Information Velocity of Cascaded Gaussian Channels With Feedback

Elad Domanovitz, *Member, IEEE*, Anatoly Khina, *Member, IEEE*, Tal Philosof, and Yuval Kochman, *Member, IEEE*

***Abstract*—We consider a line network of nodes, connected by additive white noise channels, equipped with local feedback. We study the velocity at which information spreads over this network. For transmission of a data packet, we give an explicit positive lower bound on the velocity, for any packet size. Furthermore, we consider streaming, that is, transmission of data packets generated at a given average arrival rate. We show that a positive velocity exists as long as the arrival rate is below the individual Gaussian channel capacity, and provide an explicit lower bound. Our analysis involves applying pulse-amplitude modulation to the data (successively in the streaming case), and using linear mean-squared error estimation at the network nodes. For general white noise, we derive exponential error-probability bounds. For single-packet transmission over channels with (sub-)Gaussian noise, we show a doubly-exponential behavior, which reduces to the celebrated Schalkwijk–Kailath scheme when considering a single node. Viewing the constellation as an "analog source", we also provide bounds on the exponential decay of the mean-squared error of source transmission over the network.**

***Index Terms*—Information velocity, relay networks, combined source–channel coding, Gaussian channels, low-latency communication.**

I. Introduction

The vision of ubiquitous wireless connectivity and the Internet of Things calls for developing low-latency communication technology for growing networks of smaller interconnected units. A central problem of interest in this context, is transmission over a cascade of channels interconnected by relaying nodes. Beyond serving as a building block for more complex networks, such line networks are found, for example, in saltatory conduction in the brain, where action potentials propagate along myelinated axons with the nodes of Ranvier serving as relays [2], [3], [4], [5], as well as in vehicle platooning, where wireless communication between geographically distant vehicles that travel together in a coordinated fashion is carried over intermediate vehicles which act as relays [6], [7].

From an information-theoretic perspective, without delay constraints, assuming that each channel may be used the same number of times, the maximal reliable communication rate is equal to the minimum of the individual channel capacities. In a more practical scenario, the end-to-end delay of the network is constrained. Since it is wasteful for a node to wait to decode a long block code, the nodes should opt to apply causal operations to their measurements instead. However, determining the maximal rate of reliable communication over such networks, let alone determining the error probability behavior, turns out to be very challenging.

If we fix the message size and number of channels and let the number of time steps grow, decoding with high probability is guaranteed, and one seeks the optimal error exponent. Determining the optimal error exponent turned out to be difficult even for the simple case of single-bit transmission over a tandem of binary symmetric channels [8], [9], and was eventually proved by Ling and Scarlett [10] to equal to that of a single channel. They further extended the scope to any finite number of messages in [11].

The behavior of large networks is better expressed by taking the number of channels to grow linearly with the number of time steps. The maximum ratio between the two, such that the error probability may be arbitrarily small, was termed *Information Velocity* (IV) by Polyanskiy (see [8], [12]) in the single-bit context. The same term was used earlier by Iyer and Vaze [13] in a related setting of spatial wireless networks.[1]

Specifically, for the case of transmitting a single bit B, we denote its estimation in relay r at time t as $\hat{B}_r(t)$. Hence, The error probability of this bit at relay r at time t is $P_e(r,t) \triangleq \Pr\left(\hat{B}_r(t) \neq B\right)$. The IV for transmitting this bit is the maximal v such that

$$\lim_{r\to\infty} P_e\left(r, \left\lfloor \frac{r}{v} \right\rfloor\right) = 0. \tag{1}$$

This definition can be extended for the transmission of a single packet (any fixed number of bits) and for the transmission of an infinite stream of bits which we present below.

[1] Similar concepts also exist in other disciplines, e.g., in physics, in neuroscience, epidemic spread in networks, and in marketing and finance.

Manuscript received 23 November 2023; revised 14 April 2024; accepted 30 May 2024. Date of publication 18 June 2024; date of current version 2 August 2024. This work was supported by the Israel Science Foundation under Grant 2077/20 and Grant 1558/17. The work of Elad Domanovitz and Anatoly Khina was supported by the Tel Aviv University Center for AI and Data Science (TAD). The work of Yuval Kochman was supported by the Quantum Communication Consortium through the Israel Ministry of Economy and Industry. This work was presented in part at the 2024 IEEE International Symposium on Information Theory, Athens, Greece, July 2024 [1]. *(Corresponding author: Anatoly Khina.)*

Elad Domanovitz and Anatoly Khina are with the School of Electrical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: domanovi@eng.tau.ac.il; anatolyk@eng.tau.ac.il).

Tal Philosof is with Samsung Research, Tel Aviv 6492103, Israel (e-mail: tal.philosof@gmail.com).

Yuval Kochman is with the Rachel and Salim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel (e-mail: yuvalko@cs.huji.ac.il).

Digital Object Identifier 10.1109/JSAIT.2024.3416310

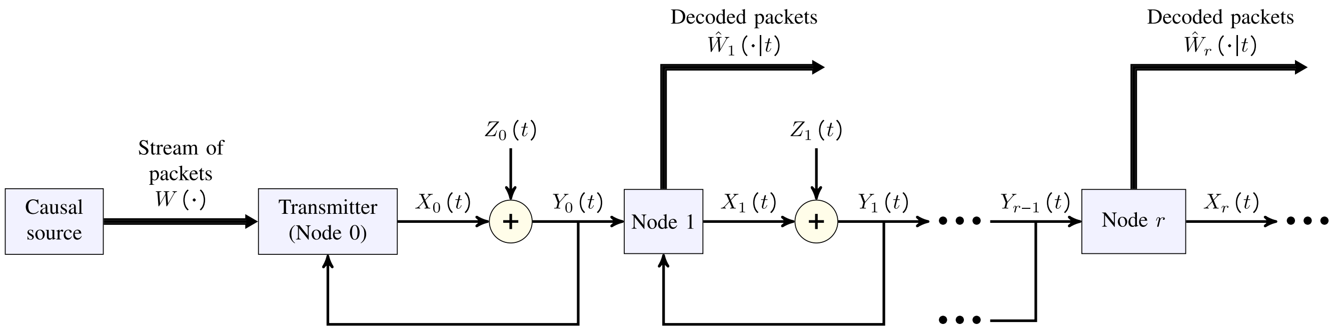

Fig. 1. Block diagram of the system. $X_r(t)$, $Y_r(t)$, and $Z_r(t)$ are the channel input, output, and noise, respectively, at node r at time t. Each T time steps, a new packet is generated. At every time step, all the nodes decode all the hitherto arrived packets.

It is not a priori clear whether such a positive IV exists at all for a given network. Rajagopalan and Schulman [12], while considering protocol simulation, answered this in the affirmative for a finite number of bits transmitted over binary symmetric channels (and hence also for any binary-input output symmetric channels via slicing). Improving upon this work, Ling and Scarlett [14] gave tighter upper and lower bounds in the limits of crossover probabilities that go to 0 or $1/2$.

Analyzing the IV of the transmission of a single message does not capture the entire behavior of the network. As the IV is meaningful when both the time and number of relays go to infinity, it might suggest that when another message arrives, the network is still occupied with processing the previous message. Hence, analyzing the IV while transmitting an infinite stream of messages is of interest. For transmission of a stream of messages over a cascade of packet-erasure channels (with instantaneous ACK/NACK feedback), the existence of a positive IV can be derived from stochastic network calculus [15], [16], [17], [18] (these results also hold for a single packet without feedback). However, only recently, in [19], the explicit IV was derived for streaming over packet-erasure channels with such feedback. This work also derived the IV of transmitting a single packet and determined the explicit error decay rate for velocities below the IV for both single-packet and streaming settings.

Despite these efforts, no explicit expressions for the IV for non-erasure channels are available even for a single bit, let alone for streaming or error probability decay rate guarantees.

In this work, we consider a cascade of additive white-noise (not necessarily Gaussian) channels with the same finite signal-to-noise ratio (SNR) $P > 0$, equipped with perfect[2] and local feedback. That is, each node is aware of the measurements of the next node in a causal manner; see Figure 1. For this setting, we derive a lower bound on both the single-packet and the streaming IVs:[3] We prove that the single-packet IV is bounded from below as

$$\mathrm{V} \geq 1 - \exp\{-2C\}, \tag{2}$$

where $C \triangleq \frac{1}{2}\log(1 + P)$ denotes the channel capacity of an individual additive white-Gaussian noise (AWGN) channel with SNR P;[4] the same bound (2) has been independently derived by Inovan [20, Ch. 5]. The streaming IV $\mathrm{V}(R)$ at data-generation rate R is bounded from below as

$$\mathrm{V}(R) \geq 1 - \exp\{-2(C - R)\} \tag{3}$$

for $R < C$; in particular, the streaming IV is positive for any $R < C$, and the bound on the streaming IV (3) reduces to that on the single-packet IV (2) in the limit of $R \to 0$.

In the context of streaming, our approach is based on the source node converting the incremental message history into a real number, as in [21]. Subsequently, this number is handled by the network in an analog linear manner. Namely, each node keeps the best linear estimator of that number based on its past measurements, as well as the estimate of the next node (known thanks to the feedback), and transmits a scaled version of the difference. Indeed, for a single data packet, the communication between the source node and the first relay reduces to the celebrated Schalkwijk–Kailath scheme for feedback communication over an additive-noise channel [22], [23], which can be viewed [24] as an application of the joint source–channel coding scheme with feedback of Elias [25]. On the other hand, the first transmission of each of the nodes reduces to scalar amplify-and-forward relaying [26].

Treating a representation of a data message as an analog source is a concept that has proved useful in relaying schemes [26], [27], [28]. Using that concept, techniques from source coding and joint source–channel coding find their way into digital communication settings. In the context of our work, the estimation at the relays are reminiscent of the Gaussian CEO problem [29], [30], [31], and in particular its sequential variant [32], [33], [34].

Due to the amount of concepts involved, we present our results gradually. We first address a single source to be transmitted, and then streaming. Within each of the above, we start with an analog source and consider the estimation mean-squared error (MSE), and in particular its decay rate, before advancing to data packets and considering error probabilities. For error probabilities, we show that for velocities lower than our achievable bound on the IV, the error probabilities

[2] Of course, the assumption of perfect feedback is not realistic. We view it as a first step towards more practical settings.

[3] These results are stated here for delayed hops, that is, a relay may only use its input in the subsequent time step. In the body of the paper we consider instantaneous hops for ease of presentation; the results for both delayed and instantaneous hops are summarized in Table I in Section VII.

[4] While our results are expressed in terms of the AWGN capacity, they hold for general additive noise.

decay at least exponentially fast, while if the noise is assumed to be Gaussian (or, more generally, sub-Gaussian), the error probabilities are shown to decay at least double exponentially, extending the known behavior of the Schalkwijk–Kailath scheme [22], [23] to multiple relays.

It should be noted that the extension from a single data packet to streaming is non-trivial. We think of an ultimate virtual "analog source" that represents the infinite stream of data bits, including those that are yet to be revealed to the source node. The way that this source is revealed to the network is as if the network were fed by a successive-refinement scheme [35], [36]. Since the relays now try to estimate a source that is yet to be fully available, a new kind of estimation error arises, which complicates the MSE analysis.

The rest of the paper is organized as follows. We formally define the problem in Section II and provide a detailed overview of this work in Section III. In Section IV, we address the MSE when communicating a single analog source sample. In Section V, we replace that source sample by a data packet, possibly of unbounded size. In Section VI, we advance to a source sample being successively revealed to the first node, and then to a stream of data packets. Finally, in Section VII we conclude with a summary of the main results, and discussion of future research directions.

Notation: Denote by $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}$ the sets of integer, non-negative integer, real, non-negative real, and natural numbers, respectively. Throughout, all logarithms and exponents are taken to the natural base. We make use of small o notation: $f(x) = o_x(g(x))$ if $\lim_{x\to\infty} f(x)/g(x) = 0$; if the argument is clear, we write $o(\cdot)$. For a scalar $a \in \mathbb{R}_{\geq 0}$, we denote

$$\bar{a} = \frac{a}{1+a}. \tag{4}$$

We denote time sequences of a signal a_k between times $t_1, t_2 \in \mathbb{Z}_{\geq 0}$ $(t_1 \leq t_2)$ by

$$a_k(t_1{:}t_2) \triangleq \left[a_k(t_1),\, a_k(t_1+1),\, \ldots,\, a_k(t_2)\right].$$

We denote Markov chains by $X_1 \leftrightarrow X_2 \leftrightarrow X_3$, namely, given X_2, X_1 is independent of X_3. exp and log denote the natural exponential and logarithm functions, respectively. We use the binary entropy and divergence functions, which are defined as

$$\begin{aligned} h(p) &\triangleq p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}, \\ \mathbb{D}(p\|q) &\triangleq p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}, \end{aligned} \tag{5}$$

respectively, for $0 \leq p, q \leq 1$, where $p\log 0 \triangleq -\infty$ for $p > 0$, and $0\log q \triangleq 0$ for all q.

II. Problem Statement

A. Communication Model

The communication model that is considered in this work is depicted in Figure 1.

Causal source. Every T time steps, a source generates a new message that comprises ℓ bits, namely, at time $t = T\tau$ for $\tau \in \mathbb{Z}_{\geq 0}$, the source generates packet

$$W(\tau) = \left[B(\tau\ell),\, B(\tau\ell+1),\, \ldots,\, B((\tau+1)\ell-1)\right]$$

comprising bits. The bits of the entire message sequence, $\left\{B(i)\middle| i \in \mathbb{Z}_{\geq 0}\right\}$, are assumed i.i.d. uniform. Define the average rate of this source, measured in nats per time instant, as

$$R = \frac{\ell}{T}\log(2). \tag{6}$$

Channels. The network is composed of a cascade of channels with channel $r \in \mathbb{Z}_{\geq 0}$ given by

$$Y_r(t) = X_r(t) + Z_r(t), \tag{7}$$

where $X_r(t)$, $Y_r(t)$, and $Z_r(t)$ are the channel input, output, and noise, respectively, at time $t \in \mathbb{Z}_{\geq 0}$. The entries of all the noises $\left\{Z_r(t)\middle| r, t \in \mathbb{Z}_{\geq 0}\right\}$ are i.i.d. zero-mean, unit variance, and independent of all channel inputs.[5] All channel inputs are subject to a mean power constraint:

$$\mathbb{E}\left[X_r^2(t)\right] \leq P \;\forall r, t \in \mathbb{Z}_{\geq 0}. \tag{8}$$

The following channel quantities will be used in the sequel:

$$C = \frac{1}{2}\log(1+P) \tag{9}$$

is the Gaussian capacity and (recalling the bar notation (4))

$$\eta \triangleq \left(1-\bar{P}\right)\cdot\exp\{2R\} = \exp\{-2(C-R)\}. \tag{10}$$

As will be formally defined in the node functions below, perfect feedback is available in each channel, from the channel output to the terminal feeding it, in the subsequent time step.

Originating transmitter (node 0). At each time $t \in \mathbb{Z}_{\geq 0}$, generates a channel input $X_0(t)$ as a function of all the packet history $W\left(0 : \left\lfloor \frac{t}{T} \right\rfloor\right)$, and of all past outputs of channel 0, $Y_0(0 : t-1)$, which are available via feedback. The input is subject to the power constraint (8).

Node r ($r \in \mathbb{N}$). At each time $t \in \mathbb{Z}_{\geq 0}$, generates the channel input $X_r(t)$ as a function of all its measurement history[6] we assume *instantaneous* hops. See Remark 3 in the sequel. $Y_{r-1}(0{:}t)$ from its feeding channel (channel $r-1$), and of all its feedback history from the subsequent channel (channel r), $Y_r(0{:}t-1)$. The input is subject to the power constraint (8). The relays also produce estimates of the message.[7] Let $\hat{B}_r(n|t)$, $\hat{B}_r(0{:}n|t)$, $\hat{W}_r(n|t)$, and $\hat{W}_r(0{:}n|t)$ denote the estimates of $B(n)$, $B(0{:}n)$, $W(n)$, and $W(0{:}n)$, respectively, at node $r \in \mathbb{N}$ at time t based on the measurement history $Y_{r-1}(0{:}t)$ and feedback history $Y_r(0{:}t-1)$.

Scheme. We refer to the collection of maps of nodes 0 to $r \in \mathbb{N}$ at times 0 to $t \in \mathbb{Z}_{\geq 0}$ as an (r, t)*-scheme*. A *scheme* is defined as a nested set collection of (r, t)-schemes with respect

[5] Most of the results hold for any noise with unit variance. When we limit our attention to Gaussian noise, we state it explicitly.

[6] The current measurement is included, that is, similarly to [37], [38] and references therein,

[7] We consider decoding at a general node $r \in \mathbb{N}$. This decoding does not affect the creation of subsequent channel inputs in any way. As a special case, one can think of a specific node where the network terminates and information is needed. See Remark 2 in the sequel.

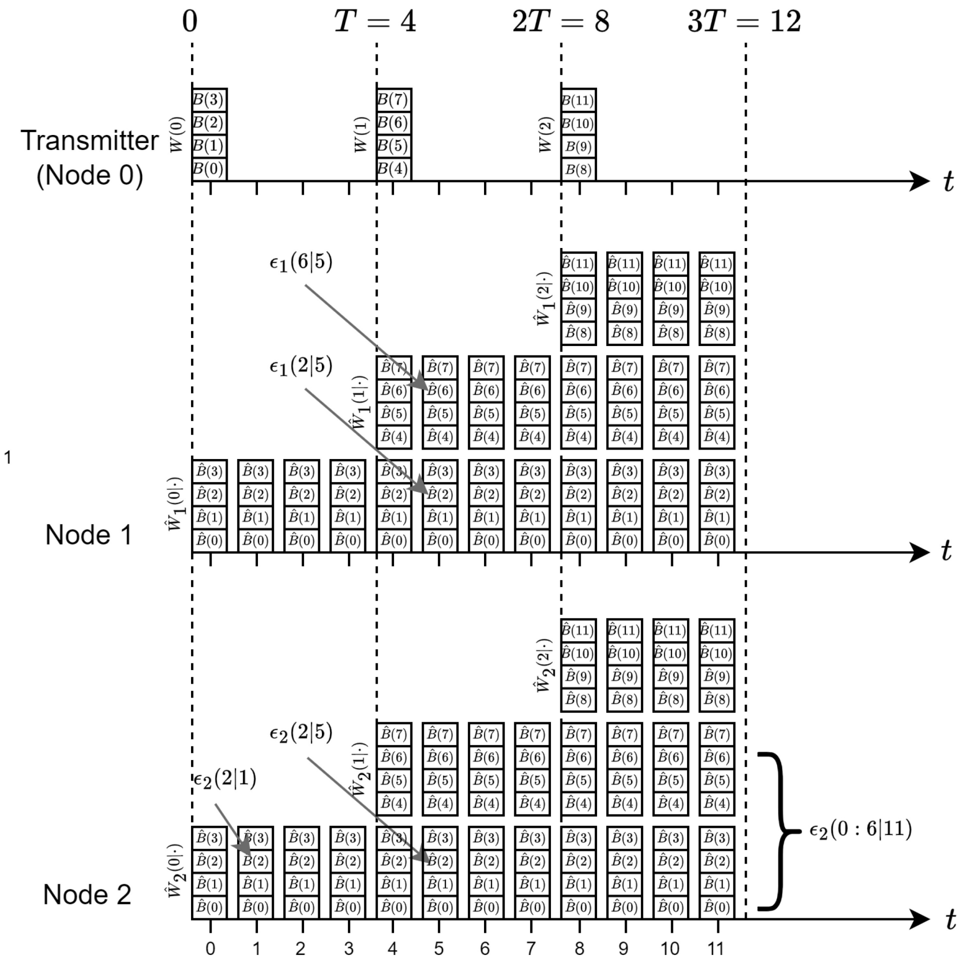

Fig. 2. Visualization of data packet streaming for instantaneous hops. Packet estimations at nodes $r = 0, 1, 2$ across time, for time period $T = 4$, and four bits per packet $\ell = 4$. The packets at the transmitter (node 0) indicate arrival time to transmitter, while the packets at nodes 1 and 2 indicate decoded packets.

to $t \in \mathbb{Z}_{\geq 0}$ for a fixed $r \in \mathbb{N}$, and with respect to $r \in \mathbb{N}$ for a fixed $t \in \mathbb{Z}_{\geq 0}$, to wit, an (r, t)-scheme equals to the union of all the $(\tilde{r}, \tilde{t})$-schemes over $\tilde{r} \leq r, \tilde{t} \leq t$.

B. Performance Measures

Error Probability. We define the *bit error probability* for bit n at relay r at time t as

$$\epsilon_r(n|t) \triangleq \Pr\Big(\hat{B}_r(n|t) \neq B(n)\Big).$$

We define the *maximal bit error probability* as a function of the detection delay Δ (the number of time steps elapsed since the generation of the bit) [19] as

$$P_e(r, \Delta) \triangleq \sup_{\tau \in \mathbb{Z}_{\geq 0}} \max_{\tau\ell \leq n \leq (\tau+1)\ell - 1} \epsilon_r(n|\tau T + \Delta). \tag{11}$$

The outer maximization (supremum) is carried over all the packets, whereas the inner maximization is carried over all the bits inside a packet, and Δ is the elapsed time (delay) after a generated bit is decoded. Thus, $P_e(r, \Delta)$ is the worst-case error probability across all the bits recovered Δ time steps after their generation.

Achievable streaming velocity. A streaming velocity $v \in \mathbb{R}_{\geq 0}$ of a source of average rate R is said to be *achievable* if there exists a scheme, such that,

$$\lim_{r \to \infty} P_e\Big(r, \Big\lfloor \frac{r}{v} \Big\rfloor\Big) = 0,$$

namely, $\Delta = \lfloor r/v \rfloor$ in P_e. This definition is a generalization of the IV of a single bit (1), where the corresponding expression for the maximal bit error probability for streaming (11) is used.

Streaming information velocity. The streaming IV of a source of average rate R, denoted by $\mathrm{V}(R)$, is defined as the supremum of all achievable streaming velocities of that source.

Figure 2 demonstrates schematically the propagation of packets along the network nodes versus time. The example in Figure 2 consists of a transmitter (node 0) and two relay nodes (nodes 1 and 2). The packet length is $\ell = 4$ bits and $T = 4$, meaning that the average rate is $R = \log 2$. The top graph indicates generation times of packets (when they arrive at node 0). Packets $W(0)$, $W(1)$, $W(2)$ arrive at node 0 at times 0, 4 and 8, respectively. The middle and bottom graphs depict the estimations of the hitherto generated packet at each time step at nodes 1 and 2, respectively: for $t \in [0, 3]$, packet $W(0)$ is estimated, namely, $\hat{W}(0|t)$; for $t \in [4, 7]$, packets $W(0{:}1)$ are estimated, namely $\hat{W}(0{:}1|t)$; for $t \in [8, 11]$, packets $W(0{:}2)$ are estimated, namely $\hat{W}(0{:}2|t)$; etc.

We also mark in Figure 2 a few instances of the error probability, e.g., $\epsilon_1(6|5)$ is the error probability at node 1 of bit $B(6)$ at time $t = 5$, and $\epsilon_2(0{:}6|11)$ is the error probability at node 2 of bits $B(0{:}6)$ at time $t = 11$.

Remark 1 (average rate): The streaming IV is set by the average rate R (6) rather than by the source parameters ℓ and T individually. To see why this is true, notice that the generation time of the bit n, which is τT where τ is the index of the packet containing n, equals $n\log(2)/R$ up to a bounded difference. This difference does not affect the IV as it is absorbed in $\Delta \to \infty$ in the definition (11).

It will become evident that we could use other transmission patterns with the same average rate without changing our results. We keep the periodic pattern for simplicity, noticing that it can approximate any desired positive rate.

Remark 2 (finite horizon): In this work, we consider a setting of an *infinite* number of cascaded nodes and analyze the velocity at which information can propagate over this network with a decaying error probability. This setting is more stringent than previous IV definitions [8], [14], [19] in which the number of nodes was a prescribed finite number with a target final receiver, which was then taken to infinity. In particular, in the previously-considered settings, all nodes were designed to attain the best possible error probability at a particular final receiver ("horizon"), whereas in our ("horizon-free") setting, tension may arise between attaining the best possible performance at different nodes. Clearly, any achievable IV or error probability guarantees carry over to the finite-horizon setting, but not necessarily the other way around.

Remark 3 (delayed hops): The described model above assumes instantaneous hops, namely, node r at time t transmits $X_r(t)$ that may depend on its measurement history $Y_{r-1}(1{:}t)$ until and *including* time t (and its feedback history $Y_r(1{:}t-1)$). The exposition of this scenario is simpler to follow and we therefore present it in detail. That said, we also present results for the parallel scenario of *delayed hops*, in which, node r at time t transmits $X_r(t)$ that can depend only on its purely causal history $Y_{r-1}(1{:}t-1)$ (and its feedback history $Y_r(1{:}t-1)$).

III. Detailed Work Overview

The setting described in Section II is the ultimate goal of this work. As mentioned in the introduction, we tackle this goal by first considering somewhat simpler settings, namely a single packet (no streaming) and source transmission (the data bits are replaced by a continuous source). Beyond setting the ground for the main result, these settings yield auxiliary results that may be of interest in their own right.

Let S denote a source sample that we want to convey to the nodes with minimal MSE, and $\hat{S}_r(t)$ denote its estimate at node r at time t. We will analyze the estimation MSE of $\hat{S}_r(t)$:

$$\mathrm{MSE}_r(t) \triangleq E\left[\left(S - \hat{S}_r(t)\right)^2\right]. \quad (12)$$

Besides being interesting in and of itself, as a joint source–channel coding (JSCC) problem, the transmission of an analog source over the same network constitutes an essential component in our proposed communication schemes. Specifically, in Section IV, the sample S is revealed to the transmitter before transmission begins.

In Section V, the error probability of transmitting a single packet over the network is derived. Transmitting a single packet may be viewed as a limiting streaming setting in which the packet inter-generation time T is infinite [and the average rate (6) goes to 0]. Therefore, all the definitions of Section II carry over to the single-packet transmission setting, except for the maximal bit error probability (11). Since the (only) packet comprises ℓ bits, $B(0{:}\ell-1)$, that are generated together at time 0, their detection delay Δ equals simply the elapsed time since the beginning of transmission t, the maximal bit error probability reduces, in this case, to

$$P_e(r,t) \triangleq \max_{0\le n\le \ell-1} \epsilon_r(n|t). \quad (13)$$

The IV is defined with respect to this error probability. Since our bounds on the single-packet IV do not depend on ℓ (and correspond to $R = 0$), we denote the single-packet IV by V without an argument.

To bound the maximal error probability of a transmission of a single packet, we think of the pulse-amplitude modulation (PAM) representation of the packet as a "source sample" and use the bounds on the MSE derived in Section IV. We subsequently analyze the packet error probability [decoding error probability of $B(0{:}\ell-1)$]:

$$\epsilon_r(0{:}\ell-1|t) = \Pr\left(\hat{B}_r(0{:}\ell-1|t) \neq B(0{:}\ell-1)\right).$$

Clearly, the packet error probability bounds from above the corresponding individual bit error probabilities since

$$\begin{aligned}\epsilon_r(0{:}\ell-1|t) &= \Pr\left(\bigcup_{i\in\{0,\dots,\ell-1\}} \left\{\hat{B}_r(i|t) \neq B(i)\right\}\right)\\ &\ge \max_{i\in\{1,\dots,\ell-1\}} \Pr\left(\hat{B}_r(i|t)\neq B(i)\right)\\ &= \max_{i\in\{1,\dots,\ell-1\}} \epsilon_r(i|t). \end{aligned} \quad (14)$$

Finally, In Section VI, we analyze the following streaming settings: in Section VI-A, we extend the JSCC setting of Section IV to a streaming-friendly variant, in which the sample S is gradually revealed (a la successive refinement) to the transmitter with quality that improves with time and derive bounds on the MSE (12) as a function of the node index r and the time t. In Section VI-B, we use the JSCC streaming solution and results of Section VI to derive bounds on the error probability for data-packet streaming by constructing a virtual source comprising the infinite concatenation of the entire data-packet sequence, where the hitherto generated data packets correspond to partial knowledge of the source S. Similar to (14), to bound the maximal bit error probability (11), we bound the more stringent prefix error probability

$$\epsilon_r(0{:}n|t) = \Pr\left(\hat{B}_r(0{:}n|t) \neq B(0{:}n)\right) \quad (15)$$

for $n \in \mathbb{N}$, by

$$\begin{aligned}P_e(r,\Delta) &\triangleq \sup_{\tau\in\mathbb{Z}_{\ge 0}} \max_{\tau\ell\le n\le(\tau+1)\ell-1} \epsilon_r(n|\tau T+\Delta)\\ &\le \sup_{\tau\in\mathbb{Z}_{\ge 0}} \epsilon_r(0{:}(\tau+1)\ell-1|\tau T+\Delta). \end{aligned} \quad (16)$$

IV. Transmission of a Single Source Sample

In this section, we consider the problem of transmitting a single zero-mean unit-variance source sample over the network under a minimum MSE (MMSE) criterion.

We propose the following simple *linear* scheme, which is a natural extension of the single-channel scheme of Elias [25] to our network setting.

Scheme 1: Initialization.

- Since the transmitter (node 0) knows S perfectly, it sets $\hat{S}_0(t) = S$ for all $t \in \mathbb{Z}_{\geq 0}$.
- Each node $r \in \mathbb{N}$ initializes its estimate before transmission begins to the mean: $\hat{S}_r(-1) = 0$.

Estimation at node $r \in \mathbb{N}$. At each time $t \in \mathbb{Z}_{\geq 0}$, constructs an estimate of S:

$$\hat{S}_r(t) = \hat{S}_r(t-1) + \gamma_r(t) Y_{r-1}(t), \tag{17}$$

where $\gamma_r(t)$ is a linear MMSE (LMMSE) estimation constant, to be specified in the sequel.

Transmission by node $r \in \mathbb{Z}_{\geq 0}$. At each time $t \in \mathbb{Z}_{\geq 0}$, transmits

$$X_r(t) = \beta_r(t)\Big[\hat{S}_r(t) - \hat{S}_{r+1}(t-1)\Big], \tag{18}$$

where $\beta_r(t)$ is a power-normalization constant, to be specified in the sequel.

Since node $r \in \mathbb{N}$ at time t knows $Y_{r+1}(0{:}t-1)$ via feedback, it can construct $\hat{S}_{r+1}(t-1)$ which is used to generate the transmit signal (18) of node r at time t. Using Scheme 1, the estimates satisfy the following properties.

Lemma 1: In Scheme 1, all channel inputs and outputs and all estimates have zero mean. Setting

$$\gamma_r(t) = \frac{\mathrm{Cov}(S, X_{r-1}(t))}{1 + \mathrm{Var}(X_{r-1}(t))}, \tag{19}$$

we have the following.

1) $\hat{S}_r(t)$ is the LMMSE estimate of S from $Y_{r-1}(0{:}t)$.
2) The outputs of the same channel at different times are uncorrelated, viz.
$$\mathrm{Cov}(Y_r(t), Y_r(\tau)) = 0 t \neq \tau.$$
3) For $r, t \in \mathbb{Z}_{\geq 0}$,
$$\mathrm{Cov}(S, X_r(t)) = \beta_r(t)\big[\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)\big].$$
4) The channel input variance equals, for $r, t \in \mathbb{Z}_{\geq 0}$,
$$\mathrm{Var}(X_r(t)) = \beta_r^2(t) \cdot \big[\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)\big].$$

The proof is based on properties of LMMSE estimation, primarily the orthogonality principle [39, Ch. 7-3]; it appears in Appendix A. Notice that the uncorrelatedness of the channel outputs means that each node receives "novel" information at each time step, and thus the channels are utilized well.

Assuming initial conditions $\mathrm{MSE}_r(-1) = 1$ for all $r \in \mathbb{Z}_{\geq 0}$, and the boundary conditions $\mathrm{MSE}_0(t) = 0$ for all $t \in \mathbb{Z}_{\geq 0}$, using the properties of Lemma 1, and choosing $\beta_r(t)$ to satisfy the power constraint with equality, we find that the constants satisfy, for all $r, t \in \mathbb{Z}_{\geq 0}$,

$$\beta_r(t) = \sqrt{\frac{P}{\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)}}, \tag{20a}$$

$$\gamma_{r+1}(t) = \frac{\sqrt{P(\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t))}}{P+1}. \tag{20b}$$

Using these values in Scheme 1 yields the relation [notice that $\beta_{r-1}(t)\gamma_r(t) = \bar{P}$ for all $r \in \mathbb{N}$ where $\bar{P}$ is defined according to the notation (4)]: For $r \in \mathbb{N}$ and $t \in \mathbb{Z}_{\geq 0}$,

$$\hat{S}_r(t) = \bar{P} \cdot \hat{S}_{r-1}(t) + \big(1 - \bar{P}\big) \cdot \hat{S}_r(t-1) + \gamma_r(t) Z_{r-1}(t).$$

Further, the resulting MSE of the scheme is given in the following lemma, whose proof appears in Appendix A.

Lemma 2: Scheme 1 with the parameters of (20) satisfies the recursion, for $r \in \mathbb{N}$ and $t \in \mathbb{Z}_{\geq 0}$,

$$\mathrm{MSE}_r(t) = \bar{P} \cdot \mathrm{MSE}_{r-1}(t) + \big(1 - \bar{P}\big) \cdot \mathrm{MSE}_r(t-1) \tag{21}$$

with the initial conditions $\mathrm{MSE}_r(-1) = 1$ for all $r \in \mathbb{Z}_{\geq 0}$, and the boundary conditions $\mathrm{MSE}_0(t) = 0$ for all $t \in \mathbb{Z}_{\geq 0}$. Furthermore, the solution of this recursion is

$$\mathrm{MSE}_r(t) = \big(1 - \bar{P}\big)^{t+1} \sum_{k=1}^{r} \binom{t+r-k}{r-k} \bar{P}^{r-k} \forall r, t \in \mathbb{N} \tag{22}$$

In order to study the asymptotic behavior of (22), we fix some velocity $v > 0$ such that $t = \lfloor r/v \rfloor$, and consider the MSE sequence as a function of the relay index.

Theorem 1: Let $v > 0$ be some fixed velocity. Then, the MSE of Scheme 1 satisfies

$$\mathrm{MSE}_r\left(\left\lfloor \frac{r}{v} \right\rfloor\right) \leq \exp\{-r E_1(v) + o(r)\}, \tag{23}$$

where $E_1(\cdot)$ is defined via $\bar{P} = \frac{P}{1+P}$ (4) as

$$E_1(v) \triangleq \begin{cases} \frac{1}{\bar{v}} \mathbb{D}\big(\bar{v} \big\| \bar{P}\big), & v < P \\ 0, & v \geq P. \end{cases} \tag{24}$$

Furthermore, the $o(r)$ correction term is independent of v.

The proof appears in Appendix A. It is based on the entropy-based bounds on the binomial coefficients. In fact, It can be shown (although it is not needed for our purposes) that the bound of (23) is exponentially tight.

Theorem 1 implies that, for any $v < P$, the MSE goes to zero with t (and $r \propto t$). That is, the source is reconstructed with arbitrarily good precision asymptotically. Indeed we can interpret this result as a lower bound on the source "reconstruction velocity".

In the next section, we will use the JSCC results of this section for the problem of transmitting a single data packet. This is achieved by mapping the data packet to a pulse amplitude modulation (PAM) symbol S, applying 1 to S by treating the latter as a source samples, and finally decoding the data packet from the resulting LMMSE estimate of S. This scheme will allow translating the JSCC results of this section to parallel results for data-packet transmission: a lower bound on the single-packet IV of $\mathrm{V} \geq P$, and an exponential decay rate of E_1 (24) for velocity $v < P$.

V. Transmission of a Single Data Packet

In this section, we consider the problem of transmitting a single data packet of ℓ i.i.d. uniform bits $B(0{:}\ell-1)$ over the network. We bound the block error probability $\epsilon_r(0{:}\ell-1|t)$ (15), which, as explained in Section III, serves as an upper bound on the worst-case individual bit error probability $P_e(r, t)$ (13). The error-probability bounds yield an achievable single-packet IV.

To that end, we map the bits $B(0{:}\ell - 1)$, which comprise the data packet, to a PAM constellation point using natural labeling:

$$S^\ell = \sqrt{3} \sum_{i=0}^{\ell-1} (-1)^{B(i)} 2^{-(i+1)}, \tag{25}$$

and apply Scheme 1 by viewing S^ℓ as source sample. Since the bits $B(0{:}\ell - 1)$ are uniformly distributed, S^ℓ is uniformly distributed over a discrete symmetric finite grid of size 2^ℓ with the spacing between two adjacent constellation points being

$$d_\ell = \sqrt{3} \cdot 2^{-\ell+1}; \tag{26}$$

in particular, S^ℓ has zero mean. In the limit of an infinite message length ℓ, S^ℓ converges to

$$S = \lim_{\ell \to \infty} S^\ell = \sqrt{3} \sum_{i=0}^{\infty} (-1)^{B(i)} 2^{-(i+1)}. \tag{27}$$

Since the bits $\{B(i)\}$ are i.i.d. uniform, S is uniformly distributed over $[-\sqrt{3}, \sqrt{3})$, meaning that it has zero mean and unit variance. For any ℓ, the variance of S^ℓ is smaller than 1. This allows us to construct the following transmission scheme for a single packet, which satisfies the power constraints (8).

Scheme 2:

1) The transmitter (node 0) maps $B(0{:}\ell - 1)$ to S^ℓ according to (25).
2) All nodes apply Scheme 1 with S^ℓ taking the role of the source S.
3) At each time step $t \in \mathbb{Z}_{\geq 0}$, each relay (node r for $r \in \mathbb{N}$) estimates $B(0{:}\ell - 1)$ from $\hat{S}^\ell_r(t)$—the estimate of S^ℓ at relay r at time t.

Remark 4: Since the variance of S^ℓ is strictly lower than 1 (converges to 1 in the limit of $\ell \to \infty$), Scheme 2 can be improved by adjusting the power of S^ℓ to 1 by multiplying it by a factor that is greater than 1. However, the gain of such an improvement is negligible in the limit of large t.

We denote the error of $\hat{S}^\ell_r(t)$ in estimating S^ℓ by

$$\mathcal{E}^\ell_r(t) \triangleq S^\ell - \hat{S}^\ell_r(t). \tag{28}$$

Since $\mathrm{Var}\left(S^\ell\right) \leq 1$,

$$\mathbb{E}\left[\left\{\mathcal{E}^\ell_r(t)\right\}^2\right] \leq \mathrm{MSE}_r(t). \tag{29}$$

Lemma 3: The packet error probability of Scheme 2 is bounded from above as

$$\epsilon_r(0{:}\ell|t) \leq \frac{1}{3} \cdot 2^{2\ell} \cdot \mathrm{MSE}_r(t)$$

with $\mathrm{MSE}_r(t)$ of Lemma 2.

Proof: Feeding $\hat{S}^\ell_r(t)$ to a nearest-neighbor decoder (slicer) of the constellation point S^ℓ results in an error probability that is bounded from above as

$$\epsilon_r(0{:}\ell|t) \leq \Pr\left(\left|\mathcal{E}^\ell_r(t)\right| > \frac{d_\ell}{2}\right) \tag{30a}$$

$$\leq \frac{\mathbb{E}\left[\left\{\mathcal{E}^\ell_r(t)\right\}^2\right]}{d_\ell^2/4} \tag{30b}$$

$$\leq \frac{1}{3} 2^{2\ell} \mathrm{MSE}_r(t), \tag{30c}$$

where (30a) follows from the nearest-neighbor decision rule and the PAM constellation points being distant by d_ℓ, (30b) follows from Chebyshev's inequality, and (30c) follows from (26) and (29). ■

Substituting the result of Theorem 1 in Lemma 3, and recalling the definition of single-packet IV (2), yields immediately the following.

Theorem 2: Let $\ell \in \mathbb{N}$ be the packet size, however large. Then, the single-packet IV (2) is bounded from below as $V \geq P$.

Moreover, for velocity $v < P$, the achievable prefix-free error probability is bounded as

$$\epsilon_r\left(0{:}\ell \middle| \left\lfloor \frac{v}{r} \right\rfloor\right) = \exp\{-E_1(v) r + o(r)\},$$

where $E_1(v)$ was defined in (24).

We notice, that the single-packet IV bound $V \geq P$ was independently derived by Inovan [20]. He also showed that essentially no better IV can be obtained by a linear scheme, and derived an achievable bound for non-homogeneous Sn Rs. However, he did not consider streaming, which is the core of this work.

For a single channel ($r = 1$) and *Gaussian noise*, Schalkwijk [23] (see also [22], [24], [40, pp. 481–482], [41, Ch. 17.1.1]) showed that the error probability of transmitting a single message decays *double exponentially* with t:

$$\epsilon_1(0{:}\ell|t) \leq \exp\left\{ - \exp\{2Ct + o(t)\}\right\}. \tag{31}$$

We next show that the doubly-exponential behavior extends also to our setting of multiple AWGN channels. To that end, we tighten the bound of Lemma 3 for AWGN channels as follows.

Lemma 4: The packet error probability of Scheme 2 over AWGN channels is bounded as

$$\epsilon_r(0{:}\ell|t) \leq 2\exp\left\{-\frac{3}{2^{2\ell+1} \cdot \mathrm{MSE}_r(t)}\right\}$$

with $\mathrm{MSE}_r(t)$ of Lemma 2.

The proof appears in Appendix B and relies on replacing the Chebyshev inequality with respect to the estimation error $\mathcal{E}^\ell_r(t)$ in (30b) with a Chernoff–Hoeffding bound with respect to a (sub-)Gaussian $\mathcal{E}^\ell_r(t)$ in this case.

Substituting the result of Theorem 1 in Lemma 4 yields immediately the following.

Theorem 3: Assume AWGN channels and let $\ell \in \mathbb{N}$ be the packet size, however large. Then, for velocity $v < P$, the achievable prefix-free error probability is bounded from above as

$$\epsilon_r\left(0{:}\ell \middle| \left\lfloor \frac{r}{v} \right\rfloor\right) = \exp\left\{ - \exp\{E_1(v) r + o(r)\}\right\},$$

where $E_1(v)$ was defined in (24).

Remark 5: The result of Theorem 3 extends to sub-Gaussian noises, since the proof of Lemma 4 relies on the sub-Gaussianity of the noise.

The result of Theorem 3 may be rewritten in terms of t as

$$\epsilon_r\left(0{:}\ell\middle|\left\lfloor \frac{r}{v} \right\rfloor\right) = \exp\left\{ -\exp\{E_1(v) \cdot vt + o(t)\}\right\}.$$

This expression means that the *doubly-exponential* behavior (31) of the celebrated scheme of Schalkwijk and Kailath [22], [23] extends also to communication at a fixed velocity across multiple relays as long as the velocity is below P. Indeed, since the error probability for $r = o(t)$ (and $r = 1$ in particular) can only be better than the limit of Theorem 3 as $v \to 0$, the result of Theorem 3 reduces exactly to (31) in this case, by noticing that $\lim_{v\to 0} E_1(v)v = 2C$ with C of (9).

Remark 6: Gallager [40, pp. 481–482] (see also [24]) rederived the doubly-exponential decay rate of Schalkwijk and Kailath by introducing a variant of their scheme that sends in the first time step a PAM constellation point, extracts the first Gaussian noise sample via feedback, and then applies over the following time steps the JSCC scheme of Elias [25] for conveying this Gaussian sample. In the analysis of Scheme 2 in Appendix B, we show that in fact the JSCC scheme can be used from the start to describe the constellation point directly; the analysis of Gallager extends to this case by identifying that the noise is sub-Gaussian with variance proxy that equals the variance.

Remark 7: The super exponential behavior of the Schalkwijk–Kailath scheme was shown by Gallager and Nakiboğlu [24] to extend beyond a second-order exponential decay, where the order of the exponential decay is commensurate with the blocklength n, namely, a tetration-like decay. This behavior seems to extend also to the network setting with a constant velocity $v < P$, but it is beyond the scope of this work.

The bound of Lemma 3 deteriorate exponentially with an increase in ℓ. We next derive a bound on the prefix error probability $\epsilon_r(0{:}n|t)$, that holds uniformly for all ℓ (assuming $n \le \ell$), and hence serves as a stepping stone for the derivation of similar bounds for streaming in Section VI.

Lemma 5: The prefix error probability of Scheme 2, for any $n \le \ell$, is bounded from above as

$$\epsilon_r(0{:}n|t) \le \frac{2}{\sqrt{3}} \cdot 2^n \cdot \sqrt{\mathrm{MSE}_r(t)}$$

with $\mathrm{MSE}_r(t)$ of Lemma 2.

This result is proven in Appendix C. In conjunction with Theorem 1, it allows to prove that any velocity below P is achievable in agreement with Theorem 2. However, as the bound of Lemma 5 is proportional to the square-root of the bound of Lemma 3, it attains only half the error exponent $E_1(v)$. This factor two in the exponent is thus the price of decoding in the presence of "interference" from an unbounded number of bits that are not decoded.

VI. Streaming

In this section, we consider the packet streaming transmission problem as introduced in Section II. Like in single packet transmission, we start with the counterpart JSCC problem, namely, conveying, over the network, a source that is only gradually revealed to the transmitter, and then apply the MSE results to packet streaming.

A. Successively Refined Source

Suppose that the transmitter node does not have full knowledge of the source at the start ($t = 0$), but rather it is given an estimate $\hat{S}_0(t)$ at time t, such that these estimates form a Markov chain $\hat{S}_0(0) \leftrightarrow \hat{S}_0(1) \leftrightarrow \ldots \leftrightarrow S$. Thus, we can think of them as being the reconstructions obtained from a successive refinement scheme feeding the network. In particular, we will assume that these transmitter inputs have MSE

$$\mathrm{MSE}_0(t-1) = \exp\{-2Rt\} \ \forall t \in \mathbb{N} \tag{32}$$

for some $R > 0$. This is the dependence that we will need in the sequel, when the "source" is the PAM constellation corresponding to a digital message that arrives at R nats per sample (neglecting rounding issues). It is also relevant in itself, as a source that is gradually revealed to the transmitter via a fixed-rate successive refinement scheme.[8]

The scheme that we use is almost identical to Scheme 1, except the boundary condition which is not 0 for all t but rather improves according to (32).

Scheme 3: Transmitter initialization. The transmitter sets the given $\hat{S}_0(t)$ for all $t \in \mathbb{Z}_{\ge 0}$.

The rest of the scheme (initialization of nodes $r \in \mathbb{N}$, estimation at nodes $r \in \mathbb{N}$, transmission by nodes $r \in \mathbb{Z}_{\ge 0}$) is exactly as Scheme 1.

The power-normalization constant $\beta_r(t)$ and the LMMSE estimation constant $\gamma_r(t)$ are set according to (20).

Remark 8: One may wonder, why the transmitter does not apply preprocessing, when estimating the source from its inputs. Indeed, due to the Markov structure, the best estimate of S from $\hat{S}_0(0{:}t)$ only depends on $\hat{S}_0(t)$. There may still be a scalar function, e.g., scaling by a factor, that would improve the estimate; we assume that it is already applied before the estimate is given to the transmitter.

In the following theorem, we evaluate the asymptotic behavior of the MSE for a fixed velocity. Namely, we fix some velocity $v > 0$, set time $t = \lfloor t/v \rfloor$, and consider the MSE sequence as a function of relay r.

Theorem 4: Let $v > 0$ be some fixed velocity. Then, the MSE of Scheme 3 satisfies

$$\mathrm{MSE}_r\left(\left\lfloor \frac{r}{v} \right\rfloor\right) \le \exp\{-rE_S(v) + o(r)\},$$

where

$$E_S(v) = \begin{cases} \frac{1}{1-\eta}\mathbb{D}\left(1-\eta \middle\| \bar{P}\right) + 2R\left(\frac{1}{v} - \frac{\eta}{1-\eta}\right), & 0 \le v \le \frac{1-\eta}{\eta} \\ \frac{1}{v}\mathbb{D}\left(\bar{v} \middle\| \bar{P}\right), & \frac{1-\eta}{\eta} < v \le P \\ 0, & P < v \end{cases} \tag{33}$$

and $\eta \triangleq (1-\bar{P}) \cdot \exp\{2R\} = \exp\{-2(C-R)\}$ as in (10). Furthermore, the $o(r)$ correction term is independent of v.

The proof appears in Appendix D. It is similar to the proofs of Lemma 2 and Theorem 1, as the MSE recursion is the same.

[8] This is the MSE that is attained by a sequence of greedy optimal quantizers of rate R (neglecting rounding issues) that are applied to a uniformly distributed source sample. Moreover, for continuous sources, this decay is exponentially optimal in t as it matches the exponent of the distortion–rate function by the Shannon's lower bound and the Gaussian source being the "least compressable" [42, Prob. 10.8 and Th. 10.4.1], [41, Prob. 3.18, Ch. 3.9].

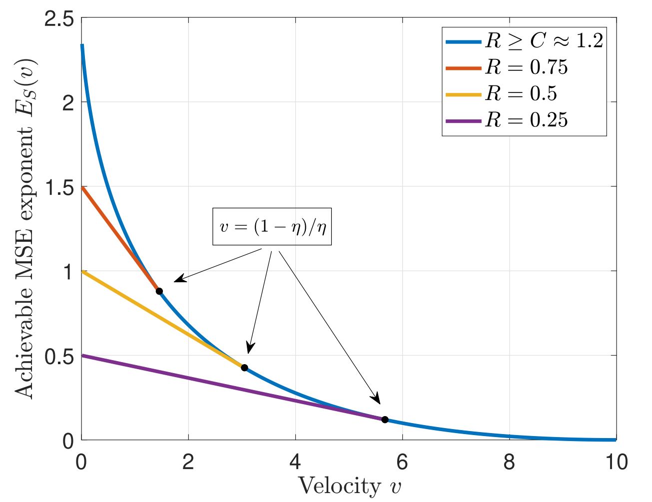

Fig. 3. Successively refined source: The achievable MSE exponent. The plot depicts the exponent as a function of velocity, at fixed SNR $P = 10$, as a function of the refinement rate R. The upper curve is $E_1(v)$, which is $E_S(v)$ for any rate above the capacity $C = \log(11)/2 \cong 1.2$ nats. The rest of the curves correspond to rates below capacity, where, for each rate ($R = 0.25, 0.5, 0.75$), the exponent $E_1(v)$ is depicted

However, the initial conditions for the MSE are different. In the first region of (33), the initial error at the transmitter node is (exponentially) dominant, while, in the second region, the lack of any source knowledge at the relay nodes at $t = 0$ is dominant and the exponent is identical to $E_1(v)$ (24).

The result in Theorem 4 means that, for any $v < P$, the MSE of the source sample goes to zero for an exponentially decaying boundary condition (32). That is, the "reconstruction velocity" is the same as it were with full a priori source knowledge (Section IV).

As the rate grows, the exponent $E_S(v)$ becomes the exponent with full source knowledge at the transmitter at $t = 0$, $E_1(v)$ (24). This is to be expected, as, in the limit $R \to \infty$, the initial MSE drops immediately. Moreover, for all $R \geq C$, the exponents are already equal, as the first region of (33) is empty in that limit. See Figure 3.

Another way to consider $E_S(v)$, evident in Figure 3, is the following. Clearly, even at zero velocity, any exponent above $2R$ is not achievable, as this is the exponent at which the transmitter node learns the source. Then, the exponent in the first region is the tangent to $E_1(v)$ at $v = \frac{1-\eta}{\eta}$, originating at $E_S(0) = 2R$.

B. Packet Streaming

We now finally reach our target scenario as described in Section II. We combine the PAM mapping with the successively-refined source scheme, as follows. Recall that, at time $t = \tau T$ for $\tau \in \mathbb{Z}_{\geq 0}$, the τ^{th} packet is made available at the transmitter. Thus, at time $t = \tau T$, it has access to bits $B(0{:}\tau\ell - 1)$. By (25), the transmitter can then map these bits to the corresponding MAP constellation point

$$S^{\tau\ell} = \sqrt{3} \sum_{i=0}^{\tau\ell-1} (-1)^{B(i)} 2^{-(i+1)}. \tag{34}$$

As in Section V, we define S to be the infinite-constellation limit:

$$S = \lim_{\ell\to\infty} S^{\ell} = \lim_{\tau\to\infty} S^{\tau\ell}; \tag{35}$$

recall that S has zero mean and unit variance. Clearly, the nested constellations have all zero mean, and they form a Markov chain $S^0 \leftrightarrow S^{\ell} \leftrightarrow S^{2\ell} \leftrightarrow S^{3\ell} \leftrightarrow \ldots \leftrightarrow S$. Furthermore, they form the LMMSE (even MMSE) estimates of S given the available bits, since the resulting estimation error $S - S^{\tau\ell}$ is independent of $B(0{:}\tau\ell - 1)$. Thus, we can use Scheme 3 for the successively refined source S, with $S^{\tau\ell}$ taking the role of $\hat{S}_0(\ell)$, as follows.

Scheme 4:

1) At instants $t = \tau T$ for $\tau \in \mathbb{Z}_{\geq 0}$, the transmitter maps $B(0{:}\tau\ell - 1)$ to $S^{\tau\ell}$ according to (34) and updates $\hat{S}_0(\tau T) = \hat{S}_0(\tau T + 1) = \hat{S}_0(\tau T + 2) = \cdots = \hat{S}_0((\tau + 1)T - 1)$.
2) The transmitter and relays apply Scheme 3 with respect to the (virtual) source S (35).
3) At each time step $t \in \mathbb{Z}_{\geq 0}$, each relay (node r for $r \in \mathbb{N}$) estimates the hitherto generated bits $B\big(0{:}\big(\lfloor \frac{t}{T} \rfloor + 1\big)\ell - 1\big)$ from $\hat{S}_r(t)$.

Remark 9: $\hat{S}_0(t)$ follow an MSE profile in t that is strictly better (smaller) than (32) for $t \neq \tau T$ with $\tau \in \mathbb{Z}_{\geq 0}$, and satisfies (32) with equality for $t = \tau T$. Thus, Scheme 4 can be improved. However, the gain from such an improvement becomes negligible in the limit of large t.

We can now invoke the MSE of Theorem 4 and the bound on the error probability of Lemma 5, which holds uniformly for all packet sizes, to obtain the following.

Lemma 6: In Scheme 3 above, for any average rate $R > 0$ and for all $v < P$, $P_e(r, \lfloor r/v \rfloor)$ is bounded from above by

$$\exp\left\{ - \inf_{t_0 \in \mathbb{Z}_{\geq 0}} \left[\frac{t_0 + \Delta}{2v} E_S\left(\frac{v\Delta}{t_0 + \Delta} \right) - t_0 R \right] + o(r) \right\},$$

where the worst-bit error probability $P_e(r, \Delta)$ was defined in (11), and $E_S(\cdot)$ is given by (33).

The proof appears in Appendix E. It is based upon combining the MSE bound of Theorem 4 with the error-probability bound of Lemma 5. We note that, in Theorem 4, the velocity is measured from time 0, while in the definition of $P_e(r, \Delta)$ in (11), the velocity of a particular packet is measured with respect to the elapsed time since its generation. Since each packet is generated at a different time, at a particular node r at a particular time t, each packet has a different velocity for the purpose of $P_e(r, \Delta)$. Hence, the key step in the proof of Lemma 6 is translating the bound of Theorem 4 using an affine transformation of the velocity.

Having proved this, we are ready to state our main result.

Theorem 5: Let $R < C$ be some average rate (6) where $C \triangleq \frac{1}{2}\log(1 + P)$ as in (9). Then, the streaming IV is bounded from below as

$$\mathrm{V}(R) \geq \exp\{2(C - R)\} - 1.$$

The proof appears in Appendix E. It is based upon substituting $E_S(v)$ (33) in the result of Lemma 6. Noticing that the argument of $E_S(\cdot)$ in (33) is at most v, for v below the claimed

TABLE I
SUMMARY OF MAIN RESULTS, AND THEIR TRANSLATION TO DELAYED HOPS. RECALL THAT $\bar{P} = P/(1+P)$, $\bar{v} = v/(1+v)$ AND $\eta = (1-\bar{P})\exp\{2R\} = \exp\{-2(C-R)\}$

Setting	Reference	Lower bound on	Instantaneous Hops	Delayed Hops
Single source sample	Thm. 1	MSE exponent	For $v \in [0, P)$: $\frac{1}{\bar{v}}\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right)$	For $v \in [0, \bar{P})$: $\frac{1}{v}\mathbb{D}\left(v\middle\|\bar{P}\right)$
Single data packet	Thm. 2	single-packet IV	P	$\bar{P}$
Successively-refined source	Thm. 4	MSE exponent	For $v \in \left[0, \frac{1-\eta}{\eta}\right)$: $\frac{1}{1-\eta}\mathbb{D}\left(1-\eta\middle\|\bar{P}\right)$ $+2\left(\frac{1}{v} - \frac{\eta}{1-\eta}\right)R$ For $v \in \left[\frac{1-\eta}{\eta}, P\right)$: $\frac{1}{\bar{v}}\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right)$	For $v \in [0, \eta)$: $\frac{1}{1-\eta}\mathbb{D}\left(1-\eta\middle\|\bar{P}\right)$ $+2\left(\frac{\bar{v}}{1-\bar{v}} - \frac{\eta}{1-\eta}\right)R$ For $v \in [\eta, \bar{P})$: $\frac{1}{v}\mathbb{D}\left(v\middle\|\bar{P}\right)$
Packet streaming	Thm. 5	Streaming IV	$\frac{1-\eta}{\eta}$	$1-\eta$

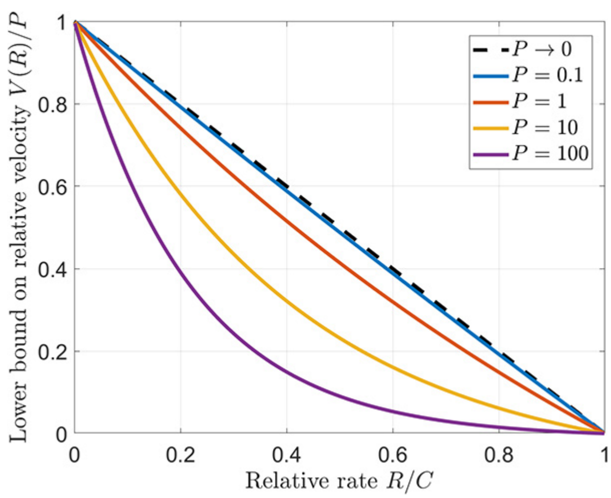

Fig. 4. Lower bound on the streaming IV as a function of the rate average rate R (6) for SNRs $P = 0.1, 1, 10, 100$ and the linear limit $P \to 0$. Relative coordinates are used: the lower bound on the streaming IV of Theorem 5 is normalized by the lower bound on the single-packet IV P of Theorem 2, whereas the average rate R is normalized by C (9).

bound we always take the first region in (33) which gives the desired positive error exponent (it becomes independent of t_0 after the substitution, making the minimization redundant).

Remark 10: For $v > \exp\{2(C-R)\} - 1$ and large enough Δ, the argument of $E_S(\cdot)$ in (33) falls inside the second region of (33) and the error exponent becomes negative; we do not provide a proof of this fact, as it is not required for our achievability result. Hence, we believe that the IV bound of Theorem 5 is tight for our Scheme 4.

As expected, the IV bound approaches the single-packet IV bound P as the rate goes to zero, and—to zero as the rate goes to C. It is also worth noting that, for any fixed P, the achievable velocity is a convex function of R. In the low-SNR limit $P \to 0$, the curve approaches the linear function (See Figure 4):

$$V(R) \cong \left(1 - \frac{R}{C}\right)P.$$

In the high-SNR limit, on the other hand, for any fixed R, our bound grows linearly with P:

$$V(R) \cong P\exp\{-2R\}.$$

VII. Conclusion and Extensions

We have derived achievable IV and error-probability bounds for both single-packet and streaming communications, and exponential MSE bounds for conveying a source sample, both when it is available upfront and when it is revealed successively to the transmitter.

In Table I, we summarize our main results, and translate them to the setting of delayed hops, which may be more familiar to some of the readers. The translation is immediate, by a linear transformation of the velocity: if the velocity with instantaneous hops is v, then with delayed hops it becomes $\bar{v} = v/(1+v)$, and in particular it is at most 1. We have not included the results on decoding error probability which are longer, but they may be translated in the same way.

This work raises many interesting directions for further research. First, we have only presented achievability results. We believe that our analysis is tight for linear schemes, but is it possible to improve them using a different approach? Specifically, it seems plausible that no positive IV is possible for rates $R > C$, but we are less certain about the IV at lower rates.

Secondly, since our analysis is local both in the time axis and in the relays axis, it presents a rather wide framework that allows for variations. For conveying a source, we considered one that is either known from the start or conveyed at a constant refinement rate; the MSE recursion was the same, and only the initial conditions differed. Other scenarios, including variable-rate refinement and side information of various types in relays along the way may be treated under the same framework. As for the channels, we assumed that they all enjoy the same SNR. Yet, the proposed scheme readily extends to heterogeneous SNRs across the different channels, by simple adjustments to the appropriate difference equation (as was done for packet-erasure channels in [19]). Indeed, an achievable single-packet IV for this scenario was derived in [20, Ch. 5].

Lastly, in this work, we have assumed perfect feedback. Such feedback is not practical. In particular, in a line network it does not make sense that, all along a line network, the backward channels enjoy much better conditions than the forward ones. Extending our results to settings with noisy feedback is of major interest. The feedback scheme may be linear, or using modulo-lattice modulation as Ben-Yishai and Shayevitz suggested for a single channel [43].

Appendix A
Proofs for Section IV

We start with two auxiliary lemmas that will be used in the sequel. They refer to an *generalized Scheme 1*, in which the estimator $\hat{S}_r(t)$ at node $r \in \mathbb{N}$ at time $t \in \mathbb{N}$ is taken to be the LMMSE estimator of S from its entire measurement history $Y_{r-1}(0{:}t)$; we will prove that $\hat{S}_r(t)$ is of the form of (17) (part 1 of Lemma 1). We denote the estimation error by

$$\mathcal{E}_r(\tau) \triangleq S - \hat{S}_r(\tau). \tag{36}$$

Lemma 7: In the generalized Scheme 1, the outputs of the same channel at different times are uncorrelated, viz. $\mathrm{Cov}(Y_r(t), Y_r(\tau)) = 0$ for $t \neq \tau$.

Proof: Assume without loss of generality that $0 \le \tau < t$. Then,

$$\begin{aligned}
\mathrm{Cov}(Y_r(t), Y_r(\tau)) &= \mathrm{Cov}(X_r(t) + Z_r(t), Y_r(\tau)) && \text{(37a)}\\
&= \mathrm{Cov}(X_r(t), Y_r(\tau)) && \text{(37b)}\\
&= \mathrm{Cov}\Big(\beta_r(t)\Big[\hat{S}_r(t) - \hat{S}_{r+1}(t-1)\Big], Y_r(\tau)\Big) && \text{(37c)}\\
&= \mathrm{Cov}\big(\beta_r(t)\big[\mathcal{E}_{r+1}(t-1) - \mathcal{E}_r(t)\big], Y_r(\tau)\big) && \text{(37d)}\\
&= \beta_r(t)[\mathrm{Cov}(\mathcal{E}_{r+1}(t-1), Y_r(\tau))\\
&\quad - \mathrm{Cov}(\mathcal{E}_r(t), Y_r(\tau))] && \text{(37e)}
\end{aligned}$$

where (37a) follows from (7), (37b) holds since $Z_r(t)$ is independent of $Y_r(0{:}t-1)$ and hence uncorrelated with its entries, (37c) follows from (18), (37d) follows from adding and subtracting S and (36), and (37e) follows from the bilinearity of the covariance. Now, to complete the proof, we prove that both of the resulting elements equal zero as follows.

1) Since $\mathcal{E}_{r+1}(t-1)$ is the LMMSE estimation error given $Y_r(0{:}t-1)$, by the orthogonality principle [39, Ch. 7-3], $\mathrm{Cov}(\mathcal{E}_{r+1}(t-1), Y_r(0{:}t-1)) = 0$. In particular, since $\tau \le t-1$, $\mathrm{Cov}(\mathcal{E}_{r+1}(t-1), Y_r(\tau)) = 0$.
2) $Y_r(\tau)$ is a linear combination of $Y_{r-1}(0{:}\tau)$ and $Z_r(0{:}\tau)$. By the orthogonality principle, $\mathrm{Cov}(\mathcal{E}_r(t), Y_{r-1}(0{:}t))$. $\mathcal{E}_r(t)$ is independent of (and hence uncorrelated with) the noise process Z_r of the subsequent channel. Since $\tau \le t-1$, $\mathrm{Cov}(\mathcal{E}_r(t), Y_r(\tau)) = 0$. ■

Lemma 8: In the generalized Scheme 1,

$$\mathrm{Cov}(\mathcal{E}_r(t), \mathcal{E}_{r+1}(t-1)) = \mathrm{Var}(\mathcal{E}_r(t)) \forall r, t \in \mathbb{Z}_{\ge 0}.$$

Proof: Let $r, t \in \mathbb{Z}_{\ge 0}$. Then,

$$\begin{aligned}
\mathrm{Cov}(\mathcal{E}_r(t), \mathcal{E}_{r+1}(t-1)) &= \mathrm{Cov}\Big(\mathcal{E}_r(t), S - \hat{S}_{r+1}(t-1)\Big) && \text{(38a)}\\
&= \mathrm{Cov}(\mathcal{E}_r(t), S) && \text{(38b)}\\
&= \mathrm{Cov}\Big(\mathcal{E}_r(t), \hat{S}_r(t) + \mathcal{E}_r(t)\Big) && \text{(38c)}\\
&= \mathrm{Var}(\mathcal{E}_r(t)) && \text{(38d)}
\end{aligned}$$

where (38a) and (38c) follow from (36); and (38b) and (38d) follow from $\mathrm{Cov}(\mathcal{E}_r(t), Y_{r-1}(0{:}t)) = 0$ and $\mathrm{Cov}(\mathcal{E}_r(t), Y_{r-1}(0{:}t)) = 0$, respectively, which were proved at the end of the proof of Lemma 7, by noting that $\hat{S}_{r+1}(t-1)$ and $\hat{S}_r(t)$ are linear combinations of $Y_r(0{:}t-1)$ and $Y_{r-1}(t)$, respectively. ■

Proof of Lemma 1: All means are zero since all the quantities in the scheme are linear combinations of the zero-mean source S and the entries of th zero-mean noise processes $\{Z_r \big| r \in \mathbb{Z}_{\ge 0}\}$. We are now ready to prove the four parts of the lemma.

1) Consider $\hat{S}_r(t)$ for $r \in \mathbb{N}$ of the generalized Scheme 1. By Lemma 7, the entries of $Y_{r-1}(0{:}t)$ are uncorrelated, meaning that the LMMSE estimator $\hat{S}_r(t)$ equals
$$\begin{aligned}
\hat{S}_r(t) &= \sum_{\tau=0}^{t} \frac{\mathrm{Cov}(S, Y_{r-1}(\tau))}{\mathrm{Var}(Y_{r-1}(\tau))} \cdot Y_{r-1}(\tau) && \text{(39a)}\\
&= \sum_{\tau=0}^{t} \frac{\mathrm{Cov}(S, X_{r-1}(\tau))}{1 + \mathrm{Var}(X_{r-1}(\tau))} \cdot Y_{r-1}(\tau), && \text{(39b)}
\end{aligned}$$
where the second equality follows from (7), the independence of S and the process Y_{r-1}, and $\mathrm{Var}(Z_{r-1}(\tau)) = 1$ for all τ. Eq. (39b) suggests the following recursive form:
$$\begin{aligned}
\hat{S}_r(t) &= \hat{S}_r(t-1) + \frac{\mathrm{Cov}(S, X_{r-1}(t))}{1 + \mathrm{Var}(X_{r-1}(t))} \cdot Y_{r-1}(\tau)\\
&= \hat{S}_r(t-1) + \gamma_r(t) Y_{r-1}(\tau),
\end{aligned}$$
where the second equality follows from (19).
This proves that Scheme 1 and the generalized Scheme 1 are in fact identical.
2) Follows immediately from the proof of part 1 of the lemma and from Lemma 7.
3) By (18), $\mathrm{Cov}(S, X_r(t))$ equals
$$\beta_r(t)\Big\{\mathrm{Cov}\Big(S, \hat{S}_r(t)\Big) - \mathrm{Cov}\Big(S, \hat{S}_{r+1}(t-1)\Big)\Big\}.$$
the proof follows by substituting
$$\begin{aligned}
\mathrm{Cov}\Big(S, \hat{S}_r(t)\Big) &= \mathrm{Cov}(S, S) - \mathrm{Cov}(S, \mathcal{E}_r(t)) && \text{(40a)}\\
&= \mathrm{Var}(S) - \mathrm{Cov}\Big(\hat{S}_r(t) + \mathcal{E}_r(t), \mathcal{E}_r(t)\Big) && \text{(40b)}\\
&= \mathrm{Var}(S) - \mathrm{Var}(\mathcal{E}_r(t)) && \text{(40c)}\\
&= \mathrm{Var}(S) - \mathrm{MSE}_r(t), && \text{(40d)}
\end{aligned}$$
where (40a) and (40b) holds by the definition of $\mathcal{E}_r(t)$ (36), (40c) follows from the orthogonality principle, and (40d) holds by the definitions (12) and (36).
4) The proof of this part is as follows.

$$\begin{aligned}
\frac{\mathrm{Var}(X_r(t))}{\beta_r^2(t)} &= \mathrm{Cov}\Big(\hat{S}_r(t) - \hat{S}_{r+1}(t-1), \hat{S}_r(t) - \hat{S}_{r+1}(t-1)\Big) && \text{(41a)}\\
&= \mathrm{Cov}(\mathcal{E}_r(t) - \mathcal{E}_{r+1}(t-1), \mathcal{E}_r(t) - \mathcal{E}_{r+1}(t-1)) && \text{(41b)}\\
&= \mathrm{Var}(\mathcal{E}_r(t)) + \mathrm{Var}(\mathcal{E}_{r+1}(t-1))\\
&\quad - 2\mathrm{Cov}(\mathcal{E}_r(t), \mathcal{E}_{r+1}(t-1)) && \text{(41c)}
\end{aligned}$$

$$= \mathrm{Var}(\mathcal{E}_r(t)) - \mathrm{Var}(\mathcal{E}_{r+1}(t-1)), \tag{41d}$$

where (41a) follows from (18), (41b) follows from (36), and (41d) follows from Lemma 8. ∎

Proof of Lemma 2: Let $r \in \mathbb{N}$ and $t \in \mathbb{Z}_{\geq 0}$. Then,

$$\mathrm{MSE}_r(t) = 1 - \sum_{\tau=1}^{t} \frac{\mathrm{Cov}^2(S, Y_{r-1}(\tau))}{\mathrm{Var}(Y_{r-1}(\tau))} \tag{42a}$$

$$= 1 - \sum_{\tau=1}^{t} \frac{\mathrm{Cov}^2(S, X_{r-1}(\tau))}{P+1}, \tag{42b}$$

where (42a) follows from the LMMSE expression [39, eq. (7-85)] and parts 1 and 2 of Lemma 1; and (42b) follows from (7), the independence of S and the unit-variance noise process Z_{r-1}, and by choosing $\left\{\beta_{r-1}(\tau)\middle|\tau \in \mathbb{Z}_{\geq 0}\right\}$ such that $\left\{X_{r-1}(\tau)\middle|\tau \in \mathbb{Z}_{\geq 0}\right\}$ satisfies the power constraint (8) with equality. This can be rewritten recursively as

$$\mathrm{MSE}_r(t-1) - \mathrm{MSE}_r(t) = \frac{\mathrm{Cov}^2(S, X_{r-1}(t))}{P+1} \tag{43a}$$

$$= \beta_{r-1}^2(t) \cdot \frac{\left[\mathrm{MSE}_r(t-1) - \mathrm{MSE}_{r-1}(t)\right]^2}{P+1} \tag{43b}$$

$$= \bar{P} \cdot \left[\mathrm{MSE}_r(t-1) - \mathrm{MSE}_{r-1}(t)\right], \tag{43c}$$

where (43b) follows from part 3 of Lemma 1, and (43c) follows from (20a). This proves (21).

We next solve the recursion (21) explicitly. Due to the linearity of the recursion, $\mathrm{MSE}_r(t)$ equals the sum of the effects of all initial conditions $\mathrm{MSE}_k(-1) = 1$ for $1 \leq k \leq r$ (the zero boundary conditions $\mathrm{MSE}_0(t) = 0$ have zero effect). Consider all the trajectories from $(k, -1)$ to (r, t), where at each step along the trajectory either the time index increases by 1—corresponding to a multiplication by $1 - \bar{P}$, or the relay index increases by 1—corresponding to a multiplication the MSE by $\bar{P}$. There are $t+1$ time steps and $r-k$ relay steps. Multiplying by the combinatorial expression for the number of trajectories, we have that the effect of a single condition is

$$\binom{t+r-k}{r-k}\left(1-\bar{P}\right)^{t+1}\bar{P}^{r-k}.$$

Eq. (22) then follows by summing over all relevant k. ∎

Proof of Theorem 1: The claim is trivial for $v \geq P$ since $\mathrm{MSE}_r(t) \leq \mathrm{Var}(S) = 1$. Thus, we assume $v < P$ and bound the LMMSE at node $r \in \mathbb{N}$ at time $t \in \mathbb{Z}_{\geq 0}$ as follows.

$$\mathrm{MSE}_r(t) = \left(1-\bar{P}\right)\sum_{k=1}^{r}\binom{t+r-k}{r-k}$$

$$* \quad \cdot\exp\left\{-(t+r-k)\left[\mathbb{D}\left(\frac{r-k}{t+r-k}\middle\|\bar{P}\right) + h\left(\frac{t}{t+r-k}\right)\right]\right\}$$

$$\leq \exp\left\{-\min_{k\in\{1,2,\ldots,r\}}(t+r-k)\ \mathbb{D}\left(\frac{r-k}{t+r-k}\middle\|\bar{P}\right) + \log r\right\},$$

where the equality follows from (22) from Lemma 2 and the definitions (5), and the inequality follows from bounding the average by the maximum and from [42, Ch. 11.1]

$$\binom{n}{i} \leq \exp\left\{-n\,h\left(\frac{i}{n}\right)\right\}.$$

By taking $t = \lfloor r/v \rfloor$ as in the theorem statement and defining $\lambda \triangleq \frac{r-k}{t}$, we have

$$\mathrm{MSE}_r\left(\left\lfloor\frac{r}{v}\right\rfloor\right) \leq \exp\left\{-\left\lfloor\frac{r}{v}\right\rfloor \inf_{\lambda\in(0,v)}(1+\lambda)\mathbb{D}\left(\frac{\lambda}{1+\lambda}\middle\|\bar{P}\right) + \log r\right\} \tag{44a}$$

$$\leq \exp\left\{-\left(\frac{r}{v}-1\right)(1+v)\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right) + \log r\right\} \tag{44b}$$

$$\leq \exp\left\{-\frac{r}{\bar{v}}\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right) - (1+P)\log\bar{P} + \log r\right\}, \tag{44c}$$

where (44b) holds since the expression in the infimum is monotonically decreasing in λ (for $v < P$) and by using $\bar{v} = \frac{v}{1+v}$ (4) and $\lfloor r/v \rfloor \geq r/v - 1$, and (44c) holds since $\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right) \leq -\log\bar{P}$ for all $v < P$. ∎

Appendix B
Proof of Lemma 4

To prove Lemma 4, we first introduce auxiliary definitions and results. We will make use of the following notion of sub-Gaussianity as it appears in [44, Ch. 2.1].[9]

Definition 1: An RV X with mean μ is sub-Gaussian with *variance proxy* $\sigma^2(X) > 0$ if

$$\mathbb{E}\left[\exp\{\lambda(X-\mu)\}\right] \leq \exp\left\{\frac{\sigma^2(X)\lambda^2}{2}\right\}\forall\lambda \in \mathbb{R}.$$

Clearly, if X is sub-Gaussian with variance proxy $\sigma^2(X)$, it is also sub-Gaussian with any variance proxy greater than $\sigma^2(X)$. The following well-known properties follow immediately from Definition 1 and Chernoff's bound.

Property 1:
1) A Gaussian RV X with variance σ^2 is sub-Gaussian with $\sigma^2(X) = \sigma^2$.
2) A Rademacher RV (uniformly distributed over $\{\pm 1\}$) X is sub-Gaussian with $\sigma^2(X) = 1$.
3) Let $X_1, \ldots, X_n$ be independent sub-Gaussian RVs with variance proxies $\sigma_1^2, \ldots, \sigma_n^2$, respectively. Let $a_1, \ldots, a_n \in \mathbb{R}$ be some constants. Then, $\sum_{i=1}^n a_i X_i$ is sub-Gaussian with $\sum_{i=1}^n a_i^2\sigma_i^2$.
4) *Chernoff–Hoeffding bound:* Let X be a zero-mean sub-Gaussian RV with variance proxy $\sigma^2(X) > 0$, and let a be some positive constant. Then,

$$\Pr(|X| \geq a) \leq 2\exp\left\{-\frac{a^2}{2\sigma^2(X)}\right\}.$$

We now return to Scheme 2 and prove the following property on $\mathcal{E}_r^\ell(t)$ of (28).

Lemma 9: $\mathcal{E}_r^\ell(t)$ is sub-Gaussian with variance proxy that satisfies $\sigma^2\left(\mathcal{E}_r^\ell(t)\right) \leq \mathrm{MSE}_r(t)$.

Proof: The inputs to the overall system are the noise samples $\left\{Z_r(t)\middle|r, t \in \mathbb{Z}_{\geq 0}\right\}$ and S^ℓ, all of which are independent. The noise samples are Gaussian, whereas S^ℓ equals a linear combination of independent Rademacher RVs by (25) (which are also independent of the Gaussian noise samples), and hence also $\mathcal{E}_r^\ell(t)$. By parts 1 and 2 of Property 1, for all the elements in the linear combination, the variance proxy equals

[9]Other definitions that are equivalent up to constants exist, e.g., in [45, Ch. 2.5].

the variance, while by part 3 the same holds for $\mathcal{E}_r^\ell(t)$. Thus, $\mathcal{E}_r^\ell(t)$ is sub-Gaussian with

$$\sigma^2\Big(\mathcal{E}_r^\ell(t)\Big) = \mathbb{E}\bigg[\Big\{\mathcal{E}_r^\ell(t)\Big\}^2\bigg] \leq \mathrm{MSE}_r(t),$$

where the inequality follows from (29). ■

Proof of Lemma 4:

$$\begin{aligned}\epsilon_r(0{:}\ell|t) &\leq \Pr\Big(\Big|\mathcal{E}_r^\ell(t)\Big| > d_\ell/2\Big) && \text{(45a)}\\ &\leq 2\exp\left\{-\frac{d_\ell^2/4}{2\mathrm{MSE}_r(t)}\right\} && \text{(45b)}\\ &\leq 2\exp\left\{-\frac{3}{2^{2\ell+1}\cdot \mathrm{MSE}_r(t)}\right\}, && \text{(45c)}\end{aligned}$$

where (45a) follows from (30a), (45b) follows from the Chernoff–Hoeffding bound (recall Property 1) and Lemma 9, and (45c) follows from (26). ■

Appendix C
Proof of Lemma 5

To make the analysis devoid of ℓ, we will analyze the error probability of S^n in the infinite-constellation limit S (27). Due to the linearity of all the components of the scheme until the final decoding of the bits, one can always attain the performance of decoding S^n from a finite-constellation $\ell \in \mathbb{N}$ by adding, to the term that is proportional to S^ℓ in the decoding error, a uniform noise U^ℓ over $[-d_\ell/2, d_\ell/2)$ that is independent of S^ℓ with the same coefficient, such that S is uniform over $[-\sqrt{3}, \sqrt{3})$:

$$S = S^\ell + U^\ell, \tag{46}$$

with S^ℓ as in (25) and

$$U^\ell = \sqrt{3}\sum_{i=\ell}^{\infty}(-1)^{B(i)}2^{-(i+1)}; \tag{47}$$

this will be made precise in the sequel in (51).

Similarly, the decomposition (46) applies for any n:

$$S = S^n + U^n, \tag{48}$$

where U^n, given in (47) with ℓ replaced by n, is independent of S^n and is uniformly distributed over $[-d_n/2, d_n/2)$ with

$$d_n = \sqrt{3}\cdot 2^{-n+1} \tag{49}$$

being the spacing between two constellation levels of S^n, in agreement with (26).

Now notice that in Scheme 1, the estimate of S^ℓ at node $r \in \mathbb{N}$ at time $t \in \mathbb{Z}_{\geq 0}$, is given by

$$\hat{S}_r^\ell(t) = \alpha_r(t)\cdot S^\ell + Z_r^{\mathrm{eff}}(t), \tag{50}$$

where $Z_r^{\mathrm{eff}}(t)$ is a combination of channel noises, independent of S^ℓ, and by the properties of MMSE estimation (recall Lemma 1) $\alpha_r(t) \in (0,1)$. By (46), adding $\alpha_r(t)U^\ell$ to $\hat{S}_r^\ell(t)$ results in

$$\hat{S}_r(t) = \alpha_r(t)\cdot S + Z_r^{\mathrm{eff}}(t). \tag{51}$$

To bound the error probability of decoding S^n at node r at time t, we will analyze a suboptimal decoder that

1) generates the estimate $\hat{S}_r^\ell(t)$ of step 3 of Scheme 2;
2) generates a randomly generated noise U^ℓ as in (47), which is uniformly distributed over $[-d_\ell/2, d_\ell/2)$ and is independent of S^ℓ and $\{Z_r(t)\}$;
3) adds $\alpha_r(t)U^\ell$ to $\hat{S}_r^\ell(t)$ [recall (50)] to generate $\hat{S}_r(t)$ of (51);
4) decodes S^n from $\hat{S}_r(t)$ using a nearest-neighbor decoder (slicer) of S^n.

The prefix error probability of S^n is bounded from above as follows.

$$\begin{aligned}\epsilon_r(0{:}\ell|t) &\leq \Pr\Big(\Big|S^n - \hat{S}_r(t)\Big| \geq d_n/2\Big) && \text{(52a)}\\ &= \Pr\bigg(\Big|(1-\alpha_r(t))S^n - \alpha_r(t)U^n - Z_r^{\mathrm{eff}}(t)\Big| \geq \frac{d_n}{2}\bigg) && \text{(52b)}\\ &= \int_{-\frac{d_n}{2}}^{\frac{d_n}{2}} \Pr\bigg(\Big|(1-\alpha_r(t))S^n - \alpha_r(t)u - Z_r^{\mathrm{eff}}(t)\Big| \geq \frac{d_n}{2}\bigg)\frac{\mathrm{d}u}{d_n} && \text{(52c)}\\ &\leq \int_{-\frac{d_n}{2}}^{\frac{d_n}{2}} \Pr\bigg(\Big|(1-\alpha_r(t))S^n - Z_r^{\mathrm{eff}}(t)\Big| \geq \frac{d_n}{2} - \alpha_r(t)|u|\bigg)\frac{\mathrm{d}u}{d_n} && \text{(52d)}\end{aligned}$$

where (52a) follows from the nearest-neighbor decision rule and since the values that S^n can take are distant by at least d_n, (52b) follows from (48) and (51), (52c) holds since U^n is uniformly distributed over $[-d_n/2, d_n/2)$ and is independent of $\big(Z_r^{\mathrm{eff}}(t), S^n\big)$, (52d) follows from the triangle inequality by noting that $0 < \alpha_r(t) < 1$.

Now define $\mathcal{E}_r^n(t) = (1-\alpha_r(t))S^n - Z_r^{\mathrm{eff}}(t)$. Note that $\mathcal{E}_r^n(t)$ can be interpreted as the estimation error $S^n - \hat{S}_r^n(t)$ as in (28), where $\hat{S}_r^n(t)$ is the estimate of S^n if the packet were of size n. Hence, $\epsilon_r(0{:}\ell|t)$ can be further bounded as follows.

$$\begin{aligned}\epsilon_r(0{:}\ell|t) &= \frac{2}{d_n}\int_{[1-\alpha_r(t)]d_n/2}^{d_n/2} \Pr\Big(\big|\mathcal{E}_r^n(t)\big| \geq x\Big)dx && \text{(53a)}\\ &\leq \frac{2}{d_n}\int_0^\infty \Pr\Big(\big|\mathcal{E}_r^n(t)\big| \geq x\Big)dx && \text{(53b)}\\ &\leq \frac{2}{d_n}\int_0^\infty \min\left\{1, \frac{\mathrm{MSE}_r(t)}{x^2}\right\}dx && \text{(53c)}\\ &\leq \frac{2}{d_n}\left[\int_0^{\sqrt{\mathrm{MSE}_r(t)}} 1dx + \int_{\sqrt{\mathrm{MSE}_r(t)}}^\infty \frac{\mathrm{MSE}_r(t)}{x^2}dx\right] && \text{(53d)}\\ &= \frac{4}{d_n}\sqrt{\mathrm{MSE}_r(t)} && \text{(53e)}\\ &= \frac{2}{\sqrt{3}}\cdot 2^n\cdot\sqrt{\mathrm{MSE}_r(t)}, && \text{(53f)}\end{aligned}$$

where (53a) follows from (52) by noting that the integrand is an even function, by the definition of $\mathcal{E}_r^n(t)$, and by integration by substitution with $x = \frac{d_n}{2} - \alpha u$; (53b) holds since the integrand is non-negative (and $0 < \alpha_r(t) < 1$); (53c) holds by Chebyshev's inequality and (29) (the same argument applies here for n in lieu of ℓ), and since the probability is trivially bounded from above by 1; and (53f) follows from (49). ■

APPENDIX D
PROOF OF THEOREM 4

When analyzing Scheme 3, all of the analysis that we applied to Scheme 1 up to the recursion relation (21) in Lemma 2 remains valid. In particular, with the parameters $\beta_r(t)$ and $\gamma_r(t)$ of (20), $\{\hat{S}_r(t)\}$ are the LMMSE estimators from all available channel outputs (which are uncorrelated), and the power constraint is always satisfied with equality. To see this, one may go through the proofs and see that the assumption on the nature of $\{\hat{S}_0(t)\}$ was never used. Consequently, the MSE recursion of (21) remains valid for Scheme 3, albeit with different boundary conditions (32) [and the same initial conditions $\mathrm{MSE}_r(-1) = 1$, for all $r \in \mathbb{Z}_{\geq 0}$]. We first provide an explicit expression for the MSE, parallel to (22):

$$\mathrm{MSE}_r(t) = \mathrm{MSE}_r^{\mathrm{I}}(t) + \mathrm{MSE}_r^{\mathrm{II}}(t), \tag{54a}$$

where

$$\mathrm{MSE}_r^{\mathrm{I}}(t) = \left(1-\bar{P}\right)^{t+1} \sum_{k=1}^{r} \binom{t+r-k}{r-k} \bar{P}^{r-k}, \tag{54b}$$

$$\mathrm{MSE}_r^{\mathrm{II}}(t) = \bar{P}^r \sum_{s=0}^{t-1} \exp\{-2R(t-s)\} \binom{r+s-1}{s} \left(1-\bar{P}\right)^s. \tag{54c}$$

This MSE follows by summing the contributions of all the initial conditions, as done when proving (22). $\mathrm{MSE}_r^{\mathrm{I}}(t)$ is the sum of contributions of the initial conditions $\mathrm{MSE}_r(-1) = 1$ for $r > 0$, and indeed it equals the MSE of (22). $\mathrm{MSE}_r^{\mathrm{II}}(t)$ is the sum of contributions of the boundary conditions $\mathrm{MSE}_0(t) = 1$ for $t > 0$, which can be derived in a dual way to the first contribution, substituting the roles as follows: $r \to t$, $t \to r-1$, $\bar{P} \to 1-\bar{P}$, taking into account the different boundary conditions.

Now we proceed to the second stage, asymptotics. Since $\mathrm{MSE}_r^{\mathrm{I}}(t)$ is identical to single-source MSE of (22), by Theorem 1, we have that

$$\mathrm{MSE}_r^{\mathrm{I}}\left(\left\lfloor \frac{r}{t} \right\rfloor\right) \leq \exp\{-rE_1(v) + o(r)\},$$

where $E_1(v)$ is defined in (24).

For $\mathrm{MSE}_r^{\mathrm{II}}(t)$, we proceed in a similar manner to the proof of Theorem 1:

$$\begin{aligned}\mathrm{MSE}_r^{\mathrm{II}}(t) &= \bar{P}\exp\{-2Rt\} \sum_{s=0}^{t-1} \binom{r+s-1}{s} \exp\{2Rs\} \\ &\quad \cdot \exp\left\{-(r+s-1)\left[\mathbb{D}\left(\frac{s}{r+s-1}\middle\| 1-\bar{P}\right) + h\left(\frac{s}{r+s-1}\right)\right]\right\} \\ &\leq \exp\{-2Rt\} \\ * &\quad \cdot \sum_{s=0}^{t-1} \exp\left\{-\left[(s+r-1)\mathbb{D}\left(\frac{s}{s+r-1}\middle\| 1-\bar{P}\right) - 2Rs\right]\right\}.\end{aligned}$$

Denote the summands as $a_s = \exp\left\{-(r-1)\tilde{E}\left(\frac{s}{r-1}\right)\right\}$ for

$$\tilde{E}(\delta) = (1+\delta)\mathbb{D}\left(\bar{\delta}\middle\| 1-\bar{P}\right) - 2R\delta. \tag{55}$$

Proceeding as in the proof of Theorem 1 would result in a correction term $o(t)$. Instead, we derive a bound with a correction term $o(r)$ that would prove more favorable. To that end, notice that $\tilde{E}(\delta)$ is a convex function, with derivative

$$\tilde{E}'(\delta) = \log\left(\frac{\bar{\delta}}{1-\bar{P}}\right) - 2R = \log\left(\frac{\bar{\delta}}{\eta}\right),$$

which attains its (unconstrained) minimum at $\delta^* = \frac{\eta}{1-\eta}$ where η was defined in (10).

Now, set s^* to be the index of the maximal element a_s in the sum. We have two cases:

1) If $(r-1)\delta^* \geq t-1$ the maximum is attained on the edge $s^* = t-1$, and

$$\sum_{s=0}^{t-1} a_s \leq t a_{s^*} \leq [(r-1)\delta^* + 1]\, a_{s^*}.$$

2) The maximum is attained at an internal point such that $|s^* - (r-1)\delta^*| \leq 1$, and

$$\sum_{s=0}^{t-1} a_s \leq \sum_{s=0}^{\infty} a_s = \sum_{s=0}^{2s^*+1} a_s + \sum_{s=2(s^*+1)}^{\infty} a_s. \tag{56}$$

The first term in (56) is bounded from above as

$$\sum_{s=0}^{2s^*+1} a_s \leq 2\left(s^*+1\right)a_{s^*} \leq 2\left[(r-1)\delta^* + 2\right]a_{s^*}.$$

To bound the second term, notice that, by convexity,

$$\tilde{E}\left(\frac{s}{r-1}\right) \geq \tilde{E}\left(\frac{2(s^*+1)}{r-1}\right) + \frac{s-2(s^*+1)}{r-1} \tilde{E}'\left(\frac{2(s^*+1)}{r-1}\right);$$

by invoking convexity and $|s^* - (r-1)\delta^*| \leq 1$, we can further bound the derivative as

$$\tilde{E}'\left(\frac{2(s^*+1)}{r-1}\right) \geq \tilde{E}'\left(2\delta^*\right) = \log\frac{2}{1+\eta}.$$

Consequently, all the elements in the second sum in (56) are bounded from above as

$$a_s \leq a_{2(s^*+1)} \left(\frac{1+\eta}{2}\right)^{s-2(s^*+1)},$$

and the infinite sum itself is bounded as

$$\begin{aligned}\sum_{s=2(s^*+1)}^{\infty} a_s &\leq \sum_{s=2(s^*+1)}^{\infty} a_{2(s^*+1)} \left(\frac{1+\eta}{2}\right)^{s-2(s^*+1)} \\ &\leq \frac{2}{1-\eta} a_{2s^*},\end{aligned}$$

where the geometric sum is finite since $\eta < 1$ by definition (10). Since a_{2s^*} decays exponentially with r with an exponent that is greater than that of a_s^*, $a_{2s^*} = o_r(a_{s^*})$.

Hence, there exists a *linear* function $c(r)$ that does not depend on v (for both cases), such that $\sum_{s=0}^{t-1} a_s \leq c(r) a_{s^*}$. By defining $E_2(v) \triangleq v \inf_{\delta \in (0,\frac{1}{v}]} \tilde{E}(\delta) - 2R$ with $\tilde{E}(\cdot)$ of (55), it follows that

$$\mathrm{MSE}_r^{\mathrm{II}}\left(\left\lfloor \frac{r}{v} \right\rfloor\right) \leq \exp\{-rE_2(v) + o(r)\},$$

where the redundancy term $o(r)$ is independent of v.

Since the infimum is δ^* for $v < \eta/(1-\eta)$ and the edge $1/v$ otherwise, $E_2(v) = E_S(v)$ for all v. Since $E_2(v) \le E_1(v)$ for all v, we obtain the desired result:

$$\begin{aligned}\mathrm{MSE}_r\left(\left\lfloor \frac{r}{v} \right\rfloor\right) &\le \exp\{-rE_1(v) + o(r)\} + \exp\{-rE_2(v) + o(r)\} \\ &= \exp\{-r\min(E_1(v), E_2(v)) + o(r)\} \\ &= \exp\{-rE_S(v) + o(r)\},\end{aligned}$$

where the correction term $o(r)$ is independent of v. ■

Appendix E
Proofs for Section VI-B

Proof of Lemma 6: Consider first the boundary condition. By the construction of Scheme 4,

$$\begin{aligned}\mathrm{MSE}_0(t) &= \mathbb{E}\left[\left[S - \hat{S}_0(t)\right]^2\right] = \mathbb{E}\left[\left[S - S^{\ell\lfloor \frac{t}{T} \rfloor}\right]^2\right] \\ &= \mathbb{E}\left[\left(U^{\ell\lfloor \frac{t}{T} \rfloor}\right)^2\right] = \frac{\left(d_{\ell\lfloor \frac{t}{T} \rfloor}\right)^2}{12} = 2^{-2\ell\lfloor \frac{t}{T} \rfloor},\end{aligned}$$

where U^n was defined in (47) for any $n \in \mathbb{N}$ and is uniformly distributed over $[-d_n/2, d_n/2)$, and d_n was defined in (49). Recalling the rate definition (6), we have

$$\mathrm{MSE}_0(t) \le \exp\{-2Rt\}.$$

Thus, indeed the MSE at node r and time t is at most $\mathrm{MSE}_r(t)$ of Scheme 3 with source refinement rate R that equals our packet streaming rate R. This MSE, $\mathrm{MSE}_r(t)$, in turn, is given by (54), and is bounded as in Theorem 4.

We now bound the maximal bit error probability (11) as follows.

$$\begin{aligned}P_e(r, \Delta) &\le \sup_{\tau \in \mathbb{Z}_{\ge 0}} \epsilon_r(0{:}(\tau+1)\ell - 1 | \tau T + \Delta) &\text{(57a)} \\ &\le \sup_{\tau \in \mathbb{Z}_{\ge 0}} \frac{2}{\sqrt{3}} \cdot 2^{(\tau+1)\ell - 1} \cdot \sqrt{\mathrm{MSE}_r(\tau T + \Delta)}, &\text{(57b)}\end{aligned}$$

where (57a) follows from (16); and (57b) follows from Lemma 5 with $\mathrm{MSE}_r(t)$ of Scheme 3, by nothing that its proof in Appendix C holds for any boundary conditions $\hat{S}_0(t)$.

To obtain the required asymptotic bound, we substitute $r = \lceil v\Delta \rceil$ in (57):

$$\begin{aligned}P_e(\lceil v\Delta \rceil, \Delta) &\le \sup_{\tau \in \mathbb{Z}_{\ge 0}} \frac{2^{(\tau+1)\ell}}{\sqrt{3}} \sqrt{\mathrm{MSE}_{\lceil v\Delta \rceil}(\tau T + \Delta)} &\text{(58a)} \\ &= \sup_{\tau \in \mathbb{Z}_{\ge 0}} \exp\left\{R\tau T - \frac{\tau T + \Delta}{2v} E_S\left(\frac{v\Delta}{\tau T + \Delta}\right) + o(\Delta)\right\} &\text{(58b)} \\ &\le \sup_{t_0 \in \mathbb{Z}_{\ge 0}} \exp\left\{Rt_0 - \frac{t_0 + \Delta}{2v} E_S\left(\frac{v\Delta}{t_0 + \Delta}\right) + o(\Delta)\right\}, &\text{(58c)}\end{aligned}$$

where (58a) follows from (57); (58b) follows from (6), Theorem 4, and noting that $E_s(\cdot)$ is a smooth function; and (58c) holds since $T\mathbb{Z}_{\ge 0} \subset \mathbb{Z}_{\ge 0}$. The required result follows by taking the $o(\Delta)$ correction term out of the optimization. This holds true since the correction term is independent of the argument of $E_S(\cdot)$, as shown in Theorem 4. ■

Proof of Theorem 5: Recalling (10), we bound $P_e(v, \lfloor r/v \rfloor)$, for

$$v \le \exp\{2(C - R)\} - 1 = \frac{1-\eta}{\eta} \tag{59}$$

as follows.

$$\begin{aligned}&P_e\left(r, \left\lfloor \frac{r}{v} \right\rfloor\right) \\ &\le \exp\left\{-\inf_{t_0 \in \mathbb{Z}_{\ge 0}} \left[\frac{t_0 + \Delta}{2v} E_S\left(\frac{v\Delta}{t_0 + \Delta}\right) - t_0 R\right] + o(r)\right\} \\ &= \exp\left\{-\left[\frac{1}{1-\eta}\left(\frac{\mathbb{D}(1-\eta\|\bar{P})}{2} - \eta R\right) + R\right]\Delta + o(r)\right\},\end{aligned}$$

where the inequality follows from Lemma 6, and the equality holds for v that satisfies (59) by noting that the argument of $E_S(\cdot)$ is at most v in this case. Now note that the exponent in the second expression at $v = \frac{1-\eta}{\eta}$ is positive:

$$\frac{1}{1-\eta}\left(\frac{\mathbb{D}(1-\eta\|\bar{P})}{2} - \eta R\right) + R\Big|_{v = \frac{1-\eta}{\eta}} = \frac{\mathbb{D}(1-\eta\|\bar{P})}{2(1-\eta)} > 0.$$

Since $P_e(r, \lfloor r/v \rfloor)$ is monotonically increasing with v, $\lim_{\Delta \to \infty} P_e(r, \lfloor r/v \rfloor) = 0$ for all v that satisfy (59). Thus all these velocities are achievable, and the proof is concluded. ■

References

[1] E. Domanovitz, A. Khina, T. Philosof, and Y. Kochman, "Information velocity of cascaded AWGN channels with feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2024, pp. 1–35.

[2] S. G. Waxman, J. D. Kocsis, and P. K. Stys, *The Axon: Structure, Function, and Pathophysiology*. Oxford, U.K.: Oxford Univ., 1995.

[3] J.-A. Girault and E. Peles, "Development of nodes of Ranvier," *Current Opin. Neurobiol.*, vol. 12, no. 5, pp. 476–485, 2002.

[4] T. Berger and W. B. Levy, "A mathematical theory of energy efficient neural computation and communication," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 852–874, Feb. 2010.

[5] P. Suksompong and T. Berger, "Capacity analysis for integrate-and-fire neurons with descending action potential thresholds," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 838–851, Feb. 2010.

[6] J. Heinovski and F. Dressler, "Platoon formation: Optimized car to platoon assignment strategies and protocols," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, 2018, pp. 1–8.

[7] T. Nawaz, M. Seminara, S. Caputo, L. Mucchi, F. S. Cataliotti, and J. Catani, "IEEE 802.15. 7-compliant ultra-low latency relaying VLC system for safety-critical ITS," *IEEE Trans. Veh. Tech.*, vol. 68, no. 12, pp. 12040–12051, Dec. 2019.

[8] W. Huleihel, Y. Polyanskiy, and O. Shayevitz, "Relaying one bit across a tandem of binary-symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, 2019, pp. 2928–2932.

[9] V. Jog and P.-L. Loh, "Teaching and learning in uncertainty," *IEEE Trans. Inf. Theory*, vol. 67, no. 1, pp. 598–615, Jan. 2021.

[10] Y. H. Ling and J. Scarlett, "Optimal rates of teaching and learning under uncertainty," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7067–7080, Aug. 2021.

[11] Y. H. Ling and J. Scarlett, "Multi-bit relaying over a tandem of channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 6, pp. 3511–3524, Jun. 2023.

[12] S. Rajagopalan and L. Schulman, "A coding theorem for distributed computation," in *Proc. ACM Symp. Theory Comput. (STOC)*, 1994, pp. 790–799.

[13] S. K. Iyer and R. Vaze, "Achieving non-zero information velocity in wireless networks," in *Proc. Symp. Model. Opt. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, 2015, pp. 584–590.

[14] Y. H. Ling and J. Scarlett, "Simple coding techniques for many-hop relaying," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7043–7053, Jun. 2022.

[15] M. Fidler, "Survey of deterministic and stochastic service curve models in the network calculus," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 1, pp. 59–86, 1st Quart., 2010.

[16] M. Fidler and A. Rizk, “A guide to the stochastic network calculus,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 92–105, 1st Quart., 2015.

[17] C.-S. Chang, *Performance Guarantees in Communication Networks*. London, U.K.: Springer, 2000.

[18] M. Fidler, “An end-to-end probabilistic network calculus with moment generating functions,” in *Proc. IEEE Int. Workshop Qual. Service (IWQoS)*, 2006, pp. 261–270.

[19] E. Domanovitz, T. Philosof, and A. Khina, “The information velocity of packet-erasure links,” in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2022, pp. 1–11.

[20] R. Inovan, “On speed and advantage: Results in information velocity and monitoring problems,” Ph.D. dissertation, School Comput. Commun. Sci., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, May 2024.

[21] A. Lalitha, A. Khina, T. Javidi, and V. Kostina, “Real-time binary posterior matching,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, 2019, pp. 2239–2243.

[22] J. P. M. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback–I: No bandwidth constraint,” *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[23] J. P. M. Schalkwijk, “A coding scheme for additive noise channels with feedback–II: Band-limited signals,” *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, Apr. 1966.

[24] R. G. Gallager and B. Nakiboğlu, “Variations on a theme by Schalkwijk and Kailath,” *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Dec. 2009.

[25] P. Elias, “Channel capacity without coding,” *Proc. IRE*, vol. 45, no. 3, pp. 381–381, Jan. 1957.

[26] B. Schein and R. G. Gallager, “The Gaussian parallel relay channel,” in *Proc. Int. Symp. Inf. Theory (ISIT)*, Sorrento, Italy, 2000, pp. 194–198.

[27] J. N. Laneman, D. N. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Nov. 2004.

[28] Y. Kochman, A. Khina, U. Erez, and R. Zamir, “Rematch-and-forward: Joint source/channel coding for parallel relaying with spectral mismatch,” *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 605–622, Jan. 2014.

[29] H. Yamamoto and K. Itoh, “Source coding theory for multiterminal communication systems with a remote source,” *IECE Tran. Jpn.*, vol. E63, no. 10, pp. 700–706, 1980.

[30] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem [multiterminal source coding],” *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[31] A. B. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic Gaussian two-encoder source-coding problem,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[32] S. C. Draper and G. W. Wornell, “Successively structured CEO problems,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Lausanne, Switzerland, 2002, p. 65.

[33] S. C. Draper and G. W. Wornell, “Side information aware coding strategies for sensor networks,” *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 966–976, Aug. 2004.

[34] J. Chen, X. Zhang, T. Berger, and S. Wicker, “Rate allocation in distributed sensor network,” in *Proc. Allerton Conf. Commun., Control, Comput.*, vol. 41, 2003, pp. 531–540.

[35] W. H. R. Equitz and T. M. Cover, “Successive refinement of information,” *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 851–857, Mar. 1991.

[36] L. Lastras and T. Berger, “All sources are nearly successively refinable,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 918–926, Mar. 2001.

[37] E. Domanovitz, A. Khisti, W.-T. Tan, X. Zhu, and J. Apostolopoulos, “Streaming erasure codes over multi-hop relay network,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 497–502.

[38] S. Singhvi, R. Gayathri, and P. V. Kumar, “Rate-optimal streaming codes over the three-node decode-and-forward relay network,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2022, pp. 1957–1962.

[39] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill Educ., 2002.

[40] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[41] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ., 2011.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[43] A. Ben-Yishai and O. Shayevitz, “The Gaussian channel with noisy feedback: Improving reliability via interaction,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, pp. 2500–2504.

[44] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. 48. Cambridge, U.K.: Cambridge Univ., 2019.

[45] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ., 2018.

Elad Domanovitz (Member, IEEE) received the B.Sc. (cum laude), M.Sc., and Ph.D. degrees in electrical engineering from Tel Aviv University, Israel, in 2005, 2011, and 2020, respectively, where he is currently a Postdoctoral Fellow with the School of Electrical Engineering. His research interests include information theory, communication theory, and statistical signal processing.

Anatoly Khina (Member, IEEE) received the B.Sc. (cum laude), M.Sc. (cum laude), and Ph.D. degrees from Tel Aviv University in 2006, 2010, and 2016, respectively, all in electrical engineering, where he is a Faculty Member with the School of Electrical Engineering. Alongside his studies, he worked as an engineer in various roles focused on algorithms, data science, software, and hardware. He was a Postdoctoral Scholar with the Department of Electrical Engineering, California Institute of Technology from 2015 to 2018, and a Research Fellow with the Simons Institute for the Theory of Computing, University of California Berkeley, Berkeley, during the Spring of 2018. He currently serves as an Officer of the IEEE Information Theory Israel Section Chapter.

Tal Philosof received the B.Sc. degree (cum laude) in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from Tel Aviv University, Israel, in 2003 and 2010, respectively. From 2013 to 2017, he was a Senior Researcher with the General Motors Advanced Research Center, focusing on wireless communication and connected vehicles. Since 2017, he has been with Samsung Research Center.

Yuval Kochman (Member, IEEE) received the B.Sc. (summa cum laude), M.Sc. (summa cum laude), and Ph.D. degrees in electrical engineering from Tel Aviv University in 1993, 2003, and 2010, respectively. From 2009 to 2011, he was a Postdoctoral Associate with the Signals, Information and Algorithms Laboratory, MIT. Since 2012, he has been with the School of Computer Science and Engineering, Hebrew University of Jerusalem. Outside academia, he has worked in the areas of radar and digital communications. His research interests include information theory, communications, and signal processing.