The relation between block length and reliability for a cascade of AWGN links

Author(s):
Subramanian, Ramanan

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-007050461

Rights / License:
In Copyright - Non-Commercial Use Permitted
The Relation Between Block Length and Reliability for a Cascade of AWGN Links

Ramanan Subramanian
Institute for Telecommunications Research, University of South Australia, Mawson Lakes, SA 5095, Australia
Email: Ramanan.Subramanian@unisa.edu.au

Abstract—This paper presents a fundamental information-theoretic scaling law for a heterogeneous cascade of links corrupted by Gaussian noise. For any set of fixed-rate encoding/decoding strategies applied at the links, it is claimed that a scaling of $\Theta(\log n)$ in block length is both sufficient and necessary for reliable communication. A proof of this claim is provided using tools from the convergence theory for inhomogeneous Markov chains.

I. INTRODUCTION

Consider a line network in which a message originating at a certain node is conveyed to the intended destination through a series of hops. The reception at every hop is impaired by zero-mean additive white Gaussian noise (AWGN) of a certain variance. The originating node as well as the intermediate nodes are subject to an average transmit-power constraint $P_0$ for the duration of transmission of any message. The messages, drawn from a certain set of size $M$, are encoded and transmitted as codewords in $\mathbb{R}^N$ that satisfy the transmit power constraint. Every intermediate hop decodes (possibly erroneously) the original message from its noisy observation and re-encodes it before transmitting to the next hop. One can form the (row-stochastic) probability transition matrix $P_i$ for the $i$th hop in which each row gives the conditional probability distribution of the message decoded at hop $i$, given a certain message sent by the previous hop:

$$P_i = \begin{pmatrix}
    p_i(1|1) & p_i(2|1) & \cdots & p_i(M|1) \\
p_i(1|2) & p_i(2|2) & \cdots & p_i(M|2) \\
    \vdots & \vdots & \ddots & \vdots \\
p_i(1|M) & p_i(2|M) & \cdots & p_i(M|M)
\end{pmatrix}.$$

Assume that a network of the type outlined above consists of $n$ hops. If the encoder-decoder pairs are identical at the originating node. For decisions made on signals corrupted by Gaussian noise, $P$ has the property that any column is either all-zero or contains only non-zero entries. For such a matrix $P$, the rows of $P^n$ tend to be identical for large $n$ (see Theorem 4.9 in [1]), if $M$, $N$, $P_0$, and $\sigma$ are independent of $n$. In other words, the probability distribution of $W_n$ will be almost independent of the original message $W$. This implies that the mutual information $I(W; W_n)$ between the original message and the decoded message tends to zero with $n$. Hence, communication in the network is highly unreliable if $N$ and $M$ do not vary with the number of hops $n$.

On the other hand, if $N$ is large and if one chooses $M = 2^{NR}$ for any $R < \frac{1}{2}\log(1 + \frac{2}{\sigma^2})$, random-coding theorem states that there exists a channel code of block length $N$ such that the probability of error during decoding at any link is bounded above by $e^{-N\epsilon(R,P/\sigma^2)}$, where $E$ is the random coding exponent for the corresponding AWGN channel [2], [3]. By employing such a code for each link in the line network model, choosing any $f(n) = o(1)$, and letting $N$ vary with $n$ as $\frac{1}{\log n} f(n)$, one obtains the following union bound on the probability of communication failure in the network:

$$P[\hat{W}_n \neq W] \leq nP_B \leq n e^{-N\epsilon(R,P/\sigma^2)} = e^{-f(n)} \rightarrow 0.$$ (1)

Hence, it is possible to transmit messages with very high reliability in the network for a fixed code rate $R$, if the code length $N$ satisfies $N(n) = o(\log n)^{1}$. On the other hand, the previous discussion about diminishing $I(W; W_n)$ implies that any fixed-rate fixed-length coding scheme is not scalable for the line network. This leads to the natural question whether the scaling $\Theta(\log n)$ for $N$ is indeed optimal. In this paper, we show that for any scaling of $N$ satisfying $N(n) = o(\log n)^{1}$, the mutual information $I(W; \hat{W}_n)$ diminishes to zero with $n$ for any set of fixed-rate codes. This implies that the scaling order $\Theta(\log n)$ for $N$ is necessary for reliable communication in the network. Hence, the sufficiency criterion for $N$ given by the union bound in Eqn. (1) is indeed optimal up to a constant factor.

While this scaling problem has been solved partially in the past for cascades of homogeneous DMCs by Niesen et al. in [4], it has also been pointed out in the same work that the tools developed there apply neither to non-homogeneous cascades nor to cascades of continuous-input channels (as is the case in the current paper). In contrast, the results developed here are applicable to continuous-input AWGN links where the links can be non-identical.

1For any $f(n) > 0$ and $g(n) > 0$, $f(n) \in o(g(n)) \implies \lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$, $f(n) \in \Theta(g(n)) \implies \exists c > 0 \text{ s.t. } \lim_{n \to \infty} \frac{f(n)}{g(n)} = c.$
II. NOTATIONS AND DEFINITIONS

The set of all real numbers is denoted by \( \mathbb{R} \) and the set of all natural numbers by \( \mathbb{N} \). Natural logarithms are assumed in the notations, unless the base is specified. The notation \( \| \cdot \| \) represents \( L^2 \) norm throughout.

We employ the following terminologies in this paper. Let \( N \in \mathbb{N} \), called 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 \( \mathcal{M} \triangleq \{ 1, 2, 3, \ldots , 2^{NR} \} \), called the message alphabet. Let \( P_0 > 0 \) be the input power constraint for transmission.

Definition 1: A rate-\( R \) length-\( N \) code \( \mathcal{C} \) with a power constraint of \( P_0 \) is an ordering of \( M = 2^{NR} \) elements from \( \mathbb{R}^N \), called code words that satisfies

\[
\mathcal{C} = \{ \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots , \mathbf{x}_M \} \quad \text{s.t.} \quad \forall \mathbf{w} \in \mathcal{M}, \quad \frac{1}{N} \| \mathbf{x}_w \|^2 \leq P_0. \tag{2}
\]

Definition 2: A rate-\( R \) length-\( N \) decision rule \( \mathcal{R} \) is an ordering of \( M = 2^{NR} \) sub-sets of \( \mathbb{R}^N \), called decision regions, spanning \( \mathbb{R}^N \) and pairwise disjoint:

\[
\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2, \ldots , \mathcal{R}_M) \quad \text{s.t.} \quad \cup_{\mathbf{w} \in \mathcal{M}} \mathcal{R}_w = \mathbb{R}^N, \quad \text{and} \quad \mathcal{R}_{w_1} \cap \mathcal{R}_{w_2} = \emptyset. \tag{3}
\]

Definition 3: The encoding function \( \mathcal{E} \triangleq \mathcal{E}_i : \mathcal{M} \rightarrow \mathbb{R}^N \) for a code \( \mathcal{C} \) is defined by:

\[
\mathcal{E}_i(\mathbf{w}) = \mathbf{x}_w, \tag{4}
\]

where \( \mathbf{x}_w \) is the \( w \)-th code word in \( \mathcal{C} \).

Definition 4: The decoding function \( \mathcal{D} \triangleq \mathcal{D}_{\mathcal{R}} : \mathbb{R}^N \rightarrow \mathcal{M} \) for a decision rule \( \mathcal{R} \) is defined by:

\[
\mathcal{D}_{\mathcal{R}}(\mathbf{y}) = \sum_{i \in \mathcal{R}} w |_{\mathcal{R}_i}(\mathbf{y}), \tag{5}
\]

where \( \mathbf{I} \) is the indicator function and \( \mathcal{R}_w \) is the \( w \)-th subset in \( \mathcal{R} \).

III. NETWORK MODEL

The line network model to be considered is described 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 \( \sigma_i^2 \), \( 1 \leq i \leq n \). For encoding, nodes 0, 1, \ldots , \( n - 1 \) choose \( \mathcal{C}_0, \mathcal{C}_1, \ldots , \mathcal{C}_{n-1} \) respectively to transmit, each code being rate-\( R \) length-\( N \) with a power constraint of \( P_0 \). For decoding, nodes 1, 2, \ldots , \( n \) choose rate-\( R \) length-\( N \) decision rules \( \mathcal{R}_1, \mathcal{R}_2, \ldots , \mathcal{R}_n \) for reception. From here on, we write \( \mathcal{E}_i \) and \( \mathcal{D}_{\mathcal{R}} \) in place of \( \mathcal{E}_{\mathcal{C}_i} \) and \( \mathcal{D}_{\mathcal{R}_i} \) respectively for simplicity. Node 0 generates a random message \( \mathbf{W} \in \mathcal{M} \) with probability distribution \( p_{W}(w) \) and intends to convey the same to node \( n \) through the cascade of noisy links in a multihop fashion. Each of the \( n - 1 \) intermediate nodes estimates the message (as sent by the node in the previous hop) from its own noisy observation, re-encodes the decoded message, 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

\[
\mathbf{X}_i = \mathcal{E}_i(\hat{\mathbf{W}}_i), \tag{6}
\]

where \( \hat{\mathbf{W}}_i \) is the estimate of the message at node \( i \) after decoding (Note that \( \hat{\mathbf{W}}_0 = \mathbf{W} \) 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:

\[
p_{Y_i|X_{i-1}}(y|x) = \frac{1}{(2\pi \sigma_i^2)^{\frac{N}{2}}} e^{-\frac{(y-x)^2}{2\sigma_i^2}}. \tag{7}
\]

The above density function follows from the assumptions of AWGN noise and memoryless-ness in the channel. The message \( \hat{\mathbf{W}}_i \) decoded by node \( i \) is given by

\[
\hat{\mathbf{W}}_i = \mathcal{D}_{\mathcal{R}_i}(\mathbf{Y}_i). \tag{8}
\]

Note that the random variable \( \hat{\mathbf{W}}_n \) represents the message decoded by the final destination, as per the above notation.

IV. AN UPPER BOUND ON THE RELIABILITY OF THE NETWORK

The key result presented here is summarized by the following theorem and its corollary:

Theorem 1: In a line network consisting of a cascade of \( n \) AWGN links, let for \( m \leq n \) of the links the noise variances satisfy \( \sigma_j^2 \geq \sigma_0^2 > 0 \). Then, for any choice of codes \( \mathcal{C}_0, \mathcal{C}_1, \ldots , \mathcal{C}_{n-1} \) and decision rules \( \mathcal{R}_1, \mathcal{R}_2, \ldots , \mathcal{R}_n \) with rate \( R \) length \( N \), and for any \( \epsilon > 0 \), the following holds for sufficiently large \( m \):

\[
I(\mathbf{W}; \hat{\mathbf{W}}_n) \leq 2^{NR} \left( 1 - \frac{N}{2} \frac{\sigma_0^2}{\Gamma(\frac{N}{2}+1)} - NE'(P_0/\sigma_0^2) \right)^{m(1-\epsilon)},
\]

where

\[
E'(S) = \exp \left\{ \cosh^{-1} \left( \frac{S}{2} + 1 \right) \right\} + \cosh^{-1} \left( \frac{S}{2} + 1 \right).
\]

The corollary to the above theorem leads to the upper bound on the reliability vs. block-length scaling as claimed in the introductory section:

Corollary 1: Let \( S > 0 \), \( \sigma_0 > 0 \), and \( 1 \geq r > 0 \) be independent of \( n \). In a line network of \( n \) links, let for all hops the noise variance be at least \( \sigma_0^2 > 0 \). Then, for any \( N(n) \in o(\log n) \), \( \lim_{n \to \infty} I(\mathbf{W}; \hat{\mathbf{W}}_n) = 0 \).

Proof of Theorem 1: At any hop \( i \), the process of channel encoding, noisy reception, followed by decoding induces the conditional probabilities \( p_{\hat{W}_i|\hat{W}_{i-1}}(k|j), \forall j, k \in \mathcal{M} \). For every hop \( i \), let \( \mathbf{P}_i \) be the \( M \times M \) row-stochastic matrix whose entry in row \( j \) and column \( k \) is \( p_{\hat{W}_i|\hat{W}_{i-1}}(k|j) \). Note that the \( j \)-th row in \( \mathbf{P}_i \) gives the conditional probability mass function on the estimate of message \( \mathbf{W} \) at hop \( i \) (i.e., \( \hat{\mathbf{W}}_i \)) given that the message sent by hop \( i - 1 \) was \( j \). Let

\[
\mathbf{P} \triangleq \prod_{i=1}^{n} \mathbf{P}_i. \tag{9}
\]

Then, \( \mathbf{P} \) clearly represents the row-stochastic probability transition matrix between the original message \( \mathbf{W} \) and the message decoded at the final destination, \( \hat{\mathbf{W}}_n \). 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 \( \tilde{W}_n \). Our goal is to find an upper bound on \( I(W;\tilde{W}_n) \), with the constraints as given in the statement of Theorem 1.

For any \( M \times M \) row-stochastic matrix \( Q = \{Q_{jk}\} \), let us consider the following measures of deviation:

\[
\delta(Q) \triangleq \max_{j_1, j_2} \max_k \{Q_{j_1k} - Q_{j_2k}\}, \tag{10}
\]

\[
\lambda(Q) \triangleq 1 - \min_{j_1, j_2} \min_{k=1}^M (Q_{j_1k}, Q_{j_2k}). \tag{11}
\]

Both the above deviations measure the degree to which the rows of \( Q \) differ. The measure \( \lambda(Q) \) is known as Dobrushin’s coefficient of ergodicity and plays a role in the convergence of non-homogeneous stochastic processes [5]. Moreover, for any stochastic matrix \( Q \),

\[
0 \leq \lambda(Q) \leq 1. \tag{12}
\]

Now consider the total variational distance between the joint distribution of \( W \) and \( \tilde{W}_n \) and the product of the respective marginal distributions:

\[
\tau(W;\tilde{W}_n) \triangleq \sum_{w_1, \tilde{w} \in \mathcal{M}} |p_W(w_1)p_{\tilde{W}_n}(\tilde{w}) - p_W(\tilde{w})|. \tag{13}
\]

In the above expression, substituting \( p_W(w_1)p_{\tilde{W}_n}(\tilde{w}|w_1) \) for \( p_{W\tilde{W}_n}(w_1, \tilde{w}) \), \( \sum_{\tilde{w}_2 \in \mathcal{M}} p_{\tilde{W}_n}(\tilde{w}_2|w_1) p_W(w_2) \) for \( p_{W\tilde{W}_n}(w_1, \tilde{w}) \) and by using the triangle inequality, one obtains the following bound:

\[
\tau(W;\tilde{W}_n) \leq \sum_{w_1, \tilde{w}} p_W(w_1) \sum_{w_2} p_W(w_2) \delta(P) = M \delta(P) = 2^{N\delta} \delta(P). \tag{14}
\]

Recalling the definition of \( P \) in our case, and from Theorem 2 in [6] (equivalently, Lemma 2 in [7]), we have:

\[
\delta(P) = \delta \left( \prod_{i=1}^n P_i \right) \leq \prod_{i=1}^n \lambda(P_i). \tag{15}
\]

The second inequality in the above follows by applying Eqn. (12) for those hops for which \( \sigma_i < \sigma_0 \). Note that as per the statement of the theorem, we have at least \( m \) terms in the product in the last expression above. Now consider \( \lambda(P_i) \) for any hop \( i \) such that \( \sigma_i \geq \sigma_0 \). Suppose \( P_i \) has \( L \) distinct columns with indices \( k_1, k_2, \ldots, k_L \) such that

\[
\forall j, P_{jk_1} = p_{W|\tilde{W}_n}(k_1|j) \geq p_l \tag{15}
\]

\[
\vdots \tag{15}
\]

\[
\forall j, P_{jk_L} = p_{W|\tilde{W}_n}(k_L|j) \geq p_L. \tag{16}
\]

Then from the definition in Eqn. (11), it follows that

\[
1 - \lambda(P_i) \geq \sum_{\ell=1}^L p_\ell. \tag{17}
\]

We now show that the conditions given by Eqn. (15)-(16) hold true in our case, so that we can apply Eqn. (17) to our bound. Let \( \alpha > 0 \). Consider one particular hop \( i \) satisfying \( \sigma_i \geq \sigma_0 \). Define the spherical regions

\[
S_1 = \left\{ y \in \mathbb{R}^N : |y| \leq \sqrt{NP_0} \right\}, \tag{18}
\]

\[
S_\alpha = \left\{ y \in \mathbb{R}^N : |y| \leq \alpha \sqrt{NP_0} \right\}. \tag{19}
\]

Clearly, any code word in \( S_{\alpha-1} \) (transmitted by the previous hop) has to be in \( S_1 \). Next, observe that \( S_\alpha \) has to contain decision regions, say \( L \) in number, spanning \( S_\alpha \) entirely. Let us designate them as \( R_{k_1}, R_{k_2}, \ldots, R_{k_L} \). Also define for each \( 1 \leq \ell \leq L \), \( V_\ell \triangleq R_{k_\ell} \cap S_\alpha \). Hence, we have:

\[
\bigcup_{\ell=1}^L V_\ell = S_\alpha, \quad \text{and} \quad V_1 \cup V_2 \cup \cdots \cup V_L = \emptyset. \tag{20}
\]

Let \( \tilde{W}_{i-1} = j \) be the message as decoded by the previous hop. Hence, the code word \( x_i \) from \( \ell_{i-1} \) was transmitted by the node \( i-1 \). The probability that this was decoded as message \( k_\ell \) at hop \( i \) is given by:

\[
p_{W|\tilde{W}_n}(k_\ell|\tilde{W}_{i-1}) \overset{(a)}{=} \int_{\mathbb{R}^{N\delta}} p_{Y_i X_i}(y|x_i) \, dy \quad \overset{(b)}{=} \int_{\mathbb{R}^{N\delta}} e^{-\frac{|y-x_i|^2}{2\sigma_i^2}} \, dy \quad \overset{(c)}{=} \int_{\mathbb{R}^{N\delta}} e^{-\frac{(y(1+\sigma_i^2)^2}{2\sigma_i^2}} \, dy \quad \overset{(d)}{=} e^{-\frac{NP_0}{2\sigma_i^2} (|V_\ell|). \tag{21}
\]
where $|V_t|$ denotes the $N$-dimensional volume of $V_t$. In the above series of inequalities, step (a) follows from Eqn. (7), (b) from the triangle inequality, and (c) from the fact that any $y \in V_t$ must satisfy $\|y\| \leq \alpha \sqrt{N}P_0$, since $V_t \subseteq S_0$, and from the fact that $x_j$, a code word in $V_j$, naturally satisfies the power constraint criterion given by $\|x_j\| \leq \sqrt{N}P_0$. Note that the lower bound given by Eqn. (19) is independent of the message $j$ decoded by the previous hop. Hence, the conditions given by Eqns. (15)-(16) are satisfied for

$$p_t = \frac{1}{(2\pi\sigma^2)^2} e^{-\frac{(1+\sigma^2\lambda_0^2)\alpha^2}{2\sigma^2}} |V_t|. \quad (20)$$

Hence, the bound given by Eqn. (17) holds true. In our case, it gives:

$$1 - \lambda(P_t) \geq \sum_{t=1}^{L} p_t = \frac{1}{(2\pi\sigma^2)^2} e^{-\frac{(1+\sigma^2\lambda_0^2)\alpha^2}{2\sigma^2}} \sum_{t=1}^{L} |V_t|$$

$$= \frac{1}{(2\pi\sigma^2)^2} e^{-\frac{1+\sigma^2\lambda_0^2}{2\sigma^2}} \left( \frac{\alpha^2 NP_0}{\sigma^2} \right)^{1.5} \quad (21)$$

The last equality follows from the fact that $V_t$ are mutually disjoint and span the entire $N$-dimensional sphere $S_0$ (Eqn. (18)). The volume of $S_0$ is computed as:

$$|S_0| = \frac{\pi^{\frac{3}{2}}}{\Gamma(\frac{3}{2}+1)} \left( \frac{\alpha \sqrt{NP_0}}{\sigma} \right)^N. \quad (22)$$

The ideas in the above argument are outlined in Figs 2(a) and 2(b).

![Diagram](image)

(a) Intersecting regions of $S_0$ with the decision rule $\mathcal{A}$.

(b) Integrating the density function of noise centered at $x_j$ over a subset $V_t$ of $S_0$.

Fig. 2: Geometrical argument used for bounding $\lambda(P_t)$.

Next, note that the bound given by Eqn. (21) holds true for any $\alpha > 0$. Hence, we can maximize the right hand side of Eqn. (21) over all $\alpha > 0$ to obtain

$$1 - \lambda(P_t) \geq \left( \frac{\alpha^2 NP_0}{\sigma^2} \right)^{1.5} e^{-\frac{1+\sigma^2\lambda_0^2}{2\sigma^2}}, \quad (23)$$

where the function $E'(S)$ is as given in the statement of the theorem. Noting further that $E(S)$ is a non-decreasing function in $S$, we will have for hops satisfying $\sigma_t \geq \sigma_0$,

$$\lambda(P_t) \leq 1 - \left( \frac{\alpha^2 NP_0}{\sigma^2} \right)^{1.5} e^{-\frac{1+\sigma^2\lambda_0^2}{2\sigma^2}}. \quad (24)$$

Since at least $m$ hops satisfy the minimum noise variance criterion $\sigma_t \geq \sigma_0$, we can now combine Eqn. (24) with Eqn. (14) to obtain

$$\delta(P) \leq \left( 1 - \left( \frac{\alpha^2 NP_0}{\sigma^2} \right)^{1.5} e^{-\frac{1+\sigma^2\lambda_0^2}{2\sigma^2}} \right)^m. \quad (25)$$

Applying the above to Eqn. (13), we will have

$$\tau(W; \hat{W}_n) \leq 2^{NR} \left( 1 - \left( \frac{\alpha^2 NP_0}{\sigma^2} \right)^{1.5} e^{-\frac{1+\sigma^2\lambda_0^2}{2\sigma^2}} \right)^m. \quad (26)$$

We can finally bound $I(Z; W_n)$ using Lemma 2.7 of Chapter 1 in [8] by:

$$I(Z; W_n) \leq \tau(W; \hat{W}_n) \log \frac{2^{NR}}{\tau(W; \hat{W}_n)}. \quad (27)$$

The bound given by the statement of the theorem then follows from combining the above with Eqn. (26) and by observing that for sufficiently large $m$, $\tau(W; \hat{W}_n) \leq e^{-1}$, that $-x \log x$ is non-decreasing for $x \leq e^{-1}$, and finally from the fact that for any $\epsilon > 0$, $x^2 \log x$ for sufficiently large $x$.

Proof of Corollary 1: For the corollary, we have $m = nr$. Moreover, $\Gamma(\frac{3}{2}+1) = (\frac{5}{3})^{\frac{3}{2}}$. Hence, the bound given by Theorem 1 reduces to:

$$I(Z; \hat{W}_n) \leq 2^{NR} \left( 1 - e^{-\frac{1}{2\sigma^2} N^2 \frac{\alpha^2 \lambda_0^2}{\sigma^2}} \right) e^{-\frac{nr(1-\epsilon)}{2}} \quad (28)$$

$$= 2^{NR} e^{-nr(1-\epsilon)} e^{-\frac{nr(\alpha^2 \lambda_0^2)}{2\sigma^2}} \quad (29)$$

$$= e^{NR \log 2 - \frac{nr(1-\epsilon)}{2}}. \quad (30)$$

The last expression diminishes to 0 with $n \to \infty$ if $N = N(n) \in o(\log n)$. Hence, we will have $I(Z; W_n) \to 0$ for any $N = o(\log n)$.

References


