

# Energy-Efficient Communication in the Presence of Synchronization Errors

Yu-Chih Huang, Urs Niesen, and Piyush Gupta

## Abstract

Communication systems are traditionally designed to have tight transmitter-receiver synchronization. This requirement has negligible overhead in the high-SNR regime. However, in many applications, such as wireless sensor networks, communication needs to happen primarily in the energy-efficient regime of low SNR, where requiring tight synchronization can be highly suboptimal.

In this paper, we model the noisy channel with synchronization errors as a duplication/deletion/substitution channel. For this channel, we propose a new communication scheme that requires only loose transmitter-receiver synchronization. We show that the proposed scheme is asymptotically optimal for the Gaussian channel with synchronization errors in terms of energy efficiency as measured by the rate per unit energy. In the process, we also establish that the lack of synchronization causes negligible loss in energy efficiency. We further show that, for a general discrete memoryless channel with synchronization errors and a general input cost function admitting a zero-cost symbol, the rate per unit cost achieved by the proposed scheme is within a factor two of the information-theoretic optimum.

## I. INTRODUCTION

Traditionally, data transmission in a communication system is based on tight synchronization between the transmitter and the receiver. This tight synchronization is usually achieved through either of two strategies. In the first strategy, synchronization is achieved through periodic transmission of pilot signals, followed by transmission of information over the synchronized channel (see, e.g., [1, Chapter 6.3]). In the second strategy, data bits are differentially encoded and then modulated (e.g., differential pulse-position-modulation) [1, Chapter 4.3.2], which implicitly achieves tight synchronization.

The above strategies work well at high signal-to-noise ratios (SNRs) as the energy overhead of achieving tight synchronization is negligible compared to that of data transmission. However, in many applications, such as wireless sensor networks, space communication, or in general any communication system requiring high energy efficiency, communication by necessity has to primarily take place in the low-SNR regime (due to the concavity of the power-rate function). In such scenarios, the energy overhead to achieve tight synchronization becomes significant and can render the aforementioned strategies highly suboptimal in terms of energy efficiency. In fact, requiring tight transmitter-receiver synchronization can have arbitrarily large loss in performance in terms of energy efficiency (see Example 3 in Section III).

To mitigate this, in this paper, we develop and analyze a framework to perform data transmission while only requiring loose synchronization between the transmitter and the receiver. To focus on the energy-efficiency aspect, we choose the rate per unit cost (with energy being a prime example of the cost) as our performance metric. We model synchronization errors through channel duplications/deletions—an approach introduced in [2]. To motivate this model, consider a transmitter-receiver pair with unsynchronized clocks, as illustrated in Fig. 1. Due to the absence of synchronization, the value of the clock at the receiver exhibits drift and jitter with respect to the value of the reference clock at the transmitter. This leads to the receiver sampling the transmitted signal either faster than the transmitter, leading to channel duplications, or slower, leading to channel deletions.

This work was supported in part by the Air Force Office of Scientific Research, Arlington, VA, USA, under grant FA9550-09-1-0317.

This paper was presented in part at the 2013 IEEE International Symposium on Information Theory.

Y.-C. Huang is with the Department of Communication Engineering, National Taipei University, New Taipei City, 23741, Taiwan (e-mail: ychuang@mail.ntpu.edu.tw). U. Niesen and P. Gupta were with Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974, USA. They are now with the Qualcomm New Jersey Research Center, Bridgewater, NJ 08807, USA (email: urs.niesen@ieee.org, p.gupta@ieee.org).

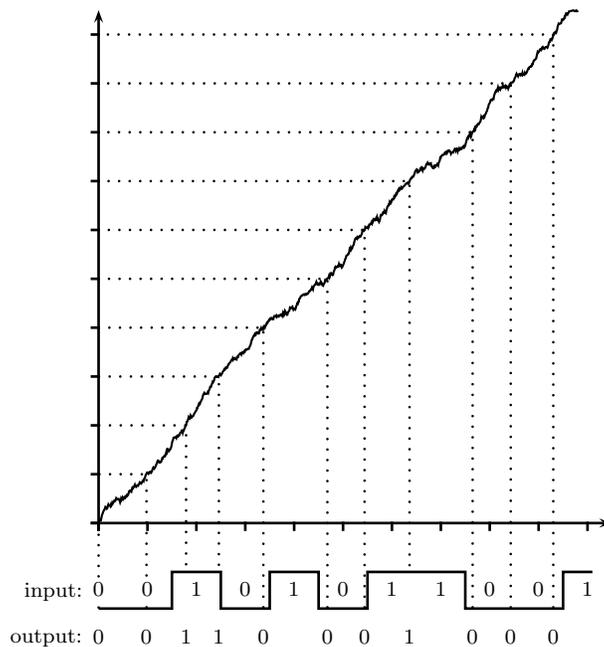


Fig. 1. An example of unsynchronized transmitter-receiver clocks. The figure plots the value of the receiver clock ( $y$ -axis) as a function of the value of the reference clock at the transmitter ( $x$ -axis). The drift and jitter of the receiver clock are visible. For a transmitted input sequence, the lack of synchronization leads to duplications/deletions in the corresponding sampled output sequence at the receiver (illustrated here for the case without receiver noise).

Before we describe the contributions of this work in more detail in Section I-B, we provide a brief overview of related work on energy-efficient communication and on channels with synchronization errors.

### A. Related Work

It is well known that the capacity per unit energy of a Gaussian channel with noise variance  $\eta^2$  is  $1/(2\eta^2 \ln 2)$  and that this can be achieved with appropriately designed pulse-position modulation [3]. For a general discrete memoryless channel (DMC), [4] has analyzed the reliability function of the rate per unit cost. Subsequently, [5] has obtained a succinct single-letter characterization for the capacity per unit cost of a DMC with a general cost function. These results, however, strongly depend on the channel being memoryless. As discussed next, synchronization errors introduce memory in the channel, and thus the aforementioned results do not apply.

The duplication/deletion/substitution channel was introduced as a model for a channel with synchronization errors by Dobrushin in 1967 [2]. Despite significant research effort since then, the capacity of this channel is still not known [6]–[13]. Indeed, even for one of the simplest versions of this problem, the noiseless binary deletion channel, only loose bounds on the capacity are known. For example, the recent paper [7] provides an approximation of the capacity of the binary deletion channel to within a factor 9 in general. Tighter bounds have been obtained for some specific regimes of the deletion probability: For instance, in [12] [13], upper and lower bounds are provided and vanishing gap is shown in the asymptotically small deletion-probability regime; while improved upper bounds based on a numerical approach are obtained in [11]. The main difficulty in analyzing these channels arises from the channel memory introduced by the duplications and deletions, which prevents a direct application of the standard information-theoretic tools.

It is worth emphasizing that the synchronization errors considered here are those at the symbol level. There are other types of synchronization issues. One such issue is frame synchronization, where errors are caused by the incorrect identification of the location of the “sync word” in the frame [14]. Thus, frame synchronization deals with synchronization errors at the level of the block size as compared to

symbol-level synchronization considered here. Energy-efficient communication in the presence of such frame asynchronism has been investigated in [15].

### B. Summary of Results

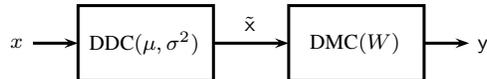


Fig. 2. The end-to-end communication channel between an unsynchronized transmitter-receiver pair is modeled by concatenating a duplication/deletion channel  $DDC(\mu, \sigma^2)$  with a discrete memoryless channel  $DMC(W)$ .

In this paper, we consider communication channels which, in addition to synchronization errors, exhibit receiver noise. We model the end-to-end communication channel between the transmitter and the receiver as the concatenation of two sub-channels, as illustrated in Fig. 2. The first sub-channel is a duplication/deletion channel (DDC), which models synchronization errors. This DDC is a sub-class of the more general insertion/deletion/substitution channels considered by Dobrushin [2] in that *random* insertions (as opposed to duplications only) are not considered here. The second sub-channel is a noisy memoryless channel, which models errors due to the receiver noise. The concatenation of the two channels is a duplication/deletion/substitution channel. The details of this model are discussed in Section II.

We first study communication systems with synchronization errors operating over Gaussian channels. We propose a new communication scheme that requires only loose synchronization between the transmitter and the receiver. Specifically, the scheme deals with the lack of transmitter-receiver synchronization by incorporating guard spaces into traditional pulse-position-modulation signaling. Decoding at the receiver is based on a sequence of independent hypothesis tests. When the aforementioned durations are chosen appropriately, we show that the scheme asymptotically achieves the information-theoretically optimal performance in terms of the energy efficiency, i.e., the capacity per unit energy. In the process, we also establish that in regimes of practical interest the lack of transmitter-receiver synchronization causes negligible loss in terms of energy efficiency. We also consider the compound setting where even the statistical properties of synchronization errors are not known precisely a priori. For this setting, we show similar results by developing new pulse-position-modulation waveforms, where the signal energy is spread over increasing intervals and guard spaces of increasing lengths are introduced.

We then analyze communication systems with synchronization errors operating over general DMCs and with a general cost function admitting a zero-cost symbol. We generalize the proposed achievable scheme for the Gaussian case to DMCs, and we show that the scheme achieves a rate per unit cost within a factor two of the information-theoretic optimum. Thus, while only loose bounds are known for the *capacity* of the general duplication/deletion/substitution channel, we provide here a tight approximation for its *capacity per unit cost*. To establish this, we obtain an upper bound on the capacity per unit cost of the channel in Fig. 2 by considering the effect of the DDC as a specific way of encoding for the DMC with an appropriately modified cost function. The upper bound is then obtained by utilizing the characterization of the capacity per unit cost for memoryless channels in [5].

### C. Organization

The remainder of the paper is organized as follows. Section II provides the detailed description of the channel model and the problem formulation. The main results of the paper are summarized in Section III. Section IV describes the proposed scheme for a general discrete memoryless channel with synchronization errors. Section V derives an upper bound on its capacity per unit cost. Section VI discusses the proposed scheme for the Gaussian channel with synchronization errors and establishes its asymptotic optimality. Lastly, Section VII analyzes the performance of the scheme for the Gaussian channels with synchronization errors where even the statistical properties of errors are not known precisely a priori (i.e., the compound setting).

## II. CHANNEL MODEL AND PROBLEM STATEMENT

We consider a duplication/deletion/substitution channel with a cost constraint. The duplication/deletion/substitution channel consists of a DDC connected to a DMC as shown in Fig. 2 in Section I. The duplication/deletion part of the channel models synchronization errors (see Fig. 1 in Section I), the substitution part models noise.

The DDC maps the input sequence  $(x[1], \dots, x[T]) \in \mathcal{X}^T$  to the output sequence  $(\tilde{x}[1], \dots, \tilde{x}[L]) \in \mathcal{X}^L$  for some random length  $L$ , where here and in the following we use sans-serif font to denote random variables. The actions of the DDC are governed by the i.i.d. sequence of states  $(s[1], \dots, s[T]) \in \{0, 1, 2, \dots\}^T$ . State  $s[t]$  describes how many times input symbol  $x[t]$  appears at the output of the DDC.

Formally, the total number of output bits is given by

$$L \triangleq \sum_{t=1}^T s[t].$$

Observe that  $L$  is a random variable depending on the state sequence of the DDC. Define for each  $\ell \in \{1, 2, \dots, L\}$  the random variable

$$t[\ell] \triangleq \min \left\{ \tau : \sum_{j=1}^{\tau} s[j] \geq \ell \right\}.$$

The relationship between the input and output of the DDC is then given by

$$\tilde{x}[\ell] \triangleq x[t[\ell]].$$

We illustrate the operation of the DDC with an example.

### Example 1.

$$\begin{array}{rcccccc} \mathbf{x} & = & x[1] & x[2] & x[3] & x[4] & x[5] & x[6] \\ \mathbf{s} & = & 1 & 1 & 2 & 1 & 0 & 2 \\ \mathbf{t} & = & 1 & 2 & 3 & 3 & 4 & 6 & 6 \\ \tilde{\mathbf{x}} & = & x[1] & x[2] & x[3] & x[3] & x[4] & x[6] & x[6] \end{array}$$

In the example,  $T = 6$  and  $L = 7$ . Here  $\mathbf{x}$  is the vector of inputs and  $\tilde{\mathbf{x}}$  is the vector of outputs of the DDC.  $\mathbf{s}$  is the vector of states and the corresponding vector of sampling times is  $\mathbf{t}$ . For concreteness, assume

$$\mathbf{x} = (0 \ 0 \ 1 \ 0 \ 1 \ 0).$$

Then the output of the DDC is

$$\tilde{\mathbf{x}} = (0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0),$$

see also Fig. 1 in Section I. ◇

We denote by

$$\mu \triangleq \mathbb{E}(s[1])$$

and

$$\sigma^2 \triangleq \text{var}(s[1])$$

the mean and variance of the duplication/deletion process, respectively, and we refer to any DDC with those parameters as  $\text{DDC}(\mu, \sigma^2)$ . Here,  $\mu$  and  $\sigma^2$  can be interpreted as capturing the drift and jitter of the receiver clock, respectively. In most situations arising in practice, the parameter  $\mu$  is close to 1: For instance, in 3GPP-Long Term Evolution (LTE), one of the modes (LTE-TDD) requires clock accuracy of 50 parts per billion [16], i.e.,  $\mu = 1 \pm 5 \cdot 10^{-8}$ ; even in a popular compact wireless sensor node, Mica2 mote, the clock accuracy is specified as 40 parts per million [17], i.e.,  $\mu = 1 \pm 4 \cdot 10^{-5}$ .

Our results will be presented for arbitrary duplication/deletion processes with finite mean and variance. For illustrative purposes, we present a commonly used special case of this setting.

**Example 2.** One commonly used definition of the state process is

$$s[t] = \begin{cases} 1, & \text{w.p. } 1 - d, \\ 0, & \text{w.p. } d, \end{cases}$$

for some parameter  $d$ . This results in the so-called *deletion channel*, which deletes each input symbol independently with probability  $d$ .  $\diamond$

The output of the  $\text{DDC}(\mu, \sigma^2)$  is then fed into a discrete memoryless channel  $\text{DMC}(W)$  described by the distribution  $W(\cdot|\tilde{x})$  of the channel output  $y \in \mathcal{Y}$  given the channel input  $\tilde{x} \in \mathcal{X}$ . The duplication/deletion/substitution channel is the (random) mapping from  $x$  to  $y$  described by the concatenation of the  $\text{DDC}(\mu, \sigma^2)$  and the  $\text{DMC}(W)$ .

*Remark:* This model is motivated by a widely-used digital demodulation architecture consisting of an envelope detector followed by matched filters (see for example [1, pages 302–304]). The model captures the duplication/deletion of symbols in such an architecture due to the lack of transmitter-receiver synchronization. However, this model does not account for the fact that over or under sampling of the received signal introduces different signal attenuations. As a consequence, the model as defined here has physical meaning only when  $\mu \approx 1$ , which is exactly the scenario motivating the analysis in this paper. For other regimes of  $\mu$ , the model may need to be suitably modified (for example by normalizing the channel output) to account for the signal attenuation due to over or under sampling.

The goal is to maximize the number of bits reliably transmitted per unit cost over this duplication/deletion/substitution channel formed by the concatenation of the  $\text{DDC}(\mu, \sigma^2)$  with the  $\text{DMC}(W)$ . We adopt the framework of [4], [5]. The *cost function*  $c: \mathcal{X} \rightarrow \mathbb{R}_+$  associates with each input symbol  $x \in \mathcal{X}$  the cost  $c(x)$  incurred by transmitting  $x$  over the channel. We make the assumption that  $\mathcal{X}$  contains a free input symbol, and without loss of generality we label this symbol as 0. In other words,  $0 \in \mathcal{X}$  and  $c(0) = 0$ . For an input sequence  $(x[1], \dots, x[T]) \in \mathcal{X}^T$ , the cost is given by

$$c((x[1], \dots, x[T])) \triangleq \sum_{t=1}^T c(x[t]).$$

A  $(T, M, P, \varepsilon)$  *code* consists of an encoder-decoder pair  $f^T$  and  $\{\varphi^\ell\}_{\ell=0}^\infty$  and  $M$  codewords. The encoder  $f^T$  maps the message  $m \in \{1, 2, \dots, M\}$  to a codeword

$$(x_m[1], \dots, x_m[T]),$$

of length  $T$  and cost at most  $P$ . The decoder consists of several sub-decoders  $\{\varphi^\ell\}_{\ell=0}^\infty$ , one for each possible realization  $\ell$  of  $L$ , and has average (assuming equiprobable messages) probability of decoding error at most  $\varepsilon$ .

**Definition.** A rate  $\hat{R}$  per unit cost is *achievable* if for every  $\varepsilon > 0$  and every large enough  $M$  there exists a  $(T, M, P, \varepsilon)$  code satisfying<sup>1</sup>  $\log(M)/P \geq \hat{R}$ . The *capacity per unit cost*  $\hat{C}$  is the supremum of achievable rates per unit cost.<sup>2</sup>

It can be seen from the above definition that, since the receiver knows the realization of  $L$  (i.e., it knows the random length of the transmission), the present paper considers the scenario of a one-shot transmission.

<sup>1</sup>Throughout this paper,  $\log(\cdot)$  and  $\ln(\cdot)$  denote the logarithms to the base 2 and  $e$ , respectively.

<sup>2</sup>In general,  $\hat{C}$  might take the value of  $-\infty$  in case that the set of achievable rates is empty. Hence,  $\hat{C}$  is well defined on the extended real line  $\mathbb{R}$ . For the case of a zero-cost input symbol as considered here, we will see later that  $\hat{C} \geq 0$ .

Throughout this paper, we are interested in  $\hat{C}(\mu, \sigma^2, W)$ , the capacity per unit cost of the duplication/deletion/substitution channel given by the DDC( $\mu, \sigma^2$ ) concatenated with the DMC( $W$ ). We also consider the compound capacity per unit cost  $\hat{C}([\mu_1, \mu_2], [\sigma_1^2, \sigma_2^2], W)$ , for which the encoder and decoder have to be able to operate on any DDC( $\mu, \sigma$ ) with  $\mu \in [\mu_1, \mu_2]$  and  $\sigma \in [\sigma_1^2, \sigma_2^2]$  without knowledge of the actual values of  $\mu$  and  $\sigma^2$ . This compound setting is of practical relevance, since usually the mean clock drift  $\mu$  is only specified as an interval (and might indeed be slowly time varying) and is hence not known exactly at the transmitter or the receiver. We also treat the Gaussian version of the problem, where the output of the DDC( $\mu, \sigma^2$ ) is subject to additive Gaussian noise of mean zero and variance  $\eta^2$ . The cost function is in this case the signal energy, i.e.,  $c(x) = x^2$ . With slight abuse of notation, we refer to the capacity per unit energy in this case as  $\hat{C}(\mu, \sigma^2, \mathcal{N}(0, \eta^2))$ .

### III. MAIN RESULTS

In this section, we summarize the main results; their proofs are discussed in subsequent sections. We start with the results for Gaussian channels with synchronization errors for the case where the statistics  $\mu$  and  $\sigma^2$  of the duplication/deletion channel are known at the transmitter and the receiver.

**Theorem 1.** *The Gaussian channel with synchronization errors having duplication/deletion process with mean  $\mu$  and variance  $\sigma^2$  and having noise power  $\eta^2$  has capacity per unit energy*

$$\hat{C}(\mu, \sigma^2, \mathcal{N}(0, \eta^2)) = \frac{\mu}{2\eta^2 \ln 2}.$$

Recall that the capacity per unit energy of the Gaussian channel is  $1/(2\eta^2 \ln 2)$ . Furthermore, as discussed in Section II, the mean  $\mu$  of the duplication/deletion process is typically close to 1. Hence, Theorem 1 implies that the lack of synchronization results in only negligible loss in the capacity per unit energy.

To establish achievability, we propose a new communication scheme that jointly performs data modulation and loose synchronization. To this end, we develop signaling waveforms where the signal energy is spread over increasing intervals and guard spaces of increasing lengths are introduced. Decoding at the receiver is based on a sequence of independent hypothesis tests, which are carefully chosen to account for the uncertainty arising due to the lack of tight synchronization. In Section VI, we show that, by appropriately choosing the aforementioned durations, the probability of error can be made arbitrarily small for any rate per unit energy up to  $\hat{C}$ . The details of the analysis of this scheme are reported in Section VI. The upper bound in Theorem 1 follows as a special case of the upper bound derived for a general DMC with synchronization errors, discussed in Theorem 3 below.

*Remark:* From the above results, it follows that the capacity per unit cost  $\hat{C}(\mu, \sigma^2, W)$  depends on the distribution of the duplication/deletion process  $s$  only through its mean. This is because  $\sigma^2$  is finite and we have a zero-cost symbol. Hence, one can use a long block of the zero-cost symbol to smooth the variation of the duplication/deletion process.

Next, consider the case where the exact statistical properties of the duplication/deletion process are not known a priori. Instead, we only know a range for each parameter, i.e., the mean  $\mu$  is in  $[\mu_1, \mu_2]$  and the variance in  $[0, \sigma^2]$ . We are interested in a communication scheme that works simultaneously for every possible set of parameters in this range. As pointed out in Section II, this compound setting is of practical relevance, since the precision of the transmitter and receiver clocks are usually only known to lie within some range.

**Theorem 2.** *The class of Gaussian channels with synchronization errors having duplication/deletion process with mean  $\mu \in [\mu_1, \mu_2]$  and variance upper bounded by  $\sigma^2$  and having noise power  $\eta^2$  has compound capacity per unit energy*

$$\hat{C}([\mu_1, \mu_2], [0, \sigma^2], \mathcal{N}(0, \eta^2)) = \frac{\mu_1}{2\eta^2 \ln 2}.$$

Comparing Theorems 1 and 2, we see that

$$\hat{C}([\mu_1, \mu_2], [0, \sigma^2], \mathcal{N}(0, \eta^2)) = \min_{\mu \in [\mu_1, \mu_2]} \hat{C}(\mu, \sigma^2, \mathcal{N}(0, \eta^2)).$$

Since a scheme for the compound setting must work for any possible value of  $\mu$  and  $\sigma^2$  of the duplication/deletion process, it is clear that it must work for the worst one, so that

$$\hat{C}([\mu_1, \mu_2], [0, \sigma^2], \mathcal{N}(0, \eta^2)) \leq \min_{\mu \in [\mu_1, \mu_2]} \hat{C}(\mu, \sigma^2, \mathcal{N}(0, \eta^2)) = \frac{\mu_1}{2\eta^2 \ln 2}.$$

Theorem 2 thus shows that there is no further loss beyond this resulting from the lack of precise knowledge of the duplication/deletion statistics at the transmitter and the receiver. The proof of Theorem 2 is presented in Section VII.

Finally, consider a duplication/deletion channel  $\text{DDC}(\mu, \sigma^2)$  concatenated with a general discrete memoryless channel  $\text{DMC}(W)$  specified by its transition probability matrix  $W: \mathcal{X} \rightarrow \mathcal{Y}$ . Furthermore, consider an arbitrary cost function  $c: \mathcal{X} \rightarrow \mathbb{R}_+$ . As mentioned in Section II, we assume that  $0 \in \mathcal{X}$  and  $c(0) = 0$ . Then, the following bounds hold on the capacity per unit cost.

**Theorem 3.** *The duplication/deletion/substitution channel consisting of a  $\text{DDC}(\mu, \sigma^2)$  concatenated with a  $\text{DMC}(W)$  has capacity per unit cost  $\hat{C}(\mu, \sigma^2, W)$  satisfying*

$$\frac{\mu}{2} \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x)||W(\cdot|0))}{c(x)} \leq \hat{C}(\mu, \sigma^2, W) \leq \mu \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x)||W(\cdot|0))}{c(x)}$$

where  $D(P||Q)$  is the Kullback-Leibler divergence between distributions  $P$  and  $Q$  and where  $0 \in \mathcal{X}$  is an input symbol with zero cost.

Theorem 3 approximates the capacity per unit cost of a general DMC with synchronization errors and general cost function admitting a zero-cost symbol to within a factor two. In contrast, recall from Section I that for the *capacity*, even of the noiseless deletion channel, only loose bounds are known despite over four decades since the introduction of the model in [2].

It was shown in [5] that the capacity per unit cost  $\hat{C}(W)$  of a  $\text{DMC}(W)$  with zero-cost symbol 0 is

$$\hat{C}(W) = \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x)||W(\cdot|0))}{c(x)}. \quad (1)$$

Thus, from the lower and upper bounds in Theorem 3, we obtain the following corollary, showing that the loss due to synchronization errors is within a factor between  $\mu/2$  and  $\mu$ .

**Corollary 4.** *The duplication/deletion/substitution channel consisting of a  $\text{DDC}(\mu, \sigma^2)$  concatenated with a  $\text{DMC}(W)$  has capacity per unit cost  $\hat{C}(\mu, \sigma^2, W)$  satisfying*

$$\frac{\mu}{2} \hat{C}(W) \leq \hat{C}(\mu, \sigma^2, W) \leq \mu \hat{C}(W).$$

*Remark:* The upper bound in Corollary 4 in fact holds irrespective of whether there is a zero-cost symbol or not (as will become clear from the proof in Section V).

The achievability in Theorem 3 is established by generalizing the proposed scheme for the Gaussian channel with synchronization errors to DMCs. The details are discussed in Section IV. For the upper bound on the capacity per unit cost in Theorem 3, we treat the effect of the DDC as a specific way of encoding for the DMC with an appropriately modified cost function. The upper bound is then obtained by utilizing the characterization of the capacity per unit cost for *memoryless* channels in [5]. Section V provides the details.

We conclude this section by illustrating through an example that the conventional schemes based on tight synchronization between the transmitter and the receiver can be highly suboptimal in terms of their rate per unit cost.

**Example 3.** Let us consider the simplest synchronized communication setting: a channel with binary input and no noise, i.e.,  $W(x|x) = 1$  for  $x \in \{0, 1\}$ . Further, let the cost function be the number of ones transmitted, i.e.,  $c(x) = x$ . This is a DMC, and by (1), its capacity per unit cost is

$$\hat{C}(W) = \frac{D(W(\cdot|1)||W(\cdot|0))}{c(1)} = \infty.$$

Let us now consider the scenario where the transmitter and the receiver are no longer perfectly synchronized. Specifically, the input signals are first corrupted by a deletion channel with deletion probability  $d \in (0, 1)$  (see Example 2 in Section II for a formal definition of this special case of a DDC), before being sent over the aforementioned noiseless channel  $W$ .

Consider first the operation of conventional schemes based on tight synchronization. In this example, we take this to mean any scheme that detects and corrects deletions without letting them accumulate. This definition applies to schemes using pilots [1, Chapter 6.3] as well as to schemes using differential modulation [1, Chapter 4.3.2]. To maintain tight synchronization, the channel inputs cannot contain too many consecutive zeros (since otherwise deletions would accumulate without any way of correcting for them). On average, we expect to see about one deleted bit every  $1/d$  transmitted bits. Thus, roughly every  $1/d$  channel inputs needs to be a 1 at a cost of  $c(1) = 1$ . Now, over a block of  $1/d$  binary channel uses, we can reliably transmit at most  $1/d$  bits. Hence, the rate per unit cost achieved by any scheme based on tight synchronization is at most

$$\hat{R}_{\text{sync}}(d, W) \leq \frac{1}{d} < \infty.$$

On the other hand, from Corollary 4, the communication scheme proposed in this paper achieves a rate per unit cost that is within a factor of  $\mu/2 = (1-d)/2$  of the capacity per unit cost  $\hat{C}(W)$  of the underlying DMC( $W$ ). Hence, the capacity per unit cost with synchronization errors is

$$\hat{C}(d, W) \geq \frac{1-d}{2} \hat{C}(W) = \infty.$$

Thus, even in the presence of synchronization errors, the rate per unit cost achieved by the proposed scheme is arbitrarily large. This illustrates that the improvement in the rate per unit cost achieved by the proposed scheme over schemes based on tight synchronization can be unbounded.  $\diamond$

#### IV. PROOF OF LOWER BOUND IN THEOREM 3

In this section, we propose a coding scheme that achieves the lower bound on the rate per unit cost stated in Theorem 3 for a general DMC with synchronization errors. The scheme uses a type of pulse-position modulation. To send message  $m$ , the encoder sends a burst of symbols  $x^* \in \mathcal{X} \setminus \{0\}$  at a position corresponding to this message. The decoder searches for the location of the pulse using a sliding window. Once the pulse is located, the decoder checks which decision region it is in and declares the corresponding message. In order to deal with duplications and deletions, guard spaces need to be introduced around the pulses and the decision regions need to be chosen judiciously. We proceed with a detailed description of the scheme and its analysis.

*Encoding:* Fix a target error probability  $\varepsilon \in (0, 1)$  and a number  $\delta \in (0, 1)$ . Let  $x^* \in \mathcal{X} \setminus \{0\}$  be a fixed nonzero channel input. The codeword for message  $m \in \{1, 2, \dots, M\}$  is

$$\mathbf{x}_m \triangleq (\mathbf{0}_{(m-1)N}, x^* \cdot \mathbf{1}_B, \mathbf{0}_{N-B}, \mathbf{0}_{(M-m)N}),$$

where

$$N \triangleq \lceil 36M\sigma^2/(\mu^2\varepsilon) \rceil = \Theta(M),$$

and

$$B \triangleq \left\lceil \frac{(2+\delta)\log(M)}{\mu D(W(\cdot|x^*)||W(\cdot|0))} \right\rceil = \Theta(\log(M)).$$

Thus, to communicate message  $m$ , the transmitter sends a pulse of symbols  $x^*$  at position  $(m-1)N+1$  and of duration  $B$ . Observe that between adjacent possible pulse positions is a guard space of  $N-B$  zeroes. The block length of this code is  $T=MN$  and the cost of each codeword is

$$P = Bc(x^*). \quad (2)$$

*Decoding:* Recall that the output  $\mathbf{y}$  of the channel has length  $L$ . The decoder forms the subsequences

$$\mathbf{y}_\ell \triangleq (y[\ell], y[\ell+1], \dots, y[\ell + \lfloor B\mu - \beta \rfloor - 1])$$

of length  $\lfloor B\mu - \beta \rfloor$  for  $\ell \in \{1, 2, \dots, L - \lfloor B\mu - \beta \rfloor + 1\}$  with

$$\beta \triangleq \sqrt{4B\sigma^2/\varepsilon} = \Theta(\log^{1/2}(M)).$$

Similarly, define the subsequences

$$\tilde{\mathbf{x}}_\ell \triangleq (\tilde{x}[\ell], \tilde{x}[\ell+1], \dots, \tilde{x}[\ell + \lfloor B\mu - \beta \rfloor - 1])$$

of the output of the DDC / input to the DMC (not observable at the receiver). Finally, define the open intervals  $\tilde{\mathcal{D}}_m \triangleq ((m-1)N\mu + 1 - \nu, (m-1)N\mu + 1 + \nu)$  for  $m \in \{2, \dots, M\}$  and define the decision regions

$$\mathcal{D}_m \triangleq \begin{cases} \{1\}, & \text{for } m = 1 \\ \mathbb{N} \cap \tilde{\mathcal{D}}_m, & \text{for } m \in \{2, \dots, M\} \end{cases}$$

with

$$\nu \triangleq \sqrt{4MN\sigma^2/\varepsilon} = \Theta(M).$$

In words, the decision region  $\mathcal{D}_m$  for message  $m$  consists of all integer points within distance  $\nu$  of  $(m-1)N\mu + 1$ .

The receiver performs independent hypothesis tests for each  $\mathbf{y}_\ell$  for the two hypotheses

$$\begin{aligned} H^0 &\triangleq \{\tilde{\mathbf{x}}_\ell = \mathbf{0}\}, \\ H^1 &\triangleq \{\tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}\}. \end{aligned}$$

Let  $\hat{H}_\ell$  be the decision of the hypothesis test for  $\mathbf{y}_\ell$ . The receiver declares that message  $m$  was sent if  $\hat{H}_\ell = H^1$  for *some*  $\ell \in \mathcal{D}_m$  and  $\hat{H}_\ell = H^0$  for *all*  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ . If no such  $m$  exists, an error is declared.

In order for the decoder to be well defined, we need to ensure that the decision regions are disjoint, i.e., that  $\mathcal{D}_m \cap \mathcal{D}_{m'} = \emptyset$  for  $m \neq m'$ . This is the case since, by the definitions of  $N$  and  $\nu$ ,

$$\begin{aligned} N\mu &\geq \sqrt{N} \sqrt{\frac{36M\sigma^2}{\mu^2\varepsilon}} \mu \\ &= 3\sqrt{\frac{4MN\sigma^2}{\varepsilon}} \\ &= 3\nu, \end{aligned} \quad (3)$$

so that

$$mN\mu + 1 - \nu \geq (m-1)N\mu + 1 + \nu$$

for all  $m$ .

*Error Analysis:* Assume that message  $m$  was sent. Let  $\mathcal{E}_{1,\ell}$  be the event that  $\hat{H}_\ell = H^0$ , and  $\mathcal{E}_{2,\ell}$  be the event that  $\hat{H}_\ell = H^1$ . Define the missed-detection event

$$\mathcal{E}_1 \triangleq \bigcap_{\ell \in \mathcal{D}_m} \mathcal{E}_{1,\ell}$$

and the false-alarm event

$$\mathcal{E}_2 \triangleq \bigcup_{m' \neq m} \bigcup_{\ell \in \mathcal{D}_{m'}} \mathcal{E}_{2,\ell}.$$

The probability of decoding error for message  $m$  is equal to  $\mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2)$ , where  $\mathbb{P}_m$  denotes probability conditioned on message  $m$  being sent.

We continue by upper bounding this probability. It will be convenient to define two auxiliary events isolating the behavior of the DDC. Let  $\mathcal{E}_3$  be the event that the total number of symbols in  $\tilde{\mathbf{x}}$  resulting from the first  $(m-1)N$  transmitted symbols is outside the interval  $((m-1)N\mu - \nu, (m-1)N\mu + \nu)$ , and let  $\mathcal{E}_4$  be the event that the number of symbols in  $\tilde{\mathbf{x}}$  resulting from symbols transmitted during time slots  $(m-1)N+1$  to  $(m-1)N+B$  is outside the interval  $(B\mu - \beta, B\mu + \beta)$ . We have

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2) &= \mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2 \mid \mathcal{E}_3 \cup \mathcal{E}_4) \mathbb{P}_m(\mathcal{E}_3 \cup \mathcal{E}_4) + \mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \mathbb{P}_m(\mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \mathbb{P}_m(\mathcal{E}_3 \cup \mathcal{E}_4) + \mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \mathbb{P}_m(\mathcal{E}_3) + \mathbb{P}_m(\mathcal{E}_4) + \mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) + \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c). \end{aligned} \quad (4)$$

The first two probabilities correspond to the events that the DDC( $\mu, \sigma^2$ ) is not well behaved and the last two correspond to the two possible detection errors caused by the DMC( $W$ ) conditioned on the behavior of the DDC to be as expected.

We continue by upper bounding each of the terms in (4) in turn. By Chebyshev's inequality,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_3) &= \mathbb{P}_m\left(\left|\sum_{t=1}^{(m-1)N} \mathbf{s}[t] - (m-1)N\mu\right| \geq \nu\right) \\ &\leq \frac{(m-1)N\sigma^2}{\nu^2} \\ &\leq \varepsilon/4, \end{aligned} \quad (5)$$

and

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_4) &= \mathbb{P}_m\left(\left|\sum_{t=(m-1)N+1}^{(m-1)N+B} \mathbf{s}[t] - B\mu\right| \geq \beta\right) \\ &\leq \frac{B\sigma^2}{\beta^2} \\ &= \varepsilon/4, \end{aligned} \quad (6)$$

where we have used the definitions of  $\nu$  and  $\beta$ , respectively.

We proceed with the analysis of  $\mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c)$  and  $\mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c)$ . The following is the key observation for this analysis. Conditioned on message  $m$  being sent and on  $\mathcal{E}_3^c \cap \mathcal{E}_4^c$ , the elements in the decision regions satisfy the following two properties for  $M$  large enough (not depending on  $m$ ):

- 1) For every  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ , we have  $\tilde{\mathbf{x}}_\ell = \mathbf{0}$ . Hence, the symbols in the subsequence  $\mathbf{y}_\ell$  are i.i.d. with distribution  $W(\cdot \mid 0)$ .
- 2) There exists at least one  $\ell \in \mathcal{D}_m$  such that  $\tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}$ . Hence, the symbols in the subsequence  $\mathbf{y}_\ell$  are i.i.d. with distribution  $W(\cdot \mid x^*)$ .

We start by proving property 1. By construction of the codewords, and since the DDC part of the channel can only delete and duplicate symbols but never "create" them, we only need to argue that the burst of symbol  $x^*$  sent in block  $m$  by the transmitter cannot be shifted into the decision region  $\mathcal{D}_{m'}$ .

Assume first  $m' < m$ . The right-most element of  $\mathcal{D}_{m'}$  is at position less than or equal to  $(m-2)N\mu + 1 + \nu$ , and therefore the right-most element of  $\tilde{\mathbf{x}}_\ell$  with  $\ell \in \mathcal{D}_{m'}$  is at position at most

$$(m-2)N\mu + \nu + B\mu.$$

Now, conditioned on  $\mathcal{E}_3^c$ , there are at least  $(m-1)N\mu - \nu$  symbols 0 in  $\tilde{\mathbf{x}}$  before the first symbol  $x^*$ . For there to be no overlap, it is sufficient to argue that

$$(m-2)N\mu + \nu + B\mu \leq (m-1)N\mu - \nu,$$

or, equivalently, that

$$N\mu \geq 2\nu + B\mu.$$

This holds for  $M$  large enough since we have  $N\mu \geq 3\nu$  by (3), and since  $\nu = \Theta(M)$  whereas  $B = \Theta(\log(M))$ .

Assume then that  $m' > m$ . The left-most element of any  $\tilde{\mathbf{x}}_\ell$  with  $\ell \in \mathcal{D}_{m'}$  is at position at least  $mN\mu + 1 - \nu$ . Conditioned on  $\mathcal{E}_3^c$ , there are at most  $(m-1)N\mu + \nu$  symbols 0 before the first symbol  $x^*$  in  $\tilde{\mathbf{x}}$ . Conditioned on  $\mathcal{E}_4^c$ , the burst of symbol  $x^*$  in  $\tilde{\mathbf{x}}$  is of length at most  $B\mu + \beta$ . For there to be no overlap, it is sufficient to argue that

$$(m-1)N\mu + \nu + B\mu + \beta < mN\mu + 1 - \nu,$$

or, equivalently, that

$$N\mu \geq 2\nu + B\mu + \beta.$$

This holds for  $M$  large enough by the same argument as in the last paragraph since  $\beta = \Theta(\log^{1/2}(M))$ . Together, this proves property 1.

To prove property 2, observe that, conditioned on  $\mathcal{E}_4^c$ , the burst of symbols  $x^*$  in  $\tilde{\mathbf{x}}$  is of length at least  $B\mu - \beta$ . Further, conditioned on  $\mathcal{E}_3^c$ , this burst must start at the receiver in the interval

$$\mathbb{N} \cap ((m-1)N\mu + 1 - \nu, (m-1)N\mu + 1 + \nu) = \mathcal{D}_m.$$

Since the subsequences  $\tilde{\mathbf{x}}_\ell$  have length  $\lfloor B\mu - \beta \rfloor$ , these two statements show that there exists at least one  $\ell \in \mathcal{D}_m$  such that  $\tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}$ .

The two properties allow us to analyze the events  $\mathcal{E}_{1,\ell}$  and  $\mathcal{E}_{2,\ell}$ . Recall that the hypothesis test on  $\mathbf{y}_\ell$  is performed under the assumption that either  $\tilde{\mathbf{x}}_\ell = \mathbf{0}$  or  $\tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}$ . Fix a threshold for the hypothesis test of  $\mathbf{y}_\ell$  such that the probability of missed detection satisfies

$$\mathbb{P}(H_\ell = H^0 \mid \tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}) \leq \varepsilon/4. \quad (7)$$

By Stein's lemma (see, e.g., [18, Theorem 12.8.1]), we then have that the probability of false alarm of the optimal test is upper bounded by

$$\begin{aligned} \mathbb{P}(H_\ell = H^1 \mid \tilde{\mathbf{x}}_\ell = \mathbf{0}) &\leq 2^{-\lfloor B\mu - \beta \rfloor D(W(\cdot|x^*)\|W(\cdot|0)) + o(B\mu - \beta)} \\ &= 2^{-B\mu D(W(\cdot|x^*)\|W(\cdot|0)) + o(\log(M))} \end{aligned} \quad (8)$$

as  $M \rightarrow \infty$ , and where we have used that  $B = \Theta(\log(M))$  and  $\beta = \Theta(\log^{1/2}(M))$ .

Consider then the value of  $\ell \in \mathcal{D}_m$  guaranteed by property 2. For this  $\ell$ , we have by (7),

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \mathbb{P}_m(\mathcal{E}_{1,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &= \mathbb{P}(H_\ell = H^0 \mid \tilde{\mathbf{x}}_\ell = x^* \cdot \mathbf{1}) \\ &\leq \varepsilon/4. \end{aligned} \quad (9)$$

By property 1 and (8), and using that  $|\mathcal{D}_1| \leq |\mathcal{D}_2| = |\mathcal{D}_3| = |\mathcal{D}_4| = \dots$ ,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}_m(\mathcal{E}_{2,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}(H_\ell = H^1 \mid \tilde{\mathbf{x}}_\ell = \mathbf{0}) \\ &\leq M |\mathcal{D}_2| 2^{-B\mu D(W(\cdot|x^*)\|W(\cdot|0)) + o(\log(M))}. \end{aligned}$$

Now,

$$|\mathcal{D}_2| \leq 2\nu + 1 \leq O(M)$$

as  $M \rightarrow \infty$ . Hence, using the definition of  $B$ ,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq 2^{2\log(M) - B\mu D(W(\cdot|x^*)||W(\cdot|0)) + o(\log(M))} \\ &\leq 2^{-\delta \log(M) + o(\log(M))} \\ &\leq \varepsilon/4 \end{aligned} \tag{10}$$

for  $M$  large enough.

Substituting (5), (6), (9), and (10) into (4) shows that for  $M$  large enough the probability of decoding error is upper bounded by  $\varepsilon$  for every message  $m$ . By (2), the achievable rate per unit cost of this scheme is

$$\begin{aligned} \hat{R} &= \frac{\log(M)}{P} \\ &= \frac{\log(M)}{Bc(x^*)} \\ &\geq \frac{\mu}{2 + \delta} \frac{D(W(\cdot|x^*)||W(\cdot|0))}{c(x^*)}. \end{aligned}$$

Since  $\delta > 0$  can be made arbitrarily small, this shows that

$$\hat{C} \geq \frac{\mu}{2} \frac{D(W(\cdot|x^*)||W(\cdot|0))}{c(x^*)}.$$

Taking the supremum over all  $x^* \in \mathcal{X} \setminus \{0\}$  completes the proof of the lower bound in Theorem 3.  $\blacksquare$

*Remark:* When using pulse-position modulation over a perfectly synchronized channel, the decoder knows exactly where the possible pulses are located, and thus needs to check only  $M$  possible pulse positions. However, in the proposed scheme for channels with synchronization errors, the codeword length is set to be of order  $\Theta(M^2)$  to combat synchronization errors. This translates into the number  $|\mathcal{D}_m|$  of possible pulse positions for message  $m$  being  $\Theta(M)$ . Thus, the sliding window-decoder needs to check  $M^2$  possible pulse positions due to the lack of synchronization. It is this increase from  $M$  to  $M^2$  hypothesis tests that results in the reduction of rate per unit cost by a factor 2 compared to the synchronized case.

## V. PROOF OF UPPER BOUND IN THEOREM 3

In this section, we provide an upper bound on the capacity per unit cost of the duplication/deletion/substitution channel. Since the DDC part of the channel is not memoryless, standard converse techniques are not applicable in this setting. Instead, we use a simulation argument, namely that the duplication/deletion/substitution channel can be simulated with an encoder and decoder communicating over a discrete memoryless channel. This DMC, in turn, can be analyzed and yields the desired upper bound for capacity per unit cost of the duplication/deletion/substitution channel. We now provide the details of this argument.

Let  $f^T$  and  $\{\varphi^\ell\}_{\ell=0}^\infty$  be an encoder-decoder pair achieving rate per unit cost  $\hat{C}(\mu, \sigma^2, W) - \delta$  and average probability of error at most  $\varepsilon$  for the duplication/deletion/substitution channel. Note that, since the output of the channel is of random length  $L$ , the decoder consists of several sub-decoders  $\{\varphi^\ell\}_{\ell=0}^\infty$ , one for each possible realization  $\ell$  of  $L$ .

We want to argue that the statistical behavior between the input  $m$  to the encoder  $f^T$  and the output  $\hat{m}$  of the decoder  $\varphi^L$  can be simulated over the discrete memoryless channel  $W$  alone (see Fig. 3). Consider a new encoder  $f'^T$  that consists of the concatenation of  $f^T$  with a DDC of same statistical behavior as the one in the original duplication/deletion/substitution channel. Denote by  $s'[\ell]$  the state random variables

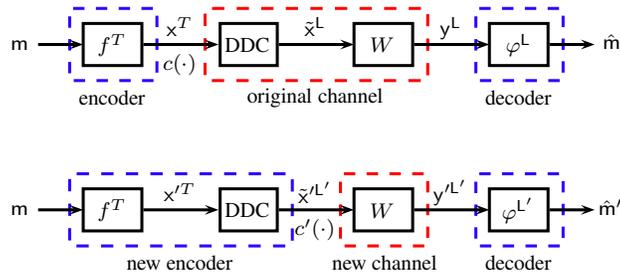


Fig. 3. The behavior of the original duplication/deletion/substitution channel (top figure) model can be simulated over the discrete memoryless channel  $W$  by modifying the encoder and the cost function (bottom figure).

describing this DDC. Observe that the encoder  $f^T$  is a randomized, variable-length encoder, mapping the message  $m$  into a random sequence  $\tilde{x}^{L'}$  of random length  $L'$ .

The output  $\tilde{x}^{L'}$  of the encoder  $f^T$  is transmitted over a DMC with the same transition probability matrix  $W$  as in the original duplication/deletion/substitution channel. Let  $y^{L'}$  be the output of this DMC. The decoder  $\varphi^{L'}$  for the DMC is equal to  $\varphi^L$ . Observe that this is a variable-length decoder, and denote by  $\hat{m}'$  its output.

By construction, for the same message  $m$ , the distributions of  $\hat{m}$  and  $\hat{m}'$  are identical. In particular, the average probability of error of both systems is the same. Hence, the average probability of error of  $f^T$  and  $\{\varphi^{\ell}\}_{\ell=0}^{\infty}$  over  $W$  is at most  $\varepsilon$ .

Define the new cost function

$$c'(\cdot) \triangleq \frac{1}{\mu} c(\cdot)$$

for the simulating DMC. With respect to this cost function, and assuming a uniformly distributed message  $m$ , the expected cost of using the variable-length encoder  $f^T$  over  $W$  is

$$\begin{aligned} \mathbb{E}c'(\tilde{x}^{L'}) &\stackrel{(a)}{=} \frac{1}{\mu} \mathbb{E} \left( \sum_{\ell=1}^{L'} c(\tilde{x}[\ell]) \right) \\ &= \frac{1}{\mu} \mathbb{E} \left( \sum_{t=1}^T s'[t] c(x'[t]) \right) \\ &\stackrel{(b)}{=} \frac{1}{\mu} \mathbb{E}(s'[1]) \sum_{t=1}^T \mathbb{E}c(x'[t]) \\ &\stackrel{(c)}{=} \frac{1}{\mu} \mathbb{E}(s[1]) \sum_{t=1}^T \mathbb{E}c(x[t]) \\ &= \mathbb{E}c(x^T). \end{aligned}$$

Here, (a) follows from the definition of  $c'(\cdot)$ ; (b) follows from the fact that the duplication/deletion process  $\{s'[t]\}$  is identically distributed and independent of the channel inputs; and (c) follows since  $s'[t]$  and  $s[t]$  have the same distribution, and since  $x'[t]$  and  $x[t]$  have the same distribution.

Hence, encoder  $f^T$  used over the DMC( $W$ ) has the same expected cost with respect to the cost function  $c'(\cdot)$  as the encoder  $f^T$  used over the duplication/deletion/substitution channel with respect to the cost function  $c(\cdot)$ . Observe that, while the two encoders have the same *expected* cost, the encoder  $f^T$  satisfies the stronger *per-codeword* cost constraint, whereas the encoder  $f^T$  does not.

The arguments in the last two paragraphs show that there exists a randomized variable-length encoder

$f'^T$  and variable-length decoder  $\{\varphi'^\ell\}_{\ell=0}^\infty$  achieving a rate per average unit cost of

$$\begin{aligned}\hat{R}' &= \frac{\log(M)}{\mathbb{E}c'(\tilde{x}'^L)} \\ &= \frac{\log(M)}{\mathbb{E}c(x^T)} \\ &\geq \hat{C}(\mu, \sigma^2, W) - \delta\end{aligned}$$

and average probability of error at most  $\varepsilon$ . Since this is just one possible coding scheme, as  $M \rightarrow \infty$ ,  $\hat{R}'$  must be upper bounded by  $\hat{C}'(W)$ , the capacity per unit cost of the DMC( $W$ ) subject to expected cost constraint, and allowing randomized variable-length codes. Thus, letting  $\delta \rightarrow 0$ ,

$$\hat{C}(\mu, \sigma^2, W) \leq \hat{C}'(W). \quad (11)$$

It remains to analyze  $\hat{C}'(W)$ . By [19, Exercise 6.28], we have that for DMCs under expected cost and with *deterministic* variable-length codes

$$\hat{C}'(W) = \max_{\tilde{x}'} \frac{I(\tilde{x}'; y)}{\mathbb{E}c'(\tilde{x}')}. \quad (12)$$

In Appendix A, we show that randomized encoders do not improve capacity per unit cost by arguing that for any randomized encoder, there exists a deterministic encoder that can achieve a not much higher error probability with asymptotically the same cost. This shows that (12) is valid for randomized encoders as well.

Note that up to this point we do not require the zero-cost symbol. Now, assuming there is one, we get by [5, Theorems 2 and 3] that

$$\begin{aligned}\max_{\tilde{x}'} \frac{I(\tilde{x}'; y)}{\mathbb{E}c'(\tilde{x}')} &= \sup_{\tilde{x}' \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|\tilde{x}') \| W(\cdot|0))}{c'(\tilde{x}')} \\ &= \mu \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x) \| W(\cdot|0))}{c(x)},\end{aligned}$$

so that

$$\hat{C}'(W) = \mu \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x) \| W(\cdot|0))}{c(x)}. \quad (13)$$

Combining (11) and (13) shows that

$$\hat{C}(\mu, \sigma^2, W) \leq \mu \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{D(W(\cdot|x) \| W(\cdot|0))}{c(x)},$$

completing the proof. ■

## VI. PROOF OF THEOREM 1

The upper bound in Theorem 1 follows from the upper bound for DMCs. Indeed, by Corollary 4,

$$\hat{C}(\mu, \sigma^2, \mathcal{N}(0, \eta^2)) \leq \mu \hat{C}(\mathcal{N}(0, \eta^2)) = \frac{\mu}{2\eta^2 \ln(2)},$$

yielding the desired upper bound.

For the lower bound, we adapt the achievable scheme for the DMC described in Section IV to the Gaussian case. For simplicity, we consider the case of noise power  $\eta^2 = 1$  and point out in the end how to extend the result for arbitrary values of  $\eta^2$ .

*Encoding:* Fix a target error probability  $\varepsilon \in (0, 1)$  and a number  $\delta \in (0, 1)$ . Let  $x^* = x^*(M) > 0$  be a nonzero channel input. Unlike the DMC case, we will choose  $x^*(M)$  as a function of  $M$  here. The codeword for message  $m \in \{1, 2, \dots, M\}$  is again given by

$$\mathbf{x}_m \triangleq (\mathbf{0}_{(m-1)N}, x^* \cdot \mathbf{1}_B, \mathbf{0}_{N-B}, \mathbf{0}_{(M-m)N}),$$

with the same

$$N \triangleq \lceil 36M\sigma^2/(\mu^2\varepsilon) \rceil = \Theta(M)$$

as before. However, here we choose the burst length  $B$  as

$$B \triangleq \lfloor \sqrt{MN\sigma^2/\mu^2} \rfloor = \Theta(M),$$

as opposed to  $\Theta(\log(M))$  in the DMC case. It can be verified that  $B \leq N$ , and hence the codewords are well defined. The block length of this code is  $T = MN$  and the cost of each codeword is

$$P = B(x^*)^2. \quad (14)$$

*Decoding:* Consider again the subsequences

$$\mathbf{y}_\ell \triangleq (y[\ell], y[\ell+1], \dots, y[\ell + \lfloor B\mu - \beta \rfloor - 1])$$

and

$$\tilde{\mathbf{x}}_\ell \triangleq (\tilde{x}[\ell], \tilde{x}[\ell+1], \dots, \tilde{x}[\ell + \lfloor B\mu - \beta \rfloor - 1])$$

of length  $\lfloor B\mu - \beta \rfloor$  for  $\ell \in \{1, 2, \dots, L - \lfloor B\mu - \beta \rfloor + 1\}$  with

$$\beta \triangleq \sqrt{4B\sigma^2/\varepsilon} = \Theta(\sqrt{M}).$$

Define the open intervals  $\tilde{\mathcal{D}}_m \triangleq ((m-1)N\mu + 1 - \nu, (m-1)N\mu + 1 + \nu)$  for  $m \in \{2, \dots, M\}$ . Define the decision regions

$$\mathcal{D}_m \triangleq \begin{cases} \{1\}, & \text{for } m = 1 \\ (\lfloor M/\log(M) \rfloor \cdot \mathbb{N}) \cap \tilde{\mathcal{D}}_m, & \text{for } m \in \{2, \dots, M\} \end{cases}$$

with

$$\nu \triangleq \sqrt{4MN\sigma^2/\varepsilon} = \Theta(M).$$

In words, the decision regions consist of  $\{1\}$  (for  $m = 1$ ) or of all multiples of  $\lfloor M/\log(M) \rfloor$  between  $(m-1)N\mu + 1 - \nu$  and  $(m-1)N\mu + 1 + \nu$  (for  $m > 1$ ). This differs from the DMC case, where the boundaries of the decision regions are the same, but there the regions contain every *integer* between them. Using the same arguments as in the DMC case shows that these decision regions are disjoint.

The receiver independently performs the hypothesis test

$$\frac{1}{\sqrt{\lfloor B\mu - \beta \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle \underset{H^0}{\overset{H^1}{\gtrless}} \sqrt{(2 + \delta) \ln(M)}$$

for each  $\ell \in \mathcal{D}_m$ ,  $m \in \{1, \dots, M\}$ , and where  $\langle \cdot, \cdot \rangle$  denotes the inner product. Let  $\hat{H}_\ell$  be the decision of the hypothesis test for  $\mathbf{y}_\ell$ . As in the DMC case, the receiver declares that message  $m$  was sent if  $\hat{H}_\ell = H^1$  for *some*  $\ell \in \mathcal{D}_m$  and  $\hat{H}_\ell = H^0$  for *all*  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ . If no such  $m$  exists, an error is declared.

*Error Analysis:* Assume that message  $m$  was sent. We define the same events as in the DMC case. Let  $\mathcal{E}_{1,\ell}$  be the event that  $\hat{H}_\ell = H^0$ , and  $\mathcal{E}_{2,\ell}$  be the event that  $\hat{H}_\ell = H^1$ . Define the missed-detection event

$$\mathcal{E}_1 \triangleq \bigcap_{\ell \in \mathcal{D}_m} \mathcal{E}_{1,\ell}$$

and the false-alarm event

$$\mathcal{E}_2 \triangleq \bigcup_{m' \neq m} \bigcup_{\ell \in \mathcal{D}_{m'}} \mathcal{E}_{2,\ell}.$$

The probability of decoding error for message  $m$  is then equal to  $\mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2)$ , where again  $\mathbb{P}_m$  denotes probability conditioned on message  $m$  being sent.

We again define the two auxiliary events describing the behavior of the DDC. Let  $\mathcal{E}_3$  be the event that the total number of symbols in  $\tilde{\mathbf{x}}$  resulting from the first  $(m-1)N$  transmitted symbols is outside  $((m-1)N\mu - \nu, (m-1)N\mu + \nu)$ , and let  $\mathcal{E}_4$  be the event that the number of symbols in  $\tilde{\mathbf{x}}$  resulting from symbols transmitted during time slots  $(m-1)N+1$  to  $(m-1)N+B$  is outside  $(B\mu - \beta, B\mu + \beta)$ . Using the same argument as for the DMC case, we can upper bound

$$\mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2) \leq \mathbb{P}_m(\mathcal{E}_3) + \mathbb{P}_m(\mathcal{E}_4) + \mathbb{P}_m(\mathcal{E}_1 | \mathcal{E}_3^c \cap \mathcal{E}_4^c) + \mathbb{P}_m(\mathcal{E}_2 | \mathcal{E}_3^c \cap \mathcal{E}_4^c). \quad (15)$$

Using Chebyshev's inequality as in the analysis of the DMC case, we obtain

$$\mathbb{P}_m(\mathcal{E}_3) \leq \varepsilon/4 \quad (16)$$

and

$$\mathbb{P}_m(\mathcal{E}_4) \leq \varepsilon/4. \quad (17)$$

We proceed with the analysis of  $\mathbb{P}_m(\mathcal{E}_1 | \mathcal{E}_3^c \cap \mathcal{E}_4^c)$  and  $\mathbb{P}_m(\mathcal{E}_2 | \mathcal{E}_3^c \cap \mathcal{E}_4^c)$ . Conditioned on message  $m$  being sent and on  $\mathcal{E}_3^c \cap \mathcal{E}_4^c$ , the elements in the decision regions satisfy the following two properties for  $M$  large enough (not depending on  $m$ ):

- 1) For every  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ , we have  $\langle \tilde{\mathbf{x}}_\ell, \mathbf{1} \rangle = 0$ . Hence,

$$\frac{1}{\sqrt{\lfloor B\mu - \beta \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle$$

is Gaussian with mean zero and variance one.

- 2) There exists at least one  $\ell \in \mathcal{D}_m$  such that

$$\langle \tilde{\mathbf{x}}_\ell, \mathbf{1} \rangle \geq x^* (\lfloor B\mu - \beta \rfloor - M/\log(M)).$$

Hence,

$$\frac{1}{\sqrt{\lfloor B\mu - \beta \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle$$

is Gaussian with mean at least

$$x^* \sqrt{\lfloor B\mu - \beta \rfloor} \left( 1 - \frac{M}{\lfloor B\mu - \beta \rfloor \log(M)} \right)$$

and variance one.

The first property follows by the same arguments as for the DMC case, using that  $B \leq \nu/(2\mu)$  and  $\beta = o(B(M))$  as  $M \rightarrow \infty$ . For the second property, note that by the arguments for the DMC case, there exists at least one  $\ell' \in \{1\}$  for  $m=1$  or  $\ell' \in \mathbb{N} \cap ((m-1)N\mu + 1 - \nu, (m-1)N\mu + 1 + \nu)$  for  $m > 1$  such that  $\tilde{\mathbf{x}}_{\ell'} = x^* \cdot \mathbf{1}$ . However, this value of  $\ell'$  may not be a multiple of  $\lfloor M/\log(M) \rfloor$ , and hence may not be an element of  $\mathcal{D}_m$ . Let  $\ell$  be the closest multiple of  $\lfloor M/\log(M) \rfloor$  to  $\ell'$  that is in  $\mathcal{D}_m$ ; such a  $\ell$  exists for  $M$  large enough since  $\nu = \Theta(M)$ . Since  $|\ell - \ell'| \leq M/\log(M)$ , this implies that

$$\begin{aligned} \langle \tilde{\mathbf{x}}_\ell, \mathbf{1} \rangle &\geq \langle \tilde{\mathbf{x}}_{\ell'}, \mathbf{1} \rangle - x^* M/\log(M) \\ &= x^* (\lfloor B\mu - \beta \rfloor - M/\log(M)), \end{aligned}$$

as required.

The two properties allow us to analyze the events  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . By property 1, and using that  $|\mathcal{D}_1| \leq |\mathcal{D}_2| = |\mathcal{D}_3| = |\mathcal{D}_4| = \dots$ ,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}_m(\mathcal{E}_{2,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}_m \left( \frac{1}{\sqrt{\lfloor B\mu - \beta \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle \geq \sqrt{(2 + \delta) \ln(M)} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c \right) \\ &\leq M |\mathcal{D}_2| Q(\sqrt{(2 + \delta) \ln(M)}). \end{aligned}$$

Using the Chernoff bound  $Q(a) \leq \exp(-a^2/2)$  for the  $Q$ -function, we have

$$Q(\sqrt{(2 + \delta) \ln(M)}) \leq M^{-(1+\delta/2)}.$$

Moreover, since  $\nu = \Theta(M)$ ,

$$|\mathcal{D}_2| \leq \frac{2\nu}{M/\log(M) - 1} + 1 \leq O(\log(M))$$

as  $M \rightarrow \infty$ . Hence,

$$\mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \leq O(M^{-\delta/2} \log(M)) \leq \varepsilon/4 \quad (18)$$

for  $M$  large enough.

Consider then the value of  $\ell \in \mathcal{D}_m$  guaranteed by property 2. For this  $\ell$ ,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \mathbb{P}_m(\mathcal{E}_{1,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \mathbb{P}_m \left( \frac{1}{\sqrt{\lfloor B\mu - \beta \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle \leq \sqrt{(2 + \delta) \ln(M)} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c \right) \\ &\leq Q \left( x^* \sqrt{\lfloor B\mu - \beta \rfloor} \left( 1 - \frac{M}{\lfloor B\mu - \beta \rfloor \log(M)} \right) - \sqrt{(2 + \delta) \ln(M)} \right). \end{aligned}$$

Recall that  $B = \Theta(M)$  and  $\beta = \Theta(\sqrt{M})$ , so that

$$\sqrt{\lfloor B\mu - \beta \rfloor} \left( 1 - \frac{M}{\lfloor B\mu - \beta \rfloor \log(M)} \right) = \sqrt{B\mu} (1 - o(1))$$

as  $M \rightarrow \infty$ . By choosing

$$x^* \triangleq (1 + \delta) \sqrt{(2 + \delta) \ln(M) / (B\mu)}, \quad (19)$$

we obtain

$$\mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \leq Q \left( (\delta - o(1)) \sqrt{(2 + \delta) \ln(M)} \right) \leq \varepsilon/4 \quad (20)$$

for  $M$  large enough.

Substituting (16), (17), (18), and (20) into (15) shows that for  $M$  large enough the probability of decoding error is upper bounded by  $\varepsilon$  for every message  $m$ . By (14) and (19), The power required by this scheme is

$$\begin{aligned} P &= B(x^*)^2 \\ &= (1 + \delta)^2 (2 + \delta) \ln(M) / \mu. \end{aligned}$$

Hence, the achievable rate per unit cost for this scheme is

$$\begin{aligned} \hat{R} &= \frac{\log(M)}{P} \\ &\geq \frac{\mu}{(1 + \delta)^2 (2 + \delta) \ln(2)}. \end{aligned}$$

Since  $\delta > 0$  can be made arbitrarily small, this shows that, for noise power  $\eta^2 = 1$ ,

$$\hat{C} \geq \frac{\mu}{2 \ln(2)}.$$

Assume then that  $\eta^2 \neq 1$ . By scaling the channel input at the transmitter by a factor  $\eta$  and the channel output at the receiver by a factor  $1/\eta$ , we can transform the channel to one with unit noise power. Since this increases the energy of the transmitted symbols by a factor  $\eta^2$ , but does not change the probability of error, this shows that

$$\hat{C} \geq \frac{\mu}{2\eta^2 \ln(2)},$$

concluding the proof. ■

## VII. PROOF OF THEOREM 2

This section adapts the coding scheme for the Gaussian duplication/deletion/substitution channel with known value of  $\mu$  described in Section VI to the compound setting with  $\mu$  only known to be in the range  $[\mu_1, \mu_2]$ . As before, we will first assume that the noise power is  $\eta^2 = 1$  and then generalize the result for arbitrary values of  $\eta^2$ .

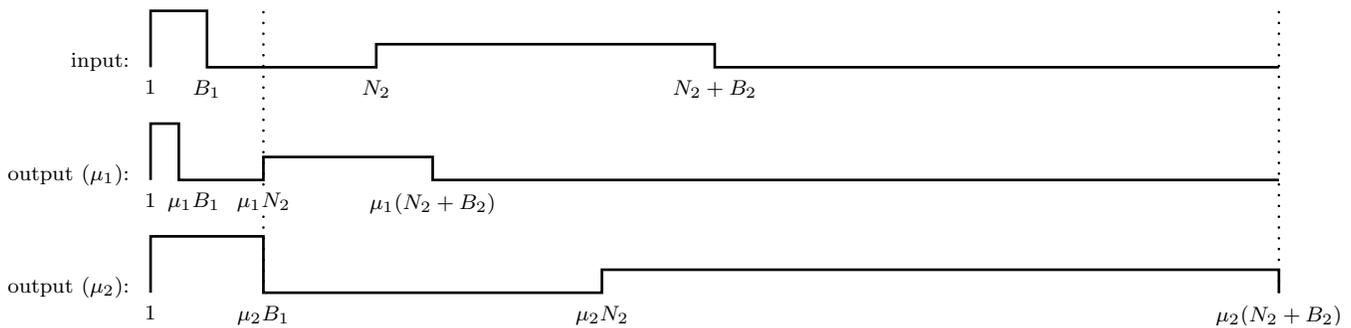


Fig. 4. Construction of input waveforms for the Gaussian compound duplication/deletion/substitution channel. In the figure,  $\mu_1 = 1/2$  and  $\mu_2 = 2$ . For simplicity,  $\delta$  is set to 0.  $N_2$  is chosen such that  $\mu_1 N_2 = \mu_2 B_1$ , ensuring that pulses corresponding to different messages (indicated by the dotted lines) are nonoverlapping at the receiver for all values of  $\mu \in [\mu_1, \mu_2]$  and under expected behavior of the DDC.

*Encoding:* Fix a target error probability  $\varepsilon \in (0, 1)$  and a number  $\delta \in (0, \mu_1)$ . Let  $x^* = x^*(M) > 0$  be a nonzero channel input. The codeword for message  $m \in \{1, 2, \dots, M\}$  is

$$\mathbf{x}_m \triangleq (\mathbf{0}_{N_m}, x^* B_m^{-1/2} \cdot \mathbf{1}_{B_m}, \mathbf{0}_{T-B_m-N_m}),$$

where

$$N_m \triangleq \begin{cases} 0, & \text{if } m = 1 \\ \lceil \frac{\mu_2 + \delta}{\mu_1 - \delta} (N_{m-1} + B_{m-1}) \rceil, & \text{if } m > 1 \end{cases}$$

and

$$B_m \triangleq \begin{cases} \lfloor \log(M) \rfloor, & \text{if } m = 1 \\ \lfloor (\mu_2 - \mu_1 + 2\delta) N_m \rfloor, & \text{if } m > 1. \end{cases}$$

Observe that here, unlike the case with known  $\mu$ , the value of the nonzero channel input is  $x^* B_m^{-1/2}$ , which depends on the message  $m$ . This construction is illustrated in Fig. 4. The block length of this code is  $T = N_M + B_M$ , and the cost of codeword  $m$  is

$$P = B_m (x^* B_m^{-1/2})^2 = (x^*)^2. \quad (21)$$

*Decoding:* Define the open intervals  $\tilde{\mathcal{D}}_m \triangleq ((\mu_1 - \delta)N_m + 1, (\mu_2 + \delta)N_m + 1)$  for  $m \in \{2, \dots, M\}$  and define the decision regions

$$\mathcal{D}_m \triangleq \begin{cases} \{1\}, & \text{for } m = 1 \\ (\lfloor N_m / \log(M) \rfloor \cdot \mathbb{N}) \cap \tilde{\mathcal{D}}_m, & \text{for } m \in \{2, \dots, M\}. \end{cases}$$

Note that, unlike the case with known value of  $\mu$ , the decision regions here have increasing width as a function of  $m$ . However, each decision region contains approximately the same number  $(\mu_2 - \mu_1 + 2\delta) \log(M)$  of points. It is easy to verify that the decision regions are disjoint.

For  $\ell \in \mathcal{D}_m$ , define the subsequences

$$\mathbf{y}_\ell \triangleq (y[\ell], y[\ell + 1], \dots, y[\ell + \lfloor (\mu_1 - \delta)B_m \rfloor - 1])$$

and

$$\tilde{\mathbf{x}}_\ell \triangleq (\tilde{x}[\ell], \tilde{x}[\ell + 1], \dots, \tilde{x}[\ell + \lfloor (\mu_1 - \delta)B_m \rfloor - 1])$$

of length  $\lfloor (\mu_1 - \delta)B_m \rfloor$ . We point out that here, unlike the case with known value of  $\mu$ , the subsequences in different regions  $\mathcal{D}_m$  and  $\mathcal{D}_{m'}$  have different lengths.

The receiver independently performs the hypothesis test

$$\frac{1}{\sqrt{\lfloor (\mu_1 - \delta)B_m \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle \underset{H^0}{\overset{H^1}{\geq}} \sqrt{(2 + \delta) \ln(M)}$$

for each  $\ell \in \mathcal{D}_m$ ,  $m \in \{1, \dots, M\}$ , and where  $\langle \cdot, \cdot \rangle$  denotes the inner product. Let  $\hat{H}_\ell$  be the decision of the hypothesis test for  $\mathbf{y}_\ell$ . As in the case of known  $\mu$ , the receiver declares that message  $m$  was sent if  $\hat{H}_\ell = H^1$  for some  $\ell \in \mathcal{D}_m$  and  $\hat{H}_\ell = H^0$  for all  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ . If no such  $m$  exists, an error is declared.

*Error Analysis:* Assume that message  $m$  was sent. We define the same error events as in the case of known  $\mu$ . Let  $\mathcal{E}_{1,\ell}$  be the event that  $\hat{H}_\ell = H^0$ , and  $\mathcal{E}_{2,\ell}$  be the event that  $\hat{H}_\ell = H^1$ . Define the missed-detection event

$$\mathcal{E}_1 \triangleq \bigcap_{\ell \in \mathcal{D}_m} \mathcal{E}_{1,\ell}$$

and the false-alarm event

$$\mathcal{E}_2 \triangleq \bigcup_{m' \neq m} \bigcup_{\ell \in \mathcal{D}_{m'}} \mathcal{E}_{2,\ell}.$$

The probability of decoding error for message  $m$  is then equal to  $\mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2)$ , where as before  $\mathbb{P}_m$  denotes probability conditioned on message  $m$  being sent.

We again define two auxiliary events describing the behavior of the DDC. Let  $\mathcal{E}_3$  be the event that the total number of symbols in  $\tilde{\mathbf{x}}$  resulting from the first  $N_m$  transmitted symbols is outside  $((\mu_1 - \delta)N_m, (\mu_2 + \delta)N_m)$ , and let  $\mathcal{E}_4$  be the event that the number of symbols in  $\tilde{\mathbf{x}}$  resulting from symbols transmitted during time slots  $N_m + 1$  to  $N_m + B_m$  is outside  $((\mu_1 - \delta)B_m, (\mu_2 + \delta)B_m)$ . We can again upper bound the probability of error as

$$\mathbb{P}_m(\mathcal{E}_1 \cup \mathcal{E}_2) \leq \mathbb{P}_m(\mathcal{E}_3) + \mathbb{P}_m(\mathcal{E}_4) + \mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) + \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c). \quad (22)$$

We start with the analysis of  $\mathbb{P}_m(\mathcal{E}_3)$  and  $\mathbb{P}_m(\mathcal{E}_4)$ . Using Chebyshev's inequality together with the upper bound  $\sigma^2$  on the variance of the states  $s[t]$ , we obtain similarly to the case with known value of  $\mu$

$$\mathbb{P}_m(\mathcal{E}_3) \leq \varepsilon/4 \quad (23)$$

and

$$\mathbb{P}_m(\mathcal{E}_4) \leq \varepsilon/4 \quad (24)$$

for  $M$  large enough and for any value of  $\mu \in [\mu_1, \mu_2]$ .

We proceed with the analysis of  $\mathbb{P}_m(\mathcal{E}_1 | \mathcal{E}_3^c \cap \mathcal{E}_4^c)$  and  $\mathbb{P}_m(\mathcal{E}_2 | \mathcal{E}_3^c \cap \mathcal{E}_4^c)$ . Conditioned on message  $m$  being sent and on  $\mathcal{E}_3^c \cap \mathcal{E}_4^c$ , the elements in the decision regions satisfy the following two properties for  $M$  large enough (not depending on  $m$ ):

1) For every  $\ell \in \mathcal{D}_{m'}$  with  $m' \neq m$ , we have  $\langle \tilde{\mathbf{x}}_\ell, \mathbf{1} \rangle = 0$ . Hence,

$$\frac{1}{\sqrt{\lfloor (\mu_1 - \delta) B_m \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle$$

is Gaussian with mean zero and variance one.

2) There exists at least one  $\ell \in \mathcal{D}_m$  such that

$$\langle \tilde{\mathbf{x}}_\ell, \mathbf{1} \rangle \geq x^* B_m^{-1/2} (\lfloor (\mu_1 - \delta) B_m \rfloor - N_m / \log(M)).$$

Hence,

$$\frac{1}{\sqrt{\lfloor (\mu_1 - \delta) B_m \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle$$

is Gaussian with mean at least

$$x^* \sqrt{(\mu_1 - \delta - 1/B_m)} \left( 1 - \frac{N_m}{\lfloor (\mu_1 - \delta) B_m \rfloor \log(M)} \right)$$

and variance one.

Property 2 can be proved using arguments analogous to the case with known value of  $\mu$ . For property 1, we need to argue that the burst of symbols  $x^*$  cannot be shifted into the incorrect decoding region.

Assume first  $m' < m$ . The right-most element of  $\mathcal{D}_{m'}$  is at position less than or equal to  $(\mu_2 + \delta)N_{m-1} + 1$ , and thus the right-most element of  $\tilde{\mathbf{x}}_\ell$  with  $\ell \in \mathcal{D}_{m'}$  is at position at most

$$(\mu_2 + \delta)N_{m-1} + (\mu_1 - \delta)B_{m-1} \leq (\mu_2 + \delta)(N_{m-1} + B_{m-1}).$$

Now, conditioned on  $\mathcal{E}_3^c$ , there are at least  $(\mu_1 - \delta)N_m$  symbols 0 in  $\tilde{\mathbf{x}}$  before the first symbol  $x^*$ . For there to be no overlap, it is sufficient to argue that

$$(\mu_2 + \delta)(N_{m-1} + B_{m-1}) \leq (\mu_1 - \delta)N_m$$

or, equivalently, that

$$N_m \geq \frac{\mu_2 + \delta}{\mu_1 - \delta} (N_{m-1} + B_{m-1}).$$

This holds by the definition of  $N_m$ .

Assume then that  $m' > m$ . The left-most element of any  $\tilde{\mathbf{x}}_\ell$  with  $\ell \in \mathcal{D}_{m'}$  is at position at least  $(\mu_1 - \delta)N_{m+1} + 1$ . Conditioned on  $\mathcal{E}_3^c$ , there are at most  $(\mu_2 + \delta)N_m$  symbols 0 before the first symbol  $x^*$  in  $\tilde{\mathbf{x}}$ . Conditioned on  $\mathcal{E}_4^c$ , the burst of symbol  $x^*$  in  $\tilde{\mathbf{x}}$  is of length at most  $(\mu_2 + \delta)B_m$ . For there to be no overlap, it is sufficient that

$$(\mu_2 + \delta)N_m + (\mu_2 + \delta)B_m \leq (\mu_1 - \delta)N_{m+1},$$

or, equivalently, that

$$N_{m+1} \geq \frac{\mu_2 + \delta}{\mu_1 - \delta} (N_m + B_m).$$

This holds again by the definition of  $N_{m+1}$ .

The two properties allow us to analyze the events  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . By property 1,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}_m(\mathcal{E}_{2,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \sum_{m' \neq m} \sum_{\ell \in \mathcal{D}_{m'}} \mathbb{P}_m \left( \frac{1}{\sqrt{\lfloor (\mu_1 - \delta) B_m \rfloor}} \langle \mathbf{y}_\ell, \mathbf{1} \rangle \geq \sqrt{(2 + \delta) \ln(M)} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c \right) \\ &\leq Q(\sqrt{(2 + \delta) \ln(M)}) \sum_{m'=1}^M |\mathcal{D}_{m'}|. \end{aligned}$$

Using the Chernoff bound  $Q(a) \leq \exp(-a^2/2)$  for the  $Q$ -function,

$$Q(\sqrt{(2 + \delta) \ln(M)}) \leq M^{-(1+\delta/2)}.$$

Moreover,

$$|\mathcal{D}_{m'}| \leq \frac{(\mu_2 - \mu_1 + 2\delta)N_m}{N_m / \log(M) - 1} + 1 \leq O(\log(M))$$

so that

$$\sum_{m'=1}^M |\mathcal{D}_{m'}| \leq O(M \log(M))$$

as  $M \rightarrow \infty$ . Hence,

$$\mathbb{P}_m(\mathcal{E}_2 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \leq O(M^{-\delta/2} \log(M)) \leq \varepsilon/4 \quad (25)$$

for  $M$  large enough.

Consider then the value of  $\ell \in \mathcal{D}_m$  guaranteed by property 2. For this  $\ell$ ,

$$\begin{aligned} \mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) &\leq \mathbb{P}_m(\mathcal{E}_{1,\ell} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \\ &\leq \mathbb{P}_m \left( \frac{\langle \mathbf{y}_\ell, \mathbf{1} \rangle}{\sqrt{\lfloor (\mu_1 - \delta) B_m \rfloor}} \leq \sqrt{(2 + \delta) \ln(M)} \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c \right) \\ &\leq Q \left( x^* \sqrt{(\mu_1 - \delta - 1/B_m)} \left( 1 - \frac{N_m}{\lfloor (\mu_1 - \delta) B_m \rfloor \log(M)} \right) - \sqrt{(2 + \delta) \ln(M)} \right). \end{aligned}$$

Note that

$$1 - \frac{N_m}{\lfloor (\mu_1 - \delta) B_m \rfloor \log(M)} = 1$$

for  $m = 1$ , and

$$1 - \frac{N_m}{\lfloor (\mu_1 - \delta) B_m \rfloor \log(M)} \geq 1 - \frac{1}{((\mu_1 - \delta)(\mu_2 - \mu_1 + 2\delta - 1/N_m) - 1/N_m) \log(M)} \geq 1 - o(1)$$

as  $M \rightarrow \infty$  for  $m > 1$ . Furthermore,

$$\sqrt{(\mu_1 - \delta - 1/B_m)} \geq \sqrt{(\mu_1 - \delta)}(1 - o(1))$$

as  $M \rightarrow \infty$ .

By choosing

$$x^* \triangleq (1 + \delta) \sqrt{(2 + \delta) \ln(M) / (\mu_1 - \delta)}, \quad (26)$$

we obtain

$$\mathbb{P}_m(\mathcal{E}_1 \mid \mathcal{E}_3^c \cap \mathcal{E}_4^c) \leq Q \left( (\delta - o(1)) \sqrt{(2 + \delta) \ln(M)} \right) \leq \varepsilon/4 \quad (27)$$

for  $M$  large enough.

Substituting (23), (24), (25), and (27) into (22) shows that for  $M$  large enough the probability of decoding error is upper bounded by  $\varepsilon$  for every message  $m$ . By (21) and (26), the power required by this scheme is

$$P = (x^*)^2 = (1 + \delta)^2(2 + \delta) \ln(M)/(\mu_1 - \delta).$$

Hence, the achievable rate per unit cost for this scheme is

$$\begin{aligned} \hat{R} &= \frac{\log(M)}{P} \\ &\geq \frac{\mu_1 - \delta}{(1 + \delta)^2(2 + \delta) \ln(2)}. \end{aligned}$$

Since  $\delta > 0$  can be made arbitrarily small, this shows that, for noise power  $\eta^2 = 1$ ,

$$\hat{C} \geq \frac{\mu_1}{2 \ln(2)}.$$

By scaling the input and output as in the proof of Theorem 1 in Section VI, this implies that

$$\hat{C} \geq \frac{\mu_1}{2\eta^2 \ln(2)}$$

for any value of noise power  $\eta^2$ , concluding the proof. ■

#### APPENDIX A

##### RANDOMIZATION DOES NOT INCREASE $\hat{C}'(W)$

In this appendix, we show that randomized encoders do not increase capacity per unit cost. We particularly argue that for any randomized encoder with probability of error  $\bar{\varepsilon}$  and cost  $\bar{c}'$ , there exists a deterministic encoder that has a probability of error upper-bounded by  $\sqrt{\bar{\varepsilon}}$ , with a cost no bigger than  $\bar{c}'/(1 - \sqrt{\bar{\varepsilon}})$ . Since  $\sqrt{\bar{\varepsilon}} \rightarrow 0$  and  $\bar{c}'/(1 - \sqrt{\bar{\varepsilon}}) \rightarrow \bar{c}'$  as  $\bar{\varepsilon} \rightarrow 0$ , this shows that randomized encoders and deterministic encoders achieve the same capacity per unit cost.

Consider a randomized encoder  $f'$ . We first point out that every realization  $f'_i$  of  $f'$  (corresponding to a deterministic encoder) has the same message rate  $\log(M)$ . Let  $\varepsilon(f'_i)$  and  $c'(f'_i)$  be the average (over the codebook) probability of error and average cost for a particular realization  $f'_i = f'$ , respectively. The overall probability of error and cost of this randomized encoder are then

$$\bar{\varepsilon} = \mathbb{E}(\varepsilon(f'_i))$$

and

$$\bar{c}' = \mathbb{E}(c'(f'_i)),$$

respectively. Now, by Carathéodory's theorem (see for example [20, Theorem 17.1]), any tuple  $(\bar{\varepsilon}, \bar{c}')$  achieved by a randomized encoder can be achieved by a convex combination of at most three deterministic encoders without loss of generality. Moreover, every point  $(\bar{\varepsilon}, \bar{c}')$  in the interior of the convex hull of these three points corresponding to the deterministic encoders is dominated by a point on a face of this convex hull. Therefore, in the following, we can assume without loss of generality that  $f'$  is a convex combination of two deterministic encoders  $f'_1$  and  $f'_2$  with weights  $0 \leq \lambda \leq 1$  and  $1 - \lambda$ , respectively. We thereby have

$$\bar{\varepsilon} = \lambda \varepsilon(f'_1) + (1 - \lambda) \varepsilon(f'_2) \tag{28}$$

and

$$\bar{c}' = \lambda c'(f'_1) + (1 - \lambda) c'(f'_2). \tag{29}$$

Now, if either  $f'_i$  for  $i \in \{1, 2\}$  satisfies  $\varepsilon(f'_i) \leq \bar{\varepsilon}$  and  $c'(f'_i) \leq \bar{c}'$ , then this  $f'_i$  is what we are looking for and we are done. So we can assume in the following that

- 1)  $\varepsilon(f'_1) \leq \bar{\varepsilon}$  and  $c'(f'_1) \geq \bar{c}'$ , and
- 2)  $\varepsilon(f'_2) \geq \bar{\varepsilon}$  and  $c'(f'_2) \leq \bar{c}'$ .

Assume that  $\varepsilon(f'_2) \geq \sqrt{\bar{\varepsilon}}$ ; otherwise,  $\varepsilon(f'_2) < \sqrt{\bar{\varepsilon}}$  and the deterministic encoder  $f'_2$  serves the purpose. Then, from (28), we have

$$\begin{aligned}\bar{\varepsilon} &\geq \lambda\varepsilon(f'_1) + (1 - \lambda)\sqrt{\bar{\varepsilon}} \\ &= \lambda\left(\varepsilon(f'_1) - \sqrt{\bar{\varepsilon}}\right) + \sqrt{\bar{\varepsilon}},\end{aligned}$$

and hence

$$\lambda \geq \frac{\sqrt{\bar{\varepsilon}} - \bar{\varepsilon}}{\sqrt{\bar{\varepsilon}} - \varepsilon(f'_1)}.$$

Plugging this bound into (29) results in

$$\begin{aligned}c'(f'_1) &= \frac{1}{\lambda}\left(\bar{c}' - (1 - \lambda)c'(f'_2)\right) \\ &\leq \bar{c}'/\lambda \\ &\leq \bar{c}'\frac{\sqrt{\bar{\varepsilon}} - \varepsilon(f'_1)}{\sqrt{\bar{\varepsilon}} - \bar{\varepsilon}} \\ &\leq \frac{\bar{c}'}{1 - \sqrt{\bar{\varepsilon}}}.\end{aligned}$$

Thus, the deterministic encoder  $f'_1$  satisfies  $\varepsilon(f'_1) \leq \bar{\varepsilon}$  and  $c'(f'_1) \leq \bar{c}'/(1 - \sqrt{\bar{\varepsilon}})$ , completing the proof. ■

## REFERENCES

- [1] J. G. Proakis, *Digital Communications*. McGraw-Hill, fourth ed., 2001.
- [2] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems Inform. Transm.*, vol. 3, no. 4, pp. 11–26, 1967.
- [3] M. J. E. Golay, "Note on the theoretical efficiency of information reception using PPM," *Proc. IRE*, vol. 37, p. 1031, Sept. 1949.
- [4] R. Gallager, "Energy limited channels: Coding, multiaccess, and spread spectrum," in *Proc. Conf. Inform. Sci. Syst.*, Mar. 1988.
- [5] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.
- [6] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1126–1237, Mar. 2006.
- [7] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4657–4660, Oct. 2006.
- [8] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for the deletion channel," in *Proc. IEEE ISIT*, pp. 1716–1720, June 2007.
- [9] E. Drinea and M. Mitzenmacher, "Improved lower bounds for the capacity of i.i.d. deletion and duplication channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2693–2714, Aug. 2007.
- [10] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, pp. 86–102, Jan. 2010.
- [11] D. Fertoni and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2753–2765, June 2010.
- [12] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. IEEE ISIT*, pp. 1002–1006, June 2010.
- [13] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probabilities," in *Proc. IEEE ISIT*, pp. 997–1001, June 2010.
- [14] J. Massey, "Optimum frame synchronization," *IEEE Trans. Commun.*, vol. 20, pp. 115–119, Apr. 1972.
- [15] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 59, pp. 1213–1226, Mar. 2013.
- [16] D. Bladsjö, M. Hogan, and S. Ruffini, "Synchronization aspects in LTE small cells," *IEEE Commun. Mag.*, vol. 51, pp. 70–77, Sept. 2013.
- [17] S. Ganeriwal, R. Kumar, and M. B. Srivastava, "Timing-sync protocol for sensor networks," in *Proc. ACM SenSys*, pp. 138–149, Nov. 2003.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, second ed., 2011.
- [20] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1997.