An Improved Bound on Information Loss due to Finite Block Length in a Gaussian Line Network

Ramanan Subramanian, Badri N. Vellambi, and Ingmar Land

Institute for Telecommunications Research, University of South Australia, Australia.

{ramanan.subramanian, badri.vellambi, ingmar.land}@unisa.edu.au

Abstract—**A bound on the maximum information transmission rate through a cascade of Gaussian links is presented. The network model consists of a source node attempting to send a message drawn from a finite alphabet to a sink, through a cascade of Additive White Gaussian Noise links each having an input power constraint. Intermediate nodes are allowed to perform arbitrary encoding/decoding operations, but the block length and the encoding rate are fixed. The bound presented in this paper is fundamental and depends only on the design parameters namely, the network size, block length, transmission rate, and signal-to-noise ratio.**

I. Introduction

Transmission of messages through a series of links corrupted by noise is a situation that occurs frequently in communication networks. When the transmission block length is allowed to be arbitrarily large, it is quite simple to show (using the data-processing inequality) that the maximum information transfer rate is equal to the capacity of the weakest link. The possibilities in the finite block length regime are far less clear. Past work by Niesen *et al.* in [1] and by us in [2] have addressed this question for the Discrete Memoryless Channel (DMC) case and the Additive White Gaussian Noise (AWGN) case, respectively. These results are *asymptotic* and provide scaling laws for the block length in terms of the number of nodes.

In this paper, we provide a universal *non-asymptotic* bound on the maximum rate of information transfer for a line network consisting of a cascade of AWGN links. This complements and improves the asymptotic scaling results derived in [2]. The bound derived here is universal in the following sense:

1) While we assume that the block length and encoding rates are constant for all the nodes, we do not assume any particular structure for the channel codes and decision rules employed at any of the nodes.

2) In addition, no assumption is made on the network size and the block length.

It is to be noted that the analysis in [2] was found to be unsuitable to our requirement that the bound be non-asymptotic, and hence we take a totally new approach here.

The rest of the paper is organized as follows. In Section II, we introduce the notations and definitions used in the rest of the paper. In Section III, we introduce the network and the signal transmission models. We then provide our main result followed by its derivation in Section IV, followed by a short discussion in Section V

that includes a comparison of our current results in relation to our previous results in [2].

II. Notations and Definitions

Let $\mathbb{R}$ be the set of all real numbers and $\mathbb{N}$ be the set of all natural numbers. Natural logarithms are assumed unless the base is specified. The notation $\|\cdot\|$ represents $\mathcal{L}^2$ norm throughout. $\mathscr{S}_{M \times M}$ denotes the set of all $M \times M$ row-stochastic matrices, and $\mathscr{S}^*_{M \times M}$ denotes the set of all $M \times M$ row-stochastic matrices whose rows are identical.

Let $N \in \mathbb{N}$ denote the *code length* or *block length* of the transmission scheme. A *code rate* $R > 0$ is a real number such that 2^{NR} is an integer. Let $\mathscr{M} \triangleq \{1, 2, 3, \ldots, 2^{NR}\}$ be the *message alphabet*.

Definition 1: For a certain $P_0 \geq 0$, a rate R length N code $\mathscr{C}$ with power constraint P_0 is an ordered set of $M = 2^{NR}$ elements from $\mathbb{R}^N$, called *codewords*, such that the power of any codeword is at most P_0:

$$\mathscr{C} = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \ldots, \mathbf{c}_M) \text{ s.t. } \forall w \in \mathscr{M}, \frac{1}{N}\|\mathbf{c}_w\|^2 \leq P_0.$$

Definition 2: A rate R dimension N *decision rule* $\mathscr{R} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_M)$ is an ordered partition of $\mathbb{R}^N$ of size $M = 2^{NR}$.

Definition 3: The *encoding function* $\mathrm{ENC}_\mathscr{C} : \mathscr{M} \to \mathscr{C}$ for a code $\mathscr{C}$ is defined by $\mathrm{ENC}_\mathscr{C}(w) = \mathbf{c}_w$, where $\mathbf{c}_w$ is the wth codeword in $\mathscr{C}$.

Definition 4: The *decoding function* $\mathrm{DEC}_\mathscr{R} : \mathbb{R}^N \to \mathscr{M}$ for a decision rule $\mathscr{R}$ is defined by:

$$\mathrm{DEC}_\mathscr{R}(\mathbf{y}) = w, \forall \mathbf{y} \in \mathcal{R}_w,$$

where $\mathcal{R}_w$ is the wth partition in $\mathscr{R}$. Let $\Omega_0 = \frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)}$, the solid angle of a N-sphere. Here, $\Gamma(\cdot)$ is the standard gamma function given by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t$. Let $Z_1, \ldots, Z_N$ be i.i.d. zero-mean unit-variance Gaussian random variables. Let for any $\gamma \geq 0$,

$$\Phi_\gamma \triangleq \begin{cases} \cot^{-1}\left(\frac{\sqrt{N\gamma}+Z_1}{\sqrt{\sum_{l=2}^N Z_l^2}}\right), & \text{if } \sum_{l=2}^N Z_l^2 > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Note that Φ_γ is a random variable whose probability distribution function has N and γ as parameters. Now, let $g : [0, \pi] \to [0, \Omega_0]$ be defined by

$$g(v) \triangleq \frac{(N-1)\pi^{\frac{N-1}{2}}}{\Gamma\left(\frac{N+1}{2}\right)}\int_0^v (\sin\theta)^{N-2}\,\mathrm{d}\theta$$
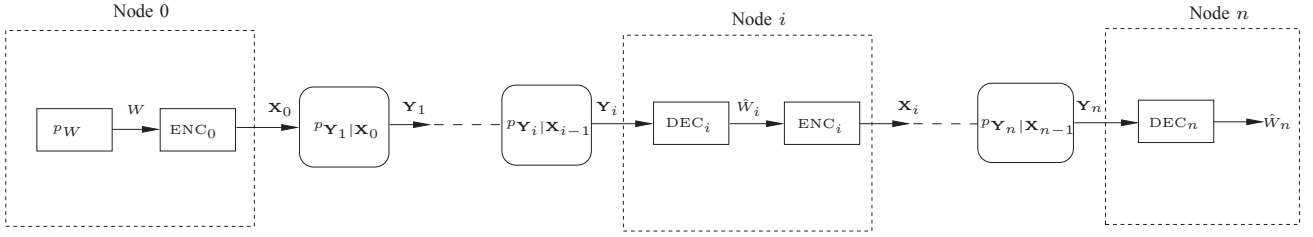
Fig. 1: Line Network Model.

The inverse cotangent function is assumed to have $[0, \pi]$ for its range so that it is continuous. Define the solid angle of a cone in N dimensions to be the area of the region cut out by the cone from a unit N-sphere centered at its apex. Then, the quantity $g(v)$ for any $v \in [0, \pi]$ is equal to the N-dimensional solid angle of a cone whose half-angle at the apex (called *cone angle* from here on) is v. We then define the following function for $x \in [0, \Omega_0]$:

$$\mathcal{Q}(x, N, \gamma) \triangleq \Pr\left[g\left(\Phi_\gamma\right) \geq x\right]. \qquad (1)$$

The above function is the same as $Q^*(\cdot)$ used by Shannon in [3]. To explain further, let $\mathbf{x}$ be a signal point satisfying $\|\mathbf{x}\|^2 = N\gamma$. Let $\mathcal{K}(\mathbf{x}, \xi)$ be the infinite right-circular cone of solid angle ξ with the origin as the apex, and the with the axis running through $\mathbf{x}$. Then, $\mathcal{Q}(\xi, N, \gamma)$ gives the probability that $\mathbf{x}$ is displaced outside $\mathcal{K}(\mathbf{x}, \xi)$ by a noise vector whose components are i.i.d. zero-mean unit-variance Gaussian random variables. Noting that $\cot^{-1} x$ will then be a decreasing function of x, we have the following remark about the monotonicity of the $\mathcal{Q}$-function w.r.t. γ:

Remark 1: For any $\gamma_1, \gamma_2 > 0$ s.t. $\gamma_1 \geq \gamma_2$,

$$\mathcal{Q}(x, N, \gamma_2) \geq \mathcal{Q}(x, N, \gamma_1).$$

III. NETWORK MODEL

The line network model to be considered is given in Fig. 1. There are $n+1$ nodes in the network identified by the indices $\{0, 1, 2, \ldots, n\}$. The n hops in the network are each associated with noise variances $\sigma_i^2 \geq \sigma_0^2 > 0, 1 \leq i \leq n$. In other words, the noise variances can be different for each link, but they are equal to or greater than a certain minimum σ_0^2 that is strictly positive. Nodes $0, 1, \ldots, n-1$ choose codes $\mathscr{C}_0, \mathscr{C}_1, \ldots, \mathscr{C}_{n-1}$ respectively to transmit, and Nodes $1, 2, \ldots, n$ choose decision rules $\mathscr{R}_1, \mathscr{R}_2, \ldots, \mathscr{R}_n$ for reception. From here on, for the sake of simplicity, we let ENC_i and DEC_i to denote $\text{ENC}_{\mathscr{C}_i}$ and $\text{DEC}_{\mathscr{R}_i}$, respectively. All the codes and decision rules have the same rate R and block length N. Node 0 generates a random message $W \in \mathscr{M}$ with probability distribution $p_W(w)$ and intends to convey the same to Node n through the noisy multihop path in the network. Each node estimates the message sent by the node in the previous hop from its noisy observation, encodes the message as a codeword, and transmits the resulting codeword to the next hop. The codeword transmitted by Node i, for any $0 \leq i \leq n-1$ is given by $\mathbf{X}_i = \text{ENC}_i(\hat{W}_i)$, where $\hat{W}_i$ is the estimate of the message at Node i after decoding (Note that $\hat{W}_0 = W$ in this notation). The observation received by Node i, for any $1 \leq i \leq n$ is given by $\mathbf{Y}_i$, which follows a conditional density function that depends on the codeword $\mathbf{X}_{i-1}$ sent by the previous node:

$$p_{\mathbf{Y}_i|\mathbf{X}_{i-1}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma_i^2)^{\frac{N}{2}}} e^{-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{2\sigma_i^2}}. \qquad (2)$$

The above density function follows from the assumptions of AWGN noise and memorylessness of the channel. The message $\hat{W}_i$ decoded by Node i is given by $\hat{W}_i = \text{DEC}_i(\mathbf{Y}_i)$. Note that the random variable $\hat{W}_n$ represents the message decoded by the final sink.

IV. THE MAIN RESULT AND ANALYSIS

The following theorem summarizes our main result.

Theorem 1: In a line network employing any choice of rate R length N codes $\mathscr{C}_0, \mathscr{C}_1, \ldots, \mathscr{C}_{n-1}$ and rate R dimension N decision rules $\mathscr{R}_1, \mathscr{R}_2, \ldots, \mathscr{R}_n$,

$$\mathcal{I}(W; \hat{W}_n) \leq NR\left[1 - M\mathcal{Q}\left(\frac{M-1}{M}\Omega_0, N, \frac{P_0}{\sigma_0^2}\right)\right]^n.$$

We now delve into the proof of Theorem 1. Let $p_{\hat{W}_i|\hat{W}_{i-1}}(k \mid j), \forall j, k \in \mathscr{M}$ denote the conditional probabilities induced by channel encoding, noisy reception, and decoding at the ith hop. For each hop i, let $\mathbf{P}_i$ be the $M \times M$ row-stochastic matrix whose entry in row j and column k is $p_{\hat{W}_i|\hat{W}_{i-1}}(k \mid j)$. Note that the jth row in $\mathbf{P}_i$ gives the conditional probability mass function on the estimate $\hat{W}_i$ of the original message W at hop i, given that the message sent by Node $i-1$ is j. Let

$$\mathbf{P} \triangleq \prod_{i=1}^{n} \mathbf{P}_i.$$

Then, $\mathbf{P}$ clearly represents the row-stochastic probability transition matrix between the original message W and the message decoded at the sink $\hat{W}_n$. The transition matrix $\mathbf{P}$ along with p_W (the probability mass function of the original message W) together induce a joint distribution between W and $\hat{W}_n$. Our goal is to find an upper bound on $\mathcal{I}(W; \hat{W}_n)$, with the constraints given in Section III.

For any $M \times M$ row-stochastic matrix $\mathbf{Q}$, define $\psi(\mathbf{Q} \mid p_W) \triangleq \mathcal{I}(W; \tilde{W})$, where $\tilde{W}$ is a random variable conditionally dependent on W according to the probability transition matrix $\mathbf{Q}$ and W is drawn according to the distribution p_W (which is the distribution of the message at Node 0). For simplicity, we just write $\psi(\mathbf{Q})$ instead of $\psi(\mathbf{Q} \mid p_W)$ for the rest of the paper, assuming throughout that the specific distribution p_W is used. Ultimately, our final bound is independent of p_W. We now have

$$\mathcal{I}(W; \hat{W}_n) = \psi\left(\prod_{i=1}^{n} \mathbf{P}_i\right). \qquad (3)$$

Before proceeding further to bound $\mathcal{I}(W;\hat{W}_n)$, we introduce the following useful lemma:

Lemma 1: For any $\mathbf{Q}_1 \in \mathscr{S}^*_{M \times M}$ and any $\mathbf{Q}_2 \in \mathscr{S}_{M \times M}, \psi\left(\mathbf{Q}_1 \mathbf{Q}_2\right) = 0$.

Proof: The result follows from noting that for any $\mathbf{Q} \in \mathscr{S}^*_{M \times M}$, $\psi\left(\mathbf{Q}\right) = 0$ and $\mathbf{Q}_1 \mathbf{Q}_2 \in \mathscr{S}^*_{M \times M}$ for $\mathbf{Q}_1 \in \mathscr{S}^*_{M \times M}$ and $\mathbf{Q}_2 \in \mathscr{S}_{M \times M}$. ∎

Now for each i, consider $\beta_i \in [0, 1]$ such that

$$\mathbf{P}_i = \beta_i \mathbf{P}_i^{(1)} + \bar{\beta}_i \mathbf{P}_i^{(2)}, \tag{4}$$

where $\bar{\beta}_i = 1 - \beta_i$, $\mathbf{P}_i^{(1)} \in \mathscr{S}_{M \times M}$ and $\mathbf{P}_i^{(2)} \in \mathscr{S}^*_{M \times M}$. In other words for each i, $\mathbf{P}_i$ be expressed as a convex combination of two row-stochastic matrices, one of them being a steady-state matrix. From (3),

$$
\begin{aligned}
\mathcal{I}\left(W;\hat{W}_n\right) = \psi\left(\prod_{i=1}^{n} \mathbf{P}_i\right) &= \psi\left(\left(\beta_1 \mathbf{P}_1^{(1)} + \bar{\beta}_1 \mathbf{P}_1^{(2)}\right) \prod_{i=2}^{n} \mathbf{P}_i\right) \\
&\overset{(a)}{\le} \beta_1 \psi\left(\mathbf{P}_1^{(1)} \prod_{i=2}^{n} \mathbf{P}_i\right) + \bar{\beta}_1 \psi\left(\mathbf{P}_1^{(2)} \prod_{i=2}^{n} \mathbf{P}_i\right) \\
&\overset{(b)}{\le} \beta_1 \psi\left(\prod_{i=2}^{n} \mathbf{P}_i\right),
\end{aligned}
$$

where (a) follows from the convexity property of mutual information w.r.t. the probability transition function, and (b) follows from applying the data processing inequality to the first term and Lemma 1 to the second term. By induction, we have:

$$\mathcal{I}\left(W;\hat{W}_n\right) \le \left(\prod_{i=1}^{n} \beta_i\right) \psi\left(I_M\right) \le NR \prod_{i=1}^{n} \beta_i, \tag{5}$$

where I_M is the $M \times M$ identity matrix. The above procedure is based on a key idea developed in the proof of Theorem V.1 in [1] in a different context. We have applied the same to facilitate a useful intermediate result given by (5). The remaining portion of the analysis that enables us to obtain the final bound involves novel steps.

We now need to determine how each $\mathbf{P}_i$ is to be split in the form given by (4) in an optimal manner, to obtain the best possible bound using this approach. Specifically, we need β_i to be as small as possible for each i. Consider the following choice:

$$
\begin{aligned}
\beta_i &= 1 - \sum_{k=1}^{M} \min_j p_{\hat{W}_i \mid \hat{W}_{i-1}}\left(k \mid j\right) \\
P_{i;j,k}^{(2)} &= \frac{1}{1 - \beta_i} \min_{j'} p_{\hat{W}_i \mid \hat{W}_{i-1}}\left(k \mid j'\right), \tag{6}
\end{aligned}
$$

where $P_{i;j,k}^{(2)}$ denotes the element on j^{th} row and k^{th} column of the matrix $\mathbf{P}_i^{(2)}$. Note that this matrix consists of identical rows, where each entry in any row is equal to the smallest element in the corresponding column of $\mathbf{P}_i$ scaled by a normalizing factor. The other matrix $\mathbf{P}_i^{(1)}$ is determined by substituting these expressions for β_i and $\mathbf{P}_i^{(2)}$ into (4). Note that the two matrices $\mathbf{P}_i^{(1)}$ and $\mathbf{P}_{\bar{\beta}_i}$ determined thus will be stochastic for any i, and that $\beta_i \in [0, 1]$. Hence, we can obtain $\mathbf{P}_i$ as a convex combination of two stochastic matrices in this manner for any i. The following lemma shows that the value of β_i

provided in (6) is the best possible value for the purpose of the bound in (5).

Lemma 2: Let $\mathbf{Q} = [Q_{jk}] \in \mathscr{S}_{M \times M}$, $\mathbf{Q}_1 \in \mathscr{S}_{M \times M}, \mathbf{Q}_2 \in \mathscr{S}^*_{M \times M}$, and let $\beta \in [0, 1]$ be chosen such that $\mathbf{Q} = \beta \mathbf{Q}_1 + (1 - \beta) \mathbf{Q}_2$. Then, $\beta \ge 1 - \sum_{k=1}^{M} \min_j Q_{jk}$.

Proof: Since $\mathbf{Q} = \beta \mathbf{Q}_1 + (1 - \beta) \mathbf{Q}_2$, every element of the matrix $(1 - \beta) \mathbf{Q}_2$ must be smaller than the corresponding element in $\mathbf{Q}$. Consider any column k of $\mathbf{Q}_2$. All the elements in that column are equal to, say, q_k. It then follows that $(1 - \beta) q_k \le Q_{jk}$ for every j, and hence $(1 - \beta) q_k \le \min_j Q_{jk}$. Summing over all k and noting that $\sum_{k=1}^{M} q_k = 1$, we obtain the desired result. ∎

Let $\mathscr{C}_{i-1}$ be the code used by Node $i - 1$ and let $\mathscr{R}_i$ be the decision rule used by Node i. As per the argument above, the optimal choice of β_i for this link will be:

$$
\begin{aligned}
1 - \beta_i &= \sum_{k=1}^{M} \min_j p_{\hat{W}_i \mid \hat{W}_{i-1}}\left(k \mid j\right) \\
&= \sum_{\mathcal{R} \in \mathscr{R}_i} \min_{\mathbf{c} \in \mathscr{C}_{i-1}} \int_{\mathcal{R}} \frac{e^{-\|\mathbf{y} - \mathbf{c}\|^2 / 2\sigma_i^2}}{(2\pi\sigma_i^2)^{N/2}} d\mathbf{y}. \tag{7}
\end{aligned}
$$

In other words, we can write

$$\beta_i = 1 - \mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}_i\right), \tag{8}$$

where for any $\sigma > 0$, rate R length N code $\mathscr{C}$, and rate R dimension N decision rule $\mathscr{R}$,

$$\mu_\sigma\left(\mathscr{C}, \mathscr{R}\right) \triangleq \sum_{\mathcal{R} \in \mathscr{R}} \min_{\mathbf{c} \in \mathscr{C}} \int_{\mathcal{R}} \frac{e^{-\|\mathbf{y} - \mathbf{c}\|^2 / 2\sigma^2}}{(2\pi\sigma^2)^{N/2}} d\mathbf{y}. \tag{9}$$

We would like to find a lower bound on $\mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}_i\right)$ that depends only on the parameters N, R, P_0 and σ_0. To do so, we need the following lemmas. Lemma 3 removes the dependency of the bound on the choice of the decision rule. Lemma 4 shows that the bound is unaffected if we restrict our choice of codes to a particular class. For a code from such a class, Lemma 5 gives a bound that depends solely on N, R, P_0 and σ_0, as desired. For brevity, we use $\eta = (2\pi\sigma_0^2)^{\frac{N}{2}}$ from here on.

Lemma 3: Let $\mathscr{C} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M)$ be a given rate R length N code. Further, let $\mathscr{R}^*\left(\mathscr{C}\right)$ be the decision rule given by $(\mathcal{R}_1^*, \mathcal{R}_2^*, \ldots, \mathcal{R}_M^*)$ where for $1 \le i \le M$,

$$\mathcal{R}_i^* = \left\{\mathbf{y} \in \mathbb{R}^N \mid i = \arg\max_{i'} \|\mathbf{y} - \mathbf{c}_{i'}\|\right\}.$$

Then, for any rate R dimension N decision rule $\mathscr{R}$,

$$\mu_\sigma\left(\mathscr{C}, \mathscr{R}\right) \ge M \mu_\sigma\left(\mathscr{C}, \mathscr{R}^*\left(\mathscr{C}\right)\right).$$

Proof: For any code $\mathscr{C}$ and decision rule $\mathscr{R}$, we have:

$$
\begin{aligned}
\eta \mu_\sigma\left(\mathscr{C}, \mathscr{R}\right) &= \sum_{\mathcal{R} \in \mathscr{R}} \min_j \int_{\mathcal{R}} e^{-\frac{\|\mathbf{y} - \mathbf{c}_j\|^2}{2\sigma^2}} d\mathbf{y} \\
&\ge \sum_{\mathcal{R} \in \mathscr{R}} \int_{\mathcal{R}} \min_j e^{-\frac{\|\mathbf{y} - \mathbf{c}_j\|^2}{2\sigma^2}} d\mathbf{y} \\
&= \int_{\mathbb{R}^N} \min_j e^{-\frac{\|\mathbf{y} - \mathbf{c}_j\|^2}{2\sigma^2}} d\mathbf{y} \\
&= \sum_{k=1}^{M} \int_{\mathcal{R}_k^*} \min_j e^{-\frac{\|\mathbf{y} - \mathbf{c}_j\|^2}{2\sigma^2}} d\mathbf{y}
\end{aligned}
$$

$$\overset{(a)}{=} \sum_{k=1}^{M} \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_k\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$\overset{(b)}{=} \sum_{k=1}^{M} \min_j \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_j\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$= \eta\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right). \quad (10)$$

Here, (a) and (b) follow from the definition of $\mathscr{R}^*(\mathscr{C})$: for any $\mathbf{y} \in \mathcal{R}_k^*$, $\arg\min_j e^{-\frac{\|\mathbf{y}-\mathbf{c}_j\|^2}{2\sigma^2}} = k$. ∎

It is useful to note that the decision rule $\mathscr{R}^*(\mathscr{C})$ given by the above lemma is the same as the $(M-1)^{\text{th}}$-order Voronoi partitioning (called "farthest-point Voronoi partitioning", see Section 3.3 in [4]) of $\mathbb{R}^N$ w.r.t. $\mathscr{C}$.

Lemma 4: Let $\mathscr{C}$ be a code satisfying $\|\mathbf{c}\|^2 \leq NP_0, \forall \mathbf{c} \in \mathscr{C}$. Then, there exists a code $\mathscr{C}'$ such that $\forall \mathbf{c}' \in \mathscr{C}', \|\mathbf{c}'\|^2 = NP_0$ and $\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right) \geq \mu_\sigma\left(\mathscr{C}', \mathscr{R}^*(\mathscr{C}')\right)$.

Proof: Let $\mathscr{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)$ and let $\mathscr{R}^*(\mathscr{C}) = (\mathcal{R}_1^*, \ldots, \mathcal{R}_M^*)$. Consider a codeword that lies strictly inside the ball $\|\mathbf{x}\|^2 < NP_0$. If no such codeword exists, the statement of the lemma is trivially true with $\mathscr{C}' = \mathscr{C}$. For the non-trivial case, we can assume that such a codeword exists. Let $\mathbf{c}_{k_0}$ be that codeword. Consider the decision region $\mathcal{R}_{k_0}^* = \left\{\mathbf{y} \in \mathbb{R}^N \mid k_0 = \arg\max_{i'} \|\mathbf{y}-\mathbf{c}_{i'}\|\right\} \in \mathscr{R}^*(\mathscr{C})$. The region $\mathcal{R}_{k_0}^*$ (if non-empty) is convex since $\mathscr{R}^*(\mathscr{C})$ is a Voronoi tesselation of order $M-1$ and since Voronoi cells of any order are convex regions (see [4, Sec. 3.2, Prop. OK.1]). Hence, there exists a unique point $\mathbf{z}_{k_0}$ in $\mathcal{R}_{k_0}^*$ nearest to $\mathbf{c}_{k_0}$. By moving the codeword at $\mathbf{c}_{k_0}$ along the line joining $\mathbf{c}_{k_0}$ and $\mathbf{z}_{k_0}$ away from the latter, the distance from the codeword to every point in $\mathcal{R}_{k_0}^*$ is increased. We continue thus until the codeword is moved to the surface of the power constraint sphere, at say $\mathbf{c}'_{k_0}$. Let us call the resulting code $\mathscr{C}_1$. Note that $\mathscr{C} \setminus \{\mathbf{c}_{k_0}\} = \mathscr{C} \setminus \{\mathbf{c}'_{k_0}\}$. Now consider the following argument for $\Delta_{\mathscr{C}} \triangleq \eta\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right)$,

$$\Delta_{\mathscr{C}} = \sum_{k=1}^{M} \min_j \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_j\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$\overset{(c)}{=} \sum_{k=1}^{M} \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_k\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$= \sum_{\substack{k=1 \\ k \neq k_0}}^{M} \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_k\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y} + \int_{\mathcal{R}_{k_0}^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_{k_0}\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$\overset{(d)}{\geq} \sum_{\substack{k=1 \\ k \neq k_0}}^{M} \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}_k\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y} + \int_{\mathcal{R}_{k_0}^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}'_{k_0}\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$\geq \sum_{\substack{k=1 \\ k \neq k_0}}^{M} \min_{\mathbf{c} \in \mathscr{C}_1} \int_{\mathcal{R}_k^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y} + \min_{\mathbf{c} \in \mathscr{C}_1} \int_{\mathcal{R}_{k_0}^*} e^{-\frac{\|\mathbf{y}-\mathbf{c}\|^2}{2\sigma^2}} \mathrm{d}\mathbf{y}$$

$$= \eta\mu_\sigma\left(\mathscr{C}_1, \mathscr{R}^*(\mathscr{C})\right) \overset{(e)}{\geq} \eta\mu_\sigma\left(\mathscr{C}_1, \mathscr{R}^*(\mathscr{C}_1)\right).$$

Here, (c) follows from (b) in the proof of Lemma 3, (d) follows from the construction of $\mathbf{c}'_{k_0}$ so that for every $\mathbf{y} \in \mathcal{R}_{k_0}^*$, $\|\mathbf{y}-\mathbf{c}_{k_0}\| \leq \|\mathbf{y}-\mathbf{c}'_{k_0}\|$, and (e) follows from Lemma 3. Note also that we have only treated the case where $\mathcal{R}_{k_0}^*$ is non-empty. If on the other hand, that decision region was empty, we can move the codeword at $\mathbf{c}_{k_0}$ along any arbitrary direction. For such a case inequality (b) becomes an equality since the integrals over $\mathcal{R}_{k_0}^*$ would be zero. From a given code $\mathscr{C}$, we can thus obtain a code $\mathscr{C}_1$ having one more codeword on the surface of the power constraint sphere, also satisfying $\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right) \geq \mu_\sigma\left(\mathscr{C}_1, \mathscr{R}^*(\mathscr{C}_1)\right)$. We can repeat this process several times to eventually obtain a code $\mathscr{C}'$ with all codewords on the power constraint sphere. ∎

Lemma 5: For any rate R length N code $\mathscr{C}$ satisfying the power constraint $\forall \mathbf{c} \in \mathscr{C}, \|\mathbf{c}\|^2 = NP_0$,

$$\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right) \geq \mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma^2}\right).$$

Proof: For any code $\mathscr{C}$ that satisfies the requirements of the lemma, the decision regions in $\mathscr{R}^*(\mathscr{C})$ consists of pyramids with their apex at the origin and extending out to infinity (see Appendix A for a proof). Assume that each of these regions $\{\mathcal{R}_k^*\}_{k=1}^{M}$ cut out a surface of area Ω_k on the unit N-sphere centered at the origin. Note that for each codeword $\mathbf{c}_k \in \mathscr{C}$ corresponding to message k, the decision region $\mathcal{R}_k^*$ contains the point $-\mathbf{c}_k$. Consider any term in the summation form of $\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right)$:

$$\min_j \int_{\mathcal{R}_k^*} \frac{e^{-\|\mathbf{y}-\mathbf{c}_j\|^2/2\sigma^2}}{(2\pi\sigma^2)^{\frac{N}{2}}} \mathrm{d}\mathbf{y} = \int_{\mathcal{R}_k^*} \frac{e^{-\|\mathbf{y}-\mathbf{c}_k\|^2/2\sigma^2}}{(2\pi\sigma^2)^{\frac{N}{2}}} \mathrm{d}\mathbf{y}.$$

The right hand side of the above equation is equal to the probability of the event E_1 that the transmitted codeword in $\mathbb{R}^N$ located at $\mathbf{c}_k$ on the sphere $\|\mathbf{x}\|^2 = NP_0$ is displaced by the noise vector into a specific region $\mathcal{R}_k^*$ (which contains the point $-\mathbf{c}_k$). Now consider the probability of the event E_2 that the same transmitted codeword is displaced into the N-dimensional circular cone $\mathcal{C}_k^*$ that has its apex at the origin, axis running through $-\mathbf{c}_k$, and cutting out a surface of area Ω_k on the unit sphere centered at the origin (i.e., the solid angle of the N-dimensional circular cone is Ω_k). We claim that the probability of E_1 cannot be smaller than the probability of E_2. This is due to the fact that for a specified solid angle Ω_k, the shape of the surface cut from the power constraint sphere by the decision region that minimizes the probability of the event E_1 is a spherical cap (for further details, see [3]). The probability of the event E_2 is equal to $\mathcal{Q}\left(\Omega_0 - \Omega_k, N, P_0/\sigma^2\right)$, from the definition of the $\mathcal{Q}$-function in Definition **??**. Hence,

$$\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right) = \sum_{k=1}^{M} \min_j \int_{\mathcal{R}_k^*} \frac{e^{\|\mathbf{y}-\mathbf{c}_j\|^2/2\sigma^2}}{(2\pi\sigma^2)^{N/2}} \mathrm{d}\mathbf{y}$$

$$\geq \sum_{k=1}^{M} \mathcal{Q}\left(\Omega_0 - \Omega_k, N, \tfrac{P_0}{\sigma^2}\right). \quad (11)$$

Noting that $\mathcal{Q}$ is a convex function of $\Omega_0 - \Omega_k$ (See Section III in [3]), we apply Jensen's inequality to (11):

$$\mu_\sigma\left(\mathscr{C}, \mathscr{R}^*(\mathscr{C})\right) \geq M\frac{1}{M}\sum_{k=1}^{M} \mathcal{Q}\left(\Omega_0 - \Omega_k, N, \tfrac{P_0}{\sigma^2}\right)$$

$$\geq M\mathcal{Q}\left(\tfrac{\sum_{k=1}^{M}(\Omega_0 - \Omega_k)}{M}, N, \tfrac{P_0}{\sigma^2}\right)$$

$$= M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma^2}\right), \quad (12)$$

4

since $\sum_{k=1}^{M} \Omega_k = \Omega_0$.

We are now ready to prove Theorem 1.

Proof of Theorem 1: Consider any link i. For the code $\mathscr{C}_{i-1}$ satisfying $\|\mathbf{c}\|^2 \leq NP_0, \forall \mathbf{c} \in \mathscr{C}_{i-1}$, we can apply Lemma 4 to construct another code $\mathscr{C}'_{i-1}$ such that $\forall \mathbf{c} \in \mathscr{C}'_{i-1}, \|\mathbf{c}\|^2 = NP_0$ and $\mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}^*\left(\mathscr{C}_{i-1}\right)\right) \geq \mu_{\sigma_i}\left(\mathscr{C}'_{i-1}, \mathscr{R}^*\left(\mathscr{C}'_{i-1}\right)\right)$. We then have:

$$
\begin{aligned}
\mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}_i\right) &\overset{(a)}{\geq} \mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}^*\left(\mathscr{C}_{i-1}\right)\right) \\
&\overset{(b)}{\geq} \mu_{\sigma_i}\left(\mathscr{C}'_{i-1}, \mathscr{R}^*\left(\mathscr{C}'_{i-1}\right)\right) \\
&\overset{(c)}{\geq} M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma_i^2}\right) \\
&\overset{(d)}{\geq} M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma_0^2}\right). \quad (13)
\end{aligned}
$$

In the above chain of equations (a), (b), and (c) follow from Lemma 3, Lemma 4 (as discussed above), and Lemma 5 respectively. Inequality (d) follows from Remark 1, since $\sigma_i^2 \geq \sigma_0^2$. Recalling that $\beta_i = 1 - \mu_{\sigma_i}\left(\mathscr{C}_{i-1}, \mathscr{R}_i\right)$ and applying (13) to (5), we have the desired result:

$$
\mathcal{I}\left(W; \hat{W}_n\right) \leq NR\left[1 - M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma_0^2}\right)\right]^n.
$$

■

V. DISCUSSION

We had mentioned in Section I that the bound presented in the current paper improves and complements the bound provided by [2]. In this section, we demonstrate this fact with a comparison plot. The bound given by [2] states that for any $\epsilon > 0$ and asymptotically large n,

$$
\mathcal{I}\left(W; \hat{W}_n\right) \leq 2^{NR}\left(1 - e^{-NE(P_0/\sigma_0^2)}\right)^{n(1-\epsilon)}, \quad (14)
$$

where for any $S \geq 0$,

$$
\begin{aligned}
E\left(S\right) &\triangleq \frac{(S+2)+\sqrt{(S+2)^2-4}}{4} \\
&\quad + \frac{1}{2}\log\left\{(S+2)+\sqrt{(S+2)^2-4}\right\}.
\end{aligned}
$$

The above bound decays with n as $\left(1 - e^{-NE(P_0/\sigma_0^2)}\right)^n$ for any code rate R. In contrast, the bound given by Theorem 1 is valid for any n and decays as $\left[1 - M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, \tfrac{P_0}{\sigma_0^2}\right)\right]^n$. Now, let

$$
E_{as}\left(R, S\right) \triangleq -\lim_{N\to\infty} \frac{1}{N}\log\left[M\mathcal{Q}\left(\tfrac{M-1}{M}\Omega_0, N, S\right)\right].
$$

We investigate how the bound given by Theorem 1 compares with (14), by comparing the values of $E(S)$ and $E_{as}(R, S)$. To do so, we obtain $E_{as}(R, S)$ as a function of R and S using the asymptotic analysis in [3]. The resulting expression for $E_{as}(R, S)$ is given below:

$$
E_{as}\left(R, S\right) = S - \frac{1}{2}\sqrt{S}G\cos\theta - \log\left(G\sin\theta\right) - R\log 2,
$$

where

$$
G = \frac{1}{2}\left(\sqrt{S}\cos\theta + \sqrt{4 + S\cos^2\theta}\right),
$$

and $\theta = \pi - \sin^{-1}2^{-R}$. Figure 2 is a plot of $E_{as}(R, S)/E(S)$ as a function of S, repeated for various

R. As can be seen from the plot $E_{as}(R, S)$ is always smaller than $E(S)$ for any R and S, thus showing that the bound obtained in Theorem 1 is tighter than the one given by (14). The current bound is also seen to be better when the SNR S is not very high.
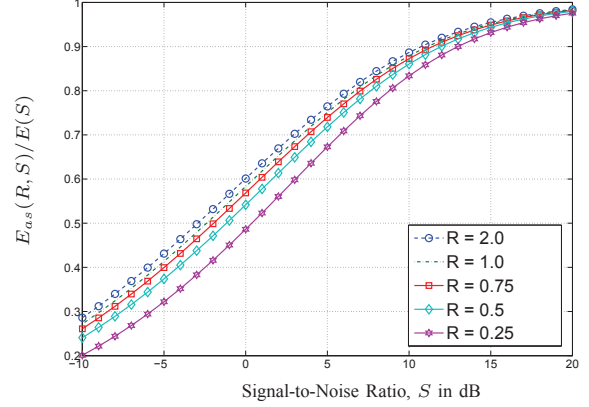

Fig. 2: Comparison of exponents for large N.

REFERENCES

[1] U. Niesen, C. Fragouli, and D. Tuninetti, "On capacity of line networks," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4039–4058, Nov. 2007.

[2] R. Subramanian, "The relation between block length and reliability for a cascade of AWGN links," *Proc. 2012 Int. Zurich Seminar on Communications (IZS)*, pp. 71–74, Feb. 29 – Mar. 2, 2012.

[3] Claude E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Claude Elwood Shannon: Collected Papers*, pp. 279–324, 1993.

[4] Atsuyuki Okabe, Barry Boots, and Kokichi Sugihara, *Spatial tessellations: concepts and applications of Voronoi diagrams*, chapter 3, John Wiley & Sons, Inc., New York, NY, USA, 1992.

APPENDIX A
THE FARTHEST-POINT VORONOI TESSELLATION FOR POINTS ON A SPHERE

Lemma 6: Given any $M, N \in \mathbb{N}$, the non-empty cells in the farthest-point Voronoi tessellation of $\mathbb{R}^N$ w.r.t. any set of M points on an N-sphere of radius $A > 0$ are all semi-infinite pyramids.

Proof: Consider the farthest-point Voronoi tessellation of $\mathbb{R}^N$ w.r.t. a code $\mathscr{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)$ s.t. $\|\mathbf{c}_i\|^2 = A^2, 1 \leq i \leq M$. Consider any $\mathbf{x} \in \mathbb{R}^N$ and assume that it is contained in $\mathcal{R}_k^*$, the farthest-point Voronoi cell for $\mathbf{c}_k$ for some k. In that case, for any i and for any $\alpha \geq 0$,

$$
\begin{aligned}
\|\mathbf{c}_k - \mathbf{x}\|^2 &\geq \|\mathbf{c}_i - \mathbf{x}\|^2 \\
\Leftrightarrow -\langle\mathbf{c}_k, \mathbf{x}\rangle &\overset{(f)}{\geq} -\langle\mathbf{c}_i, \mathbf{x}\rangle, \forall i \\
\Leftrightarrow -\langle\mathbf{c}_k, \alpha\mathbf{x}\rangle &\geq -\langle\mathbf{c}_i, \alpha\mathbf{x}\rangle, \forall i \\
\Leftrightarrow \|\mathbf{c}_k - \alpha\mathbf{x}\|^2 &\overset{(g)}{\geq} \|\mathbf{c}_i - \alpha\mathbf{x}\|^2, \forall i, \quad (15)
\end{aligned}
$$

where (f) and (g) follow from the fact that $\|\mathbf{c}_k\|^2 = \|\mathbf{c}_i\|^2 = A$. We have thus shown that for any $\mathbf{x} \in \mathbb{R}^N$ and any $\alpha > 0$, if $\mathbf{x}$ is in $\mathcal{R}_k^*$, then $\alpha\mathbf{x}$ is also in $\mathcal{R}_k^*$. Such a region is a semi-infinite pyramid by definition. ■