver $i$ on a given band, $\sigma_i^2 = \sigma^2$ is the variance of the Gaussian noise at receiver $i$ (these variances can be assumed to be equal without any loss of mathematical generality). In wireless systems such as those being implemented in recent cellular system standards, packets arrive from an upper layer (e.g. IP layer) following an arrival rate that is related to the SINR. In this chapter, assume that the packet arrival process follows a Bernoulli process with probability $q_{\mathrm{X}}(\gamma_i(\underline{p}))$ where $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$, CAR corresponding to constant arrival rate and AAR corresponding to adaptive arrival rate; this corresponds to the classical ON/OFF sources [51]. In the case of CAR, it trivially expresses as:

$$\forall i \in \mathcal{N},\ q_{\mathrm{CAR}}(\gamma_i(\underline{p})) = q \tag{3.2}$$

with $q \in [0, 1]$. This is best used for real-time applications where delay is not tolerable, however, in some applications this packet arrival model is not suitable. For instance, this is the case for applications such as file transfer or browsing. In such a situation, there is no constant stream of data and so the arrival rate can be optimized for best performance in terms of data rate and QoS. This is one of the reasons why the case of AAR is investigated for which the arrival rate is given by:

$$\forall i \in \mathcal{N},\ q_{\mathrm{AAR}}(\gamma_i(\underline{p})) = g(\Phi_{\mathrm{AAR}}(\gamma_i(\underline{p}))) \tag{3.3}$$

where $\Phi_{\mathrm{AAR}}$ is the packet loss function and $g$ is a function which is assumed to be continuous, invertible, and has an inverse function $g^{-1}$ which is twice differentiable, decreasing, and convex. To provide a specific example, the widely used and very useful approximation of the arrival rate process for the Transmission Control Protocol (TCP), which is due to [52], verifies these conditions. Therein, $g$ is merely given by $g(\Phi) = \frac{\kappa}{\sqrt{\Phi}}$, where $\kappa \in [0, 1]$ is a parameter which depends on the system design and the round trip time. The resulting rate can be interpreted as the average value for the rate.

## 3.2 A new energy-efficiency performance metric

### 3.2.1 Construction

If $\Phi_{\mathrm{X}}$, $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$, represents the packet loss due to both bad channel conditions and packet buffer finiteness (more details about this is

provided a little further), a packet is re-transmitted [1] $\frac{q_{\mathrm{X}}(\gamma_i)[1-\Phi_{\mathrm{X}}(\gamma_i)]}{f(\gamma_i)}$ times on average, the average power consumption is $b + p_i \frac{q_{\mathrm{X}}(\gamma_i)[1-\Phi_{\mathrm{X}}(\gamma_i)]}{f(\gamma_i)}$. Since the net data rate or goodput is given by $Rq_{\mathrm{X}}(\gamma_i)[1-\Phi_{\mathrm{X}}(\gamma_i)]$, now define the EE metric $\eta_{i,\mathrm{X}}(\underline{p})$ as the ratio between the average net data transmission rate and the average power consumption, which gives:

$$\eta_{i,X}(\underline{p}) = R\frac{q_{\mathrm{X}}(\gamma_i(\underline{p}))\left[1-\Phi_{\mathrm{X}}(\gamma_i(\underline{p}))\right]}{b + p_i\frac{q_{\mathrm{X}}(\gamma_i(\underline{p}))\left[1-\Phi_{\mathrm{X}}(\gamma_i(\underline{p}))\right]}{f(\gamma_i)}}. \tag{3.4}$$

This definition shows that the cross-layer design approach of power control is fully relevant in terms of EE when the transmitter has a cost, which is independent of the radiated power; otherwise, when $b = 0$ the EE function falls into the original framework of [8].

By considering the stationary regime of the queue and assuming the protocol X, the fraction of lost packets $\Phi_{\mathrm{X}}$ can be expressed as follows:

$$\Phi_{\mathrm{X}}(\gamma_i(\underline{p})) = [1 - f(\gamma_i(\underline{p}))]\Pi_{\mathrm{X}}(\gamma_i(\underline{p})) \tag{3.5}$$

where $\Pi_{\mathrm{X}}(\gamma_i)$ is the stationary probability that the packet buffer is full. Indeed, as already mentioned, each transmitter is assumed to be equipped with a device that allows the packets to be stored in a memory buffer (of size $K \geq 1$) before transmission. Packets arrive into the buffer and get transmitted through a queuing process at the buffer. Denote by $Q_{i,t}$ the size of the queue for transmitter $i$ at time slot $t$. The size of the queue $Q_{i,t}$ is a Markov process on the state space $\mathcal{Q}_i = \{0, 1, \ldots, K\}$. It is known (see [57] for example) that in the stationary regime of the stochastic process $Q_{i,t}$ the probability that the size of the queue equals $K$ is given by:

$$\Pi_{\mathrm{X}}(\gamma_i(\underline{p})) = \frac{\omega_{\mathrm{X}}^K(\gamma_i(\underline{p}))}{1 + \omega_{\mathrm{X}}(\gamma_i(\underline{p})) + \ldots + \omega_{\mathrm{X}}^K(\gamma_i(\underline{p}))} \tag{3.6}$$

with

$$\omega_{\mathrm{X}}(\gamma_i(\underline{p})) = \frac{q_{\mathrm{X}}(\gamma_i(\underline{p}))\left[1 - f(\gamma_i(\underline{p}))\right]}{\left[1 - q_{\mathrm{X}}(\gamma_i(\underline{p}))\right] f(\gamma_i(\underline{p}))} \tag{3.7}$$

where $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$.

In the case of $X = \mathrm{AAR}$, the packet arrival rate $q_{\mathrm{AAR}}$ is a function of the packet loss and the packet loss, a function of $q_{\mathrm{AAR}}$. The following proposition ensures that the AAR process achieves an average packet arrival rate according to the following proposition.

**Proposition 3.2.1** *The packet arrival rate $q_{\mathrm{AAR}}$ is obtained as the unique fixed point of these equations:*

$$\Phi_{\mathrm{AAR}}(\gamma_i(\underline{p})) = (1 - f(\gamma_i(\underline{p})))\Pi_{\mathrm{AAR}}(\gamma_i(\underline{p})) \tag{3.8}$$

1. For the sake of clarity, here and in other places in this chapter, $\underline{p}$ is omitted from the notations.

*where* $\Pi_{\mathrm{AAR}}(\gamma_i(\underline{p}))$ *has* $q_{\mathrm{AAR}}$ *as a parameter as seen from (3.7) and (3.6), and:*

$$q_{\mathrm{AAR}}(\Phi_{\mathrm{AAR}}) = g(\Phi_{\mathrm{AAR}}). \tag{3.9}$$

When the packet arrival is constant (i.e., X = CAR), the dependency of $\Pi_{\mathrm{X}}$ regarding the SINR follows a simple relation. However, when the AAR protocol is assumed, the relationship is less trivial. Indeed, the packet loss $\Phi_{\mathrm{X}}$ depends on $\omega_{\mathrm{X}}$ through (3.5) and (3.6). The quantity $\omega_{\mathrm{X}}$ depends on the arrival rate $q_{\mathrm{X}}$. But, in the AAR case, $q_{\mathrm{X}}$ also depends on the packet loss. This is the reason why under the AAR protocol assumption, its assumed that each transmitter operates at the fixed point associated with the aforementioned dependency chain. Therefore, this amounts to fixing the packet loss function to have a certain form. AAR can thus be seen as an indirect way of imposing a certain QoS on the transmission. A detailed discussion on this subject is provided in B.2.

### 3.2.2 Properties

The EE function $\eta_{i,\mathrm{X}}$ possesses a very attractive property regarding its dependency toward $p_i$. This is what the next proposition states.

**Proposition 3.2.2** *For all* $i \in \mathcal{N}$*, the EE function* $\eta_{i,\mathrm{CAR}}(\underline{p})$ *is quasi-concave w.r.t.* $p_i$ *and has a unique maximum point denoted by* $p_i^*(\underline{p}_{-i})$*.*

The proof relies, in particular, on the sigmoidness assumption for $f$ and can be found in App. B.2. This result is very useful for the NE analysis which is conducted in Sec. IV. Remarkably, the quasi-concavity property is not only available in the case of CAR but also in the case of AAR, which is not obvious a priori.

### 3.2.3 QoS constraint

As already mentioned in Sec. I, one of the recurrent problems with most works using the performance metric introduced in [8] is that EE can be maximized at a power level which does not guarantee a minimum QoS. This is why, in the case of CAR, also consider a constraint when maximizing (3.4): the packet loss rate $\Pi_{\mathrm{CAR}}[1 - f(\gamma_i)]$ has to be less than an upper bound $\epsilon$. For example, in cellular systems, typical values for $\epsilon$ are 0.1 or 0.01, based on the system requirements. Adding this constraint restricts the range of power usable by the transmitter by adding a lower bound on the power. This lower bound depends on the entry probability $q$ and on the size of the queue $K$.

## 3.3 Equilibrium analysis and distributed power control algorithm

Since it is assumed that transmitter $i$, $i \in \mathcal{N}$, can only control the variable $p_i$ of the $N-$variable function $\eta_{i,\mathrm{X}}(\underline{p})$ and is assumed to consider the energy-efficiency of his own communication, the power control problem is naturally distributed in terms of the decision. The ultimate goal of this section is to propose a power control algorithm which is distributed both in terms of the decision and information (only individual SINR feedback is required to adapt the power level). While the algorithm itself is directly inspired from existing works, its convergence analysis does not follow from a direct adaptation of existing results.

### 3.3.1 Equilibrium analysis of the associated games

A non-cooperative game under strategic form is merely given by an ordered triplet (see e.g., [40]). With the notations of this chapter it writes as

$$\mathcal{G}_{\mathrm{X}} = \left(\mathcal{N}, \{\mathcal{P}_i\}_{i\in\mathcal{N}}, \{u_{i,\mathrm{X}}\}_{i\in\mathcal{N}}\right) \tag{3.10}$$

where the set of decision-makers (DMs) or players is therefore the set of transmitters, the action space for DM $i$ is $\mathcal{P}_i = [0, P_{\max}]$, and $u_{i,\mathrm{X}}$ is the payoff function of DM $i$ when the arrival rate model is $X$. As explained in Sec. II, when CAR is assumed, a QoS constraint is imposed on the packet loss. Under this assumption, the payoff function is chosen to be:

$$u_{i,\mathrm{CAR}}(\underline{p}) = \begin{cases} \eta_{i,\mathrm{CAR}}(\underline{p}) & \text{if } \Phi_{\mathrm{CAR}}(\gamma_i(\underline{p})) \leq \epsilon \\ R\dfrac{q\left[1-\Phi_{\mathrm{CAR}}(\gamma_i(\underline{p}))\right]}{b+P_{\max}} & \text{otherwise} \end{cases}. \tag{3.11}$$

This payoff definition means that as long as the QoS constraint can be met, energy-efficiency maximization is pursued. However, if the constraint cannot be met, goodput maximization or packet loss minimization is sought. Note that, for any constraint $\epsilon$, the action space of any DM $i$ is still the interval $[0, P_{\max}]$. The constraint is instead merged into the payoff function in such a manner that as long as the constraint is not satisfied, it is always optimal to increase power. For the AAR case, the payoff function is simply defined by

$$u_{i,\mathrm{AAR}}(\underline{p}) = \eta_{i,\mathrm{AAR}}(\underline{p}). \tag{3.12}$$

A fundamental solution concept for a non-cooperative game is the Nash equilibrium. The purpose of the following propositions is to show that the two games discussed in this section possess a unique Nash equilibrium, which will further lead to the development of a simple algorithm that is guaranteed to converge to the aforementioned equilibrium. First, of all, the Nash equilibrium is formally defined in this context as:

**Definition 3.3.1** *The vector of transmit power levels* $\underline{p}_{\mathrm{X}}^{\mathrm{NE}}$ *is a pure Nash equilibrium of the game* $\mathcal{G}_{\mathrm{X}}$ *if:*

$$\forall i \in \mathcal{N},\ \forall p_i \in \mathcal{P}_i,\ u_{i,\mathrm{X}}(\underline{p}^{\mathrm{NE}}) \geq u_{i,\mathrm{X}}(p_i, \underline{p}_{-i,\mathrm{X}}^{\mathrm{NE}}). \tag{3.13}$$

**Proposition 3.3.2** *For* $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$*, the game* $\mathcal{G}_{\mathrm{X}}$ *admits at least one pure Nash equilibrium.*

In the current state of the art, all related works on energy-efficient power control use utilities which are continuous with the power profile $\underline{p}$. Interestingly, a relevant power control game in which continuity is not available can be exhibited for the case of $X = \mathrm{CAR}$.

**Proposition 3.3.3** *For* $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$*, the game* $\mathcal{G}_{\mathrm{X}}$ *admits a unique pure Nash equilibrium, for which the equilibrium power policy will be denoted by* $\underline{p}_{\mathrm{X}}^{\mathrm{NE}}$.

### 3.3.2 The proposed distributed interference management algorithm

The property of the previous proposition is also sufficient to guarantee convergence of some important distributed optimization algorithms. Note that the argmax set mentioned in the proof is a singleton (a scalar value), which can be checked from App. B.2. While this property is available for the scenario studied in [8] and many related works, it is seen here that, although the proposed QoS oriented cross-layer approach leads us to more complex and more general utilities, this property is still valid.

This means that for these algorithms, not only is convergence ensured, but the convergence point is also unique. This is very useful to characterize the performance of an implemented distributed power control algorithm, the best-response dynamics algorithm. This algorithm is well known in game theory [60] and draws its roots from the chapter by Cournot [61]. It has been used in [8] and is often used because convergence to the NE can be guaranteed. Let $\underline{p}_{\mathrm{X}}^{\mathrm{NE}}$ be the unique NE of $\mathcal{G}_{\mathrm{X}}$. For the algorithm, define $\underline{p}(t)$ as the power control policy in the previous time slot, and $\underline{p}(t+1)$ as the power control policy for the current time slot. Algorithm 1 implements the sequential best-response dynamics for $\mathcal{G}_X$:

Several comments are in order.

1. Its assumed that DM 1 updates first, who is followed by DM 2, etc. In fact, this order can be arbitrary provided it is fixed.
2. To update the power levels $m$ times, a duration corresponding to $mN$ time-slots is required.
3. The quantity $\delta > 0$ corresponds to the accuracy level wanted for the stopping criteria in terms of convergence to the NE.

**Algorithm 1** Sequential best-response dynamics

$\Delta \leftarrow 2\delta$ ▷ Initialize the observed difference in power levels over time, $\delta$ is the tolerance.

$\underline{p}^0 \leftarrow (P_{\max}, P_{\max}, \ldots, P_{\max})$ ▷ The starting power is uniform power with $P_{\max}$.

$t \leftarrow 0$ ▷ The starting time is 0.

**while** $\Delta \geq \delta$ **do** ▷ The outer loop that iterates till the power policies converge.

**for** $i = 1 \rightarrow N$ **do** ▷ The inner loop iterating over the DM indices.

$\Gamma_i = \dfrac{\gamma_i(\underline{p}^t)}{p_i^t}$ ▷ Using the SINR feedback from its receiver, DM i calculates the interference term $\Gamma_i$ for the previous time slot.

$p^* \leftarrow \arg\max_p \left( \dfrac{Rq_{\mathrm{X}}(p\Gamma_i)(1-\Phi_{\mathrm{X}}(p\Gamma_i))}{b + \dfrac{p}{f(p\Gamma_i)} q_{\mathrm{X}}(p\Gamma_i)(1-\Phi_{\mathrm{X}}(p\Gamma_i))} \right)$ ▷ Calculate the optimal power that maximizes the EE.

**if** X=CAR **then**

$p_+ \leftarrow \min(p; \Phi_{\mathrm{CAR}}(p\Gamma_i) \geq \epsilon)$ ▷ Calculate the minimum power to satisfy the QoS constraint.

$p_i^{t+1} \leftarrow \min(\max(p^*, p_+), P_{\max})$ ▷ Choose the optimal power for CAR if less than $P_{\max}$ and more than $p_+$.

**else**

$p_i^{t+1} \leftarrow \min(p^*, P_{\max})$ ▷ Choose the optimal power for AAR if less than $P_{\max}$.

**end if**

**end for**

$\Delta \leftarrow \max_i(|p_i^{t+1} - p_i^t|)$

$t \leftarrow t + 1$

**end while**

4. The algorithm 1 is completely distributed in the sense that to update his power, a DM only needs to know the SINR corresponding to his chosen power level, i.e., $\mathrm{BR}_{i,\mathrm{X}}(\underline{p}_{-i})$ can be calculated by knowing $\gamma_i$ for some $p_i$. This is typically achieved using a feedback mechanism and does not require a central entity that provides knowledge of the channel conditions or power levels chosen by the other DMs.

## 3.4 Numerical results

### 3.4.1 General setup

Unless explicitly stated otherwise, the following choices and parameters are assumed for all the simulations provided here:

- The number of users or transmitters is set to two ($N = 2$). This scenario was chosen because the behavior of various metrics like the price of anarchy (PoA) can be easily analyzed in this situation. The case of "high interference", as defined below, is also studied to compensate for this choice. In addition, some specific figures also study the case with more interferers.
- The block success rate function is chosen as in [11]: $f(\gamma_i) = \exp\left[-\left(\frac{2^{\frac{R}{R_0}}-1}{\gamma_i}\right)\right]$ where $R_0 = 1$ MHz is the bandwidth used and the gross data rate is $R = 1$ bit/s.
- When the adaptive arrival rate scenario is considered, it it is assumed that $g(\phi) = \frac{0.1}{\sqrt{\phi}}$.
- Define the low (resp. high) interference scenario as: $\mathbb{E}(g_{ii}) = 2.5$ and $\mathbb{E}(g_{ij}) = 0.5$ for $j \neq i$ (resp. $\mathbb{E}(g_{ii}) = 2.5$ and $\mathbb{E}(g_{ij}) = 2$ for $j \neq i$). For some simulations, the channel gains will be assumed to be fixed while for the others it will follow classical block Rayleigh fading. The values indicated will be the instantaneous channel fading when the scenario considered is static and otherwise will indicate the variance.
- The noise level is set to $\sigma^2 = 1$ mW; the maximum power $P_{\max} = 1000$ mW; buffer size of $K = 10$; $\epsilon = 1$ (packet loss constraint) and the fixed power consumption $b = 1000$ mW.
- To measure the global efficiency of the interference network with respect to the centralized solution, the price of anarchy is used. [62] gave a definition of PoA where the optimal situation correpsponds to a POA equal to 1, while other situations correspond to a PoA$>1$:

$$\forall \mathrm{X} \in \{\mathrm{CAR}, \mathrm{AAR}\},\ \mathrm{PoA}_{\mathrm{X}} = \frac{\max\limits_{\underline{p}} \sum\limits_{i} u_{i,\mathrm{X}}(\underline{p})}{\sum\limits_{i} u_{i,\mathrm{X}}(\underline{p}^{\mathrm{NE}})}. \tag{3.14}$$

### 3.4.2 About the considered EE performance metric

Assume a single-user scenario i.e., $N = 1$, a fixed channel gain (namely $g_{11} = 2.5$), and the arrival rate to be fixed (CAR scenario). Fig. 3.1 depicts the EE (3.4) as a function of the chosen radiated power for different values of the fixed consumption cost $b$ and packet arrival rate $q$. First, the figure illustrates what has already been proved through Prop. 3.2 namely, EE is quasi-concave w.r.t. the radiated power. Second, fix $q$ to one and assess the influence of $b$. As $b$ increases, the curve becomes less peaky. In fact, if $b$ becomes very high, EE tends to merely becomes a packet success rate function. This means that power control becomes irrelevant since it merely boils down to transmitting at maximum power whatever the channel conditions. Now fix $b$ to 1000 mW. By moving from the arrival rate of $q = 1$ (framework of [53]) to $q = 0.6$ (with a buffer size of $K = 10$), it is seen that the EE curve is quite significantly changed and the optimal radiated power changes from 460 mW to 320 mW. In the next section, the gain in terms of radiated power brought by the cross-layer approach is quantified in a more general scenario.

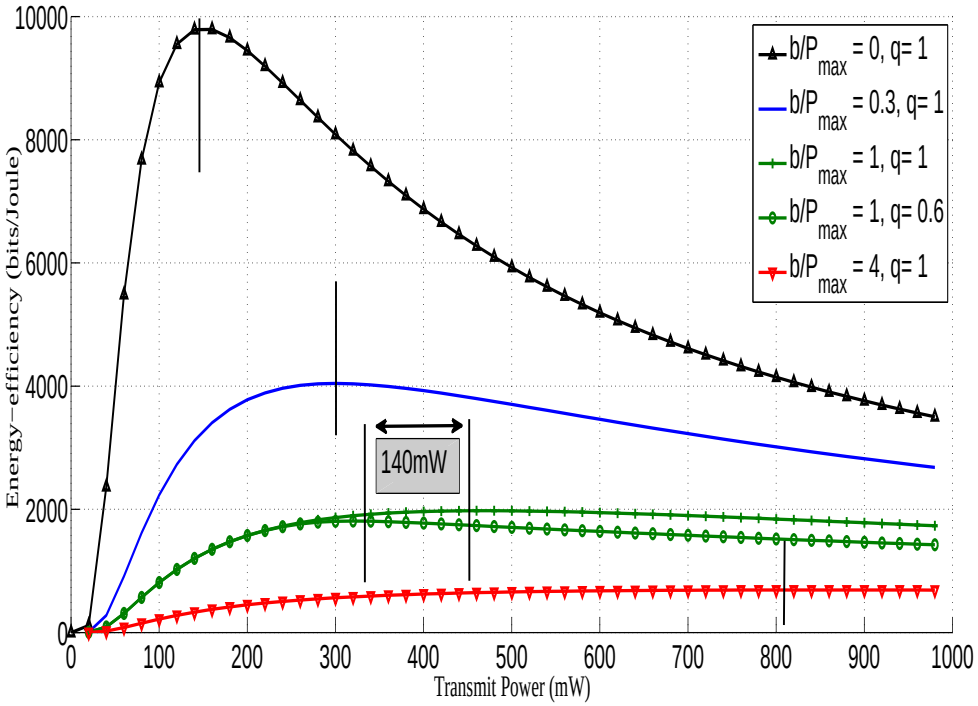


Figure 3.1: CAR: EE against $p_1$, i.e., the energy efficiency as a function of the transmit power for various values of the constant power ($b$) and packet arrival rate ($q_{CAR}$).

### 3.4.3 Influence of the packet arrival rate in the CAR scenario

The low interference scenario; For $K = 10$, Fig. 3.2 represents the gains in dB in terms of radiated power which is brought by the proposed cross-layer approach (after convergence of the proposed distributed power control algorithm) w.r.t. the conventional approach in which it is (implicitly) assumed that $q \to 1$ [53]. The gain is therefore defined by $10 \log_{10} \left( \frac{p_i^{\mathrm{NE}}[q \to 1]}{p_i^{\mathrm{NE}}[q]} \right)$, for a given $i \in \{1, 2, 3\}$, say $i = 1$ (the gain is the same for the different transmitters since the average channel gains are identical). The gain is represented as a function of the packet arrival rate. It is seen that, for different numbers of transmitter-receiver pairs ($N = 2$ or $N = 3$) and a raw packet error rate of $\epsilon = 0.1$ (by raw it is meant before re-transmission), the gain is significant if the arrival rate is typically less than 0.5. Gains as high as 10 dB (with $N - 1 = 2$ interfering users on the same band) or 30 dB (with $N - 1 = 1$ interfering user on the same band). If the raw QoS constraint is relaxed ($\epsilon = 1$), quite similar observations can be made. These gains are not

in terms of energy consumed by the whole transmit device but they mean that transmitters use much less radiated power and therefore create much less interference, while reaching the same QoS.

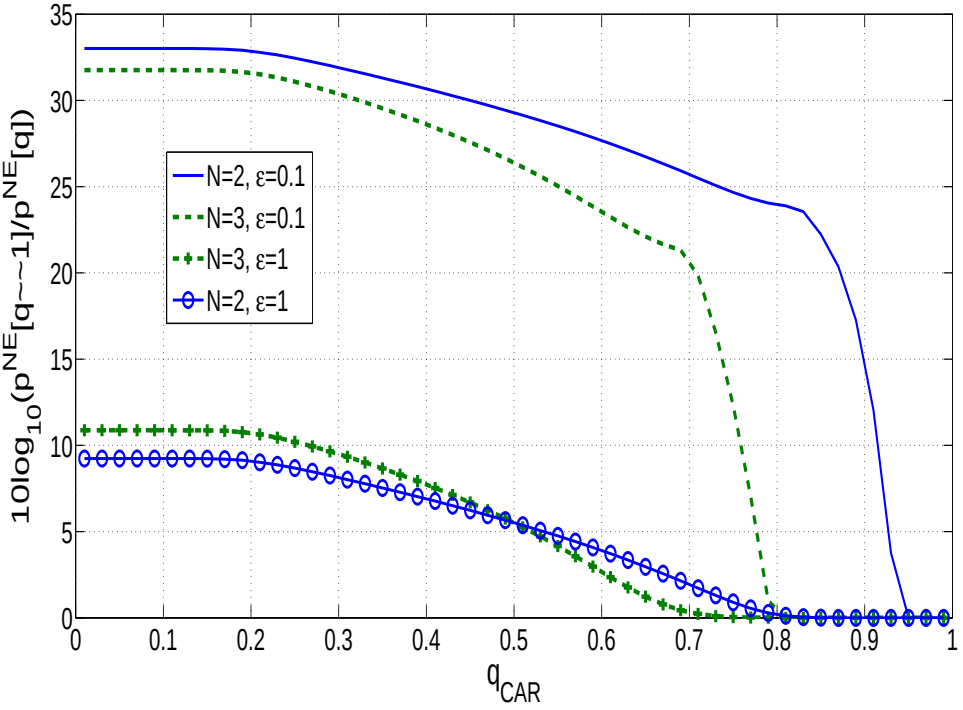


Figure 3.2: CAR: $10\log_{10}\left(\frac{p_1^{\mathrm{NE}}[q\to 1]}{p_1^{\mathrm{NE}}[q]}\right)$ against $q$, i.e., the ratio of equilibrium power levels in the cross-layer case to the case where the buffer is ignored and arrival rate is one. Interestingly, this cross-layer approach does not only allow the EE to be maximized but also allows significant gains in terms of radiated power. The transmit power for the cross-layer approach is always lower than for the purely physical layer approach, and this difference is more prominent when a packet loss constraint is imposed.

### 3.4.4 Gains in terms of energy brought by the cross-layer approach w.r.t. the state-of-the art

As explained in the introduction, maximizing energy efficiency and minimizing energy are in fact equivalent in communications systems where retransmissions are allowed. Exploiting this interpretation here to go further than just assessing the gains in terms of EE as done classically. Indeed, the gain in terms of energy or average total power brought by the proposed

cross-layer approach over the closest state-of-the art solution which is given in [53] is assessed (the latter is obtained by assuming $q \to 1$ whatever the actual value of $q$). For $q = 0.5$ and $q = 0.3$, Fig. 3.3 shows that it is possible to have improvements in terms of energy consumed by the device and not just EE. This (relative) gain can be as high as 28% for $q = 0.5$ and 42% for $q = 0.3$ in the setting under consideration. Interestingly, this gain can be obtained under the same information assumption as [53] namely, only individual SINR feedback is needed to implement the power control algorithm which provides the NE performance (after convergence). Note that in this case, $q = 1$ offers no gain as the situation is identical to that treated in [53] while $q \to 0$ would offer maximum gain.

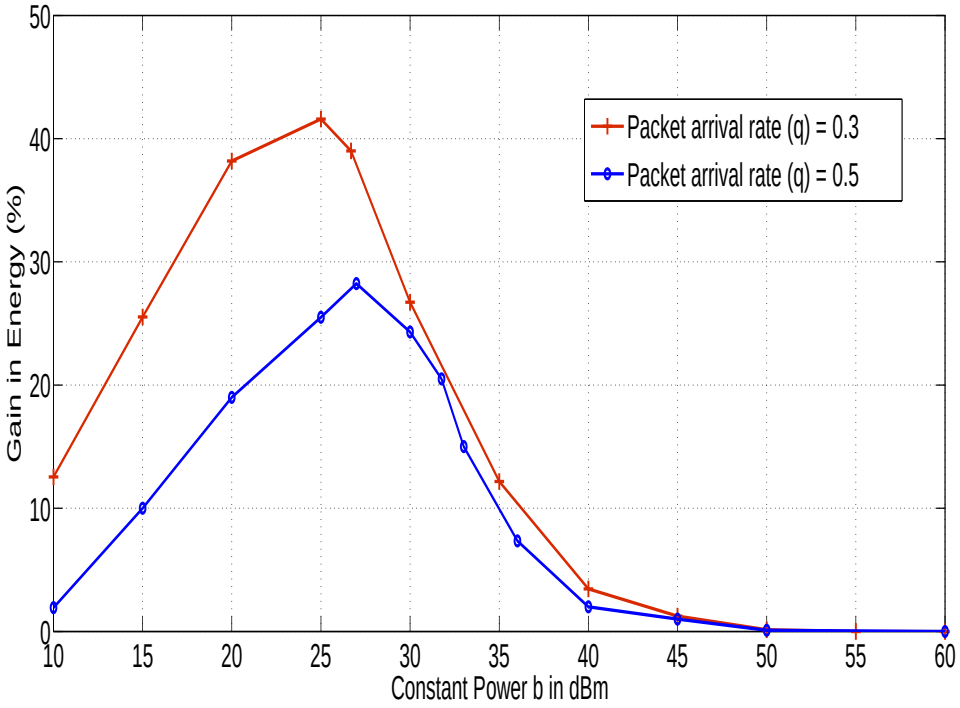


Figure 3.3: CAR: Plotting the energy consumed against $b$ with $q = 0.6$ and $q = 0.3$. Compare the performance of the proposed algorithm against using the best-response dynamics algorithm from [53] where the presence of the queue is ignored.

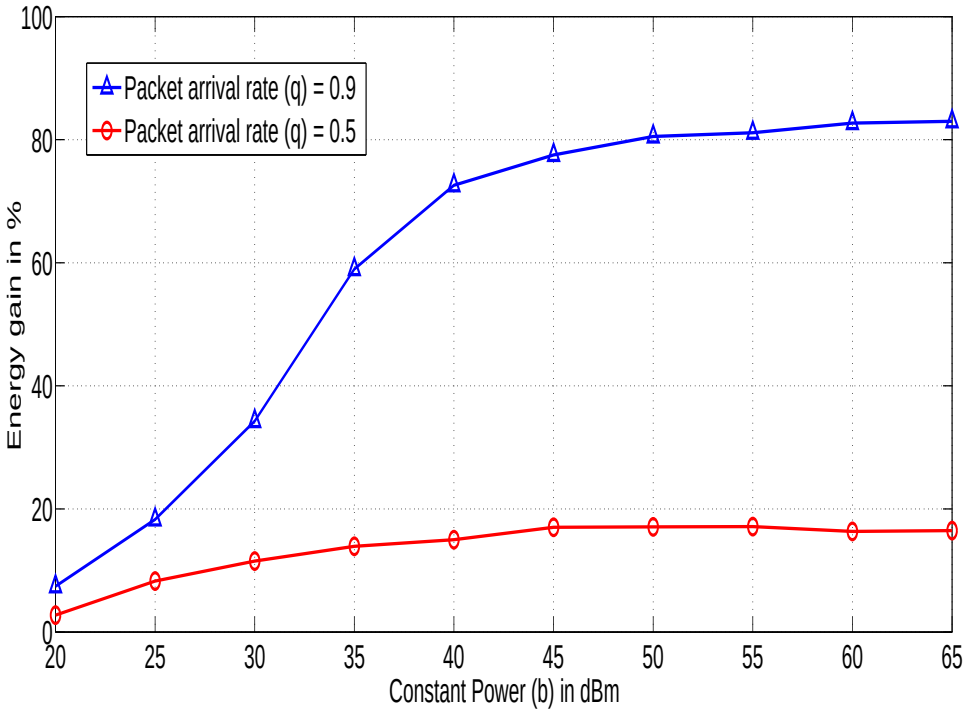


Figure 3.4: CAR: Plotting the energy consumed against $b$ with $q = 0.5$ and $q = 0.9$. Compare the performance of the proposed algorithm against a scheme that just minimizes the transmit power such that the SINR $\geq$ 25 dB.

As a second comparison in terms of energy, the energy consumed by a transmitter when optimizing (3.4) is compared with what would obtained by just minimizing the radiated power under an SINR constraint, which is a classical approach. Fig. 3.4 corresponds to the relative gain in terms of saved energy as a function of the fixed consumption cost $b$, for $q = 0.5$ and $q = 0.9$, $R = 8$ Mbps and an SINR target of 25 dB for both approaches in the single user case (interference can make achieving such a target impossible). It is seen that an energy gain of up to 80% can be achieved for sufficiently high values of $b$, which is a quite significant gain and can be easily attained in practice (e.g., maximum radiated power for femto base stations is of the order of one watt while the fixed consumption cost is typically of about a few watts). Note that the gain observed here is maximum when $q = 1$ as the highest transmit power is used in this case.

### 3.4.5 Influence of the packet buffer size in the AAR scenario

So far, CAR scenario has been treated. In particular, this has allowed the study of the influence of the parameter $q$. But, for AAR $q$ is not fixed and varies with the SINR. Fig. 3.5 represents, for different numbers of transmitters ($N \in \{2, 3, 8\}$), the network sum-payoff versus the buffer size for a static channel. The influence of interference (e.g., inter-cell interference) on global energy-efficiency clearly appears. As an important comment, as this simulation shows and many other simulations confirmed this observation (including all simulations assuming CAR instead of AAR), when the buffer size is greater than 10 typically, the asymptotic regime in terms of buffer size can be assumed to be approximately reached. In practice, this means that, when K is large enough, power control policies might be approximated by implementing the power control policies obtained by assuming $K \to +\infty$, which corresponds to switching between Cases 1 and 2 (in Sec. 3.2.2), depending on the current SINR.

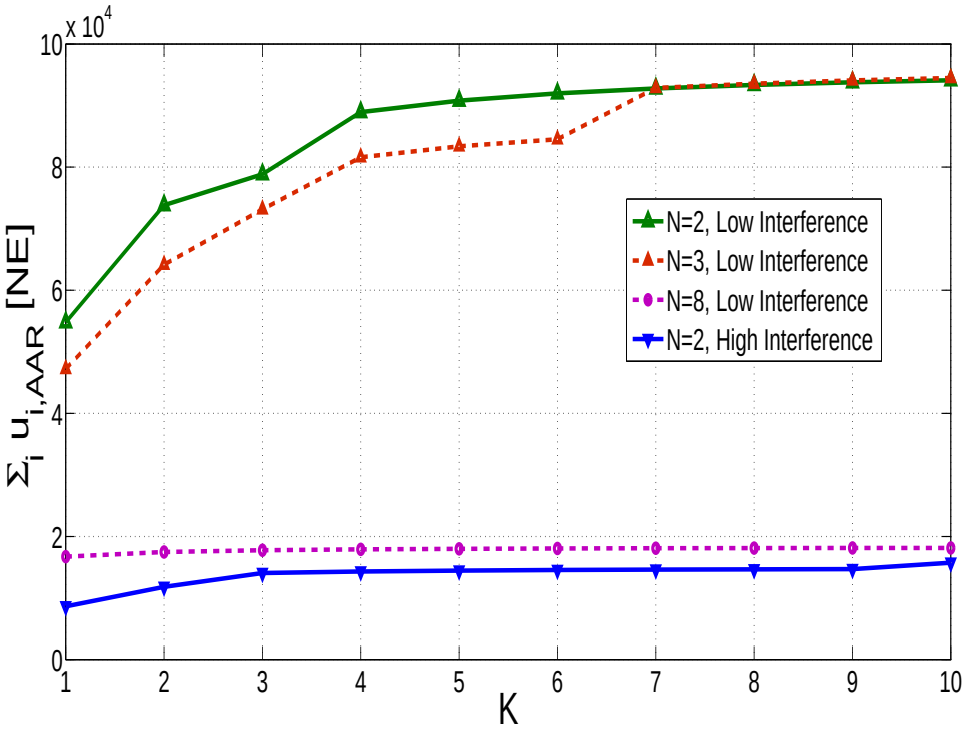


Figure 3.5: AAR: Observe that the AAR sum-payoff at the NE is sensitive to the interference level, as seen from the large difference between the two user low and high interference case. With a low interference level, $N = 8$ has a higher sum-payoff than for $N = 2$ with a high interference level.

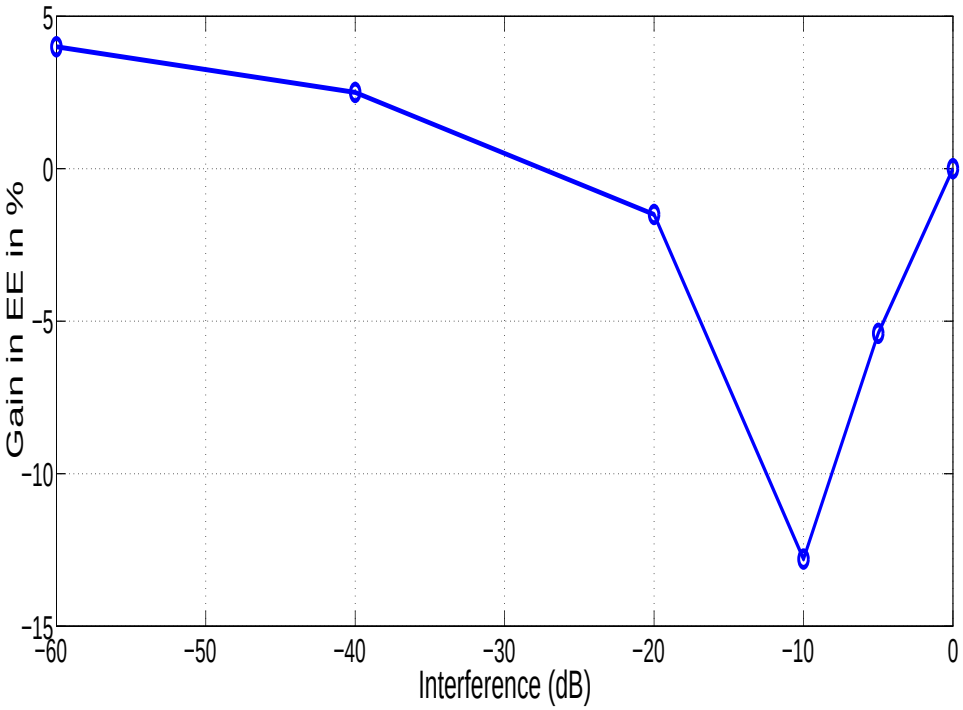


Figure 3.6: AAR: Plot of percentage gain in EE v.s $g_{i,j}, i \neq j$ (keeping $g_{i,i} = 1$), where the gain is calculated by comparing the EE achieved using the proposed AAR algorithm to the EE at the NE achieved by using the algorithm ignoring the packet level. Observe that in the very low interference regime, the proposed scheme outperforms the other algorithm. However in the low-medium interference region, the NE is inefficient with a high PoA and this results in poor performance.

Fig. 3.6 studies the average gain in EE (averaged over the channel fading) when compared to that of a power control algorithm ignoring the packet level versus the interference. Here see that when the interference is very low, the NE of the proposed scheme performs better than an algorithm that ignores the packet level. However, when under interference, the strategy under the AAR scheme would be to use a very high power as EE is individually optimized when the AAR achieves a higher packet rate. This results in a sub-optimal NE as seen in the figure when the interference is in the $[-25, 0]$ dB range. This effects indicates that the cross-layer approach might induce some performance loss w.r.t. the classical approach. This negative but quite surprising result indicates that in distributed networks, refining the modeling aspect can sometimes induce a performance loss; similar to other known paradoxes in distributed networks such as the Braess paradox [63].

# Chapter 4

# Other techniques to improve the energy-efficiency

In this chapter, techniques that improve the energy efficiency of both centralized and decentralized communication systems are discussed. Specifically, for the case of centralized systems, the energy efficiency metric is redefined by considering the dynamic nature of users that arrive and leave the network based on individual data needs. Clearly, this approach can not be carried over to a de-centralized problem (uplink for example) as a single user does not have information on the other users in a completely distributed system. Therefore, in the case of decentralized systems, a technique of exchanging CSI information through power level coding is proposed that can significantly improve system efficiency.

## 4.1 Centralized systems

The contributions of this section are summarized as follows:

1. Consider a new energy efficiency metric that accounts for the overall power consumption of the base station, including common channel and fixed consumption parts.
2. Derive an optimal power allocation scheme that maximizes the energy efficiency, while preserving Quality of Service (QoS).
3. Show that the power allocation that considers the dynamic behavior of users is significantly different from the scheme optimized locally for each state of the network. In addition to that, the former performs better than the latter.

### 4.1.1 System model

#### System description

Consider a transmitting base station with buffers of infinite (or very large) size. The base station sends packets into a queue for each user which is stored in these buffers. The packets arrive at each time slot $T_P$ (expressed in seconds), each packet being of size $S_p$ (expressed in bits). The data rate $R_p$ is equal to $\frac{S_p}{T_P}$. The throughput when using all the available bandwidth is denoted by $R(\rho)$ (expressed in bits per second), when the receiver has an average signal to interference plus noise ratio (SINR) of $\rho$. This SINR depends directly on the transmit power $P$ (expressed in Watts) as $\rho = \frac{P}{\sigma^2}$. Here $\sigma^2$ represents the average noise for a given radio condition (expressed in Watts) and it depends on the distance of the receiver from the base station.

All packets of a user are assumed of the same size and the average throughput on the radio interface, when the queue for the corresponding user is active, is denoted by $R_a(\rho)$ (expressed in bits per second) which depends on the bandwidth available. When all the packets in the queue are transmitted the queue becomes empty and inactive. Each packet stored in the buffer is a collection of frames that are transmitted over the symbol time $T_s$ (expressed in seconds). Each frame is transmitted or retransmitted till it goes through and an acknowledgment is received.

$$T_d = \frac{S_p}{R_a(\rho)} \tag{4.1}$$

If this duration exceeds $T_P$, the time by which the next packet arrives, the queue size becomes infinite and the transmitter is always on. Otherwise, the probability of the transmitter to be active ($\Phi(\rho)$) is given by the ratio of $T_d$ to $T_P$. Thus:

$$\Phi(\rho) = \max\left(\frac{R_p}{R_a(\rho)}, 1\right) \tag{4.2}$$

The values taken for $R(\rho)$ from [64], are in fact, averaged over the fast fading and are thus suitable for this model. When there are several users in the network, the available bandwidth is divided among the active users. Assume the bandwidth allocation to be equal among all users and this implies that if $N$ users are all active and experience the same radio conditions, the throughput is reduced to $\frac{R(\rho)}{N}$.

#### Proposed performance metric

In the broadcast channel there are multiple users that have to be served. In practice, users arrive randomly, and depart once they finish downloading their requested data. New arrivals are blocked when the total number of users crosses a certain limit defined by the base station. Each user may experience a different radio condition from its peers.

For convenience, the area covered by the base station is divided into "zones". Every user in the same zone, experiences the same radio conditions. This implies that if the base station transmits at a certain power, then all the users in the same zone experience the same SINR. The radio conditions are determined by the average distance of the zone to the base station. If there are $M$ zones in total, define $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\}$ as the channel conditions for each zone. Define the "state" of the system $\mathbf{s} = \{N_1, N_2, \cdots, N_M\}$. The state $\mathbf{s}$ represents the number of users in each zone. For example if there are two zones, and there are no users the state is $\{0, 0\}$. When a user arrives to zone one, the state becomes $\{1, 0\}$.

For a state $\mathbf{s} = \{N_1, N_2, \cdots, N_M\}$, the power allocation scheme defined as $\mathbf{P} = \{P_1, P_2, \cdots, P_M\}$ results in an SINR distribution of $\boldsymbol{\rho} = \{\rho_1, \rho_2, \cdots \rho_M\}$ among the zones 1 to $M$, where $\rho_j = \frac{P_j}{\sigma_j^2}$.

First, define the notion of energy-efficiency for a given state or the "local" energy-efficiency. This is useful as in practice, the base station can easily measure this quantity only for a given state as it is unable to predict when a new user will arrive. The "global" energy-efficiency defined as the average of the energy-efficiency in each state weighted by their probabilities.

If there is always one and only one user, the energy-efficiency can be defined based on [8] and other works as

$$\eta_{SU} = \frac{R(\rho)\Phi(\rho)}{b + P\Phi(\rho)} \tag{4.3}$$

where $b$ is the constant power consumed by the base station while serving at least one user [1]. The proposed form is easy to interpret as $R(\rho)$ represents the average throughput when the transmitter is active and $P$ is the cost when the transmitter is active.

When the system is state $\mathbf{s}$, the energy-efficiency is defined as:

$$\eta_{\mathbf{s}}(\mathbf{P}) = \frac{\bar{R}_{\mathbf{s}}(\boldsymbol{\rho})}{\bar{P}_{\mathbf{s}}(\mathbf{P})} \tag{4.4}$$

where $\bar{R}_{\mathbf{s}}$ and $\bar{P}_{\mathbf{s}}$ represent the total throughput and power consumed respectively in state $\mathbf{s}$.

When the number of users is random, then the global energy-efficiency function is defined as:

$$\hat{\eta} = \sum_{\mathbf{s}} \frac{\pi(\mathbf{s})\bar{R}_{\mathbf{s}}}{\bar{P}_{\mathbf{s}}} \tag{4.5}$$

Where $\pi(\mathbf{s})$ is the probability of finding the base station at state $\mathbf{s}$ of user distribution. The global energy-efficiency could alternately be defined as

1. This cost can have several origins like energy spent on the power amplifier, computation, cooling mechanisms etc. Details of the power consumption model are given in [45].

ratio of the total throughput over all states to the total power over all states. However, in practice, calculating the energy-efficiency for each state and taking the average, is easier and more reasonable. The goal of this work is to improve the above defined energy-efficiency of a transmitting base station.

This metric can be physically interpreted as the average number of bits that can be transmitted by spending one Joule of energy. Alternately, the average power cost of the base station can be written an $\frac{\text{Traffic}}{\eta}$. Hence, optimizing the global energy-efficiency amounts to minimizing the average power consumption of the base station.

### 4.1.2 Optimal power allocation for a fixed number of users

In this section, consider the case where the number of users is fixed. Refer to the optimization of the metric defined in this section as "local" optimization as it deals with the optimization of a single state of the wireless network. When the state of the network is given, the number of users in each zone is known and can thus be used to calculate the relevant information required to obtain and optimize the energy-efficiency. Assume a knowledge of the average noise levels for each zone, i.e $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\}$ are known.

#### Homogeneous radio conditions

First, consider the problem where all users experience the same average SINR, as the model is easier to be understood; the case of heterogeneous SINRs will be exposed next. Let the total number of users in the cell be $N$. As all the users experience the same radio conditions, $\mathbf{s} = \{N\}$. Define the average throughput experienced by any queue as $R_a$, to derive:

$$R_a(\rho) = \sum_{i=0}^{N-1} \binom{N-1}{i} \Phi(\rho)^i (1 - \Phi(\rho))^{N-1-i} \frac{R(\rho)}{i+1} \tag{4.6}$$

where $\Phi(\rho)$ denotes the probability that any of the $N$ users are actively being served and is given as in equation 4.2. The summation is upto $N-1$ as $R_a$ is the throughput experienced by an active user, and so consider the remaining $N-1$ users. The $R_a$ for every user is identical as all users experience the same SINR for the same transmit power. This symmetry can be exploited to conclude that the transmit power for each user will be equal when optimized. Note that $R_a(\rho)$ depends on $\Phi(\rho)$ and $\Phi(\rho)$ depends on $R_a(\rho)$ leading to a fixed point equation.

Clearly if $N$ is large enough, then the demand in data rate will exceed the maximum available throughput and $\Phi(\rho)$ becomes 1. On the other hand, if $N$ is small enough, the users may transmit their data faster than the packet arrival speed causing the queue to empty occasionally. In this period, other users can take advantage of the excess bandwidth.

From $\Phi(\rho)$, the total power consumed can be calculated as

$$\bar{P}_{\mathbf{s}} = b + P(1 - (1 - \Phi(\rho))^N) \tag{4.7}$$

Here $(1 - \Phi(\rho))^N$ is the probability of all queues being empty. If any queue is active the power consumed is $P$. The total throughput is $\bar{R}_{\mathbf{s}} = N\Phi(\rho)R_a$ leading to an energy-efficiency of

$$\eta_{\mathbf{s}} = \frac{N\Phi(\rho)R_a(\rho)}{b + P(1 - \Phi(\rho))^N} \tag{4.8}$$

**Heterogeneous radio conditions**

Consider a more realistic setting where users experience different radio conditions in each zone. Denoting the average throughput experienced by zone $j$ as $R_{a:j}$, compute

$$\begin{aligned} R_{a:j}(\boldsymbol{\rho}) = R(\rho_j) &\sum_{i_1=0}^{N_1} \sum_{i_2=0}^{N_2} \cdots \sum_{i_j=0}^{N_j-1} \cdots \sum_{i_M=0}^{N_M} \binom{N_1}{i_i} \\ &\times \binom{N_2}{i_2} \cdots \times \binom{N_j - 1}{i_j} \times \cdots \times \binom{N_M}{i_M} \times (\Phi_1(\boldsymbol{\rho}))^{i_1} \\ &\times (\Phi_2(\boldsymbol{\rho}))^{i_2} \times \cdots \times (\Phi_M(\boldsymbol{\rho}))^{i_M} \times (1 - \Phi_1(\boldsymbol{\rho}))^{N_1 - i_1} \\ &\times (1 - \Phi_2(\boldsymbol{\rho}))^{N_2 - i_2} \times \cdots \times (1 - \Phi_j(\boldsymbol{\rho}))^{N_j - i_j - 1} \\ &\times \cdots \times (1 - \Phi_M(\boldsymbol{\rho}))^{N_M - i_M} \times \frac{1}{i_1 + i_2 + \cdots + i_M + 1} \end{aligned} \tag{4.9}$$

where

$$\Phi(\boldsymbol{\rho})_j = \max\left(\frac{R_p}{R_{a:j}(\boldsymbol{\rho})}, 1\right) \tag{4.10}$$

Leading to a set of fixed point equations that can be solved to calculate all $R_{a:j}(\boldsymbol{\rho})$ for a given $\mathbf{P}$. Equation (4.9) is similar to (4.6), but considers the presence of users in other zones as well. The average power can be calculated as

$$\begin{aligned} \bar{P}_{\mathbf{s}}(\mathbf{P}) = b + &\sum_{i_1=0}^{N_1} \cdots \sum_{i_M=0}^{N_M} (1 - \delta(i_1 + \cdots + i_M)) \\ &\times (\Phi_1(\boldsymbol{\rho}))^{i_1} \times \cdots \times (\Phi_M(\boldsymbol{\rho}))^{i_M} \times \frac{P_1 i_1 + \cdots + P_M i_M}{i_1 + \cdots + i_M} \\ &\times (1 - \Phi_1(\boldsymbol{\rho}))^{N_1 - i_1} \times \cdots \times (1 - \Phi_M(\boldsymbol{\rho}))^{N_M - i_M} \end{aligned} \tag{4.11}$$

Where the $\delta$ function is used to exclude the state where all zones are empty ($\delta(x) = 0$ for all real $x$ but 0, and $\delta(0) = 1$). The energy-efficiency in this state can be calculated with $\bar{R}_{\mathbf{s}}(\boldsymbol{\rho}) = \sum_{i=1}^{M} N_i \Phi(\boldsymbol{\rho})_i R_{a:i}$ and total power from equation (4.11).

### 4.1.3 Optimal power allocation considering the dynamic behavior of users

In the previous section, the energy-efficiency for fixed numbers of users was optimized. To analyze the impact of power allocation on the network performance and account for the users arrivals and departures, a flow-level capacity analysis is required. The arrival rate can be modeled through a Poisson process (of intensity $\lambda_i$ in zone $i$) and users leave when they finish streaming a file of average size $F$. When the total number of users exceed a given threshold $N_{max}$, new user arrivals are blocked.

#### Processor sharing analysis

When users with a finite workload are considered, the number of users is not constant but varies dynamically during time. The distribution of the number of users is determined by the traffic intensity within the cell. Indeed, if the traffic intensity is large, more users connect to the system per unit time and the average number of active users increases.

The heterogeneity in radio conditions translates into a larger service time for cell edge users. When the system is in state $\mathbf{s} = \{N_1, N_2, \cdots, N_M\}$, the total number of users in the cell is $N(\mathbf{s}) = N_1 + \cdots + N_M$. Based on [65], this can be modeled as a Generalized Processor Sharing queue, whose evolution is just described by the overall number of users in a cell. The solution of the Markov process has the simple form

$$\pi(\mathbf{s}) = \frac{1}{\Gamma} \frac{N(\mathbf{s})!}{\prod_{i=1}^{M} N_i!} \prod_{c=1}^{M} \frac{\Omega_c^{N_c}}{\prod_{j=1}^{N_c} j \Phi_{c;\mathbf{s}(N_c=j)} R_{a:c;\mathbf{s}(N_c=j)}} \tag{4.12}$$

where $\Omega_c = S\lambda_c$ and $\Gamma$ is a normalizing constant. The notation $\mathbf{s}(N_c = j)$ is used to take the $\Phi$ and $R_a$ for the state $\mathbf{s}$ with $j$ users in zone $c$.

In this model, the user blocking rate can be calculated as $\alpha = \sum_i \lambda_i \sum_x \pi(x)$, $x$ such that the system is full ($N(x) = N_{max}$). Quality of service (QoS) is measured through the user blocking rate. The QoS constraint is thus $\alpha \leq \epsilon$, where $\epsilon$ is the maximum tolerable blocking rate.

#### Optimal power allocation

The steady-state probabilities defined in the previous section are calculated knowing the throughputs for each state of the network. This throughput will of course depend on the power allocation as explained in Sections II and III. The power allocation has thus to be optimized taking into account the dynamics of users. A power allocation policy $\hat{\mathbf{P}}$ is defined as a set of actions for each of the possible states:

$$\hat{\mathbf{P}} = \bigcup_{\mathbf{s}} \mathbf{P}_{\mathbf{s}} \tag{4.13}$$

The global energy efficiency; knowing the policy $\hat{\mathbf{P}}$, is given by:

$$\hat{\eta}(\hat{\mathbf{P}}) = \sum_{\mathbf{s}} \frac{\pi(\mathbf{s})\bar{R}_{\mathbf{s}}(\boldsymbol{\rho})}{\bar{P}_{\mathbf{s}}(\mathbf{P}_{\mathbf{s}})} \tag{4.14}$$

The optimization problem can be defined as

$$\hat{\mathbf{P}}^* = \arg\max[\hat{\eta}(\hat{\mathbf{P}})] \tag{4.15}$$

And the maximum global energy-efficiency possible is $\hat{\eta}(\hat{\mathbf{P}}^*)$.

The idea behind this global optimization is that the power allocation does not depend uniquely on the actual state of the network, but takes also into account the future evolutions of the network. For instance, a power allocation decision that is taken at one moment may have an influence on the evolution of the state of the network by favoring a subset of users by a better throughput. In the next section, the difference between this global policy maximization and a local one, as defined in section III are studied.

### 4.1.4 Numerical results

In this section, simulations and numerical calculations are used to study the properties of the energy-efficiency function and obtain the power allocation that maximizes it. Consider the receiver and the transmitter to have two antennas each forming a $2 \times 2$ MIMO system. The data rates for this configuration which are LTE compliant are taken from [64] and are given as a function of the SINR. For the single zone case, take $\sigma^2 = 1$ mW while for the two zone case, take $\{\sigma_1^2, \sigma_2^2\} = \{1, \frac{1}{8}\}$ mW.

#### Numerical results for the local optimization

In figure 4.1, the energy-efficiency as a function of the transmit power is shown. Here, due to symmetry, all the users use the same power. The results show that the energy efficiency begins by increasing with the transmit power increases, as users are able to reach higher throughputs. However, starting from one point, users reach the maximal throughput they are able to reach as, in LTE, modulation schemes are limited; the energy efficiency begins thus decreasing as throughputs remain constant while power consumption increases.

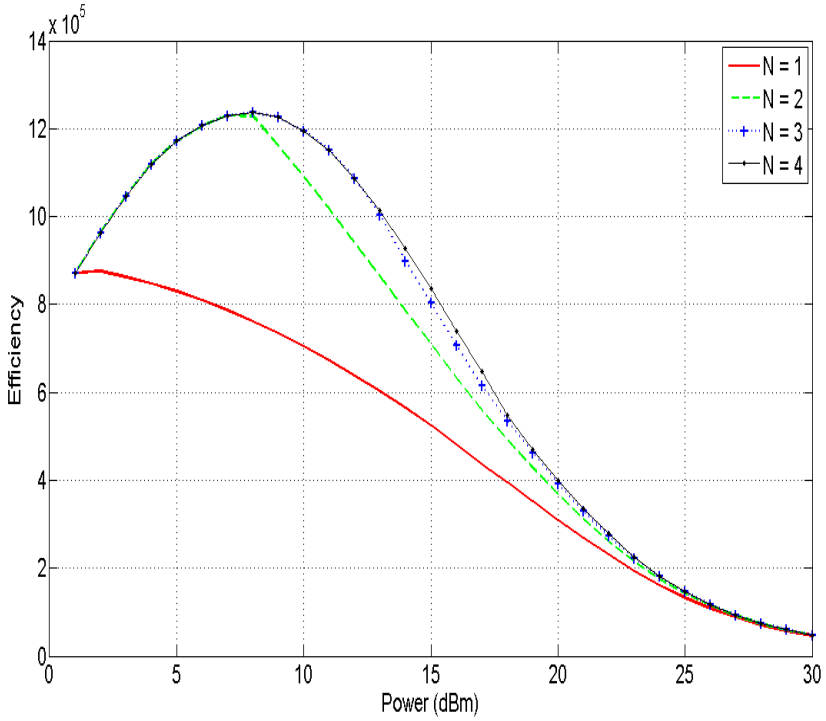


Figure 4.1: $\eta$ vs $P$ with $\frac{b}{\sigma^2} = 100$ (20dB). Note that the energy-efficiency is peaked at higher powers with additional users.

### Numerical results for the global optimization

In this setting, the power allocation is not determined for a fixed number of users, but for a given traffic intensity. the number of users is thus a random variable whose distribution depends on the traffic intensity. The optimal power allocation is the one that maximizes the energy efficiency while maintaining a constraint on the QoS. Note that this optimal power allocation is a matrix that gives, for each state of the network composed of the number of users in the cell, the power allocation for each of the users.

Initially consider the cell with homogeneous radio conditions, i.e. suppose that all the users experience the same SINR on average. In this setting, if $N_{max}$ is the maximum number of users allowed, optimization is performed over $N_{max}$ variables, i.e. the power used in each state. For the single zone case take $\sigma^2 = 1$ mW. The optimal power allocation is shown in figure 4.2. Note that, in this case, the power allocation is a vector and not a matrix, as all users experience the same radio conditions and have, by symmetry, the same allocated power.

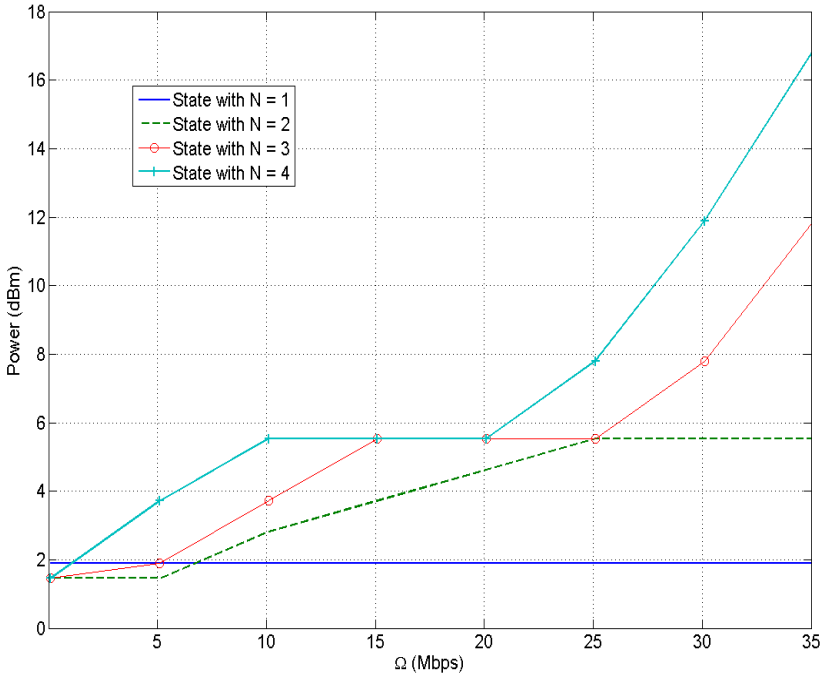


Figure 4.2: The power allocation scheme $(P_1, \cdots, P_4)$ plotted against the traffic $\Omega$ when $\hat{\eta}$ is optimized. Also note that the QoS constraint of maintaining the blocking rate below 0.01 is satisfied.

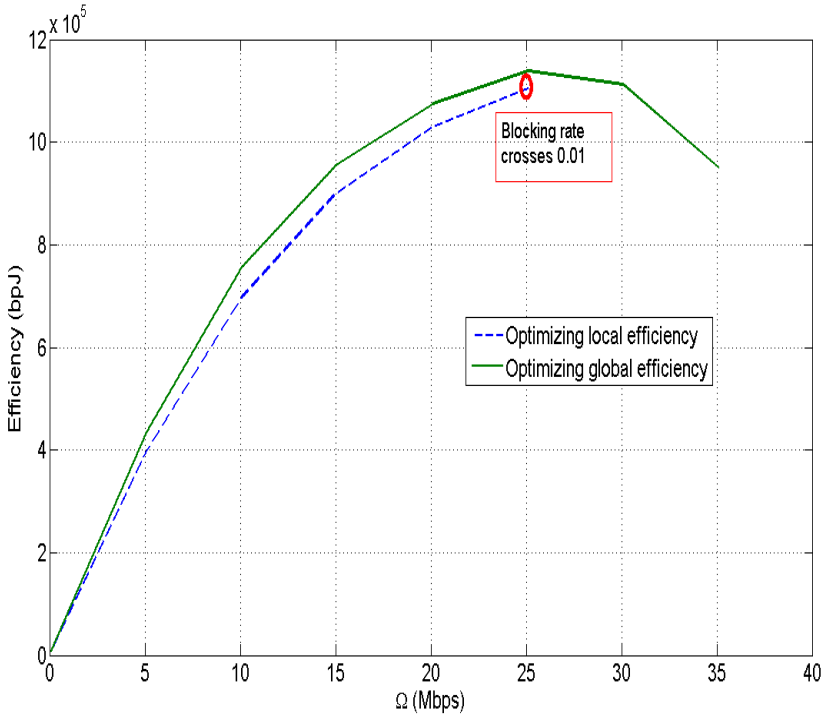


Figure 4.3: $\hat{\eta}$ plotted against the traffic $\Omega$ when $\hat{\eta}$ is optimized and when $\eta$ is optimized for each state separately.

Figure 4.3 compares the energy-efficiency obtained for the local and the global optimizations. As seen from the simulations (Figure 4.3), using a global optimization does not seem to yield much gains in the energy-efficiency for the single zone case. This is because the throughput, and thus service times, are the same for all users. Next, moving on to the two-zone case (cell center and cell edge): Here, consider a cell divided into two concentric rings, and define the outer zone as the region when the SINR is 4.8 dB (3 times) lower than the SINR for the inner zone, when the transmit power is unchanged. The outer zone also has 3 times the area of the inner zone causing $\lambda_2 = 3\lambda_1$. With these parameters calculate the optimal global energy-efficiency and corresponding power allocation for given values of $\lambda_1$. $\{\sigma_1^2, \sigma_2^2\} = \{1, \frac{1}{3}\}$ mW. Figure 4.4 shows the energy efficiencies corresponding to local and global optimizations. It is obvious that global optimization yields much higher efficiency when users have heterogeneous radio conditions. This is because, in the local optimization setting, the notion of call duration cannot be taken into account as users are considered as always active. The optimal power allocation will then tend to favor cell center users in order to maximize throughput. However, when the dynamic behavior of

users is taken into consideration, it is sometimes better to use more power on cell edge users in order to let them finish their service quickly and quit the system. Applying the policy obtained from the local optimization will lead to users accumulating at the cell edge as they are not able to finish their transfers.

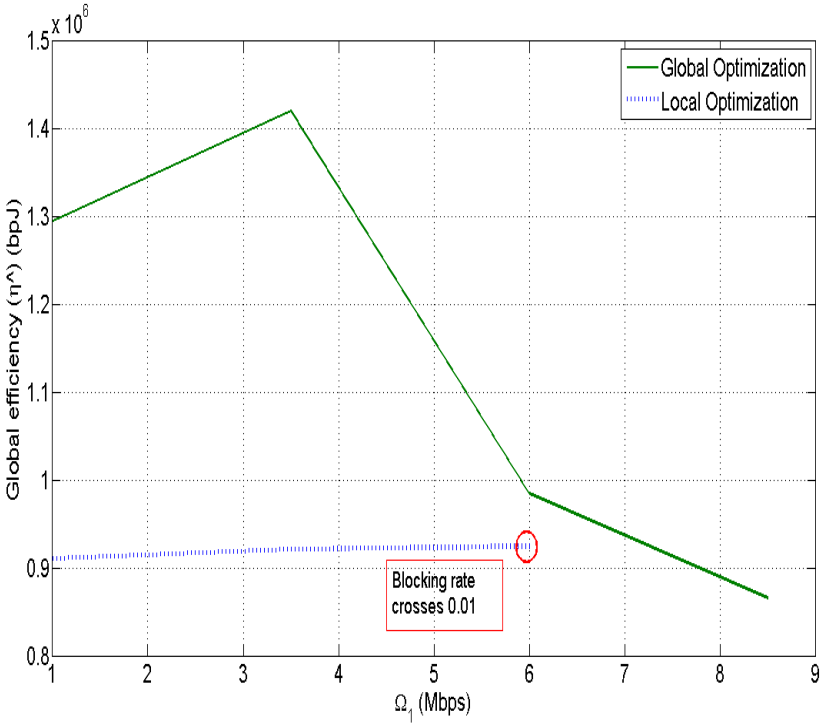


Figure 4.4: $\hat{\eta}$ plotted against the traffic $\Omega_1 = \lambda_1 S$ when $\hat{\eta}$ is optimized and when $\eta$ is optimized for each state separately. Also note that the QoS constraint of maintaining the blocking rate below 0.01 is satisfied at all points shown.

## 4.2 De-centralized systems

Interference networks have been a primary focus of interest to researchers since the conception of wireless networks. Typically, in an interference network, when there is no direct line of communication between the transmitters; the transmitters use a distributed or selfish strategy and work at a sub-optimal level of performance. For example, in the case of a distributed interference network with multiple-carriers, the iterative water-filling algorithm is considered to be the state of the art technique [66]. In [67] the authors introduce the idea of coded power control where one decision maker

that has knowledge of the future and global CSI (channel state information) provides information of the optimal action to the second decision maker through it's actions. This work focuses on a more practical situation where only imperfect local CSI is available at both the transmitters.

### 4.2.1 Motivation

Clearly, in a non-coperative setting (wi-fi for example), there seems to be no motivation for the individual decision makers in a wireless network to cooperate. This leads to an inefficient Nash equilibrium as seen in Chapter 3. However, this is mainly because the game considered in Chapter 3 was the single shot power control game. In most practical cases, the game is not a single shot game as the devices keep transmitting for a long time. From the results shown in C.2, it can be seen that using the repeated game framework, a Nash equilibrium of the repeated game can be found which pareto-dominates the Nash of a single shot game.

In repeated games (RG), as the name suggests, the same game is played several times. The long-term interactions between the players in such a situation is studied under the RG framework. The players react to past experience by taking into account what happened in all previous stages and make decisions about their future choices. The resulting utility of each player is an average of the utility of each stage. A game stage $t$ corresponds to the instant in which all players choose their actions simultaneously and independently and thus a profile of actions can be defined by $\mathbf{p}(t) = (p_1(t), p_2(t), \ldots, p_N(t))$. When assuming full monitoring, this profile choice is observed by all the players and the game proceeds to the next stage. In a repeated game with complete information and full monitoring, the Folk theorem characterizes the set of possible equilibrium utilities. It states that the set of Nash equilibrium in a RG is precisely the set of feasible and individually rational outcomes of the one-shot game (non-cooperative game). Thus, it can be justified that playing any strategy that pareto-dominates the Nash, is an equilibrium of the repeated game and hence feasible.

Of course, the next step would be in developing a method of achieving such a point. From C.2, it is clear that in an interference network, achieving such a point is not so straight-forward. The main goal of this section is to enable distributed systems to communicate implicitly through their transmit power levels and achieve an efficient equilibrium.

*Contributions*: The novel contributions of this section are two-fold. First, a technique through which all the channel fading coefficients ($g_{i,j}$) can be calculated by exploiting just the SINR feedback is proposed. Second, a method to exchange the calculated channel information through power level modulation is explained. This efficiency of the proposed scheme is analyzed quite extensively through several numerical simulations.

### 4.2.2 System Model and technique overview

The system under consideration is that of $K$ pairs of interfering transmitters and receivers. Let the transmit power of user $i$ be given by $P_i$ and the channel between transmitter $j$ and receiver $i$ be $g_{j,i}$. In a many user interference channel, the SINR at receiver $i$, $\gamma_i$ is given by:

$$\gamma_i = \frac{g_{ii}P_i}{1 + \sum_{j \neq i} g_{ji}P_j} \tag{4.16}$$

where the noise level is normalized to 1. Receiver $i$ estimates this SINR at each time slot and sends to transmitter $i$ via a feedback channel $\widehat{\gamma}_i = \gamma_i + \Delta_i$, the channel estimate. The error $\Delta_i$ could be due to imperfect estimation, quantization of the SINR and through error in the feedback channel. The $K$ dimensional vector formed by these transmit power levels is given by $\underline{P}$. The transmit power levels are assumed to be discrete and are either 0 (no transmission), or belong to the finite set $\mathcal{P}$, where $\|\mathcal{P}\| = M$ is the number of available non-zero power levels. All the channel fading coefficients $g_{i,j}$ are positive real numbers and can be expressed as elements of the $K \times K$ matrix $\mathbf{G}$. The utility of the system is given by $W(\underline{P}, \mathbf{G})$, and the power levels chosen at the globally efficient point are:

$$\underline{P}^* = \arg\max_{\underline{P}} W(\underline{P}, \mathbf{G}) \tag{4.17}$$

This work provides a procedure by which the transmitters can estimate partial information on $\mathbf{G}$ and exchange this information to obtain the complete $\mathbf{G}$. Once $\mathbf{G}$ is obtained $\underline{P}^*$ can be found and the system can operate at the globally efficient point given by (4.17). Naturally, this point can be perfectly achieved when $\Delta_i = 0$ and with a non-zero noise in the SINR estimate, a scheme of achieving an approximately globally efficient point is provided. The process of achieving the power control scheme as described by (4.17) is divided into three phases. The first phase is involved in estimating all the channel coefficients that are perceived by each receiver. Receiver 1 for example would estimate $g_{1,1}, g_{2,1}, ..., g_{K,1}$. The second phase involves encoding this information into their transmit power levels as well as decoding the information received by observing the power levels of the other transmitters. The final phase would involve using all the collected information available to all the transmitters and setting the power control scheme to the one described by (4.17).

### 4.2.3 Phase 1: Channel estimation

The process of channel estimation is done by exploiting equation (4.16). The first phase lasts for a duration of $T_1$ frames. In each frame, each of

the transmitters transmit at a power level given by $P_i(t), t \in \{1, 2, .., T_1\}$. This power control scheme for phase one is denoted by the matrix $\mathbf{P}^1$ and is given by $(\underline{P}(1)^T, \underline{P}(2)^T, ...\underline{P}(T_1)^T)^T$. In order to estimate $g_{j,i}$ based on the $\mathbf{P}^1$ choose, the sequence of SINRs $\underline{\gamma} = (\gamma(1), \gamma(2), ...\gamma(T_1))^T$ is exploited. Define the matrix $\mathbf{S}_i(\underline{\gamma}, \mathbf{P}^1)$ as

$$\begin{pmatrix} -\gamma_i(1)P_1(1) & -\gamma_i(1)P_2(1) & ... & P_i(1) & ... & -\gamma_i(1)P_K(1) \\ -\gamma_i(2)P_1(2) & -\gamma_i(2)P_2(2) & ... & P_i(2) & ... & -\gamma_i(2)P_K(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\gamma_i(T_1)P_1(T_1) & -\gamma_i(T_1)P_2(T_1) & ... & P_i(T_1) & ... & -\gamma_i(T_1)P_K(T_1) \end{pmatrix} \tag{4.18}$$

This allows the channel coefficients to be rewritten using the Moore-Penrose pseudo-inverse as:

$$\underline{g}_i = \begin{pmatrix} g_{1i} \\ \vdots \\ g_{Ki} \end{pmatrix} = \left(\mathbf{S}_i(\underline{\gamma}, \mathbf{P}^1)^H \mathbf{S}_i(\underline{\gamma}, \mathbf{P}^1)\right)^{-1} \mathbf{S}_i(\underline{\gamma}, \mathbf{P}^1)^H \times \underline{\gamma} \tag{4.19}$$

However, as only $\widehat{\underline{\gamma}}$ is available to the transmitter and not the actual SINR $\underline{\gamma}$, the channel vector can be estimated as:

$$\widehat{\underline{g}_i} = \left(\mathbf{S}_i(\widehat{\underline{\gamma}}, \mathbf{P}^1)^H \mathbf{S}_i(\widehat{\underline{\gamma}}, \mathbf{P}^1) + \lambda \mathbf{I}\right)^{-1} \mathbf{S}_i(\widehat{\underline{\gamma}}, \mathbf{P}^1)^H \times \widehat{\underline{\gamma}} \tag{4.20}$$

using the regularized inverse, where $\lambda$ is the regularization factor. Of course, the expected estimation error variance $E[\|\underline{g}_i - \widehat{\underline{g}_i}\|^2]$ is minimized by the estimator (4.20) only when $S$ is independent of the feedback noise, and the channel statistics and the SINR feedback error have a specific distribution (i.i.d and Gaussian). Generally in practice, the channel statistics is exponential and the structure of the feedback noise is unknown. and so the optimal estimator for a general noise is unknown. For the purpose of simulations and a meaningful analysis, the estimator given by (4.20) is used in this section. At the end of Phase 1, each transmitter $i$ estimates the channel vector $\underline{g}_i$ as $\widehat{\underline{g}_i}$.

### 4.2.4 Phase 2: Coding and decoding the CSI to/from the power levels

In the second phase, each transmitter encodes the information $\widehat{\underline{g}_i}$ onto their power levels $P_i(t)$ and decode the power levels used by the other transmitters. This phase lasts for a duration of $T_2$ frames. As there are M power levels available and $T_2$ frames, the total amount of information that can be transferred in this phase by any transmitter is equal to $\log_2(M^{T_2}) = T_2 \log_2(M)$ bits. The overall processing of the channel information estimated and sent by transmitter $i$ is given by:

$$\underline{g}_i \xrightarrow{Phase\,1} \widehat{\underline{g}_i} \xrightarrow{Quantizer} < \underline{g}_i > \xrightarrow{Coder} P_i(t) \xrightarrow{j-th\ Decoder} {}^j\tilde{\underline{g}}_i \tag{4.21}$$

where $< \underline{g}_i >$ is the quantized channel vector and ${}^j\underline{\tilde{g}}_i$ is the channel vector $\underline{g}_i$ decoded by transmitter $j$.

*Quantization*: Thus, the first step in the second phase is for each of the transmitters to quantize the real K dimensional vector $\underline{\widehat{g}}_i$ into $M^{T_2}$ symbols. Then, each of these symbols are associated to a power level vector of size $T_2$ with $M$ possible power levels at each time. For this purpose, the classical iterative "Lloyd" algorithm could be used in order to minimize the distortion of the coded channel. However this algorithm assumes that there is no error on the information transferred but due to quantization, whereas in this scheme, the estimate $\underline{\widehat{g}}_i$ has an error before quantization as well as an error in decoding the power level (after quantization due to symbol detection errors).

In the presence of an error just before the quantizer, [68] proposed an alternative algorithm. In this work however, there could be an error both before and after the quantization. The results in [69] exploit the statistics of the error after quantization in order to minimize the distortion. Let the the PDF of the noise due to channel estimation $\underline{g}_i - \underline{\widehat{g}}_i$ be given by $\phi()$ and the probabilty of detecting the symbol $l$ as symbol $n$ (where $n, l \in \{1, 2, .., M^{T_2}\}$ be given by $\pi_{l,k}$. First select randomly $N = M^{T_2}$ sites $s_l, l \in \{1, 2, ..N\}$ from a K-dimensional space. Then, the following steps are performed iteratively until the $N$ sites converge:

1. Calculate the Vornoi diagram of each of these $N$ sites $s_l$. For each site the corresponding Vornoi region $u_l, l \in \{1, 2, ..N\}$ is defined by the set of all points closer to that site than to any other, i.e., $u_l = \{x : \|x - s_l\| < \|x - s_k\| \forall k \neq l\}$
2. Calculate the weighted centroids $v_l, l \in \{1, 2, ..N\}$ as follows [69]:

$$v_l = \frac{\int_{u_l} \underline{g}_i f(\underline{g}_i) \sum_{n=1}^{N} \pi_{n,l} \int_{u_n} \phi(y - \underline{g}_i) dy d\underline{g}_i}{\int_{u_l} f(\underline{g}_i) \sum_{n=1}^{N} \pi_{n,l} \int_{u_n} \phi(y - \underline{g}_i) dy d\underline{g}_i} \tag{4.22}$$

   where $f(\underline{g}_i)$ is the PDF of the variable to quantize $\underline{g}_i$.
3. Set the $N$ sites to be the $N$ weighted centroids $v_l$ as calculated from Step 2; $s_l = v_l$.

*Power modulation and demodulation*: This power control scheme for phase two is denoted by the matrix $\mathbf{P}^2$ given by $(\underline{P}(T_1+1)^T, \underline{P}(T_1+2)^T, ...\underline{P}(T_1+T_1)^T)^T$. Once the quantized channel is coded onto the power levels, the next step is to identify the power levels used by the other transmitters. In this step each transmitter $i$ broadcasts the quantized channel $< \underline{g}_i >$ to every other receiver. At the transmitter $j$ at time $t$, the total interference that is observed is estimated as

$$\sum_{k \neq j} P_k g_{k,j} = \frac{P_j g_{j,j}}{\widehat{\gamma}_j(t)} - 1 \tag{4.23}$$

As $g_{k,j} \forall k$ are known at $j$ and $P_k \in \mathcal{P}$, a maximum likelihood estimator (MLE) can be used to get all $P_K(t)$. Once all the $P_k(t)$ are found for all $t$ in the second phase. The power levels and the quantization codebook can be used to retrieve the channel $\underline{g}_k, \forall k$ and thus obtain $\mathbf{G}$. However as the MLE could have a different estimate at each receiver $j$ and so at each transmitter $j$, a global channel estimate ${}^{j}\tilde{\mathbf{G}}$ is obtained.

### 4.2.5 Phase 3: Working at the globally efficient system point

By the end of phase two, the transmitter $k$ possesses the distorted quantized channel information $\{{}^{k}\dot{G}_i\}$. Therefore, at this point transmitter $k$ uses the power control policy given as $P_k^{PS}$ (power control of user k in the proposed scheme), found as the $k-th$ element of ${}^{k}\underline{P}^*$, such that:

$$ {}^{k}\underline{P}^* = \arg\max_{\underline{P}} W(\underline{P}, {}^{k}\tilde{\mathbf{G}}) \tag{4.24} $$

Clearly, in this situation, each transmitter $k$ assumes that the other transmitters also work at the same perceived power control solution ${}^{k}\underline{P}^*$. However, as ${}^{k}\tilde{g}_{ji}$ is not equal for all $k$, this results in a power control policy that is a distortion of the solution in (4.17). The policy given by (4.24) becomes exactly (4.17) when the SINR estimate at the transmitter is always perfect. To analyze the efficiency of the working point $P_k^{PS}$, making use of extensive monte-carlo simulations that are described in the following section.

### 4.2.6 Numerical Analysis

For the numerical analysis, a specific choice of the utility $W$ is fixed, i.e.,

$$ W(\underline{P}, \mathbf{G}) = \frac{\sum_k \log(1+\gamma_k)}{\sum_k b_k + P_k} \tag{4.25} $$

The choice of the utility is chosen to be the average energy efficiency for this part with $b_k$ being the constant power consumed. Typical values for $b_k$ in a small cell network are chosen and $b_k = 0.9W$ is fixed with $\max(P_k) = 0.1W$. The proposed scheme is also compared to the "Nash equilibrium" point of such a system where each transmitter blindly tries to optimize it's own rate resulting in $P_k^{NE} = \max(\mathcal{P})$. A distributed system that does not implement the proposed scheme would naturally operate at this point.

Some other choices that are made for the purpose of simulations:

1. $E[g_{i,i}] = 1$ when averaged over the fast fading.
2. The error in the SINR estimate is only due to quantization of the feedback. The SINR estimate available at the transmitter is given by $\widehat{\gamma} = Q(\gamma, \beta)$ where $\beta$ is the number of quantization levels available. For example, in an eight bit feedback channel, $\beta = 2^8$.
3. $T_1 = T_2 = K$.

In Figure 4.5, the utility from using the proposed scheme over the NE $100\left(\frac{W^{PS}}{W^{NE}}-1\right)$ is plotted against the ratio of the direct channel to the cross channel. The figure shows that even when the feedback is only 4-bits, the proposed scheme performs upto 200% better than the NE in the high interference regime ($E[g_{i,i}/g_{i,j}] < 2 \forall i \neq j$. On the other hand, when the interference level is low ($E[g_{i,i}/g_{i,j}] > 6$), it is seen that the NE outperforms the proposed scheme due to the noisy feedback. Of course, when a 8 or 12 bit feedback is used, the proposed scheme starts to beat the NE. Therefore, the conclusion that the proposed scheme should be used only in the case of strongly interfering devices, is drawn.

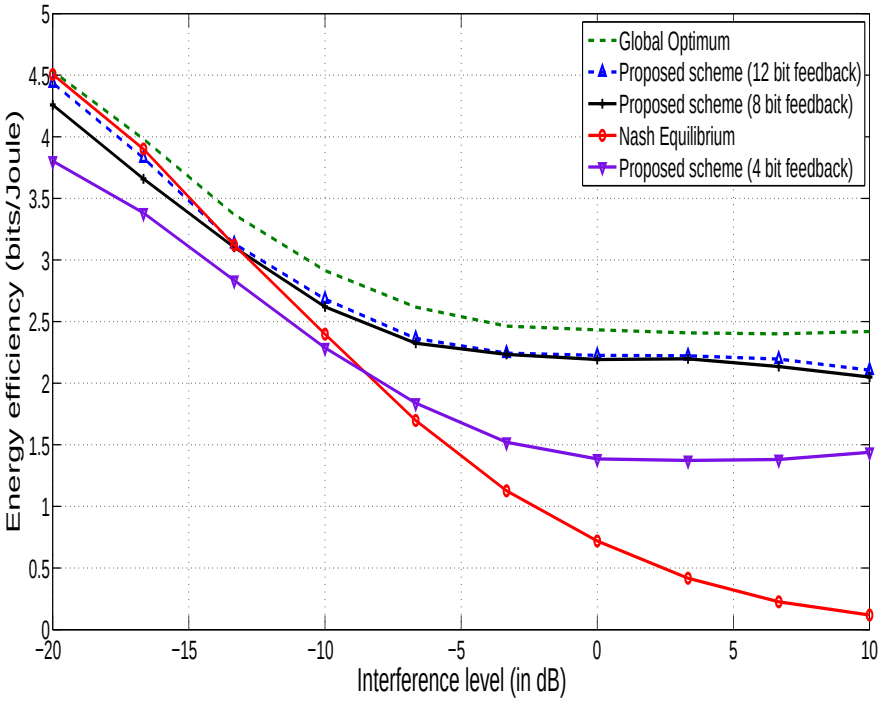


Figure 4.5: $100\left(\frac{W^{PS}}{W^{NE}}-1\right)$ v.s $E[g_{i,i}/g_{i,j}] \forall i \neq j$. The gain in utility while using the proposed scheme in phase 3 over the distributed choice at the NE is plotted in percentage.

In Figure 4.6, the utility $W$ (the sum rate in this case), is plotted against the number of power levels used in phase 2. This figure shows that in most cases using just two power levels in the best strategy.

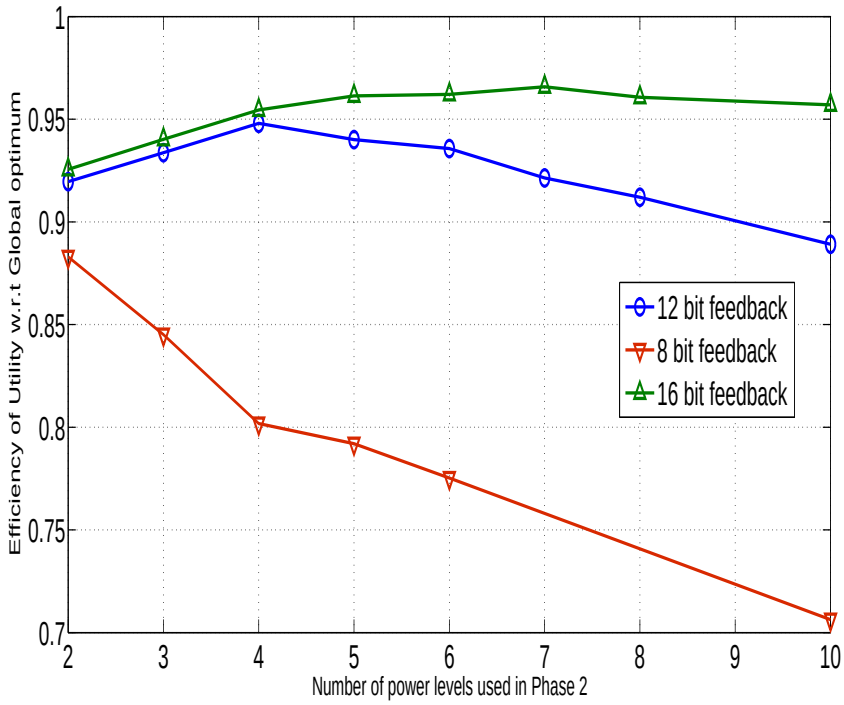


Figure 4.6: $\frac{W^{PS}}{W^{GO}}$, the efficiency of the proposed scheme w.r.t the global optimum v.s the total number of power levels used in phase 2.

# Chapter 5

# Conclusions and future research

## 5.1 Concluding remarks

Chapter 2 proposes a framework for studying the problem of energy-efficient pre-coding (which includes the problem of power allocation and control) over MIMO channels under imperfect channel state information and the regime of finite block length. As in [8], energy-efficiency is defined as the ratio of the block success rate to the transmit power. But, in contrast with [8] and the vast majority of works originating from it, an empirical choice for the success rate is not assumed, such as taking $f(x) = (1 - e^{-x})^L$, where $L$ is the block length. Instead, the numerator of the proposed performance metric is built from the notion of information, and more precisely from the average information (resp. mutual information) in the case where CSIT is available (resp. not available). This choice, in addition to giving a more fundamental interpretation to the metric introduced in [8], allows one to take into account in a relatively simple manner effects of practical interest such as channel estimation error and block length finiteness. Both in the case where (imperfect) CSIT is available and not available, it is shown that using all the available transmit power is not optimal. When CSIT is available, whereas determining the optimal power allocation scheme is a well known result (water-filling), finding the optimal total amount of power to be effectively used is non-trivial. Interestingly, the corresponding optimization problem can be shown to be quasi-convex and have a unique solution, the latter being characterized by an equation which is easy to solve numerically. Numerical results are provided to sustain the proposed analytical framework, from which interesting observations can be made which include : block length finiteness gives birth to the existence of a non-trivial trade-off between spectral efficiency and energy efficiency ; using optimal power allocation brings a large gain in terms of energy-efficiency only when

the channel has a large enough coherence time,demonstrating the value of CSIT and channel training. The proposed framework is useful for engineers since it provides considerable insights into designing the physical layer of MIMO systems under several assumptions on CSI. The next part of this chapter studied the performance of virtual MIMO systems from an energy efficiency perspective. It defines an energy efficiency metric that takes into account the capacity as well as the energy consumption. The power allocations of the different antennas is optimized and its shown that sleep mode can bring a significant energy efficiency gain. This involves deactivating antennas that do not have a significant contribution to the system capacity, for a given number of users and radio channel conditions.

Compared to the closest related works, the work reported in Chapter 3 possesses three salient features: The (possible) existence of packet buffer with finite size is taken into account; The total power consumed by the transmitter is considered; The proposed formulation considers the QoS. Remarkably, even though the derived energy-efficiency performance metric is seemingly more complex, it possesses all the main properties necessary for designing efficient distributed algorithms. Quite surprisingly, this is not only true when the packet arrival rate is constant (CAR protocol) but also when it is assumed to be adapted as a function of the SINR and the subsequent packet loss through the AAR protocol. One of the consequences of these properties is that the proposed iterative distributed power control algorithm converges towards a unique Nash equilibrium of the power control game associated with both transport protocols. While the cross-layer generalization of energy-efficient power control is supported by several key analytical results, numerical results strongly support the proposed approach as well. One of the key observations made from simulations is that a distributed power control scheme can perform as well as a centralized solution in some situations; realistic settings under which the PoA is close to one are clearly identified. Also, it is clearly explained why maximizing EE amounts to minimizing energy in communication systems with re-transmission protocols and this key interpretation is exploited to assess the gain in terms of saved energy brought by the proposed approach.

In Chapter 4, for the centralized system, it is seen that taking into account the dynamic nature of user traffic can significantly impact the average energy efficiency of a network. The notion of a "global" energy-efficiency is introduced, which is defined as the average of the energy-efficiencies of each state the cell can be in. These states represent the traffic configurations, i.e. the numbers and positions of users in the cell. Through extensive simulations its seen that optimizing the global efficiency yields a different power allocation from optimizing the efficiency of each individual state. Although this difference can be neglected when considering a cell in which all users experience the same average SINR, when considering a more realistic setting where users are subject to heterogeneous radio conditions, the global

optimization yields a considerable gain. This is because, when users are considered as static, it may be optimal to give more power to cell center in order to increase throughputs. However, when the dynamic behavior of users is taken into account, giving more power to users with bad radio conditions will allow them leaving the system faster and thus alleviating load in the future. In a decentralized system, the repeated game approach can be used to justify why even non-cooperative decision makers would prefer to communicate and achieve a more efficient solution as a Nash equilibrium of the repeated game. From the analysis conducted in the previous section, it is seen that using power modulation to implicitly communicate with other transmitters could potentially improve the performance of the system. The performance gain when compared to a purely distributed solution is studied numerically for a specific utility and the results are seen to be promising. Note that the algorithm proposed here is quite general and can be applied for any utility, not just the energy efficiency.

## 5.2 Open problems and possibilities for future research

The proposed framework in Chapter 2 opens some interesting research problems related to MIMO transmission, which include finding the optimal pre-coding matrix for the general case of i.i.d. channel matrices under no CSIT. This problem has not been solved even for the case of large MIMO systems. Extending the proposed approach to the case of Rician channels with spatial correlations,tackling the important case of multiuser MIMO channels and considering the problem of distributed energy-efficient precoding are left as open problems. The work on multi-user MIMO described in Chapter 2 is applicable only for the specific case of a small cell cluster with a centralized network and CSIT. Thus, many extensions of the work described are possible. The most relevant extension is to apply the proposed framework taking into account user traffic and a random number of users. Another natural extension of the proposed framework is of course, to study the effect of different classes of mobility on the virtual MIMO scheme and to study a distributed network.

Some of the problems left open from the work described in Chapter 3 include the cross layer energy efficiency analysis in the multi-carrier case and also the case of frequency selective channels, these extensions being potentially related. When relevant, receivers might be assumed to implement successive interference cancellation. In order to obtain more efficient equilibrium points (e.g., in the sense of the sum-payoff or a given fairness criterion), it would be of high interest to exploit a more advanced game model such as a stochastic game; this extension is especially relevant if the queue state information has to be exploited. To go further in the direction of having a

very realistic wireless network model, a less trivial, but very relevant extension would be to analyze the case of a time-varying number of users. This is definitely both of practical and theoretical interest. Finally, the case of CAR and AAR transmitters simultaneously active in the network has not been studied yet.

The results in Chapter 4 on the centralized solution are in many ways incomplete as only numerical results have been described. In addition, due to high computational complexity of the problem, even the numerical solutions can be found only for special cases. Thus, any sort of analytic results or algorithmic improvement in finding an energy efficient solution would be a significant step forward. In the distributed case, a theoretical analysis on the efficiency of phases 1 and 2 is left as an open problem and for future research. Additionally, note that the proposed method can also be easily extended to a multi-carrier system with Phase 1 and 2 remaining as it is, but with the information on each carrier channel fading matrix learned and broadcasted in parallel on each carrier.

# Appendices

# Appendix A

# Papers on MIMO

## A.1 EUSIPCO-2011

• V.S Varma, S. Lasaulce, M. Debbah and S.E. Elayoubi, "Impact of Mobility on Wireless Green Networks", European Signal Processing Conference (EUSIPCO) 2011.

# IMPACT OF MOBILITY ON MIMO GREEN WIRELESS SYSTEMS

*Vineeth S Varma[1,2], Merouane Debbah[3], Samson Lasaulce[2] and Salah Eddine Elayoubi[1]*

[1]Orange Labs
92130 Issy Les Moulineaux
France
Email: vineeth.svarma,
salaheddine.elayoubi@orange-ftgroup.com

[2]LSS, SUPELEC
91192 Gif sur Yvette
France
Email: Vineeth.VARMA,
Samson.LASAULCE@lss.supelec.fr

[3]Alcatel-Lucent Chair, SUPELEC
91192 Gif sur Yvette
France
Email: merouane.debbah@supelec.fr

## ABSTRACT

This paper studies the impact of mobility on the power consumption of wireless networks. With increasing mobility, we show that the network should dedicate a non negligible fraction of the useful rate to estimate the different degrees of freedom. In order to keep the rate constant, we quantify the increase of power required for several cases of interest. In the case of a point to point MIMO link, we calculate the minimum transmit power required for a target rate and outage probability as a function of the coherence time and the number of antennas. Interestingly, the results show that there is an optimal number of antennas to be used for a given coherence time and power consumption. This provides a lower bound limit on the minimum power required for maintaining a green network.

## 1. INTRODUCTION

The rapidly increasing demand for higher data rates has to be met by network providers. Arbitrarily increasing the signal to noise ratio (SNR) is not possible due to mainly two reasons. First, the transmit power of the Base Stations (BS) has to be constrained for energy efficiency considerations. Secondly, the transmit power has to be regulated due to electromagnetic restriction issues like in [10]. As a consequence, it is of great importance to determine the minimum power a transmitter requires to provide a certain target rate in wireless communication.

A multiple input multiple output (MIMO) channel is a well known and promising technique to increase the performance of a point to point link. In essence, MIMO system consists of multiple transmitting antennas and receiving antennas. If we have perfect channel state information (CSI), then increasing the number of antennas would always increase the rate. However, if the channel has to be estimated through training, time has to be spent for training each degree of freedom. We consider the impact of mobility in a MIMO system. In its full generality, this problem should be treated as a non coherent MIMO channel for which the capacity is hard to determine (open problem). Therefore, we restrict ourselves to a training mechanism which is suboptimal but provides tractable expressions. It was shown in [2] that in the context of point to point MIMO, there is a tradeoff between the channel estimate and the training time. A training time that maximizes the achievable rate was found in [2]. In this work we study how mobility affects the outage rates in MIMO systems.

Our objective is to study the transmit power for a target data rate, outage probability and coherence time. With an infinite coherence time, it is known that increasing the number of antennas would cause the power consumption to decrease. However, if the coherence time is small, a fraction of the time has to be used for channel training which results in a higher power consumption to transmit at the same rate. In this paper, we tackle the problem of finding the optimum number of antennas and training time for single input single output (SISO), single input multiple output (SIMO), multiple input single output (MISO) and MIMO systems.

The system model is described in Section 2. The performance metric, schemes under which we calculate the transmit power and the objectives of this paper are detailed in Section 3. The power consumed in MISO, SIMO and MIMO systems are calculated analytically and numerically in Sections 4 and 5. The power thus calculated is studied as a function of the coherence time and the number of antennas. Finally, we draw conclusions on the effect of mobility on MIMO systems.

## 2. SYSTEM MODEL

Let us consider the input-output relationship of a MIMO system given by:

$$\mathbf{y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{s} + \mathbf{z} \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^N$ is the received signal, the dimension $N$ representing the number of receive antennas. The transmitted signal is $\mathbf{s} \in \mathbb{C}^M$, where $M$ is the number of transmit antennas.$\mathbf{H} \in \mathbb{C}^{M\times N}$ represents the channel connecting the $M$ transmit to the $N$ receive antennas, and $\mathbf{z} \in \mathbb{C}^N$ represents the additive noise. The matrix $\mathbf{H}$, and the vectors $\mathbf{s}$ and $\mathbf{z}$ all comprise of i.i.d complex Gaussian entries with mean zero and variance unity, that is $h_{ij}, s_i, z_j \sim \mathcal{CN}(0,1)$. The total transmit power is proportional to $\rho$ which is the average SNR at each receive antenna.

We assume that the channel obeys the simple discrete-time block-fading law, where the channel is constant for some discrete time interval, after which it changes to an independent value that it holds for the next interval [1]. The coherence time is essentially determined by the mobility of the user and so if we calculate the dependency of the transmit power with respect to the coherence time, we will establish a relation between the mobility of a user and the corresponding transmit power. We further assume that channel estimation (via training) and data transmission is to be done within this coherence time, after which a new training sequence is done. The time for training in symbols will be given by $t$ and the coherence time in symbols by $T$. The inverse of the fraction of time that is used for data transfer will be denoted by

$\zeta = (\log_2(e)(1-\frac{t}{T}))^{-1}$. (The $\log_2(e)$ is for the conversion of natural logarithms to base two). We also denote $\frac{t}{M}$ by $\tau$.

In the training phase, all $M$ transmitting antennas broadcast orthogonal sequences of known pilot symbols of equal power $\rho$ on all antennas. The orthogonality condition imposes $\tau \geq 1$. The receiver estimates the channel $\mathbf{H}$, based on the observation of the pilot sequence, as $\hat{\mathbf{H}}$ and the error in estimation is given as $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$. The channel estimate normalized to variance one is given by $\bar{\mathbf{H}}$. From [2] we know that the rate is lowest (worst case) when $\tilde{\mathbf{H}}$ is Gaussian and then, the channel model can be rewritten as

$$\mathbf{y} = \sqrt{\frac{\rho_{eff}}{M}}\bar{\mathbf{H}}\mathbf{s} + \bar{\mathbf{I}} \tag{2}$$

where $\rho_{eff}$ is given by $\frac{\tau\rho^2}{1+\rho+\rho\tau}$ and $\bar{\mathbf{I}}$ equal to $\rho\tilde{\mathbf{H}}\mathbf{s}+\mathbf{z}$ normalized to unit variance. (2) leads to a lower bound on the mutual information and the achievable rate. Thus, all formulas derived in the following sections give lower bounds on the achievable rate and upper bounds on the outage and transmit power. This was verified to be an effective model in other works as well (See [6]). The value of $\rho$ can be calculated from $\rho_{eff}$ by inverting the equation as

$$\rho = \frac{M\rho_{\text{eff}}(1+\tau+\sqrt{(1+\tau)^2+\frac{4\tau}{\rho_{\text{eff}}}})}{2t} \tag{3}$$

## 3. METRICS

The performance metric that we consider in this paper is the outage probability for a target rate. We evaluate the transmit power as a function of the number of antennas for a given coherence time with the outage $P_{out}$ constrained for a target rate $R$. We also find the least power that can achieve the target rate and outage by optimizing over the number of antennas.

We have found in (2) an effective SNR under the assumption of worst case noise. We only deal with the model where power allocation is uniform among all the antennas and so the rate we calculate is not the channel capacity but the achievable rate. With this model we find a lower bound $\gamma$ on the achievable rate, in bits per second per channel use, from [2] as

$$\gamma \quad = \quad \zeta^{-1}\log\det(I+\rho_{eff}\frac{\bar{\mathbf{H}}\bar{\mathbf{H}}^H}{M}) \tag{4}$$

If the target data rate is represented by $R$, the outage probability $P_{out}$ is defined as the probability that the rate in a channel realization (mutual information), $\gamma$ is lower than $R$ the target rate.

$$P_{out} = P(\gamma < R) \tag{5}$$

If $\mathscr{P}$ is the threshold outage probability required to maintain the quality of service (QoS), the performance metric has to satisfy the constraint $P_{out} \leq \mathscr{P}$. The transmission power is evaluated under two schemes, the fixed power scheme where transmission is always done at a constant power and the adaptive power scheme where transmission is done at the optimal power to achieve the target outage rate. Now, we discuss these two schemes in detail:

### 3.1 Fixed power scheme

The first scheme is applicable when transmission is always done with a constant power of $\rho_0$ and we shall refer to this as *fixed power scheme*. We calculate the $\rho_0$ as $\min(\rho$, such as $P_{out}(R,\rho) \leq \mathscr{P})$ which is the lowest SNR that can achieve the given outage probability for a target rate in a MIMO system. Analyzing the behavior of $\rho_0$ with respect to changes in coherence time and the configuration of the MIMO system will be useful to determine how mobility affects the transmission power.

### 3.2 Adaptive power scheme

The second scheme assumes that power control is implemented. In this model, the estimated channel state information (CSI) is sent back to the transmitting antennas through a feedback mechanism, which we assume to be instantaneous for simplicity. This assumption a a gross simplification in reality and especially when considering small coherence times, however as the qualitative results obtained will not very much different, we ignore the feedback load. The feedback in downlink MIMO is studied and optimized in [8] and other related works and can be used to extend this paper to include its effects. Based on the feedback, the optimal power to achieve the target rate is used for transmission. If the power required is more than $\rho_0$ (The constant power defined in the previous sub-section), transmission is halted[1]. As the power in this scheme is a function of the channel, for tractable measurements, the power calculated $\rho_{min}$ in this case is the average power over all possible channel realizations, and we refer to this model as *adaptive power scheme*. This model might be harder to implement due to the feedback mechanism required and larger Peak to average power ratio (PAPR).

$$\rho_{min} = \mathbb{E}_{\mathbf{H}}(\rho_a(\mathbf{H})) \tag{6}$$

where

$$\rho_a(\mathbf{H}) = \begin{cases} \rho(\mathbf{H}) & \text{if } \rho(\mathbf{H}) \leq \rho_0 \\ 0 & \text{otherwise} \end{cases}$$

## 4. THEORETICAL ANALYSIS

In this section we find expressions relating the outage rate $P_{out}$ and, the transmit power $\rho_0$ in the fixed power scheme and $\rho_{min}$ in the adaptive power scheme.

### 4.1 MISO

For the MISO system the optimal power allocation for the best outage probability has been conjectured in [5] and later on proved in [4]. Here we look at a MISO system with uniform power allocation. In this case, $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ reduces to a real number and $\det(I+\rho_{eff}\frac{\bar{\mathbf{H}}\bar{\mathbf{H}}^H}{M})$ becomes just $1+\sum_{i=1}^{M}\bar{h}_{ij}\bar{h}_{ij}^*$.

[1]This corresponds to a system where real time data processing is required like in voice or video calls where the rate is fixed and the packets are dropped if the power is insufficient as the transfer rate is fixed. For elastic services it is possible to lower the rate and continue communication with the maximum available power.

Thus we can rewrite (4) and (5) as

$$\begin{aligned}\rho_{eff} &= \frac{M(\exp(R\zeta)-1)}{\sum_{i=1}^{M}\sum_{j=1}^{N}\bar{h}_{ij}\bar{h}_{ij}^{*}} \\ &= \frac{M(\exp(R\zeta)-1)}{\omega^2} \end{aligned} \quad (7)$$

where $\sum_{i=1}^{M}\bar{h}_{ij}\bar{h}_{ij}^{*}$ is denoted as $\omega^2$. The distribution of $\omega^2$ is the Chi square distribution with 2M degrees of freedom, $\omega^2 \sim \chi^2(2M)$ [9]. If the maximum power $\rho$ is $\rho_0$, corresponding to a $\rho_{eff}$ of $\rho_{eff0}$ then, the outage is just the probability that $\rho_{eff0} \leq \frac{M(\exp(R)-1)}{\mathbf{Trace}(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)}$. Defining $\Omega_0^2 = \frac{M(\exp(R\zeta)-1)}{\rho_{eff0}}$, the outage can be calculated as

$$\begin{aligned}P_{out} &= \frac{\int_0^{\Omega_0}\omega^{M-1}\exp(-\frac{\omega}{2})d\omega}{2^M\Gamma(M)} \\ &= \frac{\gamma(M,\frac{\Omega_0}{2})}{\Gamma(M)} \end{aligned} \quad (8)$$

where $\Gamma$ is the Gamma function and $\gamma$ is the lower incomplete Gamma function. The minimum average effective SNR, with which this outage is achieved, can be calculated from (6), (7) and (8) as

$$\begin{aligned}\rho_{\text{min-eff}} &= \frac{M\exp(R\zeta-1)}{2^M\Gamma(M)}\int_{\Omega_0}^{\infty}\omega^{M-3}\exp(-\frac{\omega}{2})d\omega \\ &= \frac{M\exp(R\zeta-1)\Gamma(M,\frac{\Omega_0}{2})}{\Gamma(M)} \end{aligned} \quad (9)$$

where $\Gamma$ is the upper incomplete Gamma function. The actual average power in the adaptive power scheme can be obtained as $\rho_{min} = \rho(\rho_{\text{min-eff}})$ using (3).

An interesting observation can be made while comparing a single input single output (SISO) system and a MISO system with $M > 1$. If we assume that the peak power can be arbitrarily large to obtain an arbitrarily small outage, then the SISO system in the adaptive power scheme will also consume an arbitrarily large power given by $\rho_{min} \propto (\lim_{\rho_0\to\infty}\mathbf{Ei}(\frac{-1}{rho_0}) = \infty)$, where **Ei** is the exponential integral that approaches infinity in this limit. However the MISO system, with $M = 2$ for instance, will only consume a finite average power $\rho_{min} \propto (\lim_{\rho_0\to\infty}\exp(\frac{-1}{rho_0}) = 1)$. Outage in both cases is $\lim_{\rho_0\to\infty} 1-\exp(\frac{-1}{rho_0}) = 0$.

### 4.2 SIMO

Let us now look at a SIMO system with uniform power allocation. In this case, $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ is a matrix of rank one, so it has only one non-zero eigen value. This non-zero eigen value is also the trace and is given as $\sum_{j=1}^{N}\bar{h}_{ij}\bar{h}_{ij}^{*}$. $\det(I+\rho_{eff}\bar{\mathbf{H}}\bar{\mathbf{H}}^H)$ becomes just $1+\sum_{j=1}^{N}\bar{h}_{ij}\bar{h}_{ij}^{*}$ as all the other eigen values are zero. Thus we can rewrite (4) and (5) as

$$\rho_{eff} = \frac{(\exp(R\zeta)-1)}{\omega_*^2} \quad (10)$$

where $\sum_{j=1}^{N}\bar{h}_{ij}\bar{h}_{ij}^{*}$ is denoted as $\omega_*^2$. The distribution of $\omega_*^2$ is also the Chi square distribution with 2N degrees of freedom, , $\omega_*^2 \sim \chi^2(2N)$. If the maximum power $\rho$ is $\rho_0$, corresponding to a $\rho_{eff}$ of $\rho_{eff0}$ then, the outage is just the probability that $\rho_{eff0} \leq \frac{(\exp(R\zeta)-1)}{\mathbf{Trace}(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)}$. Defining $\Omega_{0*}^2 = \frac{(\exp(R\zeta)-1)}{\rho_{eff0}}$, the outage can be calculated as

$$P_{out} = \frac{\gamma(N,\frac{\Omega_{0*}}{2})}{\Gamma(N)} \quad (11)$$

where $\Gamma$ is the Gamma function and $\gamma$ is the lower incomplete Gamma function. The minimum average effective SNR with which this outage can be achieved can be calculated from (6), (10) and (11) as

$$\rho_{\text{min-eff}} = \frac{\exp(R\zeta-1)\Gamma(N,\frac{\Omega 0*}{2})}{\Gamma(N)} \quad (12)$$

where $\Gamma$ is the upper incomplete Gamma function. The actual average power in the adaptive power scheme can be obtained as $\rho_{min} = \rho(\rho_{\text{min-eff}})$ using (3).

### 4.3 MIMO

Calculating the outage probability in a general MIMO system is a tedious process and also it does not give an interpretable closed-form expression like in MISO or SIMO. Therefore, here we use results from the field of large random matrices to solve this problem and and show that the approximation holds even for a finite number of antennas.

**Lemma 1** *Given* $\mathbf{H} \in \mathbb{C}^{N\times M}$ *such that* $h_{i,j} \sim C\mathcal{N}(0,1)$. *As* $M,N \to \infty$ *such that* $\lim_{N,M\to\infty}\frac{N}{M} = c$,

$$\log\det(I+\frac{\rho_{eff}}{M}(\mathbf{H}\mathbf{H}^H)) - M\mu = \psi^o \quad (13)$$

*Where* $\mu = [c\log(1+\rho_{eff}-\rho_{eff}\alpha)+\log(1+c\rho_{eff}-\rho_{eff}\alpha)-\alpha]$
*and* $\alpha = \frac{1}{2}[1+c+\rho_{eff}^{-1}-\sqrt{(1+c+\rho_{eff}^{-1})^2-4c}]$

*Then,* $\psi^o \xrightarrow{D} \psi$ *(converges in distribution) where* $\psi \sim \mathcal{N}(0,\sigma)$ *with* $\sigma = -\log(1-\frac{\alpha^2}{c})$. *The proof is in [3].*

Lemma 1 gives us a simple and tractable expression to obtain a relation between the transmit power and the achievable rate. Additionally, we know how the rate converges in distribution allowing us to calculate the outage by transforming the complexity of working with $2MN$ random variables $\bar{\mathbf{H}}$ to a single random variable $\psi$. Now let us consider a MIMO system. From Lemma 1 we know that if the transmit power is $\rho_0$, the effective SNR can be found as $\rho_{eff0}$. If we define $\sigma_0 = R\zeta - M\mu(\rho_{eff0})$, we have outage as

$$\begin{aligned}P_{out} &\approx P(\psi < R\zeta - M\mu_0) \\ &= \frac{\int_{-\infty}^{R\zeta-M\mu_0}\exp(\frac{-\psi^2}{2\sigma^2})d\psi}{\sqrt{2\pi\sigma^2}} \\ &= 1-Q(\frac{M\mu_0-R\zeta}{\sigma}) \end{aligned} \quad (14)$$

where the $Q$ function is the tail probability of the standard normal distribution. This approximation and resulting calculations are verified by simulations in the next section. Consecutively, the minimum average SNR can be calculated from (6), (13) and (14) as

$$\rho_{\min} = \frac{\int_{M\mu_0-R\zeta}^{\infty}\rho(\rho_{eff}(\psi))\exp(\frac{-\psi^2}{2\sigma^2})d\psi}{\sqrt{2\pi\sigma^2}} \quad (15)$$

### 4.4 The inverse calculation

In the previous sections we have detailed equations that express the outage probability as a function of $\rho_0$. However, our objective is to find $\rho_0$ for a given threshold outage probability. This is achieved following these series of steps.

- Given an Outage probability threshold $P_0$. Since the $Q$-function is well known and invertable, we use the inverse $Q$-function to find $X = \frac{M\mu_0 - R\zeta}{\sigma}$.
- Now we use a linear search with $\rho$, M and $t$ as the parameters to calculate $\frac{M\mu_0 - R\zeta}{\sigma}$ and match it with the $X$ such that $\rho_0 = min(\rho; \frac{M\mu_0 - R\zeta}{\sigma} \geq X)$. Where $\mu_0$ and $\sigma$ are evaluated as functions of $\rho_{eff}(\rho)$.
- Now using 15 we evaluate $\rho_{min}$.

The results of these calculations are illustrated in the following section with our numerical results.

## 5. NUMERICAL RESULTS

All theoretical results obtained in the Section 4 are verified using Monte-Carlo simulations. The theoretical results are compared to the results from simulation by a linear search over $\rho$ that yields the desired $P_{out}$. Finally, we find $\{\rho_0, M, N, \tau\}$, such that $\rho_0 = \min(\rho_0(M,N,\tau)$, such as $M,N \in \mathbb{N}, \tau \geq 1)$ and $\{\rho_{min}, M^*, N^*, \tau^*\}$, such that $\rho_{min} = \min(\rho_{min}(M^*,N^*,\tau^*)$, such as $M^*,N^* \in \mathbb{N}, \tau^* \geq 1)$.

### 5.1 MISO

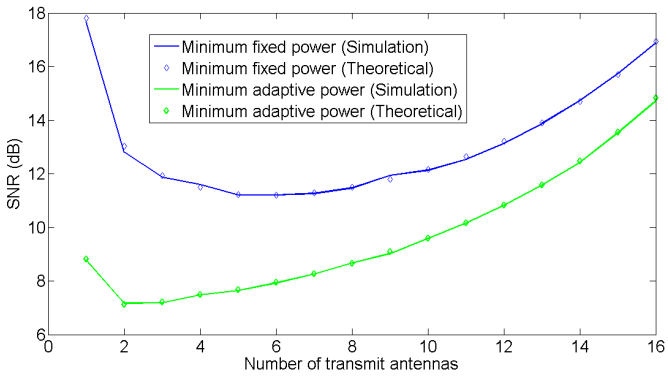


**Figure 1:** MISO system with $R = 1.44$ bps per hz, $\mathscr{P} = 5\%$ and $T = 25$ symbols.

The SNR, $\rho_0$ in the fixed power scheme and $\rho_{min}$ in the adaptive power scheme for a MISO system, with a target rate of 1.44 bps, outage rate of 5% and coherence time of 25 symbols, are plotted against $M = 1,..16$ in Figure 1. It is clear from Figure 1 that there is an optimal number of antennas for which the transmit power is minimized given a coherence time. As the number of antennas increases the gain from the additional degrees of freedom is lost due to the additional training time required. Another noteworthy fact is that the optimal number of antennas depends on the scheme of power transmission. The explanation for this is that the average SNR in the adaptive power scheme ($\rho_{min}$) increases as $M$ grows larger due to the channel hardening effect, while $\rho_0$ is still decreasing due to the gain from higher degrees of freedom. While $\rho_0$ takes its lowest value at $M = 6$, $\rho_{min}$ is minimized at $M = 2$.

### 5.2 SIMO

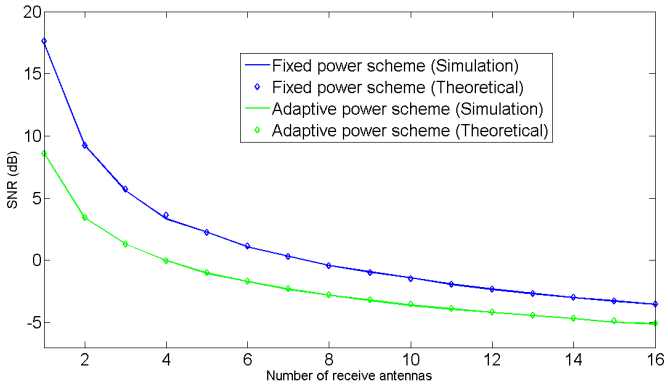


**Figure 2:** SIMO system with $R = 1.44$ bps per hz, $\mathscr{P} = 5\%$ and $T = 25$ symbols.

We also consider a SIMO system with the same parameters as the MISO system. The SNR, $\rho_0$ in the fixed power scheme and $\rho_{min}$ in the adaptive power scheme are plotted against the number of antennas in Figure 2. Here we can see that the SNR is a monotonically decreasing function of the number of antennas. However, here we ignore the computational power required and there is no additional training required for the receiving antennas, so this result is expected. Thus, increasing the number of receive antennas always improves the achievable rate thus decreasing $\rho_0$ and $\rho_{min}$.

### 5.3 MIMO

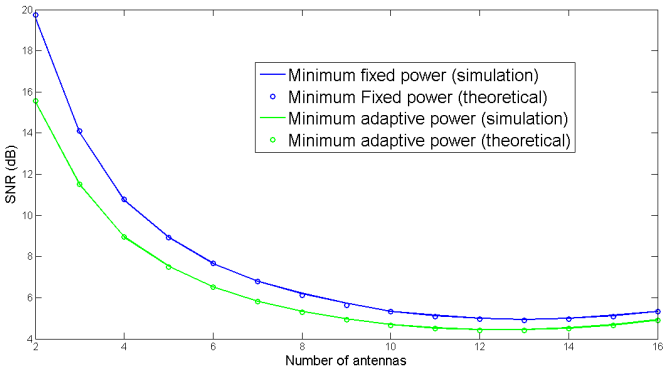


**Figure 3:** MIMO system with $R = 5.76$ bps per hz, $\mathscr{P} = 5\%$, $M = N$ and $T = 25$ symbols.

The transmit power $\rho_0$ in fixed power scheme and $\rho_{min}$ in adaptive power scheme for a MIMO system with a target rate of 5.76 bps and outage probability of 5% is plotted against the number of antennas ($N = M$) with $T = 25$ in Figure 3. We see that as the number of antennas increases the gain from the additional degrees of freedom and increased capacity is lost due to the additional training time required. The power in this case is minimized when $M = N = 13$.

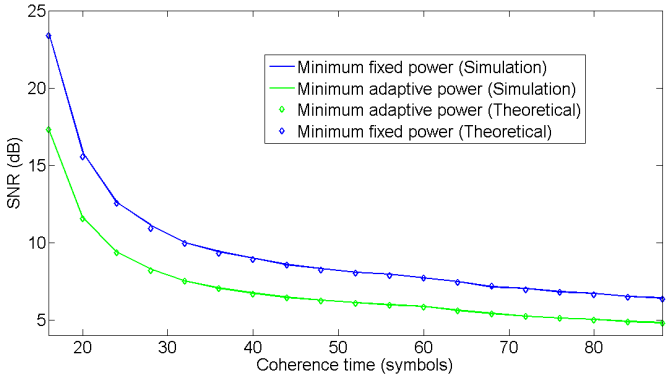


**Figure 4:** MIMO system with $R = 5.76$ bps per hz, $\mathscr{P} = 5\%$, $M = 8$ and $N = 4$.

Figure 4 plots the transmit power as a function of the coherence time for a MIMO system with $M = 8$ and $N = 4$ (target rate of 5.75 bps and outage of less than 5%). We see that as the coherence time increases the transmit power decreases as expected. The slope of the curve also decreases as $T$ increases.

### 5.4 Application

Let us consider a MIMO system operating between a single BS and a user with a mobile terminal of four receiving antennas. The long term evolution (LTE) standards specify the symbol duration to be $66.7\mu$s [7]. Let us consider a carrier frequency of 3 Ghz. If we consider a highly mobile user with a speed of 100kmph, the coherence time in symbols is about 55. Using equation (14), we find the optimal number of transmit antennas and the training time for a data rate of 16bps per hertz (LTE spectral efficiency can be up to 15 bps per hertz) and outage probability of less than 1%. Using a linear search on $M$ and $t$, we get the optimal number of antennas to be 8 and the training time to be 8 symbols corresponding to a minimum required SNR of 20 dB, while using 4 or 16 antennas would cost over 22 dB. If we consider a user with low mobility with a speed of about 10kmph, the coherence time is 550. The optimal number of antennas in this case would be 17 and the training time is 42 symbols. This corresponds to a minimum required SNR of about 16 dB while using just 4 antennas would cost over 20 dB.

## 6. CONCLUSION

In this paper, we have studied the impact of mobility on MISO, SIMO and MIMO systems. In order to have a tractable expression of the outage for a MIMO system we used recent results from the field of large random matrices. Simulations show these results to be tight even for two transmit and receive antennas. The quality of service is measured through the outage probability for a target rate in all cases and equations relating the outage probability to the transmit power were found. As a result, we see that given a coherence time, there is an optimal number of antennas for which the power required to transmit at a certain rate with a target outage probability is minimal. Studying a typical LTE system, we find that optimizing the number of antennas and training time can reduce power consumption up to 60% and that using adaptive power control can further reduce the power by 60%. Possible extensions of this work include the case of multi-user networks.

## 7. ACKNOWLEDGEMENTS

We would like to thank *LSS, SUPELEC, The Alcatel Lucent Chair and France Telecom* for providing the infrastructure and funding that helped in bringing this work to fruition. Additionally we would also like the thank J. Hoydis and V. Belmega for their insights and remarks that helped in editing and improving this work.

## REFERENCES

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", *IEEE Trans. on Information Theory*, Vol. 45, No. 1, Jan. 1999

[2] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?", *IEEE Trans. Information Theory*, Vol. 49, No. 4, APRIL 2003

[3] A. Kammoun, M. Kharouf, W. Hachem, J. Najim, "A Central Limit Theorem for the SINR at the LMMSE Estimator Output for Large-Dimensional Signals", *IEEE Trans. Information Theory*, Vol. 55, No. 11, Nov. 2009

[4] E. A. Jorswieck and H. Boche, "Outage Probability in Multiple Antenna Systems", *European Trans. Telecom.*, 2007

[5] E. Telatar. "Capacity of Multi-antenna Gaussian Channels", *European Trans. Telecommunications*, Vol. 10 (1999), pp. 585-595

[6] S. Lasaulce and N. Sellami, "On the Impact of using Unreliable Data on the Bootstrap Channel Estimation Performance", *Proc. IEEE SPAWC*, Italy, June 2003, pp. 348-352.

[7] 3GPP TR 36.814 v0.4.1, "Further Advancements for E-UTRA", *Physical Layer Aspects*, Feb. 2009.

[8] M. Kobayashi, N. Jindal, and G. Caire, "Training and Feedback Optimization for Multiuser MIMO Downlink," *IEEE Trans. Wireless Commun.*, Dec. 2009, submitted.

[9] M. Alexander, F. A. Graybill, D. C. Boes (1974). *"Introduction to the Theory of Statistics"* (Third Edition, p. 241-246). McGraw-Hill. ISBN 0-07-042864-6

[10] FCC Sec.15.249, *"Operation within the bands 902-928 MHz, 2400-2483.5 MHz, 5725-5875 MHZ, and 24.0-24.25 GHz"*

## A.2 TSP-2012

• V.S Varma, S. Lasaulce, M. Debbah and S.E. Elayoubi, "An Energy Efficient Framework for the Analysis of MIMO Slow Fading Channels", IEEE Trans. Signal Proc., Vol 61, 10, pp: 2647-2659.

# An Energy-Efficient Framework for the Analysis of MIMO Slow Fading Channels

Vineeth S. Varma, Samson Lasaulce, Merouane Debbah, and Salah Eddine Elayoubi,

***Abstract*—In this work, a new energy-efficiency performance metric is proposed for MIMO (multiple input multiple output) point-to-point systems. In contrast with related works on energy-efficiency, this metric translates the effects of using finite blocks for transmitting, using channel estimates at the transmitter and receiver, and considering the total power consumed by the transmitter instead of the radiated power only. The main objective pursued is to choose the best pre-coding matrix used at the transmitter in the following two scenarios : 1) the one where imperfect channel state information (CSI) is available at the transmitter and receiver ; 2) the one where no CSI is available at the transmitter. In both scenarios, the problem of optimally tuning the total used power is shown to be non-trivial. In scenario 2), the optimal fraction of training time can be characterized by a simple equation. These results and others provided in the paper, along with the provided numerical analysis, show that the present work can therefore be used as a good basis for studying power control and resource allocation in energy-efficient multiuser networks.**

***Index Terms*—Channel training, energy efficiency, finite block length, green communication, imperfect channel state information, MIMO.**

## I. Introduction

Over the past two decades, designing energy-efficient communication terminals has become an important issue. This is not surprising for terminals which have to be autonomous as far as energy is concerned, such as cellular phones, unplugged laptops, wireless sensors, and mobile robots. More surprisingly, energy consumption has also become a critical issue for the fixed infrastructure of wireless networks. For instance, Vodafone's global energy consumption for 2007-2008 was about 3000 GWh [1], which corresponds to emitting 1.45 million tons of CO2 and represents a monetary cost of a few hundred million Euros. This context explains, in part, why concepts like "green communications" have emerged as seen from [2], [3] and [4]. Using large multiple antennas, virtual multiple input multiple output (MIMO) systems, and small cells is envisioned to be one way of contributing to reducing energy consumption drastically. The work reported in this paper concerns point-to-point MIMO systems in which communication links evolve in a quasi-static manner, these channels are referred to as MIMO slow fading channels. The performance metric considered for measuring energy-efficiency of a MIMO communication corresponds to a trade-off between the net transmission rate (transmission benefit) and the consumed power (transmission cost).

The ultimate goal pursued in this paper is a relatively important problem in signal processing for communications. It consists of tuning the covariance matrix of the transmitted signal (called the pre-coding matrix) optimally. But, in contrast with the vast literature initiated by [5] in which the transmission rate is of prime interest, the present paper aims at optimizing the pre-coding matrix in the sense of energy-efficiency as stated in [6]. Interestingly, in [6] the authors bridge a gap between the pioneering work by Verdú on the capacity per unit cost for static channels [7] and the more pragmatic definition of energy-efficiency proposed by [8] for quasi-static single input single output (SISO) channels. Indeed, in [6], energy-efficiency is defined as the ratio of the probability that the channel mutual information is greater than a given threshold to the used transmit power. Assuming perfect channel state information at the receiver (CSIR) and the knowledge of the channel distribution at the transmitter, the pre-coding matrix is then optimized for several special cases. While [6] provides interesting insights into how to allocate and control power at the transmitter, a critical issue is left unanswered; to what extent do the conclusions of [6] hold in more practical scenarios such as those involving imperfect CSI? Answering this question was one of the motivations for the work reported here. Below, the main differences between the approach used in this work and several existing relevant works are reviewed.

In the proposed approach, the goal pursued is to maximize the number of information bits transmitted successfully per Joule consumed at the transmitter. This is different from the most conventional approach which consists in minimizing the transmit power under a transmission rate constraint: [9] perfectly represents this body of literature. In the latter and related works, efficiency is not the main motivation. [10] provides a good motivation as to how energy-efficiency can be more relevant than minimizing power under a rate constraint. Indeed, in a communication system without delay constraints, rate constraints are generally irrelevant whereas the way energy is used to transmit the (sporadic) packets is of prime interest. Rather, our approach follows the original works on energy-efficiency which includes [11], [8], [12], [13], [14]. The current state of the art indicates that, since [6], there have been no works where the MIMO case is treated by exploiting the cumulative distribution of the channel mutual information (i.e., the outage probability) at the numerator of the performance metric. As explained below, our analysis goes much further than [6] by considering effects such as channel estimation error

V. S. Varma and S. E. Elayoubi are with the Orange Labs, 92130 Issy Les Moulineaux, France (e-mail: vineethsvarma@gmail.com; salaheddine.elayoubi@orange.com).

S. Lasaulce and M. Debbah are with SUPELEC, 91192 Gif-sur-Yvette, France (e-mail: Samson.LASAULCE@lss.supelec.fr; merouane.debbah@supelec.fr).

effects. In the latter respect, several works address the issue of power allocation for outage probability minimization [15], [16], [17] under imperfect channel state information. The latter will serve as a basis for the analysis conducted in the present paper. At this point, it is possible to state the contributions of the present work.

In comparison to [6], which is the closest related work, the main contributions of the paper can be summarized as follows:

- one of the scenarios under investigation concerns he case where CSI is also available at the transmitter (only the case with CSIR and CSI distribution at the transmitter is studied in [6]).
- The assumption of perfect CSI is relaxed. In Sec. III, it is assumed that only imperfect CSIT and imperfect CSIR is available. Sec. IV considers the case with no CSIT and imperfect CSIR. In particular, this leads us to the problem of tuning the fraction of training time optimally. Exploiting existing works for the transmission rate analysis [18] and [19], it is shown that this problem can also be treated for energy-efficiency.
- The realistic assumption of finite block length is made. This is particularly relevant, since block finiteness is also a source of outage and therefore impacts energy-efficiency. Note that recent works on transmission under the finite length regime such as [20] provide a powerful theoretical background for possible extensions of this paper.
- Instead of considering the radiated power only for the cost of transmitting, the total power consumed by the transmitter is accounted for. Based on works such as [21], an affine relation between the two is assumed. Although more advanced models can be assumed, this change is sufficient to show that the behavior of energy-efficiency is also modified.

The paper is therefore structured as follows. Sec. II describes the proposed framework to tackle the aforementioned issues. Sec. III and IV treat the case with and without CSIT respectively. They are followed by a section dedicated to numerical results (Sec. V) whereas Sec. VI concludes the paper with the main messages of this paper and some relevant extensions.

## II. System model

A point-to-point multiple input and multiple output communication unit is studied in this work. In this paper, the dimensionality of the input and output is given by the numbers of antennas but the analysis holds for other scenarios such as virtual MIMO systems [22]. If the total transmit power is given as $P$, the average SNR is given by :

$$\rho = \frac{P}{\sigma^2} \tag{1}$$

where $\sigma^2$ is the reception noise variance.The signal at the receiver is modeled by :

$$\underline{y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\underline{s} + \underline{z} \tag{2}$$

where $\mathbf{H}$ is the $N \times M$ channel transfer matrix and $M$ (resp. $N$) the number of transmit (resp. receive) antennas. The entries of $\mathbf{H}$ are i.i.d. zero-mean unit-variance complex Gaussian random variables. The vector $\underline{s}$ is the $M$-dimensional column vector of transmitted symbols follows a complex normal distribution, and $\underline{z}$ is an $N$-dimensional complex white Gaussian noise distributed as $\mathcal{N}(\underline{0}, \mathbf{I})$. Denoting by $\mathbf{Q} = \mathbb{E}[\underline{s}\underline{s}^H]$ the input covariance matrix (called the pre-coding matrix), which satisfies

$$\frac{1}{M}\mathrm{Tr}(\mathbf{Q}) = 1 \tag{3}$$

where Tr stands for the trace operator. The power constraint is expressed as :

$$P \leq P_{\max} \tag{4}$$

where $P_{\max}$ is the maximum available power at the transmitter.

The channel matrix $\mathbf{H}$ is assumed to evolve in a quasi-static manner : the channel is constant for some time interval, after which it changes to an independent value that it holds for the next interval [18]. This model is appropriate for the slow-fading case where the time with which $\mathbf{H}$ changes is much larger than the symbol duration.

### A. Defining the energy efficiency metric

In this section, we introduce and justify the proposed definition of energy-efficiency of a communication system with multiple input and output antennas, and experiences slow fading.

In [8], the authors study multiple access channels with SISO links and use the properties of the energy efficiency function defined as $\frac{f(\rho)}{P}$ to establish a relation between the channel state (channel complex gain) ($h$) and the optimal power ($P^*$). This can be written as:

$$P^* = \frac{\mathrm{SNR}^*\sigma^2}{|h|^2} \tag{5}$$

where $\mathrm{SNR}^*$ is the optimal SNR for any channel state and (when $f$ is a sigmoidal/S-shaped function, i.e, it is initially convex and after some point becomes concave) is the unique strictly positive solution of

$$xf'(x) - f(x) = 0 \tag{6}$$

where $1 - f(.)$ is the outage probability. Formulating this problem in the case of MIMO channels is non-trivial as there is a problem of choosing the total transmit power as well as the power allocation.

When the same (imperfect) CSI is available at the transmitter and receiver, by estimating the channel for $t$ time, and sending the information to the transmitter for $t_f$ time, the energy-efficiency $\nu_T$ is defined as:

$$\nu_T(P, \mathbf{Q}, \hat{\mathbf{H}}) = \frac{R\left(1 - \frac{t+t_f}{T}\right) F_L\left[I_{\mathrm{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}}) - \frac{R}{R_0}\right]}{aP + b} \tag{7}$$

where $R$ is the transmission rate in bit/s, $T$ is the block duration in s, $R_0$ is a parameter which has unit Hz (e.g., the system bandwidth), and $a > 0$, $b \geq 0$ are parameters to relate the transmitter radiated power to its total consumed power ;

we define $\xi = \frac{R}{R_0}$ as the spectral efficiency. $I_{\mathrm{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}})$ denotes the mutual information with imperfect CSITR (the receiver also has the exact same CSI as the transmitter). This form of the energy-efficiency is inspired from early definitions provided in works like [8], and studies the gain in data rate with respect to the cost which is the power consumed. The numerator represents the benefit associated with transmitting namely, the net transmission rate (called the goodput in [23]) of the communication and is measured in bit/s. The goodput comprises a term $1-\frac{t+t_f}{T}$ which represents the loss in terms of information rate due to the presence of a training and feedback mechanism (for duration $t$ seconds and $t_f$ seconds resp. in a $T$s long block) [1]. The denominator of (7) represents the cost of transmission in terms of power. The proposed form for the denominator of (7) is inspired from [21] where the authors propose to relate the average power consumption of a transmitter (base stations in their case),to the average radiated or radio-frequency power by an affine model.

The term $F_L(.)$ represents the transmission success probability. The quantity $F_L(.)$ gives the probability that the "information" denoted by $\hat{I}$ as defined in [24]) is greater than or equal to the coding rate ($\xi$), i.e., it is the complementary cumulative distribution function of the information $\hat{I}$, $\mathrm{Prob}(\hat{I} \geq \xi)$. Formally, $\hat{I}$ is defined as $\hat{I} = \log \frac{\mathrm{PDF}_{X,Y}(x,y)}{\mathrm{PDF}_X(x)\mathrm{PDF}_Y(y)}$, where $\mathrm{PDF}_{X,Y}$ and $\mathrm{PDF}_X$ represents the joint and marginal probability distribution functions, $x$ and $y$ are samples of the process $X$ and $Y$, which in this case represent the transmitted and received signals. The average mutual information $I = E(\hat{I})$ is used to calculate this probability and $F_L(.)$ depends on the difference between $I$ and $\xi$. $F_L(.)$ can be verified to be sigmoidal (this is the cumulative probability distribution function of a variable with a single peaked probability distribution function) and $F_L(0) = 0.5$ (If $\xi = I$, $F_L(.)$ is the probability that a random variable is equal to or larger than its mean). When CSIT is available, it is possible to ensure that the data transmission rate is just below the channel capacity. If this is done, then there is no possibility of outage when the block length is infinite [25]. However, in most practical cases, the block length is finite and this creates an outage effect which depends on the block length $L$ [24].

The bounds on $F_L$ can be expressed as $F_L(I_{\mathrm{ICSITR}}(0, 0, \mathbf{H}) - \xi) = 0$ (no reliable communication when transmit power is zero) and as $F_L \to 1$ when $P \to \infty$. This proposed form for this function, $F_L(I_{\mathrm{ICSITR}}(P, \mathbf{Q}, \hat{\mathbf{H}}) - \xi)$, is supported by works like [24] and [26]. An approximation for this function based on the automatic repeat request protocol [27] is $F_L(x) = Q_{func}(-Tx)$, where $Q_{func}$ is the tail probability of the standard normal distribution.

Therefore, in the presence of CSI at the transmitter, outage occurs even when the mutual information is more than the targeted rate due to the noise and finite code-lengths. In this scenario, the energy-efficiency is maximized when the parameters $\mathbf{Q}$ and $P$ are optimized.

[1]In this case, we assume that the feedback mechanism is sufficient to result in perfect knowledge of $\hat{\mathbf{H}}$ at the transmitter. This is done because, assuming a different imperfect CSI at the transmitter from the receiver creates too much complexity and this problem is beyond the scope of a single paper.

In the absence of CSI at the transmitter, the earlier definition of energy efficiency is not suitable since $\mathbf{H}$ is random, $\nu_T$ is also a random quantity. Additionally, in this case, it is impossible to know if the data transmission rate is lower than the instantaneous channel capacity as the channel varies from block to block. Therefore, in this case, the source of outage is primarily the variation of the channel [28], and using (7) directly is not suitable. As the channel information is unavailable at the transmitter, define $\mathbf{Q} = \frac{\mathbf{I}_M}{M}$, meaning that the transmit power is allocated uniformly over the transmit antennas; in Sec. IV-C, we will comment more on this assumption. Under this assumption, the average energy-efficiency can be calculated as the expectation of the instantaneous energy-efficiency over all possible channel realizations. This can be rewritten as:

$$\nu_R(P,t) = \frac{R\left(1-\frac{t}{T}\right)\mathbb{E}_{\mathbf{H}}\left(F_L\left[I_{\mathrm{ICSIR}}(P,\mathbf{Q},\hat{\mathbf{H}}) - \frac{R}{R_0}\right]\right)}{aP+b}. \tag{8}$$

For large $L$, it has been shown in [28] (and later used in other works like [6]) that the above equation can be well approximated to :

$$\nu_R(P,t) = \frac{R\left(1-\frac{t}{T}\right)\mathrm{Pr}_{\mathbf{H}}\left[I_{\mathrm{ICSIR}}(P,t,\hat{\mathbf{H}}) \geq \xi\right]}{aP+b} \tag{9}$$

where $\mathrm{Pr}_{\mathbf{H}}$ represents the probability evaluated over the realizations of the random variable $\mathbf{H}$. Here, $I_{ICSIR}$ represents the mutual information of the channel with imperfect CSI at the receiver. Let us comment on this definition of energy efficiency. This definition is similar to the earlier definition in all most ways. Here the parameter $t$, represents the length of the training sequence used to learn the channel at the receiver[2]. The major difference here is that the expression for the success rate is the probability that the associated mutual information is above a certain threshold. This definition of the outage is shown to be appropriate and compatible with the earlier definition when only statistical knowledge of the channel is available [28].

Although very simple, these models allow one, in particular, to study two regimes of interest.

- The regime where $\frac{b}{a}$ is small allows one to study not only communication systems where the power consumed by the transmitter is determined by the radiated power but also those which have to been green in terms of electromagnetic pollution or due to wireless signal restrictions (see e.g., [29]).
- The regime where $\frac{b}{a}$ is large allows one to study not only communication systems where the consumed power is almost independent of the radiated power but also those where the performance criterion is the goodput.

Note that when $b = 0$, $t \to +\infty$, $T \to +\infty$, and $\frac{t}{T} \to 0$ equation (9) boils down to the performance metric investigated in [6].

[2]In this case, the optimization is done over $P$ and $t$ assuming imperfect CSI at the receiver. A parameter here not explicitly stated, but indicated nevertheless, is $M$ due to the number of transmit antennas affecting the effectiveness of training

### B. Modeling channel estimation noise

Each transmitted block of data is assumed to comprise a training sequence in order for the receiver to be able to estimate the channel; the training sequence length in symbols is denoted by $t_s$ and the block length in symbols by $T_s$. Continuous counterparts of the latter quantities are defined by $t = t_s S_d$ and $T = T_s S_d$, where $S_d$ is the symbol duration in seconds. In the training phase, all $M$ transmitting antennas broadcast orthogonal sequences of known pilot/training symbols of equal power on all antennas. The receiver estimates the channel, based on the observation of the training sequence, as $\widehat{\mathbf{H}}$ and the error in estimation is given as $\Delta\mathbf{H} = \mathbf{H} - \widehat{\mathbf{H}}$. Concerning the number of observations needed to estimate the channel, note that typical channel estimators generally require at least as many measurements as unknowns [19], that is to say $Nt_s \geq NM$ or more simply

$$t_s \geq M. \tag{10}$$

The channel estimate normalized to unit variance is denoted by $\widetilde{\mathbf{H}}$. From [19] we know that the mutual information is the lowest when the estimation noise is Gaussian. Taking the worst case noise, it has been shown in [18] that the following observation equation

$$\widetilde{\underline{y}} = \sqrt{\frac{\rho_{\text{eff}}(\rho,t)}{M}}\widetilde{\mathbf{H}}\underline{s} + \widetilde{\underline{z}} \tag{11}$$

perfectly translates the loss in terms of mutual information[3] due to channel estimation provided that the effective SNR $\rho_{\text{eff}}(\rho,t)$ and equivalent observation noise $\widetilde{\underline{z}}$ are defined properly namely,

$$\begin{cases} \widetilde{\underline{z}} & = & \sqrt{\frac{\rho}{M}}\Delta\mathbf{H}\underline{s} + \underline{z} \\ \rho_{\text{eff}}(\rho,t) & = & \frac{\frac{t}{MS_d}\rho^2}{1+\rho+\rho\frac{t}{MS_d}} \end{cases}. \tag{12}$$

As the worst case scenario for the estimation noise is assumed, all formulas derived in the following sections give lower bounds on the mutual information and success rates. Note that the lower bound is tight (in fact, the lower bound is equal to the actual mutual information) when the estimation noise is Gaussian which is true in practical cases of channel estimation. The effectiveness of this model will not be discussed here but has been confirmed in many other works of practical interest (see e.g., [31]). Note that the above equation can be utilized for the cases of imperfect CSITR and CSIR as well as the case of imperfect CSIR with no CSITR. This is because in both cases, the outage is determined by calculating the mutual information $I_{\text{ICSITR}}$ or $I_{\text{ICSIR}}$ respectively.

## III. Optimizing energy-efficiency with imperfect CSITR available

When perfect CSITR or CSIR is available, the mutual information of a MIMO system, with a pre-coding scheme $\mathbf{Q}$ and channel matrix $\mathbf{H}$ can be expressed as:

$$I_{\text{CSITR}}(P,\mathbf{Q},\mathbf{H}) = \log\left|\mathbf{I}_M + \frac{P}{M\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right| \tag{13}$$

The notation $|\mathbf{A}|$ denotes the determinant of the (square) matrix $\mathbf{A}$. With imperfect CSIT, which is exactly the same as the CSIR (i.e., both the transmitter and the receiver have the same channel estimate $\hat{\mathbf{H}}$), a lower bound on the mutual information can be found from several works like [15], [17] etc. This lower bound for $I_{\text{ICSITR}}$ is used, which is expressed as:

$$I_{\text{ICSITR}}(P,\mathbf{Q},\hat{\mathbf{H}}) = \log\left|\mathbf{I}_M + \hat{\mathbf{H}}\frac{P}{M\sigma^2(1+\rho\sigma_E^2)}\mathbf{Q}\hat{\mathbf{H}}^H\right| \tag{14}$$

where $\hat{\mathbf{H}}$ is the estimated channel and $1-\sigma_E^2$ is the variance of $\hat{\mathbf{H}}$. Considering the block fading channel model, from [15] and [19] we conclude that $\sigma_E^2 = \frac{1}{1+\rho\frac{t}{M}}$. Simplifying :

$$I_{\text{ICSITR}}(P,\mathbf{Q},\widehat{\mathbf{H}}) = \log\left|\mathbf{I}_M + \frac{\rho_{\text{eff}}}{M}\widehat{\mathbf{H}}\mathbf{Q}\widehat{\mathbf{H}}^H\right|. \tag{15}$$

Having defined the mutual information to be used for (7), we proceed with optimizing $\nu_T$.

### A. Optimizing the pre-coding matrix $\mathbf{Q}$

Studying (7) and (15), we see that varying the power allocation (or the corresponding pre-coding matrix) $\mathbf{Q}$, affects only the success rate $F_L(.)$ and the total power $P$ is the only term that is present outside $F_L(.)$. As $F_L(.)$ is known to be an increasing function, if the total power is a constant, optimizing the energy efficiency $\nu_T$ amounts to simply maximizing the mutual information $I_{\text{ICSITR}}(P,\mathbf{Q},\hat{\mathbf{H}})$. This is a well documented problem and it gives a "water-filling" type of solution [32]. Rewriting (13) as

$$I_{\text{ICSITR}}(P,\mathbf{Q},\widehat{\mathbf{H}}) = \log\left|\mathbf{I}_M + \frac{\rho_{\text{eff}}}{M}\mathbf{D}\mathbf{S}\mathbf{D}^H\right| \tag{16}$$

where the optimal covariance matrix $\mathbf{Q} = \mathbf{V}\mathbf{S}\mathbf{V}^H$ is achieved through the singular value decomposition of the channel matrix $\widehat{\mathbf{H}} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ and an optimal diagonal covariance matrix $\mathbf{S} = \text{diag}[s_1,\ldots,s_{\min(M,N)},0,\ldots,0]$. The water-filling algorithm can be performed by solving:

$$s_i = \left(\mu - \frac{1}{\rho\|d_i\|^2}\right)^+, \text{ for } i = 1,2,\cdots,\min(M,N) \tag{17}$$

where $d_i$ are the diagonal elements of $\mathbf{D}$ and $\mu$ is selected such that $\Sigma_{i=1}^{\min(M,N)} s_i = M$. Here $(x)^+ = \max(0,x)$, this implies that $s_i$ can never be negative. The actual number of non-vanishing entries in $\mathbf{S}$ depends on the values of $d_i$ as well $\rho$ (and thus $P$). Examining (17), we can see that when $\rho \to 0$, the water-filling algorithm will lead to choosing $s_j = M$ and $s_i = 0$ for all $i \neq j$, where $j$ is chosen such that $d_j = \max(d_i)$ (beamforming). Similarly for $\rho \to \infty$, $s_i = \frac{M}{\min(M,N)}$ (uniform power allocation).

[3] It is implicitly assumed that the mutual information is taken between the system input and output; this quantity is known to be very relevant to characterize the transmission quality of a communication system (see e.g. [30] for a definition).

### B. Determining the optimal total power

$\mathbf{Q}$ has been optimized in the previous section. From (7), we see that the parameters that can be optimized in order to maximize the energy efficiency are $\mathbf{Q}$ and $P$. Therefore, in this section, we try to optimize $P$, the total power. Note that for every different $P$, the optimal power allocation $\mathbf{Q}$ changes according to (17) as $\rho$ is directly proportional to $P$. Therefore optimizing this parameter is not a trivial exercise. Practically, $P$ represents the total radio power, that is, the total power transmitted by the antennas. This power determines the total consumed power $b+aP$, of base stations or mobile terminals and so, optimizing this power is of great importance.

In this section, a theorem on the properties of $\nu_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ is provided, where $\mathbf{Q}_{WF(P)}$ is the power allocation obtained by using the water-filling algorithm and iteratively solving (17) with power $P$. This procedure is said to be "iterative" because, after solving equation 17, if any $s_j < 0$, then we set $s_j = 0$ and the equation is resolved until the all solutions are positive. For optimization, desirable properties on $\nu_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ are differentiability, quasi-concavity and the existence of a maximum. The following theorem states that these properties are in fact satisfied by $\nu_T$.

*Theorem 3.1:* The energy-efficiency function $\nu_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ is quasi-concave with respect to $P$ and has a unique maximum $\nu_T(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}})$, where $P^*$ satisfies the following equation :

$$\begin{aligned} &\frac{\partial F_L[I_{\mathrm{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi]}{\partial P}\left(P^* + \tfrac{b}{a}\right) \\ &-F_L[I_{\mathrm{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi] \quad = 0 \end{aligned} \tag{18}$$

where $\frac{\partial}{\partial P}$ is the partial derivative.

The proof of this theorem can be found in Appendix A. From the above theorem and equation, we can conclude that the optimal transmit power for imperfect CSITR depends on several factors like

- the channel estimate $\widehat{\mathbf{H}}$,
- the target spectral efficiency $\xi$,
- the ratio of the constant power consumption to the radio-frequency (RF) power efficiency $\frac{b}{a}$,
- the channel training time $t$ and
- the noise level $\sigma^2$.

Note that in this model, we always assume the CSI at the transmitter to be exactly identical to CSI at the receiver. Because of this, we take the feedback mechanism to be perfect and take a constant time $t_f$. Although in practice, $t_f$ plays a role in determining the efficiency and the optimal power, in our model $t_f$ is a constant and does not appear in the equation for $P^*$. In our numerical results we focus on the impact of $\widehat{\mathbf{H}}$, $\xi$ and $\frac{b}{a}$ on $P^*$ and $\nu^*$. The impact of $t$ is not considered for this case, but is instead studied where we have no CSITR and imperfect CSIR, this choice helps in making the results presented easier to interpret and understand.

### C. An illustrative special case : SISO channels

A study on energy-efficiency in SISO systems have been studied in many works like [8] and [7]. However, the approach used in this paper is quite novel even for the SISO case and presents some interesting insights that have not been presented before. For the case of SISO, the pre-coding matrix is a scalar and $\mathbf{Q} = 1$. The optimal power can be determined by solving (18). For a SISO system with perfect CSITR and CSIR, $F_L$ can be expressed as

$$\begin{aligned} &F_L[I_{\mathrm{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi] = \\ &Q_{\mathrm{function}}\left(L(1+\|h\|^2\rho)\tfrac{\xi - \log(1+\|h\|^2\rho)}{\|h\|^2\rho}\right) \end{aligned} \tag{19}$$

from [24]. Using this expression, we can find $P^*$ maximizing $\nu_T$.

In the case of high SNR (and high $\xi$), a solution to this problem can be found as

$$\lim_{\rho\to\infty} F_L(P, 1, h) = Q_{func}\left(L[\xi - \log(1+\|h\|^2\rho])\right). \tag{20}$$

Solving (18)

$$\begin{aligned} &\tfrac{-L\|h\|^2}{\sqrt{\pi}(1+\|h\|^2\rho^*)}\exp\left(-L^2\left[\xi - \log(1+\|h\|^2\rho^*)\right]^2\right)\left(\rho^* + \tfrac{b}{a\sigma^2}\right) \\ &-Q_{func}\left(L[\xi - \log(1+\|h\|^2\rho^*])\right) = 0 \end{aligned} \tag{21}$$

From the above equation it can be deduced that if $b = 0$, for large $\xi$, $\log(1+\rho^*) \approx \xi$. While for low SNR, $\lim_{\rho\to 0} F_L(P, 1, h) = Q_{func}\left(L\tfrac{\xi - \log(1+\|h\|^2\rho)}{\|h\|^2\rho}\right)$ and so, if $b = 0$,

$$\begin{aligned} &\tfrac{1}{\sqrt{\pi}}\left[L + L\tfrac{\xi - |h\|^2\rho^*}{\|h\|^2\rho^*}\right]\exp\left(\tfrac{-1}{2}\left[L\tfrac{\xi - \|h\|^2\rho^*}{\|h\|^2\rho^*}\right]^2\right) \\ - \quad &Q_{func}\left(L\tfrac{\xi - \|h\|^2\rho^*}{\|h\|^2\rho^*}\right) = 0 \end{aligned} \tag{22}$$

Substitute $x = L\frac{\xi - \|h\|^2\rho^*}{\|h\|^2\rho^*}$ and we have

$$\frac{1}{\sqrt{\pi}}\left[L + x\right]\exp\left(\frac{-1}{2}x^2\right) - Q_{func}(x) = 0. \tag{23}$$

As seen from the above equation, the value of $x$ depends only on $L$ the block length. For example if $L = 10$, we get $x \approx -1.3$. So, $\rho^* = 1.14\frac{\xi}{\|h\|^2}$. Whereas if $L = 100$ we get $\rho^* = 1.02\frac{\xi}{\|h\|^2}$. Note that these calculations are true only for $\xi \to 0$ so that $\rho \to 0$ is satisfied.

The above equations signify that for finite block lengths, the energy efficiency at $\xi \to 0$ is lower than the value calculated in [7] (of course, a direct comparison does not make sense as in [7], infinite block lengths are assumed). This suggests that a non-zero value of $\xi$ might optimize the energy efficiency. This value is evaluated in our numerical section and we find that the energy efficiency is optimized at a non-zero power.

### D. Special Case: Infinite code-length and perfect CSITR

When a very large block is used then the achievable rate approaches the mutual information [25], i.e $\lim_{L\to\infty, I_{\mathrm{CSITR}}-\xi\to 0^+} F_L(I_{\mathrm{CSITR}} - \xi) = 1$. Therefore in this limit, we can now simplify (7) to:

$$\nu_T(P, \mathbf{Q}, \widehat{\mathbf{H}}) = \frac{R_0\left(1 - \frac{t+t_f}{T}\right) I_{\mathrm{CSITR}}(P, \mathbf{Q}, \widehat{\mathbf{H}})}{aP + b}. \tag{24}$$

This is done because we replace $\xi$ with $I_{\mathrm{CSITR}}$ to maximize efficiency as $F_L$ is 0 when $I_{\mathrm{CSITR}} < \xi$, and choosing

$\xi \to I_{\text{CSITR}}$ maximizes efficiency. Water-filling optimizes the efficiency in this situation as well, and so we use $\mathbf{Q} = \mathbf{Q}_{WF(P)}$. It can be easily verified that for $b \to 0$: $\nu_T$ is maximized for $P \to 0$. And in this case, water-filling also implies that only the antenna with the best channel is used to transmit. Interestingly, when in the domain of finite code-lengths, our simulations indicate that there is a non-zero rate and power that optimizes the energy-efficiency function.

For general $b$, $\nu_T$ is optimized for $P^*$ satisfying:

$$\frac{\partial I_{\text{CSITR}}(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})}{\partial P}(aP+b) - I_{\text{CSITR}}(P, \mathbf{Q}, \widehat{\mathbf{H}}) = 0. \tag{25}$$

The above equation admits a unique maximum because $I_{\text{CSITR}}(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ is a concave function of $P$ (can be seen from Appendix A) and is mathematically appealing. For $\lim_{b\to 0} P^* = 0$ and as $\frac{b}{a}$ increases, $P^*$ also increases. A special case of this, with $b = 0$, and perfect CSITR, for a SISO channel has been studied in [7].

## IV. Optimizing energy-efficiency with no CSIT and imperfect CSIR

This problem has already been well analyzed in [6] when perfect CSI is available at the receiver and $b = 0$. So, in this paper we focus on the case when imperfect CSI is available and is obtained through channel training. For $I_{ICSIR}(P, t, \mathbf{H})$, we use a lower bound on the mutual information obtained from the equivalent observation equation (11), derived in [19]:

$$I_{ICSIR}(P, t, \hat{\mathbf{H}}) = \log \left| \mathbf{I}_M + \frac{1}{M}\rho_{\text{eff}} \left( \frac{LP}{\sigma^2}, t \right) \hat{\mathbf{H}}\hat{\mathbf{H}}^H \right| \tag{26}$$

Note that here, $Q = \frac{\mathbf{I}_M}{M}$ is used and has been shown to be optimal in [6]. In this section our focus is to generalize [6] to a more realistic scenario where the total power consumed by the transmitter (instead of the radiated power only) and imperfect channel knowledge are accounted for.

### A. Optimal transmit power

By inspecting (9) and (26) we see that using all the available transmit power can be suboptimal. For instance, if the available power is large and all of it is used, then $\nu_R(P, t)$ tends to zero. Since $\nu_R(P, t)$ also tends to zero when $P$ goes to zero (see [6]), there must be at least one maximum at which energy-efficiency is maximized, showing the importance of using the optimal fraction of the available power in certain regimes. The objective of this section is to study those aspects namely, to show that $\nu_R$ has a unique maximum for a fixed training time length and provide the equation determining the optimum value of the transmit power.

From [33] we know that a sufficient condition for the function $\frac{f(x)}{x}$ to have a unique maximum is that the function $f(x)$ be sigmoidal. To apply this result in our context, one can define the function $f$ by

$$f(\rho_{\text{eff}}) = \Pr\left[ \log \left| \mathbf{I}_M + \frac{1}{M}\rho_{\text{eff}} \mathbf{H}\mathbf{H}^H \right| \geq \xi \right]. \tag{27}$$

For the SISO case, for a channel with $h$ following a complex normal distribution, it can be derived that $f(\rho) = \exp\left(-\frac{2^\xi - 1}{\rho}\right)$ which is sigmoidal. It turns out that proving that $f$ is sigmoidal in the general case of MIMO is a non-trivial problem, as advocated by the current state of relevant literature [6], [34], [35]. In [6], $\nu_R(P)$ under perfect CSIR, was conjectured to be quasi-concave for general MIMO, and proven to be quasi-concave for the follwing special cases:

(a) $M \geq 1$, $N = 1$;
(b) $M \to +\infty$, $N < +\infty$, $\lim_{M\to\infty} \frac{N}{M} = 0$;
(c) $M < +\infty$, $N \to +\infty$, $\lim_{N\to\infty} \frac{M}{N} = 0$;
(d) $M \to +\infty$, $N \to +\infty$, $\lim\limits_{M\to+\infty, N\to+\infty} \frac{M}{N} = \ell < +\infty$;
(e) $\sigma^2 \to 0$;
(f) $\sigma^2 \to +\infty$;

In the following proposition, we give a sufficient condition to ensure that $\nu_R(P, t)$ is quasi-concave w.r.t $P$.

*Proposition 4.1 (Optimization of $\nu_R(P, t)$ w.r.t $P$):* If $\nu_R(P)$ with perfect CSIR is quasi-concave w.r.t $P$, then $\nu_R(P, t)$ is a quasi-concave function with respect to $P$, and has a unique maximum.

This proposition is proved in Appendix B. The above proposition makes characterizing the unique solution of $\frac{\partial \nu_R}{\partial P}(P, t) = 0$ relevant. This solution can be obtained through the root $\rho^*_{\text{eff}}$ (which is unique because of [33]) of:

$$\frac{L}{\sigma^2}\left(P + \frac{b}{a}\right)\frac{\tau\rho\left[(\tau+1)\rho + 2\right]}{\left[(\tau+1)^2 + 1\right]^2} f'(\rho_{\text{eff}}) - f(\rho_{\text{eff}}) = 0 \tag{28}$$

with $\tau = \frac{t_s}{M}$. Note that $P$ is related to $\rho$ through $P = \sigma^2\rho$ and $\rho$ is related to $\rho_{\text{eff}}$ through (12) and can be expressed as

$$\rho = \frac{1}{2\tau}\rho_{\text{eff}}\sqrt{(1+\tau)^2 + \frac{4\tau}{\rho_{\text{eff}}}}. \tag{29}$$

Therefore (28) can be expressed as a function of $\rho_{\text{eff}}$ and solved numerically; once $\rho^*_{\text{eff}}$ has been determined, $\rho^*$ follows by (29), and eventually $P^*$ follows by (1). As a special case we have the scenario where $b = 0$ and $\tau \to +\infty$; this case is solved by finding the unique root of $\rho^* f'(\rho^*) - f(\rho^*) = 0$ which corresponds to the optimal operating SNR in terms of energy-efficiency of a channel with perfect CSI (as training time is infinite). Note that this equation is identical to that in [8] and in this work, we provide additional insights into the form of the function $f(.)$.

Quasi-concavity is an attractive property for the energy-efficiency as quasi-concave functions can be easily optimized numerically. Additionally, this property can also be used in multi-user scenarios for optimization and for proving the existence of a Nash Equilibrium in energy-efficient power control games [8], [36], [37].

### B. Optimal fraction of training time

The expression of $\nu_R(P, t)$ shows that only the numerator depends on the fraction of training time. Choosing $t = 0$ maximizes $1 - \frac{t}{T}$ but the block success rate vanishes. Choosing $t = T$ maximizes the latter but makes the former term go to zero. Again, there is an optimal trade-off to be found. Interestingly, it is possible to show that the function $\nu_R(P^*, t)$ is strictly concave w.r.t. $t$ for any MIMO channels in terms of

$(M, N)$, where $P^*$ is a maximum of $\nu_R$ w.r.t $P$. This property can be useful when performing a joint optimization of $\nu_R$ with respect to both $P$ and $t$ simultaneously. This is what the following proposition states.

*Proposition 4.2 (Maximization of $\nu(P^*(t), t)$ w.r.t $t$):* The energy-efficiency function $\nu_R(P^*(t), t)$ is a strictly concave function with respect to $t$ for any $P^*(t)$ satisfying $\frac{\partial \nu_R}{\partial P}(P^*, t) = 0$ and $\frac{\partial^2 \nu_R}{\partial P^2}(P^*, t) < 0$, i.e, at the maximum of $\nu_R$ w.r.t. $P$.

The proof of this proposition is provided in Appendix C. The parameter space of $\nu_R$ is two dimensional and continuous as both $P$ and $t$ are continuous and thus the set $\nu(P^*(t), t)$ is also continuous and the proposition is mathematically sound. The proposition assures that the energy-efficiency can been maximized w.r.t. the transmit power and the training time jointly, provided $\nu_R(P, t)$ is quasi-concave w.r.t $P$ for all $t$. Based on this, the optimal fraction of training time is obtained by setting $\frac{\partial \nu_R}{\partial t}(P, t)$ to zero which can be written as:

$$\left(\frac{T_s}{M} - \tau\right) \frac{\rho^2(\rho + 1)}{[\tau\rho + \rho + 1]^2} f'(\rho_{\text{eff}}) - f(\rho_{\text{eff}}) = 0 \qquad (30)$$

again with $\tau = \frac{t_s}{M}$. In this case, following the same reasoning as for optimizing the $\nu_R$ w.r.t. $P$, it is possible to solve numerically the equation w.r.t. $\rho_{\text{eff}}$ and find the optimal $t_s$, which is denoted by $t_s^*$.

Note that the energy-efficiency function is shown to be concave only when it has already been optimized w.r.t $P$. The optimization problem studied here is basically, a joint-optimization problem, and we show that once $\nu(P, t)$ is maximized w.r.t $P$ for all $t$, then, $\nu(P^*(t), t)$ is concave w.r.t $t$. A solution to (30) exists only if $\nu_R$ has been optimized w.r.t $P$. However, in many practical situations, this optimization problem might not be readily solved as the optimization w.r.t $P$ for all $t$ has to be implemented first.

The following proposition describes how the optimal training time behaves as the transmit power is very large:

*Proposition 4.3 (Optimal $t$ in the high SNR regime):* We have that: $\lim_{P \to +\infty} t_s^* = M$ for all MIMO systems.

The proof for this can be found in Appendix D.

### C. Optimal number of antennas

So far we have always been assuming that the pre-coding matrix was chosen to be the identity matrix i.e., $\mathbf{Q} = \mathbf{I}_M$. Clearly, if nothing is known about the channel, the choice $\mathbf{Q} = \mathbf{I}_M$ is relevant (and may be shown to be optimal by formulating the problem as an inference problem). On the other hand, if some information about the channel is available (the channel statistics as far as this paper is concerned), it is possible to find a better pre-coding matrix. As conjectured in [5] and proved in some special cases (see e.g., [34]), the outage probability is minimized by choosing a diagonal pre-coding matrix and a certain number of 1's on the diagonal. The position of the 1's on the diagonal does not matter since channel matrices with i.i.d. entries are assumed. However, the optimal number of 1's depends on the operating SNR. The knowledge of the channel statistics can be used to compare the operating SNR with some thresholds and lead to this optimal number. Although we consider (9) as a performance metric instead of the outage probability, we are in a similar situation to [6], meaning that the optimal pre-coding matrix in terms of energy-efficiency is conjectured to have the same form and that the number of used antennas have to be optimized. In the setting of this paper, as the channel is estimated, an additional constraint has to be taken into account that is, the number of transmit antennas used, $M$, cannot exceed the number of training symbols $t_s$. This leads us to the following conjecture.

*Conjecture 4.4 (Optimal number of antennas):* For a given coherence time $T_s$, $\nu_R$ is maximized for $M^* = 1$ in the limit of $P \to 0$. As $P$ increases, $M^*$ also increases monotonically until some $P_+$ after which, $M^*$ and $t_s^*$ decreases. Asymptotically, as $P \to \infty$, $M^* = t_s^* = 1$.

This conjecture can be understood intuitively by noting that the only influence of $M$ on $\nu_R$ is through the success rate. Therefore, optimizing $M$ for any given $P$ and $t$ amounts to minimizing outage. In [5], it is conjectured that the covariance matrices minimizing the outage probability for MIMO channels with Gaussian fading are diagonal with either zeros or constant values on the diagonal. This has been proven for the special case of MISO in [34], we can conclude that the optimal number of antennas is one in the very low SNR regime and that it increments as the SNR increases. However, the effective SNR decreases by increasing $M$ (seen from expression of $\rho_{\text{eff}}$ and $\tau$) , this will result in the optimal $M$ for each $P$ with training time lower than or equal to the optimal $M$ obtained with perfect CSI. Concerning special cases, it can be easily verified that the optimal number of antennas is 1 at very low and high SNR.

At last, we would like to mention a possible refinement of the definition in (9) regarding $M$. Indeed, by creating a dependency of the parameter $b$ towards $M$ one can better model the energy consumption of a wireless device. For instance, if the transmitter architecture is such that one radio-frequency transmitter is used per antenna, then, each antenna will contribute to a separate fixed cost. In such a situation the total power can written as $aP + Mb_0$ where $b_0$ is the fixed energy consumption per antenna. It can be trivially seen that this does not affect the goodput in any manner and only brings in a constant change to the total power as long as $M$ is kept a constant. Therefore, the optimization w.r.t $P$ and $t$ will not change but it will cause a significant impact on the optimal number of antennas to use.

## V. Numerical results and interpretations

We present several simulations that support our conjectures as well as expand on our analytical results. All simulations are performed using Monte-Carlo simulations as there is no expression available for the outage of a general MIMO system.

### A. With imperfect CSITR available

The $F_L$ we use here is based on the results in [24], $F_L = Q_{func}(\frac{\xi - I_{\text{ICSITR}}(P, \mathbf{Q}_{WF}, \mathbf{H})}{\sqrt{\frac{2\rho}{(1+\rho)L}}})$, $L$ being the code-length. This is the Gaussian approximation that is very accurate for $L$ large enough and from simulations we observe that for $L \geq 10$ the approximation is quite valid.

First of all, numerical results are presented that support and present our analytical results through figures. The first two figures shown assume imperfect CSITR obtained through training and use a $2 \times 2$ MIMO system. The quasi-concavity of the the energy-efficiency function w.r.t the transmit power is shown in Figure 1 for $\xi = 1$ and $\xi = 4$, and $t_s = 2$ and $t_s = 10$. This figure shows that for a higher target rate, a longer training time yields a better energy-efficiency. We also observe that using a higher $\xi$ can results in a better energy-efficiency as in this figure. This motivates us to numerically investigate if there is also an optimal spectral efficiency to use, given a certain $T_s$, $\frac{b}{a}$ and $L$. Figures 3 and 2 present the results of this study.

Surprisingly, we observe that our plots are quasi-concave and so there is an optimal target rate to use for each channel condition and code-length. In Figure 2, $\nu_T$ is always optimized over $P$ and $\mathbf{Q}$. Observe that $\nu_T^*(\xi)$ is also quasi-concave and has a unique maximum for each value of $d_i$ and $t_s$ (representing the channel Eigen-values as from equations (16), (17) and training time lengths). $d_i$ is ordered in an ascending order, i.e. in this case, with $d_1^2 \leq d_2^2$. The parameters used are: $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $L = 100$ and $\frac{b}{a} = 1$ mW with $t_s = 2, 10$ and 20 for $d_1^2 = 1$, $d_2^2 = 3$, and $t_s = 2$ for $d_1^2 = d_2^2 = 1$.This figure also implies that the training time and target rate can be optimized to yield the maximum energy-efficiency for a given coherence-time and channel fading. For infinite code-length the plot is maximized at the solution of (25). While for Figure 3, perfect CSIT is assumed with $b = 0$, at infinite block length, the optimal transmit rate/power is zero as expected (also seen from (25)). However, remarkably, for finite code-lengths there is a non-zero optimal rate and corresponding optimal power as seen from the figure.

Finally in Figure 4, we compare our energy efficiency function that uses optimized power allocation to uniform power allocation, and present the gain from having CSIT. In both cases, the training time and the transmit power is optimized and we plot the optimized energy efficiency v.s $P_{\max}$. Note that the optimized PA always yields a better performance when compared to UPA and at low power, UPA has almost zero efficiency while the optimal PA yields a finite efficiency. The gain observed can be considered as the major justification in using non-uniform power allocation and sending the channel state information to the transmitter. However, when the block length is small, imperfect CSIT results in a smaller gain as seen from the relatively larger gap between $T_s = 100$ and $T_s = 10000$ when compared to the size of the gap in UPA.

### B. With no CSIT

We start off by confirming our conjecture that for a general MIMO system, $\nu_R(P, t)$ has a unique maximum w.r.t $P$. We also confirm that optimal values of training lengths and transmit antennas represented by $t_s^*$ and $M^*$ are as conjectured.

Once the analytical results have been established, we explore further and find out the optimal number of antennas

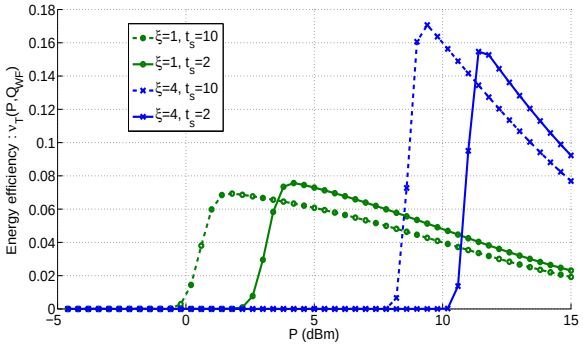


Fig. 1. Energy efficiency ($\nu_T$) in bits/J v.s transmit power (P) in dBm for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $\frac{b}{a} = 10$ mW for certain values of $\xi$ and $t_s$.

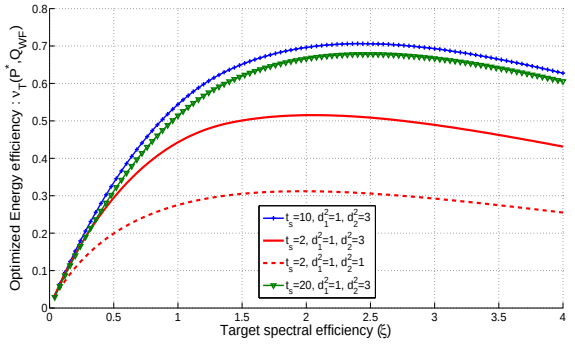


Fig. 2. Optimal energy-efficiency ($\nu_T(P^*, \mathbf{Q}_{WF})$) in bits/J v.s spectral efficiency ($\xi$) for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $T_s = 100$, $L = 100$ and $\frac{b}{a} = 1$ mW.

and training time when $\nu_R$ has been optimized w.r.t $P$. For this we use the optimized energy efficiency defined as $\nu^*(P, t) = \max\{\nu_R(p, t) \| p \in [0, P]\}$. As we know $\nu_R$ to be quasi-concave w.r.t $P$ and having a unique maximum, this newly defined $\nu^*$ will indicate what is the best energy efficiency achievable given a certain amount of transmit power $P$. Hence, plotting $\nu^*$ against $P$ for various values of $M$ or $t_s$ can be useful to determine the optimal number of antennas and training time while using the optimal power.

In the following plots we take $\frac{\sigma^2}{L} = 1mW$ so that $P$ can be expressed in $dBm$ easily. Also note that $\frac{b}{a}$ has the unit of power and is expressed in Watts (W). We also use $S_d = 15$ $\mu$s from LTE standards [38].

Figure 5 studies the energy efficiency as a function of the transmit power ($P$) for different values of $\frac{b}{a}$ and illustrates the quasi-concavity of the energy efficiency function w.r.t $P$. The parameters used are $R = 1600$, $\xi = \frac{R}{R_0} = 16$, $T_s = 55$ and $M = N = t = 4$.

Figure 6 studies the optimized energy efficiency $\nu^*$ as a function of the transmit power with various values of $t_s$. The figure illustrates that beyond a certain threshold on the available transmit power, there is an optimal training sequence length that has to be used to maximize the efficiency, when the optimization w.r.t $P$ has been done, which has been

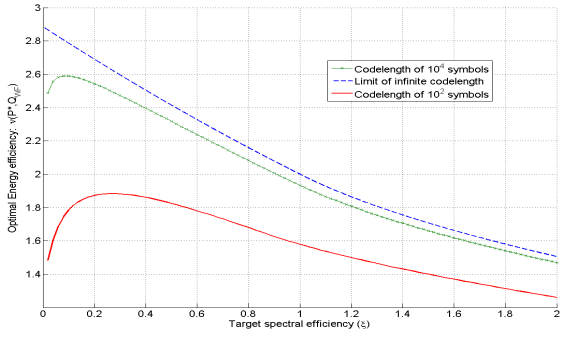


Fig. 3. Optimal energy-efficiency $(\nu_T(P^*, \mathbf{Q}_{WF}))$ v.s spectral efficiency $(\xi)$ for a MIMO system with perfect CSITR, $M = N = 2$, $R_0 = 1$bps and $\frac{b}{a} = 0$.

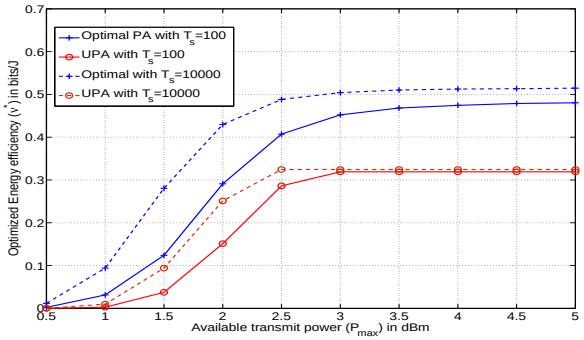


Fig. 4. Optimal energy-efficiency $(\nu_R(P^*))$ in bits/J v.s available transmit power $(\sup(P))$ for a MIMO system with imperfect CSITR, $M = N = 2$, $R_0 = 1$bps, $R = 1$bps and $\frac{b}{a} = 1$ mW.

proven analytically in proposition 4.2. The parameters are $R = 1$Mbps, $\xi = 16$, $\frac{b}{a} = 0$, $M = N = 4$, $\frac{b}{a} = 0$ and $T_s = 55$.

Figure 7 studies the optimal training sequence length $t_s$ as a function of the transmit power $P$. Note that in this case, we are not optimizing the efficiency with respect to $P$ and so this figure illustrates proposition 4.3. With $P$ large enough $t_s = M$ becomes the optimal training time and for $P$ small enough $t_s = T_s - 1$ as seen from the figure. The parameters are $R = 1600$, $\frac{b}{a} = 0$ W, $\xi = 16$ and $T_s = 10$. (We use $T_s = 10$, as if the coherence time is too large, the outage probabilities for low powers that maximize the training time, such that $t_s^* = T_s - 1$, become too small for any realistic computation.)

Figure 8 studies the optimal number of antennas $M^*$ as a function of the transmit power $P$ with the training time optimized jointly with $M$. With $P$ large enough $M = t_s = 1$ becomes the optimal number of antennas and for $P$ small enough $M = 1$ as seen from the figure. This figure illustrates conjecture 4.4. The parameters are $R_0 = 1$Mbps, $\frac{b}{a} = 10$ mW and $T_s = 100$.

From all of our theoretical and numerical results so far, we can conclude that given a target spectral efficiency $\xi$, a coherence block length $T_s$ and number of receive antennas, there is an optimal transmit power $P^*$, transmit antennas $M^*$ and training time $t_s^*$ to use that optimizes the energy efficiency.

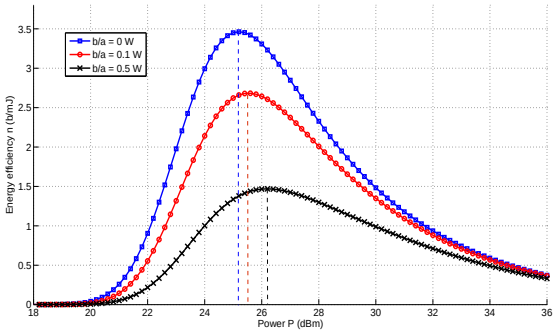


Fig. 5. Energy efficiency $(\nu_R)$ v.s transmit power (P) with $t_s = M = N = 4$,$R = 1600$bps, $\xi = \frac{R}{R_0}$=16 and $T_s = 55$ symbols.

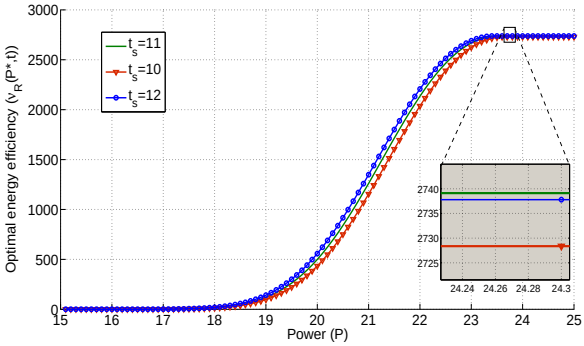


Fig. 6. Optimized efficiency $(\nu^*)$ vs. maximum transmit power $(P)$ for a MIMO system with $M = N = 4$, $R = 1$ Mbps, $\xi = \frac{R}{R_0}$=16, $T_s = 55$ and $\frac{b}{a} = 0$W.

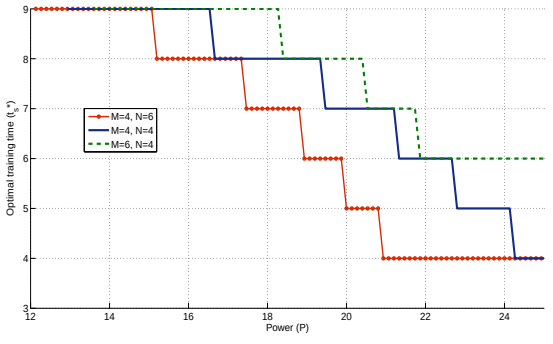


Fig. 7. Optimal training sequence length $(t_s)$ vs. Transmit Power (P) MIMO system with $\xi = \frac{R}{R_0} = 16$,$R = 1$Mbps, $N = 4$, $T_s = 10$ symbols. The discontinuity is due to the discreteness of $t_s$.

## VI. Conclusion

This paper proposes a framework for studying the problem of energy-efficient pre-coding (which includes the problem

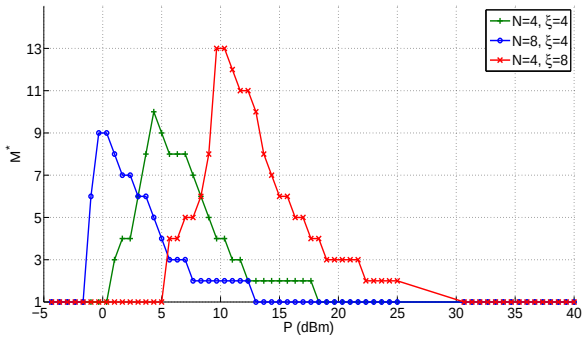


Fig. 8. Optimal number of antennas ($M^*$) vs. Transmit Power (P) in dBm for a MIMO system with $R_0 = 1$Mbps, $T_s = 100$, for certain values of $\xi$ and $N$, and $t_s$ optimized jointly with $M$. The discontinuity is due to the discreteness of $M$.

of power allocation and control) over MIMO channels under imperfect channel state information and the regime of finite block length. As in [8], energy-efficiency is defined as the ratio of the block success rate to the transmit power. But, in contrast with [8] and the vast majority of works originating from it, we do not assume an empirical choice for the success rate such as taking $f(x) = (1 - e^{-x})^L$, $L$ is the block length. Instead, the numerator of the proposed performance metric is built from the notion of information, and more precisely from the average information (resp. mutual information) in the case where CSIT is available (resp. not available). This choice, in addition to giving a more fundamental interpretation to the metric introduced in [8], allows one to take into account in a relatively simple manner effects of practical interest such as channel estimation error and block length finiteness. Both in the case where (imperfect) CSIT is available and not available, it is shown that using all the available transmit power is not optimal. When CSIT is available, whereas determining the optimal power allocation scheme is a well known result (water-filling), finding the optimal total amount of power to be effectively used is a non-trivial choice. Interestingly, the corresponding optimization problem can be shown to be quasi-convex and have a unique solution, the latter being characterized by an equation which is easy to solved numerically. When CSIT is not available, solving the pre-coding problem in the general case amounts to solving the Telatar's conjecture. Therefore, a new conjecture is proposed and shown to become a theorem in several special cases. Interestingly, in this scenario, it is possible to provide a simple equation characterizing the optimal fraction of training time. Numerical results are provided to sustain the proposed analytical framework, from which interesting observations can be made which includes : block length finiteness gives birth to the existence of a non-trivial trade-off between spectral efficiency and energy efficiency ; using optimal power allocation brings a large gain in terms of energy-efficiency only when the channel has a large enough coherence time,demonstrating the value of CSIT and channel training.

The proposed framework is useful for engineers since it provides considerable insights into designing the physical layer of MIMO systems under several assumptions on CSI. The proposed framework also opens some interesting research problems related to MIMO transmission, which includes : finding the optimal pre-coding matrix for the general case of i.i.d. channel matrices under no CSIT. Even in the case of large MIMO systems, this problem is not solved ; extending the proposed approach to the case of Rician channels with spatial correlations ; tackling the important case of multiuser MIMO channels ; considering the problem of distributed energy-efficient pre-coding.

## APPENDIX A
## PROOF OF THEOREM 3.1

In order to prove that $\nu_T(P, \mathbf{Q}_{WF(P)}, \mathbf{H})$ is quasi-concave with respect to $P$ and has a unique maximum $\nu_T(P^*, \mathbf{Q}_{WF(P^*)}, \mathbf{H})$, we exploit the result in [33] which states that if $f(x)$ is an "S"-shaped or sigmoidal function, then $\frac{f(x)}{x}$ is a quasi-concave function with a unique maximum. An "S"-shaped or sigmoidal function has been defined in [33] in the following manner. A function $f$ is "S" shaped, if it satisfies the following properties:

1) Its domain is the interval $[0, \infty)$.
2) Its range is the interval $[0, 1)$.
3) It is increasing.
4) ("Initial convexity") It is strictly convex over the interval $[0, x_f]$, with $x_f$ a positive number.
5) ("Eventual concavity") It is strictly concave over any interval of the form $[x_f, L]$, where $x_f < L$.
6) It has a continuous derivative.

Considering the non-constant terms in $\nu_T$, we see that what we have to show is that $F_L(I_{\text{ICSITR}}(P, \mathbf{Q}_{WF(P)}, \mathbf{H}) - \xi)$ is "S"-shaped w.r.t $P$. We already have that $F_L(x)$ is sigmoidal, therefore all we have to show is that $F_L(g(P))$ is also sigmoidal where $g(P) = I_{\text{ICSITR}}(P, \mathbf{Q}_{WF(P)}, \mathbf{H}) - \xi$. Trivially, when $P = 0$, $F_L(I_{\text{ICSITR}}(P)) = 0$ and $\lim_{P\to\infty} F_L = 1$. The rest can be proved using the following arguments:

- $g(P)$ is continuous: As $P$ varies, $\mathbf{Q}_{WF(P)}$ also is modified according to the iterative water-filling algorithm. This results in using one antenna for low $\rho$ to all antennas for high values of $\rho$.
  There exists certain "threshold" points of the total power, $P_i^{\text{th}}, i = \{1, \ldots, M\}$, at which the number of antennas used changes. The convention being, for $P_{i-1}^{\text{th}} \leq P \leq P_i^{\text{th}}$, $i$ number of antennas are used ($s$ for the rest are set to zero). $P_0^{\text{th}} = 0$ and $P_M^{\text{th}} = \infty$ . If $I_{\text{CSITR}}(P, \mathbf{Q}_{WF(P)}, \mathbf{H})$ is continuous at these points, then $g(P)$ is continuous. It can also be observed that in all other points, $I_{\text{CSITR}}(P, \mathbf{Q}_{WF(P)}, \mathbf{H})$ can be expressed as $\Sigma_{i=1}^{J} \log(1 + \alpha_i + \beta_i s_i), J \leq \min(M, N)$. ($\alpha$ and $\beta$ is obtained from solving (17).) A "threshold" point occurs when $P = P_j^{\text{th}}$ $s_j = 0$ is obtained by solving (17). The left hand limit is that $j - 1$ antennas are used and so, $I_{\text{CSITR}}(P, \mathbf{Q}_{WF(P)}, \mathbf{H}) = \Sigma_{i=1}^{j-1} \log(1 + \alpha_i + \beta_i s_i)$. The right hand limit will be obtained by solving (17), with $s_1 \to 0$ (assuming without loss of generality that $d_1^2$ is the smallest). This will yield a solution which can

be easily seen to be the same as the left hand limit as $p_1 \to 0$.

- We have shown $g(P)$ to be a finite sum of logarithms of a monomial expansion of $P$ in certain intervals (marked by $P_i^{\text{th}}$) . For each interval it is trivial to see that $F_L(g(P))$ is also "S"-shaped. As $g(P)$ is continuous, $F_L(g(P))$ is "S"-shaped for all $P$.
- From Lemma B proved in Appendix B we can show that $\frac{F_L(g(P))}{aP+b}$ is also "S"-shaped by a simple change of variable $x = aP + b$. Thus, we have $\nu_T(P, \mathbf{Q}_{WF(P)}, \mathbf{H})$ as a quasi-concave function with a unique maximum.
- With imperfect CSI, the only change is in $I_{\text{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}})$ now given from (15). The water-filling algorithm now replaces $\mathbf{H}$ with $\widehat{\mathbf{H}}$ and so on. This maintains the continuity of $g(P)$. However we now have $I_{\text{ICSITR}}(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}}) = \Sigma_{i=1}^{j} \log\left(1 + \frac{\alpha_i + \beta_i s_i}{1+\rho\sigma_E^2}\right)$. From [19] we have $\frac{1}{1+\rho\sigma_E^2}$ as a concave function and so even in this case, we have $F_L(g(P))$ as a sigmoidal function and $\nu_T(P, \mathbf{Q}_{WF(P)}, \widehat{\mathbf{H}})$ as quasi-concave with a unique maximum. As it is continuous and differentiable, the maximum can be found as the unique solution to the equation:

$$\begin{aligned} &\frac{\partial F_L[I_{\text{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \widehat{\mathbf{H}}) - \xi]}{\partial P}\left(P^* + \frac{b}{a}\right) \\ &-F_L[I_{\text{ICSITR}}(P^*, \mathbf{Q}_{WF(P^*)}, \mathbf{H}) - \xi] \quad = 0 \end{aligned} \tag{31}$$

where $\frac{\partial}{\partial P}$ is the partial derivative.

QED

## APPENDIX B
## PROOF OF PROPOSITION 4.1

An "S"-shaped function has been defined in [33] in the following manner. A function $f$ is "S" shaped, if it satisfies the properties as mentioned in Appendix A.

*Lemma 1:* If $f$ is a "S" shaped function, the composite function $f \circ g(x)$ is also "S" shaped if $g$ satisfies the following properties:

1) $g$ also satisfies conditions 1, 3, 4 and 6 but with $g(0) = b, b > 0$.
2) $\lim_{x\to\infty} f'(x)g'(x) = 0$.
3) $g''(x)$ is a decreasing function such that $\lim_{x\to\infty} g''(x) = 0$.

The proof for the above Lemma is at the end of this section.

In [6], the authors prove that the energy efficiency function with perfect CSI defined as the goodput ration to transmitted RF signal power is a quasi concave function by showing that the success rate function, $f(\rho)$ is "S" shaped for the following cases:

(a) $M \geq 1$, $N = 1$;
(b) $M \to +\infty$, $N < +\infty$;
(c) $M < +\infty$, $N \to +\infty$;
(d) $M \to +\infty$, $N \to +\infty$, $\lim_{M\to+\infty, N\to+\infty} \frac{M}{N} = \ell < +\infty$;
(e) $\sigma^2 \to 0$;
(f) $\sigma^2 \to +\infty$;

So, if we can show that the success rate function in our situation is also "S" shaped, our proof is complete for all the cases mentioned above. From (11) we know that the worst case mutual information in the case of imperfect CSI with training is mathematically equivalent to that of perfect CSI but with $\rho$ replaced by $\rho_{\text{eff}}$. Thus it is possible to replace $f(\rho)$, in the case of perfect CSI, by $f(\rho_{\text{eff}})$, when we study the case of imperfect CSI, and so we can study the energy efficiency function given by:

$$\nu_R(P, t) \quad = \quad R\zeta \frac{f(\rho_{\text{eff}}(p(x)))}{x} \tag{32}$$

where $x$ is a new variable that represents the total consumed power and $p(x) = \frac{L(x-b)}{a\sigma^2}$. $p'(x) > 0$ and $p''(x) = 0$ and $\rho'_{\text{eff}}(\rho) > 0$ and $\lim_{\rho\to\infty} \rho''_{\text{eff}}(\rho) = 0$. Thus $\rho_{\text{eff}}$ and $p$ satisfy the conditions on $g$ detailed in Lemma B. Hence we have proven that the numerator is "S" shaped with respect to $x$ and then it immediately follows from the results in [33] that $\nu_R$ has a unique maximum and is quasi-concave for all the specified cases.

**Proof of Lemma**

Here we show that $f \circ g$ also satisfies all the properties of the "S" function as described in [33].

1) Its domain is the domain of $g$ which is clearly the non-negative part of the real line; that is, the interval $[0, \infty)$.
2) Its range is the range of $f$, the interval $[0, 1)$.
3) It is increasing as both $f$ and $g$ are increasing.
4) ("Initial convexity"). Note that $f(g(x))'' = f''(y)g'(x) + g''(x)f'(y)$, with $y = g(x)$. As all terms in this expansion are positive in the interval $[0, x_f]$, $f \circ g$ is also convex in this interval. Also note that as $g'$ and $f'$ are strictly positive and $g''$ is decreasing, thus for $y > x_f$ once $f(g(x))'' < 0$ it stays negative till infinity. This implies that if there is an inflexion point, it is unique.
5) ("Eventual concavity") Consider $h(x) = f(g(x))' = f'(y)g'(x)$, due to the initial convexity and increasing nature of $h$, $h(x_f) = k, k > 0$. $\lim_{x\to\infty} f(g(x_f))' = 0$. As $h$ is continuous the mean value theorem imposes $h'(x) < 0$ at some point. This implies that there exists some point $x_d > 0$ such that $f \circ g$ is concave in the interval $[x_d, \infty]$ and convex before it.
6) It has a continuous derivative. (all the functions used here are continuous)

Hence, $f \circ g$ is "S" shaped.

QED

## APPENDIX C
## PROOF OF PROPOSITION 4.2

Let us consider the second partial derivative of $\nu_R$ with respect to $t$. (Note that this is possible as $t$ is a real number with the unit of time while $t_s$ is a natural number.) From (32), $\nu_R(P, t) = K^{-1}(1 - \frac{t}{T})f(\rho_{\text{eff}})$, with $K^{-1}1 = \frac{R}{x}$ a constant if $P$ is held a constant.

$$\begin{aligned} \frac{K\partial^2\nu}{\partial t^2} = \quad & (1 - \tfrac{t}{T})f''(\rho_{\text{eff}})\rho'^2_{\text{eff}}(t) + (1 - \tfrac{t}{T})f'(\rho_{\text{eff}})\rho''_{\text{eff}}(t) \\ & -\tfrac{2}{T}f'(\rho_{\text{eff}})\rho'_{\text{eff}}(t) \end{aligned} \tag{33}$$

In the above sum, it can be easily verified that the terms $f'(\rho_{\text{eff}})\rho'_{\text{eff}}(t)$ and $f'(\rho_{\text{eff}})$ are positive and that $\rho''_{\text{eff}}(t) < 0$. Thus if we have $f''(\rho_{\text{eff}}) < 0$, then $\nu_R(t)$ is strictly concave.

The only way $\nu_R$ depends on $P$ is through $\frac{f(\rho_{\text{eff}})}{aP+b}$ and $R\zeta$ stays a constant if only $P$ changes. So if we use the fact that we are working at a maximum of $\nu_R$ with respect to $\rho$, i.e $\frac{\partial \nu}{\partial \rho} = 0$ and $\frac{\partial^2 \nu}{\partial \rho^2} < 0$, we have $\frac{\partial \nu}{\partial \rho}$ as:

$$0 = f'(\rho_{\text{eff}})\rho'_{\text{eff}}(\rho)\rho^{-1} - f(\rho_{\text{eff}})\rho^{-2} \tag{34}$$

And, substituting (34) in $\frac{\partial^2 \nu}{\partial \rho^2} < 0$, that is:

$$\begin{aligned} f''(\rho_{\text{eff}})(\rho'_{\text{eff}})^2\rho^{-1} - 2f'(\rho_{\text{eff}})\rho'_{\text{eff}}\rho^{-2} + 2f(\rho_{\text{eff}})\rho^{-3} &< 0 \\ f''(\rho_{\text{eff}})(\rho'_{\text{eff}})^2\rho^{-1} + \frac{2}{\rho}(0) &< 0 \\ f''(\rho_{\text{eff}}) < 0 \end{aligned} \tag{35}$$

Thus, using (35) in (33), we have the result that $\nu_R(P^*, t)$ is strictly concave w.r.t $t$.

QED

## APPENDIX D
## PROOF OF PROPOSITION 4.3

The equation that describes the optimal training time $t_s^*$ can be written as:

$$(T - t_s^*)f'(\rho_{\text{eff}})\rho'_{\text{eff}}(t)_{t=S_d t_s^*} - f(\rho_{\text{eff}}) = 0 \tag{36}$$

Now, let us study the optimal training time $t_s^*$ in the very high SNR regime, ie. when $\rho \to \infty$. (Note that $P \to \infty$ is equivalent to $\rho \to \infty$).

High SNR regime: Applying the limit of $\rho \to \infty$ in (36), we get that

$$T_s - t_s^* = \lim_{\rho\to\infty} \frac{f(t\rho(1+t)^{-1})}{\rho f'(t\rho(1+t)^{-1})} \tag{37}$$

We know from various works including [6] that $\lim_{\rho\to\infty} f(t\rho(1+t)^{-1}) = 1$. Now let us consider $f'(\frac{t}{1+t}\rho)$.

For a MISO system we know from [39] that:

$$f_{MISO}(\rho_{\text{eff}}) \quad = \quad \frac{\gamma\left(M, \frac{2^\xi - 1}{\sqrt{\rho_{\text{eff}}}}\right)}{\Gamma(M)} \tag{38}$$

where $\Gamma$ is the Gamma function and $\gamma$ is the lower incomplete Gamma function. Now we can use the special property of the incomplete gamma function, $\lim_{x\to 0} \frac{\gamma(M,x)}{x^M} = 1/M$ detailed in [40] to determine $\lim_{\rho\to\infty} \rho f(\rho_{\text{eff}})' \propto \rho_{\text{eff}}^{-M/2}$. Plugging this into (37) we have:

$$\lim_{\rho\to\infty} T_s - t_s^* = \frac{1+t}{t}\rho_{\text{eff}}^{M/2} \tag{39}$$

Thus $\nu_R$ is optimized when $t_s \to -\infty$, but as $M \le t_s < T_s$, we have $\lim_{\rho\to\infty} t_s^* = M$ for all MISO systems.

Now for any MIMO system in general, using the eigen value decomposition of $\mathbf{HH}^H$ we have the eigenvalue decomposition $\log_2|I_M + \frac{\rho}{M}\mathbf{HH}^H| = \log_2(\Pi_{i=1}^L(1+\lambda_i))$, where $L = \min(M, N)$ and $\lambda_i$ are the eigenvalues of $\mathbf{HH}^H$. Applying the limit on $\rho$ and ignoring lower order terms we have

$$\lim_{\rho\to\infty} f(\rho_{\text{eff}}) = \Pr\left[\Pi_{i=1}^L \lambda_i \ge \frac{2^\xi}{\rho_{\text{eff}}^L}\right]. \tag{40}$$

We can observe that the above expression is a cumulative distribution function of $\Pi_{i=1}^L\lambda_i$ and so it's derivative is simply the PDF of $\Pi_{i=1}^L\lambda_i$. For $\rho \to \infty$ we have $f'(\rho_{\text{eff}}) = \Pr[\Pi_{i=1}^L\lambda_i = \frac{2^\xi}{\rho_{\text{eff}}^L}]$ As we know that the in general, if the number of transmit antennas are the same, $\Pr[\lambda_{MISO} > x] \le \Pr[\Pi_{i=1}^L\lambda_i > x]$ for any $x > 0$ [35] . Thus $f'_{MIMO}(\rho_{\text{eff}}) < f'_{MISO}(\rho_{\text{eff}})$, implying that for all MIMO systems $\lim_{\rho\to\infty} t_s^* = M$ from (37) and (39).

## REFERENCES


[1] D. Lister, "An Operator's View on Green Radio", Proc. IEEE Internat. Conf. on Commun, Workshops (ICC Workshops '09), 1st Internat. Workshop on Green Commun. (GreenComm '09) (Dresden, Ger, 2009).

[2] J. Palicot, and C. Roland, "On The Use Of Cognitive Radio For Decreasing The Electromagnetic Radiations", URSI 05, XXVIII General Assembly, New Delhi, India, October 23-29, 2005.

[3] GreenTouch, "Communications Turns Totally Green", Press Release, Jan. 11, 2010.

[4] Mobile VCE, Virtual Centre of Excellence in Mobile and Personal Communications, Core 5 Research Programme, "Green Radio - Sustainable Wireless Networks", Feb. 2009.

[5] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *European Trans. Telecommunications*, Vol. 10 (1999), pp. 585-595.

[6] E.V. Belmega, and S. Lasaulce, "Energy-efficient pre-coding for multiple-antenna terminals", IEEE. Trans. on Signal Processing, 59, 1, Jan. 2011.

[7] S. Verdú, "Spectral efficiency in the wideband regime", IEEE Trans. on Inf. Theory, vol. 48, no. 6, pp. 1319-1343, Jun. 2002.

[8] D. J. Goodman, and N. Mandayam, "Power Control for Wireless Data", IEEE Personal Communications, vol. 7, pp. 48-54, Apr. 2000.

[9] S. Cui, A. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks" , IEEE Journal on Selected Areas in Communications, Vol. 22, No. 6, pp. 1089-1098, Aug. 2004.

[10] V. Chandar, A. Tchamkerten, and D. Tse, Asynchronous Capacity Per Unit Cost , IEEE Inter. Symposium on Information Theory Proceedings (ISIT), pp. 280-284, June 2010.

[11] V. Shah, N. B. Mandayam, and D. J. Goodman, "Power control for wireless data based on utility and pricing", IEEE Proc. of the 9th Intl. Symp. Personal, Indoor, Mobile Radio Communications (PIMRC), Boston, MA, pp. 427-1432, Sep. 1998.

[12] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks", IEEE Trans. on Communications, vol. 50, No. 2, pp. 291-303, Feb. 2002.

[13] S. Yatawatta, A.P. Petropulu, and C.J. Graff, "Energy-efficient channel estimation in MIMO systems", EURASIP Journal on Wireless Communications and Networking, v.2006 n.2, pp. 317-320, April 2006.

[14] S. Buzzi, H. V. Poor, and D. Saturnino, "Energy-efficient resource allocation in multiuser MIMO systems: a game-theoretic framework, Proc. of the 16th European Signal Processing Conference (EUSIPCO 2008), Losanna (Svizzera), August 2008.

[15] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection", IEEE Journal on Selected Areas in Communications (JSAC), pp. 1478-1491, Sep. 2007.

[16] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state", IEEE Transactions on Information Theory, 56(6): 2845-2866, 2010.

[17] C. C. Gaudes, and E Masgrau, "Bounds on Capacity over Gaussian MIMO Multiaccess Channels with Channel State Information Mismatch," IEEE ICASSP06, vol. 4, pp. 93-96, May 2006

[18] T. L. Marzetta, and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", *IEEE Trans. on Information Theory*, Vol. 45, No. 1, Jan. 1999.

[19] B. Hassibi, and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?", *IEEE Trans. on Information Theory*, Vol. 49, No. 4, April, 2003.

[20] Polyanskiy, H. Vincent Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," IEEE Trans. Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
[21] F. Richter, A. J. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks", Proceedings of VTC Fall, pp. 1-5, Dresden, 2009.
[22] S. K. Jayaweera, "A virtual MIMO-based cooperative communications architecture for energy-constrained wireless sensor networks", IEEE Trans. Wireless Commun., vol. 5, pp. 984-989, May 2006.
[23] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," IEEE Trans. Commun., vol. 52, pp. 1885-1894, Nov. 2005.
[24] D. Buckingham, and M.C. Valenti, "The information-outage probability of finite-length codes over AWGN channels", IEEE Information Sciences and Systems, 2008. CISS 2008.
[25] C.E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948.
[26] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A Random Matrix Approach to the Finite Blocklength Regime of MIMO Fading Channels", IEEE International Symposium on Information Theory (ISIT), pp. 2181 - 2185, Cambridge, 2012.
[27] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-Layer Combining of Adaptive Modulation and Coding with Truncated ARQ Over Wireless Links, IEEE Trans. Wireless Commun., vol. 3, no. 5, pp. 1746-1755, Sept. 2004.
[28] L.H. Ozarow, S. Shamai, and A.D. Wyner, "Information theoretic considerations for cellular mobile radio", IEEE Vehicular Technology, vol. 43, No. 2, pp 359-378, May. 1994.
[29] FCC Sec.15.249, *Operation within the bands 902-928 MHz, 2400-2483.5 MHz, 5725-5875 MHZ, and 24.0-24.25 GHz*, Oct. 2011.
[30] T. M. Cover, and J. A. Thomas, "Elements of information theory", Wiley-Interscience, New York, NY, 1991.
[31] S. Lasaulce, and N. Sellami, "On the Impact of using Unreliable Data on the Bootstrap Channel Estimation Performance", *Proc. IEEE SPAWC*, Italy, June 2003, pp. 348-352.
[32] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems", IEEE Journal on Selected Areas Communications, vol 26, pp. 1341-1365, 2008.
[33] V. Rodriguez, "An Analytical Foundation for Ressource Management in Wireless Communication", IEEE Proc. of Globecom, San Francisco, CA, USA, pp. 898-902, Dec. 2003.
[34] E. A. Jorswieck, and H. Boche, "Outage probability in multiple antenna systems", European Transactions on Telecommunications, vol. 18, pp. 217-233, 2006.
[35] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices", PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.
[36] S. Lasaulce, and H. Tembine, "Game Theory and Learning for Wireless Networks: Fundamentals and Applications", Academic Press, pp. 1-336, Aug. 2011, ISBN 978-0123846983.
[37] S. Lasaulce, M. Debbah, and E. Altman, Methodologies for analyzing equilibria in wireless games, IEEE Signal Processing Magazine, vol. 26, no: 5, pp. 41-52, Sep. 2009.
[38] 3GPP TR 36.814 v0.4.1, "Further Advancements for E-UTRA", *Physical Layer Aspects*, Feb. 2009.
[39] M. Alexander, F. A. Graybill, and D. C. Boes (1974). *"Introduction to the Theory of Statistics"* (Third Edition, p. 241-246). McGraw-Hill. ISBN 0-07-042864-6.
[40] Digital Library of Mathematical Functions, Incomplete Gamma functions, 8.11. [Online] Available: http://dlmf.nist.gov/8.11#i

**Vineeth S Varma** was born in Tripunithura, India. He is currently doing his PhD at LSS under the supervision of Dr. Samson Lasaulce. He obtained his Bachelors in Physics with Honors from Chennai Mathematical Institute, India in 2008. He then proceeded to join the Erasmus Mundus course in Optics and obtained Masters in Science and Technology from Friedrich-Schiller-University of Jena in 2009 and Warsaw University of Technology in 2010. His areas of interest are Physics, Applied mathematics and Telecommunication. Since December 2010 he is working on his PhD aimed at finding the fundamental limits of energy on telecommunication and optimizing energy efficiency in networks. He is the recipient of the 2012 ACM VALUETOOLS best student paper award.

**Samson Lasaulce** received his BSc and Agrégation degree in Physics from École Normale Supérieure (Cachan) and his MSc and PhD in Signal Processing from École Nationale Supérieure des Télécommunications (Paris). He has been working with Motorola Labs for three years (1999, 2000, 2001) and with France Télécom R&D for two years (2002, 2003). Since 2004, he has joined the CNRS and Supélec as a Senior Researcher. Since 2004, he is also Professor at École Polytechnique. His broad interests lie in the areas of communications, information theory, signal processing with a special emphasis on game theory for communications. Samson Lasaulce is the recipient of the 2007 ACM VALUETOOLS, 2009 IEEE CROWNCOM, 2012 ACM VALUETOOLS best student paper award, and the 2011 IEEE NETGCOOP best paper award. He is an author of the book Game Theory and Learning for Wireless Networks: Fundamentals and Applications. He is currently an Associate Editor of the IEEE Transactions on Signal Processing.

**Mérouane Debbah** entered the Ecole Normale Supérieure de Cachan (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively. He worked for Motorola Labs (Saclay, France) from 1999-2002 and the Vienna Research Center for Telecommunications (Vienna, Austria) until 2003. He then joined the Mobile Communications department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor until 2007. He is now a Full Professor at Supelec (Gif-sur-Yvette, France), holder of the Alcatel-Lucent Chair on Flexible Radio and a recipient of the ERC starting grant MORE (Advanced Mathematical Tools for Complex Network Engineering). His research interests are in information theory, signal processing and wireless communications. He is a senior area editor for IEEE Transactions on Signal Processing. Mérouane Debbah is the recipient of the "Mario Boella" award in 2005, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award as well as the Valuetools 2007, Valuetools 2008, Valuetools 2012 and CrownCom2009 best student paper awards. He is a WWRF fellow. In 2011, he received the IEEE Glavieux Prize Award.

## A.3 PIMRC-2013

• V.S Varma, S.E. Elayoubi, M. Debbah and S. Lasaulce, "On the Energy Efficiency of Virtual MIMO Systems", IEEE Int. Symposium on personal, indoor and mobile communications (PIMRC 2013).

# On the Energy Efficiency of Virtual MIMO Systems

Vineeth S Varma[1,2], Salah Eddine Elayoubi[1], Merouane Debbah[4] and Samson Lasaulce[2]

[1]Orange Labs
92130 Issy Les Moulineaux
France
salaheddine.elayoubi@orange.com

[2]LSS, SUPELEC
91192 Gif sur Yvette
France
Vineeth.varma@lss.supelec.fr
Samson.lasaulce@lss.supelec.fr

[4]Alcatel Lucent Chair
SUPELEC
91192 Gif sur Yvette
France
Merouane.debbah@supelec.fr

***Abstract*—The major motivation behind this work is to optimize the sleep mode and transmit power level strategies in a small cell cluster in order to maximize the proposed energy efficiency metric. We study the virtual multiple input multiple output (MIMO) established with each base station in the cluster equipped with one transmit antenna and every user equipped with one receive antennas each. The downlink energy efficiency is analyzed taking into account the transmit power level as well as the implementation of sleep mode schemes. In our extensive simulations, we analyze and evaluate the performance of the virtual MIMO through zero-forcing schemes and the benefits of sleep mode schemes in small cell clusters. Our results show that for certain configurations of the system, implementing a virtual MIMO with several transmit antennas can be less energy efficient than a system with sleep mode using OFDMA with a single transmitting antenna for serving multiple users.**

## I. Introduction

The energy consumed by the radio access network infrastructure is becoming a central issue for operators [1]. The goal of this work is to provide insights on how to design green radio access networks, especially in the framework of virtual MIMO systems. Indeed, classical network architectures are focused on integrated, macro base stations, where each cell covers a predetermined area, and inter-cell interference is reduced by the means of fixed frequency reuse patterns [2]. Heterogeneous Networks (HetNets) introduced a new notion of small cells where pico or femto base stations are deployed within the coverage area of the macro base stations [3]. Virtual MIMO is a step forward in this context that allows distributed systems of base stations/antennas that cover a common area and cooperate in order to increase the overall spectral efficiency [4]. This paper focuses on these latter solutions and aims at addressing the problem from an energy efficiency point of view.

For classical macro networks, early works focused on designing energy-efficient power control mechanisms [5]. Therein, the authors define the energy-efficiency of a communication as the ratio of the net data rate (called goodput) to the radiated power; the corresponding quantity is a measure of the average number of bits successfully received per joule consumed at the transmitter. This metric has been used in many works. Although fully relevant, the performance metric introduced in [5] ignored the fact that transmitters consume a constant energy regardless of their output power level [6]. The impact of this constant energy has been studied for single user point-to-point MIMO systems in [7]. Sleep mode mechanisms have thus been regarded as a solution for this issue; they consist in deactivating network resources that have low traffic load, eliminating thus both the variable and constant parts of the energy consumption [1]. This mechanism has been applied to macro networks [1], as well as to heterogeneous networks with macro and small cells [3]. Our aim in this paper is to extend this concept to virtual MIMO networks, where an antenna that is not significantly contributing to the network capacity (for a given configuration of user positions and radio channels) is put into sleep mode.

The remainder of this paper is organized as follows. In section II, we present the system model and the resource allocation scheme. Section III presents our energy efficiency metric and optimizes it for a given system and channel configuration, using sleep mode mechanisms. Section IV presents some numerical examples and section V eventually concludes the paper.

## II. System model

The wireless system under consideration is the downlink in a virtual MIMO system within a small cell cluster. To be precise, each of the small cell base stations are connected to a central processor and so they act as antennas for the virtual MIMO as shown in Fig 1. We refer to the set of these base

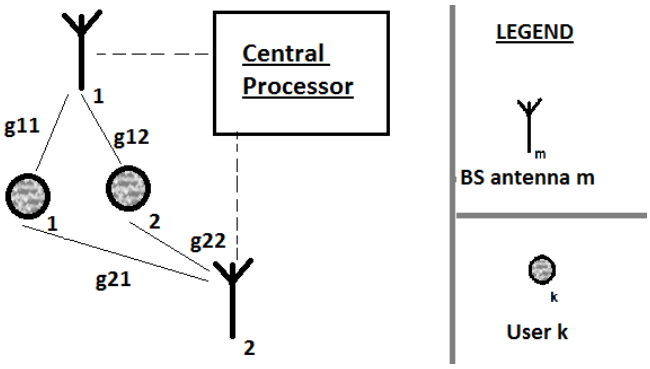


Fig. 1. An example illustration of a $2\times2$ virtual MIMO with $g_{i,j}$ representing the channel between BS antenna $i$ and user $j$.

stations as the "cluster". Each user is equipped with a single receive antenna. In order to eliminate interference zero-forcing is implemented. We consider a block-fading channel model where the channel fading stays is assumed to stay constant for the duration of the block and changes from block to block. The base stations require the channel state information available at

the user end in order to implement the zero-forcing technique. Therefore, in each block channel a training and feedback mechanism happens, after which data is transmitted. We also assume that every base station is capable of entering into a "sleep-mode". In this mode, the base station does not send any pilot signals and therefore does not perform the training or feedback actions consuming a lesser quantity of power compared to the active base stations. Let there be $M$ base stations in the cluster and $K$ users. Define $\mathcal{K} = \{1, 2, \ldots, K\}$ and $\mathcal{M} = \{1, 2, \ldots, M\}$ the sets of users and base station antennas.

### A. Power consumption model

As the transmit antennas are not co-located, each of them have an individual power budget. When a base station is active, it consumes a constant power of $b$ due to the power amplifier design and training or feedback costs. Additionally, it consumes a power $P_m\|x_m\|^2$ proportional to the radiated power, where $P_m \leq P_{\max}$ and $\|x_m\| \leq 1$ is the signal transmitted and $P_{\max}$ is the power constraint [1][6]. When it is placed on sleep mode, it is assumed that it only consumes power $c$ where $c < b$. Denote by $\mathbf{s}$ the sleep mode state vector of the cluster with elements $s_m \in \{0, 1\}$. The base station $m$ is in sleep mode when $s_m = 1$ and active when $s_m = 0$. Thus the power consumption of the m-th base station is $cs_i + (1 - s_i)(b + P_m\|x_m\|^2)$. The total power consumption of the cluster is given by:

$$P_{tot} = \sum_{m=1}^{M} cs_m + (1 - s_m)(b + P_m\|x_m\|^2) \tag{1}$$

For any given state of the cluster, we define $\omega(\mathbf{s})$ as the total number of base stations that are active. This value can be calculated as $\omega(\mathbf{s}) = M - \sum_m s_m$. If $M < K$, zero-forcing can not be used. However, if $M > K$, and $\omega(\mathbf{s}) \geq K$, then the zero-forcing technique can be implemented by choosing $K$ base stations to transmit the data signals after all $\omega(\mathbf{s})$ active base stations train and obtain feedback on their channels.

### B. The zero-forcing scheme

As there are $K$ users connected to the $\omega(\mathbf{s})$ active base stations, there are a total of $\omega(\mathbf{s}) \times K$ number of channels. Let $\zeta = \{1, 2, \ldots, \omega(\mathbf{s})\}$ be the set of active base stations. We denote by $\mathbf{G}$ with elements $g_{l,k} \in \mathbb{R}$ the path fading between base station $l \in \zeta$ and user $k \in \mathcal{K}$, while $\bar{\mathbf{H}}$ with elements $\bar{h}_{l,k} \in \mathbb{C}$ denotes the fast fading component, resulting in a signal at $k$ given by:

$$y_k = \sum_{l=1}^{\omega(\mathbf{s})} \sqrt{\frac{g_{l,k}P_l}{\sigma^2}}\bar{h}_{l,k}x_l + z \tag{2}$$

where $\mathbf{x}$ is the signal transmitted with $x_l$ as its elements; $z$ is the normalized noise and $\sigma^2$ the noise strength. Note that $g_{l,k}$ can be determined based on the user location while $\bar{h}_{l,k}$ are i.i.d. zero-mean unit-variance complex Gaussian random variables. We define $\tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}})$ as the combined channel matrix with elements $\tilde{h}_{l,k} = \sqrt{g_{l,k}}\bar{h}_{l,k}$.

In our work, as we focus on the small-cell scenario where the antennas are distributed over the cell in a dense manner, we assume that every user can have a similar level of signal strength. Define $\mathcal{N} = \{1, 2, \ldots, N\}$ as the set of transmitting antennas that perform zero forcing beam-forming. We define $\beta \in \mathcal{N} \to \zeta$ as the function that describes which base stations in $\zeta$ will be picked to transmit the data signals. Given BS $j \in \mathcal{N}$, the corresponding label for the BS in $\zeta$ is given by $\beta(j)$. Given $\omega(\mathbf{s})$ active base stations that perform training and receive feedback on $\tilde{\mathbf{H}}$, we define the effective channel matrix as $\mathbf{H}(\tilde{\mathbf{H}}, \beta)$, where its elements $h_{j,k} = \tilde{h}_{\beta(j),k}$. For zero-forcing, we require that the number of transmitting antennas is equal to the number of receiving antennas or $K = N$. With this constraint, if $\mathbf{H}$ is an invertible matrix, we define:

$$\bar{\mathbf{x}} = (\mathbf{H}(\tilde{\mathbf{H}}, \beta))^{-1}\mathbf{u} \tag{3}$$

where $\mathbf{u}$ is a vector of length $K$, where $u_k$ which is the signal received by the $k$-th user. In this work, we take $\|u_k\|^2 = 1$ so that each user obtains identical signal strengths. This constraint has several benefits:

1) This results in a very "fair" beam forming scheme as all users experience equal signal strength and thus similar data rates.
2) As the base station antennas are spread over the cell, there is no user on the "cell edge" or "cell center". In this situation, equal SINRs for all users can result in less congestion when user traffic patterns are taken into account
3) Finally, the resulting system is far less complex and easier to analyze than one with arbitary values for $\|u_k\|^2$.

With these definitions, we can now precode the transmitted signal as:

$$\mathbf{x} = \frac{\bar{\mathbf{x}}}{\alpha(\tilde{\mathbf{H}}, \beta)} \tag{4}$$

where $\alpha(\tilde{\mathbf{H}}, \beta) = \max(\bar{x}_m)$. This pre-coding works only if all the $P_j$ are equal, and so we chose $P_j = P_0$. From this point onwards, $P_0$ represents the transmit power level of the system with this pre-coding. The signal received by each of the $K$ users is given by:

$$y_k = \sqrt{\frac{P_0}{\sigma^2}}\frac{u_k}{\alpha(\tilde{\mathbf{H}}, \beta)} + z \tag{5}$$

Thus the SINR at each user is now given by

$$\gamma_k = \frac{P_0}{\|\alpha(\tilde{\mathbf{H}}, \beta)\|^2\sigma^2} \tag{6}$$

## III. Energy efficiency optimization

This work aims at minimizing the energy consumption by base stations. If each user in the network is connected to download some data, then the total energy consumed by the network is the total power consumed multiplied by the total duration for which the user stays connected. Energy efficiency (EE) is a metric that is often used to measure this, and maximizing the energy efficiency leads to minimizing the total energy consumed.

### A. Defining the EE metric

Before defining the EE, we first calculate the total power consumption of the network. From (1) and (4), the total power consumed is given by:

$$P_{tot}(P_0, \tilde{\mathbf{H}}, \beta) = \sum_{m=1}^{M} cs_m + (1 - s_m) \times \left( b + P_0 \left\| \frac{(\mathbf{H}^{-1}(\tilde{\mathbf{H}}, \beta)\mathbf{u})_{\beta^{-1}(m)}}{\alpha(\tilde{\mathbf{H}}, \beta)} \right\|^2 \right) \tag{7}$$

Here we define

$$\forall m \in \mathcal{M}; \beta^{-1}(m) = \begin{cases} j & \text{if } j \in \mathcal{N} \text{ exists } s.t\, \beta(j) = m \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

and $()_j$ is the $j-th$ element if $j \neq 0$ and is 0 if $j = 0$. In this scenario, we define the instantaneous energy efficiency as:

$$\eta(P_0, \tilde{\mathbf{H}}, \beta) = \frac{\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}, \beta))}{P_{tot}(P_0, \tilde{\mathbf{H}}, \beta)} \tag{9}$$

where $f()$ gives the effective throughput as a function of the SINR. $f(\gamma_k) = \log(1 + \gamma_k)$ for example. However when we study the base station energy efficiency for a longer duration, the effects of fast fading in $\tilde{\mathbf{H}}$ gets averaged and in this case a more reasonable definition for the EE is:

$$\bar{\eta}(P_0, \mathbf{G}, \beta) = \frac{\mathbb{E}_{\bar{\mathbf{H}}}[\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}}), \beta))]}{\mathbb{E}_{\bar{\mathbf{H}}}[P_{tot}(P_0, \tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}}), \beta)]} \tag{10}$$

### B. Optimization w.r.t the transmit power

In this section, we show some properties of our energy efficiency metric w.r.t $P_0$. If the goal of a system is to be energy efficient using power control, then one important question arises: Is there a unique power for which the energy efficiency is maximized ? We answer this question with the following proposition:

*Proposition 1:* Given a certain path loss matrix $\mathbf{G}$ and a selection of transmitting base stations $\beta$ in the virtual MIMO system, the average EE $\bar{\eta}(P_0, \mathbf{G}, \beta)$ is maximized for a unique $P_0^*$ and is quasi-concave in $P_0$.

**Proof:** Consider the SINR for each user $\gamma_k$. It can be observed that $\frac{\partial \gamma_k(P_0, \tilde{\mathbf{H}}, \beta))}{\partial P_0}$ is a constant. So if $f()$ is concave in $\gamma_k$, it is also concave in $P_0$. Now consider the numerator of (10), $\mathbb{E}_{\bar{\mathbf{H}}}[\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}(\mathbf{G}, \bar{\mathbf{H}}), \beta))]$. This is a weighted sum of several concave functions and is hence also concave itself. Similarly, $\frac{\partial P_{tot}(P_0, \tilde{\mathbf{H}}, \beta)}{\partial P_0}$ can also be verified to be a constant. Hence, $\partial \frac{\mathbb{E}_{\bar{\mathbf{H}}}[P_{tot}(P_0, \tilde{\mathbf{H}}, \beta)]}{\partial P_0}$ is a constant. Thus $\bar{\eta}(P_0, \mathbf{G}, \beta)$ is the ratio of a concave function of $P_0$ to a linear function of $P_0$. This is known to be quasi-concave and has a unique maximum $P_0^*$ satisfying:

$$\frac{\partial \bar{\eta}(P_0^*, \mathbf{G}, \beta)}{\partial P_0} = 0 \tag{11}$$

■

This proposition guarantees that given a certain choice of transmit antennas and a path fading matrix, the energy efficiency can always be optimized w.r.t the transmit power level $P_0$.

### C. Optimizing the selection of transmitting base stations

Given a certain sleep mode state $\mathbf{s}$, there are $\omega(\mathbf{s})$ base stations active that train and obtain feedback. From this set $\zeta$, $K$ base stations have to be picked for zero-forcing. This choice is mathematically expressed by $\beta$. The $\beta$ that optimizes the energy efficiency depends on the channel state $\tilde{\mathbf{H}}$. The following proposition details the method of choosing the $\beta$ that optimizes EE.

*Proposition 2:* When $\frac{P_0}{b} \rightarrow 0$, the $\beta^*$ that maximizes $\bar{\eta}(P_0, \mathbf{G}, \beta)$ is obtained by:

$$\beta^* = \arg\min[\alpha(\tilde{\mathbf{H}}, \beta);\ \beta \in \{\mathcal{N} \rightarrow \mathcal{K}\}] \tag{12}$$

**Proof:** By observing (6) it can be seen that $\gamma_k(P_0, \tilde{\mathbf{H}}, \beta)$ is maximized by picking $\beta^*$. And so $\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}, \beta))$ is maximized when $\beta = \beta^*$. When $\frac{P_0}{b} \rightarrow 0$, for $\beta^*$ and any $\beta_1$ we have:

$$\lim_{\frac{P_0}{b} \rightarrow 0} \bar{\eta}(P_0, \mathbf{G}, \beta^*) - \bar{\eta}(P_0, \mathbf{G}, \beta_1) = \tag{13}$$

$$\frac{\mathbb{E}_{\bar{\mathbf{H}}}[\sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}, \beta^*)) - \sum_k f(\gamma_k(P_0, \tilde{\mathbf{H}}, \beta_1))]}{\sum_{m=1}^{M} cs_m + (1 - s_m)b} \geq 0 \tag{14}$$

■

This implies that we pick $\beta$ such that $\alpha(\tilde{\mathbf{H}}, \beta)$ is minimized for optimizing EE when $b >> P_0$. From a practical point of view, the above result is useful as it proposes an algorithm to pick the right base stations based on the CSI obtained from all the base stations that are active. The condition $b >> P_0$ is most applicable when the users are close to the base stations resulting in a low $P_0$ being used for maximizing EE.

## IV. Numerical results

In this section we use simulations and numerical calculations to study the effectiveness of our proposal as well as the advantages offered. For the purpose of a thorough numerical study, we will analyze two kinds of system settings A and B. For both the configurations the common parameters are:

1) $c = \frac{b}{10}$ W
2) $P_{\max} = 2$ W
3) $f(\gamma) = B \log(1 + \gamma)$
4) $\sigma^2 = 1$ mW

Where $B = 10^6$ hz is the bandwidth.

The fast fading co-efficient we consider is $\bar{h}_{i,j} = o(\pi_{m,k})\Omega + 0.1\xi$. Where $\xi \in CN(0, 1)$, a is the direct line of sight factor which plays a dominant role in most small cell networks, $o_{\pi_{m,k}} \in 0, 1$ is the shadow factor and $o(pi_{m,k}) = 1$ with probability $\pi_{m,k}$. Here $\pi_{m,k}$ is the probability that the receiver $k$ has line of sight with the BS antenna $m$. We take $\pi_{m,k} = 0.5\, \forall (k, m)$ for our simulations.

The presented results study the case of two users $K = 2$ served by a small cell cluster of three base stations, i.e $M = 3$. In addition to zero-forcing, when there are two users a single base station could also alternately use Orthogonal Frequency-Division Multiple Access (OFDMA) to serve the two users and keep the other two BS in sleep mode (i.e . Our numerical simulations study all these possible scenarios and compare their respective EE performances.

In both of the settings presented below, we study two main regimes of interest:

1) $b = 1$W : This regime represents the futuristic case where power amplifier efficiencies are quite high and the constant power consumed is lower than the maximum RF output power.
2) $b = 10$W : This regime represents the more current state of the art w.r.t power amplifier efficiency where in small cell antennas, a large portion of the power is lost as a fixed cost.

We also consider two possible values of $\Omega$, the line of sight factor. The case $\Omega = 10$ is representative of pico-cells that are deployed externally, whereas the case $\Omega = 0$ represents femto-cells deployed internally and no line of sight communication is possible.

### A. Setting A

The deployment of antennas and the user locations are shown in Fig 2. In this setting we take $g_{m,k} = 1\,\forall m, k$.

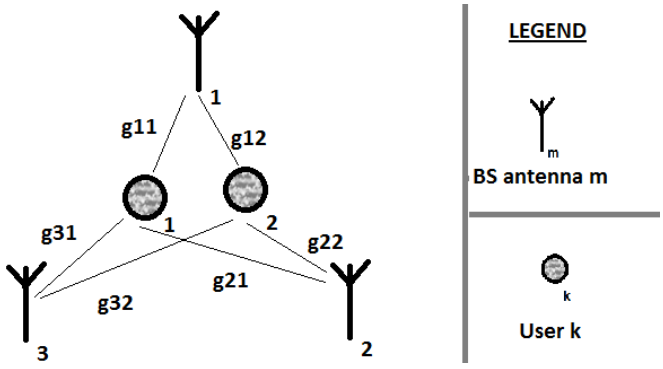


Fig. 2. Setting A schematic

In Fig 3 we study the EE of a VMIMO system with a very efficient power amplifier. In this figure, we notice that using all available base station antennas is more efficient when line of sight communications are possible. In this case, no BS is in sleep mode and all the antennas train and obtain feedback on their channels. The choice of $\beta$ is very much relevant in this scenario. However in the regime where there is no direct line of sight (users are inside buildings), it becomes more efficient to just use 2 BS antennas and put one on sleep mode. As the configuration is symmetric, the choice of the base station in sleep mode is not relevant.

In Fig 4 we study the EE of a VMIMO system with an inefficient power amplifier. In this figure, we notice that using

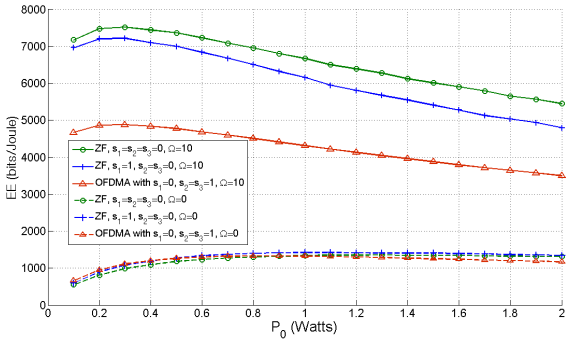


Fig. 3. Setting A: EE v.s $P_0$ for $b = 1$ W

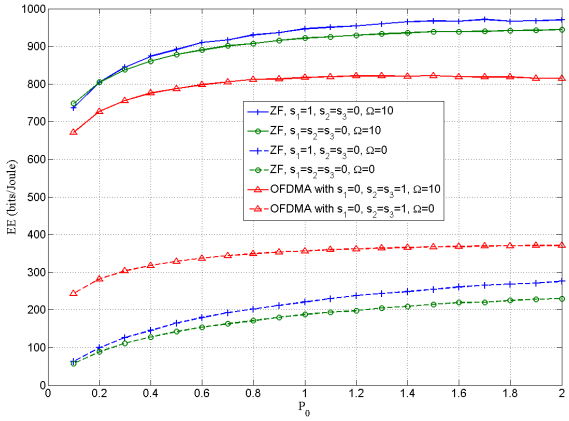


Fig. 4. Setting A: EE v.s $P_0$ for $b = 10$ W

all available base station antennas is not efficient even when line of sight communications are possible. In this case, having one BS in sleep mode and obtaining a $2\times2$ virtual MIMO with the other remaining antennas is the most efficient. Suprisingly, in the case of no line of sight, i.e $\Omega = 0$, we observe that using OFDMA with one BS active is the most efficient solution. This is explained by the relative inefficiency of zero-forcing in the low SNR regime, causing less energy to be spent by having two BS antennas in sleep mode and just one antenna transmitting for the two users in orthogonal frequencies.

### B. Setting B

The deployement of antennas and the user locations are shown in Fig 5. In this setting we take $g_{1,1} = g_{2,1} = 4$, $g_{3,1} = g_{1,2} = g_{2,2} = 0.1$ and $g_{3,2} = 10$.

In Fig 6, similarly to what was done in the previous setting, we study the EE of a VMIMO system with a very efficient power amplifier. In this figure, for both $\Omega = 0$ and $\Omega = 1$ we see that having to use 2 BS antennas and put one on sleep mode is the most efficient. In this setting, the configuration of BS and users are asymmetric and the BS to be put in sleep mode has to be chosen carefully. BS 1 and 2 are symmetric and are close to user 1, but 3 is closer to user 2. In this case

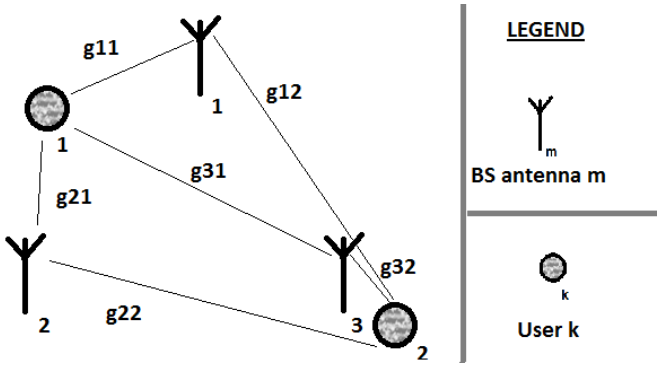


Fig. 5. Setting B schematic

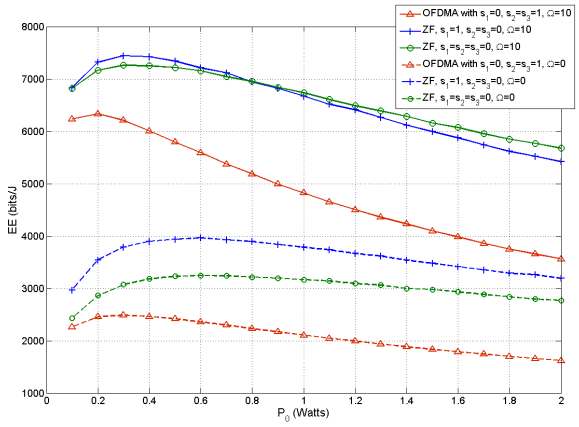


Fig. 6. Setting B: EE v.s $P_0$ for $b = 1$ W

choosing $s_1 = 1$ or $s_2 = 1$ is efficient, but $s_3 = 1$ is highly inefficient.

In this setting, we see from Fig 7 that unlike in Setting A, using OFDMA to divide resources between the two users is not as efficient as ZF due to the higher SNR when served by nearby BS antennas. Like in Fig 6, choosing $s_1 = 1$ or $s_2 = 1$ and zero-forcing is always the most efficient solution.

## V. Conclusion

This paper studies the performance of virtual MIMO systems from an energy efficiency perspective. It defines an energy efficiency metric that takes into account the capacity as well as the energy consumption, and considers both fixed and variable parts of this latter. We optimize the power allocations of the different antennas and show that sleep mode can bring a significant energy efficiency gain. This involves deactivating antennas that do not have a significant contribution to the system capacity, for a given number of users and radio channel conditions.

This work is applicable only for the specific case of a small cell cluster with a centralized network and CSIT. Thus, many extensions of the proposed work are possible. The most relevant extension is to apply the proposed framework

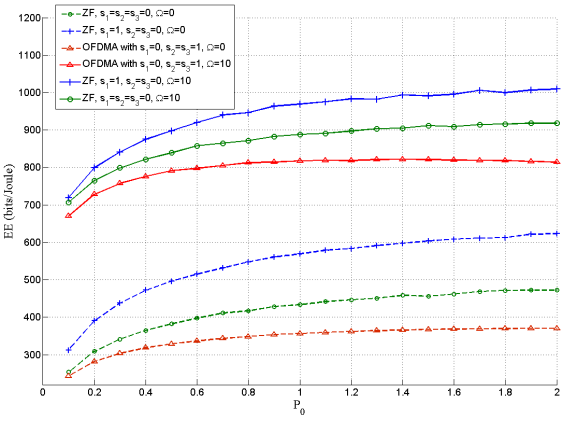


Fig. 7. Setting B: EE v.s $P_0$ for $b = 10$ W

taking into account user traffic and a random number of users. Another natural extension of the proposed framework is of course, to study the effect of different classes of mobility on the virtual MIMO scheme and to study a distributed network.

## VI. Acknowledgments

This work is a joint collaboration between Orange Labs, Laboratoire des signaux et systémes (L2S) of Supélec and the Alcatel Lucent Chair of Supélec. This work is part of the European Celtic project “Operanet2”.

## References

[1] L. Saker, S-E. Elayoubi and H-O. Scheck, *System selection and sleep mode for energy saving in cooperative 2G/3G networks*, IEEE VTC 2009-Fall.

[2] S-E. Elayoubi and O. Ben Haddada, *Uplink intercell interference and capacity in 3G LTE systems*, IEEE ICON 2007.

[3] L. Saker, S-E. Elayoubi, R. Combes and T. Chahed, *Optimal control of wake up mechanisms of femtocells in heterogeneous networks*, Selected Areas in Communications, IEEE Journal on 30 (3), 664-672.

[4] C.X Wang, X. Hong, X. Ge and X. Cheng, *Cooperative MIMO channel models: A survey*, IEEE communications magazine, Vol: 48, issue 2, 2010.

[5] D. J. Goodman, and N. Mandayam, *Power Control for Wireless Data*, IEEE Personal Communications, vol. 7, pp. 48-54, Apr. 2000.

[6] C. Desset et al. , *Flexible power modeling of LTE base stations*, IEEE Wireless Comm. and Networking Conference: Mobile and Wireless Networks, 2012.

[7] V.S Varma, S. Lasaulce, M. Debbah and S-E. Elayoubi, ”An Energy Efficient Framework for the Analysis of MIMO Slow Fading Channels”, IEEE Trans. Signal Proc., Vol 61, 10, pp: 2647-2659, 2013.

# Appendix B

# Papers on Cross Layer

## B.1 ICC-2012

• V.S Varma, S. Lasaulce, Y. Hayel, S. Eddine Elayoubi and Merouane Debbah, "Cross-Layer Design for Green Power Control",IEEE Int. Conference on Communications (ICC 2012).

# Cross-Layer Design for Green Power Control

Vineeth S Varma[1,2], Samson Lasaulce[2], Yezekael Hayel[3], Salah Eddine Elayoubi[1] and Merouane Debbah[4]

[1]Orange Labs
92130 Issy Les Moulineaux
France
salaheddine.elayoubi@orange.com

[2]LSS, SUPELEC
91192 Gif sur Yvette
France
Vineeth.varma@lss.supelec.fr
Samson.lasaulce@lss.supelec.fr

[3]University of Avignon,
84911 Avignon
France
yezekael.hayel@univ-avignon.fr

[4]Alcatel Lucent Chair
SUPELEC
91192 Gif sur Yvette
France
Merouane.debbah@supelec.fr

***Abstract*—In this work, we propose a new energy efficiency metric which allows one to optimize the performance of a wireless system through a novel power control mechanism. The proposed metric possesses two important features. First, it considers the whole power of the terminal and not just the radiated power. Second, it can account for the limited buffer memory of transmitters which store arriving packets as a queue and transmit them with a success rate that is determined by the transmit power and channel conditions. Remarkably, this metric is shown to have attractive properties such as quasi-concavity with respect to the transmit power and a unique maximum, allowing to derive an optimal power control scheme. Based on analytical and numerical results, the influence of the packet arrival rate, the size of the queue, and the constraints in terms of quality of service are studied. Simulations show that the proposed cross-layer approach of power control may lead to significant gains in terms of transmit power compared to a physical layer approach of green communications.**

## I. Introduction

For a long time, the problem of energy mainly concerned autonomous, embarked, or mobile communication terminals. Nowadays, with the existence of large networks involving both fixed and nomadic terminals, the energy consumed by the fixed infrastructure has also become a central issue for communications engineers [1]. The present work precisely falls into this framework. More specifically, our goal is to provide insights to researchers and engineers on how to devise power control schemes in green wireless networks. Among pioneering works on energy-efficient power control we find the paper by Goodman et al [4]. Therein, the authors define the energy-efficiency of a communication as the ratio of the net data rate (called goodput) to the radiated power; the corresponding quantity is a measure of the average number of bits successfully received per joule consumed at the transmitter. This metric has been used in many works. For example, in [16] it is applied to the problem of distributed power allocation in multi-carrier CDMA (code division multiple access systems) systems, in [3] it is used to model the users delay requirements in energy-efficient systems. In [5] it is re-interpreted as a capacity per unit cost measure in MIMO (multiple input multiple output) systems and in [6], it is used for subcarrier assignment in distributed OFDMA (orthogonal frequency division multiple access) multicellular networks. Although fully relevant, the performance metric introduced in [4] has left several issues unexplored which has motivated the work reported here.

First, in the definition of energy-efficiency given in works like [11],[4] or [5], the transmission cost corresponds to the radiated power that is, the power of the radio-frequency signals; this is very useful in situations where electromagnetic pollution has to be cut down. However, more generally, the consumption of the whole device matters (e.g., because of the power amplifier consumption). Second, in [4], the packets are lost due to bad channel conditions while, here, we propose to take into account the losses induced by the finite size of the queue at the transmitter (which can model limited memory or a certain delay constraint). Third, in [4] and related references, energy-efficiency can be maximized while having a bad quality of service (QoS) e.g., in terms of packet loss or equivalently in terms of goodput. In this paper, we show that these three issues can be, in fact, dealt with quite easily. The analysis is, however, more complicated than some analysis like the one conducted in [3] where the delay constraint is translated into a constraint on the minimum signal-to-noise ratio (SNR). This is due to a double effect, resulting from integrating a queueing model (justifying the term "cross-layer design") and considering the whole terminal power (instead of the radiated power). The queuing model is used in the spirit of [18] where a queuing model is used to reach a certain QoS in CDMA systems with multiple classes of traffic, but without energy considerations. Another cross-layer queuing model has been proposed in [19] but considering Shannon capacity under power constraints and not energy-efficiency.

At this point, it is important to note that this work focuses on point-to-point communications, which may be surprising since power control is the problem of interest. There are at least two strong reasons for this choice. First, the single-user case captures the main effects we want to emphasize and allows us to describe the proposed approach in a clear manner. Second, as advocated by the existing works on power control (see e.g., [4] and related works), once the single-user case is fixed, the multiuser case is tractable provided some conditions are met. One of these conditions is that the performance metric has to possess some desirable properties (quasi-concavity, that is shown to be verified for the proposed metric, is one of them) and reasonably complex multiuser channel models are considered (the multiple access channel is one of them).

This paper is structured in the following format. In Section II, we present the system model and define the proposed performance metric. In Section III, we conduct an analytical study of the performance metric while Section IV provides many numerical results to sustain the proposed approach. Finally, we conclude the paper and some possible extensions

to this work are provided.

## II. System model

We consider a buffer of size $K$ at the transmitter. The packets arrival follows a Bernoulli process with probability $q$, i.e., at each time slot $t$ (time is slotted) a new packet arrives in the queue with probability $q$ (this corresponds to classical ON/OFF sources). All packets are assumed to be of the same size $S$. The throughput on the radio interface equals to $R$ (bit/s) and depends on several parameters such as the modulation and coding scheme. We consider that the transmitter is always active, meaning that it always transmits its packet while the buffer is not empty. Each packet transmitted on the channel is received without any errors with a probability which depends on the quality of the channel and transmission power. We denote the transmission power by $p$ and we have $f(p)$ as the success probability of transmission of the packet on the channel. $f(p)$ is just assumed to be a sigmoidal function in our derivations, in practice, good approximations for the success rate function are $f(p) = \exp(-(2^{\frac{R}{R_0}}-1)\frac{\sigma^2}{p})$, for an unknown channel, and $f(p) = Q(K\frac{R}{R_0} - K\log[1+\frac{hh^*p}{\sigma^2}])$, with $Q$ being the "Q" function and $K$ a constant, for a known channel. The channel fading due to path loss is not treated separately but is integrated into the average power of noise $\sigma^2$. The success probability depends in fact on the $\mathrm{SNR} = \frac{p}{\sigma^2}$. However, based on the block fading channel assumption, we make a slight abuse of notations by using the notation $f(p)$ instead of $f(\mathrm{SNR})$. In some places in this paper, we even remove the variable $p$ for the sake of clarity and use the notation $f$. We denote by $Q_t$ the size of the queue at the transmitter at time slot $t$. The size of the queue $Q_t$ is a Markov process on the state space $Q = \{0, \ldots, K\}$. We have the following transition probabilities $\forall i, j \in Q$, $P_{i,j} := I\!P(Q(t+1) = i | Q(t) = j)$ given by:

- $P_{0,0} = 1 - q + qf$,
- $P_{K,K} = (1-q)(1-f) + q$,
- for any state $i \in \{0, \ldots, K-1\}$, $P_{i,i+1} = q(1-f)$,
- for any state $i \in \{1, \ldots, K\}$, $P_{i,i-1} = (1-q)f$,
- for any state $i \in \{1, \ldots, K-1\}$, $P_{i,i} = (1-q)(1-f) + fq$.

A new packet is lost if the queue is full when it comes in and the transmission of the packet currently on the radio interface failed on the same time slot. Indeed, we consider that a packet is in service (occupying the radio interface) until it is transmitted successfully. Thus, a packet in service blocks the queue during $\frac{1}{f(p)}$ time slots on the average. We assume that an arrival of a packet in the queue and a departure (successful transmission) at the same time slot can occur.

Given the transition probabilities above, the stationary probability of each state is given by (see e.g., [13]):

$$\forall s \in S, \quad \Pi_s = \frac{\rho^s}{1 + \rho + \ldots + \rho^K}, \tag{1}$$

with

$$\rho = \frac{q(1-f)}{(1-q)f}. \tag{2}$$

When a packet arrives and finds the buffer full (meaning that the packet currently on the radio interface is not transmitted successfully), it is blocked and this event is considered as a packet loss. The queue is full in the stationary regime with probability $\Pi_K$ :

$$\Pi_K = \frac{\rho^K}{1 + \rho + \ldots + \rho^K} = \frac{\rho^K(\rho - 1)}{\rho^{K+1} - 1}. \tag{3}$$

### A. Proposed performance metric

In order to evaluate the performance of this system, we first determine the expression for the packet loss probability. A packet is lost (blocked) only if a new packet arrives when the queue is full and, on the same time slot, transmission of the packet on the radio interface failed. Note that these two events are independent because the event of "transmit or not" for the current packet on the radio interface, does not impact the current size of the queue, but only the one for the next time slot. This amounts to considering that a packet coming at time slot $t$, is rejected at the end of time slot $t$, the packet of the radio interface having not been successfully transmitted. We consider the stationary regime of the queue and then, the fraction of lost packets, $\Phi$, can be expressed as follows:

$$\Phi(p) = [1 - f(p)]\Pi_K(p). \tag{4}$$

Thus the average data transmission rate is $q[1 - \Phi(p)]R$. Now, let us consider the cost of transmitting. For each packet successfully transmitted, there have been $\frac{1}{f(p)}$ attempts on an average [4]. For each time slot, irrespective of whether transmissions occur, we assume that the transmitter consumes energy. A simple model which allows one to relate the radiated power to the total device consumed power is provided in [14] is given by $P_{\mathrm{device}} = ap + b$, where $a \geq 0, b \geq 0$ are some parameters; $b$ precisely represents the consumed power when the transmit power is zero. The average power consumption is in our case $b + \frac{pq(1-\Phi)}{f(p)}$ (we assume without loss of generality that $a = 1$). We are now able to define the energy-efficiency metric $\eta(p)$ as the ratio between the average net data transmission rate and the average power consumption, which gives:

$$\eta(p) = \frac{q[1 - \Phi(p)]R}{b + \frac{pq[1-\Phi(p)]}{f(p)}}. \tag{5}$$

The above expression shows that the cross-layer design approach of power control is fully relevant when the transmitter has a cost which is independent of the radiated power; otherwise (when $b = 0$), one falls into the original framework of [4].

### B. Constraints on QoS and maximum power

As already mentioned in Section I, of the recurrent problems with most works using the performance metric introduced in [4] is that energy-efficiency can be maximized at a power level which does not guarantee a minimum QoS. This is why we also consider a constraint when maximizing (5): we assume $\Pi_K[1 - f(p)]$ has to be less than $\epsilon$ where $\epsilon$ is the upper bound on the packet loss. For example, in cellular systems, typical

values for $\epsilon$ are $0.1$ or $0.01$, based on the system requirements. Adding this constraint restricts the range of power usable by the transmitter by adding a lower bound on the power. This lower bound depends on the entry probability $q$ and on the size of the queue $K$. An upper bound on the usable transmit power $P_{\max}$ can also be added to model realistic situations when there is a limitation on the maximum power that can be utilized.

## III. Analytical results

Having defined the energy efficiency function, we will now examine its properties.

### A. The impact of the packet entry probability

First let us study the special case when $q = 1$: Here we have, $\lim_{q\to 1} \Pi_K = 1$, then $\Phi = 1 - f(p)$ and we have a simplified expression of the energy efficiency function $\eta = \frac{Rf(p)}{b+p}$. This energy efficiency is a more general form of the metric introduced in [4]. This particular case is in fact identical to the situation when the system is modeled with a purely physical layer approach. As the queue is always full, transmission always takes place and so the energy efficiency of the entire system is just the energy efficiency of transmission.

The next part of this section examines $\eta$ as $q$ decreases. Logically, as $q$ decreases, the average duration when the buffer is empty increases causing a wasted consumption of the fixed power during which no data is transmitted. Here, we have a proposition that formulates and proves this reasoning mathematically.

*Proposition 1:* The energy efficiency function is an increasing function of $q$, i.e., $\frac{\mathrm{d}\eta}{\mathrm{d}q} > 0$.

**Proof :** $\eta = \frac{1}{\frac{b}{(1-\Phi)q} + \frac{p}{f(p)}}$. If $\frac{\mathrm{d}\Phi}{\mathrm{d}q} < \frac{1-\Phi}{q}$, then $\mathrm{d}\frac{(1-\Phi)q}{\mathrm{d}q} > 0$ and from this, it follows that $\frac{\mathrm{d}\eta}{\mathrm{d}q} > 0$.

To prove this, we first calculate $\frac{\mathrm{d}\rho}{\mathrm{d}q} = \frac{1-f(p)}{f(p)} \frac{-1}{(1-q)^2}$. Now let us consider $\frac{\mathrm{d}\Phi}{\mathrm{d}q} = -\Phi^2 \frac{\mathrm{d}(\Phi^{-1})}{\mathrm{d}q}$. The term $\Phi^{-1} = 1 + \frac{1}{\rho} + ..+ \frac{1}{\rho^K}$ and so we have $\frac{\mathrm{d}\Phi^{-1}}{\mathrm{d}q} = (\frac{1}{\rho^2} + .. + \frac{K}{\rho^{K+1}}) \frac{1-f(p)}{f(p)} \frac{-1}{(1-q)^2}$. Simplifying and using inequalities we have $\frac{\mathrm{d}\Phi}{\mathrm{d}q} \leq \frac{1-\Phi}{q}$. ◇

### B. The limiting case of infinite queue size

Consider the extreme case of an infinite queue, i.e., $K \to \infty$.

- Case 1: $f(p) < q$; i.e., $\rho > 1$. We have that $\lim_{K\to\infty} \Pi_K = \frac{\rho-1}{\rho}$ and a simplification yields $\Phi = \frac{1-f(p)}{q}$. Thus the energy efficiency becomes $\eta = \frac{Rf(p)}{b+p}$. These expressions make sense as in the steady state, due to a higher probability of entrance than exit, the queue size blows up and there are always packets to transmit.
- Case 2: $f(p) \geq q$; i.e., $\rho \leq 1$. If $f(p) = q$, then $\Pi_K = \frac{1}{K}$ and $\lim_{K\to\infty} \Pi_K = 0$. For $f(p) > q$, we have also that $\lim_{K\to\infty} \Pi_K = 0$ and then simplification yields $\Phi = 0$. Thus the energy efficiency becomes $\eta = \frac{R}{\frac{b}{q} + \frac{p}{f(p)}}$. These expressions also make sense as in the steady state due to a higher probability of exit, the buffer is never full and there is no packet loss.

### C. Optimizing the energy efficiency

In this paragraph, we prove that there exists a unique power where the energy efficiency function is maximized when the transmission rate is a sigmoidal or "S"-shaped function of $p$. In [11], it was shown that having a sigmoidal success rate $f(p)$ implies quasi-concavity and a unique maximum for $\frac{f(p)}{p}$. This assumption was shown to be highly relevant from a practical viewpoint in [4] as well as from an information theoretical viewpoint in [5].

*Theorem 1:* The energy efficiency function $\eta$ is quasi-concave with respective to $p$ and has a unique maximum denoted by $\eta(p^*)$ if the efficiency function $f(p)$ has a sigmoidal shape.

**Proof :** Consider the asymptotic cases when $p \to 0$ and $p \to \infty$, we have the limiting cases as $\lim_{p\to 0} f(p) = 0$ and $\lim_{p\to\infty} f(p) = 1$ respectively.

• When $p \to 0$: We have $\lim_{p\to 0} \Phi = 1$ trivially and $\lim_{p\to 0} \eta = 0$.

• When $p \to \infty$: $\lim_{p\to\infty} \Phi = 0$ and $\lim_{p\to\infty} \eta = 0$.

Thus from the extension of the mean value theorem proposed in [17], we have $\frac{\mathrm{d}\eta}{\mathrm{d}p} = 0$ for at least one $p$.

Consider the function $\frac{1}{\eta} = A(p) + B(p)$, where $A(p) = \frac{p}{f(p)R}$ and $B(p) = \frac{b}{Rq(1-\Phi)}$. From the earlier work in [11], we have that $A(p)$ is convex and that $\frac{1}{A(p)}$ is quasi-convex and has a unique maximum at $p_0^*$. $\frac{\mathrm{d}f(p)}{\mathrm{d}p} > 0$ for all $p$. We also know that for $p > p_0^*$, and $\frac{\mathrm{d}^2 f(p)}{\mathrm{d}p^2} < 0$.

Now let us study the derivatives of the function $B(p)$.

$$\frac{\mathrm{d}B(p)}{\mathrm{d}p} = \frac{b}{Rq(1-\Phi)^2} \frac{\mathrm{d}\Phi}{\mathrm{d}p} \tag{6}$$

and we have

$$\frac{\mathrm{d}^2 B(p)}{\mathrm{d}p^2} = \frac{b}{Rq(1-\Phi)^2} (\frac{\mathrm{d}^2\Phi}{\mathrm{d}p^2} + \left(\frac{\mathrm{d}\Phi}{\mathrm{d}p}\right)^2 \frac{2}{1-\Phi}\Big). \tag{7}$$

If $B(p)$ is a monotonically decreasing function and is convex for $p \geq p_0^*$, then we have $\frac{1}{A(p)+B(p)}$ to be quasi-concave [15]. So in the following section of this proof we will show that $\frac{dB(p)}{dp} < 0$ and $\frac{\mathrm{d}^2 B(p)}{\mathrm{d}p^2} > 0$.

For $\frac{\mathrm{d}B(p)}{\mathrm{d}p} < 0$, by examining equation 6, we see that showing $\frac{\mathrm{d}\Phi}{\mathrm{d}p} < 0$ is sufficient as the other terms are always positive.

Similarly, for $\frac{\mathrm{d}^2 B(p)}{\mathrm{d}p^2} > 0$, by examining equation 7, we see that showing $\frac{\mathrm{d}^2\Phi}{\mathrm{d}p^2} > 0$ is sufficient as the other terms are also always positive.

$$\frac{\mathrm{d}\Phi}{\mathrm{d}p} = (1 - f(p)) \frac{\mathrm{d}\Pi_K}{\mathrm{d}p} - \Pi_K \frac{\mathrm{d}f(p)}{\mathrm{d}p}. \tag{8}$$

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}p^2} = (1 - f(p)) \frac{\mathrm{d}^2\Pi_K}{\mathrm{d}p^2} - 2\frac{\mathrm{d}\Pi_K}{\mathrm{d}p} \frac{\mathrm{d}f(p)}{\mathrm{d}p} - \Pi_K \frac{\mathrm{d}^2 f(p)}{\mathrm{d}p^2}. \tag{9}$$

For $\frac{\mathrm{d}\Phi}{\mathrm{d}p} < 0$, by examining equation 8, we see that showing $\frac{\mathrm{d}\Pi_K}{\mathrm{d}p} < 0$ is sufficient.

We have $\rho = \frac{q}{1-q}\frac{1-f(p)}{f(p)}$ and $\frac{\mathrm{d}\rho}{\mathrm{d}p} = \frac{-q}{(1-q)f(p)^2}\frac{\mathrm{d}f(p)}{\mathrm{d}p}$ which is negative. Express $\frac{1}{\Pi_K} = 1 + \frac{1}{\rho} + ..\frac{1}{\rho^K}$. Differentiating with respect to $p$, we have

$$\frac{\mathrm{d}\Pi_K}{\mathrm{d}p} = \Pi_K^2 \frac{\mathrm{d}\rho}{\mathrm{d}p}(\frac{1}{\rho^2} + ..\frac{K}{\rho^{K+1}}). \tag{10}$$

Similarly, for $\frac{\mathrm{d}^2\Phi}{\mathrm{d}p^2} > 0$, by examining equation 9, we see that showing $\frac{\mathrm{d}^2\Pi_K}{\mathrm{d}p^2} > 0$ is sufficient as $\frac{\mathrm{d}^2 f(p)}{\mathrm{d}p^2} < 0$ for $p > p_0^*$ and $\frac{\mathrm{d}\Pi_K}{\mathrm{d}p} < 0$. And from equation 10 we observe that as $p$ increases $\frac{\mathrm{d}\Pi_K}{\mathrm{d}p}$ increases. Thus following the argument from the start, we have $\eta(p)$ to be quasi-concave.
Since there exists some power $p$ for which $\eta(p)$ is maximized, we have proved that there exists a unique $p^*$ for which the energy efficiency is optimized. ◇

We are then able to determine the optimal power $p^*$ which maximize the energy efficiency function, by solving the following equation:

$$0 = \frac{-\mathrm{d}\Phi}{\mathrm{d}p}\{b + \frac{pq(1-\Phi)}{f(p)}\} + (1-\Phi)\{\frac{\mathrm{d}\Phi}{\mathrm{d}p}\frac{p}{f(p)} + \frac{\mathrm{d}(p/f(p))}{\mathrm{d}p}\}. \tag{11}$$

### *D. Behavior of $p^*$ with respect to $q$*

When $q \to 0$, the optimization problem is reduced to finding $p$ that maximizes $\frac{f(p)}{p}$. This can be obtained by calculating the derivative of $p^*$ from equation 11 and applying the limit on $q$. Indeed we have that $\lim_{q\to 0}\frac{\mathrm{d}\Phi}{\mathrm{d}p} = 0$ and then equation (11) is reducing to $\frac{\mathrm{d}(p/f(p))}{\mathrm{d}p} = 0$.

When $q \to 1$, the optimization problem is reduced to finding $p$ that maximizes $\frac{f(p)}{p+b}$. This can be obtained by simply using same ideas as previously.

### *E. Power Control with the QoS and maximum power constraint*

The QoS constraint requires that $\Phi \leq \epsilon$ and then we have to find the new properties of the energy efficiency function satisfying this constraint. We define $p_0 := \min(p|\Phi(p) \leq \epsilon)$.

*Proposition 2:* For all $p > p_0$, the constraint is satisfied, i.e., $\Phi(p > p_0) \leq \epsilon$.

**Proof :** This is quite easy to see because from our earlier proof we have that $\frac{\mathrm{d}\Phi}{\mathrm{d}p} < 0$ and so $\Phi(p) \leq \Phi(p_0) \leq \epsilon$ for all $p > p_0$. ◇

Additionally we also have another proposition which gives the properties of $p_0$ when the arrival probability, $q$, changes:

*Proposition 3:* If $q_2 > q_1$, then $p_0(q_2) \geq p_0(q_1)$.

**Proof :** This result can also be easily proved. From our earlier proof we have $\frac{\mathrm{d}\Phi}{\mathrm{d}q} > 0$ and so with the power $p_0(q_2)$, we have $\Phi(q_1) \leq \Phi(q_2) \leq \epsilon$. Thus $p_0(q_1) \leq p_0(q_2)$. ◇

With these results, we show in the following proposition that the energy efficiency function with the constraint, denoted by $\eta_*$, can still be optimized and has a unique maximum.

*Proposition 4:* Given $\eta(p)$ with a unique $p^*$ and a constraint on $\Phi$ as $\Phi \leq \epsilon$, satisfied by $p \geq p_0$; the modified energy efficiency $\eta_*$ has a unique $p_*^* = \min[\max(p_0, p^*), p_{\max}]$.

**Proof :** To proceed with our proof we solve the problem using the KKT conditions (see e.g., [15]). The Lagrangian is defined by:

$$\mathcal{L}(p, \lambda_1, \lambda_2, \lambda_3) = -\eta(p) + \lambda_1(\Phi - L) + \lambda_2(p - P_{\max}) - \lambda_3 p. \tag{12}$$

The KKT conditions applied to the above quasi-convex optimization problem yield

$$\begin{array}{rcl} \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} & = & 0 \\ \Phi - L & \leq & 0 \\ p - P_{\max} & \leq & 0 \\ -p & \leq & 0 \\ \lambda_1(\Phi - L) & = & 0 \\ \lambda_2(p - P_{\max}) & = & 0 \\ \lambda_3(-p) & = 0 & \\ \lambda_1 & \geq 0 & \\ \lambda_2 & \geq 0 & \\ \lambda_3 & \geq 0 & \end{array} \tag{13}$$

◇

## IV. Numerical results

In this section we use simulations and numerical calculations to study the effectiveness of our proposal as well as the advantages offered.

### *A. Convergence to steady state*

So far in our model, we assume the system to be in the steady state. However, in reality, it takes some time for the system to reach the steady state. In order to study the relationship between the observed values for packet loss, and the theoretical values, we devise the following simulation: Using random number generators and a virtual queue we study the fraction of packets lost for a fixed packet count (representing time) to see how fast the simulated queue converges to the steady state. For each total packet count, the simulation is iterated $10^6$ times for a queue size of $K = 10$. We observe that with a packet count of about one thousand, the simulated values of $\Phi$ are on an average the theoretically predicted value ($\pm 4\%$).

### *B. Energy efficiency and power control*

In this section we present some numerical results that bring to focus the advantage of this cross layer approach over a purely physical layer approach.

In the following section we consider the transmitter-receiver pair to be a single input single output link with the success function $f(p) = \exp(-(2^{\frac{R}{R_0}} - 1)\frac{\sigma^2}{p})$. We also always consider a queue of maximum size $K = 10$, $P_{max} = 35$dBm,$P_{min} = 10dB$, $R = 4000$bps and $R_0 = 1000$. Note that as we have $q = 1$ case being identical in theory to the case where we just model the energy efficiency with a purely physical layer approach as in [4] (after accounting for the fixed power consumption of the transmitter $b$); if we optimize the energy efficiency it will be optimal to use the power $p^*(q = 1)$. However if we consider the cross layer model the energy efficiency is optimized at a different $p^*$ based on the transition

probability $q$. As $p^* \leq p^*(q=1)$, using the cross layer optimization we have a gain which can be expressed as $10\log_{10}(\frac{p^*(q=1)}{p^*})$ in dB.

In Figure 1 we study the energy efficiency of a system with $\frac{b}{\sigma^2} = 100$. Here we see that as $q$ decreases $p^*$ decreases. Also seen from the same figure is the quasi-concavity of the energy efficiency function and the asymptotic behavior.

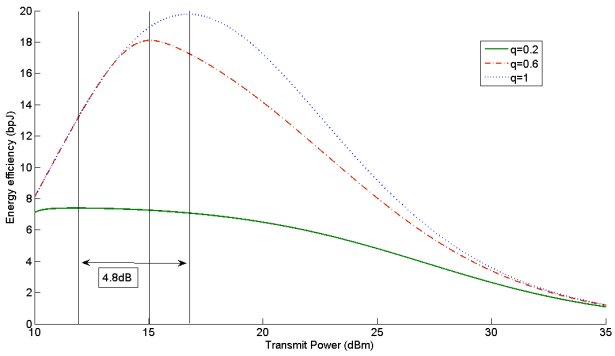


**Fig. 1:** $\eta$ vs $p$ of a system with $\frac{b}{\sigma^2} = 100$ (20dB). Observe that the function is quasi-concave for all $q$ and that $p^*$ decreases as $q$ decreases.

In Figure 2 we study the energy efficiency of a system with $\frac{b}{\sigma^2} = 100$ (20dB) and with a packet loss constraint of $L = 0.01$, i.e., $\Phi \leq 0.01$. Here we see that as $q$ decreases the minimum power required to satisfy the constraint. The quasi-concavity of the energy efficiency function is clearly preserved after the constraint and it has a unique maximum.

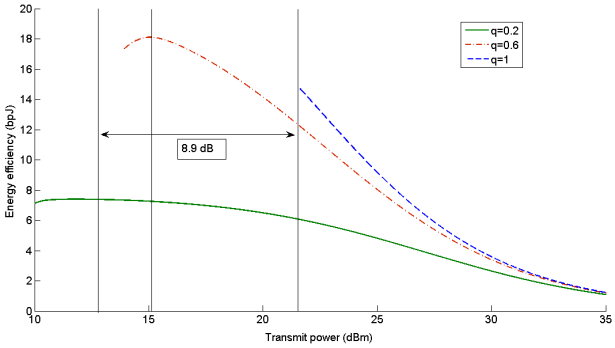


**Fig. 2:** EE of system with $\frac{b}{\sigma^2} = 100$ (20dB) and $L = 0.01$. Note that in this plot, the quasi-concavity is retained and that $p_0$ increases with $q$

In Figure 3 we study the gain in power with $q = 0.5$ plotted against $\frac{b}{\sigma^2}$. For low values of $\frac{b}{\sigma^2}$ we see that the gain for $\epsilon = 0.1$ is due to the constraint causing it to decrease with $\frac{b}{\sigma^2}$, however beyond a certain value of $\frac{b}{\sigma^2}$, even the efficiency function for $q = 0$ is optimized at $p^*$ (the constraint is met for $p \leq p^*$), the gain is due to the difference in $p^*$ which increases with $\frac{b}{\sigma^2}$ just like for the $\epsilon = 1$ case.

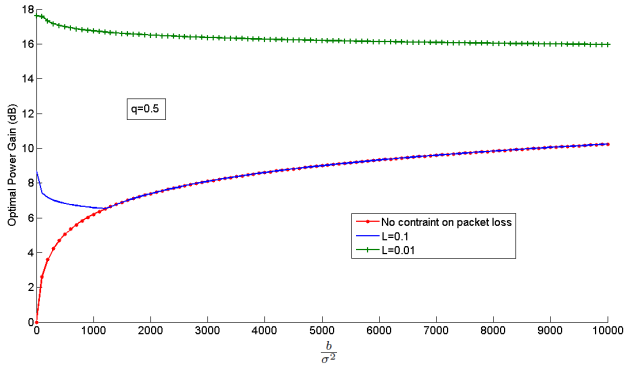


**Fig. 3:** The gain in the optimal power while using a cross layer model as opposed to a purely physical layer model i.e $\frac{p^*(q=1)}{p^*}$ plotted against $\frac{b}{\sigma^2}$.

In Figure 4 we study the gain in power with $\frac{b}{\sigma^2} = 100$ plotted against $q$. For low values of $q$ we clearly see that the gain in using the cross layer approach is the highest and decreases with $q$.

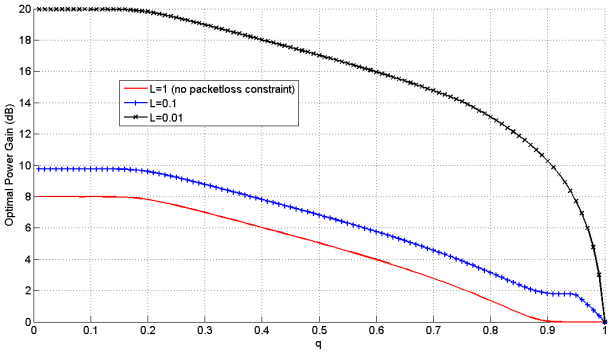


**Fig. 4:** The gain in the optimal power while using a cross layer model as opposed to a purely physical layer model i.e $\frac{p^*(q=1)}{p^*}$ plotted against $q$.

### C. Application of the results on some useful cases

In a realistic situation, when there is no packet to transmit, a base station consumes about $50\%$ of the power it consumes at full load [1]. On the other hand, the entry probability $q$ can change based on the service, protocol, traffic, etc:

- For $q = 0.5$, $R = 256$Kbps and $R_0 = 64$Kbps, our numerical calculations show that, for an SNR of 30 dB, using $p^* = 3\%$ of the transmit power is optimal. While if the user was at a distance where the received SNR is 20dB, using $p^* = 13\%$ of the power is optimal. The relationship between the optimal powers is clearly not linear with the SNR.
- Consider now a system with $q = \frac{1}{25}$ like in some streaming systems that have data sent in one out of 25

frames. In this case, for the user with a SNR of 30dB, $p^* = 1.5\%$ and for 20dB, $p^* = 15\%$. The explanation for this can be seen from the theoretical section, as $q \to 0$, optimizing $\eta$ is the same as optimizing $\frac{f(p)}{p}$ which has a solution that is linear with the SNR.

## V. Conclusion

We have examined the energy efficiency function considering the packet level dynamics of a system and incorporated the effect of the finite buffer size. We find that modeling the system in this form changes the shape of the energy efficiency function. However the energy efficiency function retains its property of quasi-concavity and has a unique maximum. In this work, we also observe that if the packet entry probability is small, the energy efficiency is deformed to a higher extent causing the optimal power to be smaller than in a model ignoring the packet level dynamics. This deformation is due to the constant power consumption of the transmitter even when it does not transmit. The effect of the constant power consumption decreases as the path loss or noise increases and in fact, it is the ratio between the constant power consumption and the noise (with path loss) that determines the deformation. If we impose a constraint on the packet loss, clearly the buffer helps in decreasing this loss which causes a further deformation in the shape of the energy efficiency function.

Many extensions of the proposed work are possible. The most relevant extension is to apply the proposed framework to the case of distributed power control over multiple access channels; we already know that the existence of a pure Nash equilibrium is guaranteed due to quasi-concavity of the energy-efficiency [20]. Another natural extension of the proposed framework is of course, the problem of resource allocation, which is known to be non-trivial, the problem of power allocation is indeed important in multi-carrier and MIMO systems. It would also be fully relevant to study distributed dynamics of the queues and power control policies leading to steady states of the system. Another interesting aspect concerns the case of variable transmission rate as a fixed transmission rate is indeed not the best scheme to minimize energy consumption.

## VI. Acknowledgements

This work is a joint collaboration between Laboratoire des signaux et systémes (L2S) of Supélec, the Alcatel Lucent Chair of Supélec, Orange Labs R&D as well as University of Avignon. This work is also partially supported by "L'Agence Nationale de la Recherche" (ANR) via the program Inter Carnot within the project ANR-09-VERS0: ECOSCELLS.

## References

[1] L. Saker and S-E. Elayoubi, "Sleep mode implementation issues in green base stations", IEEE PIMRC 2010, Istanbul, September 2010

[2] J. Kim, M. Honig, "Resource Allocation for Multiple Classes of DS-CDMA Traffic", in Transactions on Vehicular Technology, vol. 49, no. 2, 2000.

[3] F. Meshkati, H. Poor, S. Schwartz, "Energy Efficiency-Delay Tradeoffs in CDMA Networks: A Game-Theoretic Approach", in Transactions on Information Theory, vol. 55, no. 7, 2009.

[4] D. J. Goodman, and N. Mandayam, "Power Control for Wireless Data", IEEE Personal Communications, vol. 7, pp. 48-54, Apr. 2000.

[5] E.V. Belmega and S. Lasaulce, "Energy-efficient precoding for multiple-antenna terminals", IEEE. Trans. on Signal Processing, 59, 1, Jan. 2011.

[6] G. Bacci, A. Bulzomato, M. Luise, "Uplink power control and subcarrier assignment for an OFDMA multicellular network based on game theory", in Proc. Int. Conf. on Performance Evaluation Methodologies and Tools (ValueTools), Paris, France, May 2011.

[7] M. L. Honig, "Adaptive linear interference suppression for packet DS-CDMA," European Trans. Telecommun., vol. 9, no. 2, pp. 173-181, Mar.-Apr. 1998.

[8] A. Sampath, P. S. Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in Proc. IEEE PIMRC, vol. 1, Toronto, Canada, Sept. 1995, pp. 21-25.

[9] S. Yao and E. Geraniotis, "Optimal power control law for multi-media multi-rate CDMA systems," in Proc. IEEE VTC, vol. 1, Atlanta, GA, Apr. 1996, pp. 392-396.

[10] Q. Shen and W. A. Krzymien, "Power assignment in CDMA personal communication systems with integrated voice/data traffic," in Proc. IEEE GLOBECOM, London, U.K., Nov. 1996, pp. 168-172.

[11] V. Rodriguez, "An Analytical Foundation for Ressource Management in Wireless Communication", IEEE Proc. of Globecom, San Francisco, CA, USA, pp. 898-902, , Dec. 2003.

[12] A. Sampath, N. B. Mandayam, and J. M. Holtzman, "Analysis of an access control mechanism for data traffic in an integrated voice/data wireless CDMA system," in Proc. IEEE VTC, vol. 3, Atlanta, GA, Apr. 1996, pp. 1448-1452.

[13] R. W. Wolff, Stochastic Modeling and the Theory of Queues. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[14] F. Richter, A. J. Fehske, G. Fettweis, "Energy E ciency Aspects of BaseStation Deployment Strategies for Cellular Networks", Proceedings of VTC Fall'2009

[15] S. Boyd, and L. Vandenberghe, "Convex Optimization", Cambridge University Press, 2004.

[16] F. Meshkati, M. Chiang, H. Poor and S. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems ", in JSAC, vol. 24, no. 6, 2006.

[17] W. Eric, "Mean-Value Theorem". MathWorld. Wolfram Research.

[18] J. Kim and M. Hoenig, "Resource Allocation for Multiple Classes of DS-CDMA Traffic" in IEEE Transactions on vehicular technology, vol. 49, no. 2, 2000.

[19] E. Altman , K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, D. Menasche, "Constrained Stochastic Games in Wireless Networks", in proceedings of GlobeCom, 2007.

[20] J. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," Econometrica, vol. 33, pp. 520-534, 1965.

## B.2 TVT-2014

• Vineeth S Varma, Samson Lasaulce, Yezekael Hayel, Salah Eddine Elayoubi and Merouane Debbah, "A Cross-Layer Approach for Energy-Efficient Distributed Power Control", IEEE Trans. on Wireless Comm. (under review)

# A Cross-Layer Approach for Energy-Efficient Distributed Interference Management

Vineeth S. Varma[1,2], Samson Lasaulce[1], Yezekael Hayel[3], and Salah Eddine Elayoubi[2]

**Abstract**

The purpose of this paper is to design distributed energy-efficient power control policies for interference wireless networks. The problem is tackled from a cross-layer perspective. In addition to the physical layer and medium access control protocols, the presence of a finite packet buffer at the transmitter side and the impact of transport protocols are considered. This approach is relevant when the transmission cost is considered to be the total power consumed by the transmitter instead of just the radiated power as assumed usually. A generalized energy-efficiency performance metric integrating these features is constructed under two different scenarios in terms of transport layer protocols characterized by a constant or an adaptive packet arrival rate. The derived performance metric is shown to have several attractive properties, which ensures convergence of the used distributed power control algorithm to a unique point; this point is the equilibrium of a game for which the equilibrium analysis is conducted. A thorough numerical analysis is provided to illustrate the effects of the proposed approach, and provides several valuable insights in terms of designing interference management policies.

**Index Terms**

Cross-layer, distributed optimization, distributed power control, energy-efficiency, game theory, Nash equilibrium, non-cooperative game.

[1] CNRS-Supelec-Univ. Paris Sud 11, 91192 Gif-sur-Yvette, France, email: {vineeth.varma,samson.lasaulce}@lss.supelec.fr

[2] Orange Labs, 92130 Issy-les-Moulineaux, France, email: salaheddine.elayoubi@orange.com

[3] University of Avignon, 84911 Avignon, France, email: yezekael.hayel@univ-avignon.fr

## I. Introduction

Designing green wireless networks [1], [2], [3] has become increasingly important for modern wireless networks, in particular, to manage operating costs. A challenge for modern (beyond 4G and 5G) cellular networks is not only to respond to the explosion of data rates, but also to manage network energy consumption. The concept of small cell networks appears as a good candidate solution to raise such a challenge (see e.g., [4]). As small cell networks will be distributed to large extent and subject to high inter-cell interference, designing distributed energy-efficient interference management schemes appear as a natural need.

For being able to design green networks, an energy-efficiency (EE) metric is needed. In [5], the EE of a communication between a transmitter and a receiver is defined as the ratio of the net data rate to the radiated power; the corresponding quantity is a measure of the average number of bits successfully received per joule consumed at the transmitter. Quite recently, there has been a resurgence of interest in this performance metric. There are several reasons for this and we will only provide a few of them. First, the EE as defined in [5], mathematically translates in a simple manner the trade-off between the benefit of increasing the transmit power in terms of data rate, and the induced cost in terms of consumed energy or amount of created interference. Second, as motivated in [6], there are applications in which the allowable delay is not tightly constrained. Therefore, the data rate is a less relevant measure than the energy needed to transmit the information and EE naturally appears as a metric to be optimized. We furthermore explain in this paper (Sec. III-A) why maximizing EE amounts to minimizing the total energy consumed by the transmitter when packet retransmission is considered.

Remarkably, the energy-efficiency metric proposed in [5] possesses a good mathematical structure for optimization, especially for the distributed case, which partly explains why it has been applied in a large variety of scenarios of practical interest. Some examples are as follows. In [7], it is applied to design a power allocation scheme in distributed multi-carrier CDMA (code division multiple access systems) systems by using a static non-cooperative game model (just as [5]). In [8], it is used to account for the users delay requirements in energy-efficient wireless systems. In [9] and [10], also based on a static game model, the authors used the metric under consideration for sub-carrier assignment in distributed OFDMA (orthogonal frequency division multiple access) multicellular networks. In [11], the authors study the problem of energy-efficient contention-based synchronization in OFDMA systems. In [12], [13], the authors

study the problem of pre-coding in MIMO (multiple input multiple output) point-to-point communication systems. In [14], the EE metric is exploited to study the impact of sensing in terms of EE in cognitive radio networks.

Although fully relevant, the performance metric introduced in [5] and used in the related works (this, in particular, includes those cited above) has left several issues unexplored, which has motivated the work reported here. In [5], and all related references known to the authors, the numerator of EE is (up to a constant) a packet success rate which only accounts for packet losses due to bad channel conditions. In the present work, we propose a significant generalization of the metric used in the aforementioned works to the case where packets are buffered in a finite size queue. Therefore, the packet loss due to overflows is also taken into account. On the other hand, we will show in Sec. III that accounting for this effect is relevant in terms of EE, only when the transmitter has a cost in terms of consumed power independent of the radiated power; this means that the transmitter consumes power even while waiting for new packets to arrive. It turns out that this is precisely what happens for most wireless transmitters. Indeed, the transmitter energy consumption is not only induced by the radiated power but also results from other causes such as the transmitter supply consumption [16]. Note that the authors of [17] were the first to consider a transmission cost of the type "radiated power + constant" to design distributed power control strategies for multiple access channels; in their model, the constant represents the computation power at the receiver. Our approach is markedly different from [17], not only because the problem is tackled from a cross-layer perspective, but we also consider the more general case of distributed interference networks with a quality of service (QoS) constraint. For this purpose, two different models for the packet arrival rate are considered: 1) The quite simple model where the arrival rate is a constant (which is referred to as CAR for constant arrival rate). This case is useful e.g., for real-time traffic like video or streaming; 2) The more interesting model in which the arrival rate is related to the SINR (signal-to-noise plus interference ratio) through a quite generic relationship, is more suited to delay tolerant traffic like file transfer and adaptive rate services like WebRTC [18]; this case is referred to as AAR for adaptive arrival rate.

The main contributions of this paper can be summarized as follows[1]:

1) To the best of the author's knowledge, this is the first time that the EE performance metric

[1]Note that preliminary results were presented in paper [19].

originally introduced in [5] is generalized to a cross-layer approach, taking into account, the effects of the presence of a queue with finite size at the transmitter;

2) Apart from a few exceptions (which includes [17], [19], [20]), all related works using EE in the sense of [5] only consider the radiated power while, here, the total power consumed by the transmitter is accounted for. Since an affine relation between the radiated power and the total power is assumed [16], this might seem as an incremental change but the presence of this fixed cost is the key ingredient which makes the cross-layer analysis fully relevant;

3) The derived performance metric is shown to possess attractive mathematical properties. Quite surprisingly, both in the case of CAR and AAR, it can be shown to be quasi-concave with respect to the radiated power;

4) The above property is directly exploited to prove existence of a Nash equilibrium (NE) in the two static power control games under investigation namely: The game based on CAR with a constraint on the packet loss; The game based on a AAR protocol that automatically controls the packet rates by observing the packet loss. To the best of our knowledge, the former game is the first instance of an energy-efficient power control game to be identified as being semi-continuous, which allows us to prove the existence of an equilibrium by exploiting a fixed point theorem from [21];

5) Generalizing [5] and related references, the games under investigation are shown to be standard [22], [23], which guarantees both NE uniqueness and the convergence of relevant distributed optimization algorithms to this equilibrium such as those based on the iterative water-filling idea [24], [25];

6) A thorough numerical analysis is provided to assess the benefits from taking the presence of a queue with finite size into account and to give new insights into designing energy-efficient communications systems.

This paper is structured as follows. In Sec. II, we present the general system model. In Sec.

III, we construct the proposed performance metric highlighting contributions 1) and 2). In Sec. IV, we define the two power control games of interest and conduct the equilibrium analysis, which is essential to characterize the convergence of the used distributed algorithm (existence and uniqueness of the convergence points), highlighting contributions 3), 4), and 5). Sec. V highlights interesting numerical results that support the proposed approach, i.e, 6). Finally, we conclude the paper and several extensions of this work are provided.

## II. System model

The purpose of this section is to describe the communication model considered for cross-layer energy-efficient power control, which consists in expressing the SINR and packet arrival rate for a given user. A general interference network is considered with $N$ transmitter-receiver pairs, in which each transmitter communicates with its respective receiver, while under interference from the other transmitters [26]. Let $\mathcal{N} = \{1, 2, ..., N\}$ be the set of transmitters. Transmitter $i \in \mathcal{N}$ transmits with power level $p_i \in [0, P_{\max}]$, where $P_{\max} > 0$ is the maximum possible transmit power, which is identical for all transmitters (the analysis does not lose its generality with this assumption). The vector $\underline{p} = (p_1, p_2, \ldots, p_N)$ will be referred to as the power or action profile on the current data block or packet. We also denote by $\underline{p}_{-i}$, the $(N-1)$ dimensional vector obtained by removing the $i^{\text{th}}$ component from $\underline{p}$. For notational simplicity, we also sometimes represent $\underline{p}$ as $(p_i, \underline{p}_{-i})$, when the dependence of certain functions on $p_i$ has to be shown explicitly. By transmitting at $p_i$, each user $i$ has a resulting SINR $\gamma_i$ at his receiver of interest which is a function of $\underline{p}$, and is assumed to be given by:

$$\gamma_i(\underline{p}) = \frac{p_i g_{ii}}{\sigma_i^2 + \sum_{j=1, j\neq i}^{N} p_j g_{ji}} \tag{1}$$

where $g_{ji}$ represents the quasi-static or block fading channel gain of the link between transmitter $j$ and receiver $i$ on a given band, $\sigma_i^2 = \sigma^2$ is the variance of the Gaussian noise at receiver $i$ (these variances can be assumed to be equal without any loss of mathematical generality). In wireless systems such as those being implemented in recent cellular system standards, packets arrive from an upper layer (e.g. IP layer) following an arrival rate that is related to the SINR. In this paper, we assume that the packet arrival process follows a Bernoulli process with probability $q_X(\gamma_i(\underline{p}))$ where $X \in \{\text{CAR}, \text{AAR}\}$; this corresponds to the classical ON/OFF sources [27]. In the

case of CAR, it trivially expresses as:

$$\forall i \in \mathcal{N},\ q_{\mathrm{CAR}}(\gamma_i(\underline{p})) = q \tag{2}$$

with $q \in [0,1]$. This is best used for real-time applications where delay is not tolerable, however, in some applications this packet arrival model is not suitable. For instance, this is the case for applications such as file transfer or browsing. In such a situation, there is no constant stream of data and so the arrival rate can be optimized for best performance in terms of data rate and QoS. This is one of the reasons why we also investigate the case of AAR for which we assume that the arrival rate is given by:

$$\forall i \in \mathcal{N},\ q_{\mathrm{AAR}}(\gamma_i(\underline{p})) = g(\Phi_{\mathrm{AAR}}(\gamma_i(\underline{p}))) \tag{3}$$

where $\Phi_{\mathrm{AAR}}$ is the packet loss function and $g$ is a function which is assumed to be continuous, invertible, and has an inverse function $g^{-1}$ which is twice differentiable, decreasing, and convex. To provide a specific example, the widely used and very useful approximation of the arrival rate process for the Transmission Control Protocol (TCP), which is due to [28], verifies these conditions. Therein, $g$ is merely given by $g(\Phi) = \frac{\kappa}{\sqrt{\Phi}}$, where $\kappa \in [0,1]$ is a parameter which depends on the system design and the round trip time. The resulting rate can be interpreted as the average value for the rate.

*Remark 1.* The CAR protocol can also be seen as a constant piecewise approximation of any adaptive arrival rate protocol in which arrival rate variations are much more slower than channel variations. On the other hand, the AAR case aims at better understanding more complex scenarios where both arrival rate and channel variations have quite similar time-scales. This is close to WebRTC congestion control protocols, like the one proposed by Google [18], where the sending rate is adapted based on the observed packet loss [29].

*Remark 2.* It would be possible to study a more general communication scenario by considering multi-band communications, MIMO communications, or a more advanced reception scheme (e.g., interference cancellation as in [30]). There are several reasons why we do not treat these scenarios here. First, we want to emphasize in a manner as clear as possible the real contributions of this paper namely, the introduction of a queue for the problem of energy-efficient power control. Second, studying the power control problem is the main step towards these extensions. For instance, in [7] in which the authors address EE over multi-carrier multiple access channels, it is proved that the best selfish/equilibrium policy for a transmitter is to select its best carrier

(in terms of SINR) and apply the single-carrier policy to tune the power level over this carrier. Therefore, the assumed model can be understood as a single-band model (e.g., several base stations which try to mitigate inter-cell interference on a given band) or a multi-band model for which interference is managed for the selected channel or interference management is performed independently from band to band.

## III. A new energy-efficiency performance metric

### A. Construction

In [5], EE is defined as the ratio between the average net data transmission rate and the power consumed for sending a given packet. When the radiated power is considered as the transmission cost, this ratio merely equals $\frac{Rf(\gamma_i(\underline{p}))}{p_i}$. The quantity $R$ is the gross data rate on the radio interface and is a constant w.r.t. the power levels. It depends on several parameters such as the modulation and coding scheme [5], [8]. Each packet transmitted on the channel is received without any errors with a probability which depends on the quality of the communication link, the interference, and transmit power levels. The corresponding block or packet success rate (also called efficiency function) is precisely the function $f(\gamma_i(\underline{p}))$ above. The function $f : [0, +\infty) \rightarrow [0, 1]$ is a sigmoidal[2] or S-shaped function verifying $f(0) = 0$ and $\lim_{x\rightarrow\infty} f(x) = 1$ (see [41] for more details). Common examples for $f$ are $f(x) = (1-e^{-x})^M$ [8], $f(x) = e^{-\frac{c}{x}}$ [12], [13], where $M \geq 1$ is the packet length and $c > 0$ is some constant related to spectral efficiency (this relation is specified in Sec. V). Energy-efficiency is particularly relevant when packet retransmission is allowed. When there is no retransmission, the energy[3] consumed to send $V$ bits while transmitting at the power level $p_i$ is $p_i\frac{V}{R}$. Minimizing energy amounts to minimizing $p_i$. However, when retransmission is allowed (typically by using an automatic repeat request -ARQ- protocol, that is used at the physical layer independently of the architecture at the upper layer), the average duration to send a packet equals $\frac{V}{Rf(\gamma_i(\underline{p}))}$ and the energy consumed becomes $p_i\frac{V}{Rf(\gamma_i(\underline{p}))}$. Clearly, minimizing energy amounts to maximizing EE. This means that, at least in presence of re-transmissions, the classical approach which consists in minimizing $p_i$ (subject to some QoS constraints) induces a loss in terms of minimizing the energy consumption; this will be illustrated in Sec. V. In the scenario investigated in this paper, the fact that both the total power consumed by the transmitter and

[2]A sigmoidal function is a function which is initially convex for $\gamma \in [0, \gamma_+]$ and eventually concave for $\gamma \in [\gamma_+, \infty)$.

[3]Here, the energy under consideration is the energy associated with the radiated signal.

the presence of a packet buffer with finite size are considered makes the construction of energy-efficiency more involving than the aforementioned derivation.

A simple model which allows one to relate the radiated power to the total consumed power is provided in [16]; it is given by $P_{\text{total},i} = ap_i + b$, where $a \geq 0, b \geq 0$ are some parameters. We will assume without loss of generality that $a = 1$. The quantity $b$ precisely represents the consumed power when the radiated power is zero[4]. If $\Phi_X$, $X \in \{\text{CAR}, \text{AAR}\}$, represents the packet loss due to both bad channel conditions and packet buffer finiteness (more details about this is provided a little further), a packet is re-transmitted[5] $\frac{q_X(\gamma_i)[1-\Phi_X(\gamma_i)]}{f(\gamma_i)}$ times on average, the average power consumption is in our case $b + p_i\frac{q_X(\gamma_i)[1-\Phi_X(\gamma_i)]}{f(\gamma_i)}$. Since the net data rate or goodput is given by $Rq_X(\gamma_i)[1 - \Phi_X(\gamma_i)]$, we are now able to define the EE metric $\eta_{i,X}(\underline{p})$ as the ratio between the average net data transmission rate and the average power consumption, which gives:

$$\eta_{i,X}(\underline{p}) = R\frac{q_X(\gamma_i(\underline{p}))\left[1 - \Phi_X(\gamma_i(\underline{p}))\right]}{b + p_i\frac{q_X(\gamma_i(\underline{p}))\left[1-\Phi_X(\gamma_i(\underline{p}))\right]}{f(\gamma_i)}}. \tag{4}$$

This definition shows that the cross-layer design approach of power control is fully relevant in terms of EE when the transmitter has a cost, which is independent of the radiated power; otherwise, when $b = 0$, one falls into the original framework of [5]. On the other hand, when $b$ is large, the EE function behaves like a packet success rate function.

Although the efficiency function $f$ (which is assumed to be sigmoidal) can be easily related to the SINR through simple functions such as those mentioned previously, expressing the packet loss function is more involving. Relating $\Phi_X$ to the SINR is the purpose of what follows. A packet is declared to be lost (blocked) only if a new packet arrives when the packet buffer is full and, on the same time-slot, transmission of the packet on the radio interface failed. Note that these two events are independent because the event of "transmit or not" for the current packet on the radio interface, does not impact the current size of the queue, but only the one for the next time slot. This amounts to considering that a packet coming at time slot $t$, is rejected at the end of time slot $t$, the packet of the radio interface having not been successfully transmitted. By considering the stationary regime of the queue and assuming the protocol X, the fraction of

[4]This power consumption occurs even when data is not transmitted due to various causes such as pilot signaling, power amplifier consumption, cooling costs, etc.

[5]For the sake of clarity, here and in other places in the paper, $\underline{p}$ is omitted from the notations.

lost packets $\Phi_X$ can be expressed as follows:

$$\Phi_X(\gamma_i(\underline{p})) = [1 - f(\gamma_i(\underline{p}))]\Pi_X(\gamma_i(\underline{p})) \tag{5}$$

where $\Pi_X(\gamma_i)$ is the stationary probability that the packet buffer is full. Indeed, as already mentioned, each transmitter is assumed to be equipped with a device that allows the packets to be stored in a memory buffer (of size $K \geq 1$) before transmission. Packets arrive into the buffer and get transmitted through a queuing process at the buffer. Denote by $Q_{i,t}$ the size of the queue for transmitter $i$ at time slot $t$. The size of the queue $Q_{i,t}$ is a Markov process on the state space $\mathcal{Q}_i = \{0, 1, \ldots, K\}$. It is known (see [31] for example) that in the stationary regime of the stochastic process $Q_{i,t}$ the probability that the size of the queue equals $K$ is given by:

$$\Pi_X(\gamma_i(\underline{p})) = \frac{\omega_X^K(\gamma_i(\underline{p}))}{1 + \omega_X(\gamma_i(\underline{p})) + \ldots + \omega_X^K(\gamma_i(\underline{p}))} \tag{6}$$

with

$$\omega_X(\gamma_i(\underline{p})) = \frac{q_X(\gamma_i(\underline{p}))\left[1 - f(\gamma_i(\underline{p}))\right]}{\left[1 - q_X(\gamma_i(\underline{p}))\right] f(\gamma_i(\underline{p}))} \tag{7}$$

where $X \in \{\text{CAR}, \text{AAR}\}$.

In the case of $X = \text{AAR}$, the packet arrival rate $q_{\text{AAR}}$ is a function of the packet loss and the packet loss, a function of $q_{\text{AAR}}$. The following proposition ensures that the AAR process achieves an average packet arrival rate according to the following proposition. For the purpose of making the inter-dependency of the two following equations clear, we express explicitly in these equations, some of the parameters used implicitly in the rest of the paper.

*Proposition 3.1:* The packet arrival rate $q_{\text{AAR}}$ is obtained as the unique fixed point of these equations:

$$\Phi_{\text{AAR}}(\gamma_i(\underline{p})) = (1 - f(\gamma_i(\underline{p})))\Pi_{\text{AAR}}(\gamma_i(\underline{p})) \tag{8}$$

where $\Pi_{\text{AAR}}(\gamma_i(\underline{p}))$ has $q_{\text{AAR}}$ as a parameter as seen from (7) and (6), and:

$$q_{\text{AAR}}(\Phi_{\text{AAR}}) = g(\Phi_{\text{AAR}}). \tag{9}$$

*Proof:* It can be verified that the two equations are continuous and differentiable. The packet arrival rate $q_{\text{AAR}}(\gamma_i)$ ranges from $0 < g(0) \leq 1$ to $0 \leq g(1) < g(0)$ and $\Phi_{\text{AAR}}(\gamma_i)$ ranges from 0 to 1. Based on the properties of $\Phi_{\text{AAR}}$ given in App. A, $\Phi_{\text{AAR}}$ ranges from 0 to 1 as $q$ goes from 0 to 1. Now study $F(q_{\text{AAR}}) \triangleq \Phi_{\text{AAR}}(\gamma_i, q_{\text{AAR}}) - g^{-1}(q_{\text{AAR}})$. The function $F(q_{\text{AAR}})$ is a continuous and differentiable function in the interval of $q \in [0, 1]$. A point such that $F(q_{\text{AAR}}) = 0$ is a fixed

point for this set of equations. Based on the mean value theorem [32], and from the limits $\lim_{q\to 0} F(q_{\mathrm{AAR}}) \leq 0$ and $\lim_{q\to 1} F(q_{\mathrm{AAR}}) \geq 0$, we have $F(q_{\mathrm{AAR}}) = 0$ for some $q \in [0,1]$. Also note that $F(q)$ is strictly increasing and so the point where $F(q_{\mathrm{AAR}}) = 0$ is unique.

The fixed point equation can be solved as:

$$g^{-1}(q_{\mathrm{AAR}}(\gamma_i(\underline{p}))) = [1 - f(\gamma_i(\underline{p}))]\frac{\omega_i(\gamma_i(\underline{p}))^K}{\sum_{j=0}^{K} \omega_i(\gamma_i(\underline{p}))^j} \tag{10}$$

and has a unique solution. ■

*Remark 3.* For $b > 0$, it can be seen that for any $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$, $q_X \to 1 \Rightarrow \omega_X \to +\infty \Rightarrow \Pi_X \to 1 \Rightarrow \Phi_X \to 1 - f$, which means that one falls into the framework of [17].

*Remark 4.* When the packet arrival is constant (i.e., X = CAR), the dependency of $\Pi_X$ regarding the SINR follows a simple relation. However, when the AAR protocol is assumed, the relationship is less trivial. Indeed, the packet loss $\Phi_X$ depends on $\omega_X$ through (5) and (6). The quantity $\omega_X$ depends on the arrival rate $q_X$. But, in the AAR case, $q_X$ also depends on the packet loss. This is the reason why we assume that, under the AAR protocol assumption, each transmitter operates at the fixed point associated with the aforementioned dependency chain. Therefore, this amounts to fixing the packet loss function to have a certain form. AAR can thus be seen as an indirect way of imposing a certain QoS on the transmission. To be more specific, if one assumes an arrival rate process which can be approximated as in [28] (namely, $g(\Phi) = \frac{\kappa}{\sqrt{\Phi}}$) and the regime of large buffer size $K \to \infty$, the operating packet arrival rate function can be shown to be :

$$\lim_{K\to\infty} q_{\mathrm{AAR}}(\gamma_i) = f(\gamma_i)\frac{1 + \sqrt{1 + 4\left(\frac{\kappa}{f(\gamma_i)}\right)^2}}{2}. \tag{11}$$

*Remark 5.* In the above equations, we have implicitly made a symmetry assumption: the efficiency and arrival rate functions are assumed to be identical for all users. This choice allows one to gain in terms of clarity while the extension to the non-symmetric case is ready. For the same reason, the gross data rates at which the users transmit $R_i, i \in \mathcal{N}$, have been assumed to be equal (to $R$ bit/s).

*Remark 6.* As the form of the performance metric under consideration implicitly indicates (see (4)), the choice made in this paper is not to account for possible memory effects which would be due e.g., to correlated channel realizations from block to block or the state of the queue. This choice is coherent with the related literature on EE which originates from [5] and the merit of it is that the corresponding power control policies remain distributed in the sense

of the required knowledge to implement it. As seen in Sec. IV, a transmitter only needs to know its instantaneous SINR to tune its power level on the current block and therefore manage EE and created interference. Exploiting stochastic models can be seen as a relevant extension of the present paper which would lead to a better performance (provided all the additional parameters required are well estimated) but at the expense of obtaining power control policies which are (possibly much) more demanding computationally and requiring (possibly much) more information (see e.g., [33][34]). Summarizing, the proposed approach can be seen as a reasonable tradeoff between performance gain in terms of EE and ease of implementation.

### *B. Properties*

The EE function $\eta_{i,X}$ possesses a very attractive property regarding its dependency toward $p_i$. This is what the next proposition states.

*Proposition 3.2:* For all $i \in \mathcal{N}$, the EE function $\eta_{i,\mathrm{CAR}}(\underline{p})$ is quasi-concave w.r.t. $p_i$ and has a unique maximum point denoted by $p_i^*(\underline{p}_{-i})$.

The proof relies, in particular, on the sigmoidness assumption for $f$ and can be found in App. A for the CAR case and in App. A-B for the AAR case. This result is very useful for the NE analysis which is conducted in Sec. IV. Remarkably, the quasi-concavity property is not only available in the case of CAR but also in the case of AAR, which is not obvious a priori.

In order to obtain more insights about the impact of having a buffer on energy-efficiency, we now briefly analyze the case of CAR. The following result holds.

*Proposition 3.3:* Let X = CAR. For all $i \in \mathcal{N}$, the EE function $\eta_{i,X}$ is a strictly increasing function of the parameter $q$.

*Proof:* Let $\underline{p}$ be fixed and remove the dependency toward $\underline{p}$ and $\gamma_i$ from the notations. The EE function can be rewritten as $\eta_{i,\mathrm{CAR}} = \frac{1}{\frac{b}{(1-\Phi_{\mathrm{CAR}})q}+\frac{p_i}{f}}$. Clearly, if the sufficient condition $\frac{\partial \Phi_{\mathrm{CAR}}}{\partial q} < \frac{1-\Phi_{\mathrm{CAR}}}{q}$ holds, then $\frac{\partial}{\partial q}(1-\Phi_{\mathrm{CAR}})q > 0$. From this, it follows that $\frac{\partial \eta_{i,\mathrm{CAR}}}{\partial q} > 0$. Let us prove the sufficient condition. The derivative $\frac{\partial \Phi_{\mathrm{CAR}}}{\partial q}$ can also be written as $\frac{\partial \Phi_{\mathrm{CAR}}}{\partial q} = -\Phi_{\mathrm{CAR}}^2 \frac{\partial (\Phi_{\mathrm{CAR}}^{-1})}{\partial q}$ with $\Phi_{\mathrm{CAR}}^{-1} = 1 + \frac{1}{\omega_{\mathrm{CAR}}} + .. + \frac{1}{\omega_{\mathrm{CAR}}^K}$. Using

$$\frac{\partial \omega_{\mathrm{CAR}}}{\partial q} = \frac{1-f}{f}\frac{1}{(1-q)^2} \tag{12}$$

implies that

$$\frac{\partial \Phi_{\mathrm{CAR}}^{-1}}{\partial q} = \left(\frac{1}{\omega_{\mathrm{CAR}}} + .. + \frac{K}{\omega_{\mathrm{CAR}}^K}\right)\frac{1}{q(1-q)} > \frac{1}{\Pi_{\mathrm{CAR}} q(1-q)}. \tag{13}$$

The sufficient condition follows by using $\frac{1}{\Pi_{\text{CAR}} q(1-q)} > 0$ and thus $\frac{\partial \Phi_{\text{CAR}}}{\partial q} < 0 < \frac{1-\Phi_{\text{CAR}}}{q}$. ■

This proposition mathematically translates the following intuition. If the packet arrival rate $q$ decreases, the average duration during which the buffer is empty increases. Since there is a fixed transmission cost $b$, this induces a waste of energy.

To conclude this section, let us analyze the limit of large buffer size. Two sub-cases can be distinguished.

- Case 1: $q_X > f(\gamma_i(\underline{p}))$, i.e., $\omega_X > 1$. We have that the steady-state probability of having a full buffer $\lim_{K\to\infty} \Pi_X = \frac{\omega_X - 1}{\omega_X}$ and a simplification yields $\Phi_X = 1 - \frac{f(\gamma_i(\underline{p}))}{q_X}$. Thus the EE becomes $\lim_{K\to\infty} \eta_{i,X}(\underline{p}) = \frac{R f(\gamma_i(\underline{p}))}{b + p_i}$. This means that a higher probability of entrance than exit causes the queue size to blow up, and there are always packets to be transmitted, which explains why one falls into the framework of [17] in Case 1.
- Case 2: $q_X \leq f(\gamma_i(\underline{p}))$, i.e., $\omega_X \leq 1$. If $f(\gamma_i(\underline{p})) = q_X$, then $\Pi_X = \frac{1}{K}$ and $\lim_{K\to\infty} \Pi_X = 0$. For $f(\gamma_i(\underline{p})) > q_X$, we have also that $\lim_{K\to\infty} \Pi_X = 0$ and then simplification yields $\Phi_X \to 0$. Thus the EE becomes $\lim_{K\to\infty} \eta_{i,X}(\underline{p}) = \frac{R}{\frac{b}{q_X} + \frac{p_i}{f(\gamma_i(\underline{p}))}}$. This means that, even with the fixed consumption cost $b$, the EE performance metric to be optimized becomes $\frac{f(\gamma_i(\underline{p}))}{p_i}$ (i.e., the same metric as [5]). This is also quite intuitive, as in the steady state, due to a higher probability of exit, the buffer is never full and there is no packet loss due to buffer overflow.

*Remark 7.* The above special case analysis suggests that, in the regime of large buffer size, the power control policies may be obtained from an approximated payoff function which is simpler than the exact expression (4). It is seen that, depending on the current value of the SINR and arrival rate, the approximated payoff function coincides either with that of [5] or [17].

### *C. QoS constraint*

To conclude this part, we will mention how the QoS constraint is treated in our analysis. As already mentioned in Sec. I, one of the recurrent problems with most works using the performance metric introduced in [5] is that EE can be maximized at a power level which does not guarantee a minimum QoS. This is why, in the case of CAR, we also consider a constraint when maximizing (4): the packet loss rate $\Pi_{\text{CAR}}[1 - f(\gamma_i)]$ has to be less than an upper bound $\epsilon$. For example, in cellular systems, typical values for $\epsilon$ are 0.1 or 0.01, based on the system requirements. Adding this constraint restricts the range of power usable by the transmitter by

adding a lower bound on the power. This lower bound depends on the entry probability $q$ and on the size of the queue $K$. At last, our choice is not to impose this constraint for the AAR protocol since, by construction, this protocol aims at automatically adapting the packet arrival rate to congestion.

## IV. Equilibrium analysis and distributed power control algorithm

Since it is assumed that transmitter $i$, $i \in \mathcal{N}$, can only control the variable $p_i$ of the $N-$variable function $\eta_{i,X}(\underline{p})$ and is assumed to consider the energy-efficiency of his own communication, the power control problem is naturally distributed in terms of the decision. The ultimate goal of this section is to propose a power control algorithm which is distributed both in terms of the decision and information (only individual SINR feedback is required to adapt the power level). While the algorithm itself is directly inspired from existing works, its convergence analysis does not follow from a direct adaptation of existing results. We first begin by an equilibrium analysis of two non-cooperative games associated with the two considered power control scenarios (CAR and AAR), before proposing the algorithm and proving its convergence.

### *A. Equilibrium analysis of the associated games*

A non-cooperative game under strategic form is merely given by an ordered triplet (see e.g., [23]). With the notations of this paper it writes as

$$\mathcal{G}_X = \left(\mathcal{N}, \{\mathcal{P}_i\}_{i\in\mathcal{N}}, \{u_{i,X}\}_{i\in\mathcal{N}}\right) \tag{14}$$

where the set of decision-makers (DMs) or players is therefore the set of transmitters, the action space for DM $i$ is $\mathcal{P}_i = [0, P_{\max}]$, and $u_{i,X}$ is the payoff function of DM $i$ when the arrival rate model is $X$. As explained in Sec. II, when CAR is assumed, a QoS constraint is imposed on the packet loss. Under this assumption, the payoff function is chosen to be:

$$u_{i,\mathrm{CAR}}(\underline{p}) = \begin{cases} \eta_{i,\mathrm{CAR}}(\underline{p}) & \text{if } \Phi_{\mathrm{CAR}}(\gamma_i(\underline{p})) \le \epsilon \\ R\dfrac{q\left[1-\Phi_{\mathrm{CAR}}(\gamma_i(\underline{p}))\right]}{b+P_{\max}} & \text{otherwise} \end{cases}. \tag{15}$$

This payoff definition means that as long as the QoS constraint can be met, energy-efficiency maximization is pursued. However, if the constraint cannot be met, goodput maximization or packet loss minimization is sought. Note that, for any constraint $\epsilon$, the action space of any DM $i$ is still the interval $[0, P_{\max}]$. The constraint is instead merged into the payoff function in such

a manner that as long as the constraint is not satisfied, it is always optimal to increase power. For the AAR case, the payoff function is simply defined by

$$u_{i,\mathrm{AAR}}(\underline{p}) = \eta_{i,\mathrm{AAR}}(\underline{p}). \tag{16}$$

A fundamental solution concept for a non-cooperative game is the Nash equilibrium. There are at least two important reasons for this. When operating at an NE, a network possesses a form of strategic stability: a transmitter which changes his power control policy while the others keep on using the equilibrium policies, will see this payoff decreased or maintained in the best case. The second reason is that, under some conditions, important iterative distributed optimization algorithms such as the sequential best-response dynamics (called sequential iterative water-filling in the literature of distributed power allocation whose objective is to maximize the transmission rate) converge to an NE. It turns out that the two games under study verify a simple sufficient condition which allows the second feature to be exploited. All of this gives us a strong reason for conducting the equilibrium analysis for the two defined games in order to show that the two games above possess an equilibrium, to prove that it is unique and to provide a simple distributed optimization algorithm which converges to it. This is the purpose of the following propositions which follow the definition of a Nash equilibrium in our context.

*Definition 4.1:* The vector of transmit power levels $\underline{p}_X^{\mathrm{NE}}$ is a pure Nash equilibrium of the game $\mathcal{G}_X$ if:

$$\forall i \in \mathcal{N},\ \forall p_i \in \mathcal{P}_i,\ u_{i,X}(\underline{p}^{\mathrm{NE}}) \geq u_{i,X}(p_i, \underline{p}^{\mathrm{NE}}_{-i,X}). \tag{17}$$

*Proposition 4.2:* For $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$, the game $\mathcal{G}_X$ admits at least one pure Nash equilibrium.

*Proof:* The proof is based on a fixed point theorem proved in [21]. The called theorem states that if the action spaces are compact convex sets, every payoff function of the game is upper semi-continuous w.r.t. the action profile ($\underline{p}$ in our context), and for any DM $i$ the payoff function is quasi-concave w.r.t. to the individual action ($p_i$ in our context), then the game possesses a pure Nash equilibrium. For $X \in \{\mathrm{CAR}, \mathrm{AAR}\}$, the action space is $[0, P_{\max}]$ which is a compact convex space. When $X = \mathrm{CAR}$, $u_{i,X}$ is upper semi-continuous w.r.t $p_i$ whereas it is continuous when $X = \mathrm{AAR}$. From Prop. 3.2, we know that for any X, individual quasi-concavity is available. This concludes the proof. ■

To our knowledge, all related works on energy-efficient power control use utilities which are continuous with the power profile $\underline{p}$. Interestingly, a relevant power control game in which continuity is not available can be exhibited for the case of $X = \mathrm{CAR}$.

*Proposition 4.3:* For $X \in \{\text{CAR}, \text{AAR}\}$, the game $\mathcal{G}_X$ admits a unique pure Nash equilibrium, for which the equilibrium power policy will be denoted by $\underline{p}_X^{\text{NE}}$.

*Proof:* The proof of this result mainly relies on one important property of the studied games, namely both games are standard in the sense of [22]. In App. C we prove that the DMs' best-responses are always standard functions; by definition, the best-response of a DM $i$ to the (reduced) action profile $\underline{p}_{-i}$ is the set-valued function defined by $\text{BR}_{i,X}(\underline{p}_{-i}) = \arg\max_{p_i} u_i(\underline{p})$. If the best-responses of all the DM's are standard, then the game is also standard, which completes our proof. ■

*B. The proposed distributed interference management algorithm*

The property of the previous proposition is also sufficient to guarantee convergence of some important distributed optimization algorithms. Note that the argmax set mentioned in the proof is a singleton (a scalar value), which can be checked from App. B for CAR and App. A-B for AAR. While this property is available for the scenario studied in [5] and many related works, it is seen here that, although the proposed QoS oriented cross-layer approach leads us to more complex and more general utilities, this property is still valid.

This means that for these algorithms, not only is convergence ensured, but the convergence point is also unique. This is very useful to characterize the performance of an implemented distributed power control algorithm. Here, we only mention one of such algorithms: the asynchronous or sequential best-response dynamics. This algorithm is well known in game theory [35] and draws its roots from the paper by Cournot [36]. It has been used in [5] and is often used because convergence to the NE can be guaranteed. Let $\underline{p}_X^{\text{NE}}$ be the unique NE of $\mathcal{G}_X$. For the algorithm, we define $\underline{p}(t)$ as the power control policy in the previous time slot, and $\underline{p}(t+1)$ as the power control policy for the current time slot. Algorithm 1 implements the sequential best-response dynamics for $\mathcal{G}_X$:

Several comments are in order.

1) We have assumed that DM 1 updates first, who is followed by DM 2, etc. In fact, this order can be arbitrary provided it is fixed (see e.g., [37]).
2) To update the power levels $m$ times, a duration corresponding to $mN$ time-slots is required.
3) The quantity $\delta > 0$ corresponds to the accuracy level wanted for the stopping criteria in terms of convergence to the NE.

**Algorithm 1** Sequential best-response dynamics

$\Delta \leftarrow 2\delta$ ▷ Initialize the observed difference in power levels over time, $\delta$ is the tolerance.

$\underline{p}^0 \leftarrow (P_{\max}, P_{\max}, \ldots, P_{\max})$ ▷ The starting power is uniform power with $P_{\max}$.

$t \leftarrow 0$ ▷ The starting time is 0.

**while** $\Delta \geq \delta$ **do** ▷ The outer loop that iterates till the power policies converge.

**for** $i = 1 \rightarrow N$ **do** ▷ The inner loop iterating over the DM indices.

$\Gamma_i = \dfrac{\gamma_i(\underline{p}^t)}{p_i^t}$ ▷ Using the SINR feedback from its receiver, DM i calculates the interference term $\Gamma_i$ for the previous time slot.

$p^* \leftarrow \arg\max_p \left( \dfrac{Rq_X(p\Gamma_i)(1-\Phi_X(p\Gamma_i))}{b + \dfrac{p}{f(p\Gamma_i)} q_X(p\Gamma_i)(1-\Phi_X(p\Gamma_i))} \right)$ ▷ Calculate the optimal power that maximizes the EE.

**if** X=CAR **then**

$p_+ \leftarrow \min(p; \Phi_{\text{CAR}}(p\Gamma_i) \geq \epsilon)$ ▷ Calculate the minimum power to satisfy the QoS constraint.

$p_i^{t+1} \leftarrow \min(\max(p^*, p_+), P_{\max})$ ▷ Choose the optimal power for CAR if less than $P_{\max}$ and more than $p_+$.

**else**

$p_i^{t+1} \leftarrow \min(p^*, P_{\max})$ ▷ Choose the optimal power for AAR if less than $P_{\max}$.

**end if**

**end for**

$\Delta \leftarrow \max_i(|p_i^{t+1} - p_i^t|)$

$t \leftarrow t + 1$

**end while**

4) The algorithm 1 is completely distributed in the sense that to update his power, a DM only needs to know the SINR corresponding to his chosen power level, i.e., $\text{BR}_{i,X}(\underline{p}_{-i})$ can be calculated by knowing $\gamma_i$ for some $p_i$. This is typically achieved using a feedback mechanism and does not require a central entity that provides knowledge of the channel conditions or power levels chosen by the other DMs.

## V. Numerical results

### A. General setup

Unless explicitly stated otherwise, the following choices and parameters are assumed for all the simulations provided here:

- The number of users or transmitters is set to two ($N = 2$). This scenario was chosen because the behavior of various metrics like the price of anarchy (PoA) can be easily analyzed in this situation. The case of "high interference", as defined below, is also studied to compensate for this choice. In addition, some specific figures also study the case with more interferers.
- The block success rate function is chosen as in [13]: $f(\gamma_i) = \exp\left[-\left(\frac{2^{\frac{R}{R_0}}-1}{\gamma_i}\right)\right]$ where $R_0 = 1$ MHz is the bandwidth used and the gross data rate is $R = 1$ bit/s.
- When the adaptive arrival rate scenario is considered, it it is assumed that $g(\phi) = \frac{0.1}{\sqrt{\phi}}$.
- We define the low (resp. high) interference scenario as: $\mathbb{E}(g_{ii}) = 2.5$ and $\mathbb{E}(g_{ij}) = 0.5$ for $j \neq i$ (resp. $\mathbb{E}(g_{ii}) = 2.5$ and $\mathbb{E}(g_{ij}) = 2$ for $j \neq i$). For some simulations, the channel gains will be assumed to be fixed while for the others it will follow classical block Rayleigh fading. The values indicated will be the instantaneous channel fading when the scenario considered is static and otherwise will indicate the variance.
- The noise level is set to $\sigma^2 = 1$ mW; the maximum power $P_{\max} = 1000$ mW; buffer size of $K = 10$; $\epsilon = 1$ (packet loss constraint) and the fixed power consumption $b = 1000$ mW.
- To measure the global efficiency of the interference network with respect to the centralized solution, we use the price of anarchy. [38] gave a definition of PoA where the optimal situation correpsponds to a POA equal to 1, while other situations correspond to a PoA> 1:

$$\forall X \in \{\text{CAR}, \text{AAR}\},\ \text{PoA}_X = \frac{\max\limits_{\underline{p}} \sum\limits_i u_{i,X}(\underline{p})}{\sum\limits_i u_{i,X}(\underline{p}^{\text{NE}})}. \tag{18}$$

### B. About the considered EE performance metric

Here we assume a single-user scenario i.e., $N = 1$, a fixed channel gain (namely $g_{11} = 2.5$), and the arrival rate to be fixed (CAR scenario). Fig. 1 depicts the EE (4) as a function of the chosen radiated power for different values of the fixed consumption cost $b$ and packet arrival rate $q$. First, the figure illustrates what has already been proved through Prop. 3.2 namely, EE is quasi-concave w.r.t. the radiated power. Second, we fix $q$ to one and assess the influence of $b$. As $b$ increases, the curve becomes less peaky. In fact, if $b$ becomes very high, EE tends to merely

becomes a packet success rate function. This means that power control becomes irrelevant since it merely boils down to transmitting at maximum power whatever the channel conditions. Now we fix $b$ to 1000 mW. By moving from the arrival rate of $q = 1$ (framework of [17]) to $q = 0.6$ (with a buffer size of $K = 10$), it is seen that the EE curve is quite significantly changed and the optimal radiated power changes from 460 mW to 320 mW. In the next section, the gain in terms of radiated power brought by the cross-layer approach is quantified in a more general scenario.

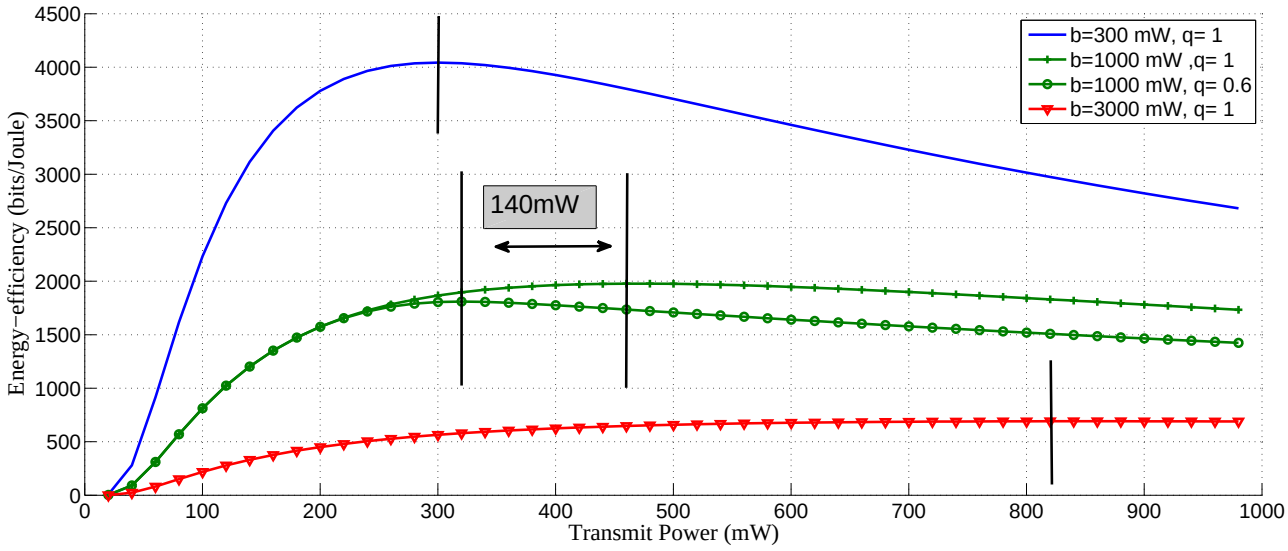


Fig. 1. CAR: EE against $p_1$, i.e., the energy efficiency as a function of the transmit power for various values of the constant power ($b$) and packet arrival rate ($q_{CAR}$).

### C. *Influence of the packet arrival rate in the CAR scenario*

Here we assume the low interference scenario. For $K = 10$, Fig. 2 represents the gains in dB in terms of radiated power which is brought by the proposed cross-layer approach (after convergence of the proposed distributed power control algorithm) w.r.t. the conventional approach in which it is (implicitly) assumed that $q \to 1$ [17]. The gain is therefore defined by $10\log_{10}\left(\frac{p_i^{\mathrm{NE}}[q \to 1]}{p_i^{\mathrm{NE}}[q]}\right)$, for a given $i \in \{1,2,3\}$, say $i = 1$ (the gain is the same for the different transmitters since the average channel gains are identical). The gain is represented as a function

of the packet arrival rate. It is seen that, for different numbers of transmitter-receiver pairs ($N = 2$ or $N = 3$) and a raw packet error rate of $\epsilon = 0.1$ (by raw it is meant before re-transmission), the gain is significant if the arrival rate is typically less than 0.5. Gains as high as 10 dB (with $N - 1 = 2$ interfering users on the same band) or 30 dB (with $N - 1 = 1$ interfering user on the same band). If the raw QoS constraint is relaxed ($\epsilon = 1$), quite similar observations can be made. These gains are not in terms of energy consumed by the whole transmit device but they mean that transmitters use much less radiated power and therefore create much less interference, while reaching the same QoS.

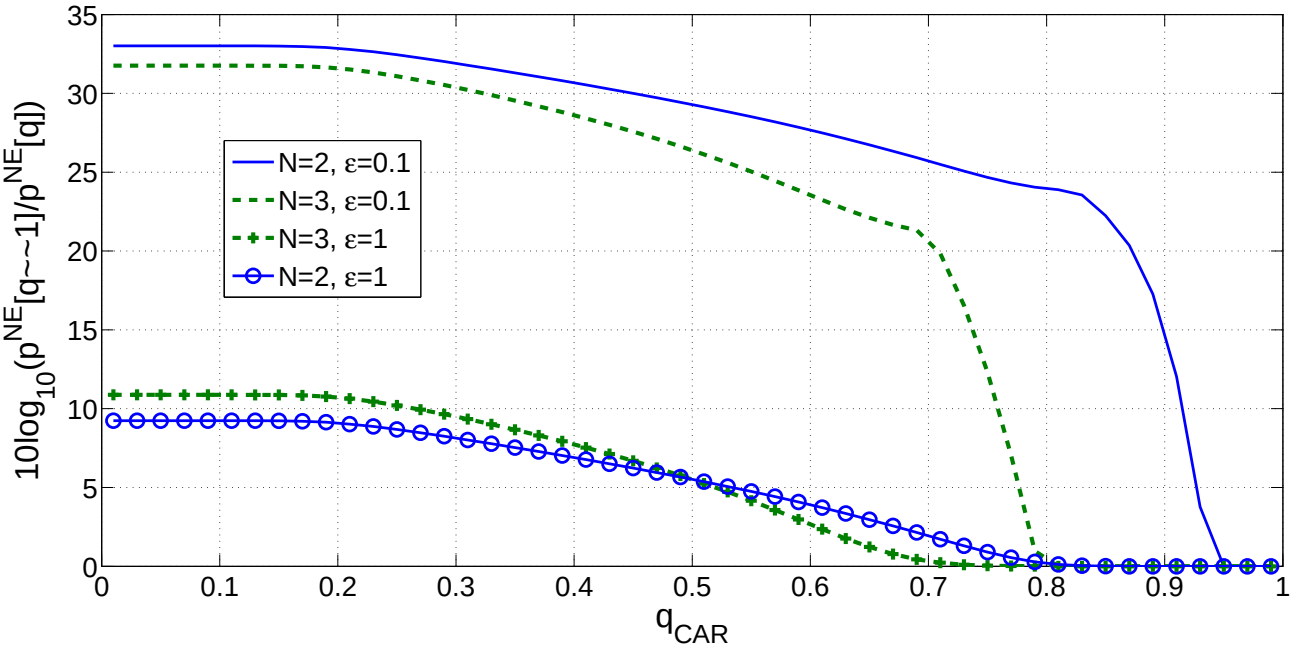


Fig. 2. CAR: $10\log_{10}\left(\frac{p_1^{\mathrm{NE}}[q \to 1]}{p_1^{\mathrm{NE}}[q]}\right)$ against $q$, i.e., the ratio of equilibrium power levels in the cross-layer case to the case where the buffer is ignored and arrival rate is one. Interestingly, our cross-layer approach does not only allow the EE to be maximized but also allows significant gains in terms of radiated power. The transmit power for the cross-layer approach is always lower than for the purely physical layer approach, and this difference is more prominent when a packet loss constraint is imposed.

In the low interference static channel scenario, Fig. 3 depicts the PoA or price of having a distributed network versus the packet arrival rate for different buffer sizes ($K = 1$ and $K = 10$) and a raw packet error rate of $\epsilon = 0.02$. In contrast with existing works on EE, the PoA can be

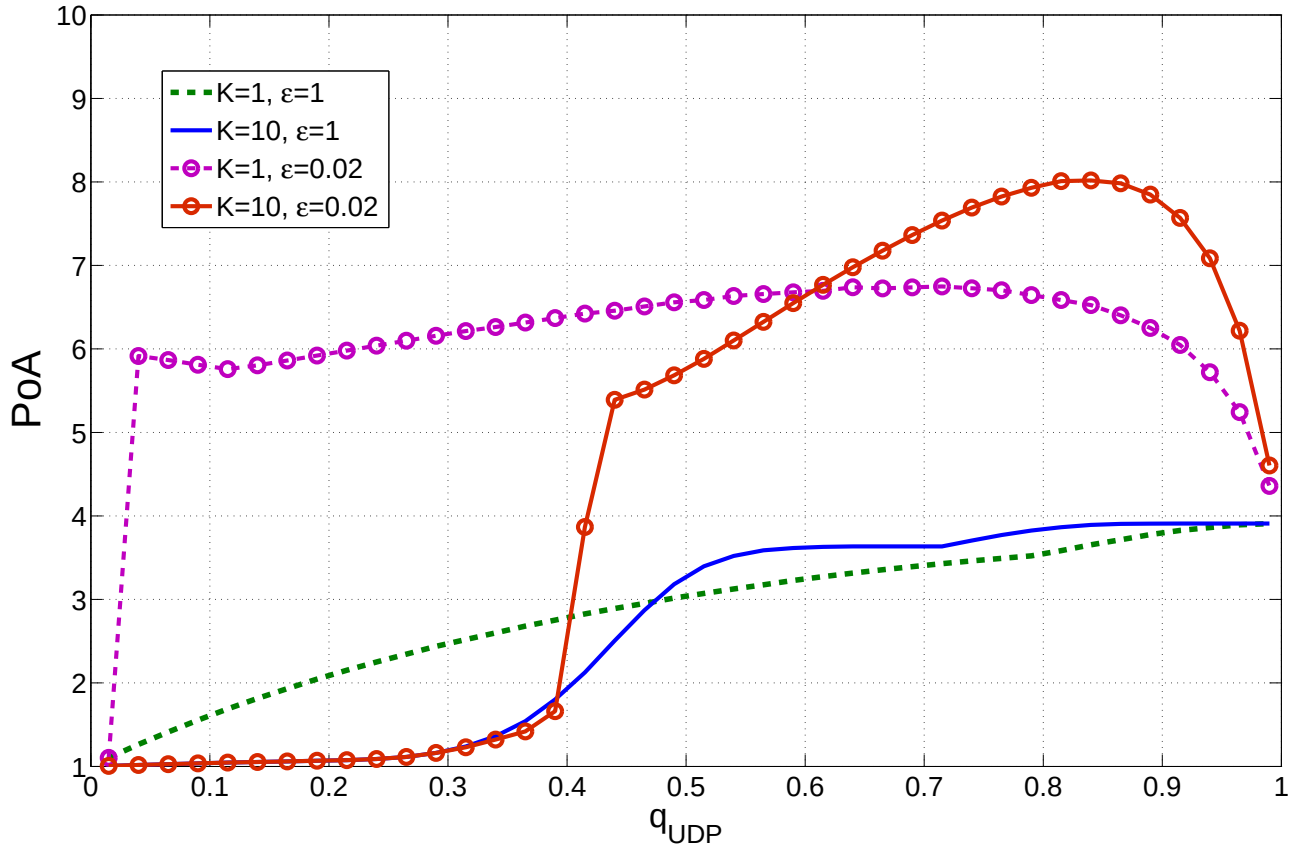


Fig. 3. CAR: Here, a given realization is assumed for the channel. In contrast with existing works on energy-efficient power control which assume $q \to 1$ and therefore always obtain a high value for the PoA, it is seen here that low values are actually reachable when the packet rate is sufficiently small. With a large enough buffer size ($K$), even for $\epsilon = 0.02$, the NE is close to centralized solution if the right $q$ is used.

small in energy-efficient interference networks. This occurs when the arrival rate is typically less than 0.4 and for a reasonably large buffer size; $K = 10$ is in fact quite small while $K = 1$ is the minimum buffer size possible and corresponds a very extreme case. The jump observed in the figure around $q = 0.4$ at low interference and $q = 0.5$ at high interference. This occurs when $q \geq f(\gamma^{NE})$. This jump in the PoA occurs when the value of $q$ crosses this threshold, as the equilibrium power control policy before the jump corresponds to a power control policy closer to the one seen in [5], while after the jump, the equilibrium is closer to the one in [17]. It is therefore worth noting that, under some realistic conditions, a distributed interference management policies can

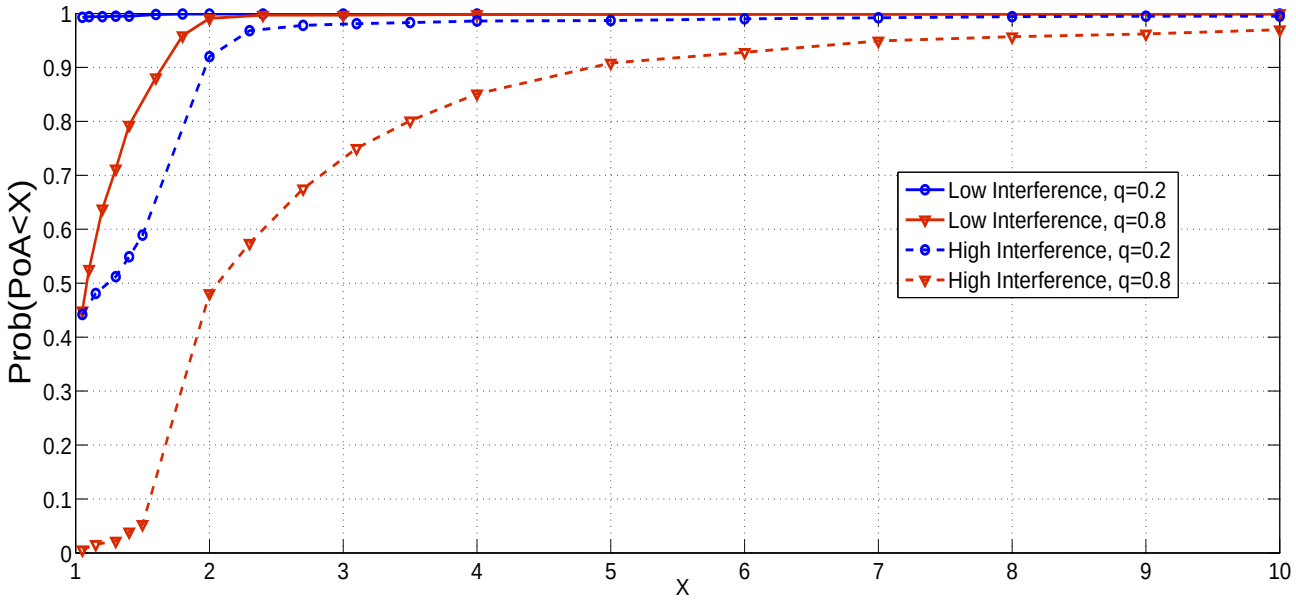


Fig. 4. Contrarily to Fig. 3 which assumes a given channel realization, this figure is obtained by averaging over channel realizations. The CDF of the PoA provides information about how often the price of having a distributed system is low or high. In this figure as well, we see that even in the high interference regime, a small arrival rate can lead to an efficient equilibrium more often.

perform as well as a centralized one. To our knowledge, this observation has not been made before in the literature originating from [5] because all the corresponding works assume that the transmitter has always packets to send while this is not the case in many real scenarios (download speeds are often limited by server speeds).

Since our observations regarding the PoA might be thought to be related to the specific realization of the channel, we now provide numerical results which have been obtained by averaging over channel realizations. Fig. 4 shows the cumulative distribution function (CDF) of the PoA for four parameter settings: Low interference scenario and $q_{\text{CAR}} = 0.2$; Low interference scenario and $q_{\text{CAR}} = 0.8$; High interference scenario and $q_{\text{CAR}} = 0.2$; and finally, high interference scenario and $q_{\text{CAR}} = 0.8$. This figure confirms that the loss on optimality induced by decentralization is rather small if transmissions are sporadic and interference is not severe.

Now, Fig. 5 represents the network sum-payoff, which is an absolute performance measure. In the high interference scenario, for $K = 10$, a raw QoS of $\epsilon = 0.02$, the figure depicts the sum-

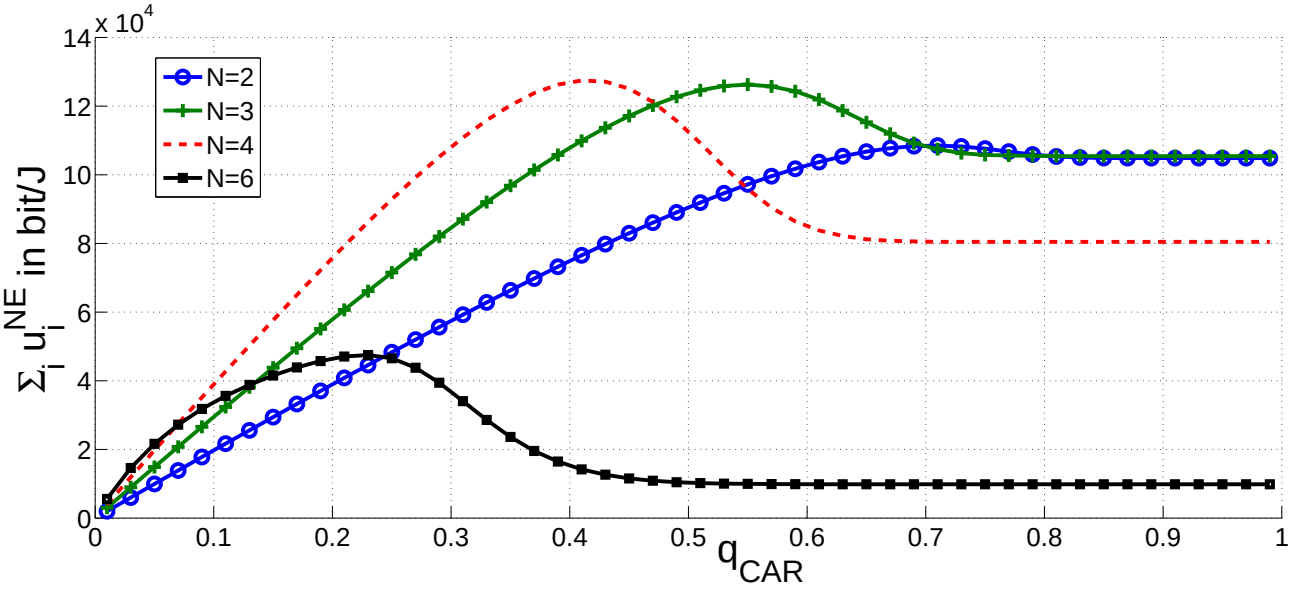


Fig. 5. CAR: Remarkably, this figure shows that a communication system can be optimized in terms of used traffic or service. Indeed, there exists an optimal packet rate at which the network EE is maximized. As the number of users increases, the $q$ that corresponds to the best equilibrium, in terms of sum-payoff, decreases. Note that this plot is for the high interference case with $\epsilon = 0.02$ resulting in a low equilibrium payoff for $N = 8$.

payoff versus $q$ for different numbers of transmitter-receiver pairs ($N \in \{2, 3, 4, 6\}$). This figure illustrates that the sum-payoff at the NE is maximized at a particular $q$ which is seen to decrease with the number of transmitters. This can be intuitively understood, as if the packet arrival rate is reduced, it is possible for more transmitters to experience the same QoS and transmit at a lower power. On the other hand, a very small $q$ implies that the network resources are not being sufficiently exploited, resulting in low efficiency.

### *D. Gains in terms of energy brought by the cross-layer approach w.r.t. the state-of-the art*

To our knowledge, existing works in the literature originating from [5] do not interpret EE maximization as energy minimization. As explained in the paper, both problems are in fact equivalent in communications systems where re-transmissions are allowed. We exploit this interpretation here to go further than just assessing the gains in terms of EE as done classically. Indeed, we assess the gain in terms of energy or average total power brought by the proposed cross-layer approach over the closest state-of-the art solution which is given in [17] (the latter is obtained by assuming $q \to 1$ whatever the actual value of $q$). For $q = 0.5$ and $q = 0.3$, Fig.

6 shows that it is possible to have improvements in terms of energy consumed by the device and not just EE. This (relative) gain can be as high as 28% for $q = 0.5$ and 42% for $q = 0.3$ in the setting under consideration. Interestingly, this gain can be obtained under the same information assumption as [17] namely, only individual SINR feedback is needed to implement the power control algorithm which provides the NE performance (after convergence). Note that in this case, $q = 1$ offers no gain as the situation is identical to that treated in [17] while $q \to 0$ would offer maximum gain.

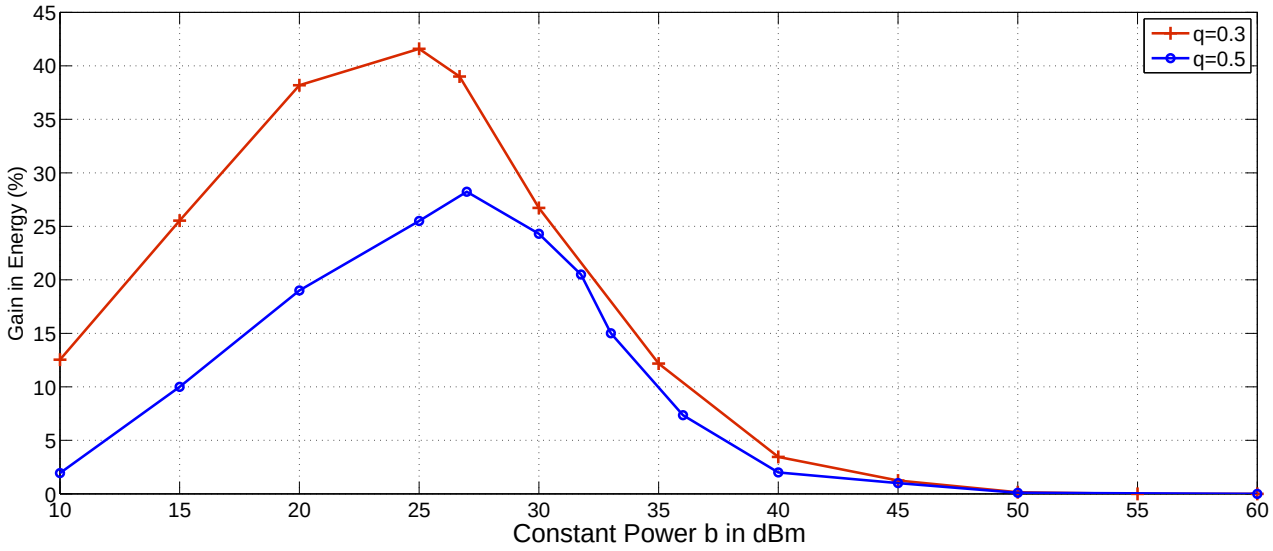


Fig. 6. CAR: Plotting the energy consumed against $b$ with $q = 0.6$ and $q = 0.3$. We compare the performance of our proposed algorithm against using the best-response dynamics algorithm from [17] where the presence of the queue is ignored.

As a second comparison in terms of energy, we compared the energy consumed by a transmitter when optimizing (4) with what would obtained by just minimizing the radiated power under an SINR constraint, which is a classical approach. Fig. 7 corresponds to the relative gain in terms of saved energy as a function of the fixed consumption cost $b$, for $q = 0.5$ and $q = 0.9$, $R = 8$ Mbps and an SINR target of 25 dB for both approaches in the single user case (interference can make achieving such a target impossible). It is seen that an energy gain of up to 80% can be achieved for sufficiently high values of $b$, which is a quite significant gain and can be easily attained in practice (e.g., maximum radiated power for femto base stations is of the order of

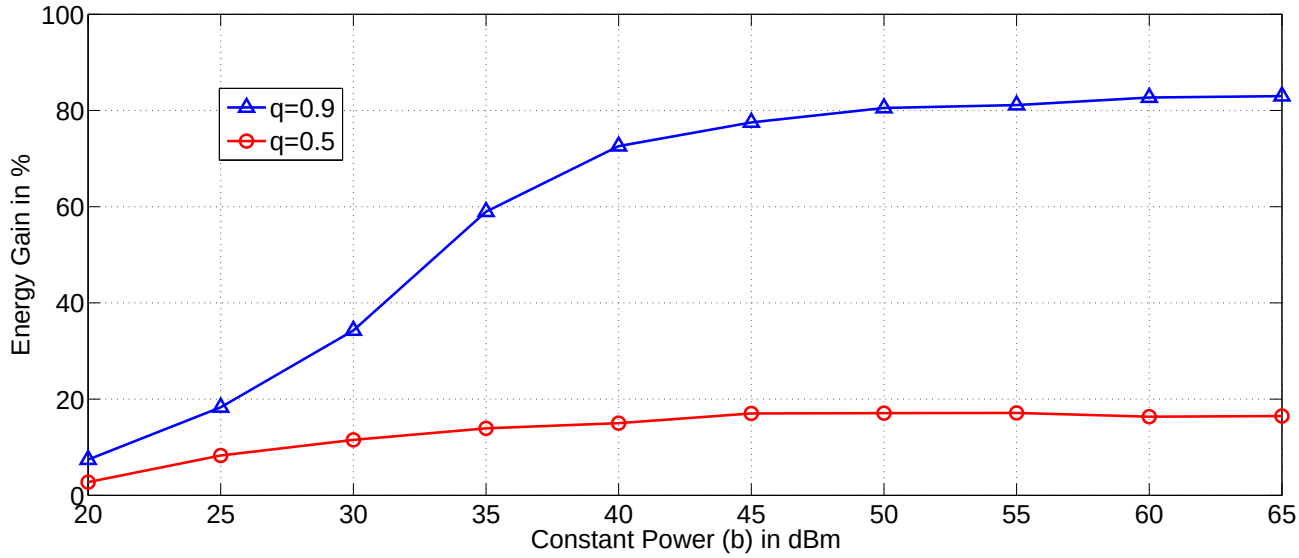


Fig. 7. CAR: Plotting the energy consumed against $b$ with $q = 0.5$ and $q = 0.9$. We compare the performance of our proposed algorithm against a scheme that just minimizes the transmit power such that the SINR $\geq 25$ dB. We show that our proposed algorithm satisfies this constraint and still consumes less energy.

one watt while the fixed consumption cost is typically of about a few watts). Note that the gain observed here is maximum when $q = 1$ as the highest transmit power is used in this case.

### *E. Influence of the packet buffer size in the AAR scenario*

So far, we have been assuming the CAR scenario. In particular, this has allowed us to study in detail the influence of the parameter $q$. But, for AAR $q$ is not fixed and varies with the SINR. Fig. 8 represents, for different numbers of transmitters ($N \in \{2, 3, 8\}$), the network sum-payoff versus the buffer size for a static channel. The influence of interference (e.g., inter-cell interference) on global energy-efficiency clearly appears. As an important comment, as this simulation shows and many other simulations confirmed this observation (including all simulations assuming CAR instead of AAR), when the buffer size is greater than 10 typically, the asymptotic regime in terms of buffer size can be assumed to be approximately reached. In practice, this means that, when K is large enough, power control policies might be approximated by implementing the power control policies obtained by assuming $K \rightarrow +\infty$, which corresponds to switching between Cases 1 and 2 (in Sec. III-B), depending on the current SINR.

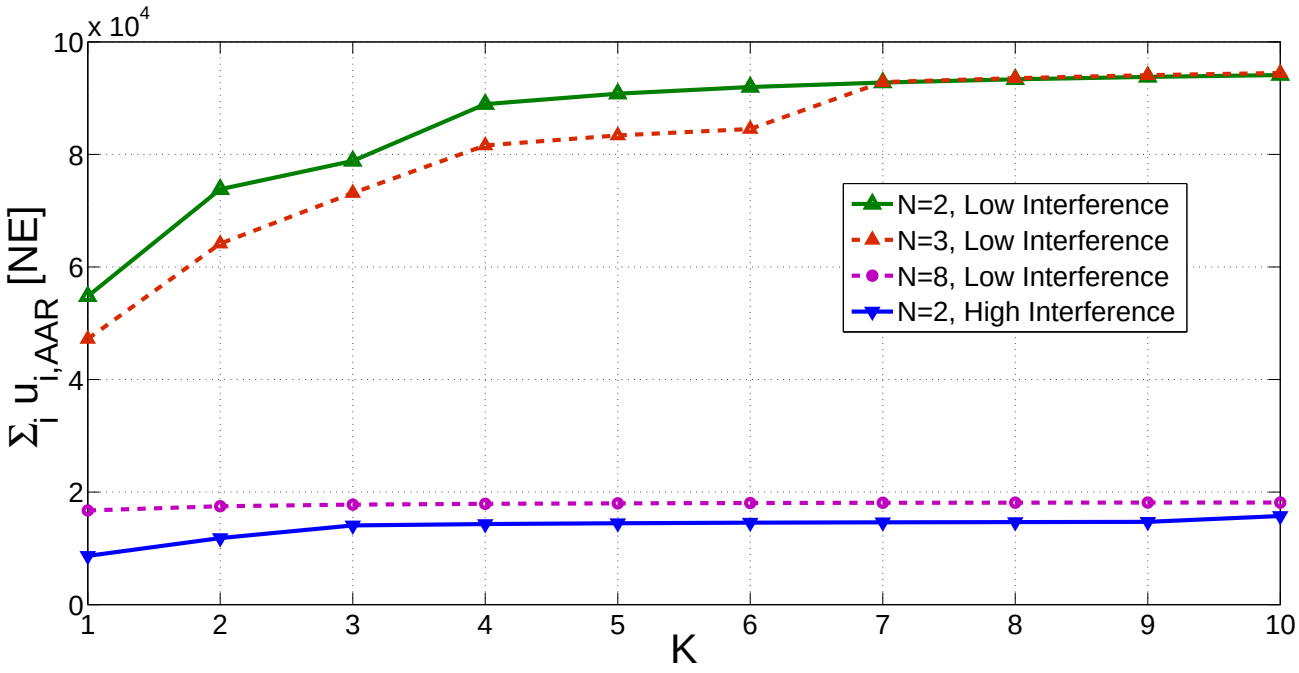


Fig. 8. AAR: We observe that the AAR sum-payoff at the NE is sensitive to the interference level, as seen from the large difference between the two user low and high interference case. With a low interference level, $N = 8$ has a higher sum-payoff than for $N = 2$ with a high interference level.

Fig. 9 studies the average gain in EE (averaged over the channel fading) when compared to that of a power control algorithm ignoring the packet level versus the interference. Here we see that when the interference is very low, the NE of the proposed scheme performs better than an algorithm that ignores the packet level. However, when under interference, the strategy under the AAR scheme would be to use a very high power as EE is individually optimized when the AAR achieves a higher packet rate. This results in a sub-optimal NE as seen in the figure when the interference is in the $[-25, 0]$ dB range. This effects indicates that the cross-layer approach might induce some performance loss w.r.t. the classical approach. The authors wanted to emphasize this negative but quite surprising result since it indicates that in distributed networks, refining the modeling aspect can sometimes induce a performance loss; this result can be related to other known paradoxes in distributed networks such as the Braess paradox [39].

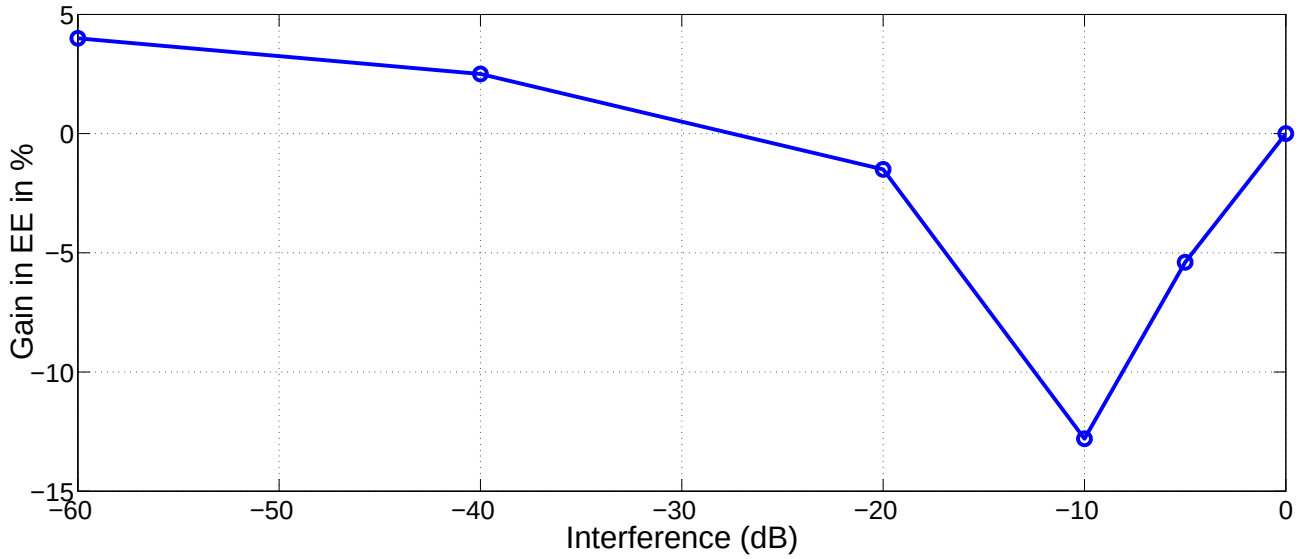


Fig. 9. AAR: Here, we plot the percentage gain in EE v.s $g_{i,j}, i \neq j$ (keeping $g_{i,i} = 1$), where the gain is calculated by comparing the EE achieved using the proposed AAR algorithm to the EE at the NE achieved by using the algorithm ignoring the packet level. We observe that in the very low interference regime, the proposed scheme outperforms the other algorithm. However in the low-medium interference region, the NE is inefficient with a high PoA and this results in poor performance.

## VI. Conclusion

Compared to the closest related works, the work reported in this paper possesses three salient features: The (possible) existence of packet buffer with finite size is taken into account; The total power consumed by the transmitter is considered; The proposed formulation considers the QoS. Remarkably, even though the derived energy-efficiency performance metric is seemingly more complex, it possesses all the main properties necessary for designing efficient distributed algorithms. Quite surprisingly, this is not only true when the packet arrival rate is constant (CAR protocol) but also when it is assumed to be adapted as a function of the SINR and the subsequent packet loss through the AAR protocol. One of the consequences of these properties is that the proposed iterative distributed power control algorithm converges towards a unique Nash equilibrium of the power control game associated with both transport protocols.

While the cross-layer generalization of energy-efficient power control is supported by several key analytical results, numerical results strongly support our approach as well. One of the key

observations made from simulations is that a distributed power control scheme can perform as well as a centralized solution in some situations; realistic settings under which the PoA is one are clearly identified. Also, it is clearly explained why maximizing EE amounts to minimizing energy in communication systems with re-transmission protocols and this key interpretation is exploited to assess the gain in terms of saved energy brought by the proposed approach.

The proposed approach might be extended in many relevant ways. To address more general wireless scenarios, the most simple extension would be to address the multi-carrier case and also the case of frequency selective channels, these extensions being potentially related. When relevant, receivers might be assumed to implement successive interference cancellation. In order to obtain more efficient equilibrium points (e.g., in the sense of the sum-payoff or a given fairness criterion), it would be of high interest to exploit a more advanced game model such as a stochastic game; this extension is especially relevant if the queue state information has to be exploited. To go further in the direction of having a very realistic wireless network model, a less trivial, but very relevant extension would be to analyze the case of a time-varying number of users. This is definitely both of practical and theoretical interest. Finally, the case of CAR and AAR transmitters simultaneously active in the network has not been studied in this paper. However, our results for quasi-concavity are independent of the protocol used by the other DMs and so the existence of the NE is guaranteed, while uniqueness is left for future extensions.

## VII. Acknowledgments

This work is a joint collaboration between Laboratoire des Signaux et Systèmes (L2S) of Supélec, Orange Labs R&D as well as University of Avignon. The work has been done as part of the Operanet 2 European project.

## Appendix A
## Proof of quasi-concavity

**Proof:** To prove that $\eta_{i,X}(\underline{p}_i)$ is quasi-concave w.r.t $p_i$, we consider its reciprocal $\frac{1}{\eta_{i,X}(\underline{p}_i)} = A_i(\underline{p}) + B_X(\gamma_i(\underline{p}))$, where;

$A_i(\underline{p}_i) = \frac{p_i}{Rf(\gamma_i(\underline{p}))}$, the physical layer factor which depends on the transmit power and the SINR.

$B_X(\gamma_i(\underline{p})) = \frac{b}{Rq_X(\gamma_i(\underline{p}))[1-\Phi_X(\gamma_i(\underline{p}))]}$, the cross-layer factor which depends on the protocol X and the SINR.

Recall that $f(\gamma)$ is increasing and initially convex for $\gamma \in [0, \gamma_+]$ and eventually concave for $\gamma \in [\gamma_+, \infty)$. So, we have that $A_i(\underline{p})$ is decreasing in the interval $\gamma_i(\underline{p}) \in [0, \gamma_+]$ and convex w.r.t $p_i$ for $\gamma_i(\underline{p}) \in [\gamma_+, \infty)$. If a function is continous and differenciable, then it is sufficient to show that it is convex at all local minima/maxima for quasi-convexity. The inverse function we consider is strictly decreasing in the interval $[0, \gamma_+]$ and thus, can not have a maxima/minima in this interval. Hence, once we prove that it is convex in the other interval, we prove quasi-concavity of the original function.

If $B_{\mathrm{X}}(\gamma_i(\underline{p}))$ is a monotonically decreasing function and is convex for $\gamma_i \geq \gamma_+$, then we have $\dfrac{1}{A_i(\underline{p}) + B_{\mathrm{X}}(\gamma_i(p))}$ quasi-concave w.r.t $p_i$ [40], [41].

From (1), $\dfrac{\partial \gamma_i(\underline{p})}{\partial p_i}$ is a constant and in the following sections, to prove that $\dfrac{\partial B_{\mathrm{X}}}{\partial p_i} < 0$ and $\dfrac{\partial^2 B_{\mathrm{X}}}{\partial p_i^2} > 0$, we just prove that:

$$\frac{\partial B_{\mathrm{X}}}{\partial \gamma_i} < 0, \quad \text{and} \quad \frac{\partial^2 B_{\mathrm{X}}}{\partial \gamma_i^2} > 0.$$

*A. The case of CAR*

In this sub-section we prove that the required conditions for quasi-concavity are satisifed under a CAR scheme. For CAR, $q_{\mathrm{CAR}}(\gamma_i(\underline{p})) = q$ is a constant. Now let us study the derivatives of the function $B_{\mathrm{CAR}}(\gamma_i)$ w.r.t $\gamma_i$;

$$\frac{\partial B_{\mathrm{CAR}}(\gamma_i)}{\partial \gamma_i} = \frac{b}{Rq(1 - \Phi_{\mathrm{CAR}}(\gamma_i))^2} \frac{\partial \Phi_{\mathrm{CAR}}(\gamma_i)}{\partial \gamma_i}, \tag{19}$$

and

$$\frac{\partial^2 B_{\mathrm{CAR}}(\gamma_i)}{\partial \gamma_i^2} = \frac{b}{Rq(1 - \Phi_{\mathrm{CAR}}(\gamma_i))^2} \left( \frac{\partial^2 \Phi_{\mathrm{CAR}}(\gamma_i)}{\partial \gamma_i^2} + \left( \frac{\partial \Phi_{\mathrm{CAR}}(\gamma_i)}{\partial p_i} \right)^2 \frac{2}{1 - \Phi_{\mathrm{CAR}}(\gamma_i)} \right). \tag{20}$$

From (19), we see that showing $\frac{\partial \Phi_{\mathrm{CAR}}}{\partial \gamma_i} < 0$ is sufficient for proving that $\frac{\partial B_{\mathrm{CAR}}(\gamma_i)}{\partial \gamma_i} < 0$ as the other terms are always positive.

Similarly, $\frac{\partial^2 \Phi_{\text{CAR}}}{\partial \gamma_i^2} > 0$ is sufficient for proving that $\frac{\partial^2 B_{\text{CAR}}(p)}{\partial \gamma_i^2} > 0$.

$$\frac{\partial \Phi_{\text{CAR}}}{\partial \gamma_i} = (1 - f(\gamma_i)) \frac{\partial \Pi_{\text{CAR}}}{\partial \gamma_i} - \Pi_{\text{CAR}} \frac{\partial f(\gamma_i)}{\partial \gamma_i} \tag{21}$$

$$\frac{\partial^2 \Phi_{\text{CAR}}}{\partial \gamma_i^2} = (1 - f(\gamma_i)) \frac{\partial^2 \Pi_{\text{CAR}}}{\partial p^2} - 2 \frac{\partial \Pi_{\text{CAR}}}{\partial \gamma_i} \frac{\partial f(\gamma_i)}{\partial \gamma_i} - \Pi_{\text{CAR}} \frac{\partial^2 f(\gamma_i)}{\partial \gamma_i^2}. \tag{22}$$

For $\frac{\partial \Phi_{\text{CAR}}}{\partial \gamma_i} < 0$, by examining (21), we see that showing $\frac{\partial \Pi_{\text{CAR}}}{\partial \gamma_i} < 0$ is sufficient.
We have $\omega_{\text{CAR}} = \frac{q}{1-q} \frac{1-f(\gamma_i)}{f(\gamma_i)}$ and so:

$$\frac{\partial \omega_{\text{CAR}}}{\partial \gamma_i} = \frac{-q}{(1-q) f(\gamma_i)^2} \frac{\partial f(\gamma_i)}{\partial \gamma_i} < 0. \tag{23}$$

It can be easily verified that $\frac{\partial^2 \omega_{\text{CAR}}}{\partial \gamma_i^2} > 0$ for $\gamma_i \geq \gamma^+$. Express $\frac{1}{\Pi_{\text{CAR}}} = 1 + \frac{1}{\omega_{\text{CAR}}} + ... + \frac{1}{\omega_{\text{CAR}}^K}$. Differentiating with respect to $\gamma_i$, we have

$$\frac{\partial \Pi_{\text{CAR}}}{\partial \gamma_i} = \Pi_{\text{CAR}}^2 \frac{\partial \omega_{\text{CAR}}}{\partial \gamma_i} \left( \frac{1}{\omega_{\text{CAR}}^2} + ... + \frac{K}{\omega_{\text{CAR}}^{K+1}} \right) < 0. \tag{24}$$

Again, it can be verified that $\frac{\partial^2 \Pi_{\text{CAR}}}{\partial \gamma_i^2} > 0$. Thus, we have:

$$\frac{\partial \Phi_{\text{CAR}}}{\partial \gamma_i} < 0 \tag{25}$$

and

$$\frac{\partial^2 \Phi_{\text{CAR}}}{\partial \gamma_i^2} > 0 \tag{26}$$

Now, following the argument from the start, we have $\eta_{\text{CAR}}(p_i, \underline{p}_{-i})$ to be quasi-concave. Since there exists some power $p_i$ for which $\eta_{i,\text{CAR}}(p_i)$ is maximized, we have proved that there exists a unique $p_i^*$ for which the EE is optimized.

■

We are able to determine the optimal power $p_i^*$ which maximize the EE function, by solving the following equation:

$$0 = -\frac{\partial \Phi_{\text{CAR}}}{\partial p_i} \left( b + \frac{p_i q (1 - \Phi_{\text{CAR}})}{f(\gamma_i(\underline{p}))} \right) + (1 - \Phi_{\text{CAR}}) \left( \frac{\partial \Phi_{\text{CAR}}}{\partial p_i} \frac{p_i}{f(\gamma_i(\underline{p}))} + \frac{\partial (p_i / f(\gamma_i(\underline{p})))}{\partial p_i} \right). \tag{27}$$

*B. The case of AAR*

In this sub-section we prove that the required conditions for quasi-concavity are satisifed under an AAR scheme. For AAR, $q_{\mathrm{AAR}}(\gamma_i(\underline{p}))$ is determined by the (10). In AAR, from (3), we know $q = g(\phi)$ and so $\Phi = g^{-1}(q)$ where $g^{-1}$ is the function inverse of $g(.)$ which is assumed to exist, be double differentiable, strictly decreasing and convex.

And so we have the following equation for $B_{\mathrm{AAR}}(\gamma_i)$:

$$B_{\mathrm{AAR}}(\gamma_i(\underline{p})) = \frac{b}{R(q_{\mathrm{AAR}}(1 - g^{-1}(q_{\mathrm{AAR}}(\gamma_i(\underline{p}))))}. \tag{28}$$

Now let us study the derivatives of the function $B_{\mathrm{AAR}}$ w.r.t $\gamma_i$ as $\frac{\partial \gamma_i}{\partial p_i} > 0$ and is a constant. So the sign of these derivatives do not change even when differentiated w.r.t $p_i$.

$$\frac{-b}{R}\frac{\partial B_{\mathrm{AAR}}(\gamma_i)}{\partial \gamma_i} = \frac{q'_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}})) - q_{\mathrm{AAR}}(\gamma_i)q'_{\mathrm{AAR}}(\gamma_i)(g^{-1})'(q_{\mathrm{AAR}})}{[q_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}}(\gamma_i)))]^2} \tag{29}$$

and

$$\begin{aligned}\frac{R}{b}\frac{\partial^2 B_{\mathrm{AAR}}(\gamma_i)}{\partial \gamma_i^2} = &\frac{[q'_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}})) - q_{\mathrm{AAR}}(\gamma_i)q'_{\mathrm{AAR}}(\gamma_i)(g^{-1})'(q_{\mathrm{AAR}})]^2}{[q_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}}(\gamma_i)))]^3} \\ &- \frac{[q''_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}})) - 2q'_{\mathrm{AAR}}(\gamma_i)^2(g^{-1})'(q_{\mathrm{AAR}})]}{[q_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}}(\gamma_i)))]^2} + \\ &\frac{q_{\mathrm{AAR}}(\gamma_i)[q''_{\mathrm{AAR}}(\gamma_i)(g^{-1})'(q_{\mathrm{AAR}}) + q'_{\mathrm{AAR}}(\gamma_i)^2(g^{-1})''(q_{\mathrm{AAR}})]}{[q_{\mathrm{AAR}}(\gamma_i)(1 - g^{-1}(q_{\mathrm{AAR}}(\gamma_i)))]^2}\end{aligned} \tag{30}$$

From the above expressions, we deduce that the requirements for $B_{\mathrm{AAR}}$ to be decreasing and convex, knowing $(g^{-1})'(q_{\mathrm{AAR}}) \leq 0$ and $(g^{-1})''(q_{\mathrm{AAR}}) \geq 0$ is that

$$\frac{\partial q_{\mathrm{AAR}}(\gamma_i)}{\partial \gamma_i} \geq 0. \tag{31}$$

$$\frac{\partial^2 q_{\mathrm{AAR}}(\gamma_i)}{\partial \gamma_i^2} \leq 0. \tag{32}$$

Now, we exploit the AAR based fixed point equation:

$$g^{-1}(q_{\mathrm{AAR}}(\gamma_i)) = \frac{1 - f(\gamma_i)}{1 + \omega_{\mathrm{AAR}}(\gamma_i)^{-1} + \omega_{\mathrm{AAR}}(\gamma_i)^{-2} + \cdots + \omega_{\mathrm{AAR}}(\gamma_i)^{-K}}. \tag{33}$$

Differentiating (33) w.r.t $\gamma_i$ once, we get that $\frac{\partial q_{\text{AAR}}(\gamma_i)}{\partial \gamma_i} \geq 0$ and differentiating twice, we get that the inequality (32) is satisfied for $\gamma_i \geq \gamma_+$.

Thus we have shown that $\eta_{i,\text{AAR}}$ is quasi-concave w.r.t $p_i$ for the AAR case.

■

## Appendix B
## Proof of existence of a pure NE in $\mathcal{G}_{\text{CAR}}$

Proof: Here we use the result in [21] (Corollary of Theorem 2 in [21]) which states that:

*Theorem 1:* $\forall i$, $\mathcal{A}_i \subset \mathbb{R}^N$ be non-empty, convex and compact, and let $U_i : \mathcal{A}_i \rightarrow \mathbb{R}$ be quasi-concave in $u_i$ and upper semi-continuous. Define $V_i(a_{-i}) = \max[U_i(a_i, a_{-i})]$. If $V_i$ is lower semi-continuous in $a_{-i}$ then, the game $(\mathcal{N}, \mathcal{A}, U)$ has a pure NE.

In this section, we prove that $u_{i,\text{CAR}}$ is upper semi-continuous and that the newly defined function $V_i(\underline{p}_{-i}) = \max[u_{i,\text{CAR}}(p_i, \underline{p}_{-i})]$ is lower semi-continuous. Here, we identify $V_i$ as the payoff of the best-response, i.e., $V_i(\underline{p}_{-i}) = u_{i,\text{CAR}}(\text{BR}_{i,\text{CAR}}(\underline{p}_{-i}), \underline{p}_{-i})$. Studying the specific cases of $\mathcal{G}_{\text{CAR}}$: Note that $\frac{q[1-\Phi_{\text{CAR}}(\gamma_i)]R}{b+P_{\max}} < \eta_{i,\text{CAR}}(\underline{p})$; $\forall (p_i < P_{\max})$. Define $p_i^+(\underline{p}_{-i}) : \Phi_{\text{CAR}}(\gamma_i(p_i^+(\underline{p}_{-i}), \underline{p}_{-i})) = \epsilon$ and $p_i^*(\underline{p}_{-i}) : \frac{\partial \eta_{\text{CAR}}(p_i^*(\underline{p}_{-i}), \underline{p}_{-i})}{\partial p_i} = 0$. There are several cases possible:

1) $p_i^+(\underline{p}_{-i}) \geq P_{\max}$: Here, $u_{i,\text{CAR}}(\underline{p})$ is a strictly increasing function and maximizes at $P_{\max}$.
2) $p_i^*(\underline{p}_{-i}) \leq p_i^+(\underline{p}_{-i}) < P_{\max}$: Here $u_{i,\text{CAR}}(\underline{p})$ is a strictly increasing function in the interval $p_i = [0, p_i^+(\underline{p}_{-i}))$ and after a point of discontinuity at $p_i^+(\underline{p}_{-i})$, is strictly decreasing in the interval $[p_i^+(\underline{p}_{-i}), P_{\max}]$. So $u_{i,\text{CAR}}$ maximizes at $p_i^+(\underline{p}_{-i})$.
3) $p_i^+(\underline{p}_{-i}) < p_i^*(\underline{p}_{-i})$: $u_{i,\text{CAR}}(\underline{p})$ is strictly increasing in the interval $[0, p_i^+(\underline{p}_{-i}))$ and after a point of discontinuity at $u_i^+$, is quasi-concave in the interval $[p_i^+(\underline{p}_{-i}), P_{\max}]$. So $u_{i,\text{CAR}}$ maximizes at $p_i^*(\underline{p}_{-i})$.

In all three cases, $u_{i,\text{CAR}}(\underline{p})$ is upper semi-continuous and quasi-concave (See Appendix A for properties of $\eta_{i,\text{CAR}}$). Also, $u_i(p_i^*(\underline{p}_{-i}), \underline{p}_{-i})$ is a continuous function in $\underline{p}_{-i}$ and for small $\underline{p}_{-i}$, $\text{BR}_{i,\text{CAR}}(\underline{p}_{-i}) = p_i^*$ . After the point where $p_i^*(\underline{p}_{-i}) = p_i^+(\underline{p}_{-i})$, as $\underline{p}_{-i}$ increases further, $\text{BR}_{i,\text{CAR}}(\underline{p}_{-i}) = p_i^+$ which is also continuous. And so $\text{BR}_{i,\text{CAR}}(\underline{p}_{-i})$ is in fact continuous and increasing in $\underline{p}_{-i}$.
$V_i$ is continuous in the interval $\underline{p}_{-i} \leq \underline{p}_{-i}^+ : p_i^+(\underline{p}_{-i}^+) = P_{\max}$ and is given by $\eta_{i,\text{CAR}}$. For $\underline{p}_{-i} > \underline{p}_{-i}^+$, $V_i$ jumps down according to the definition in (15) and is thus, lower semi-continuous. Using Theorem 1, we have the result that the Game admits a pure NE.

■

## Appendix C
## Best-responses for $\mathcal{G}_X$ are standard

This proof holds for both cases of CAR and AAR. Here, we prove that the best-responses are monotonic and scalable (standard) if $\eta_{i,X}(\underline{p})$ is quasi-concave w.r.t $p_i$. A function $F(x)$ is standard, if it satisfies the following properties:

1) $F(x_1) \geq F(x_2)$, if $x_1 \geq x_2$: Monotonic
2) $F(\lambda x) \leq \lambda F(x)$, if $\lambda \geq 1$: Scalable

Consider $\underline{P}_{-j} := \lambda \underline{p}_{-j}$, where $\lambda > 1$.

$\mathrm{BR}_{j,X}(\underline{p}_{-j})$ can be calculated by solving for $\gamma_j^*$ in

$$0 = \frac{\partial A(\rho_j^*, p_j)}{\partial p_j} + \frac{\partial B(\gamma_j^*)}{\partial p_j} \tag{34}$$

which can be simplified to

$$0 = \hat{A}(\gamma_j^*) + C(\underline{p}_{-j})\hat{B}(\gamma_j^*) \tag{35}$$

where $\hat{A}(\gamma_j^*) = \dfrac{f(\gamma_j^*) - f'(\gamma_j^*)\gamma_j^*}{f^2(\gamma_j^*)}$, $C(\underline{p}_{-j}) = \dfrac{b}{\sigma^2 + \sum_{i \neq j} h_i p_i}$ and $\hat{B}(\gamma_j^*) = \dfrac{\partial B(\gamma_j^*)}{\partial \gamma_j}$. As $A$ is convex and $\hat{B}$ negative, (proved in App. A), we can conclude that $\gamma_j^*(\underline{P}_{-j}) \leq \gamma_j^*(\underline{p}_{-j})$. Thus, $\mathrm{BR}_{j,X}(\underline{P}_{-j}) \leq \lambda \mathrm{BR}_{j,X}(\underline{p}_{-j})$ as $p_j = \gamma_j(\sigma^2 + \sum_{i \neq j} g_{ij} p_i)$. Therefore the best-responses for the game are scalable.

Now consider $\underline{P}_{-j} \geq \lambda \underline{p}_{-j}$ such that $(\sigma^2 + \sum_{i \neq j} g_{ij} P_i) = \lambda(\sigma^2 + \sum_{i \neq j} g_{ij} p_i)$ , where $\lambda > 1$. Let $\gamma_j^{**}$ (where $\mathrm{BR}_{j,X}(\underline{p}_{-j}^{\cdot}) = \gamma_j^{**}(\sigma^2 + \sum_{i \neq j} h_i \underline{P}_i)$ is the best-response) satisfy

$$0 = \hat{A}(\gamma_j^{**}) + \frac{C(\underline{p}_{-j})}{\lambda}\hat{B}(\gamma_j^{**}) \tag{36}$$

Now replace $\gamma_j^{**}$ by $\frac{\gamma_j^*}{\lambda}$ and we have

$$\hat{A}(\gamma_j^* \lambda^{-1}) + \tfrac{C(\underline{p}_{-j})}{\lambda}\hat{B}(\gamma_j^* \lambda^{-1}) \leq \tag{37}$$

$$\lambda^{-1}\hat{A}(\gamma_j^*) + \tfrac{C(\underline{p}_{-j})}{\lambda^2}\hat{B}(\gamma_j^*) \leq \tag{38}$$

$$\tfrac{\hat{A}(\gamma_j^*) + C(\underline{p}_{-j})\hat{B}(\gamma_j^*)}{\lambda} \leq 0. \tag{39}$$

The above inequalities are a result of the properties of $\hat{A}$ and $\hat{B}$ given in App. A.

Which shows that $\gamma_j^{**} \geq \frac{\gamma_j^*}{\lambda}$ and thus, $\mathrm{BR}_{j,X}(\underline{P}_{-j}) \geq \mathrm{BR}_{j,X}(\underline{p}_{-j})$ and hence the best-responses are monotonic.

As all the powers played are positive, the best-response functions satisfy the two requirements and so are standard functions. ■

## References


[1] D. Lister, "An operators view on green radio", Proc. IEEE Internet. Conf. on Comm. Workshops (ICC Workshops 2009), 1st Int. Workshop on Green Comm. (GreenComm 09) (Dresden, Ger., 2009).

[2] J. Palicot and C. Roland, "On the use of cognitive radio for decreasing the electromagnetic radiations", URSI 05, XXVIII General Assembly, New Delhi, India, October 23-29, 2005.

[3] GreenTouch, "Communications Turns Totally Green", Press Release, Jan. 11, 2010.

[4] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks", IEEE Vehicular Technology Magazine, vol. 6, no. 1, Mar. 2011.

[5] D. J. Goodman, and N. Mandayam, "Power control for wireless data", IEEE Personal Comm., vol. 7, pp. 48-54, Apr. 2000.

[6] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost.", Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on. IEEE, 2010.

[7] F. Meshkati, M. Chiang, H. Poor and S. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems", in Journal on Selected Areas in Comm., vol. 24, no. 6, 2006.

[8] F. Meshkati, H. Poor, S. Schwartz, "Energy efficiency - delay tradeoffs in CDMA networks: A game-theoretic approach", in Trans. on Information Theory, vol. 55, no. 7, 2009.

[9] S. Buzzi, G. Colavolpe, D. Saturnino, and A. Zappone, "Potential games for energy-efficient power control and subcarrier allocation in uplink multicell OFDMA systems," IEEE Journal on Selected Topics in Signal Processing (special issue on Game Theory in Signal Processing), Vol. 6, pp. 89 - 103, 2012.

[10] G. Bacci, A. Bulzomato, M. Luise, "Uplink power control and subcarrier assignment for an OFDMA multicellular network based on game theory", in Proc. Int. Conf. on Performance Evaluation Methodologies and Tools (ValueTools), Paris, France, May 2011.

[11] G. Bacci, L. Sanguinetti, M. Luise, H. V. Poor, "A game-theoretic approach for energy-efficient contention-based synchronization in OFDMA systems," IEEE Trans. Signal Process., vol. 61, no. 5, pp. 1258-1271, Mar. 2013.

[12] S. M. Perlaza, E. V. Belmega, S. Lasaulce, and M. Debbah, "On the base station selection and base station sharing in self-configuring networks", Proc. Int. Conf. on Performance Evaluation Methodologies and Tools (VALUETOOLS), Pisa, Italy, invited paper, Oct. 2009.

[13] E. V. Belmega and S. Lasaulce, "Energy-efficient precoding for multiple-antenna terminals", IEEE. Trans. on Signal Processing, 59, 1, Jan. 2011.

[14] M. Le Treust, S. Lasaulce, Y. Hayel, and G. He, "Green power control in cognitive wireless networks", IEEE Trans. on Vehicular Technology, Vol. 62, No. 4, pp. 1741–1754, April 2013.

[15] S. Verdú, "Spectral efficiency in the wideband regime", IEEE Trans. on Information Theory, vol. 48, no. 6, pp. 1319-1343, Jun. 2002.

[16] F. Richter, A. J. Fehske, G. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks", Proc. of VTC Fall'2009.

[17] S. M. Betz, H. V. Poor, "Energy efficient communications in CDMA networks: A game Theoretic analysis considering operating costs", IEEE Trans. on Signal Processing, 56(10-2): 5181-5190, 2008.

[18] H. Alvestrand, S. Holmer, and H. Lundin, A Google Congestion Control Algorithm for Real-Time Communication on the World Wide Web, 2013, IETF Internet Draft.

[19] V. S. Varma, Y. Hayel, S. Lasaulce, S.E. Elayoubi, M. Debbah, "Cross-layer design for green power control", Proc. of IEEE Int. Conf. on Comm. (ICC), Ottawa, 2012.

[20] A. Zappone, Z. Chong, E. Jorswieck, S. Buzzi, "Energy-aware competitive power control in relay-assisted interference wireless networks," IEEE Trans. on Wireless Communications, Vol. 12, pp. 1860–1871, April 2013.

[21] P. Dasgupta, and E. Maskin, "The existence of equilibrium in discontinuous economic games", I: Theory, Review of Economic Studies LIII, l-26, 1986.

[22] R. D. Yates, "A framework for uplink power control in cellular radio systems", IEEE J. Sel. Areas Comm., vol. 13, no. 9, pp. 1341-1347, Sep. 1995.

[23] S. Lasaulce, and H. Tembine, "Game Theory and Learning for Wireless Networks: Fundamentals and Applications", Academic Press, Elsevier, August 2011.

[24] W. Yu, G. Ginis, and J. M. Cio, "Distributed multiuser power control for digital subscriber lines", IEEE J. Sel. Areas Comm., vol. 20, no. 5, pp. 1105-1115, May 2002.

[25] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: a unified view", IEEE J. Sel. Areas Comm., vol. 26, no. 7, pp. 1089-1103, September 2008.

[26] A. B. Carleial, "Interference channels", IEEE Trans. on Information Theory, vol. 24, no. 1, pp. 60-70, 1978.

[27] H. Takagi, L. Kleinrock, "Output processes in contention packet broadcasting systems", IEEE Trans. on Comm., vol. 33, no. 11, pp. 1191-1199, 1985.

[28] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, "Modeling TCP Throughput: A simple model and its empirical validation", SIGCOMM Comput Comm. Rev, Volume: 28, Issue: 4, Publisher: ACM, Pages: 303-314, 1998.

[29] V. Singh, A. Lozano, and J. Ott, Performance Analysis of Receive-Side Real-Time Congestion Control for WebRTC, In Proc. of IEEE Packet Video (Vol. 2013).

[30] S. Lasaulce, Y. Hayel, R. El Azouzi, and M. Debbah, "Introducing hierarchy in energy games", IEEE Transactions on Wireless Communications," Vol. 8, No. 7, pp. 3833–3843, Jul. 2009.

[31] R. W. Wolff, "Stochastic modeling and the theory of queues", Englewood Cliffs, NJ: Prentice-Hall, 1989.

[32] W. Eric, "Mean-value theorem", MathWorld, Wolfram Research.

[33] M. Le Treust, S. Lasaulce, and M. Debbah, "Implicit cooperation in distributed energy-efficient networks", IEEE Proc. of the International Symposium on Communications, Control and Signal Processing (ISCCSP), Limasol, Greece, pp. 1–4, Mar. 2010.

[34] A. Destounis, M. Assaad, M. Debbah, B. Sayadi, and A. Feki, "On queue-aware power control in interfering wireless links: Heavy traffic asymptotic modelling and application in QoS provisioning", IEEE Transactions on Mobile Computing 99.1 (2013): 1.

[35] D. Fudenberg and J. Tirole, "Game theory", MIT Press, 1991.

[36] A. Cournot, "Recherches sur les principes mathématiques de la théorie des richesses", 1838 (Re-edited by Mac Millian in 1987).

[37] D. P. Bertsekas, "Nonlinear programming", Athena Scientific, 1995.

[38] C. Papadimitriou, "Algorithms, games, and the internet." Proceedings of the thirty-third annual ACM symposium on Theory of computing. ACM, 2001.

[39] D. Braess, "ber ein Paradoxon aus der Verkehrsplanung." Unternehmensforschung 12.1 (1968): 258-268.

[40] S. Boyd, and L. Vandenberghe, "Convex optimization", Cambridge University Press, 2004.

[41] V. Rodriguez, "An analytical foundation for resource management in wireless communication", IEEE Proc. of Global Comm. Conf. (GLOBECOM), San Francisco, CA, USA, pp. 898-902, Dec. 2003.

# Appendix C

# Papers on other techniques

## C.1 VALUETOOLS-2012

• V.S Varma, S.E. Elayoubi, S. Lasaulce and M. Debbah, "A Flow Level Perspective on Base Station Power Allocation in Green Networks", ACM International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2012).

# A Flow Level Perspective on Base Station Power Allocation in Green Networks

Vineeth S Varma[1,2], Salah Eddine Elayoubi[1], Samson Lasaulce[2] and Merouane Debbah[3]

[1]Orange Labs
92130 Issy Les Moulineaux
France
vineeth.svarma@orange.com
salaheddine.elayoubi@orange.com

[2]LSS, SUPELEC
91192 Gif sur Yvette
France
Samson.lasaulce@lss.supelec.fr

[3]Alcatel Lucent Chair
SUPELEC
91192 Gif sur Yvette
France
Merouane.debbah@supelec.fr

***Abstract*—In this work, we propose a novel power allocation mechanism which allows one to optimize the energy-efficiency of base stations operating in the downlink. The energy-efficiency refers to the amount of bits that can be transmitted by the base station per unit of energy consumed. This work studies the impact of flow-level dynamics on the energy efficiency of base stations, by considering user arrivals and departures. Our proposed power allocation scheme optimizes the energy-efficiency, accounting for the dynamic nature of users (referred to as the global energy-efficiency). We emphasize our numerical results that study the influence of the radio conditions, transmit power and the user traffic on the energy-efficiency in an LTE compliant framework. Finally, we show that the power allocation scheme that considers traffic dynamics, is significantly different from the power allocation scheme when the number of users is considered as constant, and that it has a better performance.**

## I. Introduction

For a long time, the problem of energy in the field of communications revolved around autonomous, embarked, or mobile terminals. Nowadays, with the existence of large networks involving both fixed and nomadic terminals and the larger data rates supported, the energy consumed by the fixed infrastructure has also become a central issue for communications engineers [1]. As stated by the project GreenTouch, the telecommunications industry currently account for $2\%$ of the global carbon footprint, of which the major portion comes through the energy consumed at base stations [2]. This has led to the growing awareness for the need to reduce energy consumption as well as to optimize the use of energy in order to gain maximum benefit out of every unit of energy spent. The present work falls into this framework, more specifically, our goal is to devise the power allocation schemes for base stations in green wireless networks with the focus on downlink. The novelty of this work is in treating the problem of energy-efficiency and power allocation for dynamic users, i.e for users who, like in most practical cases, arrive randomly with a finite workload and depart after finishing it.

Among the pioneering works on energy-efficient power control is the work by Goodman et al [5] in which the authors define the energy-efficiency of a communication as the ratio of the net data rate to the radiated power; the corresponding quantity is a measure of the average number of bits successfully received per joule consumed at the transmitter. This metric has motivated many works. A survey on works that deal with this metric can be found in [6]. Other works like [8] deal with the energy-efficiency metric, and it is applied to the problem of distributed power allocation in multi-carrier CDMA (code division multiple access systems) systems, in [4] it is used to model the users delay requirements in energy-efficient systems.

Summarizing the literature overview for energy-efficiency optimization, we conclude that although several works consider deal with this problem, they do not take into account several key-aspects of the network. First, in the definition of energy-efficiency, the number of users in the system is fixed, corresponding to a full buffer traffic model. In a real system, users arrive and depart and the number of users in the system is a dynamic quantity. Secondly, the transmission cost usually corresponds to the radiated power that is, the power of the radio-frequency signals. In this paper, we propose a power allocation scheme that responds to these two needs: considering the dynamic behaviour of users and taking into account the whole power consumption and not only the radiated power. This work uses a cross-layer approach, which deals with both the Media Access Control (MAC) layer, as well as the flow level (user arrivals and departures) in Orthogonal Frequency-Division Multiple Access (OFDMA) systems that are LTE compliant. Similar cross-layer approaches have been used in works like [9] and [7], but the metric used is often the capacity or data rates maximized under power constraints, while in this work we deal here with energy-efficiency optimization.

The original contributions of this paper are summarized as follows:

1) We consider a new energy efficiency metric that accounts for the overall power consumption of the base station, including common channel and fixed consumption parts.
2) We derive an optimal power allocation scheme that maximizes the energy efficiency, while preserving Quality of Service (QoS).
3) We show that the power allocation that considers the dynamic behavior of users is significantly different from the scheme optimized locally for each state of the network. In addition to that, the former performs better than the latter. To the best of our knowledge, this is the first time where such a flow level power allocation scheme is derived.

This paper is structured as follows. In Section II, we present the system model and define the proposed performance metric. In Section III, we derive the optimal power allocation scheme when supposing that the number of users is fixed. Section IV shows how to deal with the dynamic behavior of users. Section V provides numerical results comparing both approaches (local vs. global optimization). Finally, we conclude the paper and suggest some possible extensions to this work.

## II. System model

### A. System description

We consider a transmitting base station with buffers of infinite (or very large) size. The base station sends packets into a queue for each user which is stored in these buffers. The packets arrive at each time slot $T_P$ (expressed in seconds), each packet being of size $S_p$ (expressed in bits). The data rate $R_p$ is equal to $\frac{S_p}{T_P}$. The throughput when using all the available bandwidth is denoted by $R(\rho)$ (expressed in bits per second), when the receiver has an average signal to interference plus noise ratio (SINR) of $\rho$. This SINR depends directly on the transmit power $P$ (expressed in Watts) as $\rho = \frac{P}{\sigma^2}$. Here $\sigma^2$ represents the average noise for a given radio condition (expressed in Watts) and it depends on the distance of the receiver from the base station. Note that in this work, the effects of fast fading are not studied and we just consider the average SINR.

All packets of a user are assumed of the same size and the average throughput on the radio interface, when the queue for the corresponding user is active, is denoted by $R_a(\rho)$ (expressed in bits per second) which depends on the bandwidth available. When all the packets in the queue are transmitted the queue becomes empty and inactive. We assume that the transmitter always transmits packets while the queue is not empty. Each packet stored in the buffer is a collection of frames that are transmitted over the symbol time $T_s$ (expressed in seconds). Each frame is transmitted or retransmitted till it goes through and an acknowledgment is received. With these assumptions we proceed to calculate the average packet duration $T_d$ in the buffer.

$$T_d = \frac{S_p}{R_a(\rho)} \tag{1}$$

If this duration exceeds $T_P$, the time by which the next packet arrives, the queue size becomes infinite and the transmitter is always on. Otherwise, the probability of the transmitter to be active ($\Phi(\rho)$) is given by the ratio of $T_d$ to $T_P$. Thus we have:

$$\Phi(\rho) = \max\left(\frac{R_p}{R_a(\rho)}, 1\right) \tag{2}$$

In this work, we focus on an OFDMA system that suits LTE standards, and obtain the throughput $R(\rho)$ by link level simulations as described in [3]. The values taken for $R(\rho)$ from [3], are in fact, averaged over the fast fading and are thus suitable for our model. When there are several users in the network, the available bandwidth is divided among the active users. We assume the bandwidth allocation to be equal among all users and this implies that if $N$ users are all active and experience the same radio conditions, the throughput is reduced to $\frac{R(\rho)}{N}$.

### B. Proposed performance metric

In the broadcast channel there are multiple users that have to be served. In practice, users arrive randomly, and depart once they finish downloading their requested data. New arrivals are blocked when the total number of users crosses a certain limit defined by the base station. Each user may experience a different radio condition from its peers.

For convenience, we divide the area covered by the base station into "zones". Every user in the same zone, experiences the same radio conditions. This implies that if the base station transmits at a certain power, then all the users in the same zone experience the same SINR. The radio conditions are determined by the average distance of the zone to the base station. If we have $M$ zones in total, we can define $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\}$ as the channel conditions for each zone. We then define the "state" of the system $\bar{\mathbf{s}} = \{N_1, N_2, \cdots, N_M\}$. The state $\bar{\mathbf{s}}$ represents the number of users in each zone. For example if there are two zones, and there are no users the state is $\{0, 0\}$. When a user arrives to zone one, the state becomes $\{1, 0\}$.

For a state $\bar{\mathbf{s}} = \{N_1, N_2, \cdots, N_M\}$, the power allocation scheme defined as $\mathbf{P} = \{P_1, P_2, \cdots, P_M\}$ results in an SINR distribution of $\hat{\rho} = \{\rho_1, \rho_2, \cdots \rho_M\}$ among the zones 1 to $M$, where $\rho_j = \frac{P_j}{\sigma_j^2}$.

First, we define the notion of energy-efficiency for a given state or the "local" energy-efficiency. This is useful as in practice, the base station can easily measure this quantity only for a given state as it is unable to predict when a new user will arrive. The "global" energy-efficiency defined as the average of the energy-efficiency in each state weighted by their probabilities.

If there is always one and only one user, the energy-efficiency can be defined based on [5] and other works as

$$\eta_{SU} = \frac{R(\rho)\Phi(\rho)}{b + P\Phi(\rho)} \tag{3}$$

where $b$ is the constant power consumed by the base station while serving at least one user[1]. The proposed form is easy to interpret as $R(\rho)$ represents the average throughput when the transmitter is active and $P$ is the cost when the transmitter is active.

When the system is state $\bar{\mathbf{s}}$, the energy-efficiency is defined as:

$$\eta_{\bar{\mathbf{s}}}(\mathbf{P}) = \frac{\bar{R}_{\bar{\mathbf{s}}}(\hat{\rho})}{\bar{P}_{\bar{\mathbf{s}}}(\mathbf{P})} \tag{4}$$

where $\bar{R}_{\bar{\mathbf{s}}}$ and $\bar{P}_{\bar{\mathbf{s}}}$ represent the total throughput and power consumed respectively in state $\bar{\mathbf{s}}$.

[1]This cost can have several origins like energy spent on the power amplifier, computation, cooling mechanisms etc. Details of the power consumption model are given in [1].

When the number of users is random, then the global energy-efficiency function is defined as:

$$\hat{\eta} = \sum_{\bar{\mathbf{s}}} \frac{\pi(\bar{\mathbf{s}})\bar{R}_{\bar{\mathbf{s}}}}{\bar{P}_{\bar{\mathbf{s}}}} \tag{5}$$

Where $\pi(\bar{\mathbf{s}})$ is the probability of finding the base station at state $\bar{\mathbf{s}}$ of user distribution. The global energy-efficiency could alternately be defined as ratio of the total throughput over all states to the total power over all states. However, in practice, calculating the energy-efficiency for each state and taking the average, is easier and more reasonable. The goal of this work is to improve the above defined energy-efficiency of a transmitting base station.

This metric can be physically interpreted as the average number of bits that can be transmitted by spending one Joule of energy. Alternately, the average power cost of the base station can be written an $\frac{\text{Traffic}}{\eta}$. Hence, optimizing the global energy-efficiency amounts to minimizing the average power consumption of the base station.

## III. Optimal power allocation for a fixed number of users

In this section we consider the case where the number of users is fixed. We will refer to the optimization of the metric defined in this section as "local" optimization as it deals with the optimization of a single state of the wireless network. When the state of the network is given, we know the number of users in each zone and can thus calculate the relevant information required to obtain and optimize the energy-efficiency. For our calculations we assume a knowledge of the average noise levels for each zone, i.e $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\}$ are known.

### A. Homogeneous radio conditions

First, we consider the problem where all users experience the same average SINR, as the model is easier to be understood; the case of heterogeneous SINRs will be exposed next. Let the total number of users in the cell be $N$. As all the users experience the same radio conditions, $\bar{\mathbf{s}} = \{N\}$. In this case if we define the average throughput experienced by any queue as $R_a$, we can derive:

$$R_a(\rho) = \sum_{i=0}^{N-1} \binom{N-1}{i} \Phi(\rho)^i (1-\Phi(\rho))^{N-1-i} \frac{R(\rho)}{i+1} \tag{6}$$

where $\Phi(\rho)$ denotes the probability that any of the $N$ users are actively being served and is given as in equation 2. The summation is upto $N-1$ as $R_a$ is the throughput experienced by an active user, and so we consider the remaining $N-1$ users. The $R_a$ for every user is identical as all users experience the same SINR for the same transmit power. This symmetry can be exploited to conclude that the transmit power for each user will be equal when optimized. Note that $R_a(\rho)$ depends on $\Phi(\rho)$ and $\Phi(\rho)$ depends on $R_a(\rho)$ leading to a fixed point equation.

Clearly if $N$ is large enough, then the demand in data rate will exceed the maximum available throughput and $\Phi(\rho)$ becomes 1. On the other hand, if $N$ is small enough, the users may transmit their data faster than the packet arrival speed causing the queue to empty occasionally. In this period, other users can take advantage of the excess bandwidth.

From $\Phi(\rho)$, the total power consumed can be calculated as

$$\bar{P}_{\bar{\mathbf{s}}} = b + P(1-(1-\Phi(\rho))^N) \tag{7}$$

Here $(1-\Phi(\rho))^N$ is the probability of all queues being empty. If any queue is active the power consumed is $P$. The total throughput is $\bar{R}_{\bar{\mathbf{s}}} = N\Phi(\rho)R_a$ leading to an energy-efficiency of

$$\eta_{\bar{\mathbf{s}}} = \frac{N\Phi(\rho)R_a(\rho)}{b + P(1-\Phi(\rho))^N} \tag{8}$$

### B. Heterogeneous radio conditions

Consider a more realistic setting where users experience different radio conditions in each zone. Denoting the average throughput experienced by zone $j$ as $R_{a:j}$, we can compute

$$\begin{aligned} R_{a:j}(\hat{\rho}) = R(\rho_j) &\sum_{i_1=0}^{N_1} \sum_{i_2=0}^{N_2} \cdots \sum_{i_j=0}^{N_j-1} \cdots \sum_{i_M=0}^{N_M} \binom{N_1}{i_i} \\ &\times \binom{N_2}{i_2} \cdots \times \binom{N_j-1}{i_j} \times \cdots \times \binom{N_M}{i_M} \times (\Phi_1(\hat{\rho}))^{i_1} \\ &\times (\Phi_2(\hat{\rho}))^{i_2} \times \cdots \times (\Phi_M(\hat{\rho}))^{i_M} \times (1-\Phi_1(\hat{\rho}))^{N_1-i_1} \\ &\times (1-\Phi_2(\hat{\rho}))^{N_2-i_2} \times \cdots \times (1-\Phi_j(\hat{\rho}))^{N_j-i_j-1} \\ &\times \cdots \times (1-\Phi_M(\hat{\rho}))^{N_M-i_M} \times \frac{1}{i_1+i_2+\cdots+i_M+1} \end{aligned} \tag{9}$$

where

$$\Phi(\hat{\rho})_j = \max\left(\frac{R_p}{R_{a:j}(\hat{\rho})}, 1\right) \tag{10}$$

Leading to a set of fixed point equations that can be solved to calculate all $R_{a:j}(\hat{\rho})$ for a given $\mathbf{P}$. Equation (9) is similar to (6), but considers the presence of users in other zones as well. The average power can be calculated as

$$\begin{aligned} \bar{P}_{\bar{\mathbf{s}}}(\mathbf{P}) = b + &\sum_{i_1=0}^{N_1} \cdots \sum_{i_M=0}^{N_M} (1-\delta(i_1+\cdots+i_M)) \\ &\times (\Phi_1(\hat{\rho}))^{i_1} \times \cdots \times (\Phi_M(\hat{\rho}))^{i_M} \times \frac{P_1 i_1 + \cdots + P_M i_M}{i_1+\cdots+i_M} \\ &\times (1-\Phi_1(\hat{\rho}))^{N_1-i_1} \times \cdots \times (1-\Phi_M(\hat{\rho}))^{N_M-i_M} \end{aligned} \tag{11}$$

Where the $\delta$ function is used to exclude the state where all zones are empty ($\delta(x) = 0$ for all real $x$ but 0, and $\delta(0) = 1$). The energy-efficiency in this state can be calculated with $\bar{R}_{\bar{\mathbf{s}}}(\hat{\rho}) = \sum_{i=1}^{M} N_i \Phi(\hat{\rho})_i R_{a:i}$ and total power from equation (11).

## IV. Optimal power allocation considering the dynamic behavior of users

In the previous section, we optimized the energy-efficiency for fixed numbers of users. To analyze the impact of power allocation on the network performance and account for the users arrivals and departures, a flow-level capacity analysis is required. The arrival rate can be modeled through a Poisson process (of intensity $\lambda_i$ in zone $i$) and users leave when they

finish streaming a file of average size $F$ (we assume that $F$ is the same for all users). When the total number of users exceed a given threshold $N_{max}$, new user arrivals are blocked.

### A. Processor sharing analysis

When users with a finite workload are considered, the number of users is not constant but varies dynamically during time. The distribution of the number of users is determined by the traffic intensity within the cell. Indeed, if the traffic intensity is large, more users connect to the system per unit time and the average number of active users increases. In this section, we show how to compute the distribution of the number of users knowing the traffic intensity.

The heterogeneity in radio conditions translates into a larger service time for cell edge users. When the system is in state $\bar{\mathbf{s}} = \{N_1, N_2, \cdots, N_M\}$, the total number of users in the cell is $N(\bar{\mathbf{s}}) = N_1 + \cdots + N_M$. Based on [7], we can model the system as a Generalized Processor Sharing queue, whose evolution is just described by the overall number of users in a cell. The solution of the Markov process has the simple form

$$\pi(\bar{\mathbf{s}}) = \frac{1}{\Gamma} \frac{N(\bar{\mathbf{s}})!}{\prod_{i=1}^{M} N_i!} \prod_{c=1}^{M} \frac{\Omega_c^{N_c}}{\prod_{j=1}^{N_c} j \Phi_{c;\bar{\mathbf{s}}(N_c=j)} R_{a:c;\bar{\mathbf{s}}(N_c=j)}} \quad (12)$$

where $\Omega_c = S\lambda_c$ and $\Gamma$ is a normalizing constant. The notation $\bar{\mathbf{s}}(N_c = j)$ is used to take the $\Phi$ and $R_a$ for the state $\bar{\mathbf{s}}$ with $j$ users in zone $c$.

In this model, the user blocking rate can be calculated as $\alpha = \sum_i \lambda_i \sum_x \pi(x)$, $x$ such that the system is full ($N(x) = N_{max}$). Quality of service (QoS) is measured through the user blocking rate. The QoS constraint is thus $\alpha \leq \epsilon$, where $\epsilon$ is the maximum tolerable blocking rate.

### B. Optimal power allocation

The steady-state probabilities defined in the previous section are calculated knowing the throughputs for each state of the network. This throughput will of course depend on the power allocation as explained in Sections II and III. The power allocation has thus to be optimized taking into account the dynamics of users. A power allocation policy $\hat{\mathbf{P}}$ is defined as a set of actions for each of the possible states:

$$\hat{\mathbf{P}} = \bigcup_{\bar{\mathbf{s}}} \mathbf{P}_{\bar{\mathbf{s}}} \quad (13)$$

The global energy efficiency; knowing the policy $\hat{\mathbf{P}}$, is given by:

$$\hat{\eta}(\hat{\mathbf{P}}) = \sum_{\bar{\mathbf{s}}} \frac{\pi(\bar{\mathbf{s}}) \bar{R}_{\bar{\mathbf{s}}}(\hat{\rho})}{\bar{P}_{\bar{\mathbf{s}}}(\mathbf{P}_{\bar{\mathbf{s}}})} \quad (14)$$

The optimization problem can be defined as

$$\hat{\mathbf{P}}^* = \arg\max[\hat{\eta}(\hat{\mathbf{P}})] \quad (15)$$

And the maximum global energy-efficiency possible is $\hat{\eta}(\hat{\mathbf{P}}^*)$.

The idea behind this global optimization is that the power allocation does not depend uniquely on the actual state of the network, but takes also into account the future evolutions of the network. For instance, a power allocation decision that is taken at one moment may have an influence on the evolution of the state of the network by favoring a subset of users by a better throughput. We will study in the next section the difference between this global policy maximization and a local one, as defined in section III.

## V. Numerical results

In this section, we use simulations and numerical calculations to study the properties of the energy-efficiency function and obtain the power allocation that maximizes it. We consider the receiver and the transmitter to have two antennas each forming a $2 \times 2$ MIMO system. The data rates for this configuration which are LTE compliant are taken from [3] and are given as a function of the SINR. For the single zone case we take $\sigma^2 = 1$ mW while for the two zone case we have $\{\sigma_1^2, \sigma_2^2\} = \{1, \frac{1}{8}\}$ mW. We begin by illustrating the results when the network is optimized supposing that the number of users is fixed. The dynamic behavior of users is taken into account afterwards and the performance of the network is compared for both schemes.

### A. Numerical results for the local optimization

We begin by illustrating the power allocation scheme when the dynamic behavior of users is not taken into account, and when all users are subject to the same radio conditions. In figure 1, we show the energy-efficiency as a function of the transmit power. Here, due to symmetry, all the users use the same power. The results show that the energy efficiency begins by increasing with the transmit power increases, as users are able to reach higher throughputs. However, starting from one point, users reach the maximal throughput they are able to reach as, in LTE, modulation schemes are limited; the energy efficiency begins thus decreasing as throughputs remain constant while power consumption increases.

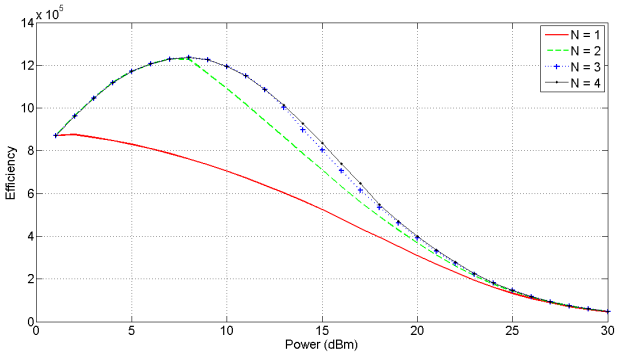


Fig. 1. $\eta$ vs $P$ with $\frac{b}{\sigma^2} = 100$ (20dB). Note that the energy-efficiency is peaked at higher powers with additional users.

In figure 2, we consider the case of two users: one in the "inner" zone (near base station) and the other in the "outer" zone (at cell edge). In this case, the system has a sufficient capacity to support both users and the energy efficiency is optimized when more power is used on the outer zone which

compensates for its lower SINR. Here the total throughput can thus be increased by using more power on the outer zone user. However in figure 3, we have three users in both the inner and outer zones. Here the throughput of the wireless network is not sufficient for all the users and so the energy-efficiency is optimized by simply putting more power in the inner zone with the higher SINR as the total throughput is not improved by putting more power into the outer zone.

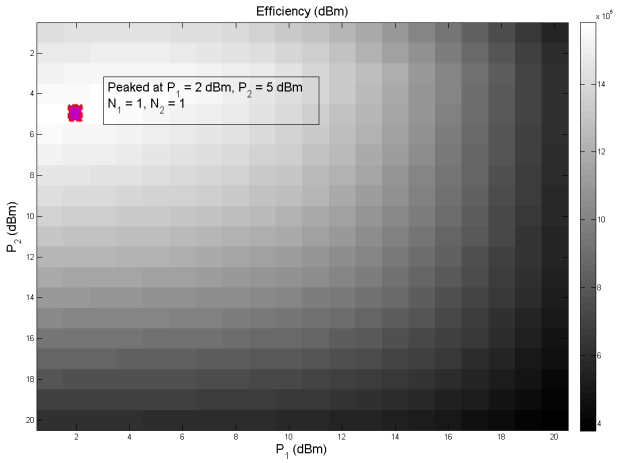


Fig. 2. $\eta$ over combinations of $P_1$ and $P_2$ with $\frac{b}{\sigma_1^2} = 100$ (20dB), $N_1 = N_2 = 1$. Zone 2 corresponds to a lower SINR and in this case the efficiency is optimized by using more power on the zone 2 user.

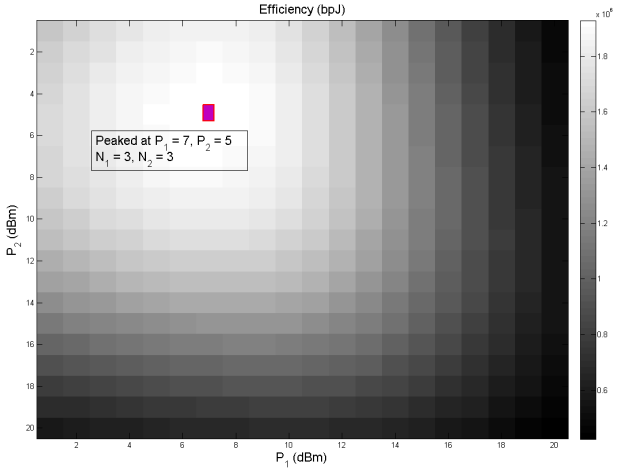


Fig. 3. $\eta$ over combinations of $P_1$ and $P_2$ with $\frac{b}{\sigma_2^2} = 100$ (20dB), $N_1 = N_2 = 3$. As before, zone 2 corresponds to a lower SINR and interestingly, in this case, the efficiency is optimized by using more power along the zone 1 user. This is because with 3 users in each zone, the demanded rate exceeds the maximum available throughput and so, optimization is done by using power on users with a better SINR.

### *B. Numerical results for the global optimization*

We have illustrated, till now, the performance of the system when the number of users is fixed. In this section, we consider the dynamic behavior of users. In this setting, the power allocation is not determined for a fixed number of users, but for a given traffic intensity. the number of users is thus a random variable whose distribution depends on the traffic intensity. The optimal power allocation is the one that maximizes the energy efficiency while maintaining a constraint on the QoS. Note that this optimal power allocation is a matrix that gives, for each state of the network composed of the number of users in the cell, the power allocation for each of the users.

Initially we consider the cell with homogeneous radio conditions, i.e. we suppose that all the users experience the same SINR on average. In this setting, if $N_{max}$ is the maximum number of users allowed, optimization is performed over $N_{max}$ variables, i.e. the power used in each state. For the single zone case we take $\sigma^2 = 1$ mW. The optimal power allocation is shown in figure 4. Note that, in this case, the power allocation is a vector and not a matrix, as all users experience the same radio conditions and have, by symmetry, the same allocated power.

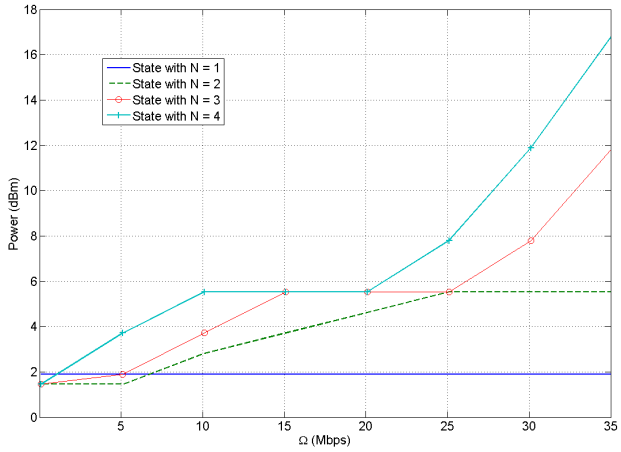


Fig. 4. The power allocation scheme $(P_1, \cdots, P_4)$ plotted against the traffic $\Omega$ when $\hat{\eta}$ is optimized. Also note that the QoS constraint of maintaining the blocking rate below $0.01$ is satisfied.

Figure 5 compares the energy-efficiency obtained for the local and the global optimizations. Recall that, by local, we mean that the optimization is done for each state independently from the others, taking into account only the observed number of users and not the future evolutions of the system. As seen from the simulations (Figure 5), using a global optimization does not seem to yield much gains in the energy-efficiency for the single zone case. This is because the throughput, and thus service times, are the same for all users. We next move on to the two-zone case (cell center and cell edge). Here we consider a cell divided into two concentric rings, and define the outer zone as the region when the SINR is $4.8$ dB (3 times)

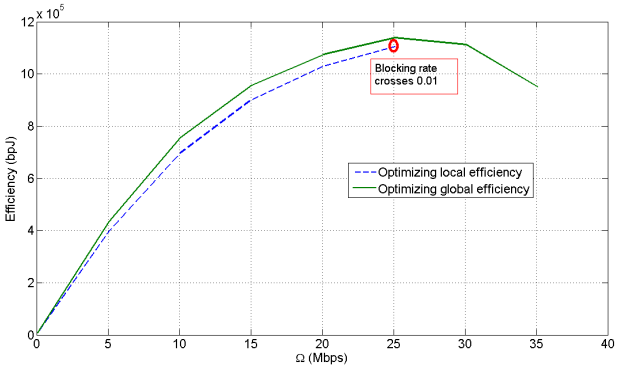


Fig. 5. $\hat{\eta}$ plotted against the traffic $\Omega$ when $\hat{\eta}$ is optimized and when $\eta$ is optimized for each state separately.

lower than the SINR for the inner zone, when the transmit power is unchanged. The outer zone also has 3 times the area of the inner zone causing $\lambda_2 = 3\lambda_1$. With these parameters we attempt to calculate the optimal global energy-efficiency and corresponding power allocation for given values of $\lambda_1$. We have $\{\sigma_1^2, \sigma_2^2\} = \{1, \frac{1}{3}\}$ mW. Figure 6 shows the energy efficiencies corresponding to local and global optimizations. It is obvious that global optimization yields much higher efficiency when users have heterogeneous radio conditions. This is because, in the local optimization setting, the notion of call duration cannot be taken into account as users are considered as always active. The optimal power allocation will then tend to favor cell center users in order to maximize throughput. However, when the dynamic behavior of users is taken into consideration, it is sometimes better to use more power on cell edge users in order to let them finish their service quickly and quit the system. Applying the policy obtained from the local optimization will lead to users accumulating at the cell edge as they are not able to finish their transfers.

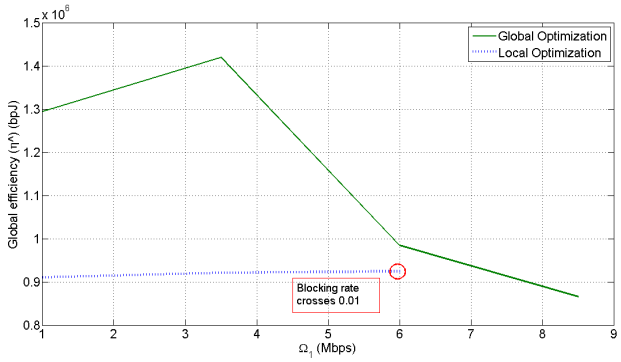


Fig. 6. $\hat{\eta}$ plotted against the traffic $\Omega_1 = \lambda_1 S$ when $\hat{\eta}$ is optimized and when $\eta$ is optimized for each state separately. Also note that the QoS constraint of maintaining the blocking rate below 0.01 is satisfied at all points shown.

## VI. Conclusion

In this work we study and optimize the flow level energy efficiency of base stations in LTE. We introduce the notion of a "global" energy-efficiency which is defined as the average of the energy-efficiencies of each state the cell can be in. These states represent the traffic configurations, i.e. the numbers and positions of users in the cell. Through extensive simulations we see that optimizing the global efficiency yields a different power allocation from optimizing the efficiency of each individual state. Although this difference can be neglected when considering a cell in which all users experience the same average SINR, when considering a more realistic setting where users are subject to heterogeneous radio conditions, the global optimization yields a considerable gain. This is because, when users are considered as static, it may be optimal to give more power to cell center in order to increase throughputs. However, when the dynamic behavior of users is taken into account, giving more power to users with bad radio conditions will allow them leaving the system faster and thus alleviating load in the future. When compared to the local optimization, it is observed that the global optimization improves the energy-efficiency up to a factor of 50%.

## Acknowledgments

This work is a joint collaboration between Orange Labs, Laboratoire des signaux et systémes (L2S) of Supélec and the Alcatel Lucent Chair of Supélec. This work is part of the European Celtic project "Operanet2".

## References

[1] L. Saker and S-E. Elayoubi, "Sleep mode implementation issues in green base stations", IEEE PIMRC 2010, Istanbul, September 2010
[2] GreenTouch at "http://www.greentouch.org/"
[3] J.B Landre, Z.E Rawas, R.Visoz, and S Bouguermouh, "Realistic Performance of LTE in a macro-cell environment", IEEE VTC 2012
[4] F. Meshkati, H. Poor, S. Schwartz, "Energy Efficiency-Delay Tradeoffs in CDMA Networks: A Game-Theoretic Approach", in Transactions on Information Theory, vol. 55, no. 7, 2009.
[5] D. J. Goodman, and N. Mandayam, "Power Control for Wireless Data", IEEE Personal Communications, vol. 7, pp. 48-54, Apr. 2000.
[6] E. V. Belmega, S. Lasaulce, and M. Debbah, "A survey on energy-efficient communications", IEEE Intl. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC 2010).
[7] R. Combes, S.E. Elayoubi, Z. Altman, "Cross-layer analysis of scheduling gains: Application to LMMSE receivers in frequency-selective Rayleigh-fading channels". WiOpt - 2011, 133-139
[8] F. Meshkati, M. Chiang, H. Poor and S. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems ", in JSAC, vol. 24, no. 6, 2006.
[9] E. Altman , K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, D. Menasche, "Constrained Stochastic Games in Wireless Networks", in proceedings of GlobeCom, 2007.

## C.2 BLACKSEACOM-2013

• M. Mhiri, V.S Varma, M.L. Truest, S. Lasaulce and A. Samet, "On the benefits of repeated game models for green cross-layer power control in small cell", BlackSeaComm 2013

# On the benefits of repeated game models for green cross-layer power control in small cells

*(Invited Paper)*

Mariem Mhiri*, Vineeth S. Varma†, Maël Le Treust‡, Samson Lasaulce† and Abdelaziz Samet*

*Tunisia Polytechnic School P.B. 743-2078, University of Carthage, La Marsa, Tunisia
mariem.mhiri@gmail.com, abdelaziz.samet@ept.rnu.tn

†L2S - CNRS - SUPELEC, F-91192 Gif-sur-Yvette, University of Paris-Sud, France
{vineeth.varma, samson.lasaulce}@lss.supelec.fr

‡EMT Centre - INRS University, Ouest Montréal (Québec) H5A 1K6, Canada
mael.le.treust@emt.inrs.ca

***Abstract*—In this paper, we consider the problem of distributed power control for multiple access channels when energy-efficiency has to be optimized. In contrast with related works, the presence of a queue at each transmitter is accounted for and globally efficient solutions are sought. To this end, a repeated game model is exploited and shown to lead to solutions which are distributed in the sense of the decision, perform well globally, and may rely on limited channel state information at the transmitter.**

***Index Terms*—distributed power control, energy efficiency, repeated game, channel state information.**

## I. Introduction

Designing energy-efficient communication systems has become a critical issue in modern day wireless networks. The problem treated in this work deals with power control when energy efficiency (EE) has to be optimized. This metric (EE) has been defined in [1] as a ratio of the net data rate (goodput) to the transmit power level. The problem was formulated as a non-cooperative game where each transmitter aims at selfishly maximizing its individual energy-efficiency. The considered solution is the Nash equilibrium (NE) which is shown to be unique but generally Pareto inefficient. To deal with this inefficiency, an operating point (OP) was proposed in [2] where repeated game was exploited. Authors in [2] showed that when playing with the developed OP according to a cooperation plan, only channel state information (CSI) is needed and transmitters can improve the social welfare (sum of utilities). Recently, a generalized EE metric has been proposed in [3] for two important transport layer protocols (Transmission Control Protocol (TCP) and User Datagram Protocol (UDP)). The new EE metric is based on a cross-layer approach and takes into account the effects of the presence of a queue with a finite size at the transmitter. An interference channel system was studied and it was shown that a unique NE exists for a non-cooperative game. In this paper, we consider the problem of distributed power control with the new EE metric according to UDP protocol developed in [3] and for multiple access channels (MAC) system. Our goal is to find another unique solution concept which is efficient and may rely on limited CSI at the transmitter. We refer to a repeated game model (RG) developed in [2] and try to apply the results on the cross-layer power control game. One of the major mathematical distinction between the two metrics used is the presence of a constant power term in the denominator of the EE metric. Although it appears to be a small change, the structure of the equilibrium solution is quite different. The optimal SINR when using the [1] metric is independent of the channel state. This property is lost when accounting for the constant power consumption, and motivates us to propose a new OP for the cross-layer metric. The main contributions of this work are:

1) Study the RG when using the cross layer EE as the utility of the game;
2) Establish the threshold on the game length beyond which the equilibrium policy can be pareto-optimal;
3) Propose a new OP that is efficient and can be reached in a distributed manner.

This paper is structured as follows. In section II-A, we introduce the system model under study. Then, we define (in section II-B) the static power control game. This is followed (section II-C) by a review of the non-cooperative one-shot game. In section III, we give the formulation of the RG model. In section IV, we introduce the new OP and an equilibrium for the finite RG is proposed. Numerical results are presented and discussed in section V. Finally, concluding remarks are proposed in section VI.

## II. Problem statement

### A. System model

The communication network under study is that of a MAC system, where $N$ small transmitters are communicating with a receiver and are operating in the same frequency band. Transmitter $i \in \{1, \ldots, N\}$ sends a signal $\sqrt{p_i}x_i$ with power $p_i \in [0, P_i^{\max}]$ where $P_i^{\max} > 0$ is the maximum transmit power. The channel gain of the link between transmitter $i$ and the destination is denoted as $g_i$. Thus, the baseband signal

received is written:

$$y_i = g_i\sqrt{p_i}x_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} g_j\sqrt{p_j}x_j + n_i, \quad (1)$$

with $n_i$ is additive white Gaussian noise (AWGN) with mean 0 and variance $\sigma_i^2$. We assume that $\sigma_i^2$ is identical for all the transmitters such that: $\sigma_i^2 = \sigma^2$. Therefore, the resulting SINR $\gamma_i$ at the receiver is given by:

$$\gamma_i(\mathbf{p}) = \frac{p_i|g_i|^2}{\sigma^2 + \frac{1}{L}\sum_{\substack{j=1 \\ j\neq i}}^{N} p_j|g_j|^2}, \quad (2)$$

where $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ is the power vector which will describe later the power actions of the $N$ transmitters and $L$ refers to the spreading factor [3].

We assume that the described system is based on the IP (Internet Protocol) stack where packets arrive from an upper layer into a finite memory buffer of size $K$ (in packets). Here, the considered protocol is UDP for which the packet arrival process follows a Bernoulli process with a constant probability $q$, independent from the SINR. This results in an effective packet loss denoted by $\Phi(\gamma_i)$ and an energy efficiency $\eta_i$ given by:

$$\eta_i(p_i, \mathbf{p}_{-i}) = \frac{Rq(1-\Phi(\gamma_i(\mathbf{p})))}{b + \frac{qp_i(1-\Phi(\gamma_i(\mathbf{p})))}{f(\gamma_i(\mathbf{p}))}}, \quad (3)$$

where $\mathbf{p}_{-i} = (p_1, .., p_{i-1}, p_{i+1}, .., p_N)$, $R$ is the used throughput (in bit/s) and $b$ represents the fixed consumed power when the radiated power is zero [3].

### B. Static power control game

The major motivation behind this work is in order to establish an efficient equilibrium point to which a completely distributed system can converge to. A non-cooperative game has been introduced in [3] where the existence of a unique Nash equilibrium was proved. Here, we are looking for more efficient solutions which are distributed in the sense of the decision making, but may rely on limited channel state information at the transmitter. As motivated in [3], the power control can be modeled by a strategic form game (see e.g., [4]).

*Definition 2.1:* The game is defined by the ordered triplet $\mathcal{G} = \left(\mathcal{N}, (\mathcal{A}_i)_{i\in\mathcal{N}}, (u_i)_{i\in\mathcal{N}}\right)$ where

- $\mathcal{N}$ *is the set of players. Here, the players of the game are the sources/transmitters,* $\mathcal{N} = \{1, \ldots, N\}$;
- $\mathcal{A}_i$ *is the set of actions. Here, the action of source/transmitter* $i$ *consists in choosing* $p_i$ *in its action set* $\mathcal{A}_i = [0, P_i^{\max}]$;
- $u_i$ *is the utility function of each user according to UDP given by:*

$$u_i(p_i, \mathbf{p}_{-i}) = \eta_i(p_i, \mathbf{p}_{-i}) \quad (4)$$

The function $f : [0, +\infty) \to [0, 1]$ is a sigmoidal efficiency function which corresponds to the packet success rate verifying $f(0) = 0$ and $\lim_{x\to+\infty} f(x) = 1$. The function $\Phi$ identifies the packet loss due to both bad channel conditions and the finiteness of the packet buffer. This can be calculated as:

$$\Phi(\gamma_i) = (1 - f(\gamma_i))\Pi_K(\gamma_i) \quad (5)$$

where $\Pi_K(\gamma_i)$ is the stationary probability that the buffer is full and is given by:

$$\Pi_K(\gamma_i) = \frac{\omega^K(\gamma_i)}{1 + \omega(\gamma_i) + \ldots + \omega^K(\gamma_i)} \quad (6)$$

with:

$$\omega(\gamma_i) = \frac{q(1 - f(\gamma_i))}{(1-q)f(\gamma_i)} \quad (7)$$

In [3], the authors prove that the non-cooperative game with rational players, $\mathcal{G}$, allows for a unique pure Nash equilibrium (NE). This NE is the set of powers from which no player has anything to gain by changing only his own strategy unilaterally. This is explained in the following section.

### C. Review of the non-cooperative game

The non-cooperative power control game has been investigated in [3] where the quasi-concavity of the utility function given in (4) was proved. Accordingly, as the NE represents the fundamental solution for a non-cooperative game, existence and uniqueness of such a solution have been studied and demonstrated as well. Thus, the optimal power denoted as $p_i^*$ is obtained by setting $\partial u_i/\partial p_i$ to zero, which leads to solve the following equation:

$$b\gamma_i'\Phi'(\gamma_i) + q\left(\frac{1-\Phi(\gamma_i)}{f(\gamma_i)}\right)^2 [f(\gamma_i) - p_i\gamma_i' f'(\gamma_i)] = 0, \quad (8)$$

where $\gamma_i' = \dfrac{\mathrm{d}\gamma_i}{\mathrm{d}p_i} = \dfrac{\gamma_i}{p_i}$, $f' = \dfrac{\mathrm{d}f}{\mathrm{d}\gamma_i}$ and $\Phi' = \dfrac{\mathrm{d}\Phi}{\mathrm{d}\gamma_i}$.

However, the NE solution is not always Pareto efficient for many scenarios. An example is presented in Fig. 1 where we stress that the NE is far from the Pareto frontier. Motivated by the need to design an efficient solution relying on limited CSI at the transmitter, we move to the repeated game framework.

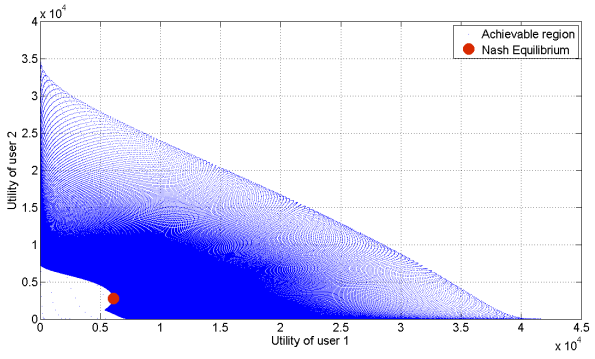


Fig. 1. Pareto inefficiency of the NE.

## III. Repeated power control game

In repeated games (RG), as the name suggests, the same game is played several times. The long-term interactions between the players in such a situation is studied under the RG framework. The players react to past experience by taking into account what happened in all previous stages and make decisions about their future choices [5], [6]. The resulting utility of each player is an average of the utility of each stage. A game stage $t$ corresponds to the instant in which all players choose their actions simultaneously and independently and thus a profile of actions can be defined by $\mathbf{p}(t) = (p_1(t), p_2(t), \ldots, p_N(t))$. When assuming full monitoring, this profile choice is observed by all the players and the game proceeds to the next stage [6]. The sequence of actions $\mathbf{p}_i(t)$ of a transmitter $i$ at time $t$ defines his history denoted as $\mathbf{h}(t) = \mathbf{p}_i(t) = (p_i(1), p_i(2), \ldots, p_i(t-1))$ and which lies in the set $\mathcal{H}_t = \mathcal{P}_i^{t-1}$. Before playing stage $t$, all histories are known by all the players [2]. According to the above descriptions, a pure strategy $\delta_{i,t}$ of player $i \in \mathcal{N}$ is a mapping from $\mathcal{H}_t$ to the action set $\mathcal{A}_i = [0, P_i^{\max}]$ specifying the action to choose after each history [2], [6]:

$$\delta_{i,t} : \left| \begin{array}{ccc} \mathcal{H}_t & \rightarrow & [0, P_i^{\max}] \\ \mathbf{h}(t) & \mapsto & p_i(t) \end{array} \right. \tag{9}$$

We define the joint strategy $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_N)$ as the vector of all the players strategies.

In this paper, we are interested in the finite repeated game, i.e the game is played for a finite number of steps ($T$ steps). The utility function of each player results from averaging over the instantaneous utilities over all the game stages. At each stage $t$, the instantaneous utility of player $i$ is a function of the profile of actions of all the players $\mathbf{p}(t)$.

*Definition 3.1:* The utility function of the $i^{th}$ player for the finite RG is the arithmetic average of the sum of the utilities for the initial $T$ first stages [6], [7]. We have [2]:

$$v_i^T(\boldsymbol{\delta}) \quad = \frac{1}{T} \sum_{t=1}^{T} u_i(\mathbf{p}(t)) \qquad \text{for the finite RG} \tag{10}$$

where $T \geq 1$ defines the number of game stages in the finite RG.

An equilibrium solution of the RG is defined in the following manner:

*Definition 3.2:* A joint strategy $\delta$ satisfies the equilibrium condition for the finite repeated game if for all players $i \in N$, for all other strategies $\delta_i'$, we have $v_i^T(\delta) \geq v_i^T(\delta_i', \delta_{-i})$. It means that no deviating strategy $\delta_i'$ can increase the utility $v_i^t(\delta_i', \delta_{-i})$ of any one player.

This equilibrium solution is exactly what we are interested in, as a strategy $\delta$ satisfying the above condition would be precisely what rational players in a RG would play. In an RG with complete information and full monitoring, the Folk theorem characterizes the set of possible equilibrium utilities [2], [6]. It states that the set of Nash equilibrium in a RG is precisely the set of feasible and individually rational outcomes of the one-shot game (non-cooperative game) [5], [6]. In an RG, interesting patterns of behavior between players can be seen and studied. This includes: rewarding and punishing, co-operation and threats, transmitting information and concealing [5].

## IV. An operating point and repeated game characterization

### A. New OP for the game $\mathcal{G}$

Consider the operating point (OP) described in [2]. It is identified by a subset of points of the achievable utility region such that $p_i|g_i|^2 = p_j|g_j|^2$ for all $(i,j) \in \mathcal{N}$. The optimal subset of powers consists of the solutions of the following system of equations:

$$\forall (i,j) \in \mathcal{N} \ , \ \frac{\partial u_i}{\partial p_i}(\mathbf{p}) = 0 \text{ with } p_i|g_i|^2 = p_j|g_j|^2 \tag{11}$$

with $u_i$ is the utility function defined in (4).

Due to the presence of the parameter $b$ which we consider different from 0, it can be observed that there will be $N$ different solutions corresponding to equation (11) in terms of $p_i$ and thus the operating point from [2] is not well defined when using the utility defined in [3]. To deal with this problem, a new OP is proposed. The new OP consists in setting $p_i|g_i|^2$ to a constant denoted as $\alpha$ that can be optimized. We propose to determine a unique optimal $\alpha$ by maximizing the expected sum utility over all the channel states as follows:

$$\tilde{\alpha} = \arg\max E_g \left[ \sum_{i=1}^{N} u_i(\alpha, g) \right] \tag{12}$$

When playing at the OP, the power played by the $i^{th}$ player, denoted as $\tilde{p}_i$, is given by:

$$\tilde{p}_i = \frac{\tilde{\alpha}}{|g_i|^2} \tag{13}$$

In the following section, we focus on the characterization of the finite RG.

### B. Repeated power control game characterization

As a first step, we determine the minimum number of stages ($T_{\min}$) corresponding to the finite RG. When the number of stages in the game is less than $T_{\min}$, the equilibrium of the RG is to simply play at the NE. However, if $T > T_{\min}$, a more efficient equilibrium point can be reached. Assuming that channel gains $|g_i|^2$ lie in a compact set $[\eta_i^{\min}, \eta_i^{\max}]$ [2], we have the following proposition:

*Proposition 4.1 (Equilibrium solution for the finite RG.):* For a finite RG, if $T > T_{\min}$, then the corresponding equilibrium solution is given by [2]:

$$\delta_{i,t} : \left| \begin{array}{ll} \tilde{p}_i & \text{for } t \in \{1, 2, \ldots, T - T_{\min}\} \\ p_i^* & \text{for } t \in \{T - T_{\min} + 1, \ldots, T\} \\ P_i^{\max} & \text{for any deviation detection} \end{array} \right. \tag{14}$$

where $T_{\min}$ is:

$$T_{\min} = \left\lceil \frac{\frac{A\eta_i^{\max}}{b\eta_i^{\min} + \tilde{\gamma}_i \sigma^2 B} - \frac{G\eta_i^{\max}}{b\eta_i^{\min} + \tilde{\alpha} H}}{\frac{E\eta_i^{\min}}{b\eta_i^{\max} + \gamma_i^* \left(\sigma^2 + \frac{1}{L}\sum_{j \neq i} p_j^* \eta_i^{\max}\right) F} - \frac{C\eta_i^{\min}}{b\eta_i^{\max} + \tilde{\gamma}_i \left(\sigma^2 + \frac{1}{L}\sum_{j \neq i} p_j^{\max} \eta_i^{\max}\right) D}} \right\rceil \tag{15}$$

The proof for this proposition is given in Appendix A, as well as the quantities $A$, $B$, $C$, $D$, $E$, $F$, $G$ and $H$. $\gamma_i^*$ is the SINR at the NE while $\bar{\gamma}_i$ and $\hat{\gamma}_i$ are the SINRs related to the utility max and the utility minmax respectively (see Appendix A).

## V. Numerical results

We consider a scenario with a MAC where $N$ transmitters are communicating with their corresponding receiver (e.g. base station). The efficiency function is assumed to be $f(x) = e^{-c/x}$ where $c = 2^{\frac{R}{R_0}} - 1$ with $R$ and $R_0$ are the throughput and the used bandwidth and supposed to be 1 Mbps and 1 MHz respectively. The other parameters are set as follows:

- $\sigma^2 = 10^{-3}$ W
- $b = 10^{-2}$ W
- $K = 10$
- $q = 0.5$
- $P_i^{\max} = P^{\max} = 10^{-1}$ W

The channel gains are assumed to be $|g_i|^2 = 1$ and $|g_j|^2 = 0.5$. Our simulations consist in showing firstly the advantage of the OP regarding the NE of the one shot game. Thus, we plot the achievable utility region, the NE and the proposed OP when considering a system of two transmitters and a spreading factor $L = 2$. In Fig. 2, the region delimited by the Pareto frontier and the minmax level defines, according to the Folk theorem, the possible set of equilibrium utilities of the RG. In addition, we highlight that the new OP dominates in terms of Pareto the NE and it is Pareto efficient.

Fig. 3 represents the ratio $\frac{w_{FRG}}{w_{NE}}$ for the finite RG as a function of the number of stages. We have:

$$\frac{w_{FRG}}{w_{NE}} = \frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T-T_{\min}} \tilde{u}_i(\mathbf{p}(t)) + \sum_{t=T-T_{\min}+1}^{T} u_i^*(\mathbf{p}(t))\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T} u_i^*(\mathbf{p}(t))} \tag{16}$$

We consider a system with 25 transmitters and a spreading factor $L = 100$. We proceed to an averaging over channel gains lying in a compact set such that $10\log_{10}\frac{\eta_{\max}}{\eta_{\min}} = 20$. According to equation (15), the minimum number of stages $T_{\min}$ is equal to 1200. According to this figure, we deduce that the social welfare can be improved when playing an RG.

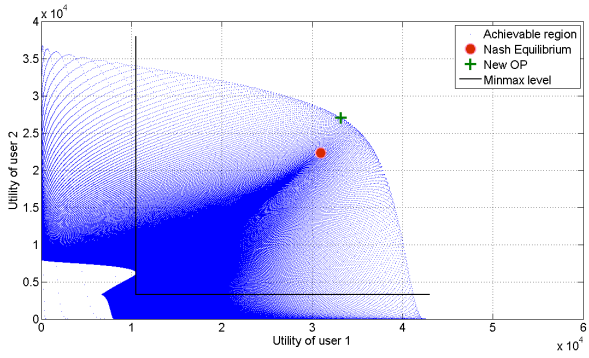


Fig. 2. Pareto dominance and Pareto efficiency of the proposed OP regarding the NE.

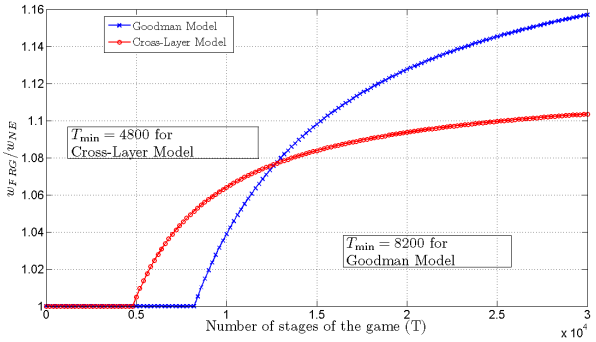


Fig. 3. Improvement of the social welfare in finite repeated game vs the Nash equilibrium. While the efficiency of the RG while using the traditional metric defined in [1] seems to be higher, it requires a longer game than in the cross layer model.

Figure 3 plots the improvement of the social welfare as defined in 16. This improvement obtained is compared for case when using the metric defined in [1] to the cross-layer metric used. The required time for profiting from the RG scenario is much lower in the cross-layer case, but the improvement seems to be relatively smaller. However, note that the NE in the cross-layer game itself is more efficient than the NE in [1] and so in absolute terms, the proposed OP is still quite efficient and can be utilized for shorter games. This validates our approach and shows that the RG formulation is a useful technique for efficient distributed power control.

## VI. Conclusion

In this paper, we study an efficient solution for a relevant game with a new EE metric considering a cross-layer approach and taking into account the effects of the presence of a queue with a finite size at the transmitter. As the NE is generally inefficient in terms of Pareto, we design a new OP and exploit a repeated game model to improve the performance of a MAC system. We contribute to express the analytic form for the minimum number of stages in a finite RG. Moreover, our approach provides an efficient solution relying on limited CSI at the transmitter when comparing to the NE and contributes to considerable gains in terms of social welfare for the finite RG.

## Appendix A
## Proposition 4.1

The utilities max and minmax are expressed respectively as follows:

$$\begin{aligned} \bar{u}_i &= \max_{\mathbf{p}_{-i}} \max_{p_i} u_i(p_i, \mathbf{p}_{-i}) \\ \hat{u}_i &= \min_{\mathbf{p}_{-i}} \max_{p_i} u_i(p_i, \mathbf{p}_{-i}) \end{aligned}$$

As a first step, we determine the power $p_i$ maximising $u_i$ and which we denote as $\dot{p}_i$. This amounts to reduce $\partial u_i/\partial p_i$ to 0. We recall that we consider the following notations: $\gamma_i' = \frac{\mathrm{d}\gamma_i}{\mathrm{d}p_i} = \frac{\gamma_i}{p_i}$, $f' = \frac{\mathrm{d}f}{\mathrm{d}\gamma_i}$ and $\Phi' = \frac{\mathrm{d}\Phi}{\mathrm{d}\gamma_i}$.

The power $\dot{p}_i$ maximising $u_i$ is then the solution of the following equation:

$$b\frac{\gamma_i}{p_i}\Phi'(\gamma_i) + q\left(\frac{1-\Phi(\gamma_i)}{f(\gamma_i)}\right)^2\left[f(\gamma_i) - \gamma_i f'(\gamma_i)\right] = 0 \tag{17}$$

Therefore, the expression of the maximum utility function writes as:

$$\dot{u}_i(\dot{p}_i,\mathbf{p}_{-i})=\frac{Rq(1-\phi(\dot{\gamma}_i))}{b+\frac{\dot{p}_i q(1-\phi(\dot{\gamma}_i))}{f(\dot{\gamma}_i)}},$$

with:

$$\dot{\gamma}_i=\frac{\dot{p}_i|g_i|^2}{\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j|g_j|^2}$$

In a second step, we are interested in studying the behavior of $\dot{u}_i(\dot{p}_i,\mathbf{p}_{-i})$ as a function of $p_j$ for $j\neq i$; which amounts to calculating the sign of $\frac{\partial \dot{u}_i(\dot{p}_i,\mathbf{p}_{-i})}{\partial p_j}$,which is shown to be negative in [3]. Therefore $\frac{\partial \dot{u}_i(\dot{p}_i,\mathbf{p}_{-i})}{\partial p_j}<0$. As $\dot{u}_i$ is a decreasing function of $p_j$, it reaches its maximum when $p_j=0$ and it is minimum when $p_j=p_j^{\max}$ (for all $j\neq i$).

### A. Expression of $\bar{u}_i$

The utility $\dot{u}_i$ reaches its maximum when $p_j=0$. When substituting $p_j=0$ in the SINR expression $\dot{\gamma}_i$, this allows the determination of the optimal power $\dot{p}_i$:

$$b\frac{|g_i|^2}{\sigma^2}\Phi'(\gamma_i)+q\Big(\frac{1-\Phi(\gamma_i(p_i))}{f(\gamma_i(p_i))}\Big)^2\big[f(\gamma_i(p_i))-\gamma_i f'(\gamma_i(p_i))\big]=0 \qquad (18)$$

As the latter equation is a function of the SINR, the solution will be in terms of SINR and will be denoted as $\bar{\gamma}_i$. The corresponding optimal power is $\bar{p}_i=\frac{\bar{\gamma}_i\sigma^2}{|g_i|^2}$. Then, we have:

$$\bar{u}_i=\frac{Rq(1-\phi(\bar{\gamma}_i))}{b+\frac{\bar{\gamma}_i\sigma^2}{|g_i|^2}\frac{q(1-\phi(\bar{\gamma}_i))}{f(\bar{\gamma}_i)}}$$

### B. Expression of $\widehat{u}_i$

We proceed as described previously and determine the optimal SINR denoted as $\widehat{\gamma}_i$ which is the solution of the following equation:

$$\frac{b|g_i|^2}{\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^{\max}|g_j|^2}\Phi'(\gamma_i)+q\Big(\frac{1-\Phi(\gamma_i)}{f(\gamma_i)}\Big)^2\big[f(\gamma_i)-\gamma_i f'(\gamma_i)\big]=0 \qquad (19)$$

Then, we have:

$$\widehat{u}_i=\frac{Rq(1-\phi(\widehat{\gamma}_i))}{b+\frac{\widehat{\gamma}_i}{|g_i|^2}\left(\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^{\max}|g_j|^2\right)\frac{q(1-\phi(\widehat{\gamma}_i))}{f(\widehat{\gamma}_i)}}$$

### C. Existence proof of $\widehat{\gamma}_i$ and $\bar{\gamma}_i$

Both equations (18) and (19) are resulting from the same equation (17) for two different forms of the SINR ($\bar{\gamma}_i$ for $p_j=0$ and $\widehat{\gamma}_i$ for $p_j=p_j^{\max}$). Showing the existence of these two solutions amounts to prove the existence of the solution of equation (17). However, according to the study established in [3], it was proved that $u_i$ is quasi-concave in $(p_i,\mathbf{p}_{-i})$ and then it exists $\gamma^+$ such that the first derivative of $u_i$ regarding to $p_i$ is strictly positive on $[0,\gamma^+]$ and strictly negative on $[\gamma^+,+\infty]$ for all $p_j\in[0,p_j^{\max}]$ : the first derivative is continous and is equal to zero in $\gamma^+$. According to the utility which we are studying (max or minmax), $\gamma^+$ is either $\bar{\gamma}_i$ (eq. (18)) or $\widehat{\gamma}_i$ (eq. (19)).

### D. Proof

From [2], we have:

$$\begin{aligned}\bar{u}_i(\mathbf{p}(t))+\sum_{s=T-T_{\min}+1}^{T}\mathbb{E}_g\{\widehat{u}_i(\mathbf{p}(s))\} \quad &\leq \\ \tilde{u}_i(\mathbf{p}(t))+\sum_{s=T-T_{\min}+1}^{T}\mathbb{E}_g\{u_i^*(\mathbf{p}(s))\} \quad &(20)\end{aligned}$$

The SINR $\tilde{\gamma}_i$ refers to the SINR when playing the new OP. In order to simplify expressions, we use the following notations:

$$\begin{aligned}A &= Rq(1-\phi(\bar{\gamma}_i))\\ B &= \tfrac{q(1-\phi(\bar{\gamma}_i))}{f(\bar{\gamma}_i)}\\ C &= Rq(1-\phi(\widehat{\gamma}_i))\\ D &= \tfrac{q(1-\phi(\widehat{\gamma}_i))}{f(\widehat{\gamma}_i)}\\ E &= Rq(1-\phi(\gamma_i^*))\\ F &= \tfrac{q(1-\phi(\gamma_i^*))}{f(\gamma_i^*)}\\ G &= Rq(1-\phi(\tilde{\gamma}_i))\\ H &= \tfrac{q(1-\phi(\tilde{\gamma}_i))}{f(\tilde{\gamma}_i)}\end{aligned}$$

The inequality (20) becomes:

$$\begin{aligned}&\tfrac{A|g_i|^2}{b|g_i|^2+\tilde{\gamma}_i\sigma^2 B}+\textstyle\sum_{s=T-T_{\min}+1}^{T}\mathbb{E}_g\left[\frac{C|g_i|^2}{b|g_i|^2+\tilde{\gamma}_i\left(\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^{\max}|g_j|^2\right)D}\right]\\ &\leq\tfrac{G|g_i|^2}{b|g_i|^2+\tilde{\alpha}H}+\textstyle\sum_{s=T-T_{\min}+1}^{T}\mathbb{E}_g\left[\frac{E|g_i|^2}{b|g_i|^2+\gamma_i^*\left(\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^*|g_j|^2\right)F}\right]\end{aligned}$$

and simplifying:

$$T_{\min}=\left\lceil\frac{\frac{A\eta_i^{\max}}{b\eta_i^{\min}+\tilde{\gamma}_i\sigma^2 B}-\frac{G\eta_i^{\max}}{b\eta_i^{\min}+\tilde{\alpha}H}}{\frac{E\eta_i^{\min}}{b\eta_i^{\max}+\gamma_i^*\left(\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^*\eta_i^{\max}\right)F}-\frac{C\eta_i^{\min}}{b\eta_i^{\max}+\tilde{\gamma}_i\left(\sigma^2+\frac{1}{L}\sum_{j\neq i}p_j^{\max}\eta_i^{\max}\right)D}}\right\rceil$$

## REFERENCES


[1] D. J. Goodman and N. B. Mandayam, "Power control for wireless data," *IEEE Personal Communications*, vol. 7, pp. 48–54, Apr. 2000.

[2] M. L. Treust and S. Lasaulce, "A repeated game formulation of energy-efficient decentralized power control," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2860–2869, Sep. 2010.

[3] V. S. Varma, S. Lasaulce, Y. Hayel, and S. E. Elayoubi, "A cross-layer approach for energy-efficient distributed power control," *submitted to IEEE Transactions on Wireless Communications*, Dec. 2012.

[4] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*. Academic Press, Elsevier, Jul. 2011.

[5] S. Hart, "Robert aumann's game and economic theory," *Scandinavian Journal of Economics*, vol. 108, no. 2, pp. 185–211, 2006.

[6] S. Sorin, "Repeated games with complete information," in *Handbook of Game Theory*, R. J. Aumann and S. Hart, Eds. Elsevier Science Publishers, 1992, pp. 72–107.

[7] R. J. Aumann and L. S. Shapley, "Long-term competition-a game-theoretic," pp. 1–26, 1976, preprint.

# Bibliography


[1] D. Lister, "An Operator's View on Green Radio", Proc. IEEE Internat. Conf. on Commun, Workshops (ICC Workshops '09), 1st Internat. Workshop on Green Commun. (GreenComm '09) (Dresden, Ger, 2009).

[2] J. Palicot, and C. Roland, "On The Use Of Cognitive Radio For Decreasing The Electromagnetic Radiations", URSI 05, XXVIII General Assembly, New Delhi, India, October 23-29, 2005.

[3] GreenTouch, "Communications Turns Totally Green", Press Release, Jan. 11, 2010.

[4] Mobile VCE, Virtual Centre of Excellence in Mobile and Personal Communications, Core 5 Research Programme, "Green Radio - Sustainable Wireless Networks", Feb. 2009.

[5] D. Lister, "An operators view on green radio", Proc. IEEE Internet. Conf. on Comm. Workshops (ICC Workshops 2009), 1st Int. Workshop on Green Comm. (GreenComm 09) (Dresden, Ger., 2009).

[6] GreenTouch, "Communications Turns Totally Green", Press Release, Jan. 11, 2010.

[7] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks", IEEE Vehicular Technology Magazine, vol. 6, no. 1, Mar. 2011.

[8] D. J. Goodman, and N. Mandayam, "Power Control for Wireless Data", IEEE Personal Communications, vol. 7, pp. 48-54, Apr. 2000.

[9] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost.", Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on. IEEE, 2010.

[10] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *European Trans. Telecommunications*, Vol. 10 (1999), pp. 585-595.

[11] E.V. Belmega, and S. Lasaulce, "Energy-efficient pre-coding for multiple-antenna terminals", IEEE. Trans. on Signal Processing, 59, 1, Jan. 2011.

[12] S. Verdú, "Spectral efficiency in the wideband regime", IEEE Trans. on Inf. Theory, vol. 48, no. 6, pp. 1319-1343, Jun. 2002.

[13] S. Cui, A. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks" , IEEE Journal on Selected Areas in Communications, Vol. 22, No. 6, pp. 1089-1098, Aug. 2004.

[14] V. Chandar, A. Tchamkerten, and D. Tse, Asynchronous Capacity Per Unit Cost , IEEE Inter. Symposium on Information Theory Proceedings (ISIT), pp. 280-284, June 2010.

[15] V. Shah, N. B. Mandayam, and D. J. Goodman, "Power control for wireless data based on utility and pricing", IEEE Proc. of the 9th Intl. Symp. Personal, Indoor, Mobile Radio Communications (PIMRC), Boston, MA, pp. 427-1432, Sep. 1998.

[16] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks", IEEE Trans. on Communications, vol. 50, No. 2, pp. 291-303, Feb. 2002.

[17] S. Yatawatta, A.P. Petropulu, and C.J. Graff, "Energy-efficient channel estimation in MIMO systems", EURASIP Journal on Wireless Communications and Networking, v.2006 n.2, pp. 317-320, April 2006.

[18] S. Buzzi, H. V. Poor, and D. Saturnino, "Energy-efficient resource allocation in multiuser MIMO systems: a game-theoretic framework, Proc. of the 16th European Signal Processing Conference (EUSIPCO 2008), Losanna (Svizzera), August 2008.

[19] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection", IEEE Journal on Selected Areas in Communications (JSAC), pp. 1478-1491, Sep. 2007.

[20] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state", IEEE Transactions on Information Theory, 56(6): 2845-2866, 2010.

[21] C. C. Gaudes, and E Masgrau, "Bounds on Capacity over Gaussian MIMO Multiaccess Channels with Channel State Information Mismatch," IEEE ICASSP-06, vol. 4, pp. 93-96, May 2006

[22] T. L. Marzetta, and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", *IEEE Trans. on Information Theory*, Vol. 45, No. 1, Jan. 1999.

[23] B. Hassibi, and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?", *IEEE Trans. on Information Theory*, Vol. 49, No. 4, April, 2003.

[24] Polyanskiy, H. Vincent Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," IEEE Trans. Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.

[25] F. Richter, A. J. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks", Proceedings of VTC Fall, pp. 1-5, Dresden, 2009.

[26] S. K. Jayaweera, "A virtual MIMO-based cooperative communications architecture for energy-constrained wireless sensor networks", IEEE Trans. Wireless Commun., vol. 5, pp. 984-989, May 2006.

[27] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," IEEE Trans. Commun., vol. 52, pp. 1885-1894, Nov. 2005.

[28] D. Buckingham, and M.C. Valenti, "The information-outage probability of finite-length codes over AWGN channels", IEEE Information Sciences and Systems, 2008. CISS 2008.

[29] C.E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948.

[30] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A Random Matrix Approach to the Finite Blocklength Regime of MIMO Fading Channels", IEEE International Symposium on Information Theory (ISIT), pp. 2181 - 2185, Cambridge, 2012.

[31] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-Layer Combining of Adaptive Modulation and Coding with Truncated ARQ Over Wireless Links, IEEE Trans. Wireless Commun., vol. 3, no. 5, pp. 1746-1755, Sept. 2004.

[32] L.H. Ozarow, S. Shamai, and A.D. Wyner, "Information theoretic considerations for cellular mobile radio", IEEE Vehicular Technology, vol. 43, No. 2, pp 359-378, May. 1994.

[33] FCC Sec.15.249, *Operation within the bands 902-928 MHz, 2400-2483.5 MHz, 5725-5875 MHZ, and 24.0-24.25 GHz*, Oct. 2011.

[34] T. M. Cover, and J. A. Thomas, "Elements of information theory", Wiley-Interscience, New York, NY, 1991.

[35] S. Lasaulce, and N. Sellami, "On the Impact of using Unreliable Data on the Bootstrap Channel Estimation Performance", *Proc. IEEE SPAWC*, Italy, June 2003, pp. 348-352.

[36] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems", IEEE Journal on Selected Areas Communications, vol 26, pp. 1341-1365, 2008.

[37] V. Rodriguez, "An Analytical Foundation for Ressource Management in Wireless Communication", IEEE Proc. of Globecom, San Francisco, CA, USA, pp. 898-902, Dec. 2003.

[38] E. A. Jorswieck, and H. Boche, "Outage probability in multiple antenna systems", European Transactions on Telecommunications, vol. 18, pp. 217-233, 2006.

[39] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices", PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[40] S. Lasaulce, and H. Tembine, "Game Theory and Learning for Wireless Networks: Fundamentals and Applications", Academic Press, pp. 1-336, Aug. 2011, ISBN 978-0123846983.

[41] S. Lasaulce, M. Debbah, and E. Altman, Methodologies for analyzing equilibria in wireless games, IEEE Signal Processing Magazine, vol. 26, no: 5, pp. 41-52, Sep. 2009.

[42] 3GPP TR 36.814 v0.4.1, "Further Advancements for E-UTRA", *Physical Layer Aspects*, Feb. 2009.

[43] M. Alexander, F. A. Graybill, and D. C. Boes (1974). *"Introduction to the Theory of Statistics"* (Third Edition, p. 241-246). McGraw-Hill. ISBN 0-07-042864-6.

[44] Digital Library of Mathematical Functions, Incomplete Gamma functions, 8.11. [Online] Available: http://dlmf.nist.gov/8.11#i

[45] L. Saker, S-E. Elayoubi and H-O. Scheck, *System selection and sleep mode for energy saving in cooperative 2G/3G networks*, IEEE VTC 2009-Fall.

[46] S-E. Elayoubi and O. Ben Haddada, *Uplink intercell interference and capacity in 3G LTE systems*, IEEE ICON 2007.

[47] L. Saker, S-E. Elayoubi, R. Combes and T. Chahed, *Optimal control of wake up mechanisms of femtocells in heterogeneous networks*, Selected Areas in Communications, IEEE Journal on 30 (3), 664-672.

[48] C.X Wang, X. Hong, X. Ge and X. Cheng, *Cooperative MIMO channel models: A survey*, IEEE communications magazine, Vol: 48, issue 2, 2010.

[49] C. Desset et al. , *Flexible power modeling of LTE base stations*, IEEE Wireless Comm. and Networking Conference: Mobile and Wireless Networks, 2012.

[50] A. B. Carleial, "Interference channels", IEEE Trans. on Information Theory, vol. 24, no. 1, pp. 60-70, 1978.

[51] H. Takagi, L. Kleinrock, "Output processes in contention packet broadcasting systems", IEEE Trans. on Comm., vol. 33, no. 11, pp. 1191-1199, 1985.

[52] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, "Modeling TCP Throughput: A simple model and its empirical validation", SIGCOMM Comput Comm. Rev, Volume: 28, Issue: 4, Publisher: ACM, Pages: 303-314, 1998.

[53] S. M. Betz, H. V. Poor, "Energy efficient communications in CDMA networks: A game Theoretic analysis considering operating costs", IEEE Trans. on Signal Processing, 56(10-2): 5181-5190, 2008.

[54] H. Alvestrand, S. Holmer, and H. Lundin, A Google Congestion Control Algorithm for Real-Time Communication on the World Wide Web, 2013, IETF Internet Draft.

[55] V. Singh, A. Lozano, and J. Ott, Performance Analysis of Receive-Side Real-Time Congestion Control for WebRTC, In Proc. of IEEE Packet Video (Vol. 2013).

[56] S. Lasaulce, Y. Hayel, R. El Azouzi, and M. Debbah, "Introducing hierarchy in energy games", IEEE Transactions on Wireless Communications," Vol. 8, No. 7, pp. 3833–3843, Jul. 2009.

[57] R. W. Wolff, "Stochastic modeling and the theory of queues", Englewood Cliffs, NJ: Prentice-Hall, 1989.

[58] F. Baccelli and B. Blaszczyszyn, Stochastic geometry and wireless networks: Theory, Foundations and Trends in Networking, vol. 3, no. 3-4, pp. 249449, 2009.

[59] F. Meshkati, H. Poor, S. Schwartz, "Energy efficiency - delay tradeoffs in CDMA networks: A game-theoretic approach", in Trans. on Information Theory, vol. 55, no. 7, 2009.

[60] D. Fudenberg and J. Tirole, "Game theory", MIT Press, 1991.

[61] A. Cournot, "Recherches sur les principes mathématiques de la théorie des richesses", 1838 (Re-edited by Mac Millian in 1987).

[62] C. Papadimitriou, "Algorithms, games, and the internet." Proceedings of the thirty-third annual ACM symposium on Theory of computing. ACM, 2001.

[63] D. Braess, "Über ein Paradoxon aus der Verkehrsplanung." Unternehmensforschung 12.1 (1968): 258-268.

[64] J.B Landre, Z.E Rawas, R.Visoz, and S. Bouguermouh, "Realistic Performance of LTE in a macro-cell environment", IEEE VTC 2012

[65] R. Combes, S.E. Elayoubi, Z. Altman, "Cross-layer analysis of scheduling gains: Application to LMMSE receivers in frequency-selective Rayleigh-fading channels". WiOpt - 2011, 133-139.

[66] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," IEEE Trans. Signal Process., Vol. 57, No. 5, pp. 1917-1935, May 2009.

[67] B. Larrousse and S. Lasaulce, "Coded power control: performance analysis", IEEE International Symposium on Information Theory, July 2013.

[68] Kurtenbach, Alfred J., and Paul A. Wintz, "Quantizing for noisy channels.", Communication Technology, IEEE Transactions on 17.2 (1969): 291-302.

[69] D. Brice, S. Lasaulce, and A. G. Klein, "Practical quantize-and-forward schemes for the frequency division relay channel.", EURASIP Journal on Wireless Communications and Networking 2007.

[70] A. Gjendemsj, D. Gesbert, G.E. Oien, and S.G. Kiani, "Binary power control for sum rate maximization over multiple interfering links", IEEE Transactions on Wireless Communications, 7(8), 3164-3173, 2008.