On the Energy-Delay Tradeoff in Streaming Data: Finite Blocklength Analysis

Mirza Uzair Baig, Lei Yu, Zixiang Xiong, *Fellow, IEEE*, Anders Høst-Madsen, *Fellow, IEEE*, Houqiang Li, *Senior Member, IEEE*, and Weiping Li, *Fellow, IEEE*

***Abstract*— This paper investigates basic trade-offs between energy and delay in wireless communication systems using finite blocklength theory. We first assume that data arrive in constant stream of bits, which are put into packets and transmitted over a communications link. Our results show that depending on exactly how energy is measured, in general energy depends on $\sqrt{d^{-1}}$ or $\sqrt{d^{-1}\log d}$, where d is the delay. This means that the energy decreases quite slowly with increasing delay. Furthermore, to approach the absolute minimum of -1.59 dB on energy, bandwidth has to increase very rapidly, much more than what is predicted by infinite blocklength theory. We then consider the scenario when data arrive stochastically in packets and can be queued. We devise a scheduling algorithm based on finite blocklength theory and develop bounds for the energy-delay performance. Our results again show that the energy decreases quite slowly with increasing delay.**

***Index Terms*— Wireless communications, delay, energy, finite blocklength, queuing, scheduling.**

I. Introduction

THE focus of this paper is to understand the relationship between delay and energy in wireless communications. With the proliferation of mobile devices, such as smart phones and tablet PCs, wireless communications are increasingly used to serve traffic with stringent delay constraints, such as video streaming, online gaming, VoIP, and video conferencing. Nowadays most of the internet traffic are video. Therefore, providing stringent delay guarantees becomes an important challenge for enhancing the quality of service (QoS) of end users.

Manuscript received April 12, 2017; revised October 25, 2019; accepted November 4, 2019. Date of publication November 19, 2019; date of current version February 14, 2020. This work was supported in part by the NSF Grant CCF-1216001, Grant CCF-1017823, Grant EECS-1546980, Grant EECS-1923751, and Grant EECS-1923803; in part by the Shenzhen Fundamental Research Fund under Grant KQTD2015033114415450 and Grant ZDSYS201707251409055; and in part by the Guangdong Province Grant 2017ZT07X152. This article was presented at the 2015 IEEE Information Theory Workshop, Jeju Island, South Korea, and at the 2015 53rd Annual Allerton Conference.

M. U. Baig and A. Høst-Madsen are with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA (e-mail: mub@hawaii.edu; ahm@hawaii.edu).

L. Yu, H. Li, and W. Li are with the Department of Electrical Engineering and Information Science, University of Science and Technology of China, Nanjing 210094, China (e-mail: yulei@ustc.edu.cn; lihq@ustc.edu.cn; wpli@ustc.edu.cn).

Z. Xiong is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zx@ece.tamu.edu).

Communicated by A. Guillen i Fabregas, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2019.2954347

On the other hand, energy consumption of communications is becoming an increasing focus under the banner of "green" communications (IEEEXplore returns thousands of hits on "green communications"). The main reason for this is the prevalence of mobile battery powered devices. Also, recently there has been an increased awareness on energy used in data centers [1], but 90% of energy consumption in cloud computing is actually in wireless access, not in data centers [2]. The two trends, delay sensitive communications and the desire for energy conservation, are basically conflicted. Hence it is important to understand the basic tradeoff between energy and delay.

There has been a large body of work on analyzing the trade-off between energy and delay using infinite-blocklength information theory. It mainly focuses on designing power allocation and scheduling algorithms with the objective of minimizing energy consumption under various types of constraints on the delay. The optimization problem has been considered under various system setups and channel conditions. The different types of delay constraints considered include: an average buffer delay constraint [3], average queuing delay constraint [4]–[6], a single hard delay constraint over M packets [7], [8], and individual hard delay constraint on each packet [9], [11], [12]. In [13], the constraint on delay is converted into one on the departure time of the packet; this approach can be used to model various QoS constraints. The algorithm design and system performance also depend on the availability of channel state information: time-variant and fading channels have been considered in [14]–[20].

One recent progress in information theory, namely the results on finite blocklength initiated by Polyanskiy, Poor and Verdú [21], has made analyzing delay more true to real world constraints. Previously, in order to use information theory to analyze delay, the packet size would have to approach infinity to use asymptotic results. This is of course a contradiction: with infinite packet size, the delay is infinite. While in some applications, the delay is so large that this gives reasonable insight, for smaller delay the result may not be accurate. With finite blocklength theory it is finally possible to precisely analyze, for example, what is the actual energy needed to meet a given delay constraint.

In terms of energy, energy per bit is the key quantity in low-power communications. Shannon [22] first demonstrated that for AWGN channels and any channel code, in the limit of the number of information bits $k \to \infty$, blocklength $n \to \infty$, error probability $\varepsilon \to 0$, and code rate $r \triangleq \frac{k}{n} \to 0$, the minimum

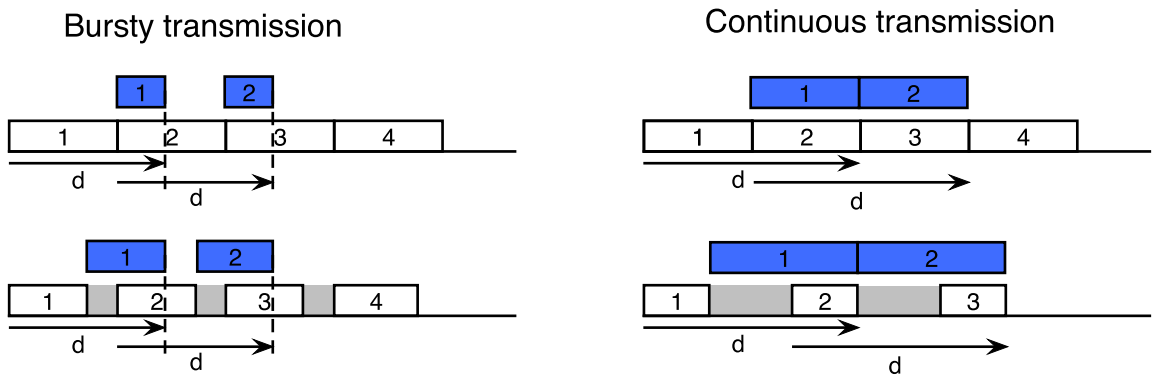

Fig. 1. Packet transmission modes. Blue are transmitted packets, grey dropped bits.

energy per bit converges to

$$(E_b/N_0)_{\min} = \ln 2 = -1.59\,\text{dB}, \tag{1}$$

where $\frac{N_0}{2}$ is the noise power spectral density. Verdú [23], [24] generalized [22] from AWGN channels to general memoryless channels. In the regime of fixed rate r and ε, non-asymptotic bounds on the minimum E_b for finite k have been studied in [21], [25], [26]. In the seminal work [21], Polyanskiy *et al.* also maximized the average throughput with ARQ for a given k and input power. That is equivalent to minimizing the average delay per bit. In [27], Polyanskiy *et al.* studied the minimum energy to transmit finite bits without delay constraint. Except for memoryless channels, energy-delay tradeoff for the communication over fading channels has been studied in [3], [29]–[33], and diversity-multiplexing-delay tradeoff or error probability-delay tradeoff for MIMO channels has been investigated in [34]–[36].

In this paper we study the energy-delay tradeoff taking into consideration finite blocklength. The paper consists of two parts. In part one we assume that bits arrive periodically in a steady stream with a maximum delay constraint. This allows us to obtain rigorous results based on finite blocklength theory. We show that depending on how energy is measured, in general energy depends on $\sqrt{d^{-1}}$ or $\sqrt{d^{-1}\log d}$, where d is the delay. As $d \to \infty$ both $\sqrt{d^{-1}}$ and $\sqrt{d^{-1}\log d}$ of course converge to zero, but the approach to zero is quite slow, e.g., compared to a linear convergence in d^{-1}. However, the type of bit arrival in part one is not common in real world systems. In part two of this paper, we allow packets to arrive at random times to a queue. This on one hand gives a more realistic picture of real-world systems; on the other hand, to analyze this complex system we can only use finite blocklength theory mainly as an approximation. Furthermore, in this case we consider average delay, as the optimum solution is found using Markov decision process theory, which becomes too complex with a maximum delay constraint. Our results again show that the energy decreases quite slowly with increasing delay.

A. Notation and Conventions

We use $\varepsilon(x)$ denote any function that satisfies $\lim_{x\to 0}\varepsilon(x) = 0$, while $o(x) = x\varepsilon(x)$.

Usually, the energy is specified in dB. In terms of asymptotic, $E_{b,dB} = 10\log_{10}(E_{b,\min} + \Delta E_b) = 10\log_{10}(E_{b,\min}) + \frac{10\Delta E_b}{E_{b,\min}\ln 10} + o(\Delta E_b)$. Since our interest is the behavior of excess energy when small, except for a proportionality constant, this is the same in absolute units and dB. We will therefore use absolute units in theoretical results as this makes equations simpler.

II. Periodic Bit Arrivals: Fundamental Theory

In this section we investigate this relationship between delay and energy for a basic problem. Consider an AWGN (additive white Gaussian noise) channel with symbol spacing T_c. An infinite stream $b[t], t = 1, 2, \ldots$ of bits arrive periodically at a transmitter with spacing T_s, i.e., arrival rate $\lambda = T_s^{-1}$; we let $R_a = \frac{T_c}{T_s} = \lambda T_c$ be the unit-less arrival rate. Equivalently, the bandwidth for transmission is $B = T_c^{-1} = R_a^{-1}T_s^{-1}$. The decoder needs to decode bit $b[t]$ no later than at time $(t + d)T_s$, where d is the (unit-less) delay. The energy critically depends on the two parameters R_a and d, and aim is to find the energy per bit required for the transmission, i.e., the function $E_b(d, R_a)$.

For a finite delay and finite energy, error-free transmission is not possible. The most natural setting for the problem we consider is clearly sequential decoding, which was studied by Fano in [44] with practical coding considered in [45]. However, in this work we will only consider block coding. There are a number of reasons for this: packet based transmission is used in most practical communications systems, practical block coding is more developed than sequential coding, and we can use the theory initiated with [21].

For packet transmission, the transmitter takes k bits from the input bit stream and packs them into a packet. This packet is then transmitted in n channel uses; all the bits must arrive at the receiver within dT_s seconds after the first bit in the packet arrived at the transmitter, see the top row in Fig. 1 (we assume zero transmission and decoding delay). We assume that a packet is either received without error or is lost with probability δ. This means that independent of the packet length k, the fraction of *bits* lost is δ. We consider δ as a fixed and given *constraint* in our system independent of R_a and d.

With packet transmission there are certain special ways energy can be saved. Usually we think of errors happening because of the random noise of the channel. However, instead the transmitter can decide not to transmit certain packet or bits [10]. To see why the former pays off, suppose that the

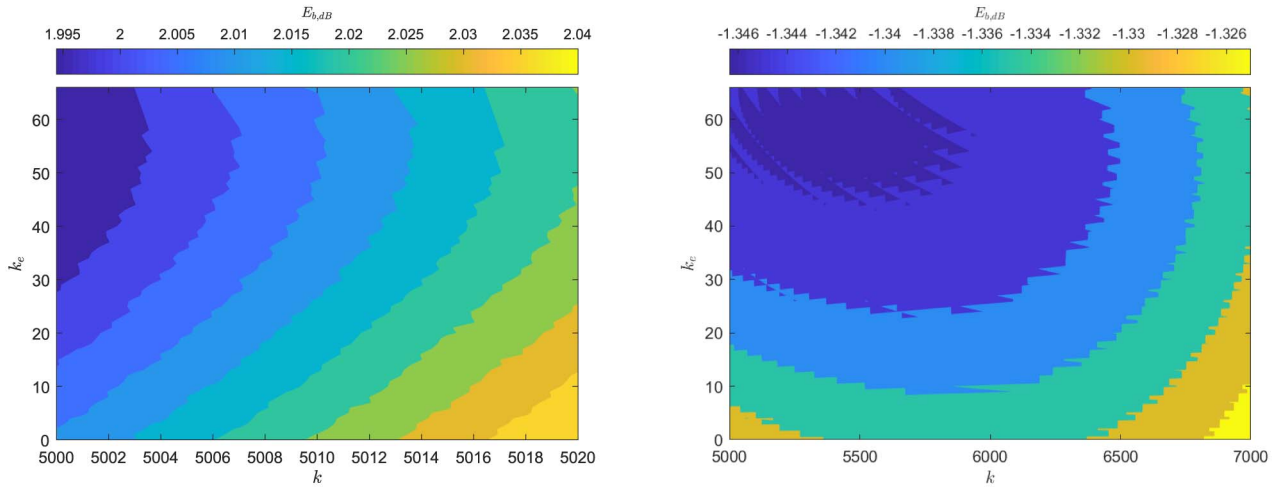

Fig. 2. Numerical solution of (9) for (a) *high rate* with delay $d = 10,000$, $R_a = 1$, $\delta = 10^{-2}$ and (b) *low rate* with delay $d = 10,000$, $R_a = 0.01$, $\delta = 10^{-2}$.

delay and packet size are large and R_a is small. Then a bit can be transmitted with an energy near $\ln 2$ (see (1)) with very small error probability. The transmitter now can decide to simply drop the last packets. If the fraction of dropped packets is $(1-\delta)$, the error probability now is δ, and the energy per bit is $(1-\delta)\ln 2 < \ln 2$; in terms of overall energy this can pay off. Of course, the transmitter can also randomly spread out the dropped packets for a more reasonable solution. Still, the solution is perhaps not so relevant from an application point of view, both because the idea of deleting packets might seem odd and because energy will be consumed in an non-uniform way. We can avoid this solution by using maximum energy per packet instead of average energy per packet. Thus, energy saved on one packet (say, by not transmitting it at all) cannot be used on other packets; that way energy is consumed at a constant rate. We will consider both average and maximum energy.

The transmitter can also decide to drop the last k_e bits in each packet. This will contribute to the error probability, but will allow longer time to transmit the packet. Consider Fig. 1 for the general relationship between parameters, which gives the following constraints

$$0 \le k_e \le \delta d \tag{2}$$

$$\frac{d-k_e}{2} \le k \le d \tag{3}$$

$$\epsilon = \frac{\delta - \frac{k_e}{d}}{1-\frac{k_e}{d}} \tag{4}$$

$$n = \frac{d-k}{R_a} \tag{5}$$

where ϵ is the required *packet error probability* to achieve a certain bit loss δ. If $k = \frac{d-k_e}{2}$ the transmission is continuous, that is the channel is kept active all the time; on the other hand, if $k > \frac{d-k_e}{2}$ transmission is bursty: the channel is idle in-between packets. If there are no dropped bits, the relationships simplify to

$$\frac{d}{2} \le k \le d; \quad n = \frac{d-k}{R_a} \tag{6}$$

The energy to transmit the packet is

$$\frac{E_b}{\frac{N_0}{2}} = \frac{nP}{k} \tag{7}$$

For simplicity we set $N_0 = 1$ so that

$$E_b = \frac{nP}{2k}. \tag{8}$$

For maximum energy transmissions per packet we can use

$$k = nC(P) - \sqrt{nV(P)}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1) \tag{9}$$

from [21], [46]. Here $Q(\cdot)$ is the Q-function,

$$V(P) = \frac{P}{2}\frac{P+2}{(P+1)^2}\log^2 e \tag{10}$$

is the channel dispersion, and

$$C(P) = \frac{1}{2}\log(1+P) \tag{11}$$

is the channel capacity. If we ignore the $O(1)$ term, we can solve this numerically with respect to P, and then use (8) to calculate the energy per bit. A few such solutions can be seen in Fig. 2. It can be seen that it seems to pay off to have $k_e > 0$, and for small R_a it seems that bursty transmission ($k > \frac{d}{2}$) pays off.

When bits arrive periodically, instead of numerical solution as above, it is possible to derive analytical solutions in a rigorous way, without ignoring the $O(1)$ term. We will, as discussed above, consider both average and maximum energy. However, we will disregard dropped bits, that is, use $k_e = 0$ as the idea that bits are dropped in a periodic and predictable fashion is hardly a reasonable communications scheme (and the dropped bits always *have* to be at the end of a packet). An additional reason is that allowing $k_e > 0$ requires optimization over ϵ

as $d \to \infty$, and finite blocklength theory assumes a fixed ϵ, and that d has to be very large for $k_e > 0$ to pay off. As a consequence, $\epsilon = \delta$, and we will therefore use δ as the packet error probability *in this section*.

A. Formal Problem Statement

As mentioned at the beginning of the section, we consider an infinite stream $b[t], t = 1, 2, \ldots$ of bits, where the decoder needs to decode bit $b[t]$ no later than at time $(t + d)T_s$. The coder divides the bit stream into blocks of size k bits, each of which forms a message $W \in \left\{1, 2, \ldots, 2^k\right\}$ transmitted in n channel uses over an AWGN using a (possibly randomized) coder $f : \left\{1, 2, \ldots, 2^k\right\} \to \mathbb{R}^n$; no simultaneous transmission is allowed, resulting in the relationship (6) between d, k, n. The decoder is a function on the received signal y^n in each block, $g : \mathbb{R}^n \to \left\{0, 1, 2, \ldots, 2^k\right\}$ (0 denotes error). If $\hat{W} = g(y^n) \neq W$ an error is declared for that block. The error probability is constrained to δ.

Let the codebook be $\mathbf{c}_j \in \mathbb{R}^n$, $j \in \left\{1, 2, \ldots, 2^k\right\}$. We consider two possible constraints on energy. For a *maximum energy constraint* we require

$$\forall j \in \left\{1, 2, \ldots, 2^k\right\} : \frac{\|\mathbf{c}_j\|_{\ell_2}^2}{2k} \leq E_b \tag{12}$$

while for an *average energy constraint*

$$2^{-k} \sum_{j=1}^{2^k} \frac{\|\mathbf{c}_j\|_{\ell_2}^2}{2k} \leq E_b \tag{13}$$

The goal is to find the relationship between E_b, d, R_a, and δ.

B. Fixed Arrival Rate R_a

We first consider the problem when we fix R_a and let $d \to \infty$. The solution in principle is simply to solve (9) with respect to P using series expansion and then using (8). What makes the proof not quite straightforward is that we have to 1) deal rigorously with the $O(1)$ term and 2) optimize over k. The solution is given by

Theorem 1. *For fixed R_a and maximum energy constraint, the energy per bit is*

$$E_b(d, \delta) = \frac{2^{2R_a} - 1}{2R_a} + \frac{2^{2R_a}\sqrt{2V\left(2^{2R_a} - 1\right)}}{\sqrt{R_a}\log e} Q^{-1}(\delta)\sqrt{d^{-1}} + o\left(\sqrt{d^{-1}}\right) \tag{14}$$

This limit can be achieved by continuous transmission i.e., with $k = \frac{d}{2}$.

For an average energy constraint, instead of (8) we use [47, (19)] with $r = \frac{k}{n}$ (with no dropped bits, $\epsilon = \delta$)

$$r = C\left(\frac{P}{1-\epsilon}\right) - \sqrt{V\left(\frac{P}{1-\epsilon}\right)}\sqrt{\frac{\log n}{n}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{15}$$

Theorem 2. *For fixed R_a and average energy constraint, the energy per bit is*

$$\frac{E_b(d, \delta)}{1-\delta} = \frac{2^{2R_a} - 1}{2R_a} + \frac{2^{2R_a}\sqrt{2V\left(2^{2R_a} - 1\right)}}{\sqrt{R_a}\log e}\sqrt{\frac{\log d}{d}} + o\left(\sqrt{\frac{\log d}{d}}\right) \tag{16}$$

This limit can be achieved by continuous transmission i.e., with $k = \frac{d}{2}$.

The proofs are in Appendix A.

C. Variable Arrival Rate R_a

It can be noticed that in the results in Section II-B the absolute minimum energy of $\ln 2$ from (1) (or $(1-\delta)\ln 2$ for average energy) is not achieved even if $d \to \infty$. To approach the theoretical minimum energy per bit, -1.59dB, we need to let $R_a \to 0$, i.e., let the bandwidth $B \to \infty$. That is, we have an infinite (or very large) bandwidth available for transmission. First, if we keep d fixed, then when $R_a = 0$ identical (i.e., exact infinite bandwidth is available) it is clear that letting $k = d$ is optimum (since the time to transmit the bits is zero), that is, extremely bursty transmission is optimum. In this case we have the following result

Proposition 3. *The minimum energy per bit for $R_a = 0$ for maximum energy constraint is given by*

$$E_b(d, \delta) = \ln 2 + Q^{-1}(\delta)\sqrt{2\ln 2}\sqrt{d^{-1}} - \frac{1}{2}\frac{\ln d}{d} + O\left(d^{-1}\right) \tag{17}$$

This is just a restatement of the results in [27] in terms of energy per bit, and the proof is simply doing suitable series expansion so we will omit it.

For average energy constraint we have the following generalization of [27]

Theorem 4. *With average energy constraint, for given δ the packet size k is given by*

$$k = \frac{E}{1-\delta}\log e - \sqrt{\frac{2E}{1-\delta}\ln\left(\frac{E}{4\pi(1-\delta)}\right)}\left(1 + \ln^{-1}\left(\frac{E}{4\pi(1-\delta)}\right)\right)\log e + \log(E) + O(1). \tag{18}$$

for E sufficiently large.

The proof is in Appendix C.

Corollary 5. *For average energy constraint, the minimum energy per bit is given by*

$$\frac{E_b(d, \delta)}{1-\delta} = \ln 2 + \sqrt{2\ln 2 d^{-1}\ln d} + o\left(\sqrt{d^{-1}\ln d}\right). \tag{19}$$

This result is straightforward to prove using appropriate series expansions in Theorem 4 and setting $k = d$ and the proof is hence omitted.

What we are interested in is how the limit (1) is approached as $R_a \to 0$ but $R_a > 0$. This is tricky to state rigorously, as we are dealing with two simultaneous limits, $R_a \to 0$ as in [24] and $d \to \infty$ as in [21]. Clearly, the results depend on how R_a and d jointly approach their limit. For example, if R_a is fixed while $d \to \infty$, we simply get the results of Section II-B; by implication, if $d \to \infty$ while R_a approaches zero very, very slowly, we should get a result very similar to this. On the other hand, if $R_a \to 0$ while d approaches infinity very, very slowly, we should get something like Proposition 3. One way to specify this rigorously is to consider R_a as function of d, $R_a(d)$. The results now depends on how $R_a(d) \to 0$ as $d \to \infty$, and are given by the following Theorem.

Theorem 6. *Depending on how R_a behaves as a function of d, we get different behavior of E_b for a maximum energy constraint.*

*1) **Non-convergence**. If $\liminf_d R_a(d) > 0$, E_b is bounded away from -1.59dB.*

*2) **Continuous transmission**. If $O\Big(d\exp(-\sqrt{d\ln 2}Q^{-1}(\delta))\Big) < R_a < o(\sqrt{d^{-1}})$ continuous transmission is optimum, and the energy is given by*

$$E_b(d,\delta) = \ln 2 + \sqrt{2}\sqrt{2\ln 2}Q^{-1}(\delta)\sqrt{d^{-1}} + o(\sqrt{d^{-1}}) \quad (20)$$

*3) **Extremely bursty transmission**. If $R_a < o(d)\exp(-\sqrt{2d\ln 2}Q^{-1}(\delta))$, the energy is given by*

$$E_b(d,\delta) = \ln 2 + \sqrt{2\ln 2}Q^{-1}(\delta)\sqrt{d^{-1}} + o(\sqrt{d^{-1}}) \quad (21)$$

The proof is in Appendix B. In light of previous results, and in particular Corollary 5, one can conjecture that similar results for average energy constraint with $\sqrt{d^{-1}}$ replaced with $\sqrt{\frac{\log d}{d}}$ holds. However, it turns out that the bounds in [47] are too weak to prove this. It might be possible to prove a stronger version of Theorem 6 with the theory in [38].

The Theorem shows that for extremely bursty transmission, the absolute lower bound of Proposition 3 is achievable. Of course this was in some sense already known from Proposition 3, but only when $R_a = 0$ identical. What Theorem 6 tells is how fast R_a has to decrease with d to achieve this lower bound.

We notice that the only difference between continuous transmission and extremely bursty transmission is a factor $\sqrt{2}$ on the $\sqrt{d^{-1}}$ term, a very small difference. But to get that slight additional gain, the bandwidth has to be enormous. For continuous transmission we only require $R_a \sim \sqrt{d^{-1}}$ (slightly smaller, to be precise). Since $\sqrt{d^{-1}} \sim \Delta E_b$ that means the bandwidth B is essentially proportional to ΔE_b^{-1}, just as in [24]. On the other hand, for extremely bursty transmission we essentially require B proportional to $\exp(\Delta E_b^{-1})$, an exploding bandwidth.

D. Numerical Results

As we have already mentioned, there are two ways to deal with equation (9). We can treat it as an exact expression, as we have done in Theorem 3, or we can treat as an approximation. It can be of interest to compare those two approaches.

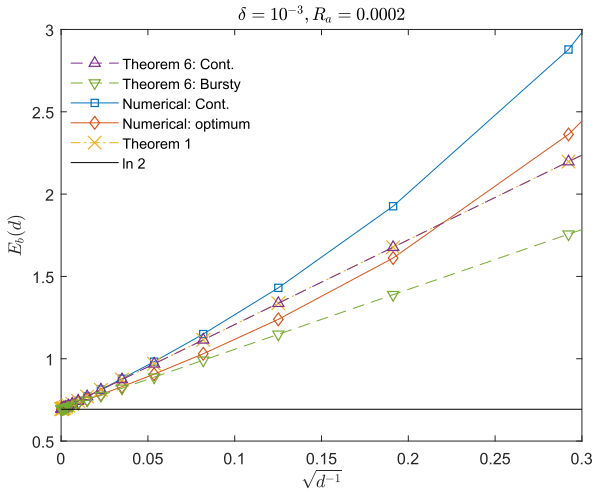

Fig. 3. The plot shows the first two terms in the energy delay relationships in Theorem 6, straight lines. This is compared with solving (9) numerically, either fixing $k = \frac{d}{2}$ ("Numerical: Cont.") or optimizing over k ("Numerical: optimum").

To use equation (9) as an approximation we ignore the $O(1)$ term, or more precisely we put $O(1) = 0$. We fix R_a and δ and next choose k subject to the condition (6), which also gives n. Finally, we numerically solve equation (9) with respect to P, which is the only unknown. We can now find the corresponding E_b from (8). We take two different approaches

- Continuous transmission (case 2 in Theorem 6): in this case put $k = \frac{d}{2}$.
- Optimum transmission (case 3 in Theorem 6): In this case we (numerically) optimize k subject to $\frac{d}{2} \le k \le d$.

The results are given in Fig. 3.

The plot shows that for large d the theoretical results and numerical results agree. Specifically, the theoretical results predict a linear relationship in $\sqrt{d^{-1}}$, and also give the slopes, and the numerical results converge to this. In particular, it confirms one conclusion from Theorem 6: for energies near the minimum of -1.59dB, bursty transmission is better than continuous transmission; however, the gain is only a minor improvement in slope. The cost is that bursty transmission uses much more bandwidth than continuous transmission.

III. Packet Arrivals: Queuing System

In this part of the paper we extend the simple model in Section II with attributes of a more realistic communication system. The network communication system we consider is shown in Fig. 4.

Rather than the source emitting a constant stream of bits at in Section II, the bits now arrives in packets (messages) at random epochs. The messages with each b bits, generated from a source node, enter a buffer (with infinite size) at the encoding node to form a queue. It is assumed that the message arrivals follow a Poisson process with *message* arrival rate λ_{msg} – the bit arrival rate is $\lambda = b\lambda_{msg}$, and therefore $R_a = \lambda T_c = b\lambda_{msg}T_c$. According to the current queue length $q \in \mathbb{N} \triangleq \{0, 1, 2, \cdots\}$, the encoder adaptively chooses the number of bits k (an integer multiple of the message size b), coding blocklength n, power P, and maximum number of

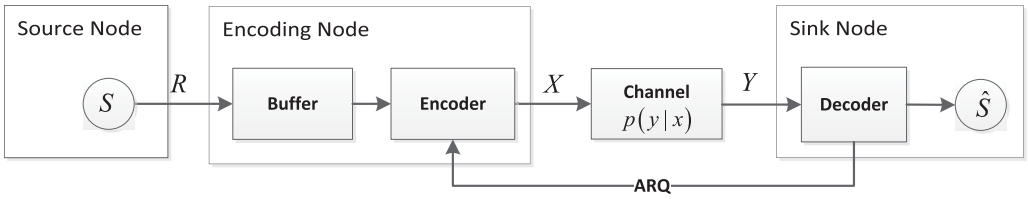

Fig. 4. The lossy network communication system considered in this work.

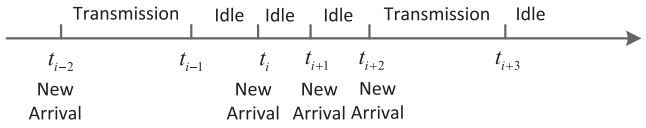

Fig. 5. Coding epochs.

transmissions L to encode the first k/b messages in the buffer into codeword $\mathbf{x}_j \in \mathbb{R}^n, 1 \leq j \leq 2^k$, with $2^{-k}\sum_{j=1}^{2^k} ||\mathbf{x}_j||^2 \leq nP$, corresponding to an average energy constraint, Section II-A. This means that the optimal action (k, n, P, L) at each encoding epoch is a vector-valued function of the current queue length q, i.e., $(k, n, P, L) = \mathbf{d}(q)$, where the function $\mathbf{d} \in \mathbb{D}$ is called the decision rule that prescribes actions for each state of the queue and $\mathbb{D}$ denotes the set of all possible decision rules. Similar to the feedback scenario considered in [27], we assume that the decoder can feedback an ARQ message to the encoder to inform whether the transmitted messages have been decoded correctly[1]. If not, the encoder retransmits the coded messages. Here the coding process is assumed to be instantaneous, i.e., without delay, and each transmission is assumed to be independent, i.e., for one block data, the current transmission does not make use of the previous unsuccessful ones. The main objective in this section is to find the decision rule $\mathbf{d}(q)$ to minimize overall energy consumption, as well as analyzing the delay-energy tradeoff for the system in Fig. 4.

We assume that during the transmission period of one block, the encoder can not code or transmit another block. Hence, as shown in Fig. 5, we set the arrival epochs (outside of the transmission periods) and transmission stop epochs to coding epochs. It is worth noting that this assumption is restrictive, e.g. it does not allow the transmitter to add additional data to a codeword before retransmitting it. Let $\mathbb{T} = \{t_i, i = 1, 2, \cdots\}$ denote the set of coding epochs. For any epoch $t_i \in \mathbb{T}$, set the system state to q_i, where q_i is the queue length (not including the messages in transmission) at epoch t_i. A policy $\pi = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_I)$ is a sequence of coding decisions $\mathbf{d}_i \in \mathbb{D}$ at different epochs. Let $(k_i, n_i, P_i, L_i) = \mathbf{d}_i(q_i)$, resulting in a certain packet error probability ϵ_i. We let $k_i = 0$ mean that the encoder does not code (or is in an idle state) at the coding epoch, and $k_i > 0$, $n_i = P_i = L_i = 0$ indicate that the encoder drops k_i bits at the coding epoch, i.e., $\epsilon_i = 1$. Furthermore, a scheme without ARQ corresponds to the case that $L_i = 1$ whenever $k_i > 0$. As opposed to the bit streaming scenario in Section II, it is not sensible to exclude bit dropping, as they are random in nature, and therefore similar to random errors; we therefore do not have $\epsilon = \delta$. Furthermore, the packet error probability is variable as opposed to the fixed error probability for streaming bits, which again makes more sense in this model. We also denote the coding rate as $r_i \triangleq \frac{k_i}{n_i}$ if $n_i > 0$. For a Poisson arrival queue, the states q_i form an embedded Markov chain, and the resulting decision process becomes an embedded Markov decision process (MDP) [37].

For a deterministic Markovian (possibly time-sharing) policy π, the message delivery success rate is given by

$$p_s^\pi = \frac{T_c}{R_a t_I}\mathbb{E}^\pi[\sum_{i=1}^{I} k_i(1-\epsilon_i^{L_i})], \tag{22}$$

where the expectation $\mathbb{E}^\pi$ is taken over the distribution of all queue length $q_i, 1 \leq i \leq I$, and ϵ_i is the error probability for each transmission of the ith block.

Little's law [39] states that the long-term average number of messages in a stable system N^π is equal to the arrival rate λ_{msg}, multiplied by the average time a message spends in the system (average delay), d^π. That is[2]

$$d^\pi = \frac{N^\pi}{\lambda_{msg} T_s}. \tag{23}$$

Hence to compute average delay, we only need to compute average number of messages in the system N^π.

From coding epoch i to coding epoch $i+1$, the average number of messages in the system (given the number of transmissions l_i) is

$$\overline{q}_i = \begin{cases} q_i, & \text{if } k_i = 0; \\ q_i + \frac{1}{2}\lambda_{msg} l_i n_i T_c, & \text{if } k_i > 0. \end{cases} \tag{24}$$

Hence N^π can be computed as

$$N^\pi = \frac{1}{t_I}\mathbb{E}^\pi[\sum_{i=1}^{I}\mathbb{E}_{l_i,\Delta t_i}\Delta t_i \overline{q}_i] \tag{25}$$

$$= \frac{1}{t_I}\mathbb{E}^\pi[\sum_{i=1}^{I}\overline{Q}_i], \tag{26}$$

[1]This assumption can be realized in the communication system with ideal feedback, where according to receiver's feedback the sender determines whether the transmitted messages have been decoded correctly, and then sends the result to the receiver by powerful channel codes (we assume that the transmission of the result is error-free and the overhead incurred is negligible, since only one bit is generated for each coding block); besides, there are also some practical error detection strategies, see [40] and [41].

[2]As in the previous section, here we also normalize the delay through dividing it by T_s. Furthermore, when it is allowed for the transmitter to drop messages, the waiting delays of dropped messages are also taken into account in the delay of the system. Ideally, we should define the delay of the system as the average delay of the messages that are correctly decoded by the receiver. However, this is difficult to compute. Hence we consider the average delay of all the messages emitted from the source, which is much simpler to compute. Furthermore, it does not seem reasonable to define the delay of the system as the average delay of the messages that are delivered to the receiver. This is because the dropped messages can be seen as being transmitted to the receiver by a special channel code (with $n = 0, P = 0$).

where similar to (22) the expectation $\mathbb{E}^{\pi}$ is also taken over the distribution of all queue length $q_i, 1 \le i \le I$, $\Delta t_i := t_{i+1} - t_i$ denotes the length of time from epoch i to epoch $i+1$, and

$$\overline{Q}_i := \mathbb{E}_{l_i,\Delta t_i} \Delta t_i \overline{q}_i = \begin{cases} \frac{q_i}{\lambda_{msg}}, & \text{if } k_i = 0; \\ q_i \bar{l}_i n_i T_c + \frac{1}{2}\lambda_{msg} \overline{l_i^2} n_i^2 T_c^2, & \text{if } k_i > 0, \end{cases} \tag{27}$$

with $\bar{l}_i = \mathbb{E}[l_i]$ and $\overline{l_i^2} = \mathbb{E}\left[l_i^2\right]$ denoting the expectation and second moment of number of transmissions for the ith block. Since the number of transmissions l_i follows the following distribution:

$$\mathbb{P}(l_i = j) = \begin{cases} (1-\epsilon_i)\epsilon_i^{j-1}, & \text{if } j \le L_i - 1; \\ \epsilon_i^{L_i - 1}, & \text{if } j = L_i, \end{cases} \tag{28}$$

we have $\bar{l}_i = \frac{1-\epsilon_i^{L_i}}{1-\epsilon_i}$ and $\overline{l_i^2} = \frac{2\left(1-\epsilon_i^{L_i}\right)}{1-\epsilon_i} - 1 - (2L_i - 1)\epsilon_i^{L_i}$. The decoding error probability ϵ_i for ith block satisfies (15). Ignoring the $O(\sqrt{n})$ term gives

$$k_i = n_i C\left(\frac{P_i}{1-\epsilon_i}\right) - \sqrt{V\left(\frac{P_i}{1-\epsilon_i}\right) n_i \log n_i}. \tag{29}$$

In this section of the paper, our focus is on deriving an optimum decision rule $\mathbf{d}(q)$ or policy π, and to do so we need a concrete relationship between (k_i, n_i, P_i) – O terms are not useable. The resulting policy therefore optimizes for the approximate energy. It is still a valid policy for the actual system, and while it might not minimize actual energy, it is almost certainly better than a policy based on infinite blocklength theory. Furthermore, as seen from the results in Section II (particularly Fig. 3), ignoring the O terms gives an approximation that is accurate in an asymptotic sense.

For policy π, the average power (per transmitted symbol)

$$P^{\pi} = \frac{T_c}{t_I}\mathbb{E}^{\pi}[\sum_{i=1}^{I} \bar{P}_i], \tag{30}$$

where

$$\bar{P}_i = \begin{cases} 0, & \text{if } k_i = 0; \\ n_i P_i \bar{l}_i, & \text{if } k_i > 0, \end{cases} \tag{31}$$

denotes the expected total energy consumed for the ith block.

We now proceed to define the power-delay-error function $P(d,\delta)$ as the minimum power consumed for the system under delay d and average message error rate δ (including loss and decoding error), which can be expressed as

$$\begin{aligned} P(d,\delta) \quad = \quad & \limsup_{I\to\infty} \min P^{\pi} && (32) \\ \text{over} \quad & \pi = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_I) \\ \text{subject to} \quad & k_i(q) \le bq, \forall i, q \in \mathbb{N} && (33) \\ & d^{\pi} \le d && (34) \\ & p_s^{\pi} \ge 1 - \delta. && (35) \end{aligned}$$

For any policy π, since the power P^{π} and energy per bit E_b^{π} are related by

$$\frac{E_b^{\pi}}{\frac{N_0}{2}} = \frac{P^{\pi} t_I / T_c}{b\lambda_{msg} t_I} = \frac{P^{\pi}}{R_a}, \tag{36}$$

we similarly define the energy-delay-error function as

$$E_b(d, \delta, R_a) = \frac{P(d, \delta, R_a)}{2R_a}. \tag{37}$$

Here we have explicitly noted the dependency on R_a of $P(d,\delta)$ in (32) to clarify that R_a (and λ_{msg}) is more than a proportionality factor. Notice that R_a is fixed as in Section II-B.

A. The Optimal Power-Delay Tradeoff

We first convert the power-delay tradeoff problem into a MDP problem subject to an expected cost constraint, and then give the optimality equation of the resulting MDP problem. Finally, we compute the optimal coding strategy and the corresponding power-delay function by a policy iteration algorithm [37].

Instead of directly computing the optimal power-delay tradeoff under a given message error rate δ defined in (32), we start from the optimal power-delay-error tradeoff, and then transform it into the optimal power-delay tradeoff by setting the error rate to be a constant δ.

By a time-sharing argument [43], the achievable power-delay-error region is convex, hence the Lagrangian method can be applied to solve the power-delay-error tradeoff problem. We minimize the Lagrangian cost

$$g^{\pi} = -p_s^{\pi} + \mu P^{\pi} + \nu N^{\pi} \tag{38}$$

with $\mu, \nu \ge 0$, since error rate is equal to $1 - p_s^{\pi}$, and delay is proportional to N^{π}.

1) Markov Decision Process and the Optimality Equation: For each policy $\pi = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_I)$, using (22) and (30), we can rewrite the Lagrangian cost in (38) as[3]

$$g^{\pi} = \limsup_{I\to\infty} \frac{1}{t_I}\mathbb{E}^{\pi}[\sum_{i=1}^{I} c_{\mathbf{d}_i}(q_i)], \tag{39}$$

where $c_{\mathbf{d}_i}(q_i) = -\frac{T_c}{R_a} k_i (1 - \epsilon_i^{L_i}) + \mu \bar{P}_i + \nu \overline{Q}_i$ is the expected total cost between two successive decision epochs, given the system occupies state $q_i \in \mathbb{N}$ at the first decision epoch and the decision maker chooses action (k_i, n_i, P_i, L_i).

Denote $P_{\mathbf{d}_i}(q_{i+1}|q_i)$ as the the transition probability that the MDP occupies state q_{i+1} at the next decision epoch, given the decision maker chooses action (k_i, n_i, P_i, L_i) in state q_i at current decision epoch. Then

$$\begin{aligned} & P_{\mathbf{d}_i}(q_{i+1}|q_i) = 1, \quad \text{if } k_i = 0,\ q_{i+1} = q_i + 1; \\ & P_{\mathbf{d}_i}(q_{i+1}|q_i) = \sum_{l=1}^{L_i - 1} (1-\epsilon_i)\epsilon_i^{l-1} p(j, \lambda_{msg} n_i l T_c) \\ & \qquad + \epsilon_i^{L_i - 1} p(j, \lambda_{msg} n_i l T_c), \\ & \qquad \text{if } k_i > 0, j \ge 0,\ q_{i+1} = q_i - \frac{k_i}{b} + j, \end{aligned} \tag{40}$$

where $p(m,\mu) = \frac{\mu^m}{m!} e^{-\mu}, m \ge 0$ is the Poisson probability mass function. To achieve the optimal coding strategy that

[3]We only deal with unichain MDP, for which the optimal cost does not depend on the initial state. We thus omit the initial state in (32), (37) and (39).

TABLE I
POLICY ITERATION ALGORITHM

1. Set $n = 0$ and select an arbitrary coding strategy $\mathbf{d}_n \in \mathbb{D}$.
2. (Policy Evaluation) Obtain a scalar g_n and an $\mathbf{h}_n$ by solving
$$\mathbf{0} = \mathbf{c}_{\mathbf{d}_n} - g\mathbf{y}_{\mathbf{d}_n} + \left(\mathbf{P}_{\mathbf{d}_n} - \mathbf{I}\right)\mathbf{h}$$
3. (Policy Improvement) Choose $\mathbf{d}_{n+1}$ to satisfy
$$\mathbf{d}_{n+1} \in \arg\min_{\mathbf{d}\in\mathbb{D}} \left\{\mathbf{c}_{\mathbf{d}} - g_n\mathbf{y}_{\mathbf{d}} + \left(\mathbf{P}_{\mathbf{d}} - \mathbf{I}\right)\mathbf{h}_n\right\}$$
setting $\mathbf{d}_{n+1} = \mathbf{d}_n$ if possible.
4. If $\mathbf{d}_{n+1} = \mathbf{d}_n$, stop and set $\mathbf{d}^* = \mathbf{d}_n$. Otherwise increment n by 1 and return to step 2.

minimizes g^π in (39), we need to introduce the following theorem from [37].

Theorem 7. *[37] A) For a unichain average cost MDP, regardless of the initial state, the optimality equation is given by*

$$\mathbf{0} = \min_{\mathbf{d}\in\mathbb{D}} \left\{\mathbf{c}_{\mathbf{d}} - g\mathbf{y}_{\mathbf{d}} + \left(\mathbf{P}_{\mathbf{d}} - \mathbf{I}\right)\mathbf{h}\right\}, \tag{41}$$

where $\mathbf{P}_{\mathbf{d}}$ and $\mathbf{I}$ are the transition probability matrix consisting of $P_{\mathbf{d}}(q_{i+1}|q_i)$ defined in (40) and unit matrix, respectively, $\mathbf{c}_{\mathbf{d}} = (c_{\mathbf{d}}(0), c_{\mathbf{d}}(1), c_{\mathbf{d}}(2), \cdots)^T$, $\mathbf{y}_{\mathbf{d}} = (y_{\mathbf{d}}(0), y_{\mathbf{d}}(1), y_{\mathbf{d}}(2), \cdots)^T$ with the q-th component being the expected length of time from this decision epoch to the next one, given the decision maker chooses action in state q at current decision epoch, $\mathbf{h}$ is also a vector, representing the bias of the Markov cost process, and g is a scalar for the expected average cost.

B) Moreover, if there exists a decision rule $\mathbf{d} \in \mathbb{D}$, constant g and vector $\mathbf{h}$ for which (41) holds, then $g = g^*$, where g^* is the optimal cost for all initial states and the policy using $\mathbf{d}$ at all coding epochs is optimal.

For our optimization problem of (39), $y_{\mathbf{d}}(q_i)$ can be expressed as

$$y_{\mathbf{d}}(q_i) = \begin{cases} \frac{1}{\lambda_{msg}}, & \text{if } k_i = 0; \\ n_i\bar{l}_i T_c, & \text{if } k_i > 0. \end{cases} \tag{42}$$

Part A) of Theorem 7 gives a necessary condition for the optimal cost and part B) states that this necessary condition is also sufficient. Hence, we can apply it to compute the optimal cost and the corresponding optimal coding strategy[4].

2) The Policy Iteration Algorithm: In order to solve the optimality equation (41), we borrow the policy iteration algorithm of [37] shown in Table I. Convergence and uniqueness of its solution are guaranteed by Theorem 8.6.6 of [37]. Hence using this algorithm, we can find the optimal coding strategy π^*. We note that in our analysis, the buffer size is infinity, but the policy iteration algorithm can only be implemented for finite state spaces. Hence this algorithm only can be used to approximately compute the optimal energy-delay tradeoff, by setting the buffer size to a sufficiently large value.

[4]To guarantee the MDP is unichain, we assume that any decision rule in $\mathbb{D}$ has the property for any $q_2 \in \mathbb{N}$, $k(q_2) > 0$ if there exists a $q_1 \in \mathbb{N}$ and $q_1 \le q_2$ such that $k(q_1) > 0$.

B. Bounds and Asymptotics

The algorithmic approach employed above to computing the optimal energy/power-delay functions is not amenable to analysis, because the optimal solution is implicitly given by the Bellman equation [37], even in the limit as $d \to \infty$. We now give theoretical bounds of the energy/power-delay functions and study their asymptotics.

1) Lower and Upper Bounds: Denote $\mathbb{A} \triangleq \{(k, n, P, L) : k/b \in \mathbb{N}\}$ as the set of action elements. Define the set of the probability distributions on $\mathbb{A}$ satisfying the following constraints on system stability, packet loss, and delay bound.

$$\begin{aligned} A_{d,\delta}^{LB} \triangleq \Big\{ & (\alpha_1, \alpha_2, \cdots, \alpha_{|\mathbb{A}|}) : 0 \le \alpha_i \le 1, \text{for} 1 \le i \le |\mathbb{A}|, \\ & \sum_{i=1}^{|\mathbb{A}|} \alpha_i = 1, \\ & \frac{\sum_{i=1}^{|\mathbb{A}|} \alpha_i n_i \bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i} \le \frac{1}{R_a}, \\ & \frac{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i (1-\epsilon_i^{L_i})}{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i} \ge 1-\delta, \\ & \frac{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i^2}{2R_a \sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i} + \frac{\sum_{i=1}^{|\mathbb{A}|} \alpha_i n_i \bar{l}_i k_i}{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i} - \frac{b}{2R_a} \le \frac{d}{R_a} \Big\}, \end{aligned} \tag{43}$$

where $\bar{l}_i$ was defined in previous subsections. Define a set of modified actions (with the parameter L replaced by the transmission success rate p_0) as

$$\begin{aligned} A_{d,\delta}^{UB} \triangleq \Big\{ & (k, n, P, p_0) : p_0 = \frac{1-\delta}{1-\epsilon} \le 1, \\ & k = bp_0k_0, k_0 \in \mathbb{N}, \\ & \rho = \frac{R_a p_0}{r} < 1, \\ & k_0 b\left(1 + \frac{\rho}{2}\right) + \frac{b(3\rho - 2)}{2(1-\rho)} \le d \Big\}, \end{aligned} \tag{44}$$

where r and ϵ were defined in previous subsections. The inequalities in definition of $A_{d,\delta}^{UB}$ respectively correspond to constraints on packet loss, system stability, and delay bound. Then we have the following bounds.

Theorem 8. *For a given arrival rate λ_{msg} the energy-delay function is bounded by*

$$E_b^{LB}(d, \delta) \le E_b(d, \delta) \le E_b^{UB}(d, \delta) \tag{45}$$

where

$$E_b^{LB}(d, \delta) = \inf_{(\alpha_1, \alpha_2, \cdots, \alpha_{|\mathbb{A}|}) \in A_{d,\delta}^{LB}} \frac{\sum_{i=1}^{|\mathbb{A}|} \alpha_i n_i P_i \bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|} \alpha_i k_i}, \tag{46}$$

$$E_b^{UB}(\tau, \delta) = \inf_{(k,n,P,p_0) \in A_{d,\delta}^{UB}} \frac{p_0 P}{r}. \tag{47}$$

The proof is given in Appendix D.

Fig. 6 shows the energy-delay function together with its bounds.

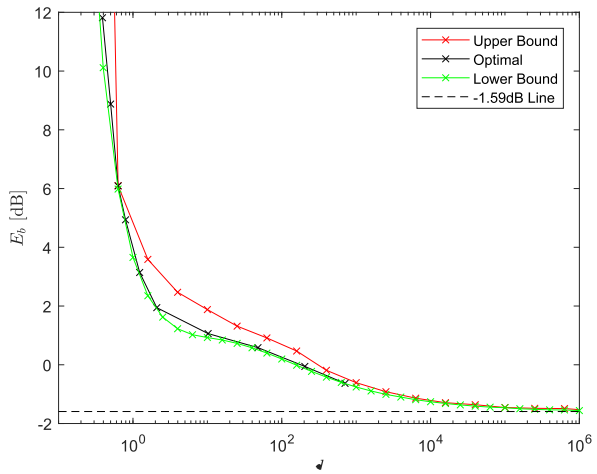

Fig. 6. The energy-delay function together with its bounds for message arrival rate message arrival rate $\lambda_{msg} = 0.01$, error probability constraint $\delta = 0.01$, bandwidth $T_c = 0.01$, and $b = 100$.

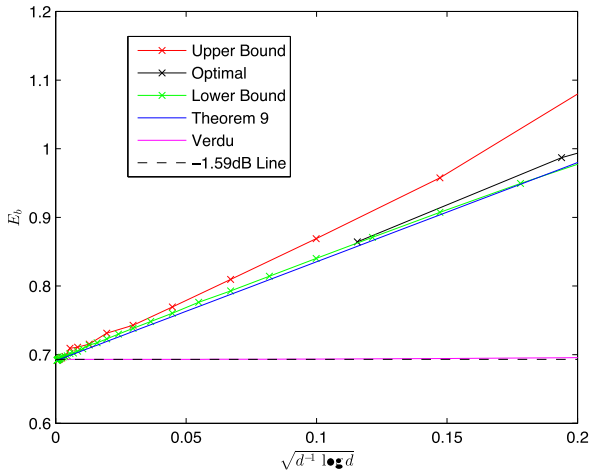

Fig. 7. Comparison of energy per bit vs $\sqrt{d^{-1}\log d}$ for finite and infinite [23] blocklength codes. In the former case, message arrival rate $\lambda_{msg} = 0.01$, error probability constraint $\delta = 0.01$, bandwidth $T_c = 0.01$, and $b = 100$.

2) Asymptotics: Similar to Section II we would like to analyze the system in Fig. 4 in terms of fundamental energy-delay trade-off. It is important to notice here that finite blocklength theory, whether [21] or [47], keeps the packet error probability ϵ constant as the blocklength $n \to \infty$, and this is not easy to get around. The analytical results there will assume that ϵ is constant as $d \to \infty$. On the other hand, in the scheduling algorithm, when (k_i, n_i, P_i, L_i) is chosen adaptively, there is no good reason to fix ϵ_i, as mentioned previously, and the same is true when numerically calculating performance. There is therefore a slight difference between the setup for exact analysis and numerical analysis. With this in mind we get

Theorem 9. *For fixed R_a and average energy, the achievable energy per bit is*

$$\frac{E_b(d,\delta)}{1-\epsilon} = \frac{2^{2R_a p_0}-1}{2R_a} + \frac{2^{2R_a p_0}\sqrt{3V\left(2^{2R_a p_0}-1\right)}}{\sqrt{2R_a}\log e}\sqrt{\frac{\log d}{d}} + o\left(\sqrt{\frac{\log d}{d}}\right). \tag{48}$$

where $p_0(1-\epsilon) = 1-\delta$. This is achieved for $\rho(d) = 1 - \frac{\log d}{d}$ for d sufficiently large. This is true with or without ARQ.

The proof is in Appendix E. Here $1 - p_0$ is the packet/bit drop probability. In Section II we did not allow bit dropping, and we can therefore compare with Theorem 2 for $p_0 = 1$. In this section we consider average delay, and we can take this into account by changing $d \longmapsto \frac{d}{2}$ in (16) to arrive at

$$\frac{E_b(d,\delta)}{1-\delta} = \frac{2^{2R_a}-1}{2R_a} + \frac{2^{2R_a}\sqrt{3V\left(2^{2R_a}-1\right)}}{\sqrt{2R_a}\log e}\sqrt{\frac{\log d}{d}} + o\left(\sqrt{\frac{\log d}{d}}\right). \tag{49}$$

Both of these expressions are achievable energies, so one should be slightly careful with concluding too much. Still, the fact that the expressions are the same *indicates* that random packet arrival does not increase energy up to the first order in delay (i.e., the second terms in (48) and (49)).

It is clear that the first term in (48) is minimized for $\epsilon = 0$ or, more precisely, $\epsilon \to 0$ as $d \to \infty$, but, as mentioned, this limit is not allowed in the analysis[5]. However, we can analyze the energy per bit in the limit, i.e., the first term separately to get

Theorem 10. *As $d \to \infty$, and letting $\epsilon \to 0$ both bounds $E_b^{LB}(d,\delta)$ and $E_b^{UB}(d,\delta)$ in* (45) *approach*

$$\frac{1}{2R_a}\left(2^{2R_a(1-\delta)}-1\right). \tag{50}$$

The proof is given in Appendix F.

C. Numerical Results

In this section we will plot some numerical results for the algorithms and bounds developed above. As a baseline, we would also like to compare with what one would expect using infinite blocklength theory. Suppose that data arrives in packets of size b, and each such packet is transmitted in n channel uses. A reasonable measure of delay[6] is therefore a scaling of n – with our normalization $d = nR_a$. The power needed for transmission is $P = 2^{2r} - 1 = 2^{2R_a b d^{-1}} - 1$, and the energy can be found from (8) (with $k = b$). Comparing with the wideband slope of [24], it can then be seen that the energy per bit (in dB) is linear in d^{-1} to the first order.

Fig. 6 depicts convergence of the lower and upper bound with d. In Fig. 7, we plot the energy per bit vs $\sqrt{d^{-1}\log d}$ to better see its asymptotic behavior. The infinite blocklength

[5]We will briefly clarify the difference between bit dropping and the packet dropping of [47]. The communication scheme in [47] allows packet dropping. However, when a packet is dropped in [47], the transmitter stays idle, which makes sense for single packet transmission. On the other hand, in streaming, when bits are dropped, some other bits could be transmitted instead, which is what we refer to a bit dropping. In terms of energy, it is more efficient to spread out transmission, and this explains why $\epsilon \to 0$ as $d \to \infty$, i.e., all errors are due to bit dropping in the limit.

[6]Since with infinite blocklength the *actual* delay is infinite, some handwaving is usually used to conclude something about delay.

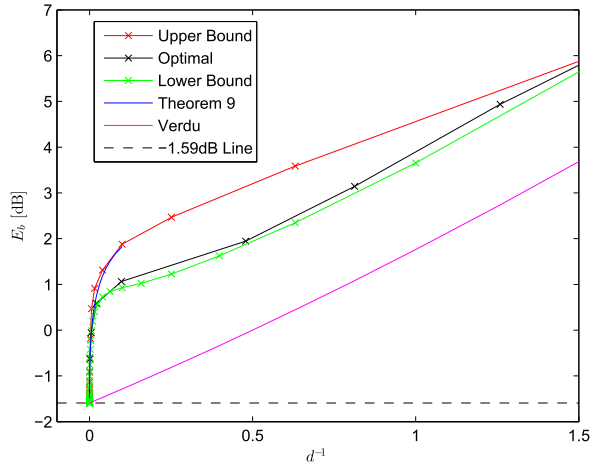

Fig. 8. Comparison of energy per bit vs d^{-1} for finite and infinite [23] blocklength codes. In the former case, message arrival rate $\lambda_{msg} = 0.01$, error probability constraint $\delta = 0.01$, bandwidth $T_c = 0.01$, and $b = 100$.

case of [24] is labeled 'Verdu'. It is seen that, for a finite blocklength code, the energy increases linearly with $\sqrt{d^{-1}\log d}$, which verifies the results of Theorem 9. In Fig. 8 we plot the same results[7], but for smaller values of d. It is seen that the energy increases very rapidly when delay moves away from infinity, as opposed to the prediction from infinite blocklength theory that energy just increases linearly with (inverse) delay. This means that it is very hard to achieve the -1.59 dB limit in practice when the actual delay is taken into account. This confirms the conclusion from Theorem 6.

IV. Conclusion

This paper has investigated the basic tradeoffs between delay and energy in streaming data. A consistent finding is that the minimum energy is approached slowly, both in terms of delay and bandwidth, much more so than predicted by infinite blocklength theory.

The results in the paper have been based on existing finite blocklength theory. This puts some limits on what can be proven. Basically, finite blocklength theory was not developed with streaming delay and energy in mind. We have therefore chosen our constraints so that they could be matched to finite blocklength theory. For example, we have only considered packet transmission, and with that a fixed packet error probability. Also, energy has been measured in a packet-based framework. It would be interesting to see generalizations of for example Theorems 1 and 3 without being constrained to packet transmission, and Theorem 9 with $\epsilon \to 0$. This requires new approaches to finite blocklength theory (and a new name: finite delay theory, as the fundamental constraint is not that the blocks should be finite, but the delay be finite). We hope to develop this in a future paper.

Appendix A
Proof of Theorems 1 and 2

The $O(\cdot)$ terms in (9), (15) are with respect to n, but depends on δ and P. Here, δ is still kept constant, but P is variable. For Theorem 1 we can therefore write

$$k = nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n + b(n,P), \tag{51}$$

where $|b(n,P)| \le M(P)$ for sufficiently large n. It is easy to see from [21], [46] that $M(P)$ is itself bounded for small variations of P. That is, we can write formally

$$\exists \delta, M > 0 : \forall n > n_0 : \forall P \in [P_0 - \delta, P_0 + \delta] : |b(n,P)| \le M \tag{52}$$

and in this region we can write (9) explicitly as

$$\begin{aligned} nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n - M \\ \le k \le nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n + M. \end{aligned} \tag{53}$$

We also know that since R_a is fixed we have

$$\lim_{d\to\infty} P = P_0 = 2^{2R_a} - 1. \tag{54}$$

Thus, we can use (52) in the limit.

Let us write $P = P_0 + \triangle P$, where $\lim_{d^{-1}\to 0} \triangle P = 0$. With this we can write $C(P), V(P)$ as

$$C = C_0 + \frac{\log e}{2(1+P_0)}\triangle P + o(\triangle P) \tag{55}$$

$$\sqrt{V} = \sqrt{V_0} + \frac{\sqrt{V_0}}{P_0(1+P_0)(2+P_0)}\triangle P + o(\triangle P), \tag{56}$$

where C_0, V_0 is C, V evaluated at P_0. Let $K_0 = \frac{\log e}{2(1+P_0)}$, $K_1 = \frac{\sqrt{V_0}}{P_0(1+P_0)(2+P_0)}$. Now $o(\triangle P) = \triangle P \varepsilon(\triangle P)$, where we use $\varepsilon(x)$ to denote any function that satisfies

$$\lim_{x\to 0} \varepsilon(x) = 0. \tag{57}$$

But we know from (54) that $\triangle P \to 0$ when $d^{-1} \to 0$ so $\varepsilon(\triangle P) = \varepsilon\left(d^{-1}\right)$. And thus we can say that $o(\triangle P) = \triangle P \varepsilon(\triangle P) = \triangle P \varepsilon\left(d^{-1}\right)$. Let $K_2 = \sqrt{2R_a}Q^{-1}(\delta)$ and set $k = (1-\alpha)d, n = d\alpha R_a^{-1}$ where $\alpha \in \left[\beta, \frac{1}{2}\right]$ for any constant β satisfying $0 < \beta < \frac{1}{2}$ by (6). We introduced β to bound α away from zero; in the end, the value of β will not matter. One consequence is $\lim_{d\to\infty} n = \infty$ so that we can use (9).

With this relationships (53) can be written as

$$\begin{aligned} (1-\alpha)d \lessgtr{} & \alpha d R_a^{-1}\left(R_a + K_0 \triangle P(1 + \varepsilon(\triangle P))\right) \\ & - \sqrt{\alpha d R_a^{-1}}Q^{-1}(\delta)\sqrt{V_0} \\ & - \sqrt{\alpha d R_a^{-1}}Q^{-1}(\delta) K_1 \triangle P(1 + \varepsilon(\triangle P)) \\ & + \frac{1}{2}\log\left(\alpha d R_a^{-1}\right) \pm M. \end{aligned} \tag{58}$$

Solving this with respect to ΔP we get and using the series expansion $\frac{1}{1-y} = 1 + y + O\left(y^2\right)$

$$\begin{aligned} \triangle P(1 + \varepsilon(\triangle P)) = {} & \frac{1-2\alpha}{\alpha}K_0^{-1}R_a \\ & + \sqrt{\alpha^{-1}d^{-1}R_a V_0}K_0^{-1}Q^{-1}(\delta) \\ & + \sqrt{\alpha^{-1}d^{-1}R_a^3}K_0^{-2}K_1\frac{1-2\alpha}{\alpha}Q^{-1}(\delta) \\ & + o\left(\sqrt{d^{-1}}\right). \end{aligned} \tag{59}$$

[7]To compare with the infinite blocklength case we plot versus d^{-1}, as in that case energy is (approximately) linear in d^{-1}.

Now we can write using the relation $\varepsilon(\triangle P) = \varepsilon\left(d^{-1}\right)$ that

$$\begin{aligned} E_b(d,\delta) &= \frac{nP}{2k} = \frac{\alpha}{1-\alpha}\frac{P}{2R_a} = \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\left[P_0+\triangle P\right] \\ &= \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\left[P_0+\frac{1-2\alpha}{\alpha}K_0^{-1}R_a\right] \\ &+ \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\sqrt{\alpha^{-1}d^{-1}R_aV_0}K_0^{-1}Q^{-1}(\delta) \\ &+ \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\sqrt{\alpha^{-1}d^{-1}R_a^3}K_0^{-2}K_1\frac{1-2\alpha}{\alpha}Q^{-1}(\delta) \\ &+ o\left(\sqrt{d^{-1}}\right) \end{aligned} \tag{60}$$

According to Lemma 63 in [21] we can minimize (60) with respect to α by minimizing the first term: all terms are continuous in α, and the minimization is over the compact set $\left[\beta,\frac{1}{2}\right]$. According to Lemma 63 in [21] this minimization results in an $o\left(\sqrt{d^{-1}}\right)$ term, which can be included in the $o\left(\sqrt{d^{-1}}\right)$ term in (60). Thus, we have to minimize

$$f(P_0,\alpha) = \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\left[P_0+\frac{1-2\alpha}{\alpha}\frac{2(1+P_0)}{\log e}R_a\right]. \tag{61}$$

Making use of the substitution $R_a = C(P_0)$ we have

$$f(P_0,\alpha) = \frac{\alpha\ln 2\left(\frac{(1-2\alpha)(2P_0+2)\ln(1+P_0)}{2\alpha}+P_0\right)}{(1-\alpha)\log(P_0+1)} \tag{62}$$

$$\frac{\partial f}{\partial\alpha} = \frac{\ln 2(P_0-(P_0+1)\ln(P_0+1))}{(\alpha-1)^2\log(P_0+1)} < 0, \forall P_0 > 0. \tag{63}$$

This tells us that for sufficiently large d, setting $\alpha = \frac{1}{2}$ is therefore optimum, independent of β, which corresponds to continuous transmission. Substituting the value of α we obtain

$$\triangle P = \frac{K_2\sqrt{V_0}}{K_0}\sqrt{d^{-1}} + o\left(\sqrt{d^{-1}}\right) \tag{64}$$

and $E_b = \frac{1}{2R_a}\left[P_0+\frac{K_2\sqrt{V_0}}{K_0}\sqrt{d^{-1}}+o\left(\sqrt{d^{-1}}\right)\right]$.

For Theorem 2 we proceed similarly but instead of (52) now have

$$\frac{k}{n} \leqslant C\left(\frac{P}{1-\delta}\right) - \sqrt{V\left(\frac{P}{1-\delta}\right)}\sqrt{\frac{\log n}{n}} \pm \frac{M}{\sqrt{n}}. \tag{65}$$

We now expand $C(P)$ and $V(P)$ around $P_0 = (1-\delta)\left(2^{2R_a}-1\right)$ and making the substitutions $k = (1-\alpha)d$ and $n = \alpha d R_a^{-1}$ where $\alpha \in \left[\beta,\frac{1}{2}\right]$ for any constant β satisfying $0 < \beta < \frac{1}{2}$. We have

$$R_a\frac{1-\alpha}{\alpha} \leqslant C_0 - V_0 + [C_1 - V_1]\triangle P + o(\triangle P) \mp \frac{M}{\sqrt{\alpha d R_a^{-1}}}, \tag{66}$$

where $C_0 = C\left(\frac{P_0}{1-\delta}\right) = R_a, C_1 = \frac{1}{2\ln 2(1-\delta+P_0)}$, $V_0 = \sqrt{V\left(\frac{P_0}{1-\delta}\right)\log\left(\alpha d R_a^{-1}\right)R_a\alpha^{-1}d^{-1}}$ and $V_1 = \frac{(\delta-1)^2\sqrt{\log\left(\alpha d R_a^{-1}\right)R_a\alpha^{-1}d^{-1}}}{\sqrt{2}\ln 2\sqrt{P_0(-2\delta+P_0+2)}(-\delta+P_0+1)^2}$. Again, we solve for ΔP and obtain

$$\begin{aligned} E_b(d,\delta) &= \frac{nP_0+n\triangle P}{2k} \\ &= \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\left[P_0+C_1^{-1}R_a\left(\alpha^{-1}-2\right)\right] \\ &+ \frac{1}{2R_a}\frac{\alpha}{1-\alpha}\left[C_1^{-1}V_0+R_a\left(\alpha^{-1}-2\right)C_1^{-2}V_1\right] \\ &+ o\left(\sqrt{d^{-1}\log d}\right). \end{aligned} \tag{67}$$

Minimizing $E_b(d,\delta)$ for sufficiently large d is equivalent to minimizing (using Lemma 63 in [21])

$$g(P_0,\alpha) = \frac{\alpha\ln 2\left(\frac{(1-2\alpha)(-2\delta+2P_0+2)\ln\left(\frac{P_0}{1-\delta}+1\right)}{2\alpha}+P_0\right)}{(1-\alpha)\ln\left(\frac{P_0}{1-\delta}+1\right)} \tag{68}$$

$$\frac{\partial g}{\partial\alpha} = \frac{(\delta-1)\left(4^{R_a}(R_a\ln 4-1)+1\right)}{2R_a(\alpha-1)^2} < 0, \quad \forall R_a > 0, \forall\delta\in(0,1). \tag{69}$$

This tells us that for sufficiently large d, setting $\alpha = \frac{1}{2}$ is optimum. Substituting the value of $\alpha = \frac{1}{2}$ in (67) we get the desired result.

Appendix B
Proof of Theorem 6

Non-convergence is just a restatement of Theorem 1.

To achieve the minimum energy limit (1) transmission in the low-power regime is required, that is, the essence of reaching the limit is that $P \to 0$. Expression (9) is for a fixed P. In the proof of Theorem 1 this was overcome by noting that the bound has bounded variations for small variations of P. However, to establish a bound for $P \to 0$, we need to examine the proof of (9) more carefully.

In [21, Appendix L] the authors explicitly state that

$$k \leq nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n + g_c(P,\delta) \tag{70}$$

We will show that $\lim_{P\to 0} g_c(P,\delta) = k$, some constant. Here

$$\begin{aligned} g_c(P,\delta) &= -2B(P)\sqrt{V}g_1(P,\delta) - \log B(P) \\ B(P) &= \frac{6E[|S_i|^3]}{V^{3/2}} \\ S_i &= \frac{P\log e}{2(1+P)}\left(Z_i^2 - 2\frac{Z_i}{\sqrt{P}} - 1\right) \\ Z_i &\sim N(0,1). \end{aligned} \tag{71}$$

It is easy to see that $\lim_{P\to 0} B(P) = k_1$, since the dominating terms in P have the same power in numerator and denominator. Now

$$g_1(P,\delta) = \min_{\theta\in[\alpha_1,1-\delta]}\frac{d}{dx}Q^{-1}(\theta). \tag{72}$$

Since $\alpha_1 < 1-\delta$, $g_1(P,\delta)$ is bounded away from $-\infty$. Therefore $g_c(P,\delta)$ is bounded as $P \to 0$ and we have

$$\begin{aligned} k &\leq nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n + b(n,P) \\ \exists P_0, M > 0 &: \forall n > n_0 : \forall P < P_0 : |b(n,P)| \leq M. \end{aligned} \tag{73}$$

In the other direction, [46] give the following lower bound on k (slightly restated)

$$k \geq nC - \sqrt{nV(P)}Q^{-1}\left(\delta - \frac{B(P)+G(P)}{\sqrt{n}} - \xi_n(P)\right) - \log\left(\frac{G(P)J(P)}{\sqrt{n}}\right) \tag{74}$$

where we have noted explicit dependence on P, $B(P)$ is given in (71), and where

$$G(P) = \frac{24}{\sqrt{2\pi}}\frac{(1+P)^{3/2}\sqrt{1+2P}}{\sqrt{P}}. \tag{75}$$

We have already proven that $B(P)$ is bounded as $P \to 0$ above, and it is easy to see from [46] that $J(P)$ is bounded (from above) and bounded away from zero as $P \to 0$. Furthermore, $\lim_{P\to 0}\xi_n(P) = 0$. The complication is that $\lim_{P\to 0} G(P) = \infty$. For the bound to even be valid (that is, relevant), we require that what is inside the Q^{-1} function to converge to δ, that is $\frac{G(P)}{\sqrt{n}} \to 0$, or $nP \to \infty$.

We will first consider continuous transmission, that is $k = \frac{d}{2}$ and therefore $n = \frac{d}{2}R_a^{-1}$. In terms of the lower bound, from (73) this solution satisfies

$$\frac{d}{2} \leq \frac{d}{2R_a}C - \sqrt{\frac{d}{2R_a}V}Q^{-1}(\delta) + \frac{1}{2}\log\frac{d}{2R_a} + M. \tag{76}$$

Using

$$\begin{aligned} C &= \frac{P}{2\ln 2} - \frac{P^2}{4\ln 2} + o\left(P^2\right) \\ \sqrt{V} &= \frac{\sqrt{P}}{\ln 2} + o\left(\sqrt{P}\right) \\ E_b &= \frac{P}{2R_a}. \end{aligned} \tag{77}$$

we can write (76) as

$$\begin{aligned} 1 \leq &\left[\frac{E_b}{\ln 2} - \frac{R_a E_b^2}{\ln 2}\left(1+\varepsilon\left(P\right)\right)\right] \\ &- \sqrt{\frac{2}{d}}Q^{-1}(\delta)\left[\frac{\sqrt{2E_b}}{\ln 2}\left(1+\varepsilon\left(P\right)\right)\right] \\ &+ \frac{\log\frac{d}{2R_a} + 2M}{d}. \end{aligned} \tag{78}$$

From the lower bound of the Theorem statement, $O\left(d\exp(-\sqrt{d\ln 2}Q^{-1}(\delta))\right) < R_a$, follows that $\frac{\log\frac{d}{2R_a}+2M}{d} = o(\sqrt{d^{-1}})$. With this and the fact that $R_a \leq o(\sqrt{d^{-1}}) = \sqrt{d^{-1}}\varepsilon\left(d^{-1}\right)$ and $\varepsilon\left(P\right) = \varepsilon\left(d^{-1}\right)$, we can write (78) as

$$1 \leq \frac{E_b}{\ln 2} - \frac{2Q^{-1}(\delta)}{\ln 2}\sqrt{E_b d^{-1}} + o\left(\sqrt{d^{-1}}\right) \tag{79}$$

We solve this inequality with respect to E_b,

$$\sqrt{E_b} \geq -\frac{B}{2}\left(1 + \sqrt{1 + \frac{4\ln 2}{B^2}}\right) \tag{80}$$

$$B = -2Q^{-1}(\delta)\sqrt{d^{-1}}\left(1+\varepsilon\left(d^{-1}\right)\right). \tag{81}$$

The square root can be expanded as

$$\sqrt{1+\frac{4\ln 2}{B^2}} = \frac{\sqrt{\ln 2}}{Q^{-1}(\delta)}\sqrt{d}\left(1+\varepsilon\left(d^{-1}\right)\right). \tag{82}$$

Then

$$\begin{aligned} \sqrt{E_b} &\geq Q^{-1}(\delta)\sqrt{d^{-1}}\left(1+\varepsilon\left(d^{-1}\right)\right) \\ &\quad\times\left(1 + \frac{\sqrt{\ln 2}}{Q^{-1}(\delta)}\sqrt{d}\left(1+\varepsilon\left(d^{-1}\right)\right)\right) \\ &= \sqrt{\ln 2} + Q^{-1}(\delta)\sqrt{d^{-1}} + o\left(\sqrt{d^{-1}}\right), \end{aligned} \tag{83}$$

which give the lower bound in (20).

For the upper bound, notice that we consider solutions with $P \to 0$. On the other hand $\frac{R_a^{-1}P}{2} = E_b \geq E_{b,\min} > 0$ so that $nP = \frac{d}{2}R_a^{-1}P \to \infty$. Therefore the conditions for using the bound (74) are satisfied. We can write this as

$$\frac{d}{2} \geq \frac{d}{2R_a}C - \sqrt{\frac{d}{2R_a}V}Q^{-1}(\delta) + \frac{1}{2}\log\frac{d}{2R_a} - \log G(P) + M. \tag{84}$$

Now use that

$$-\log G = \frac{\ln P}{2} + \tilde{M} - \frac{5}{2}P + \frac{7}{4}P^2 + o\left(P^2\right) \tag{85}$$

together with (77) to get

$$\begin{aligned} 1 \geq &\left[\frac{E_b}{\ln 2} - \frac{R_a E_b^2}{\ln 2}\left(1+\varepsilon\left(P\right)\right)\right] \\ &- 2\sqrt{\frac{1}{d}}Q^{-1}(\delta)\left[\frac{\sqrt{E_b}}{\ln 2}\left(1+\varepsilon\left(P\right)\right)\right] \\ &+ \frac{\log E_b d}{d} - \frac{10R_a E_b}{d} + \frac{14\left(R_a E_b\right)^2\left(1+\varepsilon\left(P\right)\right)}{d} + \frac{2\hat{M}}{d} \end{aligned} \tag{86}$$

or

$$1 \geq \frac{E_b}{\ln 2} - \frac{2Q^{-1}(\delta)}{\ln 2}\sqrt{E_b d^{-1}} + o\left(\sqrt{d^{-1}}\right). \tag{87}$$

This is the same as (79) with the inequality reversed. Solving it the same way for E_b results in the upper bound in (20)

We next consider solutions that allow bursty transmission. First notice that we have

$$\begin{aligned} E_b = \frac{nP}{2k} &\geq \frac{nP}{2\left(nC - \sqrt{nV}Q^{-1}(\delta) + \frac{1}{2}\log n + M\right)} \\ &= \frac{1}{f(P)} \end{aligned} \tag{88}$$

$$f(P) = 2\frac{nC}{nP} - 2\frac{\sqrt{nV}}{nP}Q^{-1}(\delta) + \frac{\log n}{nP} + \frac{2M}{nP}. \tag{89}$$

The idea is that we maximize $f(P)$ for fixed n (we will get to constraints below). We have

$$f'(P)nP^2 = -\log n - 2M + \sqrt{nP}Q^{-1}(\delta)\log e + \sqrt{n}o(\sqrt{P}). \tag{90}$$

To solve $f'(P) = 0$ is equivalent to solving

$$P\left(1+\varepsilon\left(P\right)\right) = \frac{(2M+\log n)^2}{\left(Q^{-1}(\delta)\log e\right)^2 n}. \tag{91}$$

We claim that the solution is

$$P = \frac{\log^2 n}{\left(Q^{-1}(\delta)\log e\right)^2 n} + \frac{\log^2 n}{n}\varepsilon(n^{-1}). \tag{92}$$

To show that this is the solution, we need to verify that the term $\epsilon(n^{-1})$ does in fact converge to zero. Let explicitly $\varepsilon(n^{-1}) = h(n)$. Then inserting (92) in (91)

$$h(n) = \frac{n}{\log^2 n}\frac{1}{1+\varepsilon(n^{-1})}\left(\frac{(2M+\log n)^2}{\left(Q^{-1}(\delta)\log e\right)^2 n} - \frac{\log^2 n}{\left(Q^{-1}(\delta)\log e\right)^2 n}(1+\varepsilon(n^{-1}))\right) \quad (93)$$

$$\to 0 \quad \text{as } n\to\infty. \quad (94)$$

Where we have used that $\varepsilon(P) = \varepsilon(n^{-1})$ for the given solution (92). We now insert (92) in (the Taylor series of) $f(P)$,

$$f(P) = \log e - P\frac{\log e}{2} - \frac{2Q^{-1}(\delta)\log e}{\sqrt{nP}} + \frac{3Q^{-1}(\delta)\log e}{2}\sqrt{\frac{P}{n}} + \frac{\log n + 2M}{nP} + o(P). \quad (95)$$

Taylor series now gives

$$\frac{\log n + 2M}{nP} = \frac{\left(Q^{-1}(\delta)\log e\right)^2}{\log n}\left(1+\varepsilon\left(n^{-1}\right)\right) + \frac{2M}{\log^2 n}\left(1+\varepsilon\left(n^{-1}\right)\right) \quad (96)$$

$$\frac{3Q^{-1}(\delta)\log e}{2}\sqrt{\frac{P}{n}} = \frac{3}{2}\frac{\log n}{n}\left(1+\varepsilon(n^{-1})\right) \quad (97)$$

$$-\frac{2Q^{-1}(\delta)\log e}{\sqrt{nP}} = \frac{-2\left(Q^{-1}(\delta)\log e\right)^2}{\log n}\left(1+\varepsilon\left(n^{-1}\right)\right) \quad (98)$$

$$o(P) = \frac{\log^2 n}{n}\left(1+\varepsilon\left(n^{-1}\right)\right)\varepsilon\left(n^{-1}\right), \quad (99)$$

which gives

$$f(P) = \log e - \frac{\left(Q^{-1}(\delta)\log e\right)^2}{\log n} + o\left(\frac{1}{\log n}\right). \quad (100)$$

Thus,

$$E_b \geq \frac{1}{\log e} + \frac{\left(Q^{-1}(\delta)\right)^2}{\log n} + o\left(\frac{1}{\log n}\right). \quad (101)$$

For the achievability bound, we can choose P, and based on (92) we choose

$$P = \frac{\log^2 n}{\left(Q^{-1}(\delta)\log e\right)^2 n}. \quad (102)$$

The bound (74) now is a function only of n(and δ of course). We call the resulting expression for $G(P)$ for $G(n)$ with

$$G(n) = \frac{24}{\sqrt{2\pi}}\frac{\sqrt{n}}{\log n} + o\left(\frac{\sqrt{n}}{\log n}\right). \quad (103)$$

Thus $\lim_{n\to\infty}\frac{G(n)}{\sqrt{n}} = 0$, so that we have

$$k \geq nC(n) - \sqrt{nV(n)}Q^{-1}(\delta) - \log\left(\frac{G(n)}{\sqrt{n}}\right) + O(1). \quad (104)$$

We can now upper bound the energy per bit as follows

$$E_b = \frac{nP}{2k} \leq \frac{\left(Q^{-1}(\delta)\log e\right)^{-2}\log^2 n}{2\left(nC(n) - \sqrt{nV(n)}Q^{-1}(\delta) - \log\left(\frac{G(n)}{\sqrt{n}}\right) + M\right)} \leq \frac{1}{g(n)} \quad (105)$$

$$g(n) = \frac{2nC(n) - 2\sqrt{nV(n)}Q^{-1}(\delta) - 2\log\left(\frac{G(n)}{\sqrt{n}}\right) + 2M}{\left(Q^{-1}(\delta)\log e\right)^{-2}\log^2 n}. \quad (106)$$

Here

$$\frac{2nC}{\left(Q^{-1}(\delta)\log e\right)^{-2}\log^2 n} = \log e - \frac{\left(Q^{-1}(\delta)\log e\right)^{-2}}{2}\frac{\log^2 n}{n} + o\left(\frac{\log^2 n}{n}\right) \quad (107)$$

and

$$-2\frac{\sqrt{nV(n)}Q^{-1}(\delta)}{\left(Q^{-1}(\delta)\log e\right)^{-2}\log^2 n} = -2\frac{1}{\log n} + \frac{3\left(Q^{-1}(\delta)\log e\right)^{-2}}{2}\frac{\log n}{n} + o\left(\frac{\log n}{n}\right) \quad (108)$$

$$-2\frac{\log\left(\frac{G(n)}{\sqrt{n}}\right)}{\log^2 n} + \frac{2M}{\log^2 n} = 2\frac{\log\log n}{\log^2 n} + \frac{\tilde{M}}{\log^2 n}. \quad (109)$$

Inserting this gives

$$g(n) = \log e - \frac{2}{\left(Q^{-1}(\delta)\log e\right)^{-2}}\frac{1}{\log n} + o\left(\frac{1}{\log n}\right). \quad (110)$$

From which

$$E_b \leq \frac{1}{\log e} + \frac{2\left(Q^{-1}(\delta)\right)^2}{\log n} + o\left(\frac{1}{\log n}\right). \quad (111)$$

We notice already here that in the bursty transmission regime, upper (111) and lower bounds (101) are not tight.

From (100) and (110) we also get the optimum k as

$$k^*(n) = \frac{1}{2\left(Q^{-1}(\delta)\right)^2\log e}\log^2(n) + O(\log n). \quad (112)$$

Now if $k^*(n) < \frac{d}{2}$, this means that $k = \frac{d}{2}$, the minimum, is optimum – in the sense that the lower bound is increasing with k, while it is also achieved for $k = \frac{d}{2}$. This condition is equivalent to

$$\frac{1}{2\left(Q^{-1}(\delta)\right)^2\log e}\log^2(n) + O(\log n) < \frac{d}{2} \quad (113)$$

$$n < 2^{\sqrt{d}Q^{-1}(\delta)\sqrt{\log e} + O(1)}. \quad (114)$$

That is

$$R_a > O\left(\frac{d}{2}2^{-\sqrt{d}Q^{-1}(\delta)\sqrt{\log e}}\right). \quad (115)$$

We now look at the regime of very bursty transmission. Since the energy is decreasing with n and $k(n)$ is increasing in n, we should choose $k(n) \approx d$. Specifically, we put $k(n) = d - \xi(d)$. With this we get

$$\log n = \sqrt{2\log e} Q^{-1}(\delta)\sqrt{d-\xi(d)} + O(1) \tag{116}$$

and then from (101) and (111)

$$E_b \geq \frac{1}{\log e} + \frac{Q^{-1}(\delta)}{\sqrt{2\log e}} \frac{1}{\sqrt{d-\xi(d)}} + o\left(\frac{1}{\sqrt{d-\xi(d)}}\right) \tag{117}$$

$$E_b \leq \frac{1}{\log e} + 2\frac{Q^{-1}(\delta)}{\sqrt{2\log e}} \frac{1}{\sqrt{d-\xi(d)}} + o\left(\frac{1}{\sqrt{d-\xi(d)}}\right). \tag{118}$$

We want to choose $\xi(d)$ so that it does not contribute to the first term, but is absorbed into the $o(\cdot)$ term. This is the case for $\xi(d) = o(d)$. Then

$$E_b \geq \frac{1}{\log e} + \frac{Q^{-1}(\delta)}{\sqrt{2\log e}} \frac{1}{\sqrt{d}} + o\left(\frac{1}{\sqrt{d}}\right) \tag{119}$$

$$E_b \leq \frac{1}{\log e} + 2\frac{Q^{-1}(\delta)}{\sqrt{2\log e}} \frac{1}{\sqrt{d}} + o\left(\frac{1}{\sqrt{d}}\right) \tag{120}$$

and

$$R_a = \frac{o(d)}{n} = o(d)\exp(-\sqrt{2d\ln 2}Q^{-1}(\delta)). \tag{121}$$

The lower bound (119) is lower than the lower bound of Proposition 3, so we can use that instead. The upper bound (120) on the other hand meets the lower bound, so that it is achievable.

Appendix C
Proof of Theorem 4

We use the notation of [27]. Consider codebooks with (E, M, ϵ) (with $M = 2^k$) satisfying the average energy constraint. The basic idea in both achievability and converse is to use a codebook $\left(\hat{E}, \hat{M}, \hat{\epsilon}\right)$ with $\hat{M} < M$ satisfying a maximal energy constraint, and then from this construct an average energy constraint codebook by amending the codebook with $M - \hat{M}$ $\mathbf{0}$ entries. For achievability we can assume this structure, and for the converse we need to prove it to be optimum. Let $0 < \alpha < 1$, then we set

$$M = \frac{\hat{M}}{\alpha} \tag{122}$$

$$E = \hat{E}\alpha \tag{123}$$

The average P_e is

$$\begin{aligned} P_e &= 1 \times \frac{M-\hat{M}}{M} + \hat{P}_e \times \frac{\hat{M}}{M} \\ &\leq 1 - \alpha + \hat{\epsilon} \times \alpha = \epsilon \end{aligned} \tag{124}$$

and the average energy of codewords in this codebook is

$$0 \times \frac{M-\hat{M}}{M} + \hat{E} \times \frac{\hat{M}}{M} = E. \tag{125}$$

For achievability we can directly use [27, Theorem 3] with $\left(\hat{E}, \hat{M}, \hat{\epsilon}\right)$. But instead of the final expression we use [27, (48)]

$$\begin{aligned} \log(\alpha M) &\geq -\log Q\left(\sqrt{\frac{2E}{\alpha}} + Q^{-1}\left(\frac{1-\epsilon}{\alpha}\right) + O\left(\sqrt{\frac{\alpha}{E}}\right)\right) \\ \log(M) &\geq -\log(\alpha) - \log Q\left(\sqrt{\frac{2E}{\alpha}} + Q^{-1}\left(\frac{1-\epsilon}{\alpha}\right)\right. \\ &\quad \left. + O\left(\sqrt{\frac{\alpha}{E}}\right)\right). \end{aligned} \tag{126}$$

For the converse, most of the proof of [27, Theorem 2] caries over to the case of variable energy. Specifically, [27, (21)] is still true when E is replaced by $E_j = \|\mathbf{c}_j\|^2$, and [27, (24)] is still true when E_j is used,

$$\frac{1}{M} \geq \frac{1}{M}\sum_{j=1}^{M} \beta_{P^j(g^{-1}(j))}(E_j). \tag{127}$$

Therefore,

$$\begin{aligned} \frac{1}{M} &\geq \frac{1}{M}\sum_{j=1}^{M} Q\left(\sqrt{2E_j} + Q^{-1}(1-\epsilon_j)\right) \\ &= \mathbb{E}\left[Q\left(\sqrt{2E_j} + Q^{-1}(1-\epsilon_j)\right)\right]_{1\leq j\leq M} \end{aligned} \tag{128}$$

where

$$\frac{1}{M}\sum_{j=1}^{M} E_j \leq E \tag{129}$$

$$\frac{1}{M}\sum_{j=1}^{M} \epsilon_j \leq \epsilon. \tag{130}$$

A valid choice here is to set $\epsilon_j = 1$ ([27, (21)] is still valid), and for those terms we get $Q\left(\sqrt{2E_j} + Q^{-1}(1-\epsilon_j)\right) = 0$ independent of E_j, so that setting $E_j = 0$ is optimum. Let us assume that the $(1-\alpha)M$ last codewords have $\epsilon_j = 1$. Then

$$\frac{1}{\alpha M}\sum_{j=1}^{\alpha M} Q\left(\sqrt{2E_j} + Q^{-1}(1-\epsilon_j)\right) \tag{131}$$

subject to

$$\begin{aligned} \frac{1}{\alpha M}\sum_{j=1}^{\alpha M} E_j &\leq \frac{E}{\alpha} \\ \frac{1}{\alpha M}\sum_{j=1}^{\alpha M} \epsilon_j + (1-\alpha) &\leq \epsilon. \end{aligned} \tag{132}$$

It is clear that (131) is minimized for the constraints (132) are satisfied with equality. Therefore let $E_1 = ME - \sum_{j=2}^{\alpha M} E_j$ and $\epsilon_1 = \alpha M(\epsilon - (1-\alpha)) - \sum_{j=2}^{\alpha M}\epsilon_j$. Taking derivatives of the right hand side of (131), it is then seen that a local minimum

therefore must satisfy

$$-\frac{1}{2}\left(\sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right)\right)^2 = -\frac{1}{2}\left(\sqrt{2E_1}+Q^{-1}\left(1-\epsilon_j\right)\right)^2 \tag{133}$$

$$2\left[Q^{-1}\left(1-\epsilon_j\right)\right]^2-\frac{1}{2}\left(\sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right)\right)^2 = 2\left[Q^{-1}\left(1-\epsilon_1\right)\right]^2-\frac{1}{2}\left(\sqrt{2E_1}+Q^{-1}\left(1-\epsilon_j\right)\right)^2. \tag{134}$$

We divide the codes into those that have $\infty > \sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right) > 0$ and those with $\sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right)\le 0$. We can assume that the former are the first $\beta\alpha M$ indices. For those j we can conclude from (133) that we must have $\sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right)$ independent of j. Then we can conclude from (134) that ϵ_j is independent of j, and therefore also E_j independent of j. Then

$$\frac{1}{\alpha M}\ge\beta\frac{1}{\beta\alpha M}\sum_{j=1}^{\beta\alpha M}Q\left(\sqrt{2E_j}+Q^{-1}\left(1-\epsilon_j\right)\right)+\frac{1}{2}\left(1-\beta\right) \tag{135}$$

$$\ge\beta Q\left(\sqrt{\frac{2E}{\alpha}}+Q^{-1}\left(\frac{1-\epsilon}{\alpha}\right)\right)+\frac{1}{2}\left(1-\beta\right). \tag{136}$$

Here, for sufficiently large E, the Q-function is less than half, so that $\beta=1$ is optimum. Thus

$$\frac{1}{M}\ge\alpha Q\left(\sqrt{\frac{2E}{\alpha}}+Q^{-1}\left(\frac{1-\epsilon}{\alpha}\right)\right) \ge\left(1-\epsilon\right)Q\left(\sqrt{\frac{2E}{\alpha}}+Q^{-1}\left(\frac{1-\epsilon}{\alpha}\right)\right) \tag{137}$$

as $\alpha\ge 1-\epsilon$ follows from (132). Note that (137) is very similar with (126).

We need to optimize the converse with respect to α; we can then use the same value for achievability. We let $z=Q^{-1}\left(\frac{\alpha+\epsilon-1}{\alpha}\right)$, $\alpha=\frac{1-\epsilon}{1-Q(z)}$, and then bound

$$\max_{z>0}\sqrt{\frac{2E}{1-\epsilon}\left(1-Q\left(z\right)\right)}-z. \tag{138}$$

To maximize the last expression with respect to z, we take the derivative and equate to zero. Let $K_0^{-1}=4\pi\left(1-\epsilon\right)$.

$$z^2=\ln\left(K_0E\right)-\ln\left(1-Q\left(z\right)\right) \tag{139}$$

$$=\ln\left(K_0E\right)-\ln\left(1-Q\left(\sqrt{\ln\left(K_0E\right)-\ln\left(1-Q\left(z\right)\right)}\right)\right) \tag{140}$$

$$=\ln\left(K_0E\right)-\ln\left(1-Q\left(\sqrt{\ln\left(K_0E\right)+o\left(1\right)}\right)\right). \tag{141}$$

In (140) we substitute for z in $Q\left(z\right)$ from equation (139) and in (141) we make use of the known fact that $z\to\infty$ (for we know that $\alpha\to 1-\epsilon$) as $E\to\infty$. Also note that $\lim_{E^{-1}\to 0}Q\left(z\right)\to 0$. Using the well known series $Q\left(x\right)=K_1e^{-\frac{x^2}{2}}\left(x^{-1}+o\left(x^{-1}\right)\right)$ for sufficiently large x, where $K_1^{-1}=\sqrt{2\pi}$ we have

$$z^2=\ln\left(K_0E\right)-\ln\left(1-\frac{K_1e^{-\ln K_0E+o(1)}}{\sqrt{\ln\left(K_0E\right)+o\left(1\right)}}\left[1+\varepsilon\left(E^{-1}\right)\right]\right) \tag{142}$$

$$=\ln\left(K_0E\right)+\frac{K_1K_0^{-1}E^{-1}}{\sqrt{\ln\left(K_0E\right)+o\left(1\right)}}\left[1+\varepsilon\left(E^{-1}\right)\right] \tag{143}$$

$$=\ln\left(K_0E\right)+K_1K_0^{-1}E^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right] \tag{144}$$

$$=\ln\left(K_0E\right)+o\left(1\right). \tag{145}$$

Let $K_2=\frac{2}{1-\epsilon}$. Now it is easy to verify that

$$Q\left(z\right)=K_1e^{-\ln\left(\sqrt{K_0E}\right)-K_12^{-1}K_0^{-1}E^{-1}\sqrt{\ln^{-1}(K_0E)}\left[1+\varepsilon\left(E^{-1}\right)\right]}\times\left(z^{-1}+o\left(z^{-1}\right)\right) \tag{146}$$

$$=\sqrt{\frac{K_1^2K_0^{-1}E^{-1}}{z^2}}\times\left(1-K_12^{-1}K_0^{-1}E^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right]\right) \tag{147}$$

$$=\sqrt{\frac{K_1^2K_0^{-1}E^{-1}}{\ln\left(K_0E\right)+K_1K_0^{-1}E^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right]}}\times\left(1-K_12^{-1}K_0^{-1}E^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right]\right) \tag{148}$$

$$=\sqrt{K_1^2K_0^{-1}E^{-1}\ln^{-1}\left(K_0E\right)}\times\left(1-K_12^{-1}K_0^{-1}E^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right]\right) \tag{149}$$

and thus we can write $EQ\left(z\right)=\sqrt{K_1^2K_0^{-1}E\ln^{-1}\left(K_0E\right)}+o\left(1\right)$. Furthermore,

$$\sqrt{E\left(1-Q\left(z\right)\right)z^2}=\sqrt{E\ln\left(K_0E\right)+K_1K_0^{-1}\sqrt{\ln^{-1}\left(K_0E\right)}\left[1+\varepsilon\left(E^{-1}\right)\right]}-\sqrt{2^{-1}K_1^2K_0^{-1}}\left[1+\varepsilon\left(E^{-1}\right)\right] \tag{150}$$

$$=\sqrt{E\ln\left(K_0E\right)}-\sqrt{2^{-1}K_1^2K_0^{-1}}\left[1+\varepsilon\left(E^{-1}\right)\right]. \tag{151}$$

and

$$\left[\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z\right]^2=K_2E-K_2EQ\left(z\right)+z^2-2\sqrt{K_2E\left(1-Q\left(z\right)\right)}z \tag{152}$$

$$=K_2E-K_2\sqrt{K_1^2K_0^{-1}E\ln^{-1}(K_0E)}-2\sqrt{K_2E\ln\left(K_0E\right)}-\sqrt{2K_2K_1^2K_0^{-1}} \tag{153}$$

$$+\ln\left(K_0E\right)+o\left(1\right). \tag{154}$$

Using the above expansions we have

$$\begin{aligned}&\log\left[\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z\right]\\&=\frac{1}{2}\log\left[\left(\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z\right)^2\right]\\&=2^{-1}\log\left[K_2E\left[1-O\left(\sqrt{E^{-1}\ln E}\right)\right]\right]\\&=2^{-1}\log\left(K_2E\right)+o\left(1\right).\end{aligned}\tag{155}$$

From the series expansion of the Q function ([27, (42)]), we get for the converse expression

$$\begin{aligned}&\log M\\&\leq 2^{-1}\left[K_2E-K_2\sqrt{K_1^2K_0^{-1}E\ln^{-1}\left(K_0E\right)}\right.\\&\left.-2\sqrt{K_2E\ln\left(K_0E\right)}\right]\log e\\&+\log E+O(1).\end{aligned}\tag{156}$$

$$\begin{aligned}&=\frac{E}{1-\epsilon}\log e\\&-\sqrt{\frac{2E}{1-\epsilon}\ln\frac{E}{4\pi\left(1-\epsilon\right)}}\left(1+\ln^{-1}\left(\frac{E}{4\pi\left(1-\epsilon\right)}\right)\right)\log e\\&+\log E+O(1).\end{aligned}\tag{157}$$

For achievability, we can write $O\left(\sqrt{\frac{\alpha}{E}}\right)=O\left(\sqrt{E^{-1}}\right)$ and

$$\begin{aligned}&\left[\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z+O\left(\sqrt{E^{-1}}\right)\right]^2\\&=\left[\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z\right]^2+O\left(E^{-1}\right)\\&+O\left(\sqrt{\left(1-Q\left(z\right)\right)}\right)-O\left(z\sqrt{E^{-1}}\right)\end{aligned}\tag{158}$$

$$=\left[\sqrt{K_2E\left(1-Q\left(z\right)\right)}-z\right]^2+O\left(1\right).\tag{159}$$

As above we now have

$$\begin{aligned}&\log M\\&\geq\frac{E}{1-\epsilon}\log e\\&-\sqrt{\frac{2E}{1-\epsilon}\ln\frac{E}{4\pi\left(1-\epsilon\right)}}\left(1+\ln^{-1}\left(\frac{E}{4\pi\left(1-\epsilon\right)}\right)\right)\log e\\&+\log\left(E\right)+O\left(1\right).\end{aligned}\tag{160}$$

Appendix D
Proof of Theorem 8

The energy lower bounds: Assume that π^* achieves the optimal delay-power tradeoff (δ,P), then the corresponding coding action (k_i,n_i,P_i,L_i) chosen by the encoder at each coding epoch satisfies (33)-(35). We start from constraint (33), i.e., $0\leq k_i/b\leq q$, and relax it to $k_i/b\geq 0$, which is one of the constraints put on $\mathbb{A}$ in (46) for $E_b^{LB}\left(d,\delta\right)$.

Since the above mentioned coding action satisfies (33), it automatically belongs to $\mathbb{A}$. Denote α_i the probability for the encoder to choose this coding action from $\mathbb{A}$, then

$$E_b\left(d,\delta\right)=\frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_in_iP_i\bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i}.\tag{161}$$

After a long enough period of t (that lasts for N coding epochs), the average number of messages that enter the buffer is

$$N\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i=\lambda_{msg}bt\tag{162}$$

and the average number of output symbols is $N\sum_{i=1}^{|\mathbb{A}|}\alpha_in_i\bar{l}_i$. Due to bandwidth constraint, we have

$$N\sum_{i=1}^{|\mathbb{A}|}\alpha_in_i\bar{l}_i\leq\frac{t}{T_c}.\tag{163}$$

Combining (162) and (163), we have

$$\frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_in_i\bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i}\leq\frac{1}{\lambda_{msg}bT_c}=\frac{1}{R_a}.\tag{164}$$

Now we consider the delay constraint. The waiting time consists of two parts: the waiting time for the arrival of the whole batch (k_i messages), and the waiting time for the start of the service. We only consider the first part, and then get a lower bound of waiting time, which is also a lower bound of the total time. Therefore, the average total time should be low bounded by $\frac{N\sum_{i=1}^{|\mathbb{A}|}\alpha_i\frac{1}{2\lambda_{msg}}\left(\frac{k_i}{b}-1\right)\frac{k_i}{b}}{\lambda_{msg}t}$.

On the other hand, the average total delay is constrained by d, hence we have

$$\frac{N\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i\left(\frac{k_i}{b}-1\right)}{2\lambda_{msg}^2bt}+\frac{N\sum_{i=1}^{|\mathbb{A}|}\alpha_in_i\bar{l}_iT_c\frac{k_i}{b}}{\lambda_{msg}t}\leq dT_s.\tag{165}$$

Combining (162) and (165) gives us

$$\frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i^2}{2\lambda_{msg}b\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i}-\frac{1}{2\lambda_{msg}}+\frac{T_c\sum_{i=1}^{|\mathbb{A}|}\alpha_in_i\bar{l}_ik_i}{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i}\leq dT_s.\tag{166}$$

Furthermore, the message success rate can be expressed as

$$p_s=\frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i(1-\epsilon_i^{L_i})}{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i},\tag{167}$$

which is constrained by $1-\delta$.

Therefore, combining the above, we have $\left(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|}\right)\in A_{d,\delta}^{LB}$, with $A_{d,\delta}^{LB}$ given in (43).

If we define

$$E_b^{LB}\left(d,\delta\right)=\inf_{\left(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|}\right)\in A_{d,\delta}^{LB}}\frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_in_iP_i\bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|}\alpha_ik_i},\tag{168}$$

then $E_b^{LB}\left(d,\delta\right)$ lower bounds $E_b\left(d,\delta\right)$, i.e.,

$$E_b(d,\delta)\geq E_b^{LB}(d,\delta).\tag{169}$$

The energy upper bound: We consider the following time-sharing strategy: *The encoder waits till the number in the queue reaches p_0k_0, then all the p_0k_0 messages are served (encoded and transmitted) in a batch with fixed $k=p_0k_0b$, n, P and $L=1$ such that $r<C$. After that, the encoder drops $\left(1-p_0\right)k_0$ messages successively. Then the encoder repeats these two operations indefinitely.*

Next we will show that the energy upper bound could be achieved by applying the above coding scheme to the original

system in Fig. 4. To prove this, we need to introduce another two queuing systems with infinite buffer as well. The first new system adopts a batch serving strategy: if the number of messages in the queue is less than k_0, the encoder waits till the number of messages in the queue reaches k_0; otherwise, it serves all the k_0 messages in a batch within a fixed service time of nT_c.

The second new system adopts a similar strategy: if the number of messages in the queue is less than k_0, the encoder waits till the number of messages in the queue reaches k_0; otherwise, it sends a message within a fixed service time of $\frac{nT_c}{k_0}$. Note that different from the first system, the second new system serves only *one* message at each time.

For the first new system, we consider the following two cases: a) consider the departure epoch of a whole batch as the departure epoch of each message in the batch; b) consider the "real" departure epoch of a message as the departure epoch of that message.

Denote the average delays for the original system, the first new system for case a), the first new system for case b), and the second new system as $d^{total}, d_{1,a}^{total}, d_{1,b}^{total}$, and d_2^{total}[8], respectively. Then we have the following Lemma

Lemma 1. *Consider three same message streams are fed into the original system and the two new systems, respectively. For the first new system, if we consider the case a), then the original system has a lower average delay than the first new system; and in turn, if we consider the case b), the first new system has a lower average delay than the second new system. That is,*

$$d^{total} \le d_{1,a}^{total} \tag{170}$$

$$d_{1,b}^{total} \le d_2^{total}. \tag{171}$$

Proof: Obviously the original system has a lower average delay than the first new system, since it does not need to wait all the k_0 messages before starting service. Next we show the first new system has a lower average delay than the second new system.

We claim that all the messages depart from the first new system earlier than from the second new system.

We use mathematical induction to prove the above claim. Denote the arrival epoch of the ith message as t_i^a, and the departure epochs for the first and second new system as t_i^{d1} and t_i^{d2}, respectively. We also assume that both the systems are initially empty. Next we will show $t_i^{d1} \le t_i^{d2}$.

1) Before the first k_0 messages arrive, both systems are in the idle state. After they arrive and enter the buffers, the second system serves the first message, and the first system serves the whole batch (of all k_0 messages). Hence the departure epoch of the first message is $t_{k_0}^a + \Delta T, \Delta T \triangleq \frac{nT_c}{k_0}$ for both the systems, i.e., $t_1^{d1} = t_1^{d2}$.

2) Assume that $t_i^{d1} \le t_i^{d2}, i \ge 1$ holds. To prove $t_{i+1}^{d1} \le t_{i+1}^{d2}$, we need to enumerate all possible cases for the original system.

[8]All of them denote real (unnormalized) delays.

Case 1: If the ith message and the $(i+1)$th message are served in the same batch in the first system, then $t_{i+1}^{d1} = t_i^{d1} + \Delta T$. Hence $t_{i+1}^{d2} \ge t_i^{d2} + \Delta T \ge t_i^{d1} + \Delta T = t_{i+1}^{d1}$.

Case 2: If the ith message and the $(i+1)$th message are served in the different batches in the first system, then $t_{i+1}^{d1} = \max\left\{t_i^{d1} + \Delta T, t_{i+k_0}^a + \Delta T\right\}$. Hence $t_{i+1}^{d2} = \max\left\{t_i^{d2} + \Delta T, t_{i+k_0}^a + \Delta T\right\} \ge t_{i+1}^{d1}$.

Combining both the cases above, we have $t_{i+1}^{d1} \le t_{i+1}^{d2}$. Therefore, $t_i^{d1} \le t_i^{d2}$ for all $i \ge 1$, i.e., all the messages depart from the original system earlier than from the new system. This implies that the original system has lower average delay. ∎

On the other hand, for the second new system we divide the buffer into two segments: the first consisting of the first $k_0 - 1$ message space, and the second consisting of the other (infinite) message space. Then by observing the strategy applied to the second new system, we can find that the first segment will be always full after $k_0 - 1$ message arrivals and that the second segment is equivalent to an M/D/1 queuing system with arrival rate λ_{msg} and service time $\frac{nT_c}{k_0}$. After a sufficiently long period of time t, the incurred average delay d_2^{total} for the second new system can be expressed as

$$d_2^{total} = \frac{(k_0-1)\,t + \lambda_{msg} t\left(d_2 + \frac{nT_c}{k_0}\right)}{\lambda_{msg} t} = \frac{k_0-1}{\lambda_{msg}} + d_2 + \frac{nT_c}{k_0} \tag{172}$$

where $d_2 = \dfrac{\rho}{2\mu(1-\rho)} = \dfrac{\rho^2}{2\lambda_{msg}(1-\rho)}$, with $\frac{1}{\mu} = \frac{nT_c}{k_0}$ and $\rho = \frac{\lambda_{msg} nT_c}{k_0} = \frac{R_a p_0}{r}$, is the average *waiting* delay of the M/D/1 queuing system. To guarantee stability of the system, $\rho < 1$ should hold.

For the first new system, we can observe that the average delay for cases a) and b) are only different in service (transmission) delay part. Moreover, we can express them as

$$d_{1,a}^{total} = d_1^{wait} + d_{1,a}^{serv} \tag{173}$$

$$d_{1,b}^{total} = d_1^{wait} + d_{1,b}^{serv}, \tag{174}$$

where d_1^{wait} denotes the waiting delay for both cases a) and b), and

$$d_{1,a}^{serv} = nT_c \tag{175}$$

$$d_{1,b}^{serv} = \frac{k_0+1}{2}\frac{nT_c}{k_0} \tag{176}$$

denote the service delay for cases a) and b), respectively.

From (170)-(176) we have

$$d^{total} \le d_{1,a}^{total} = d_1^{wait} + d_{1,a}^{serv} \tag{177}$$

$$= d_{1,b}^{total} - d_{1,b}^{serv} + d_{1,a}^{serv} \tag{178}$$

$$\le d_2^{total} - d_{1,b}^{serv} + d_{1,a}^{serv} \tag{179}$$

$$= \frac{k_0-1}{\lambda_{msg}} + d_2 + \frac{nT_c}{k_0} - \frac{k_0+1}{2}\frac{nT_c}{k_0} + nT_c \tag{180}$$

$$= \frac{k_0}{\lambda_{msg}}\left(1 + \frac{\rho}{2}\right) + \frac{3\rho - 2}{2\lambda_{msg}(1-\rho)}. \tag{181}$$

In addition, the message delivery success rate of the original system is

$$p_s = p_0\left(1-\epsilon\right). \tag{182}$$

and the energy per bit is

$$E_{b,0} = \frac{nP}{k_0 b} = \frac{p_0 P}{r}. \tag{183}$$

Hence we have

$$E_b\left(d,\delta\right) \leq \min_{\left(k,n,P,p_0\right)\in A_{d,\delta}^{UB}} \frac{p_0 P}{r}. \tag{184}$$

Appendix E
Proof of Theorem 9

The energy is minimized by maximizing block length, so we set $k_0 = \frac{\lambda_{msg} dT_s}{1+\frac{\rho}{2}} - \frac{3\rho-2}{2\left(1+\frac{\rho}{2}\right)(1-\rho)}$ in (44)[9] to get

$$n = \frac{k}{r} = \frac{b}{R_a p_0} k_0 \rho = \frac{d}{R_a p_0}\frac{\rho}{1+\frac{\rho}{2}} - \frac{b}{R_a p_0}\frac{\left(3\rho-2\right)\rho}{2\left(1+\frac{\rho}{2}\right)\left(1-\rho\right)}. \tag{185}$$

As in Appendix A, we now substitute n and r in

$$r = C\left(\frac{P_0+\triangle P}{1-\epsilon}\right) - \sqrt{V\left(\frac{P_0+\triangle P}{1-\epsilon}\right)}\sqrt{\frac{\log n}{n}} + \frac{M}{\sqrt{n}} \tag{186}$$

where $P = P_0 + \triangle P$ and $\frac{P_0}{1-\epsilon} = 2^{2R_a p_0} - 1$. Since $\Delta P \to 0$ as $n \to \infty$ $(d \to \infty)$ we have

$$\frac{R_a p_0}{\rho\left(d\right)} = C_0 - V_0 + \left[C_1 - V_1\right]\triangle P + o\left(\triangle P\right) + \frac{M}{\sqrt{n\left(d\right)}} \tag{187}$$

where $C_0 = C\left(\frac{P_0}{1-\epsilon}\right), C_1 = \frac{1}{2\ln 2(1-\epsilon+P_0)}$, $V_0 = \sqrt{V\left(\frac{P_0}{1-\epsilon}\right)\log\left(n\right)n^{-1}}$ and $V_1 = \frac{(\epsilon-1)^2\sqrt{\log(n)n^{-1}}}{\sqrt{2}\ln 2\sqrt{P_0(-2\epsilon+P_0+2)}(-\epsilon+P_0+1)^2}$. Solving for ΔP in (187) gives us

$$\triangle P\left(1+\varepsilon\left(\Delta P\right)\right) = C_1^{-1}\left(\frac{R_a p_0}{\rho\left(d\right)} - C_0\right) + C_1^{-1}V_0 + \left(\frac{R_a p_0}{\rho\left(d\right)} - C_0\right)C_1^{-2}V_1 + o\left(\sqrt{n^{-1}\log n}\right).$$

We will now choose an optimum $\rho\left(d\right)$ such that both achievable $E_b\left(d,\delta\right)$ is minimized and simultaneously ensuring that $n \to \infty$ as $d \to \infty$ so that (186) can be used in the limit. The latter condition would also ascertain that both $V_0, V_1 \to 0$ as $d \to \infty$. Using the above equation we have

$$\begin{aligned} E_b\left(d,\delta\right) &= \frac{p_0 P}{2r} = \frac{\rho\left(d\right)}{2R_a}\left(P_0+\Delta P\right) \\ &= 2^{-1}C_1^{-1}R_a^{-1}\Bigg[\rho C_1 P_0 + \left(R_a p_0 - C_0\rho\right) + V_0\rho \\ &\quad + \left(R_a p_0 - C_0\rho\right)C_1^{-1}V_1 + o\left(\sqrt{n^{-1}\log n}\right)\Bigg]. \end{aligned}$$

[9]As $k_0 \to \infty$, the integer constraint can be ignored

Optimizing $E_b\left(d,\delta\right)$ for sufficiently large n (equivalently d), is equivalent to selecting a suitable $\rho\left(d\right)$ such that

$$h\left(P_0,\rho\left(d\right)\right) := \frac{R_a p_0}{2C_1 R_a} + \frac{P_0 - C_1^{-1}C_0}{2R_a}\rho\left(d\right) \tag{188}$$

is minimized. As

$$C_1^{-1}C_0 - P_0 = \left(1-\epsilon\right)2^{2y}\left(y2\ln 2 - 1 + 2^{-2y}\right) \geq 0,\ y = R_a p_0 \tag{189}$$

this means our choice for $\rho\left(d\right) \to 1$ as $d \to \infty$. With this insight, let $\rho\left(d\right) = 1 - \alpha\left(d\right)$ where $\alpha\left(d\right) \in \left(0,1\right)$ and $\alpha\left(d\right) \to 0$ as $d \to \infty$. This gives us $n\left(d\right) = \frac{1-\alpha}{3-\alpha}\left[\frac{b}{R_a p_0}\frac{(3\alpha-1)}{\alpha} + \frac{2}{R_a p_0}d\right]$ and

$$E_b\left(d,\delta\right) = 2^{-1}C_1^{-1}R_a^{-1}\Bigg[C_1 P_0 + \left(C_1^{-1}C_0 - P_0\right)\alpha\left(d\right) + V_0 + \left[C_0 C_1^{-1}V_1 - V_0\right]\alpha\left(d\right) + o\left(\sqrt{n^{-1}\log n}\right)\Bigg].$$

Finally, we now select $\alpha(d) \in \varepsilon\left(d^{-1}\right)$ such that both $o\left(\sqrt{n^{-1}\log n}\right)$ and $\alpha\left(d\right)$ get absorbed into an $o\left(\sqrt{d^{-1}\log d}\right)$ term – otherwise, these terms contribute positively to energy, i.e., make the energy larger. This is satisfied by choosing $\alpha\left(d\right) = d^{-1}\log d$ giving us $\sqrt{n^{-1}\log n} = \sqrt{\frac{3}{2}R_a d^{-1}\log d} + o\left(\sqrt{d^{-1}\log d}\right)$ and $V_1\alpha\left(d\right), V_0\alpha\left(d\right) \in o\left(\sqrt{d^{-1}\log d}\right)$. Thus

$$\begin{aligned} \frac{E_b\left(d,\delta\right)}{1-\epsilon} &= \frac{2^{2R_a p_0}-1}{2R_a} \\ &\quad + \frac{2^{2R_a p_0}}{\log e}\sqrt{\frac{3}{2R_a}V\left(2^{2R_a p_0}-1\right)d^{-1}\log d} \\ &\quad + o\left(\sqrt{d^{-1}\log d}\right). \end{aligned}$$

This completes the proof.

Appendix F
Proof of Theorem 10

Lower Bounds: Denote $\mathbb{A}^+ = \left\{i : 1 \leq i \leq \left|\mathbb{A}\right|, n_i > 0\right\}$. For any $i \notin \mathbb{A}^+$, it holds that $n_i = L_i = 0$. Hence,

$$\begin{aligned} A_{\infty,\delta}^{LB} &= \lim_{d\to\infty} A_{d,\delta}^{LB} = \\ &\Bigg\{\left(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|}\right) : 0 \leq \alpha_i \leq 1, \text{for}\, 1 \leq i \leq \left|\mathbb{A}\right|, \\ &\sum_{i=1}^{|\mathbb{A}|}\alpha_i = 1, \\ &\frac{\sum_{i\in\mathbb{A}^+}\alpha_i n_i \bar{l}_i}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \leq \frac{1}{R_a}, \\ &\frac{\sum_{i\in\mathbb{A}^+}\alpha_i k_i\left(1-\epsilon_i^{L_i}\right)}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \geq 1-\delta\Bigg\}. \end{aligned} \tag{190}$$

On the other hand,

$$\lim_{d\to\infty} E_b^{LB}(d,\delta) \geq \inf_{(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|})\in A_{\infty,\delta}^{LB}} \frac{\sum_{i\in\mathbb{A}^+}\alpha_i n_i \bar{l}_i(1-\epsilon_i)\frac{N_0}{2}\left(2^{2r_i}-1\right)}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \tag{191}$$

$$= \inf_{(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|})\in A_{\infty,\delta}^{LB}} \frac{\frac{N_0}{2}\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})\left(2^{2r_i}-1\right)}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \tag{192}$$

$$\geq \inf_{(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|})\in A_{\infty,\delta}^{LB}} \frac{\frac{N_0}{2}\left(\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})\right)\left(2^{\frac{\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})r_i}{\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})}}-1\right)}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \tag{193}$$

$$\geq \inf_{(\alpha_1,\alpha_2,\cdots,\alpha_{|\mathbb{A}|})\in A_{\infty,\delta}^{LB}} \frac{\frac{N_0}{2}\left(\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})\right)\left(2^{\frac{(1-\delta)\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i}{\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})}}-1\right)}{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i} \tag{194}$$

$$= \inf_t \frac{N_0}{2}\frac{2^{2(1-\delta)t}-1}{t}, \tag{195}$$

where

$$t = \frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i}{\sum_{i\in\mathbb{A}^+}\alpha_i n_i(1-\epsilon_i^{L_i})} \geq \frac{\sum_{i=1}^{|\mathbb{A}|}\alpha_i k_i}{\sum_{i\in\mathbb{A}^+}\alpha_i n_i \bar{l}_i} \geq R_a, \tag{196}$$

and (193) follows from the fact that $2^{2x}-1$ is convex in x. Since $\frac{2^{2t}-1}{t}$ is increasing in t, by (195) we have

$$\lim_{\tau\to\infty} E_b^{LB}(\tau,\delta) \geq \frac{N_0}{2R_a}\left(2^{2R_a(1-\delta)}-1\right). \tag{197}$$

Upper Bounds: Fix r and P, since $r < C$, let $n \to \infty$ ($d \to \infty$), then $\epsilon \to 0$. Therefore,

$$\begin{aligned}
\lim_{d\to\infty} E_b^{UB}(d,\delta) &\leq \lim_{\tau\to\infty}\inf \frac{p_0 P}{r} \\
\text{over} \quad & (r,n,P) \\
\text{subject to} \quad & p_0 = \frac{1-\delta}{1-\epsilon}, \\
& k = bp_0k_0, k_0 \in \mathbb{N}, \\
& r < C, \\
& \rho = \frac{R_a p_0}{r} < 1, \\
& \frac{k_0}{\lambda_{msg}}\left(1+\frac{\rho}{2}\right) + \frac{3\rho-2}{2\lambda_{msg}(1-\rho)} \leq dT_s \\
= & \inf_{r,P} \frac{(1-\delta)P}{r} \\
\text{subject to} \quad & R_a p_0 < r < C \\
\leq & \inf_r \frac{(1-\delta)N_0}{2r}\left(2^{2r}-1\right) \\
\text{subject to} \quad & r > R_a p_0(1-\delta) \\
= & \frac{N_0}{2R_a}\left(2^{2R_a p_0(1-\delta)}-1\right).
\end{aligned} \tag{198}$$

In addition,

$$\lim_{d\to\infty} E_b^{UB}(d,\delta) \geq \lim_{d\to\infty} E_b^{LB}(d,\delta). \tag{199}$$

Combining this with (197) and (198) gives us

$$\lim_{d\to\infty} E_b^{UB}(d,\delta) = \lim_{d\to\infty} E_b^{LB}(d,\delta) = \frac{N_0}{2R_a}\left(2^{2R_a p_0(1-\delta)}-1\right). \tag{200}$$

REFERENCES

[1] G. Cook, "How clean is your cloud?" Greenpeace, Amsterdam, The Netherlands, Tech. Rep., 2012. [Online]. Available: http://www.greenpeace.org/international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf

[2] *The Power of Wireless Cloud*, Bell Labs, CEET, Univ. Melbourne, Melbourne, VIC, Australia, 2013. [Online]. Available: http://www.ceet.unimelb.edu.au/publications/ceet-white-paper-wireless-cloud.pdf

[3] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[4] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 1999, pp. 709–717.

[5] P. Nuggehalli, V. Srinivasan, and R. R. Rao, "Delay constrained energy efficient transmission strategies for wireless devices," in *Proc. IEEE INFOCOM*, vol. 3, Jun. 2002, pp. 1765–1772.

[6] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.

[7] B. Prabhakar, E. U. Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. INFOCOM*, vol. 1, Apr. 2001, pp. 386–394.

[8] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 487–499, Aug. 2002.

[9] W. Chen, M. J. Neely, and U. Mitra, "Energy-efficient transmissions with individual packet delay constraints," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2090–2109, May 2008.

[10] M. J. Neely, "Delay analysis for max weight opportunistic scheduling in wireless systems," *IEEE Trans. Autom. Control*, vol. 54, no. 9, pp. 2137–2150, Sep. 2009.

[11] X. Zhong and C.-Z. Xu, "Delay-constrained energy-efficient wireless packet scheduling with QoS guarantees," in *Proc. GLOBECOM*, vol. 6, Nov./Dec. 2005, pp. 3336–3340.

[12] M. A. Khojastepour and A. Sabharwal, "Delay-constrained scheduling: Power efficiency, filter design, and bounds," in *Proc. INFOCOM*, vol. 3, Mar. 2004, pp. 1938–1949.

[13] M. A. Zafer and E. Modiano, "A calculus approach to minimum energy transmission policies with quality of service guarantees," in *Proc. IEEE INFOCOM*, vol. 1, Mar. 2005, pp. 548–559.

[14] N. Abuzainab and A. Ephremides, "Energy/delay tradeoffs in data transmission over a time varying wireless link," in *Proc. Allerton*, Sep. 2011, pp. 1326–1333.

[15] R. A. Berry, "Optimal power-delay tradeoffs in fading channels—Small-delay asymptotics," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3939–3952, Jun. 2013.

[16] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proc. INFOCOM*, vol. 2, Mar./Apr. 2003, pp. 1095–1105.

[17] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1866–1875, Apr. 2009.

[18] A. Sharifkhani and N. C. Beaulieu, "Dynamic power allocation over block-fading channels with delay constraint," in *Proc. GLOBECOM*, Nov./Dec. 2009, pp. 1–7.
[19] M. Zafer and E. Modiano, "Delay-constrained energy efficient data transmission over a wireless fading channel," in *Proc. ITA*, Jan./Feb. 2007, pp. 289–298.
[20] S. Zuo, H. Deng, and I.-H. Hou, "Energy efficient algorithms for real-time traffic over fading wireless channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1881–1892, Mar. 2017.
[21] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
[22] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. 37, no. 1, pp. 10–21, Jan. 1949.
[23] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
[24] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
[25] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.
[26] R. G. Gallager and B. Nakiboglu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.
[27] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4880–4902, Aug. 2011.
[28] T. Javidi and A. Goldsmith, "Dynamic joint source-channel coding with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 16–20.
[29] S. Goel and R. Negi, "The queued-code in finite-state Markov fading channels with large delay bounds," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, USA, Jul. 2006, pp. 30–34.
[30] J. Lee and N. Jindal, "Asymptotically optimal policies for hard-deadline scheduling over fading channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2482–2500, Apr. 2013.
[31] J. Cao and E. M. Yeh, "Power-delay tradeoff analysis for communication over fading channels with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 614–618.
[32] S. Xu, T.-H. Chang, S.-C. Lin, C. Shen, and G. Zhu, "Energy-efficient packet scheduling with finite blocklength codes: Convexity analysis and efficient algorithms," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5527–5540, Aug. 2016.
[33] V. Kostina, Y. Polyanskiy, and S. Verdú, "Joint source-channel coding with feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 3502–3515, Jun. 2017.
[34] T. Holliday, A. Goldsmith, and H. V. Poor, "The impact of delay on the diversity, multiplexing, and ARQ tradeoff," in *Proc. IEEE ICC*, vol. 4, Jun. 2006, pp. 1445–1449.
[35] H. E. Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.
[36] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with Bursty and delay-limited information," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 746–763, Feb. 2009.
[37] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2005.
[38] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "A beta-beta achievability bound with applications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2669–2673.
[39] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2008.
[40] A. R. Williamson, T.-Y. Chen, and R. D. Wesel, "Reliability-based error detection for feedback communication with low latency," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2552–2556.
[41] J. C. Fricke and P. A. Hoeher, "Reliability-based retransmission criteria for hybrid ARQ," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2181–2184, Aug. 2009.
[42] U. Erez, M. D. Trott, and G. W. Wornell, "Rateless coding for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 530–547, Feb. 2012.
[43] E. Altman and A. Shwartz, "Time-sharing policies for controlled Markov chains," *Oper. Res.*, vol. 41, no. 6, pp. 1116–1124, 1993.
[44] R. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inf. Theory*, vol. 9, no. 4, pp. 64–74, Apr. 1963.
[45] S. V. Maiya, D. J. Costello, and T. E. Fuja, "Low latency coding: Convolutional codes vs. LDPC codes," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1215–1225, May 2012.
[46] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," 2013, *arXiv:1311.2337*. [Online]. Available: https://arxiv.org/abs/1311.2337
[47] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, Sep. 2015.

Mirza Uzair Baig received the B.S. degree in electrical engineering from the Lahore University of Management Sciences, Pakistan, in 2012, and the M.A. degree in mathematics and the Ph.D. degree in electrical engineering from the University of Hawai'i, Honolulu, HI, USA, in 2017 and 2019, respectively.

Since August 2019, he has been working as an Experienced Researcher at Ericsson AB, Sweden. His main research interests include information theory, mathematical theories, and their application to interference networks and wireless communications.

Lei Yu received the B.E. and Ph.D. degrees from the University of Science and Technology of China (USTC), in 2010 and 2015, respectively, both in electronic engineering. From 2015 to 2017, he was a Postdoctoral Researcher at the Department of Electronic Engineering and Information Science (EEIS), USTC. He is currently a Research Fellow at the Department of Electrical and Computer Engineering, National University of Singapore. His research interests lie in the intersection of information theory, probability theory, and combinatorics.

Zixiang Xiong (S'91–M'96–SM'02–F'07) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1996.

Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University, where he is currently a Professor and also an Associate Department Head. His main research interest lies in image/video processing, computer vision, virtual/augmented reality, big data, and communications. He received an NSF Career Award in 1999, an ARO Young Investigator Award in 2000, and an ONR Young Investigator Award in 2001. He is a co-recipient of the 2006 IEEE Signal Processing Magazine Best Paper Award, top 10% paper awards at the 2011 and 2015 IEEE Multimedia Signal Processing Workshops, an IBM Best Student Paper Award at the 2016 IEEE International Conference on Pattern Recognition, and the Best Demo Paper Award at the 2018 IEEE International Conference on Multimedia and Expo. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (1999–2005), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002–2005), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002–2006), the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (Part B) (2005–2009), and the IEEE TRANSACTIONS ON COMMUNICATIONS (2008–2013). He is currently an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA.

Anders Høst-Madsen (M'95–SM'02–F'13) received the M.Sc. degree in engineering and the Ph.D. degree in mathematics from the Technical University of Denmark, in 1990 and 1993, respectively.

From 1993 to 1996, he was with Dantec Measurement Technology A/S, Copenhagen, Denmark. From 1996 to 1998, he was an Assistant Professor at the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. From 1998 to 2000, he was an Assistant Professor at the University of Calgary, Calgary, AB, Canada, and also a Staff Scientist at TRLabs, Calgary. Since 2001, he has been with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, and has been a Professor since 2009. He was a Visiting Professor at Seoul National University in 2017, and at the Shenzhen Research Institute of Big Data (SRIBD), The Chinese University of Hong Kong, Shenzhen, in 2018 and 2019, respectively. He was a Founder and CTO of Kai Medical, Inc., from 2007 to 2008, which is making equipment for non-contact heart monitoring. His research interests are in statistical signal processing, data science, information theory, and wireless communications, including ad hoc networks, wireless sensor networks, heart monitoring, marine mammal signal processing, big data, and learning theory.

Dr. Høst-Madsen received the *Eurasip Journal of Wireless Communications and Networks* (JWCN) Best Paper Award in 2006, and the College of Engineering Faculty Research Award in 2019. He was the General Co-Chair of ISITA 2012 and IEEE ISIT 2014. He has served as an Editor for *Multiuser Communications* for the IEEE TRANSACTIONS ON COMMUNICATIONS and as an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Houqiang Li (SM'12) received the B.S., M.Eng., and Ph.D. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 1992, 1997, and 2000, respectively, where he is currently a Professor with the Department of Electronic Engineering and Information Science. His research interests include video coding and communication, multimedia search, and image/video analysis. He was a recipient of the Best Paper Award for Visual Communications and Image Processing Conference in 2012. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2010 to 2013, and has been with the Editorial Board of the *Journal of Multimedia* since 2009.

Weiping Li (F'00) received the B.S. degree from the University of Science and Technology of China (USTC) in 1982, and the M.S. and Ph.D. degrees from Stanford University, in 1983 and 1988, respectively, all in electrical engineering. He was an Assistant Professor, Associate Professor with Tenure, and a Professor of Lehigh University from 1987 to 2001. He worked in several high-tech companies in the Silicon Valley with technical and management responsibilities from 1998 to 2010. He has been a Professor with USTC since 2010. He has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, a founding member of the Board of Directors of MPEG-4 Industry Forum, and several other positions in the IEEE and SPIE.