2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619138

Information Velocity of Cascaded AWGN Channels with Feedback

Elad Domanovitz, Anatoly Khina, Tal Philosof, and Yuval Kochman

***Abstract*—We consider a line network of nodes connected by additive white Gaussian noise channels and equipped with local feedback. We study the velocity at which information spreads over this network. For the transmission of a data packet, we derive an explicit positive lower bound on the velocity for any packet size. Furthermore, we consider streaming, that is, transmission of data packets that is 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. Due to the analog-linear nature of the scheme, the results extend to any additive noise. For general noise, we derive exponential error-probability bounds. Moreover, for (sub-)Gaussian noise, we show doubly-exponential behavior, which reduces to the celebrated Schalkwijk–Kailath scheme when considering a single node. By 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.**

I. Introduction

The demand for real-time communication over large-scale networks is steadily increasing due to the growing size and the distributed nature of modern-day technology. A central problem of interest in this context is transmission over a cascade of channels interconnected by relaying nodes.

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 [1], [2], and was eventually proved by Ling and Scarlett [3] to equal to that of a single channel. They further extended the scope to any finite number of messages in [4].

Setting	Theorem	Instantaneous Hops	Delayed Hops
Single packet	Th. 2	P	$\frac{P}{1+P}$
Packet streaming	Th. 5	$\exp\{2(C-R)\}-1$	$1-\exp\{-2(C-R)\}$

TABLE I: Main results. $C \triangleq \frac{1}{2}\log(1+P)$ denotes the channel capacity of an individual additive white-Gaussian noise channel with SNR P, and R is the data-generation rate in the streaming setting.

The behavior of large networks can be expressed by taking the number of channels to grow linearly with the number of time steps. The minimum ratio between the two, such that the error probability may be arbitrarily small, was termed *Information Velocity* (IV) by Polyanskiy (see [1], [5]) in the single-bit context. The same term was used earlier by Iyer and Vaze [6] in a related setting of spatial wireless networks.[1]

Analyzing the IV of the transmission of a single message does not capture the entire behavior of the network: In case of a stream of messages, there is an inherent tension between the velocity across the network and the error probability of different messages. Analyzing the IV for an infinite stream of messages is, therefore, more challenging. 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 [7]–[10] (these results also hold for a single packet without feedback). In [11], 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 the 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 (say Gaussian) channels with the same signal-to-noise ratio (SNR) $P > 0$, equipped with perfect and local feedback. That is, each node is aware of the measurements of the next node

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

T. 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, Hebrew University of Jerusalem, Jerusalem 9190401, Israel (email: yuvalko@cs.huji.ac.il).

The work of E. Domanovitz and A. Khina was supported by the Israel Science Foundation (grant No. 2077/20) and supported by a grant from the Tel Aviv University Center for AI and Data Science (TAD).

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

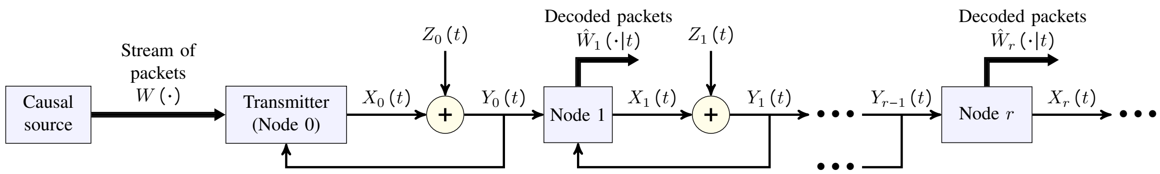

Fig. 1: Block diagram of the system. $X_r(t)$, $Y_r(t)$, and $Z_r(t)$ are the channel input, output, and noise, respectively, at node r at time t. Each T time steps, a new packet is generated. At every time step, all the nodes decode all the hitherto arrived packets.

in a causal manner; see Figure 1. For this setting, we derive lower bounds on both the single-packet and the streaming IVs, as summarized in Table I. In particular, the streaming IV is positive for any $R < C$, and the bound on the streaming IV reduces to that on the single-packet IV in the limit of $R \to 0$.[2]

Reminiscent of [13], our approach is based on the source node translating the incremental message history into a real number. 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 Schalkwijk–Kailath scheme for feedback communication over an additive-noise channel [14], [15], which can be viewed [16] as an application of the joint source–channel coding scheme with feedback of Elias [17]. On the other hand, the first transmission of each of the nodes reduces to scalar amplify-and-forward relaying [18].

Treating a representation of a data message as an analog source is a concept that has proved useful in relaying schemes [18]–[20]. 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 is reminiscent of the sequential Gaussian CEO problem [21]–[23]. In our solutions, the source node builds its scalar representation of the data stream as if an analog source were revealed to it via successive refinement [24].

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 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 [14], [15] to multiple relays.

The proofs of all the statements are available in [25], along with plots and a more detailed exposition.

II. Problem Statement

$\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}$ denote the sets of integer, non-negative integer, real, non-negative real, and natural numbers, respectively. Logarithms and exponents are taken to the natural base. $f(x) = o(g(x))$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. We denote $\bar{a} = \frac{a}{1+a}$ for $a \in \mathbb{R}_{\geq 0}$. For sequence a and for $t_1, t_2 \in \mathbb{Z}_{\geq 0}$, $t_1 \leq t_2$, denote $a(t_1 : t_2) \triangleq [a(t_1), a(t_1+1), \dots, a(t_2)]$. The binary entropy and binary Kullback–Leibler divergence functions are denoted by $h(p)$ and $\mathbb{D}(p\|q)$. $X \leftrightarrow Y \leftrightarrow Z$ denotes a Markov triplet, viz. X is independent of Z given Y.

A. Communication Model

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 $W(\tau) = B(\tau\ell : (\tau+1)\ell - 1)$ where $B(i) \in \{0, 1\}$. The bits of the entire message sequence, $\{B(i) | i \in \mathbb{Z}_{\geq 0}\}$, are assumed i.i.d. uniform. Define the average rate of this source as

$$R = \frac{\ell}{T} \log(2). \tag{1}$$

Channels. The network is composed of a cascade of channels; channel $r \in \mathbb{Z}_{\geq 0}$ is an additive noise channel:

$$Y_r(t) = X_r(t) + Z_r(t),$$

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 noise sequences $\{Z_r(t) | r, t \in \mathbb{Z}_{\geq 0}\}$ are i.i.d. zero-mean, unit variance, and independent of all channel inputs. All channel inputs are subject to a mean power constraint: $\mathbb{E}\left[X_r^2(t)\right] \leq P, \ \forall r, t \in \mathbb{Z}_{\geq 0}$.

The following channel quantities will be used in the sequel: $C = \frac{1}{2}\log(1+P)$ is the capacity of an AWGN channel with SNR P, and $\eta \triangleq (1-\bar{P}) \cdot \exp\{2R\} = \exp\{-2(C-R)\}$.

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

Originating transmitter (node 0). At each time $t \in \mathbb{Z}_{\geq 0}$, node 0 generates a channel input $X_0(t)$ as a function of the packet history $W(0 : \lfloor t/T \rfloor)$, and of all past outputs of

[2] In the body of the paper, we consider instantenous hops for ease of presentation; the results for both dealyed and instantaneous hops are summarized in Table I. The translation is immediate, by a linear transformation of the velocity, from v in instantaneous hops to $\bar{v} = v/(1+v)$ in delayed hops. The single-packet bound for delayed hops was independently derived by Inovan [12, Ch. 5].

channel 0, $Y_0(0:t-1)$, which are available via feedback. The input is subject to the power constraint P.

Node r ($r \in \mathbb{N}$). At each time $t \in \mathbb{Z}_{\geq 0}$, node r generates the channel input $X_r(t)$ as a function of all its measurement history[3] $Y_{r-1}(0:t)$ from its feeding channel (channel $r-1$), and from all its feedback history from the subsequent channel (channel r), $Y_r(0:t-1)$. The input is subject to the power constraint P. The relays also produce estimates of the message.[4] 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 at time t based on the measurement history $Y_{r-1}(0:t)$ and with respect to (w.r.t.) the 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 w.r.t. $t \in \mathbb{Z}_{\geq 0}$ for a fixed $r \in \mathbb{N}$, and w.r.t. $r \in \mathbb{N}$ 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 for Streaming

Error Probability. We define the *bit error probability* for bit n at relay r at time t as $\epsilon_r(n|t) \triangleq \Pr\left(\hat{B}_r(n|t) \neq B(n)\right)$. 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) [11] as

$$P_e(r,\Delta) \triangleq \sup_{\tau \in \mathbb{Z}_{\geq 0}} \max_{\tau\ell \leq n \leq (\tau+1)\ell - 1} \epsilon_r(n|\tau T + \Delta). \quad (2)$$

The outer maximization (supremum) is carried over all the packets, whereas the inner maximization is carried over all the bits inside a packet. 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, $\lim_{r\to\infty} P_e\left(r, \left\lfloor \frac{r}{v} \right\rfloor\right) = 0$, namely, $\Delta = \lfloor r/v \rfloor$ in (2).

Streaming IV. 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. For simplicity, we assume a periodic arrival pattern, to wit, R nats are generated by the source every time step.

III. Transmission of a Single Source Sample

In this section, we consider the problem where instead of the data source, we transmit a single zero-mean unit-variance source sample S over the network under a minimum MSE criterion. Let $\hat{S}_r(t)$ denote the source estimate at node r at time t. We will analyze $\mathrm{MSE}_r(t) \triangleq E\left[\left(S - \hat{S}_r(t)\right)^2\right]$. We propose the following simple *linear* scheme, which is a natural extension of the single-channel scheme of Elias [17].

[3]The current measurement included, that is, we assume *instantaneous* hops.

[4]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.

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: $\hat{S}_r(t) = \hat{S}_r(t-1) + \gamma_r(t) Y_{r-1}(t)$, where $\gamma_r(t)$ is a linear MSE (LMMSE) constant.
Transmission by node $r \in \mathbb{Z}_{\geq 0}$. At each time $t \in \mathbb{Z}_{\geq 0}$

$$X_r(t) = \beta_r(t)\left[\hat{S}_r(t) - \hat{S}_{r+1}(t-1)\right], \quad (3)$$

where $\beta_r(t)$ is a power-normalization constant.

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 (3) of node r at time t.

Lemma 1. *In Scheme 1, all channel inputs and outputs and all estimates have zero mean. Setting $\gamma_r(t) = \frac{\mathrm{Cov}(S, X_{r-1}(t))}{1+\mathrm{Var}(X_{r-1}(t))}$, we have the following for $r,t,\tau \in \mathbb{Z}_{\geq 0}$.*

1) *$\hat{S}_r(t)$ is the LMMSE estimate of S from $Y_{r-1}(0:t)$.*
2) *$\mathrm{Cov}(Y_r(t), Y_r(\tau)) = 0$ for $t \neq \tau$.*
3) *$\mathrm{Cov}(S, X_r(t)) = \beta_r(t)\left[\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)\right]$.*
4) *$\mathrm{Var}(X_r(t)) = \beta_r^2(t)\left[\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)\right]$.*

The proof is based on properties of LMMSE estimation, primarily the orthogonality principle [26, Ch. 7-3].

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 $r,t \in \mathbb{Z}_{\geq 0}$,

$$\beta_r(t) = \sqrt{\frac{P}{\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t)}}, \quad (4a)$$

$$\gamma_{r+1}(t) = \frac{\sqrt{P(\mathrm{MSE}_{r+1}(t-1) - \mathrm{MSE}_r(t))}}{P+1}. \quad (4b)$$

Further calculations yield the MSE of the scheme, as follows.

Lemma 2. *Scheme 1 with the parameters of* (4) *satisfies the recursion for all $r \in \mathbb{N}$, $t \in \mathbb{Z}_{\geq 0}$*

$$\mathrm{MSE}_r(t) = \bar{P}\,\mathrm{MSE}_{r-1}(t) + (1-\bar{P})\,\mathrm{MSE}_r(t-1),$$

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 for all $r,t \in \mathbb{N}$, is

$$\mathrm{MSE}_r(t) = \left(1-\bar{P}\right)^{t+1} \sum_{k=1}^{r} \binom{t+r-k}{r-k} \bar{P}^{r-k}. \quad (5)$$

In order to study the asymptotic behavior of (5), 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 $\mathrm{MSE}_r\left(\left\lfloor \frac{r}{v} \right\rfloor\right) \leq \exp\{-rE_1(v) + o(r)\}$, where the $o(r)$ term is independent of v, $E_1(\cdot)$ is defined as*

$$E_1(v) \triangleq \begin{cases} \frac{1}{\bar{v}}\mathbb{D}\left(\bar{v}\middle\|\bar{P}\right), & v < P; \\ 0, & v \geq P. \end{cases} \quad (6)$$

Theorem 1 implies that, for any $v < P$, the MSE goes to zero with 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".

IV. Transmission of a Single Data Packet

In this section, the error probability of transmitting a single packet over the network is derived. All the definitions of Section II carry over to the single-packet transmission setting, except for the maximal bit error probability (2), which is defined as $P_e(r,t) \triangleq \max_{0\le n\le \ell-1} \epsilon_r(n|t)$. The IV is defined w.r.t. 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.

We will 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.

We map the bits $B(0:\ell-1)$, which comprise the data packet, to a PAM constellation point using natural labeling:

$$S^\ell = \sqrt{3}\sum_{i=0}^{\ell-1}(-1)^{B(i)}2^{-(i+1)}, \tag{7}$$

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

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

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

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

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 P.

Scheme 2. 1) The transmitter (node 0) maps $B(0:\ell-1)$ to S^ℓ according to (7).
2) All nodes apply Scheme 1 with S^ℓ instead of S.
3) At each time step $t \in \mathbb{Z}_{\ge 0}$, 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.

Denote by $\mathcal{E}^\ell_r(t) \triangleq S^\ell - \hat{S}^\ell_r(t)$ the error of $\hat{S}^\ell_r(t)$ in estimating S^ℓ. Since $\mathrm{Var}\left(S^\ell\right) \le 1$, $\mathbb{E}\left[\left\{\mathcal{E}^\ell_r(t)\right\}^2\right] \le \mathrm{MSE}_r(t)$. Combining this with the Chernoff bound yields the following.

Lemma 3. *The packet error probability of Scheme 2 is bounded, with* $\mathrm{MSE}_r(t)$ *of Lemma 2, as*

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

Substituting Theorem 1 in Lemma 3, yields

Theorem 2. *Let $\ell \in \mathbb{N}$ be the packet size. Then, the single-packet IV is bounded as $V \ge P = 1 - \exp\{-2C\}$.*
Moreover, for velocity $v < P$ and $E_1(v)$ of (6),

$$\epsilon_r(0:\ell|\lfloor v/r\rfloor) = \exp\{-E_1(v)r + o(r)\}.$$

For a single channel ($r = 1$) and *Gaussian noise*, Schalkwijk and Kailath [15] (see also [14], [16], [27, pp. 481–482], [28, Ch. 17.1.1]) showed that the error probability of transmitting a single message decays *double exponentially* with t:

$$\epsilon_1(0:\ell|t) \le \exp\big\{-\exp\{2Ct + o(t)\}\big\}. \tag{8}$$

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 5 for AWGN channels.

Lemma 4. *Scheme 2 over AWGN channels achieves*

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

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

The proof relies on replacing the Chernoff inequality w.r.t. the estimation error $\mathcal{E}^\ell_r(t)$ in Lemma 3 with a Chernoff–Hoeffding bound w.r.t. a (sub-)Gaussian $\mathcal{E}^\ell_r(t)$.

Substituting Theorem 1 in Lemma 4, yields

Theorem 3. *Assume AWGN channels and let $\ell \in \mathbb{N}$ be the packet size. Then, for velocity $v < P$, the achievable prefix-free error probability is bounded from above as Scheme 2 achieves*

$$\epsilon_r(0:\ell|\lfloor r/v\rfloor) = \exp\big\{-\exp\{E_1(v)r + o(r)\}\big\}, \tag{9}$$

where $E_1(v)$ was defined in (6). In particular, $\mathrm{V} \ge P$.

Remark 1. Theorem 3 extends to sub-Gaussian noises, since proving Lemma 4 relies on the sub-Gaussianity of the noise.

Eq. (9) may be rewritten in terms of time t as

$$\epsilon_r(0:\ell|\lfloor r/v\rfloor) = \exp\big\{-\exp\{E_1(v)\cdot vt + o(t)\}\big\}. \tag{10}$$

Eq. (10) means that the *doubly-exponential* behavior (8) of the Schalkwijk–Kailath scheme [14], [15] extends also to communication at a fixed velocity across multiple relays as long as the velocity is below P.

The bound of Lemma 3 deteriorates 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 V. The "cost" of universality is a slower decay as a function of the MSE, yet this result would also suffice for achieving $\mathrm{V} = P$.

Lemma 5. *The prefix error probability of Scheme 2, for any $n \le \ell$, with* $\mathrm{MSE}_r(t)$ *of Lemma 2, is bounded as*

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

V. 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

$$\mathrm{MSE}_0(t-1) = \exp\{-2Rt\} \qquad \forall t \in \mathbb{N} \qquad (11)$$

for some $R > 0$. We can think of this as a digital message that arrives at R nats per sample (neglecting rounding issues), or as a source that is gradually revealed to the transmitter via a fixed-rate successive refinement scheme. [5]

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 (11).

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

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

The power-normalization constant $\beta_r(t)$ and the LMMSE constant $\gamma_r(t)$ are set according to (4).

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 $t = \lfloor r/v \rfloor$, 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 $\mathrm{MSE}_r\left(\left\lfloor \frac{r}{v} \right\rfloor\right) \leq \exp\{-rE_S(v) + o(r)\}$, where the correction term is independent of v,*

$$E_S(v) = \begin{cases} \frac{1}{1-\eta}\mathbb{D}\left(1-\eta \| \bar{P}\right) + 2R\left(\frac{1}{v} - \frac{\eta}{1-\eta}\right), & 0 \leq v \leq \frac{1-\eta}{\eta}; \\ \frac{1}{\bar{v}}\mathbb{D}\left(\bar{v} \| \bar{P}\right), & \frac{1-\eta}{\eta} < v \leq P; \\ 0, & P < v; \end{cases} \qquad (12)$$

and $\eta \triangleq (1-\bar{P}) \cdot \exp\{2R\} = \exp\{-2(C-R)\}$.

As the rate grows, the exponent $E_S(v)$ becomes the exponent with full source knowledge at the transmitter at $t = 0$, $E_1(v)$ (6). 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 (12) is empty in that limit.

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 packet τ is made available at the transmitter. Thus, at time $t = \tau T$ it has access to bits $B(0 : \tau\ell - 1)$. By (7), the transmitter can then map these bits to the corresponding MAP constellation point

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

As in Section IV, we define S to be the infinite-constellation limit: $S = \lim_{\ell\to\infty} S^{\ell} = \lim_{\tau\to\infty} S^{\tau\ell}$; 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 use Scheme 3 for the successively refined source S, with $S^{k\ell}$ taking the role of $\hat{S}_0(\ell)$.

Scheme 4.
1) At instants $t = \tau T$ for $\tau \in \mathbb{Z}_{\geq 0}$, the transmitter maps $B(0 : \tau\ell - 1)$ to $S^{\tau\ell}$ and updates
$$\hat{S}_0(\tau T) = \hat{S}_0(\tau T + 1) = \cdots = \hat{S}_0((\tau+1)T - 1).$$
2) The transmitter and relays apply Scheme 3 w.r.t. the (virtual) source S.
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 $B\left(0 : \left(\left\lfloor \frac{t}{T} \right\rfloor + 1\right)\ell - 1\right)$ from $\hat{S}_r(t)$.

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 $v < P$,*

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

where the worst-bit error probability $P_e(r, \Delta)$ was defined in (2)*, and $E_S(\cdot)$ is given by* (12).

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* (1) *where $C \triangleq \frac{1}{2}\log(1 + P)$. Then, the streaming IV is bounded from below as $\mathrm{V}(R) \geq \exp\{2(C - R)\} - 1$.*

The proof is based upon substituting $E_S(v)$ (12) in the result of Lemma 6. Noticing that the argument of $E_S(\cdot)$ in (12) is at most v, for v below the claimed bound we always take the first region in (12) which gives the desired positive error exponent. (it becomes independent of t_0 after the substitution, making the minimization redundant).

As expected, the IV bound tends to the single-packet IV bound P as the rate goes to zero, and to zero as the rate goes to C. It is 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: $V(R) \cong \left(1 - \frac{R}{C}\right)P$. In the high-SNR limit, on the other hand, for any fixed R, our bound grows linearly with P: $V(R) \cong P\exp\{-2R\}$.

[5] This is the MSE that is attained by a sequence of rate-R greedy optimal quantizers (up to rounding issues) that are applied to a uniform source sample. For continuous sources, this decay is exponentially optimal in t.

References

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

[2] V. Jog and P.-L. Loh, “Teaching and learning in uncertainty,” *IEEE Trans. Inf. Theory*, vol. 67, no. 1, pp. 598–615, 2020.

[3] 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, August 2021.

[4] ——, “Multi-bit relaying over a tandem of channels,” *IEEE Trans. Inf. Theory*, vol. 69, no. 6, pp. 3511–3524, June 2023.

[5] S. Rajagopalan and L. Schulman, “A coding theorem for distributed computation,” in *Proc. ACM Symp. Theory of Computing (STOC)*, 1994, pp. 790–799.

[6] S. K. Iyer and R. Vaze, “Achieving non-zero information velocity in wireless networks,” in *Proc. Symp. Model. and Opt. in Mobile, Ad hoc, and Wireless Net. (WiOpt)*, 2015, pp. 584–590.

[7] M. Fidler, “Survey of deterministic and stochastic service curve models in the network calculus,” *IEEE Communications Surveys Tutorials*, vol. 12, no. 1, pp. 59–86, 2010.

[8] M. Fidler and A. Rizk, “A guide to the stochastic network calculus,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 92–105, 2015.

[9] C.-S. Chang, *Performance Guarantees in Communication Networks*. Springer Science & Business Media, 2000.

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

[11] E. Domanovitz, T. Philosof, and A. Khina, “The information velocity of packet-erasure links,” in *IEEE Int. Conf. Comp. Comm. (INFOCOM)*, May 2022.

[12] R. Inovan, “On speed and advantage: Results on information velocity and adversarial hypothesis testing,” Ph.D. dissertation, School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, 2024.

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

[14] 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. 12, pp. 172–182, Apr. 1966.

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

[16] 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.

[17] P. Elias, “Channel capacity without coding,” in *Proceedings of the IRE*, vol. 45, no. 3, Jan. 1957, pp. 381–381.

[18] B. Schein and R. G. Gallager, “The Gaussian parallel relay channel,” in *Proc. Int. Symp. Info. Theory (ISIT)*, Sorrento, Italy, June 2000, p. 22.

[19] 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.

[20] 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.

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

[22] ——, “Side information aware coding strategies for sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 966–976, 2004.

[23] J. Chen, X. Zhang, T. Berger, and S. Wicker, “Rate allocation in distributed sensor network,” in *Proc. Allerton Conf. on Comm., Control, and Comput.*, vol. 41, no. 1, Monticello, IL, October 2003, pp. 531–540.

[24] 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.

[25] E. Domanovitz, A. Khina, T. Philosof, and Y. Kochman, “Information velocity of cascaded Gaussian channels with feedback,” *arXiv preprint arXiv:2311.14223*, 2023.

[26] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. Tata McGraw-Hill Education, 2002.

[27] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[28] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.