Impact of Processing Energy on the Capacity of Wireless Channels

by

Pamela Youssef-Massaad

BE in Computer and Communications Engineering
American University of Beirut, 2003

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Masters of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2005

Author . . . .
Department of Electrical Engineering and Computer Science
anuary, 2005

Certified by . . . .
Muriel Médard
Esther & Harold E. Edgerton Associate Professor
Thesis Supervisor

Certified by . . . .
g Zheng
rofessor
pervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Impact of Processing Energy on the Capacity of Wireless Channels

by

Pamela Youssef-Massaad

Submitted to the Department of Electrical Engineering and Computer Science
on January, 2005, in partial fulfillment of the
requirements for the degree of
Masters of Science in Electrical Engineering and Computer Science

Abstract

Power efficiency is a capital issue in the study of mobile wireless nodes owing to constraints on their battery size and weight. A careful examination of the power consumption in low-power nodes shows that, as the total power available to such nodes decreases, the ratio of power consumed for transmission purposes to the power consumed on other non-transmission processes also decreases. The latter power therefore constitutes a considerable fraction of the total power available to such devices.
We perform our study in terms of power per unit of time, or energy. Traditional information theoretic energy constraints consider only the energy used for transmission purposes. We study optimal transmission strategies by explicitly taking into account the energy expended by processes other than transmission, that run when the transmitter is in the 'on' state. We term this energy by 'processing energy'.
We first derive the capacity of a single user Additive White Gaussian Noise (AWGN) channel in the presence of processing energy. We prove that, unlike the case where only transmission energy is taken into account, achieving capacity may require intermittent, or 'bursty', transmission.
We show that in the low Signal to Noise Ratio (SNR) regime, burstiness becomes optimal when the processing energy is greater than half the square of the total energy available to the transmitter.
This analysis is extended to the AWGN multiple access channel with M senders and a single receiver. We first show that, under a processing energy constraint, Time Division Multiple Access (TDMA) outperforms other known multiple access techniques in the maximization of the sum rate. We prove that, for that same purpose, burstiness is capacity achieving in the low SNR regime when the sum of the ratios of total energy to processing energy is less than unity. Moreover, we present numerical results that show the improvement in the shape of the general two-user achievable rate region obtained with a bursty transmission scheme. We compare the rates obtained by bursty signaling to the rates that can be achieved by TDMA and to the Cover-Wyner region under a processing energy constraint.
Finally, we show that, in low SNR regime, a time-variable channel can be analyzed

as a channel with no variability but with processing energy. In fact, the results we obtain for the bursty signaling in time-variable channels from the processing energy correspondence match those of other studies ([3], [4], [5]) that do not make use of the processing energy argument.
This leads to posit a model in which energy can be regarded as a unifying cost or penalty for various communication impediments.

Thesis Supervisor: Muriel Médard
Title: Associate Professor

Thesis Supervisor: Lizhong Zheng
Title: Associate Professor

Acknowledgments

I thank my advisor Professor Muriel Médard for her guidance and support. It is a great privilege to work with such a highly talented and devoted person who is always very supportive. Being her student has greatly helped in my development as a graduate student as well as a person.

I also thank Professor Lizhong Zheng for co-supervising my thesis; his smart comments and sense of synthesis have greatly helped in my understanding of the material.

I thank Prof. Munther Dahleh for his support and his helpful suggestions that have helped me throughout my masters.

I thank my parents, Gaby and Jacky Massaad. Without them, I would have never arrived to this stage of my life. Thanks mum for your love, affection, prayers and your continuous support as a friend and as a mother, despite the distance. Thanks pap for your love, devotion and sacrifices; I learn a lot from your advice and your support gives me confidence in myself. Kisses to my brother Elie and my sister Joey that I love very much; they give me the nice feeling of being surrounded by a perfect family.

I would also like to thank Habib Haddad for sharing my life with its good and bad times. Thanks for being by my side Beebz!

My thanks also go out to all my friends, especially Maya, Johnny, Rita, Fang, Angelina, my friends in the lab as well as my friends abroad.

Contents

List of Figures

Chapter 1

Introduction

1.1 Computer and Sensor Nodes Literature for Energy Efficiency

Portable and hand-held computers must be careful not to waste the scarce energy resources in their batteries. Even though battery technology is improving continuously and processors and displays are rapidly improving in terms of power consumption, power-aware designs and algorithms applied to various aspects of the communication process continue to be a crucial way to optimize consumption.

Conservation of power has been addressed using many techniques such as variable speed CPUs, flash memory and so on. While power-aware components and algorithms are essential building blocks for portable systems ([19, 28, 29, 30, 20, 31]), saving power in the transmission process is crucial. Consider, for instance, low-power nodes. We note that, when the total power available to such nodes decreases, the ratio of power spent for transmission purposes as compared to power spent for non-transmission purposes becomes significantly small. The latter power constitutes a considerable fraction of the total power available to such devices.

For instance, in the computer and sensor nodes literature, various techniques have been proposed to reduce the mobile host's power consumption during operation:

In [32], Chandrakasan et al. studied power-aware techniques to minimize power consumption of wireless microsensor systems.

In [19], the authors proposed techniques to reduce energy consumption for mobile computers by using extra dedicated low-power modules to cut on processor cycles of the main CPU. Moreover, because networking consumes a large part of the battery resources, the authors used network interfaces with a power aware network protocol that minimizes the 'on'-time of network interfaces.

In [29], the authors studied mobile power management for maximum battery life in mobile wireless communication networks. Their objective is to minimize mean energy consumption subject to maintaining a given quality of service. This was done by unbalancing power levels so that users see alternating periods of interference and decide on transmission.

For laptop computers where the energy use is dominated by the display and the disk, [22] and [21] examined the use of techniques to reduce this consumption by turning the devices off after a period of no use.

For smaller computing devices which often have no disk and which eliminate the display-related power, power consumed by the CPU is significant. In [20], the authors considered reducing the energy used for executing instructions and dynamically varying chip speed so as to reduce energy consumption.

In the case of video applications, encoding and decoding become expensive in terms of power; Low power encoding and decoding schemes have been researched extensively under limited power constraints, like for instance in [27].

When minimizing the total energy, it is fundamental to consider, besides the energy spent on transmission purposes, non-transmission energies. In particular, the energy cost incurred with the state of the channel being 'on', appears to constitute an important fraction of the total energy expended in wireless devices.

At the intersection between the communication theory and the networking fields:

El Gamal et al. proposed an optimal scheduling algorithm to minimize transmission energy by maximizing the transmission time for buffered packets, [24].

In [23], Cui et al. considered wireless applications, where nodes operate on batteries, and analyzed the best modulation strategy to minimize the total energy consumption, when error-control codes are used.

In [25], the authors analyzed the best modulation strategy to minimize the total energy consumption, while satisfying throughput and delay requirements.

The last 3 rows of table (1-1) show the relatively high amount of power consumed in a PDA and a laptop by the mere fact of being 'on'. This fact was also recognized by

Device	Sleep (mW)	Idle (mW)	Wakeup (mW)	Turn-on Time
Wavelan (2.4 Ghz)	177.3284	1318.857	N/A	100ms
Wavelan (915 Mhz)	143.0023	1148.601	N/A	100ms
Metricom	93.50762	346.984	N/A	5 sec
Infrared	N/A	349.607	431.034	100ms
Newton PDA	164.1871	1187.75	N/A	N/A
Magic Link PDA	312.03	700	N/A	N/A
Laptop	N/A	8000	N/A	N/A

Figure 1-1: Power consumption at idle times

Kravertz et al. who argued that: "when inserted, many wireless communication devices consume energy continuously" and that "this energy consumption can represent over 50% of total system power for current hand-held computers and 10% for laptop computers". Kravertz et al. then suggested software-level techniques to suspend the mobile host's device during idle periods of the communication ([11]).

The information theoretic literature has also extensively considered the energy cost of being 'on'. In particular, Verdú investigated in [10], the minimum cost incurred by the transmission of one bit of information through a noisy channel, characterizing the most economical way to communicate reliably. This approach is well suited for continuous changes and presents very general results. Applying those to include a cost associated with the 'on' state, would not lead, however, to continuous results since, in this case, a step function would be involved.

In this thesis, we take explicitly into account the energy of being 'on', which we term 'processing energy', in addition to the energy used for transmission purposes. Under this new constraint, we derive the capacity of an Additive White Gaussian Noise (AWGN) channel. We prove that, unlike the case where only transmission energy is taken into account, achieving capacity may require intermittent, or 'bursty', transmission. We define the burstiness of signaling as the ratio of time in which transmission occurs. We distinguish two transmission schemes, bursty vs. non-bursty. Each of these schemes can become optimal given a certain amount of processing energy and a given total energy constraint. For instance, we show that in the low Signal to Noise Ratio (SNR) regime, burstiness becomes optimal when the processing energy is greater than half the square of the total energy available to the transmitter.

This analysis is extended to the AWGN multiple access channel with M senders and a single receiver. We first show that, under a processing energy constraint, Time Division Multiple Access outperforms other multiple access techniques for the purpose of maximizing the sum rate. We prove that, for the same purpose, burstiness is capacity achieving in the low SNR regime when the sum of the ratios of total energy to processing energy is less than unity. Next, we present numerical results that show the improvement in the shape of the general two-user achievable rate region under a processing energy constraint as compared to the TDMA curve on one hand and to the Cover-Wyner region obtained under processing energy constraint, on the other hand. Moreover, we show that, in low SNR regime, a channel with a certain time variability as well as a channel with processing energy lead to the same intuition regarding burstiness of signaling, since the variability of a channel can simply be treated as an energy penalty, or a processing cost, as in our study. We show that the same expression of burstiness as a function of the SNR is obtained for both channels.

Throughout this study, we neglect the energy cost associated with the transitioning to the 'on' state; the effect of state transition energy cost would only be manifested if a delay metric, such as an error exponent, was used.

1.2 Thesis Outline

The remainder of the thesis is organized as follows:

Chapter 2 deals with the single user case. In section 2.2, we modify the energy constraint of a single user AWGN channel to include the processing energy. Then, we derive the capacity of the channel in terms of the optimal transmitted energy and the optimal burstiness of signaling. In section 2.3, we determine the conditions that make burstiness capacity achieving, in terms of the total available energy and the processing energy.

The first three subsections of section 3.1 give a brief literature on Gaussian multiple access channels. In subsection 3.1.4 we briefly review some multiple access techniques. In section 3.2, we include the processing power in the energy constraint in the multiple user case. Then, for the purpose of maximizing the total sum rate, we derive the condition that makes burstiness capacity achieving in the low SNR regime. We show that TDMA outperforms other multiple access techniques. Section 3.4 shows a graphical illustration of the effect of processing energy on TDMA. Section 3.5 gives numerical results about the shape of the capacity region in the presence of processing energy as compared to the Cover-Wyner region obtained by merely subtracting the processing costs.

Chapter 4 gives a brief background on how to characterize the channel variability of channels with an intermediate coherence level then shows the correspondence between the effect of processing energy from one side and channel variability on the other side on the burstiness of signaling.

Chapter 5 summarizes our conclusions and presents possible future work.

Finally, a review of the Lambert$W$ function can be found in the Appendix.

Chapter 2

Processing Energy in the Single Access Case

2.1 The Model

The classical information theory problem of characterizing capacity for an Additive White Gaussian Noise (AWGN) channel under an energy constraint strives to answer the question: How much information can be transmitted through a channel, and how? This traditional approach considers that all the available energy without inefficiency or overhead is consumed as radiated energy for transmission. Owing to the concavity of $\log(1 + x)$, transmission should have constant energy. Unlike the case when transmission accounts for all the expended energy, we take explicitly into account the processing energy. We show that the processing energy may lead, instead, to making bursty transmission capacity achieving.

We consider a bandlimited AWGN channel, as shown in figure 2-1, with one sender and one receiver.

For this channel, we use a sampled time that we denote by the variable i. The input signal is constrained in energy and in bandwidth and is distorted by additive and bandlimited white Gaussian noise. The noise samples are mutually independent and identically distributed Gaussian random variables with zero-mean and variance σ^2: $Z(i) \sim \mathcal{CN}(0, \sigma^2)$.

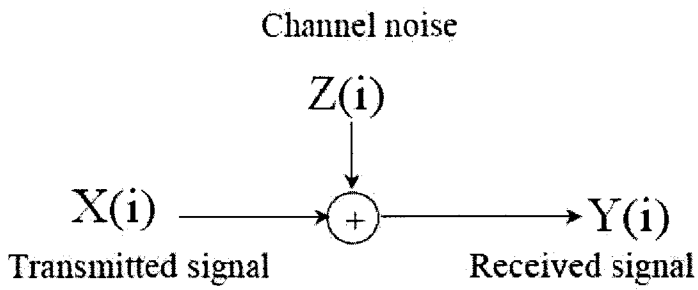

Figure 2-1: AWGN channel

The output of the channel at time i is given by:

$$Y(i) = X(i) + Z(i). \tag{2.1}$$

2.2 Inclusion of Processing Energy in Energy Constraint

2.2.1 Capacity in Terms of Transmission Energy

Let us first review the capacity of an AWGN channel [9], with energy constraint $\mathcal{E}$, noise variance σ^2 (product of bandwidth and noise spectral density), and under no processing energy:

$$C = \log(1 + \frac{\mathcal{E}}{\sigma^2}). \tag{2.2}$$

The energy constraint over n samples is given by:

$$\frac{\sum_{i=1}^{n} P_i}{n} \leq \mathcal{E}, \tag{2.3}$$

where $P_i = \overline{X^2(i)}$ and $X(i)$ is the input to the channel at time i. Please refer to the comments on n in the remarks later.

In what follows, we let 1_i be the indicator function taking the value 0 or 1 according to whether at time i the transmitter is sending or not, P_i be the actual energy used for transmission at time i and $\mathcal{E}$ be the total energy available per channel use. When the processing energy is taken into account and is given by ϵ, the system that we seek to solve becomes:

$$\max \frac{1}{n}\sum_{i=1}^{n} 1_i \log(1 + \frac{P_i}{\sigma^2}) \tag{2.4}$$

$$\text{s.t.} \ \frac{\sum_{i=1}^{n} 1_i(P_i + \epsilon)}{n} \leq \mathcal{E}. \tag{2.5}$$

Remarks:

- We are considering arbitrarily large time n while maximizing the sum in (2.4).
- For a fixed n, this is equivalent to dividing the channel into n subchannels with power constraints $P_1 ... P_n$ respectively and assuming independent coding over the subchannels.
- In writing (2.4), we assumed that the transmitter and the receiver have agreed on the time at which transmission should occur.
- Setting ϵ to 0, (2.5) leads back to the constraint in (2.3) and the corresponding maximized value of (2.4) would be given by (2.2).
- Since burstiness may be capacity achieving, we model the channel as $Y = aX + N$, where a is a binary random variable taking the value 1 or 0 according to whether the sender is 'on' or 'off', all other variables being Gaussian random variables,[9].

  In order to find the capacity of the channel, we aim to maximize

  $$I(a, X; Y) = I(a; Y) + I(X; Y|a). \tag{2.6}$$

  When $a = 0$, $Y = N$ and $I(X; Y|a = 0) = 0$ since the noise is independent of the input.

Let $\Theta = P(a = 1)$. Then,

$$\begin{aligned}
I(X;Y|a) &= \Theta I(X;Y|a=1) + (1-\Theta) I(X;Y|a=0) && (2.7)\\
&= \Theta I(X;Y|a=1) && (2.8)\\
\text{so, } I(a,X;Y) &= I(a;Y) + \Theta I(X;Y|a=1) && (2.9)\\
&= I(a;Y) + \Theta \log(1 + \frac{\mathsf{SNR}}{\Theta} - \epsilon) && (2.10)\\
&\leq H(a) + \Theta \log(1 + \frac{\mathsf{SNR}}{\Theta} - \epsilon) && (2.11)\\
&\leq 1 + \Theta \log(1 + \frac{\mathsf{SNR}}{\Theta} - \epsilon). && (2.12)
\end{aligned}$$

since a is a discrete random variable. For high SNR, clearly the second term dominates as compared to 1.

For the low SNR case, we need to express $I(a,Y)$ as $I(a;Y) = H(a) - H(a|Y)$. In this case, a more complicated proof is needed to prove that for low SNR, the order of this term is less than the order of $\mathsf{SNR}^{1+\alpha}$ where $\alpha > 0$.

We have neglected $I(a;Y)$ in both low and high SNR regimes. The result we obtain is therefore, at least, an achievable rate, the converse not being necessarily true. However, as will be shown later, the scheme we propose readily gives great improvements over the rate region obtained when no burstiness is allowed.

When we transmit, i.e. when $1_i = 1$, the concavity of the function $\log(1+x)$ implies that P_i should, at any time i for which $1_i = 1$, be equal to a constant, say ν.

We denote the *burstiness of signaling* by

$$\Theta \stackrel{def}{=} \frac{1}{n} \sum_{i=0}^{n} 1_i. \tag{2.13}$$

When $\Theta = 1$, a constant transmission strategy is being used; when $\Theta < 1$, bursty signals are being transmitted. The smaller the Θ, the burstier the transmission.

Given a certain total amount of energy $\mathcal{E}$, how would an increase of the processing energy, ϵ, affect the optimal transmission mode? Clearly, since the optimal strategy should avoid paying too much overhead owing to the processing energy cost, sending more bursty signals would result in higher rates. On the other hand, transmission should not be too bursty, since placing all the energy in one time slot may result in a loss of rate. Therefore, there is a certain tradeoff between sending bursty signals and adopting a continuous transmission strategy. This tradeoff is controlled by the values of $\mathcal{E}$ vs. ϵ.

Back to the constraint in (2.5), since additional total energy can only be beneficent, we take the constraint to equality:

$$\Theta(\nu + \epsilon) = \mathcal{E}. \tag{2.14}$$

Hence, we denote (3.15) in terms of ν by,

$$C_1(\nu) \stackrel{def}{=} \frac{\mathcal{E}}{\nu + \epsilon} \log(1 + \frac{\nu}{\sigma^2}). \tag{2.15}$$

Taking the derivative with respect to ν and setting it to 0, we obtain because of convexity the optimal value of the energy for which the capacity is maximized in the presence of processing energy

$$\nu_{opt} = \frac{\epsilon - \sigma^2}{W(\frac{\epsilon - \sigma^2}{e\sigma^2})} - \sigma^2. \tag{2.16}$$

As seen in the expression of ν_{opt}, we have used the Lambert$W$ function, which we denoted by W. A brief review of the function can be found in the appendix.

Note that the optimal value of the transmission energy is independent of the total available energy $\mathcal{E}$ and is merely dependent on the processing energy ϵ.

The maximum value of (3.15) that can be achieved in the presence of processing energy is

$$C_1(\nu_{opt}) = \frac{\mathcal{E}}{\nu_{opt} + \epsilon} \log(1 + \frac{\nu_{opt}}{\sigma^2}), \tag{2.17}$$

where ν_{opt} is given by (2.16). We examine the second derivative constraint in section 2.2.2.

There are two constraints on the value of the energy,

$$\nu \geq 0 \tag{2.18}$$

and

$$\nu \geq \mathcal{E} - \epsilon \text{ from } \Theta \leq 1. \tag{2.19}$$

For unit noise variance, we can show that $\nu_{opt} \geq 0$ for all $\epsilon \geq 0$. The functions $\epsilon - 1$ and $W(\frac{\epsilon-1}{e})$ are plotted in figure 2-2. For $0 < \epsilon < 1$, both functions are negative and $\epsilon - 1 < W(\frac{\epsilon-1}{e})$, so the ratio is positive and greater than unity, consequently $\nu_{opt} > 0$. For $\epsilon > 1$, both functions are positive and $\epsilon - 1 > W(\frac{\epsilon-1}{e})$, so $\nu_{opt} > 0$. Values of ν_{opt} are illustrated in figure 2-3.

For the second constraint, we should find the values of ϵ and $\mathcal{E}$ for which $\nu_{opt} \geq \mathcal{E} - \epsilon$. Substituting for ν_{opt} from (2.16) and simplifying, we obtain the equivalent constraint in terms of $\mathcal{E}$ and ϵ, i.e.

$$\mathcal{E} \leq \frac{(\epsilon - 1)(1 + W(\frac{\epsilon-1}{e}))}{W(\frac{\epsilon-1}{e})}. \tag{2.20}$$

When this condition is not met, $\nu_{opt} = \mathcal{E} - \epsilon$ becomes the optimal value that maximizes the rate.

For instance, in the case of $\mathcal{E} = 1$, it can be shown that $\nu_{opt} \geq \mathcal{E} - \epsilon$ for all $\epsilon >\approx 0.237$. In that case, the value in (2.16) achieves capacity. Again, we omit the proof for the sake of brevity.

2.2.2 Capacity in Terms of Burstiness of Signaling

We can also maximize the capacity with respect to Θ. Hence, using (2.14), we denote (2.4) by:

$$C_2(\Theta) \stackrel{def}{=} \Theta \log(1 + \frac{\frac{\mathcal{E}}{\Theta} - \epsilon}{\sigma^2}). \tag{2.21}$$

Note that, although different functions, C_2 above and C_1 in (2.17) are describing the same entity, namely the maximum achievable rate in the presence of processing energy, but in terms of different variables. Calculating the derivative with respect to

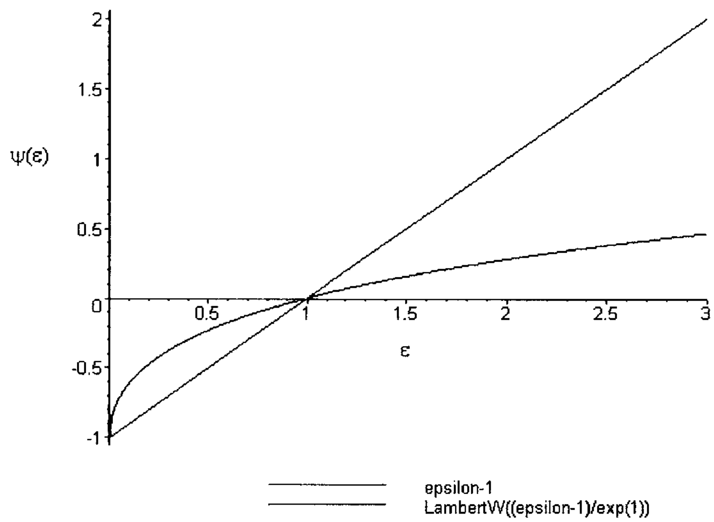

Figure 2-2: Position of the two graphs for all ϵ

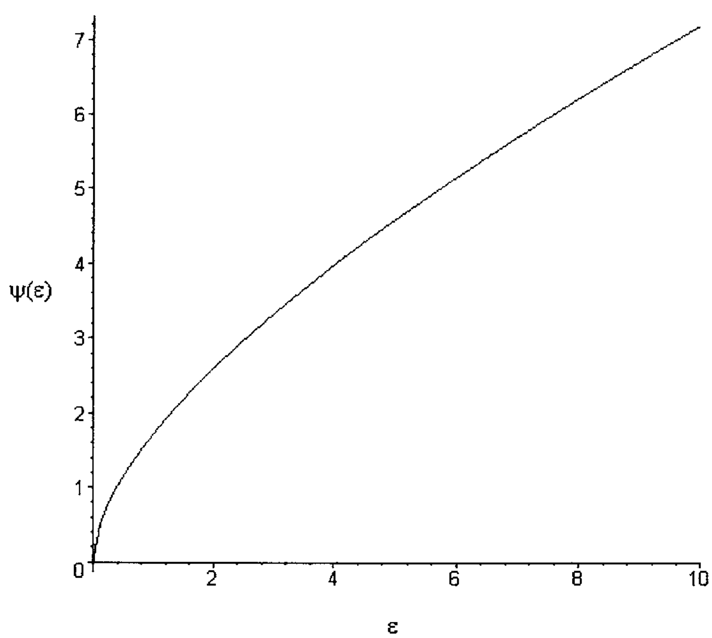

Figure 2-3: Positivity of the optimal energy of transmission for all ϵ

Θ, then, setting it to 0, we obtain the corresponding Θ_{opt} that maximizes the capacity in the presence of processing energy.

$$\Theta_{opt} = \frac{\mathcal{E} W(\frac{\epsilon - \sigma^2}{\sigma^2 \times e})}{(\epsilon - \sigma^2)(W(\frac{\epsilon - \sigma^2}{\sigma^2 \times e}) + 1)}. \tag{2.22}$$

For the remainder of the paper, we assume $\sigma^2 = 1$; under this assumption, the Signal to Noise Ratio (SNR) given by $\frac{\mathcal{E}}{\sigma^2}$ is simply given by $\mathcal{E}$. Similarly, we define the Normalized Cost (NC) as $\frac{\epsilon}{\sigma^2}$, which simply equals the processing energy or ϵ. Then, Θ_{opt} can be reexpressed merely in terms of the SNR and the NC by:

$$\Theta_{opt} = \mathcal{E} \frac{W(e^{-1}(\epsilon - 1))}{(\epsilon - 1)(W(e^{-1}(\epsilon - 1)) + 1)}. \tag{2.23}$$

From this equation we note two things. First, the optimal burstiness of signaling is directly proportional to the SNR. Therefore, the bigger the total energy available, the larger the Θ. Therefore, the channel should be less bursty in the transmission since ϵ becomes negligible as compared to $\mathcal{E}$ and continuous transmission becomes more efficient.

Second, we can know how $\Theta_{opt}/\mathsf{SNR}$ varies with the NC, i.e. with ϵ. Figure 2-4 shows the plot of $\log_{10}(\Theta_{opt}/\mathcal{E})$ vs. ϵ.

As illustrated in figure 2-4, for a given SNR, the burstiness of signaling decreases with increasing NC. This agrees with our intuition, since the higher the cost of being 'on', the lesser the fraction of time the user should transmit, i.e. the burstier the signaling. Now, from the definition of Θ in (2.13), $0 \leq \Theta_{opt} \leq 1$.

We can show that $\Theta_{opt} \geq 0$ for all $\epsilon \geq 0$. For $\epsilon > 0$, $\epsilon - 1 > -1$ and $e^{-1}(\epsilon - 1) > -e^{-1}$. Since the LambertW function is increasing and $W(-e^{-1}) = -1$, we obtain that $W(e^{-1}(\epsilon - 1)) > -1$. Also, from above, $\epsilon - 1 > W(e^{-1}(\epsilon - 1))$.

Next, we seek the values of $\mathcal{E}$ and ϵ for which $\Theta_{opt} \leq 1$. Let us illustrate the effect of these constraints when $\mathcal{E} = 1$ and check the conditions on ϵ. We plot in figure 2-5, $(\Theta_{opt} - 1)$ versus ϵ. As seen in 2.2.1, for $\epsilon \gtrapprox 0.237$, $\Theta_{opt} \leq 1$ and the optimal value is Θ_{opt} as given in (2.22), otherwise $\Theta_{opt} = 1$.

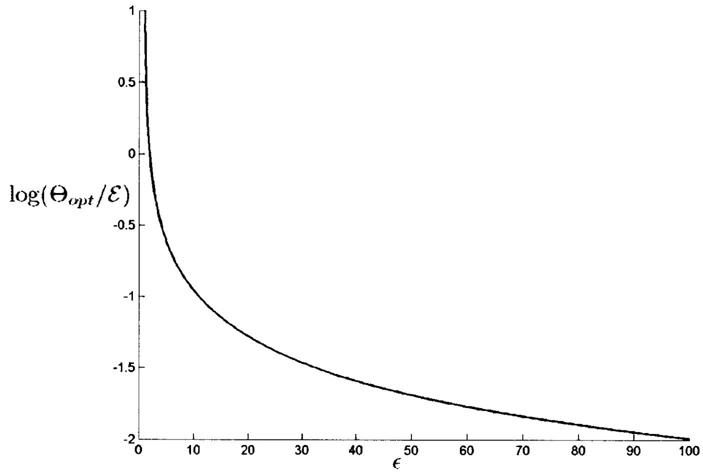

Figure 2-4: Plot of $\log_{10}(\Theta_{opt}/\mathcal{E})$ vs. ϵ

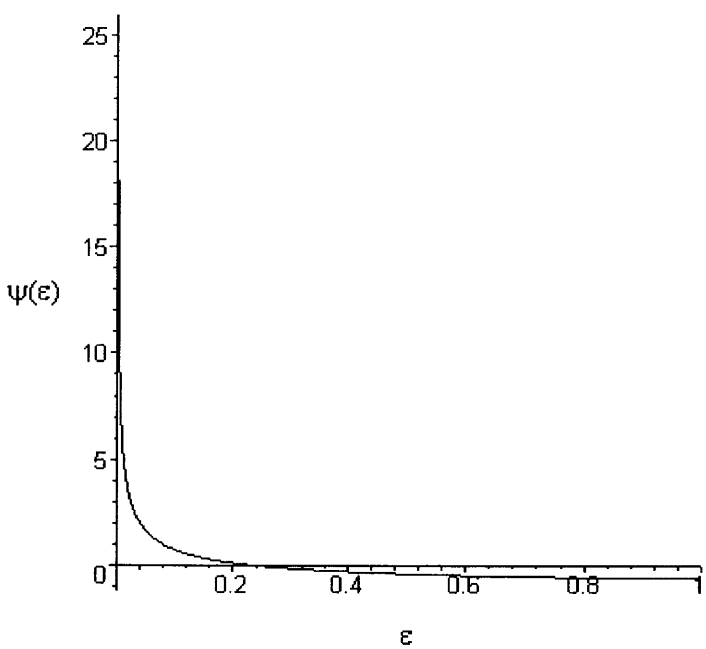

Figure 2-5: $(\Theta_{opt} - 1)$ vs. ϵ

Given certain values of ϵ and $\mathcal{E}$, we calculate the second derivatives of (2.17) and (2.21) and verify that they are non-positive when evaluated at $\nu = \nu_{opt}$ and $\Theta = \Theta_{opt}$ respectively. If this test fails, one of the boundary values of ν_{opt} and Θ_{opt} would maximize the sum in (2.4).

The above example in which we used $\mathcal{E} = 1$ and unit noise variance, illustrates one case in which Θ_{opt} is not always equal to 1, unlike the case where no processing energy is considered. Therefore, burstiness is, in some cases, capacity achieving. We investigate the conditions of such a transmission mode in the following section.

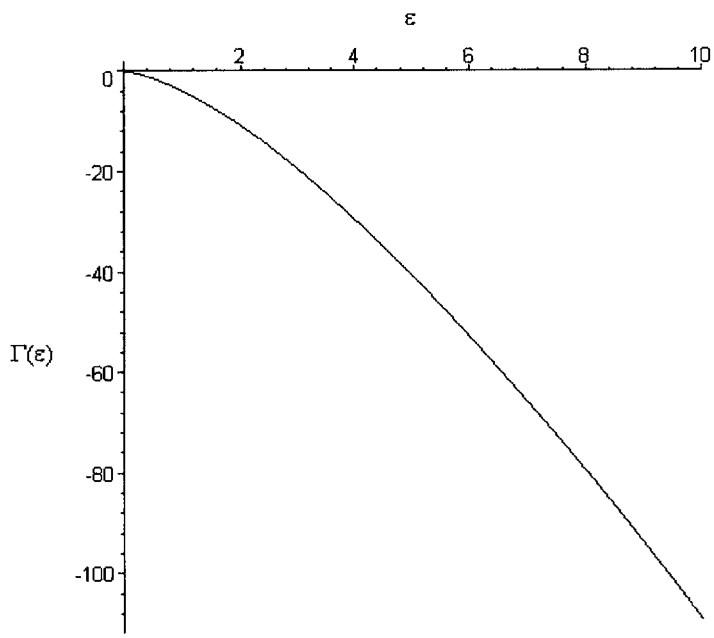

Figure 2-6: Second derivative of Θ_{opt} as a function of ϵ

2.3 Bursty Transmission

In this section, we determine when achieving the capacity requires bursty transmission. For this purpose, we would like to find a relationship between the SNR ($\mathcal{E}$) and the processing energy (ϵ) that allows the transition from the bursty transmission

regime to the constant transmission one. This transition occurs at $\Theta_{opt} = 1$.
Recall the expression of Θ_{opt} in (2.22). We have that

$$\Theta_{opt} = 1 \Leftrightarrow \mathcal{E} = \frac{(\epsilon - 1)(W(\frac{\epsilon-1}{e}) + 1)}{W(\frac{\epsilon-1}{e})}. \tag{2.24}$$

We consider the curve $(\mathcal{E}, \epsilon)$ which separates the regions in which burstiness and constant transmission become capacity achieving. As seen in figure 2-9, the region below the function $\Theta_{opt} = 1$ denotes the region where we would like to be bursty. The region where $\Theta_{opt} = 1$ is where continuous transmission is appropriate and the processing energy can be neglected. We are then back to the original idealized model of wireless channels where there is no need to consider the cost of being 'on'.
Next, we would like to find numerical bounds between these two models.
From the definition of the LambertW function, we recognize that,

$$\lim_{x \to -e^{-1}} W(x) = -1. \tag{2.25}$$

We can refer to Appendix A for the plot of the LambertW function. Then, we have

$$\lim_{x \to 0} W(x - e^{-1}) = -1.$$

We can also show that,

$$\lim_{x \to 0} \frac{W(x - e^{-1}) + 1}{\sqrt{x}} = \sqrt{2e}. \tag{2.26}$$

We can then use 2.26 to approximate the expression in (2.24) as $\epsilon \to 0$, or $\Theta \to 0$ as follows.
Let $y = ex$. Use change of variables on (2.26) to obtain:

$$\lim_{y \to 0} W(e^{-1}(y - 1)) - \sqrt{2y} = -1, \tag{2.27}$$

Therefore, (2.27) shows that for arbitrarily small ϵ, we can substitute $W(e^{-1}(\epsilon - 1))$ by $(\sqrt{2\epsilon} - 1)$.

(2.24) can then be approximated by:

$$\mathcal{E} \approx \frac{\sqrt{2\epsilon}(\epsilon - 1)}{\sqrt{2\epsilon} - 1}. \tag{2.28}$$

(2.28) can be furthermore approximated by using:

$$\lim_{\epsilon \to 0} \frac{\sqrt{2\epsilon}(\epsilon - 1)}{\sqrt{2\epsilon} - 1} - \sqrt{2\epsilon} = 0. \tag{2.29}$$

Therefore, for low SNR we obtain:

$$\mathcal{E} \approx \sqrt{2\epsilon}. \tag{2.30}$$

We illustrate in figure 2-7, the equation in (2.24) and its approximation in (2.30). In general, (2.22) and (2.29) give:

$$\mathcal{E} \approx \Theta_{opt}\sqrt{2\epsilon}. \tag{2.31}$$

Replacing $\Theta_{opt} \leq 1$ in (2.31), we obtain

$$\mathcal{E} \leq \sqrt{2\epsilon}. \tag{2.32}$$

The above equation shows that burstiness is capacity achieving when the total energy available to the channel is less than the square root of double the processing energy, or equivalently when the SNR is less than the square root of double the Normalized Cost (NC). This fact is illustrated in figure 2-9 for small ϵ.

Moreover, the slope of (2.30) at $\epsilon = 0$ goes to ∞. Therefore, in the low SNR regime, when the processing energy is *any percentage* of the total energy, burstiness is capacity achieving. On the other hand, to remain non-bursty, ϵ and $\mathcal{E}$ should satisfy the relationship: $\epsilon \leq \frac{\mathcal{E}^2}{2}$. Under this constraint, constant transmission over the channel is

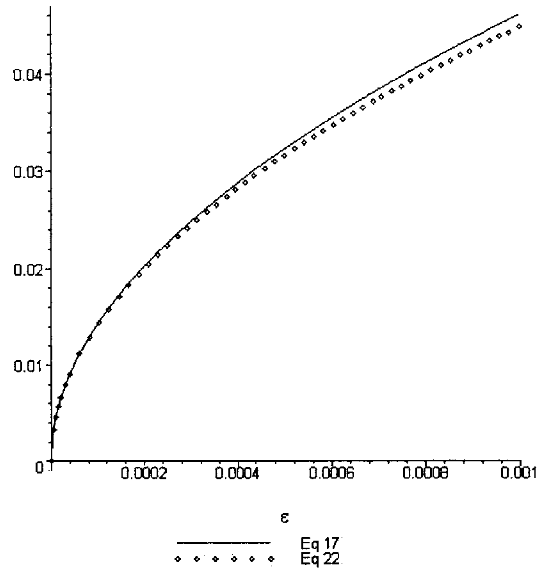

Figure 2-7: Equation (2.24) and its approximation in (2.30) in low SNR regime

optimal. For large SNR, when $\epsilon \to \infty$,

$$\frac{d\mathcal{E}}{d\epsilon} = 1 + \frac{1}{W(\frac{\epsilon-1}{e})} - \frac{1}{W(\frac{\epsilon-1}{e})(1 + W(\frac{\epsilon-1}{e}))} \to 1. \tag{2.33}$$

Thus, the slope of the curve that separates the bursty transmission regime from the non-bursty region tends to 1. In figure 2-8, we plot $\frac{d\mathcal{E}}{d\epsilon}$ as a function of ϵ. Therefore, for large SNR, given a certain ratio $\frac{\mathcal{E}}{\epsilon}$, we can directly deduce the region in which we are situated in figure 2-9 and decide whether burstiness is optimal or not. For instance, for $\epsilon = 0.01\mathcal{E}$, we are clearly above the curve $(\mathcal{E}, \epsilon)$, hence we want to have continuous transmission. This, of course, agrees with our intuition that for a small cost of being 'on', processing energy can be neglected and transmission should remain constant.

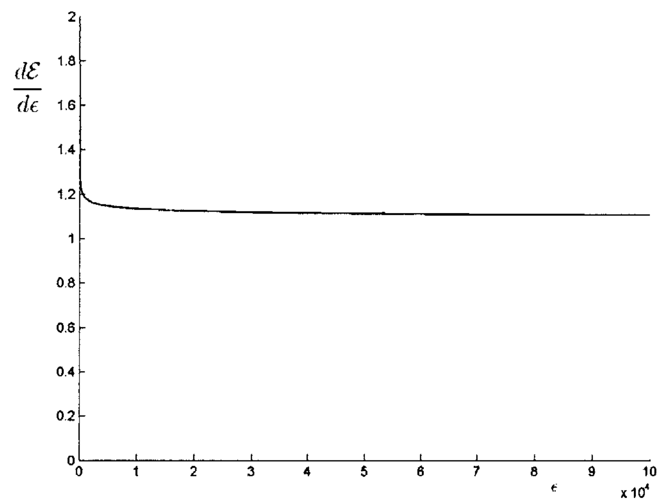

Figure 2-8: Plot of equation (2.34) vs ϵ

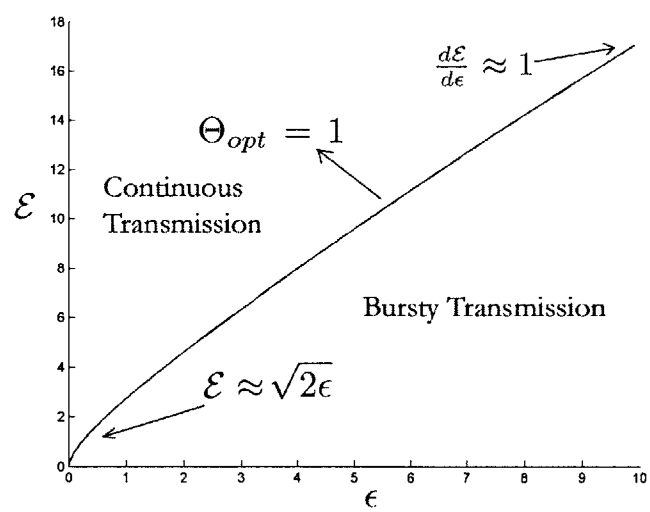

Figure 2-9: Bursty and continuous transmission regimes

Chapter 3

Processing Cost in the Multiple Access Case

3.1 Background on Multiple Access Channels

In this section, we present a brief overview of the classical results on the information-theoretic aspects of multiple access channels (MAC). We confine our discussion to the basic MAC model which is discrete, memoryless, without no channel side information or feedback, and where the two (or more) senders are synchronized. Subsection 3.1.1 introduces the channel model while subsection 3.1.2, establishes the capacity region for a general MAC. Corresponding results for the Gaussian MAC are presented in subsection 3.1.3.

3.1.1 The Channel Model

A multiple access channel is a channel via which two (or more) senders send information to a common receiver. More precisely, in the case of two users:

Definition 1 *(Discrete memoryless MAC) A two-user discrete memoryless MAC consists of three alphabets, $X1$, $X2$ and Y, and a probability transition matrix $p(y|x_1, x_2)$. [9]*

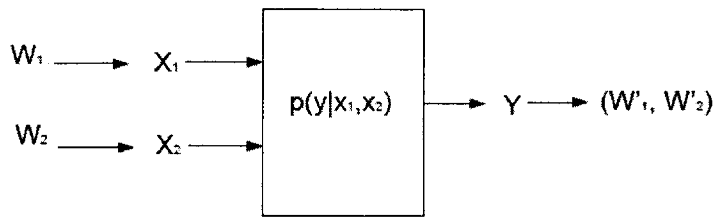

Figure 3-1: The discrete memoryless multiple access channel

The channel is memoryless, i.e.

$$p(y^n|x_1^n, x_2^n) = \prod_{i=1}^{n} p(y^i|x_1^i, x_2^i) \tag{3.1}$$

3.1.2 Capacity Region

Consider two users sharing the same medium, with user i having rate R_i. In 1971, Liao and Alshwede independently derived the conditions that should be satisfied by the rates R_1 and R_2. This is summarized in the following theorem.
Given any two distributions $p_1(x_1)$ and $p_2(x_2)$ over X_1 and X_2, we have

Theorem 1 *(MAC capacity region for given $p_1(x_1)$ and $p_2(x_2)$) For any given $p_1(x_1)$ and $p_2(x_2)$, all pairs (R_1, R_2) satisfying*

$$R_1 < I(X_1; Y|X_2), \tag{3.2}$$

$$R_2 < I(X_2; Y|X_1), \tag{3.3}$$

$$R_1 + R_2 < I(X_1, X_2; Y) \tag{3.4}$$

belong to the capacity region of the MAC.

Theorem 2 *(MAC capacity region) The capacity region C of the MAC is the convex closure of the regions established in Theorem 1 for every possible product distribution $p_1(x_1)p_2(x_2)$ over $\mathcal{X}_1 \times \mathcal{X}_2$.*

Furthermore, the capacity region, $\mathcal{C}$, established above is equivalent to the region specified in the following corollary:

The corresponding capacity region for the two-user case is shown in figure (3-2).

Corollary 1 *The capacity region of the MAC is also given by the convex closure of all* (R_1, R_2) *pairs satisfying*

$$R_1 < I(X_1; Y|X_2, Q), \tag{3.5}$$

$$R_2 < I(X_2; Y|X_1, Q), \tag{3.6}$$

$$R_1 + R_2 < I(X_1, X_2; Y|Q) \tag{3.7}$$

Here Q *is some auxiliary random variable such that* $X_1 \to Q \to X_2$ *and* $Q \to (X_1, X_2) \to Y$ *are two Markov chains. Furthermore,* $|Q|$ *can be as small as* 2.

Note that $\mathcal{C}$ is a closed and convex set.

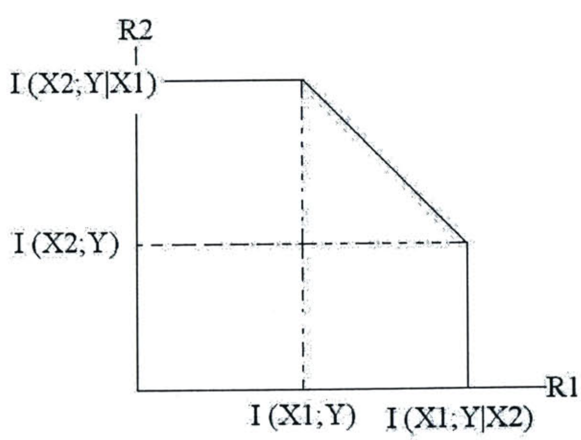

Figure 3-2: Capacity region $\mathcal{C}$ of the MAC

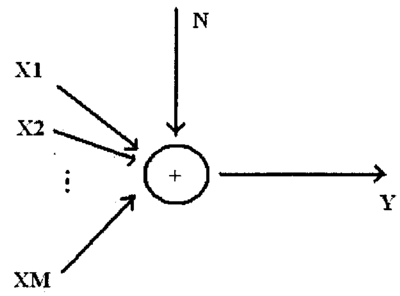

Figure 3-3: Gaussian multiple access channel

3.1.3 Gaussian Multiple Access Channels

The capacity region of the time sampled AWGN channel above with energy constraint $\mathcal{E}^i$ for user i under no processing energy is the subset of R_M containing the rate-tuples $(R_1, .., R_M)$ with nonnegative components satisfying

$$\sum_{i \in S} R^i \leq \log(1 + \frac{\sum_{i \in S} \mathcal{E}^i}{\sigma^2}), \forall S \subseteq \{1, ..., M\}. \tag{3.8}$$

Some points in this capacity region are known to be achievable by successive cancellation. These are the points which satisfy

$$R^i \leq \log(1 + \frac{\mathcal{E}^i}{\sigma^2 + \sum_{j<i} \mathcal{E}^j}). \tag{3.9}$$

The energy constraint for user i is given by

$$\frac{\sum_{k=1}^{n} \mathcal{E}_k^i}{n} \leq \mathcal{E}^i \tag{3.10}$$

where $\mathcal{E}_k^i = E[X_k^{i2}]$ and X_k^i is the input to the channel from sender i at time k, and n is the largest time index. For the two-user case, we show the capacity region in figure (3-4). More background on multiple access channels and intuition about the Cover-Wyner region can be found in [1], [18] and [8].

For the time-sampled AWGN channel shown in figure (3-3), we can use the equations

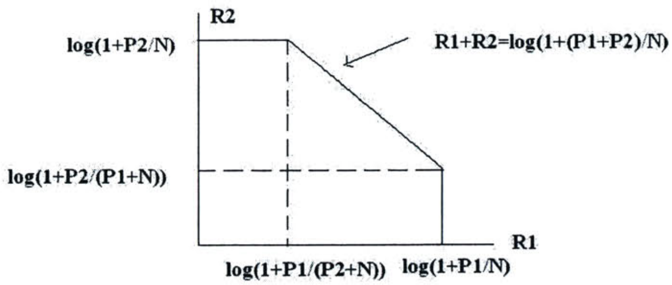

Figure 3-4: Capacity region of the two-user Gaussian MAC with power constraint P_i and noise variance N

in Theorem 1 and obtain the Cover-Wyner region in the Gaussian case for two users as shown in (3-4).

The dominant face of the pentagon corresponds to:

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + \frac{\mathcal{E}^1 + \mathcal{E}^2}{\sigma^2}). \tag{3.11}$$

The two corner points have the following rate pairs:

$$(R_1 = \frac{1}{2}\log(1 + \frac{\mathcal{E}^1}{\sigma^2}); R_2 = \frac{1}{2}\log(1 + \frac{\mathcal{E}^2}{\sigma^2 + \mathcal{E}^1})), \tag{3.12}$$

$$(R_1 = \frac{1}{2}\log(1 + \frac{\mathcal{E}^1}{\sigma^2 + \mathcal{E}^2}); R_2 = \frac{1}{2}\log(1 + \frac{\mathcal{E}^2}{\sigma^2})). \tag{3.13}$$

Those points have the characteristic that they can be achieved by interference cancellation,[8]. To achieve points on the dominant face, time-sharing betwen theses corners can be done. Anoher method, known as rate-splitting, may be used to achieve points on the diominant sum-rate face. In this method, one of the users (say user 1) would be split into two virtual users (virtual user 1 and virtual user 3) by splitting the rate and the energy between these two virtual users, [8]. Splitting one of the users into two gives an extra degree of freedom as to what fraction of power the two virtual users will represent out of the total energy of the original user and allows to move on the dominant face.

3.1.4 Multiple Access Techniques

There are several multiple access techniques for accessing a shared AWGN channel. In this section, we will briefly review TDMA, FDMA and CDMA:

In Time Division Multiple Access (TDMA), different users are assigned to different time slots. Let α be the fraction of time that user 1 transmits. User 1 will then transmit with an energy $\frac{\mathcal{E}^1}{\alpha}$. Therefore, $R_1 = \alpha \log(1 + \frac{\mathcal{E}^1}{\alpha\sigma^2})$. User 2 will then transmit with an energy $\frac{\mathcal{E}^1}{1-\alpha}$.

The capacity region and the TDMA rates are shown in figure 3-5.

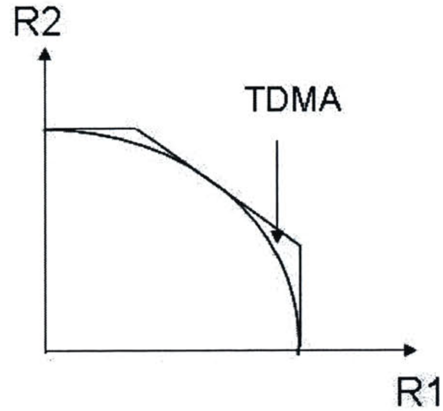

Figure 3-5: Rates achievable by TDMA as compared to the Cover-Wyner region

In Frequency Division Multiple Access (FDMA), different users are assigned different frequencies. If we assign user 1 to frequency W_1, then $R_1 = \frac{W_1}{2} \log(1 + \frac{\mathcal{E}^1}{W_1\sigma^2})$.

In TDMA and FDMA, the maximum sum rate is achieved when the fraction of time or frequency assigned to the user is proportional to the proportion of energies available to the users.

3.2 Inclusion of the Processing Energy in the Energy Constraint

Let $\mathcal{A}$ be the set of active users. For the case of two users, $\mathcal{A} = \{00, 01, 10, 11\}$. For instance, '01' means that user 1 is 'off' while user 2 is 'on'.

Let $Q_k^{\mathcal{A}}$ be the indicator function taking the value 1 if at time k, the given combination of active users out of the set $\mathcal{A}$ is true. $\mathcal{E}^i$ is the total energy available for user i and $P_k^{\mathcal{A}}(i)$ is the actual transmitted energy of this user at time k, given the set of active users.

Now, if the processing energy is taken into account and is given by ϵ^i for user i, the system that we would like to solve becomes

$$\max \frac{1}{n} \sum_{k=1}^{n} \sum_{\mathcal{A}} Q_k^{\mathcal{A}} \log(1 + \frac{\sum_{i=1}^{M} Q_k^{\mathcal{A}} P_k^{\mathcal{A}}(i)}{\sigma^2}) \tag{3.14}$$

$$\text{s.t. } \frac{1}{n} \sum_{k=1}^{n} \sum_{\mathcal{A}} Q_k^{\mathcal{A}} (P_k^{\mathcal{A}}(i) + \epsilon^i) \leq \mathcal{E}^i \text{ for each } i. \tag{3.15}$$

the above system has i constraints.

Note that in writing (3.15), we assumed that the transmitter and the receiver have agreed on the time at which transmission should occur.

When we transmit, i.e. when $Q_k^{\mathcal{A}} = 1$, the concavity of the function $\log(1+x)$ implies that $P_k^{\mathcal{A}}(i)$ should, at any time k for which $Q_k^{\mathcal{A}} = 1$, be equal to a constant, say $\nu^{\mathcal{A}}(i)$.

We denote the *burstiness of signaling* by

$$\Theta^{\mathcal{A}} = \frac{1}{n} \sum_{k=1}^{n} Q_k^{\mathcal{A}}. \tag{3.16}$$

The constraint for user i in (3.15), relaxed to equality, becomes

$$\sum_{\mathcal{A}} \Theta^{\mathcal{A}} (\nu^{\mathcal{A}}(i) + \epsilon^i) = \mathcal{E}^i. \tag{3.17}$$

Finally, an additional constraint would be, that the sum of the fractions of transmission time for the different combinations of users should be less than 1, i.e.

$$\sum_{\mathcal{A}} \Theta^{\mathcal{A}} \leq 1, \tag{3.18}$$

with each $0 \leq \Theta^{\mathcal{A}} \leq 1$.

We illustrate in the following section how the system and the constraints would be written in the two-user case.

3.3 Superiority of TDMA for Maximizing the Sum Rate

As explained in section 3.1.4, in TDMA, different users are assigned to different time slots. Let α be the fraction of time that user 1 transmits. User 1 will then transmit with an energy

$$\frac{\mathcal{E}^1}{\alpha}. \tag{3.19}$$

Therefore,

$$R_1 = \alpha \log(1 + \frac{\mathcal{E}^1}{\alpha\sigma^2}). \tag{3.20}$$

Similarly, for user 2,

$$R_2 = (1 - \alpha) \log(1 + \frac{\mathcal{E}^2}{(1-\alpha)\sigma^2}). \tag{3.21}$$

The TDMA rates are suboptimal except at one point where the TDMA curve touches the dominant face.

In what follows, we shall prove that, when we take into account the processing energy, TDMA achieves a better sum rate than the 45 degree line of the Cover-Wyner region that we obtain by merely removing the processing energy from the total available energy.

For simplicity, we will consider the two-user case:

$\Theta^{\mathcal{A}} = \Theta^{00}, \Theta^{01}, \Theta^{10}, \Theta^{11}$ depending on which senders are transmitting at a time.
User 1 will have a rate that he achieves by himself (R_{1alone}) and a rate achieved when both users are transmitting ($R_{12shared}$). These two rates can be expressed as follows:

$$R_{1alone} = \Theta^{10} \log(1 + \frac{\nu^{10}(1)}{\sigma^2}) \text{ and } R_{12shared} = \Theta^{11} \log(1 + \frac{\nu^{11}(1) + \nu^{11}(2)}{\sigma^2}) \quad (3.22)$$

A similar reasoning can be applied to user 2. Therefore the sum rate would be given by:

$$R_1 + R_2 = \Theta^{10} \log(1 + \frac{\nu^{10}(1)}{\sigma^2}) + \Theta^{01} \log(1 + \frac{\nu^{01}(1)}{\sigma^2}) + \Theta^{11} \log(1 + \frac{\nu^{11}(1) + \nu^{11}(2)}{\sigma^2}) \quad (3.23)$$

In order to find the maximum sum rate, we shall solve the following system:

$$\begin{aligned} \max \quad & R_1 + R_2 && (3.24)\\ \text{s.t.} \quad & \Theta^{10} (\nu^{10}(1) + \epsilon^1) + \Theta^{11}(\nu^{11}(1) + \epsilon^1) = \mathcal{E}^1 && (3.25)\\ & \Theta^{01} (\nu^{01}(2) + \epsilon^2) + \Theta^{11}(\nu^{11}(2) + \epsilon^2) = \mathcal{E}^2 && (3.26)\\ & \Theta^{01} + \Theta^{10} + \Theta^{11} \leq 1. && (3.27) \end{aligned}$$

The last equation is basically equation (3.18) obtained by removing $\Theta^{00} \geq 0$, the fraction of time where the two users are 'off'.
The other energy constraints are equation (3.17) applied to each of the two users.
Consider the case where both senders are ON, i.e. when $\Theta^{11} > 0$. The rate achieved is

$$\log(1 + \frac{\nu^{11}(1) + \nu^{11}(2)}{\sigma^2}). \quad (3.28)$$

In the Cover-Wyner region, this rate is the one obtained from the 45 degree line with user i having an energy of $\nu^{11}(i)$. At the same time, this rate can be obtained by using TDMA with each sender transmitting a fraction α of the time, where

$$\alpha = \frac{\nu^{01}(1)}{\nu^{10}(2) + \nu^{01}(1)}. \quad (3.29)$$

This proof is given in [6]. This argument proves that obtaining this rate does not really necessitate that both users be ON since we can alternate their transmissions with the convenient shares of energies and achieve the same total rate. Therefore, we can increase Θ^{10} and Θ^{01} and set Θ^{11} to 0.

Now, we may investigate the effect of processing energy on the maximum sum rate. TDMA gives an advantage over the non-processing scenario since TDMA will save us part of the processing energy.

We are assuming that the noise is complex Gaussian; for simplicity, we assume unit noise variance and symmetric users having the same $(\mathcal{E}, \epsilon)$.

Since the users are symmetric, they have the same total energy available to them. At the same time, each of the two users will be spending one processing energy for being 'on', therefore the sum rate can be written as:

$$R_1 + R_2 = \log(1 + 2\mathcal{E} - 2\epsilon). \tag{3.30}$$

Consider that, instead of allowing the users to transmit together, TDMA is used. Then, the rates are as follows

$$R_1 = \alpha \log(1 + \frac{\mathcal{E}}{\alpha} - \epsilon), R_2 = (1 - \alpha) \log(1 + \frac{\mathcal{E}}{1 - \alpha} - \epsilon). \tag{3.31}$$

Since we are in the symmetric case, TDMA achieves the maximum sum rate when $\alpha = 0.5$. Therefore, the total sum rate is

$$\log(1 + 2\mathcal{E} - \epsilon). \tag{3.32}$$

Comparing the sum rate alone to the sum of the rates in (3.31), we see that TDMA achieves a sum rate larger than the one achieved by allowing Θ^{11} to be non-zero. Therefore, we can always achieve a better sum rate by using TDMA in the presence of processing energy.

In the M-user case, the more we allow users to send together, i.e. $\Theta^{11 \ldots 1} > 0$, the more we expend unnecessary processing energy. Using TDMA will allow only one user to send at a time and only the processing energy of this user to be expended. Therefore, the benefit of TDMA will increase as more and more users share the medium as it is saving (M-1) times the processing energy. At the same time, the effect of processing energy will decrease as the number of users increases since only the processing energy of one user is expended with TDMA, owing to the fact that we can have at most one user transmitting at one time.

Recognizing that for the purpose of maximizing the sum rate, Θ^{11} should be set to 0, we can go back and solve the original system of equations for the two-user case, in (3.24), which is repeated below.

$$\max\ \Theta^{10}\log(1+\frac{\nu^{10}(1)}{\sigma^2})+\Theta^{01}\log(1+\frac{\nu^{01}(1)}{\sigma^2}) \tag{3.33}$$

$$\text{s.t. } \Theta^{10}(\nu^{10}(1)+\epsilon^1)\le\mathcal{E}^1 \tag{3.34}$$

$$\Theta^{01}(\nu^{01}(2)+\epsilon^2)\le\mathcal{E}^2 \tag{3.35}$$

$$\Theta^{01}+\Theta^{10}\le 1. \tag{3.36}$$

Using Lagrange multipliers, the first constraints will be biting. Therefore,

$$\nu^{10}=\frac{\mathcal{E}^1}{\Theta^{10}}-\epsilon^1 \tag{3.37}$$

and

$$\nu^{01}=\frac{\mathcal{E}^1}{\Theta^{01}}-\epsilon^2 \tag{3.38}$$

. Taking the derivative with respect to Θ^{10} and Θ^{01} and setting them to 0, we obtain because of concavity the optimal burstiness of signaling for each of the users:

$$\Theta^{10}=\frac{\mathcal{E}^1\mathrm{W}(\frac{\epsilon^1-\sigma^2}{\sigma^2 e})}{(\epsilon^1-\sigma^2)(\mathrm{W}(\frac{\epsilon^1-\sigma^2}{\sigma^2 e})+1)}\text{ and }\Theta^{01}=\frac{\mathcal{E}^2\mathrm{W}(\frac{\epsilon^2-\sigma^2}{\sigma^2 e})}{(\epsilon^2-\sigma^2)(\mathrm{W}(\frac{\epsilon^2-\sigma^2}{\sigma^2 e})+1)}. \tag{3.39}$$

As before, we have denoted by W the Lambert$W$ function which is defined as:

$$\mathrm{Lambert}W(x)\times e^{\mathrm{Lambert}w(x)}=x. \tag{3.40}$$

But, from the last constraint, we need to satisfy $\Theta^{01} + \Theta^{10} \leq 1$. Recognizing that,

$$\lim_{x \to -e^{-1}} \mathrm{W}(x) = -1, \tag{3.41}$$

$$\lim_{x \to 0} \mathrm{W}(x - e^{-1}) = -1, \text{and} \tag{3.42}$$

$$\lim_{x \to 0} \frac{\mathrm{W}(x - e^{-1}) + 1}{\sqrt{x}} = \sqrt{2e}, \tag{3.43}$$

we obtain the condition for burstiness in the low SNR regime, i.e. as $\epsilon^i \to 0$, or $\Theta^i \to 0$

$$\frac{\mathcal{E}^1}{\sqrt{2\epsilon^1}} + \frac{\mathcal{E}^2}{\sqrt{2\epsilon^2}} < 1. \tag{3.44}$$

This result can be easily generalized to the M-user case. In the low SNR regime, burstiness becomes capacity achieving when

$$\sum_{i=1}^{M} \frac{\mathcal{E}^i}{\sqrt{2\epsilon^i}} < 1. \tag{3.45}$$

Remarks:

- If (3.45) is verified, then the capacity region is rectangular. Consider a time slot from 0 to 1. Then, if user 1 only needs 25% of the time slots and user 2 needs 45%, and as long as the two percentages sum to less than 1, the two users do not interfere together, as each one of them will occupy different time slots.

- For high SNR, the capacity region without processing is almost triangular, and TDMA asymptotically follows the 45 degree line, covering almost the whole capacity region. So, even with no processing energy, TDMA can be used with almost no loss in capacity. In the presence of processing energy, TDMA will be favored at the maximum sum rate, i.e. for the whole 45 degree line which constitutes most of the capacity region. So taking TDMA to be the capacity region under processing energy is reasonable.

3.4 Graphical Illustration

We illustrate the effect of processing energy on the system where all users are only allowed to transmit at all times, under the constraint of additional processing energy, i.e. the Cover-Wyner region obtained by simply subtracting the processing energy from the total energy available. This is the region obtained with maximum rates

$$\log(1 + \mathcal{E}^1 - \epsilon^1), \tag{3.46}$$

$$\log(1 + \mathcal{E}^2 - \epsilon^2), \tag{3.47}$$

and sum rate

$$\log(1 + \mathcal{E}^1 + \mathcal{E}^2 - \epsilon^1 - \epsilon^2). \tag{3.48}$$

For convenience, we will refer to the region formed by these rates as the 'Traditional Cover-Wyner region under processing energy constraint'. We use the word 'Traditional' to refer to the case in which burstiness of signaling is not allowed, i.e. when continuous (or non-bursty) transmission is used. We will illustrate in this section how the use of bursty signaling improves the achievable rate region.

The TDMA curve is obtained by using

$$(R_1, R_2) = (\alpha \log(1 + \frac{\mathcal{E}^1}{\alpha} - \epsilon^1), (1-\alpha)\log(1 + \frac{\mathcal{E}^2}{1-\alpha} - \epsilon^2)) \tag{3.49}$$

and varying α between 0 and 1.

We refer to the region obtained out of these rates by varying α, as the 'TDMA under processing energy constraint'.

As argued before, when we compare the Traditional Cover-Wyner region under processing energy constraint and the TDMA under processing energy constraint, we notice that the maximum sum rate is achieved by TDMA in the presence of processing energy. This fact is illustrated in figure (3-6).

In the high-SNR case, the Cover-Wyner region is almost triangular, the dominant face being actually the dominant part of the region; The TDMA technique will result

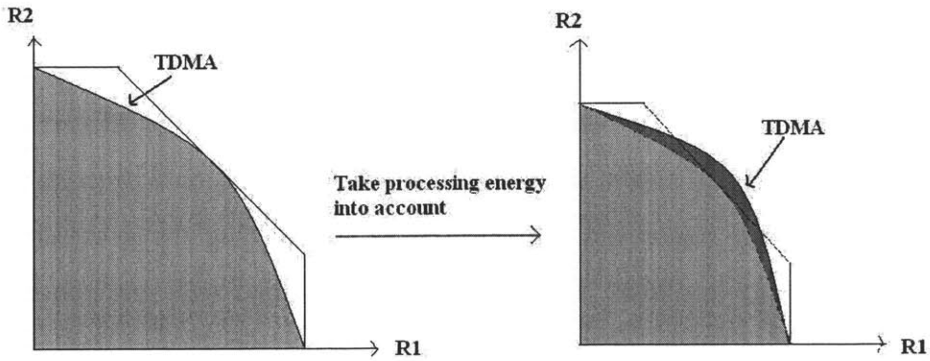

Figure 3-6: RHS: Processing energy is neglected. Cover-Wyner region and rates achievable by TDMA. LHS: Processing energy is taken into account. Traditional Cover-Wyner under processing energy constraint and rates achieved by TDMA under processing energy.

in rates that are asymptotic to the dominant face and are almost overlapping with it, therefore, the effect of processing energy is clearly going to make the TDMA technique achieve a higher sum rate. This is illustrated in figures (3-7) and (3-8).

After including the processing power, the Traditional Cover-Wyner region under processing energy constraint has lower boundary rates than the Cover-Wyner region obtained when processing energy is not taken inot account. This is because, when no processing energy is considered, more energy will be allocated for transmission and consequently higher rates will be obtained. Before the processing is taken into account, TDMA touches the dominant face at one point but is otherwise lower than the region. With processing, TDMA gives higher sum rate than the Traditional Cover-Wyner under processing energy constraint. This illustrates that the Cover-Wyner region is no more optimal. In fact, we have proved in section 2.2.2 that the optimal region is obtained by using bursty signaling. However, TDMA under processing energy constraint will still achieve the maximum sum rate. The impact of processing energy in the low-SNR case is illustrated in figure (3-9).

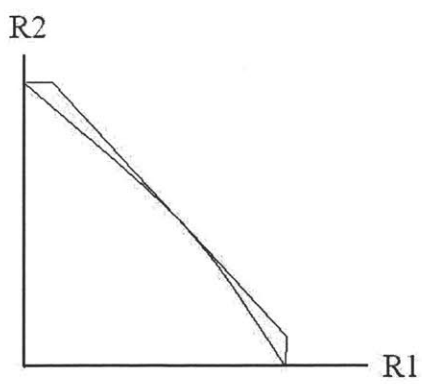

Figure 3-7: Cover-Wyner region and TDMA rate curve at high SNR: At high-SNR, when the processing energy is neglected, the rates achieved by TDMA almost overlap with the dominant face of the Cover-Wyner region.

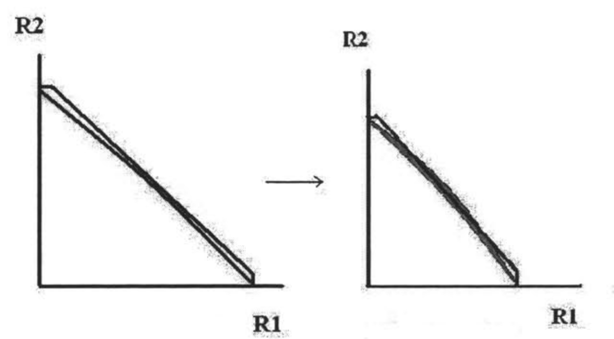

Figure 3-8: Effect of processing energy in the high SNR case

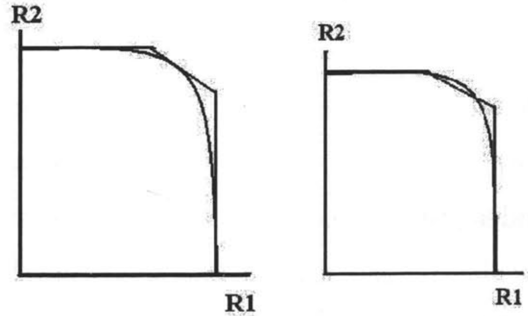

Figure 3-9: Effect of processing energy in the low SNR case

3.5 Capacity Region

In this section, we aim to find the shape of the whole capacity region in the presence of processing energy. For this purpose, we fix R_2 and aim to maximize R_1. We know the maximum value that R_2 can take (say R_{2max}), which is achieved when user 2 transmits alone. Let the fixed value of R_2 be x, where $x < R_{2max}$. We approximate the rates R_1 and R_2 by

$$R_1 = \Theta^{10}\log(1+\nu^{10}(1)) + \beta\Theta^{11}\log(1+\nu^{11}(1)+\nu^{11}(2)) \quad (3.50)$$

$$R_2 = \Theta^{01}\log(1+\nu^{01}(2)) + (1-\beta)\Theta^{11}\log(1+\nu^{11}(1)+\nu^{11}(2)) = x. \quad (3.51)$$

Now, using (3.51), we can express R_1 in terms of x as follows

$$R_1 = \Theta^{10}\log(1+\nu^{10}(1)) + \frac{\beta}{1-\beta}(x - \Theta^{01}\log(1+\nu^{01}(2))). \quad (3.52)$$

R_1 should be maximized subject to $\Theta^{01} + \Theta^{10} + \Theta^{11} < 1$ and $\Theta^{10}(\nu^{10}(1) + \epsilon_1) + \Theta^{11}(\nu^{11}(1) + \epsilon_1) \leq \mathcal{E}^1$. Using Lagrange multipliers, this maximization problem can be solved and a closed form solution could be obtained for the maximum R_1 in terms of R_2. The equations, while compact, do not present closed-form evaluations, so we have resorted to numerical methods.

Let us consider an example where $\mathcal{E}^1 = 0.8$, $\mathcal{E}^2 = 0.7$ and $\epsilon^1 = 0.3$ and $\epsilon^2 = 0.2$.

Numerically, we solve for the maximum rates that can be obtained while allowing the Θs to go from 0 to 1. In all figures, we plot in blue the different couples (R_1, R_2); the envelope of the blue area constitutes the maximum achievable rate area when burstiness is allowed. We compare this region to the Traditional Cover-Wyner region under processing energy constraint. We also plot the maximum achievable rates obtained by TDMA under processing energy constraint i.e. as described in 3.4.

Figure 3-14 shows the achievable rate region, given by the envelope of the region drawn by the points in blue. The TDMA under processing energy constraint is shown in red. The Cover-Wyner region when no burstiness is allowed is shown is black. This region is obtained by simply subtracting the processing power from each of the rates, as in section 3.4. As shown, the new rate region gives great improvements over the Cover-Wyner region with no burstiness allowed. The red curve shows that TDMA still allows achieving the maximum sum rate, obtained by the new region.

The improvement obtained is intuitive because when the users are given an extra degree of freedom by allowing them to transmit at a certain fraction of time, instead of continuously, they can transmit in a way to save the most of the power: by backing away, even when information is available for transmission, users can save part of the processing energy incurred whenever they are 'on'. We have proven why TDMA allows achieving the maximum sum rate and why it outperforms in this case other multiple access techniques.

For more insight, we show several figures that allow to compare the other methods. Figure (3-10) for instance, shows the TDMA with and without burstiness allowed. Other figures show the TDMA, the achievable regions with and without burstiness allowed, all under a processing energy constraint.

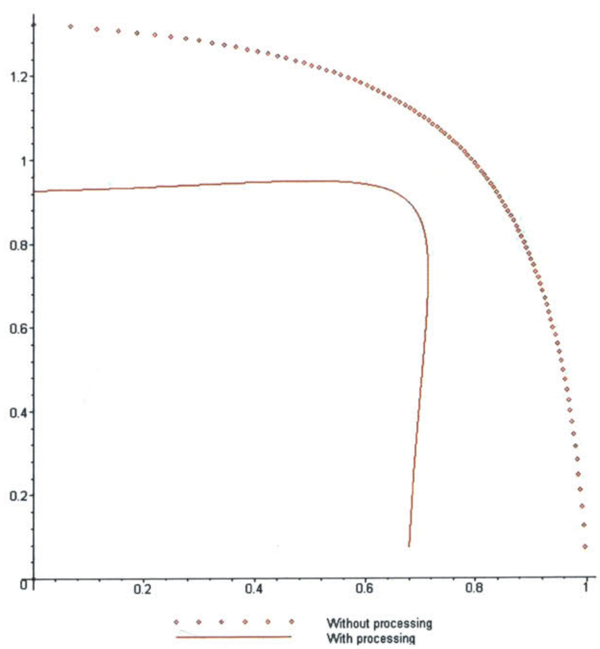

Figure 3-10: Boundary of the region of rates achievable with TDMA, with and without processing energy

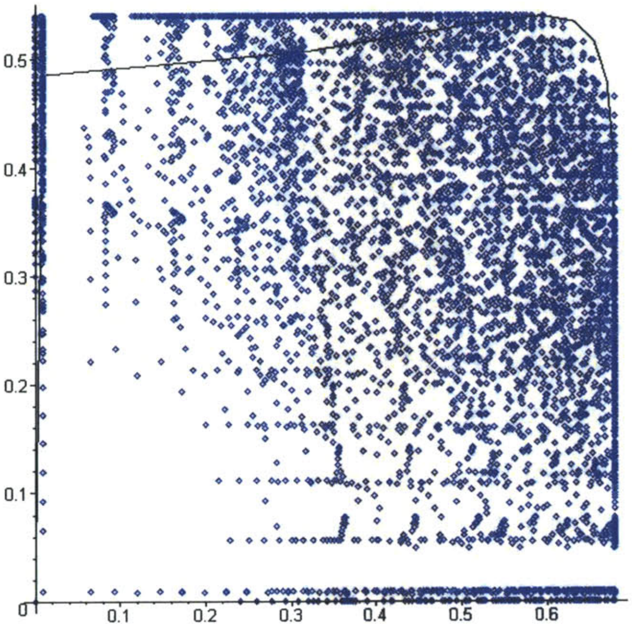

Figure 3-11: Achievable rates by bursty signaling: The black line shows the rate achievable by TDMA under processing energy. We can see how the maximum sum rate of the region in blue is achieved by TDMA.

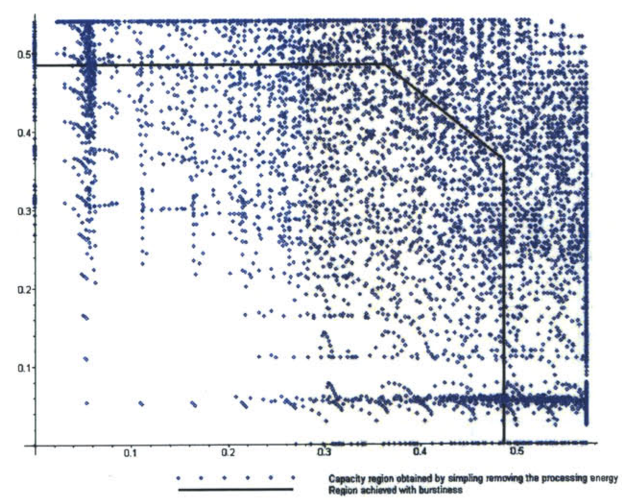

Figure 3-12: Bursty vs. non- bursty signaling: This figure shows the improvement obtained by the region in blue (when burstiness is allowed) as compared to the Traditional Cover-Wyner under processing energy.

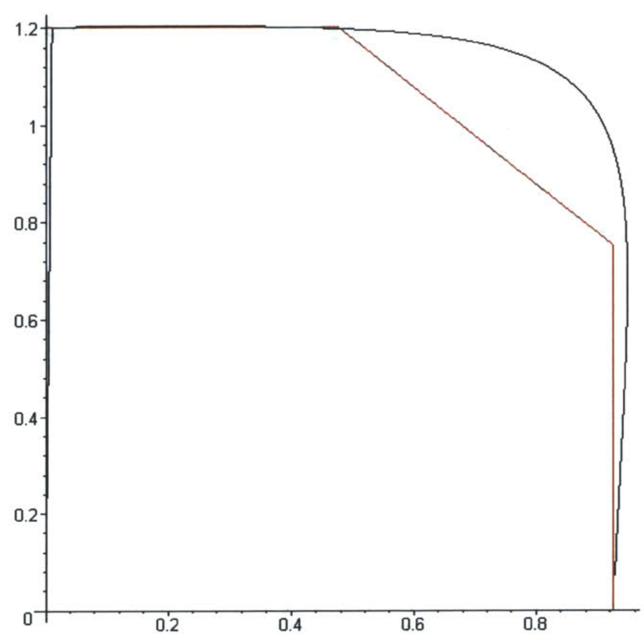

Figure 3-13: Traditional Cover-Wyner region and rates achieved by TDMA, both under processing energy constraint.

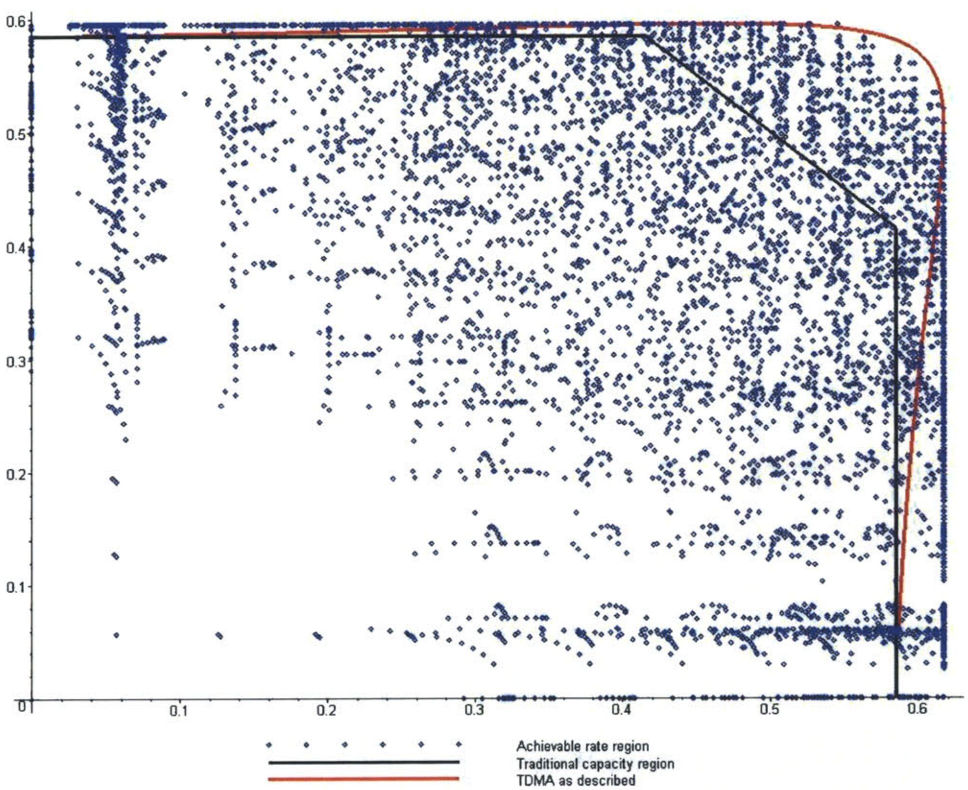

Figure 3-14: Summary of results: Traditional Cover-Wyner under processing energy, rate achieved by TDMA and the achievable rate region when burstiness is allowed.

Chapter 4

Correspondence Between Channel Variability and Processing Energy

4.1 Background on Non-Coherent Channels

This section refers to some of the results in [3], [4] and [5]. We consider wideband wireless communication over a channel where the transmitter has no channel state information (CSI). The coherent model assumes perfect channel knowledge; and the non-coherent i.i.d. fading model assumes not only no CSI at the receiver, but also that the channel changes so rapidly that there is no hope at all to obtain even partial channel knowledge based on the previously received signal.

From a signaling point-of-view, at low SNR, the estimation of the fading coefficients becomes difficult. It is therefore desirable to use peaky signaling, that is, to use only a small fraction of the available degrees of freedom, and avoid estimating the other fading coefficients. In contrast, if the fading coefficients are known, it is desirable to spread the transmitted energy over all the available time-frequency slots, in order to reduce the SNR per degree of freedom and obtain the optimal energy efficiency. Thus, the choice of the optimal peakiness of the input signals, which we denote by δ, is a key difference between the coherent and the non-coherent cases.

Then, as in [3],

$$C(\mathsf{SNR}) = \mathsf{SNR} - \Delta(\mathsf{SNR}). \tag{4.1}$$

Intuitively, in the case that partial channel information is available, the channel estimation becomes easier; when the channel changes over time slowly, a channel estimate can be used for a longer time. The effect of channel variation on the relation between capacity and SNR has been shown in [3] to be captured by a coefficient, β, which is termed coherence level.

The coherent channel with perfect CSI at the receiver corresponds to the case that $\beta = 1$, and the i.i.d. non-coherent channel corresponds to the case that $\beta = 0$.

We consider block fading channels with an intermediate coherence level $\beta \in [0, 1]$. The rate of channel time variation is characterized by the coherence time, denoted by T. In a block fading model, the channel fading coefficients are assumed to remain constant within a block of T symbol periods.

The parameter T is therefore used to indicate how slow the channel changes over time.

It has been shown in [3], that for the special case of $\beta = 1$,

$$T = \mathsf{SNR}^{-2}. \tag{4.2}$$

It has also been shown that for any $\beta \in [0, 1]$,

$$T = \mathsf{SNR}^{-2\beta}. \tag{4.3}$$

and

$$\delta = \mathsf{SNR}^{1-\beta}. \tag{4.4}$$

As β decreases, transmission must become increasingly peaky to be capacity achieving.

4.2 Correspondence Between Processing Energy and Channel Variability

In 4.1, we have shown that processing energy and channel variability both require peaky distribution for achieving capacity.

In this section, we show that, in the low SNR regime, there is a simple correspondence between processing energy and channel variability.

When T is large, the lack of knowledge is small and the cost we pay for not knowing the channel is small. Inversely, when T is small, the channel changes more often and the cost to pay for this variability increases. Therefore, the changing rate of the channel can be quantified by $1/T$. We can picture the lack of knowledge of the channel as an extra cost to pay and, most importantly, represent this cost as an energy penalty.

Let us consider side-by-side two channels, channel 1 and channel 2.

The first is characterized by a processing energy ϵ. The second is characterized by a level of coherence β and a coherence time T. Both channels have the same SNR. Let us assume that

$$\frac{1}{\epsilon} = T. \tag{4.5}$$

For channel 1, when $\Theta_{opt} = 1$, we have from (2.30)

$$\mathsf{SNR} = \mathcal{E} = \sqrt{2\epsilon}. \tag{4.6}$$

If we neglect the $\sqrt{2}$ factor, we obtain

$$\mathsf{SNR} \approx \sqrt{\epsilon}. \tag{4.7}$$

For channel 2, when $\beta = 1$,(i.e. $\delta = 1$), from (4.3),

$$T = \mathsf{SNR}^{-2}, \tag{4.8}$$

or

$$\mathsf{SNR} = \frac{1}{\sqrt{T}}. \tag{4.9}$$

Consequently,

$$\mathsf{SNR} \approx \sqrt{\epsilon}, \tag{4.10}$$

and

$$\mathsf{SNR} = \frac{1}{\sqrt{T}}. \tag{4.11}$$

which, by our assumption of similar SNR in both channels, only holds when (4.5) holds.

On the other hand, from (2.30), again neglecting the $\sqrt{2}$ factor, we have for channel 1

$$\Theta_{opt} = \frac{\mathsf{SNR}}{\sqrt{2\epsilon}} \approx \frac{\mathsf{SNR}}{\sqrt{\epsilon}}. \tag{4.12}$$

But, for channel 2, using (4.5) and (4.3) when $\beta \in [0, 1]$,

$$\sqrt{\epsilon} = \sqrt{\frac{1}{T}} = \mathsf{SNR}^{\beta}. \tag{4.13}$$

Therefore, substituting (4.13) in (4.12), we obtain the burstiness of signaling of channel 1 as a function of the level of coherence for channel 2:

$$\Theta_{opt} = \mathsf{SNR}^{1-\beta}. \tag{4.14}$$

Finally, for channel 2, recall that from (4.3) the optimal peakiness

$$\delta = \mathsf{SNR}^{1-\beta}. \tag{4.15}$$

Hence,

$$\Theta_{opt} = \mathsf{SNR}^{1-\beta} = \delta = \mathsf{SNR}^{1-\beta}. \tag{4.16}$$

So the burstiness of signaling as defined for channels with additional processing energy is the same as the peakiness of signaling as defined for channels with a certain variability or a coherence level.

On the other hand, the lack of knowledge of the channel can be interpreted as an energy penalty.

Chapter 5

Conclusions and Future Work

We study the optimal strategy in using the total energy when the processing energy of the transmitter being 'on' is also taken into account. We derive the capacity of an AWGN channel in terms of the total energy constraint $\mathcal{E}$ and the processing energy ϵ. We find the optimal value of transmission energy (ν_{opt}) for which capacity is maximized. We distinguish two modes for transmission: constant transmission and bursty transmission. We calculate the optimal burstiness of signaling(Θ_{opt}) at which channels should be transmitting. We describe the conditions that render bursty transmission capacity achieving in terms of the total energy available to the input and the processing energy.

In low SNR regime, we prove that achieving capacity requires burstiness when the processing energy is greater than half the square of the total energy. In particular, if the processing energy is any percentage of the total available energy, bursty transmission achieves capacity.

At high SNR, we show that the slope of the curve $(\mathcal{E}, \epsilon)$, which tends to 1 at infinity defines a clear-cut between the continuous transmission scheme and the bursty one.

In this analysis, we only consider the mutual information conditioned on the indicator of when we actually transmit although it is in fact possible to use the indicator itself to convey information; this can be done by having a random place to turn on and off and using the detection of this to carry information. The results we obtain give at least an achievable capacity region which readily gives great improvements over the

Cover-Wyner region with no burstiness allowed.

Next, we extend the results to the AWGN multiple access channel with M senders and a single receiver. We first show that, under a processing energy constraint, TDMA outperforms other multiple access techniques for the purpose of maximizing the sum rate. We prove that, for the same purpose, burstiness is capacity achieving in the low SNR regime when the sum of the ratios of total energy to processing energy is less than unity. Finally, we present numerical results under general conditions and illustrate the shape of the achievable capacity region under processing energy. We illustrate the improvement in rate, obatined by using burstiness, over the rate obatined by the traditional Cover-Wyner region with no burstiness allowed.

We show that, in low SNR regime, a channel with a certain time variability as well as a channel with processing energy lead to the same intuition regarding burstiness of signaling, since the variability of a channel can simply be treated as an energy penalty, or a processing cost, as in our study. We show that the same expression of burstiness as a function of the SNR is obtained for both channels.

This direct correspondence between the processing cost and the variability of the channels allows to interpret the lack of perfect coherence as an energy penalty.

This opens a new horizon in which one can think about the energy as a currency, that one needs to pay whenever there is a lack of knowledge of some aspects of the channel. It would be interesting to investigate what are the factors that can be solved for by paying extra energy. Future work could include then a generalization of the concept of energy as a unifying cost for various communication impediments.

Appendix A

LambertW Function

The Lambert$W(x)$ function [26], also called the Omega function, is the inverse function of $f(w) = w \times e^w$ for complex numbers w; This means that for every complex number x, we have

$$\text{Lambert}W(x) \times e^{\text{Lambert}W(x)} = x. \tag{A.1}$$

Since the function f is not injective in $(-\infty, 0)$, the Lambert$W$ is multivalued in $[-\frac{1}{e}, 0)$.
For $x \geq -e^{-1}$, the funtion $W(x)$ is real and single valued.
For example, $W(0) = 0$ and $W(-e^{-1}) = -1$.

Furthermore, the plot of $W(x)$ as a function of x, for $x \geq -e^{-1}$, is shown in figure A-1.

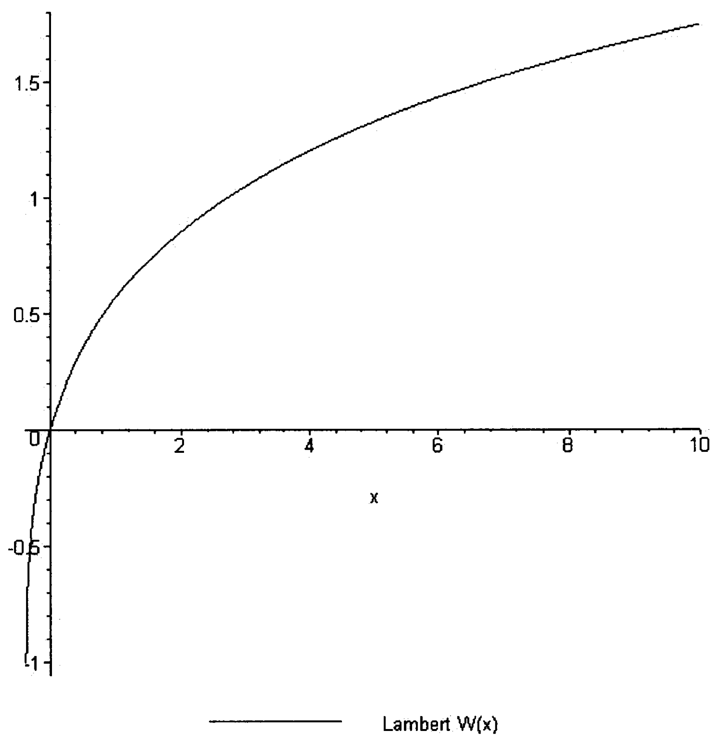

Figure A-1: W(x) as a function of x

Bibliography

[1] H. Liao, "Multiple access channels," PhD Thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972

[2] S. V. Hanly, "Information capacity of radio networks," Ph.D. dissertation, University of Cambridge, 1993

[3] L. Zheng, D. N.C. Tse, M. Médard,"Channel coherence in the low SNR regime," in preparation

[4] L. Zheng, D. N.C. Tse, M. Médard,"Channel coherence in the low SNR regime," invited paper, *38th Annual Conference on Information Sciences and Systems*, Princeton, 2004.

[5] L. Zheng, M. Médard, D.N.C. Tse, C. Luo, "On channel coherence in the low SNR regime," *41st Allerton Annual Conference on Communication, Control and Computing*, 2003.

[6] S. V. Hanly, "Information capacity of radio networks," Ph.D. dissertation, University of Cambridge, 1993.

[7] G. Caire, D. Tuninetti, and S. Verdú, "Suboptimality of TDMA in the low-power regime," *IEEE Trans. on Info. Theory*, vol. 50, no. 4, 2004.

[8] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. on Info. Theory*, Vol. 42, No. 2, 1996.

[9] T. M. Cover and J. A. Thomas, *Elements of information theory*, New York, Wiley, 1991.

[10] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. on Info. Theory*, Vol 36, No 5, 1990.

[11] R. Kravertz, P. Krishman, "Power management techniques for mobile communication," *Proceedings of The Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking*, 1998.

[12] R. Ahlswede, "Multi-way communication channels", at *Proc 2nd Int. Symp. Information Theory*, Armenian S.S.R.:23-52, 1971.

[13] A. El Gamal, T.M. Cover,"Multiple user information theory," *Proceedings of the IEEE* 68:12:1466-1483, 1980.

[14] A. D. Wyner, " Recent results in the Shannon theory," *IEEE Transactions on. Information Theory*, vol. IT-20., 1974.

[15] Csiszti and J. Kilner, "Information Theory: coding theorems for discrete memoryless systems," New York: Academic Press, 1981.

[16] R. E. Blahut, "Principle and practice of information theory," Reading, MA: Addison-Wesley, 1987.

[17] T. M. Cover, A. Balakrishnan and A. Viterbi, "Some advances in broadcast channels," in *Advances in Communications Systems Theory and Applications* , Eds. New York: Academic-Press, 1975.

[18] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124-142, 1985.

[19] P. J.M. Havinga, G. J.M. Smit, "Minimizing energy consumption for wireless computers in Moby Dick,"*Proceedings IEEE International Conference on Personal Wireless Communication*, 1997.

[20] M. Weiser, B. Welch, A. Demers, and S. Shenker, "Scheduling for reduced cpu energy," *Proceedings of the First Symposium on Operating System Design and Implementation* , 1994.

[21] F. Douglis, P. Krishnan, B. Marsh,"Thwarting the power-Hungry disk," *Proc. of Winter 1994 USENIX Conference*, 1994.

[22] K. Li, R. Kumpf, P. Horton, T. Anderson,"A quantitative analysis of disk drive power management in portable computers," *Proc. of Winter 1994 USENIX Conference*,1994

[23] S. Cui, A.J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization for coded systems," In *Globecom*, 2003.

[24] A.El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," In *Proc. IEEE Infocom*, 2002.

[25] S. Cui, A.J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO Techniques in sensor networks,"*IEEE journal on selected areas on communications*, 2004.

[26] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth, "On the Lambert W function," In *Advances in Computational Mathematics*, vol.5, pp. 329–359, 1996.

[27] A.Chandrakasan, S. Sheng, R. Brodersen "Low power digital CMOS design," In *Kluwer Academic Publishers*, Norwell, MA, 1995.

[28] P. Agrawal, J-C. Chen, S. Kishore, P. Ramanathan, and K. Sivalingam,"Battery power sensitive video processing in wireless networks," In *Proc. IEEE PIMRC*, 1998.

[29] J.M. Rulnick and N. Bambos, "Mobile power management for maximum battery life in wireless communication networks," In *Proc. of IEEE Infocom 96*, 1996.

[30] A. Sampath, P.S. Kumar, and J. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," In *The Sixth International Symposium on Personal, Indoor and Mobile Radio Communications*, 1995.

[31] M. Stemm and R. Katz,"Reducing power consumption of network interfaces in hand-held devices," In *Third International Workshop on Mobile Multimedia Communications*, 1996.

[32] A. Chandrakasan, R. Min, M. Bhardwaj, S-H Cho, and A. Wang, "Power aware wireless microsensor systems," *Keynote Paper ESSCIRC*, 2002.