Bursty Transmission and Glue Pouring: On Wireless Channels with Overhead Costs

Pamela Youssef-Massaad, Lizhong Zheng, and Muriel Médard

***Abstract*—Power efficiency is a capital issue in the study of mobile wireless nodes owing to constraints on their battery size and weight. In practice, especially for low-power nodes, it is often the case that the power consumed for non-transmission processes is not always negligible. In this paper, we consider the channels with a special form of overhead: a processing energy cost whenever a non-zero signal is transmitted. We show that under certain conditions, achieving the capacity of such channels requires intermittent, or 'bursty', transmissions. Thus, an optimal sleeping schedule can be specified for wireless nodes to achieve the optimal power efficiency. We show that in the low SNR regime, there is a simple relation between the optimal burstiness and the overhead cost: one should use a fraction of the available degrees of freedom at an SNR level of $\sqrt{2\epsilon}$, where ϵ is the normalized overhead energy cost. We extend this result to use bursty Gaussian transmissions in multiple parallel channels with different noise levels. Our result can be intuitively interpreted as a "glue pouring" process, generalizing the well-known water pouring solution. We then use this approach to compute the achievable rate region of the multiple access channel with overhead cost.**

***Index Terms*—Low power algorithms and protocols, multimedia, networks and systems, UWB, transmission technology, medium access control, multimedia, networks and systems.**

I. INTRODUCTION

WHEN minimizing the total energy, it is fundamental to consider, besides the energy spent on transmission purposes, non-transmission energies. In this paper, we refer to such non-transmission energy costs as "processing energy", or simply "overhead" of transmissions. It is intuitive that a large overhead cost might change the nature of the optimal signaling.

In the computer and sensor nodes literature, various techniques have been proposed to reduce the mobile host's power consumption during operation. Recognizing the fact 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 handheld computers and up to 10% for high-end laptops," Kravertz *et al*. proposed in [7] software-level techniques to suspend the mobile host's device during idle periods of the communication. In [14], Chandrakasan *et al*. studied power-aware techniques to minimize power consumption of wireless microsensor systems. 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, [8]. In [11], Rulnick and Bambos studied mobile power management for maximum battery life in wireless communication networks. In [15], 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 [16], the authors analyzed the best modulation strategy to minimize the total energy consumption, while satisfying throughput and delay requirements. Various other interesting power-aware components and algorithms for wireless networks can be found in ([10], [9], [12], [13]).

The information theoretic result by Verdú [6] on capacity per unit cost can be applied to study energy limited systems with the signal processing energy taken into consideration. Here, one can view the cost of transmitting a symbol as the sum of transmission energy and the corresponding non-transmission energy. Verdú's result states that the most cost efficient way of signaling is peaky binary signaling with position detections.

In this paper, we consider using Gaussian channels with a special form of processing energy cost, the energy of being 'on'. That is, we model the energy cost of operating the device as a constant, ϵ, added to the transmission power, whenever a non-zero signal is being transmitted. This simplified model is a reasonable approximation to reality in many applications, and allows insightful results to be derived. Here, one can directly apply Verdú's result, simply letting the cost of sending symbol x be $|x|^2 + \epsilon \cdot 1_{\{x\neq 0\}}$. It is intuitively clear that with a significant processing energy cost, the optimal signaling should only transmit in a small fraction, Θ, of the time to save energy, i.e., to peaky signaling instead of i.i.d. Gaussian input as for the Gaussian channel with no processing energy cost.

We derive the optimal burstiness in terms of the LambertW function. For the low SNR case, we show that there is a very simple relation between the optimal burstiness and the processing cost,

$$\Theta^* \approx \mathcal{E}/\sqrt{2\epsilon}$$

where $\mathcal{E}$ is the average energy constraint at the transmitter, normalized by the noise variance. We also show that it is near optimal to transmit and receive Gaussian signals in a prescribed Θ^* fraction of the time, and thus on-off signaling is not the unique way to achieve the optimal energy efficiency. We also observe that there is an interesting similarity of the optimal signaling as well as the achieved throughput between this problem and the non-coherent block fading channel studied in [5]. We thus interpret the lack of channel information in

Manuscript received February 21, 2008; revised May 30, 2008; accepted August 17, 2008. The associate editor coordinating the review of this paper and approving it for publication was X.-G. Xia. This research is supported by HP, Advanced Concepts in Wireless Networking, Research Alliance Agreement Award 008542-008 and NSF NRT Award ANI-0335217.

The authors are with the Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technolgy, Cambridge, MA (e-mail: {pmassaad, medard, lizhong}@mit.edu).

Digital Object Identifier 10.1109/T-WC.2008.070939

the non-coherent channel as equivalently an energy overhead.

We generalize these results to the case when multiple parallel Gaussian channels, with different noise levels, are available. We derive the optimal power allocation in this case. Our result can be intuitively thought as a "glue pouring" process, in contrast to the well-known water pouring solution. The conventional water pouring can be viewed as a procedure to successively allocate the total signal energy to parallel channels. During this process, each increment of signal energy is spent in the sub-channel with the lowest sum of the noise power and the signaling power already allocated. Thus a new degree of freedom starts to be used only if its corresponding noise level is lower than the signal-plus-noise level of the sub-channels that are already used. When processing cost presents, this condition becomes more stringent: a new degree of freedom is used only if the signal plus noise level in the channels already used is strictly larger than $(1+\nu)$ times the noise level, where $\nu > 0$ is a constant that we will explicitly compute. Intuitively, the presence of overhead costs makes it less profitable to start using a new channel. We further extend our analysis to the multiple access channel with overhead, where the optimal achievable rate region by bursty Gaussian transmissions is found.

The remainder of the paper is organized as follows. In Section II, we consider the case of a single user when the energy constraint of an AWGN channel to include the processing energy. We derive the optimal power allocation and burstiness of signaling. We also study the power allocation in parallel Gaussian channels, and generalize the well-known water-pouring solution to the case with processing overhead. In Section III, we apply these results to find the optimal rate region for 2-user Gaussian multiple access channels with overhead costs. In particular, we show that carefully scheduled turning off users gives a gain in the achievable rate region.

II. Single User Capacity with Processing Energy Cost

We consider a single user additive white Gaussian noise channel. Let i be the time index. The noise samples are independent and identically distributed Gaussian random variables with zero-mean and variance σ^2: $Z_i \sim \mathcal{CN}(0, \sigma^2)$. The output of the channel at time i is given by

$$Y_i = X_i + Z_i. \tag{1}$$

The capacity of this channel is

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

where $\mathcal{E}$ denotes the average transmitted signal energy constraint, i.e.,

$$\frac{\sum_{i=1}^n |X_i|^2}{n} \leq \mathcal{E}, \tag{3}$$

In what follows, we will set the noise variance to be 1, and still used $\mathcal{E}$ as the normalized energy constraint. We model the processing cost as a constant amount whenever there is non-zero signal being transmitted. For convenience, we also normalize the processing energy cost by the noise variance, and write it as ϵ. Now the energy constraint at the transmitter (3) is replaced by

$$\frac{1}{n}\sum_{i=1}^n \left[|X_i|^2 + \epsilon \cdot 1_{\{X_i \neq 0\}}\right] \leq \mathcal{E},$$

where $1_{\{\cdot\}}$ is the indicator function.

This modeling of the processing energy cost is clearly a simplification from the reality. First, we ignore the energy cost of transition between the 'on' state and the 'off' state of the device, which can sometime be significant; secondly, we assume that whenever the transmitter is sending $X = 0$, it not only appears silent, but also can turn off the support circuits immediately to save energy. In cases the signaling requires frequent and fast turning on and off, such as with pulse position modulation, our assumption would fall apart. However, we will show later that it is nearly optimal to use a signaling scheme that remains in the on and off stages for long periods of time, (while keeping the ratio between the "on time" and the "off time"). Thus, the cost of transition between the on and off states is amortized over time, making our simplified model a good approximation to reality.

The capacity of this channel can be computed, numerically in most cases. The case of interests is the one with very limited total energy per channel use, or equivalently, the case with abundant bandwidth. For this limiting case, one can directly apply Verdu's result on capacity per unit cost [6], where the cost of sending a symbol x is simply $C(x) = |x|^2 + \epsilon \cdot 1_{\{x \neq 0\}}$. Note that the symbol $x = 0$ is a zero-cost symbol to the channel, and thus the capacity per unit energy is given by the optimization

$$x^* = \arg\max_x \frac{D(N(x,1)||N(0,1))}{C(x)}, \tag{4}$$

where $N(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. This data rate can be achieved by using on-off signaling where the optimal value x^* is transmitted as the "on" signal, and the position of which is used to carry information.

It is easy to verify that the the optimal signaling is the pulse position modulation (PPM), with an arbitrarily high peak. The resulting capacity is linear in the total energy, and the same as if the processing energy cost is not present. Intuitively, bursty transmission reduces the processing energy cost by reducing the number of degrees of freedom used in sending non-zero signals. At the low energy limit, the signaling is so peaky that the signal power, whenever transmitted, is much higher than the processing cost.

This simple result, however, has some limitations when applied in reality. First, in order to detect the PPM signal, the receiver has to be on for all the time, resulting in highly unbalanced duty cycles at the transmitter and the receiver; secondly, it is often the case that the peak transmission power is limited by the hardware, thus it is usually unrealistic to transmit at high enough energy such that the processing cost becomes ignorable.

To address the above issue, we study a specific signaling that is sub-optimal in the sense of [6]. Here, we let the transmitter and the receiver be turned "on" for a prescribed Θ fraction of the time, and transmit Gaussian signal in this period of time, with a fixed average signal power, ν, satisfying the total energy constraint. We chose the parameter Θ and ν by solving the following optimization problem.

$$\max_{\Theta,\nu:\Theta(\nu+\epsilon)\leq\mathcal{E}} \quad \Theta\log(1+\nu) \tag{5}$$

Remarks:

- In capacity computation, we consider coding over arbitrarily long blocks. Here, even though the time that the transmitter is on is a small fraction Θ of the overall coding time, we still assume it is indeed a long period of time (possibly with multiple segments), over which one can code to achieve the Gaussian capacity $\log(1+\nu)$;
- Using concavity, it is clear that when the transmitter is on, one should always use a same average signal power ν, instead of varying ν from time to time;
- When Gaussian signaling is transmitted, strictly speaking, it is possible that a 0 symbol is drawn from the Gaussian ensemble, at which time it is potentially possible to turn off the transmitter to save energy. We ignore this possibility;
- One main difference between the proposed bursty Gaussian signaling and PPM is that the fraction of time when the transmitter is on is predetermined. Thus no information is conveyed by the position of the transmitted signals. This allows the receiver to be turned on in the same fraction of time, also makes the signaling more suitable to multi-user applications;
- At the low energy limit, the capacity result (4) leads to a data rate that is linear in the total energy, the bursty Gaussian signaling yields a rate that is strictly sub-linear in signal energy. The optimization in (5) can be understood as a tradeoff: on one hand, reducing Θ and use bursty signaling avoids paying the processing energy cost, on the other hand, accumulating too much energy reduces the energy efficiency, as the $\log(\cdot)$ function becomes sub-linear as ν gets large.

The optimization problem (5) can be easily solved. We write $\Theta = \mathcal{E}/(\nu+\epsilon)$, and set

$$\frac{\partial}{\partial\nu}\left[\frac{\mathcal{E}}{\nu+\epsilon}\log(1+\nu)\right] = 0,$$

which implies

$$\frac{1}{1+\nu} = \frac{1}{\nu+\epsilon}\log(1+\nu). \tag{6}$$

Together with the constraint that $\Theta \leq 1$, we have that he optimal values

$$\Theta_{\mathsf{opt}} = \min\left\{1, \frac{\mathcal{E}\cdot W(e^{-1}(\epsilon-1))}{(\epsilon-1)(W(e^{-1}(\epsilon-1))+1)}\right\}; \tag{7}$$

and

$$\nu_{\mathsf{opt}} = \frac{\mathcal{E}}{\Theta_{\mathsf{opt}}} - \epsilon = \frac{\epsilon-1}{W((\epsilon-1)e^{-1})} - 1. \tag{8}$$

where we denoted by W the LambertW function, where $W(x)$ is the solution of $W\cdot e^W = x$. Reviews of the properties of LambertW function can be easily found on the web and in [17].

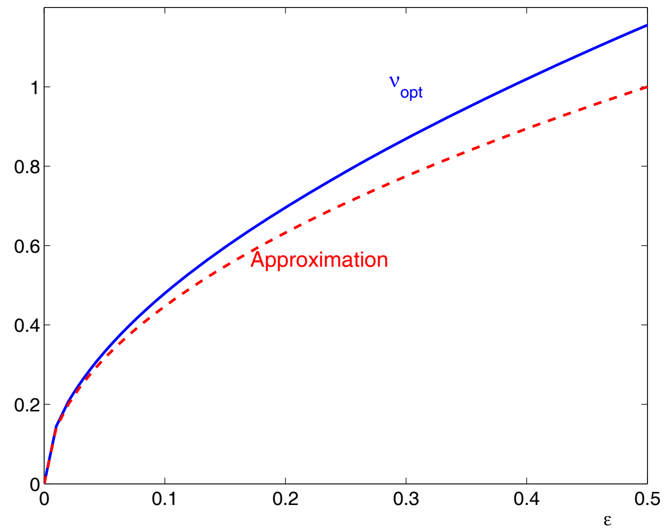

Fig. 1. Approximation to ν_{opt}.

It is interesting to observe that when $\Theta_{\mathsf{opt}} < 1$, i.e., when it is optimal to transmit bursty signals, the optimal signal power ν_{opt} does not depend on the total energy $\mathcal{E}$. That is, there is an optimal power level, depending only on the processing cost ϵ. When $\mathcal{E} > \nu_{\mathsf{opt}} + \epsilon$, the total energy is spread out over all degrees of freedom and the transmission is not peaky; when $\mathcal{E} < \nu_{\mathsf{opt}} + \epsilon$, it is optimal to transmit only Θ_{opt} fraction of the time, with $\Theta_{\mathsf{opt}} = \mathcal{E}/(\nu_{\mathsf{opt}}+\epsilon)$, to maintain the desired signal power per channel use.

We are particularly interested in the case where both $\mathcal{E}$ and ϵ are small. In this case, the interaction between the total energy and the processing cost can be characterized by the following approximation of (8), for small values of ϵ. It can be shown that

$$\lim_{t\to 0} W(e^{-1}(t-1)) - \sqrt{2t} = -1, \tag{9}$$

thus we have an approximation to (8)

$$\nu_{\mathsf{opt}} \approx \sqrt{2\epsilon}.$$

The approximation is more precise when ϵ is small, as shown in Figure II. Note that all notions of signal energy in this paper are normalized by the unit noise level, and are thus energy ratios; hence the expression does not cause any contradiction in the units.

A simple observation to be made is that when transmitting at the optimal value ν_{opt}, the ratio between the processing cost and the transmission cost, $\epsilon/\sqrt{2\epsilon}$, becomes small as $\epsilon \to 0$.

In summary, for the system of interest with small values of total energy $\mathcal{E}$ and small processing cost ϵ, suppose the input is chosen from the family of bursty Gaussian signaling described above, it is optimal to use strictly bursty signal, $\Theta < 1$, when $\mathcal{E} < \sqrt{2\epsilon}$. In such cases, the Θ should be chosen as

$$\Theta_{\mathsf{opt}} \approx \mathcal{E}/\sqrt{2\epsilon} \tag{10}$$

to maintain the desired transmission power.

In the same asymptotic regime, the achieved data rate by the bursty Gaussian signaling with the optimal choice of Θ_{opt}

can be computed using Taylor expansion as follows:

$$\begin{aligned} & R(\mathcal{E}, \epsilon) \\ = & \Theta_{\text{opt}} \log\left(1 + \frac{\mathcal{E}}{\Theta_{\text{opt}}} - \epsilon\right) \\ = & \frac{\mathcal{E}}{\sqrt{2\epsilon}} \log(1 + \sqrt{2\epsilon} - \epsilon) \\ = & \frac{\mathcal{E}}{\sqrt{2\epsilon}} \cdot \left[\sqrt{2\epsilon} - \epsilon - \frac{1}{2}\left(\sqrt{2\epsilon} - \epsilon\right)^2 + O\left(\epsilon^{\frac{3}{2}}\right)\right] \\ = & \mathcal{E} - \mathcal{E} \cdot O(\sqrt{\epsilon}). \end{aligned}$$

The key observation is that when $\mathcal{E}$ and ϵ are both small, the achieved data rate with the proposed signaling scheme is close to $\mathcal{E}$, which is the best possible energy efficiency, even when there is no processing cost. Thus we conclude that the bursty Gaussian signaling used in this asymptotic regime is close to optimal in the achieved data rate.

Discussion: Connections to Non-coherent Channel Capacity:

In [5], we studied the capacity of non-coherent block fading channel in the low SNR regime. It is observed that when the channel state information (CSI) is not available at the receiver, it is desirable to use bursty signaling. Intuitively, the burstiness of the transmission is chosen from the tradeoff between avoiding the cost of estimating the channel and keeping signal power low (thus energy efficiency high). The result there has some nice connections to the current paper. We summarize that result as follows.

Consider the block fading channel

$$\mathbf{y} = \sqrt{\mathsf{SNR}} \cdot \mathbf{h}\mathbf{x} + \mathbf{w},$$

where $\mathbf{x}, \mathbf{w}, \mathbf{y} \in \mathcal{C}^l$; $\mathbf{h} \sim \mathcal{CN}(0,1)$ is the scalar fading coefficient, which is assumed to remain constant through the block before changing into independent realization in the next block. The block length l is the *coherence time* of the channel. $\mathbf{w}$ is the additive noise with i.i.d. $\mathcal{CN}(0,1)$ entries. The power constraint is also normalized such that each entry of $\mathbf{x}$ has unit average power. We are interested in the low SNR regime with $\mathsf{SNR} \to 0$. The main result of [5] says that the optimal signaling for this channel is bursty Gaussian signaling, with joint channel estimation and decoding at the receiver; and the optimal burstiness depends on the relation between the SNR and the channel coherence time as follows.

Let the coherence time l satisfy for some $\alpha \in [0,1]$, $\lim_{\mathsf{SNR}\to 0} \frac{\log l}{\log \mathsf{SNR}} = -2\alpha$. The capacity of this channel can be written as $C_l(\mathsf{SNR}) = \mathsf{SNR} - \Delta(\mathsf{SNR})$, with $\Delta(\mathsf{SNR}) \approx \mathsf{SNR}^{1+\alpha}$. This capacity can be achieved by using Gaussian signal in $\delta(\mathsf{SNR}) \approx \mathsf{SNR}^{1-\alpha}$ fraction of the available time slots. Put it another way, the burstiness of the signaling scales as $\delta \approx \mathsf{SNR} \cdot \sqrt{l}$.

Compare this result with (10), we observe a clear correspondence between the two different channels.

$$\begin{array}{ccc} \mathcal{E} & \longleftrightarrow & \mathsf{SNR} \\ \epsilon & \longleftrightarrow & \frac{1}{l} \\ \Theta_{\text{opt}} \approx \mathcal{E}/\sqrt{2\epsilon} & \longleftrightarrow & \delta \approx \mathsf{SNR} \cdot \sqrt{l} \end{array}$$

Exchanging the parameters on the left from the problem of communication with overhead costs with the ones on the right from [5], we found that the optimal signaling and the achieved capacity of the two problems are very similar, except that in the current paper, the results are more precise. For example, the factor of 2 in the above expression cannot be captured by the asymptotic analysis used in [5]. The general approach in the current paper can be used to study non-asymptotic cases. On the other hand, in [5], it is shown that at the low SNR limit, bursty Gaussian is near optimal: not only is the low SNR capacity limit achieved but also is that limit approached at the fastest rate. This result can indeed be extend to the problem in the current paper. It can be shown that bursty Gaussian signaling is capacity achieving at the limit that both $\mathcal{E}$ and ϵ approach 0, with $\log \mathcal{E}/\log \epsilon$ remains constant. The analysis is however rather repetitive to [5], and is thus omitted.

The correspondence between the two papers gives more insights to both problems. Intuitively, for a non-coherent block fading channel, there is an *implicit* cost of obtaining channel knowledge, [1] which can be amortized over the block of l symbols corresponding to one channel realization. In effect, this cost is equivalent as an extra energy cost of $\epsilon \sim 1/l$ per symbol time. The optimal signal peakiness Θ_{opt} and δ are thus chosen by the same tradeoff between the efforts to avoid the overhead cost by transmitting over fewer channels, and to improve energy efficiency by spreading signal energy over more channels.

It is interesting to generalize the above results into the case with multiple parallel Gaussian channels. It is well known that water-pouring gives the optimal power allocation and thus the capacity over parallel Gaussian channels. The water-pouring solution in the case with processing energy is, however, slightly different.

It is instructional to understand the water-pouring solution as a process of differential power allocation. In brief, it is capacity optimal, for Gaussian channels, to break the total signal energy into, say K, small pieces, each used to carry a sub-message. At the receiver, each sub-message is decoded in turn, viewing the yet-not-decoded sub-messages as interference, and then subtracted from the received signals before decoding the next message [3]. In a single Gaussian channel, this process can be mathematically written as

$$\log\left(1 + \frac{P}{N}\right) = \sum_{j=1}^{K} \log\left(1 + \frac{\Delta P_j}{N + \sum_{i=1}^{j-1} \Delta P_i}\right)$$

with $\sum_{j=1}^{K} \Delta P_j = P$. In the case with parallel channels, this differential power allocation process always allocate increments of signal energy in the channel with the lowest interference level, and thus keep the signal plus noise level at all channels balanced.

In the presence of processing energy, the differential power allocation process is somewhat different. This can be best understood reexamining equation (6), which is rewritten below for convenience.

$$\frac{1}{(1 + \nu_{\text{opt}})} = \frac{1}{\epsilon + \nu_{\text{opt}}} \log(1 + \nu_{\text{opt}}).$$

[1]Note we emphasize that this cost of channel knowledge is implicit, since the capacity results in [5] do not assume any explicit channel estimation.

To interpret this equation, suppose there are many parallel Gaussian channels with noise variance $N = 1$, and suppose only a fraction of them are used. Let the signal to noise ratio in these used channels be ν. Now suppose we have an increment of signal power ΔP. We can use this power in either or the combination of the following two ways:

- We can spend ΔP in a channel that is already used. This does not incur an extra cost of processing energy, but the transmission has an interference level of $1 + \nu$, and the increase in data rate is
$$dR_1 = \log(1 + \nu + \Delta P) - \log(1 + \nu) \approx \frac{\Delta P}{1 + \nu}.$$
- Alternatively, we can spend ΔP in some unused channels. Suppose we use these channels at the same SNR level of ν. Counting the processing cost, we can use $\Delta P/(\nu + \epsilon)$ channels, and obtain a rate increase of
$$dR_2 = \frac{\Delta P}{\nu + \epsilon} \log(1 + \nu).$$

Equation (6) states that the optimal SNR to use a channel with processing energy cost is ν_{opt} that balances these two options, satisfying $dR_1 = dR_2$. Intuitively, if there is any channel used at an SNR level lower than ν_{opt}, then it is more profitable to put more signal power in this channel before using a new one, with the extra overhead. On the other hand, if there is any unused channel, there is no point to use any channel at an SNR level higher than ν_{opt}. For the case that the noise variance of all channels are N, (instead 1), (6) can be rewritten as
$$\frac{1}{N \cdot (1 + \nu_{\mathsf{opt}})} = \frac{1}{\epsilon + N \cdot \nu_{\mathsf{opt}}} \log(1 + \nu_{\mathsf{opt}}). \qquad (11)$$
Clearly, the optimal signal-to-noise ratio ν_{opt} depends on the relation between the overhead cost ϵ and the noise level N. Now the water-pouring problem is simply a generalization where channels have different noise levels.

The simplest case for water-pouring is for two parallel channels. However, to facilitate our discussion on the multiple access channels in the next section, we consider a slightly more general case. Let there be K parallel channels. It will be clear that the number K does not affect the result, and can be thought as the channels used over multiple time slots. Let $\alpha_1 K$ channels have noise level of N_1, and the other $\alpha_2 K$ of N_2, where $\alpha_1 + \alpha_2 = 1$, and $N_2 > N_1$, w.l.o.g.. Suppose that the total signal power is $K \cdot \mathcal{E}$, and the access overhead is ϵ for all channels. It is clear that within each sub-group of channels with the same noise level, it is always optimal to send Gaussian signals over $\Theta_i \in [0, 1]$ fraction of the channels with the same SNR level ν_i, for $i = 1, 2$. Thus the optimal signal can be described as a power allocation function between the two groups. For convenience, we denote
$$(\Theta_1, \nu_1, \Theta_2, \nu_2) = F(\mathcal{E}, \epsilon, \alpha_1, \alpha_2, N_1, N_2). \qquad (12)$$
We also define, following (11), $\nu_{\mathsf{opt},i}, i = 1, 2$ to satisfy
$$\frac{1}{N_i \cdot (1 + \nu_{\mathsf{opt},i})} = \frac{1}{\epsilon + N_i \cdot \nu_{\mathsf{opt},i}} \log(1 + \nu_{\mathsf{opt},i}),$$
which can also be written as
$$\nu_{\mathsf{opt},i} = \frac{\frac{\epsilon}{N_i} - 1}{W((\frac{\epsilon}{N_i} - 1)e^{-1})} - 1$$
or appoximately at low SNR, $\nu_{\mathsf{opt},i} \approx \sqrt{2\frac{\epsilon}{N_i}}$,

Now using the idea of differential power allocation, we can determine F as follows.

- If $\mathcal{E} \le \alpha_1 \cdot (\nu_{\mathsf{opt},1} \cdot N_1 + \epsilon)$,
$$F = \left(\frac{\mathcal{E}}{\alpha_1(\nu_{\mathsf{opt},1} \cdot N_1 + \epsilon)}, \nu_{\mathsf{opt},1}, 0, 0 \right).$$
That is, only a fraction of the channels with noise N_1 are used at an SNR level of $\nu_{\mathsf{opt},1}$, and other channels are not used. Note this is different from the normal water-pouring solution in that a peaky signaling is used even among the good channels.
- If $\mathcal{E} \in (\alpha_1 \cdot (\nu_{\mathsf{opt},1} \cdot N_1 + \epsilon), \alpha_1 \cdot (\nu_{\mathsf{switch}} \cdot N_1 + \epsilon))$ with ν_{switch} satisfying
$$\frac{1}{N_1 \cdot (1 + \nu_{\mathsf{switch}})} = \frac{1}{\epsilon + N_2 \cdot \nu_{\mathsf{opt},2}} \log(1 + \nu_{\mathsf{opt},2}). \qquad (13)$$
This means that when all the signal power used in group 1, it is not enough to have SNR level of ν_{switch}. In this case, all the signal energy should be used evenly over group 1, i.e.,
$$F = \left(1, \frac{\mathcal{E}/\alpha_1 - \epsilon}{N_1}, 0, 0\right).$$
Here, (13) is a straightforward generalization of (11): only when the channels in group 1 have SNR level of ν_{switch}, it starts to become more profitable to use extra signal power in the channels in group 2, whose noise level is higher.
- If $E \in (\alpha_1 \cdot (\nu_{\mathsf{switch}} \cdot N_1 + \epsilon), \alpha_1 \cdot (\nu_{\mathsf{switch}} \cdot N_1 + \epsilon) + \alpha_2 \cdot (\nu_{\mathsf{opt},2} \cdot N_2 + \epsilon))$,
$$F = (1, \nu_{\mathsf{switch}}, \Theta_2, \nu_{\mathsf{opt},2}),$$
where Θ_2 fraction of the channels in group 2 are used at SNR level of $\nu_{opt,2}$, and the expression of Θ_2 is omitted.
- If $E > \alpha_1 \cdot (\nu_{\mathsf{switch}} \cdot N_1 + \epsilon) + \alpha_2 \cdot (\nu_{\mathsf{opt},2} \cdot N_2 + \epsilon)$ As all the channels are used, the problem reduces into a normal water-pouring problem, with total power of $\mathcal{E} - \epsilon$ per channel. Notice that (13) and the definition of $\nu_{\mathsf{opt},2}$ imply that
$$N_1(1 + \nu_{\mathsf{switch}}) = N_2(1 + \nu_{\mathsf{opt},2}),$$
thus when Θ_2 increases to 1, the signal plus noise level of all channels are equal.

Now to summarize, when there is an overhead of using a new channel, the water-pouring process uses a new channel only if the signal plus noise level of the channels already used is strictly higher than a certain threshold. The overhead makes it less profitable to start using a new degree of freedom. This process can thus be better phrased as "glue pouring".

The above solution can be readily used in a multiple access

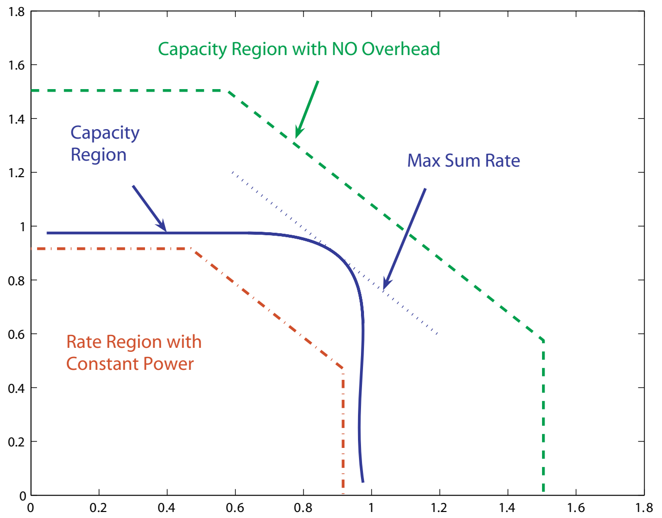

Fig. 2. Achievable Rate Region with Overhead, $N = 1, \epsilon = 2, \mathcal{E}_1 = \mathcal{E}_2 = 3.5$.

channel with processing energy.

III. Multiple Access Channels with Processing Cost

The capacity region for Gaussian multiple access channel is well known. For simplicity, we will focus on the case with only 2 users. Let $\mathcal{E}_i$ denote the total power constraint for user i, the capacity region is determined by the inequalities

$$\begin{aligned} R_i &\leq \log(1+\mathcal{E}_i), \qquad \text{for } i = 1,2; \\ R_1 + R_2 &\leq \log(1+\mathcal{E}_1+\mathcal{E}_2). \end{aligned}$$

It is well-known that any point in the capacity region can be achieved by using time-sharing or rate-splitting [2], [3], and successive cancelation receivers.

Most of the points on the frontier of the capacity region are only achievable if both users are transmitting all the time. TDMA approach, allocating times slots to individual users, is in general sub-optimal. The only exception is that the maximum sum-rate in the capacity region can indeed be achieved by time sharing.

In the presence of processing overhead, the time sharing approach, with one user turned off at sometimes, saves in the overhead cost, and thus is more profitable. Consider the achievable sum rate for example. If we, like in the conventional case, use constant signal power levels for both users, the signal powers are $\mathcal{E}_i - \epsilon$, and the achievable sum rate is

$$R_1 + R_2 = \log(1+\mathcal{E}_1+\mathcal{E}_2 - 2\epsilon).$$

The entire rate region achieved this way is reduced accordingly, as shown in Figure III. In contrast, we can let user i be turned on only in Θ_i fraction of the time, and set $\Theta_i = \frac{\mathcal{E}_i}{\mathcal{E}_1+\mathcal{E}_2}$ for $i = 1, 2$, a higher sum rate is easily achievable.

$$\begin{aligned} R_1 + R_2 &= \sum \Theta_i \log\left(1 + \frac{\mathcal{E}_i}{\Theta_i} - \epsilon\right) \\ &= \log(1+\mathcal{E}_1+\mathcal{E}_2 - \epsilon). \end{aligned}$$

For a general 2-user multiple access problem, one can easily imagine a slight generalization of the time-sharing approach, specified by parameters Θ_1, Θ_2, and Θ_{12}. Here, Θ_i denotes the fraction of time that only user i is on, and Θ_{12} denotes the time that both users are transmitting. Now, the achieved rate for user 1 is simply the rate transmitted in the Θ_1 time period, plus a part in the Θ_{12} period, depending on the decoding order and the time-sharing factor.

Similar to the single user case, the above scheme in concerned only with power allocation, and ignores the possibility to convey information by the randomness of users being on and off. Thus strictly speaking, we will describe only an achievable performance in the sequel, where the focus is on the allocation of degrees of freedom in the channel and signal power.

The power allocation optimization, including the choice of Θ_1, Θ_2, and Θ_{12}, together with the signal power in each period of time. The solution to this problem can be explicitly written out, although cumbersome. In the following, we will describe this solution using the water-pouring result in the previous section.

First, we know that even if there is only one user in the system, due to the overhead cost of accessing a channel, it is sometimes optimal to use only a fraction of the channels. Let $\Theta_{\text{opt},i}$ be the single user optimal fraction for user i given in (7). In the case that $\Theta_{\text{opt},1} + \Theta_{\text{opt},2} \leq 1$, it is obvious that each user should operate exactly the same as the single user case. The achievable rate region is thus a rectangle, specified by the two single user rates.

The more interesting case is when the two users compete for the available degrees of freedom, i.e., $\Theta_{\text{opt},1} + \Theta_{\text{opt},2} > 1$. Suppose that user 1 is decoded first and subtracted from the received signal. Now from user 2's point of view, he occupies the channel all by himself. We use an extra parameter α_2 as a tradeoff factor. That is, we assume that user 2 uses α_2 fraction of the degrees of freedom, instead of $\Theta_{\text{opt},2}$. By doing this, while user 2's data rate is reduced, it can be potentially beneficial to user 1's throughput. The achieved rate region, in terms of the tradeoff of the data rates of the two users, can thus be parameterized by α_2.

Now it is optimal for user 2 to simply spread the signal power in these available channels. The resulting SNR per d.o.f. is

$$\nu_2 = \frac{\mathcal{E}}{\alpha_2} - \epsilon,$$

and achieved data rate is

$$R_2(\alpha_2) = \alpha_2 \log(1+\nu_2).$$

Viewing the signals from user 2 as interference, user 1's channel is precisely the parallel Gaussian channel studied in the previous section, where α_2 fraction of the channels have noise level of $N_2 = 1+\nu_2$, and the other $\alpha_1 = 1-\alpha_2$ fraction of channels have noise level of $N_1 = 1$. The power allocation for this case is fully solved. Let

$$(t_1, \nu_1, t_2, \nu_2) = F(\mathcal{E}_1, \epsilon, \alpha_1, \alpha_2, N_1 = 1, N_2 = 1+\nu_2),$$

for $F(\cdot)$ defined in (12). We can directly compute $\Theta_1 = t_1\alpha_1$, and $\Theta_{12} = t_2\alpha_2$, as well as the optimal SNR levels. The resulting data rate is thus

$$R_1 = \Theta_1 \log(1+\nu_1) + \Theta_{12} \log(1+\nu_2).$$

The above scheme always decodes user 1 before user 2, we can switch the positions of the two users and allow time-sharing between different schemes, which gives the entire rate region, as plotted in Figure III. It is clear from the example that as overhead cost reduces the capacity region, carefully allocating the degrees of freedoms yields better achievable rates, especially the sum rate.

IV. Conclusion

In this paper, we studied the channel capacity with a processing energy overhead. We show that bursty transmission is optimal in the case with large overhead. We interpret the bursty transmission in terms of "glue pouring", and thus described the optimal power allocation for the case of parallel channels with different noise levels. We also used this result to find the optimal rate region of multiple access channels with overhead.

References

[1] P. Youssef-Massaad, M. Médard, and L. Zheng, "Impact of processing energy on the capacity of wireless channels," in *Proc. International Symposium on Info. Theory and its Applications*, Oct. 2004

[2] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, Mar. 1996

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991

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

[5] L. Zheng, D. N.C. Tse, and M. Médard, "Channel coherence in the low SNR regime," *IEEE Trans. Inform. Theory*, Mar. 2007

[6] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, Sept. 1990

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

[8] A. E. Gamal and T. M. Cover,''Multiple user information theory," *Proc. IEEE* vol. 68, no. 12, pp. 1466-1483, 1980

[9] P. Agrawal, J.-C. Chen, S. Kishore, P. Ramanathan, and K. Sivalingam, ''Battery power sensitive video processing in wireless networks," in *Proc. IEEE PIMRC'98*, Sept. 1998

[10] P. J. M. Havinga and G. J. M. Smit, ''Minimizing energy consumption for wireless computers in Moby Dick," in *Proc. IEEE International Conference on Personal Wireless Communication* ICPWC'97, Dec. 1997.

[11] J. M. Rulnick and N. Bambos, ''Mobile power management for maximum battery life in wireless communication networks," in *Proc. IEEE INFOCOM 1996*.

[12] A. Sampath, P. S. Kumar, and J. Holtzman, "Power control and resource management for a multimedia cdma wireless system," in *Proc. Sixth International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'95)*.

[13] M. Stemm and R. Katz,"Reducing power consumption of network interfaces in hand-held devices," in *Proc. Third International Workshop on Mobile Multimedia Communications (MoMuc-3)*, Dec. 1996.

[14] A. Chandrakasan, R. Min, M. Bhardwaj, S.-H. Cho, and A. Wang, "Power aware wireless microsensor systems," keynote paper ESSCIRC, 2002.

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

[16] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Select. Areas Commun.*, 2004.

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