Variable-length Coding Error Exponents for the AWGN Channel with Noisy Feedback at Zero-Rate

Kenneth Palacio-Baus*†, Natasha Devroye*
*University of Illinois at Chicago, {kpalac2, devroye}@uic.edu
†University of Cuenca, Ecuador

***Abstract*—A one-way additive white Gaussian noise (AWGN) channel with active feedback sent over another AWGN feedback channel is considered. Achievable error exponents are presented in the finite message / zero-rate regime for a variable length coding (VLC) scheme. This coding scheme uses a form of round-robin scheduling of messages, and a simplex-based feedback code to obtain reliable feedback and remain synchronized, despite the noise in the feedback link. Our results show that this new VLC scheme under an almost-sure power constraint achieves an error exponent similar to an achievable exponent attained using a fixed block length scheme under a much more relaxed expected block power constraint, and is larger than that achieved by schemes without feedback.** [1]

I. Introduction

Variable length coding (VLC) techniques have been shown to achieve larger error exponents than those achieved under fixed block length coding. They use feedback to inform the transmitter about tentative decoding decisions at the receiver, allowing the source to determine whether a retransmission is necessary to correct potential decoding errors.

VLC schemes have mainly been considered in the presence of perfect output (or *noiseless*) feedback – including the work of Burnashev [1] , Yamamoto and Itoh [2] and Forney [3], among others. In the context of variable-length coding, the literature on error exponents under *noisy* feedback has been much more sparse. For discrete memoryless channels, schemes able to achieve error exponents between Forney's and Burnashev's noiseless feedback error exponent bounds have been proposed by Draper and Sahai in [4], where an anytime synchronization code and a round-robin message scheduling technique were used for synchronization and reliable feedback.

Synchronization is an important issue in proposed VLC achievability schemes under *noisy* feedback: such schemes rely on re-transmissions to correct errors, and hence the transmitter and receiver must agree on whether the message being transmitted is a new one or a re-transmission. When the feedback link is noiseless, terminals may be easily synchronized. When the feedback is noisy, synchronization cannot be taken for granted and is a topic explicitly addressed in [4] and [5].

All prior work on error exponents for VLC with noisy feedback has focussed on positive rates. As mentioned in [5], the insertion of a few bits alongside the message to denote a sort of message serial number may aid in synchronization. At positive rates, these bits do not decrease the rate considerably. One question is what happens to the achievable error exponents in the limit of zero-rate, where the number of such serial numbered bits may be on the same order as the messages themselves. Are new, more creative solutions available to synchronize VLC schemes at zero-rate with noisy feedback? Here, we consider an AWGN channel with AWGN feedback and characterize an achievable error exponent for VLC for a finite number of messages. We propose a communication scheme that takes ideas proposed in the literature for both DMC [3], [4] and discrete time AWGN channels [6], [7], and uses them in a new VLC scheme able to achieve error exponents higher than those attained under fixed block length transmission (with or without feedback) under a similar power constraint, and comparable to known results under a more relaxed power constraint, briefly described next.

Fixed block length comparison points. Under fixed block length coding, the reliability function attained at rate R is defined as: $E(R) = \lim_{N\to\infty} -\frac{1}{N}\log \mathsf{P}_e^{(N)}$, where $\mathsf{P}_e^{(N)}$ is the smallest probability of error achieved by a code of rate R and block length N. Under the almost sure power constraint (AS), where channel inputs satisfy $\sum_{k=1}^{N} X_k^2 \leq NP$, the error exponent of the transmission of $|\mathcal{W}|$ messages over an AWGN channel with signal to noise ratio $\frac{P}{\sigma^2}$ without feedback is [8]:

$$\text{AS power, no FB: } E\left(|\mathcal{W}|, P, \sigma^2\right) = \frac{|\mathcal{W}|}{4(|\mathcal{W}|-1)}\frac{P}{\sigma^2}, \quad (1)$$

which for two messages becomes $E\left(2, P, \sigma^2\right) = \frac{P}{2\sigma^2}$. Pinsker [9] studied this problem assuming that perfect feedback (PFB) is available. He demonstrated that under a similar power constraint, noiseless feedback may improve the error exponent of the transmission of $|\mathcal{W}| \geq 2$ messages up to

$$\text{AS power, perfect FB: } E^{\text{PFB}}\left(|\mathcal{W}|, P, \sigma^2\right) = \frac{P}{2\sigma^2}. \quad (2)$$

For $|\mathcal{W}| = 2$, feedback does not result in any error exponent gains. The use of noisy feedback with fixed block length codes has been investigated in [6], [7], [10]–[13]. We highlight an achievable error exponent for $|\mathcal{W}| = 2$ under an expected block power constraint (EXP) $\mathsf{E}\left[\sum_{k=1}^{N} X_k^2\right] \leq NP$ [6]:

$$\text{EXP power, noisy FB: } E^{\text{BB}}\left(2, P, \sigma^2\right) \geq \frac{2P}{\sigma^2}. \quad (3)$$

This error exponent can be achieved through a *building block* (BB) scheme [6, Section VII-A, Equation (138)], which takes

[1]This work was partially supported by NSF under awards 1053933 and 1645381. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

advantage of very high amplitude transmissions that correct potential errors occurring with exponentially small probability.

II. Problem Statement and Main Result

One terminal (the transmitter) wishes to transmit a stream of messages to another terminal (the receiver). Each message, M, is selected uniformly from the finite set of messages $\mathcal{W} = \{1, 2, 3, ..., |\mathcal{W}|\}$ and transmitted over the channel shown in Figure 1. The forward and backward directions correspond to AWGN channels characterized at channel use n by

$$Y_n = X_n + N_n,\ N_n \sim \mathcal{N}(0, \sigma^2) \tag{4}$$
$$Z_n = U_n + N_{\text{FB}_n},\ \ N_{\text{FB}_n} \sim \mathcal{N}(0, \sigma^2_{\text{FB}}) \tag{5}$$

where channel inputs X_n, U_n and outputs Y_n, Z_n take on real values, and the noise in the forward channel N_n is independent and identically distributed for every n, as is the noise in the feedback channel, N_{FB_n}. Active feedback is permitted.

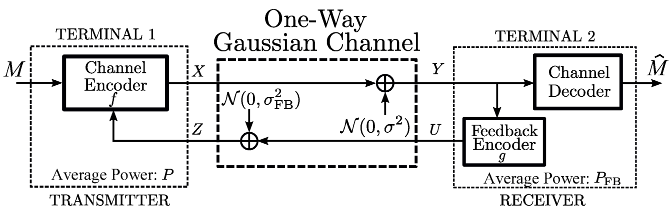

Fig. 1. One-way AWGN channel with active feedback.

Let $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{Z}$ be the set of reals. A *variable length code* $\mathcal{C}_{vl}\left(|\mathcal{W}|, P, \sigma^2, P_{FB}, \sigma^2_{FB}, \Delta, N\right)$ for the transmission of $|\mathcal{W}|$ messages uniformly selected from $\mathcal{W}$ over an AWGN channel with forward and feedback almost sure power constraints of P and P_{FB}, and noise variances of σ^2 and σ^2_{FB} consists of:

1. Forward and feedback encoding functions:

$$f_n : \mathcal{W} \times \mathcal{Z}^{n-1} \to \mathcal{X}, \tag{6}$$
$$g_n : \mathcal{Y}^n \to \mathcal{U}, \quad \text{for } n = 1, 2, 3, \cdots \tag{7}$$

leading to the n-th channel inputs $X_n = f_n\left(M, Z^{n-1}\right)$ and $U_n = g_n\left(Y^n\right)$, satisfying the per block (time-slot) almost sure (AS) or energy power constraint on both channel inputs:

$$\sum_{k=1}^{N} X_k^2 \le NP \quad \text{and} \quad \sum_{k=1}^{N} U_k^2 \le NP_{\text{FB}}; \tag{8}$$

2. Decoding functions for each channel use $n = 1, 2, 3, \cdots$:

$$\phi_n : \mathcal{Y}^n \to \mathcal{W} \cup \{0\}, \tag{9}$$

where 0 corresponds to an *erasure* (message not decoded);

3. A non-negative transmission time Δ (a random variable) defined as the first n for which a message is successfully decoded and not declared as an erasure, [4]:

$$\Delta = n, \text{s.t.} \begin{cases} \phi_n(y^n) \neq 0, \\ \phi_n(y^{n'}) = 0,\ \forall n' < n, \end{cases} \tag{10}$$

where $\mathsf{E}[\Delta] \le N$.

Let $\mathsf{P}_e\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta, N\right)$ be the probability of decoding error. Then, the variable length error exponent for the transmission of $|\mathcal{W}|$ messages over an AWGN channel with noisy feedback is defined as

$$E_{\text{vl}}\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta\right) = \limsup_{N\to\infty} \frac{-\mathsf{P}_e\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta, N\right)}{\mathsf{E}[\Delta]} \tag{11}$$

taken over all possible variable length codes $\mathcal{C}_{\text{vl}}\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta, N\right)$.

Main result. We next introduce our main result, presented for ease of notation for $|\mathcal{W}| = 2$, with a remark on how this may be generalized to any finite $|\mathcal{W}|$ in (33).

Theorem 1: An achievable error exponent subject to the use of a variable length code for the transmission of two messages over an AWGN channel under an AS power constraint is

$$E_{\text{vl}}\left(|\mathcal{W}|, \frac{P}{\sigma^2}\right) \ge \frac{2P}{\sigma^2}. \tag{12}$$

Interestingly, we note that the error exponent for $|\mathcal{W}| = 2$ achieved under an AS power constraint and noisy feedback and VLC is the same as that of fixed block coding (3) and noisy feedback under the more relaxed EXP power constraint (note that we are comparing achievable error exponents). Both of these are four times higher than those achievable for fixed block coding under the AS power constraint and perfect feedback (2). This in some way suggests that the flexibility in power allowed by the EXP power constraint (using very high amplitude retransmissions that correct highly unlikely potential decoding mistakes immediately) and the flexibility of allowing variable decoding times (potential, but highly unlikely very long delays) and re-transmissions (correcting potential decoding mistakes in a different fashion) are in some way related from an error exponent perspective.

III. Scheme Description and Error Analysis

The general operation of the achievability scheme we propose is presented in Figures 2 and 3. The initial transmission of a message over a time-slot of length N in the forward channel is decoded a) correctly, or b) incorrectly, or c) as an erasure. The receiver indicates which of these occurred by setting a bit every time a message is successfully *decoded* (either correctly or incorrectly); otherwise the bit remains unset indicating an *erasure*. This bit is transmitted to the source via the noisy feedback channel, where it is decoded as either "Decoded" or "Erasure". Note that if this feedback bit is incorrectly decoded, an irreparable synchronization error results, since a retransmission request could be interpreted as new message request, and vice versa. The resulting *out of synchronization* events are denoted as $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{E}_3$, and analyzed later – but the main idea is that we will make the probability of synchronization error decay fast enough so as not to be the dominant term in the overall error exponent. The transmission of the decoding status bit is based on redundancy, thus, as we explain later, multiple transmissions of this bit are performed at different time instances allowing the transmitter to recover from temporary synchronization losses. Note that synchronization here is meant at the message level – as in [4], synchronization at the level of channel uses is assumed.

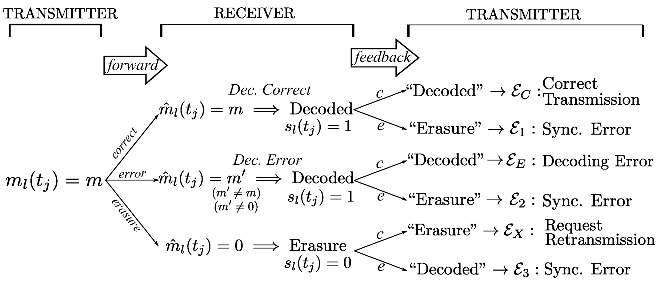

Fig. 2. Scheme operation: events tree. (c = correct and e = error)

This work uses some ideas originally proposed in [4] for Discrete Memoryless Channels (DMC), in which a stream of messages belonging to L virtual data stacks are interleaved and transmitted over a common channel using a round-robin message scheduling technique. Since we focus on AWGN channels, our scheme makes use of a simplex code for both directions, which allowed us to define an erasure decoding rule and to jointly encode the feedback messages of $L-1$ stacks. Figure 3 illustrates a block diagram of our adapted round-robin scheme. We use the term *time instance* to denote a round in which each data stack $l \in \{1, 2, ..., L\}$ has transmitted or retransmitted one message, i.e. during time instant t_j, each stack has a transmission chance (time-slot) using codewords of length N. The following subsections present a detailed explanation of the VLC protocol followed by an analysis of the probability of error, the computation of the expected transmission time, and finally, the proof of our main result.

A. Forward message transmission

A message $m_l(t_j) = m$ for each of the L stacks is encoded using a simplex code of size $|\mathcal{W}|$, and sent over the channel in sequence; each time instance has a duration of LN channel uses, and each stack has a new chance to retransmit after $(L-1)N$ channel uses. The receiver decodes each received sequence $y_l^N(t_j)$ using erasure decoding as follows:

$$\hat{m}_l(t_j) = \begin{cases} m, & \text{if } y_{m_l}^N(t_j) \in \mathcal{A}_m \\ m', & \text{if } y_{m_l}^N(t_j) \in \mathcal{A}_{m' \neq m} \\ 0, & \text{if } y_{m_l}^N(t_j) \in \left(\bigcup_{i=1}^{|\mathcal{W}|} \mathcal{A}_i\right)^c \end{cases} \tag{13}$$

where $\mathcal{A}_m$ denotes the decoding region corresponding to message m. The receiver keeps track of the decoding result at time instance t_j using a status bit for each data stack:

$$s_l(t_j) = \begin{cases} 0, & \text{if } \hat{m}_l(t_j) = 0 \\ 1, & \text{if } \hat{m}_l(t_j) \neq 0 \end{cases} \tag{14}$$

This bit is fed back to the transmitter as we describe next.

B. Transmitting decoding status bits to the source

Status bits are fed back to the source over the noisy backward channel. We propose a feedback communication scheme inspired in [4], in which the reliability of the transmission of erasure/decoding status bits increases with the number of virtual data stacks, L (an even number).

Consider, without loss of generality, that a message from stack $l = 1$, $m_1(t_j) = m$, has been transmitted and that the corresponding status bit $s_1(t_j)$ has been determined by the receiver. Our strategy for communicating this bit consists of first: repeatedly transmitting it $L-1$ times (coded, together with in a packet of $L-1$ of these decision bits corresponding to the other stacks) over the backward channel every time a message from a different stack is sent (decision bit $s_1(t_j)$ cannot be transmitted at the same time as message of stack 1 is coming in, but can be transmitted over all other $L-1$ stacks), and second: using a majority vote to make the final decision. As shown in the block diagram of Figure 3, the transmission of the binary vector $S_1(t_j) = \{s_1(t_j), s_3(t_{j-1}), s_4(t_{j-1}), ..., s_L(t_{j-1})\}$ (where the sub-index in S_1 indicates the stack number which updated its status bit after the last decoding operation) occurs simultaneously to the transmission of message $m_2(t_j)$. We have adopted uppercase letter S to denote status vectors and lower case s for the individual bits. Status vectors S_l for $l = 1, 2, ..., L$ are thus updated every time a new message is decoded by the receiver, and encoded in codewords of length N using a simplex code of size 2^{L-1} symbols subject to the backward power constraint.

We obtain $L-1$ estimates of each status bit, i.e. the transmitter ends up with the estimates $\hat{s}_{1_k}(t_j)$ (for $k \in \{2, 3, 4, ..., L\}$). Here, k identifies the data stack that is transmitting a message in the forward channel while a feedback codeword is being sent in the backward channel. The final decision of each status bit is obtained from a majority vote decoding rule:

$$\hat{s}_1(t_j) = \begin{cases} 0, & \text{if } \hat{s}_{1_k}(t_j) = 0 \text{ for } \frac{L}{2} \text{ or more times} \\ 1, & \text{otherwise} \end{cases}, \tag{15}$$

which is used to decide whether to retransmit when $\hat{s}_1(t_j) = 0$, or to proceed with a new message, when $\hat{s}_1(t_j) = 1$.

C. Probability of error analysis

Let $m_l(t_j) = m$ (instance of random variable M) for ease of notation. Errors occur if the message is incorrectly decoded $\hat{M} \neq M$, or if the transmitter and receiver become unsynchronized:

$$\begin{aligned} &\mathsf{P}_e\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma_{\text{FB}}^2, \Delta, N\right) \\ &\quad = \mathsf{P}\left(\left(\hat{M} \neq M \cap \text{Synched}\right) \cup (\text{Out of Sync.})\right) \\ &\quad \leq \underbrace{\mathsf{P}\left(\hat{M} \neq M \mid \text{Synched}\right)}_{\mathsf{P}(\mathcal{E}_E)} \mathsf{P}(\text{Synched}) + \underbrace{\mathsf{P}(\text{Out of Sync.})}_{\mathsf{P}(\mathcal{E}_S)} \\ &\quad \leq \mathsf{P}(\mathcal{E}_E) + \mathsf{P}(\mathcal{E}_S), \end{aligned} \tag{16}$$

where $\mathcal{E}_S = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3$ and $\mathcal{E}_E$ correspond to the events described in Figure 2. Then,

$$\begin{aligned} E_{\text{vl}}\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma_{\text{FB}}^2, \Delta\right) &= \limsup_{N\to\infty} -\frac{1}{\mathsf{E}[\Delta]} \ln\left(\text{Eq.}(16)\right) \\ &\geq \limsup_{N\to\infty} -\frac{1}{\mathsf{E}[\Delta]} \max\left\{\ln \mathsf{P}(\mathcal{E}_E), \ln \mathsf{P}(\mathcal{E}_S)\right\}. \end{aligned} \tag{17}$$

Next, we analyze the occurrence of events $\mathcal{E}_E$ and $\mathcal{E}_S$.

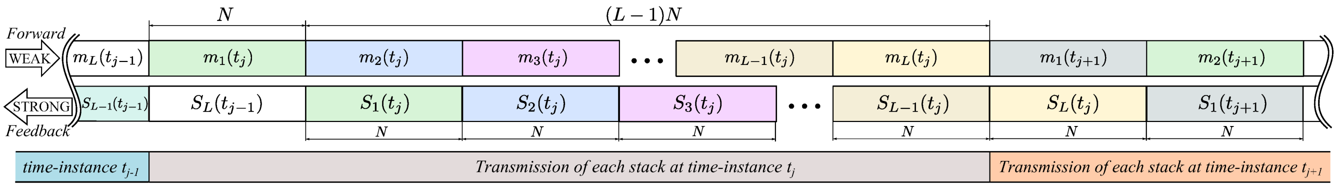

Fig. 3. Block diagram of the proposed scheme, where: $S_l(t_j) = \{s_1(t_j), s_2(t_j), ..., s_{l-1}(t_j), s_{l+1}(t_{j-1}), ..., s_{L-1}(t_{j-1}), s_L(t_{j-1})\}$

1) Message decoding error, $\mathcal{E}_E$: Forney [3] introduced the concept of erasure decoding for the transmission of $|\mathcal{W}|$ symbols, in which the decision region designated for each symbol is smaller than that when the complete decoding space is partitioned into $|\mathcal{W}|$ non overlapping regions. To illustrate how messages transmitted over the forward channel in each time slot of length N are decoded at the receiver, we consider the case of $|\mathcal{W}| = 2$ thus, $M = \{1, 2\}$. This erasure decoding method is based on that proposed in [6, Section VII-A] for the transmission of two messages using a block code of length N:
Forward channel encoding rule: Messages are sent using antipodal signaling for $i = 1, 2, 3, ..., N$, satisfying the power constraint imposed in equation (8) as:

$$X_i = \begin{cases} +\sqrt{P}, & \text{if } M = 1, \\ -\sqrt{P}, & \text{if } M = 2, \end{cases} \tag{18}$$

Received sequence decoding rule: At the receiver side, $\hat{M}$ is determined by first computing:

$$Z = \frac{1}{N}\sum_{i=1}^{N} y_i \sim \begin{cases} \mathcal{N}\left(+\sqrt{P}, \frac{\sigma^2}{N}\right), & \text{if } M = 1, \\ \mathcal{N}\left(-\sqrt{P}, \frac{\sigma^2}{N}\right), & \text{if } M = 2. \end{cases} \tag{19}$$

Thus, the receiver uses the erasure decoding rule of Equation (13), for an arbitrarily small positive parameter δ such that:

$$\mathcal{A}_1 = \{y^N : Z \geq (1-\delta)\sqrt{P}\} \tag{20}$$
$$\mathcal{A}_2 = \{y^N : Z \leq -(1-\delta)\sqrt{P}\} \tag{21}$$

Probability of decoding error: A message decoding error occurs when $\hat{M} \neq M$. Assuming that $M = 1$ has been sent, and by symmetry:

$$\begin{aligned} \mathsf{P}(\text{Dec.err.}) &= \mathsf{P}\left(\hat{M} \neq M | M = 1\right) = \mathsf{P}\left(y^N \in \mathcal{A}_2 | M = 1\right) \\ &= \mathsf{P}\left(Z < -(1-\delta)\sqrt{P} \mid M = 1\right) \\ &= Q\left(\frac{(2-\delta)\sqrt{P}}{\sigma/\sqrt{N}}\right) \leq \frac{1}{2}\exp\left(-N\frac{(2-\delta)^2 P}{2\sigma^2}\right) \end{aligned} \tag{22}$$

Probability of erasure: Given (13), the receiver declares an erasure, $\hat{M} = 0$, if $y^N \in (\mathcal{A}_1 \cup \mathcal{A}_2)^c$, which yields:

$$\begin{aligned} \mathsf{P}(\text{Erasure}) &= \mathsf{P}(\hat{M} = 0 | M = 1) = \mathsf{P}\left(y^N \in (\mathcal{A}_1 \cup \mathcal{A}_2)^c\right) \\ &= \mathsf{P}\left(|Z| < (1-\delta)\sqrt{P} | M = 1\right) \\ &= Q\left(\frac{\delta\sqrt{P}}{\sigma/\sqrt{N}}\right) - Q\left(\frac{(2-\delta)\sqrt{P}}{\sigma/\sqrt{N}}\right) \\ &\doteq \exp\left(-N\frac{\delta^2 P}{2\sigma^2}\right) < \beta \end{aligned} \tag{23}$$

where $a_n \doteq b_n$ denotes that $\frac{1}{n}\ln(a_n/b_n) \to 0$ as $n \to \infty$. Note from (23), that an erasure is declared with exponentially small probability and that this result is exactly the same as [6, Equation (126)]. Next, note from Figure 2, and (23) that:

$$\begin{aligned} \mathsf{P}(\mathcal{E}_E) &= \mathsf{P}(\text{Dec.err.})\mathsf{P}(\text{“Decoded”}|\text{Decoded}) \\ &\leq \mathsf{P}(\hat{M} \neq M \mid M = 1) \overset{(22)}{\leq} \frac{1}{2}\exp\left(-N\frac{2P}{\sigma^2}\right). \end{aligned} \tag{24}$$

2) Synchronization loss, $\mathsf{P}(\mathcal{E}_S)$: This event results from the occurrence of either of the loss of synchronization events described in Figure 2: $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$. Thus:

$$\begin{aligned} \mathsf{P}(\mathcal{E}_S) &= \mathsf{P}\left(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3\right) \leq \mathsf{P}\left(\mathcal{E}_1\right) + \mathsf{P}\left(\mathcal{E}_2\right) + \mathsf{P}\left(\mathcal{E}_3\right) \\ &= \mathsf{P}\left(\text{Dec.corr.}\right)\mathsf{P}\left(\text{“Erasure”} \mid \text{Decoded}\right) + \\ &\quad \mathsf{P}\left(\text{Dec.err.}\right)\mathsf{P}\left(\text{“Erasure”} \mid \text{Decoded}\right) + \\ &\quad \mathsf{P}\left(\text{Erasure}\right)\mathsf{P}\left(\text{“Decoded”} \mid \text{Erasure}\right) \\ &= \mathsf{P}\left(\hat{M} = M \mid M = 1\right)\mathsf{P}\left(\hat{s}_l(t_j) \neq s_l(t_j)\right) + \\ &\quad \mathsf{P}(\hat{M} \neq M \mid M = 1)\mathsf{P}\left(\hat{s}_l(t_j) \neq s_l(t_j)\right) + \\ &\quad \mathsf{P}(\hat{M} = 0 \mid M = 1)\mathsf{P}\left(\hat{s}_l(t_j) \neq s_l(t_j)\right) \\ &= \mathsf{P}\left(\hat{s}_l(t_j) \neq s_l(t_j)\right) \end{aligned} \tag{25}$$

Feedback messages $S_l(t_j)$ are encoded using a simplex code of size 2^{L-1} messages, whose probability of error was derived in [8] and given in (1). Each status bit is extracted from the decoded message $\hat{S}_l(t_j)$. To analyze the probability of error of the estimation of bit $s_1(t_j)$ for the single k-th vector transmission, note that message $S_1(t_j)$ may be incorrectly decoded as one out of $2^{L-1} - 1$ possible messages, that is $\hat{S}_1(t_j) \neq S_1(t_j)$. Also, observe that for these $2^{L-1} - 1$ possibilities, only $\frac{2^{L-1}}{2}$ actually produce an error in decoding bit $s_{1_k}(t_j)$ specifically. Then, the probability of error on decoding a status bit may be bounded as:

$$\begin{aligned} &\mathsf{P}\left(\hat{s}_{1_k}(t_j) \neq s_{1_k}(t_j)\right) \\ &\leq \underbrace{\frac{\frac{2^{L-1}}{2}}{2^{L-1}-1}}_{\hat{s}_{1_1}(t_j) \neq s_{1_1}(t_j)} \underbrace{\exp\left(-N\frac{2^{L-1}}{4\left(2^{L-1}-1\right)}\frac{P_{\text{FB}}}{\sigma^2_{\text{FB}}}\right)}_{2^{L-1}\text{–simplex code: } \hat{S}_1(t_j) \neq S_1(t_j)} \end{aligned} \tag{26}$$

Recalling the majority voting strategy, a status bit is incorrectly decoded if $L/2$ estimates are incorrectly decoded. Since each estimate is independent from the others, and that we select either $L/2$ out of the $L-1$ estimations, we have

$$\mathsf{P}\left(\hat{s}_1(t_j) \neq s_1(t_j)\right) \leq \binom{L-1}{L/2}\left(\mathsf{P}\left(\hat{s}_{1_k}(t_j) \neq s_{1_k}(t_j)\right)\right)^{L/2} \tag{27}$$

where k corresponds to the k-th transmission of a feedback codeword including bit $s_1(t_j)$. Then, from (26) and (27)

$$\mathsf{P}(\mathcal{E}_S) \leq \mathsf{P}\left(\hat{s}_1(t_j) \neq s_1(t_j)\right) \leq \binom{L-1}{L/2}\left(\frac{2^{L-2}}{2^{L-1}-1}\right)^{\frac{L}{2}} \exp\left(-NL\left(\frac{2^{L-4}}{2^{L-1}-1}\right)\frac{P_{\text{FB}}}{\sigma^2_{\text{FB}}}\right) \quad (28)$$

D. Expected transmission time

The expected transmission time must satisfy $E[\Delta] \leq N$, and can be computed from the probability of occurrence of event $\mathcal{E}_X$, shown in Figure 2, which denotes a successful retransmission request detected at both, transmitter and receiver:

$$\begin{aligned}\mathsf{P}(\mathcal{E}_X) &= \mathsf{P}\left(\text{“Erasure”} \cap \text{Erasure}\right)\\ &= \mathsf{P}\left(\text{“Erasure”} \mid \text{Erasure}\right)\mathsf{P}\left(\text{Erasure}\right)\\ &= \left(1 - \mathsf{P}\left(\hat{s}_1(t_j) \neq s_1(t_j)\right)\right)\mathsf{P}(\hat{M} = 0 \mid M = 1)\\ &\leq \mathsf{P}(\text{Erasure}) < \beta, \end{aligned} \quad (29)$$

where the last inequality is for ease of notation, and comes from (23) as the probability of an erasure is exponentially small with N, so, we upper bound it by a (decaying in N) β.

Note that the first transmission of a message $m_l(t_j)$ has a transmission time of N. If a retransmission is necessary, the transmission duration is $(L+1)N$. If two retransmissions are necessary, then this duration becomes $(2L+1)N$, and so

$$\begin{aligned}\mathsf{E}[\Delta] &= N + \beta LN + \beta^2 LN + \beta^3 LN + ...\\ &= N + LN\sum_{i=1}^{\infty}\beta^i = N + LN\frac{\beta}{1-\beta}\end{aligned} \quad (30)$$

Note that since $\lim_{N\to\infty} LN\frac{\beta}{1-\beta} = 0$, the expected transmission time is $\mathsf{E}[\Delta] = N$, as $\beta \to 0$ exponentially fast in N.

E. Proof of Theorem 1

Equation (12) results from the expected transmission time in (30), and using (24) and (28) in (17)

$$\begin{aligned}&E_{\text{vl}}\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta\right)\\ &\geq \min\left\{\underbrace{L\left(\frac{2^{L-4}}{2^{L-1}-1}\right)\frac{P_{\text{FB}}}{\sigma^2_{\text{FB}}}}_{\text{from (28)}}, \underbrace{\frac{2P}{\sigma^2}}_{\text{from (24)}}\right\}.\end{aligned} \quad (31)$$

Note in Equation (31) that the following relation must hold so that the error exponent is dominated by the decoding error event $\mathcal{E}_E$ rather than the synchronization error:

$$L\left(\frac{2^{L-4}}{2^{L-1}-1}\right)\frac{P_{\text{FB}}}{\sigma^2_{\text{FB}}} > \frac{2P}{\sigma^2} \quad (32)$$

The above condition yields (12) in Theorem 1, and may be used to obtain a lower bound on the choice of L for given forward and backward channel SNRs. For example, for $\frac{P}{\sigma^2} = 1$ and $\frac{P_{\text{FB}}}{\sigma^2_{\text{FB}}} = 2$, it may be verified numerically that selecting $L \geq 8$ will satisfy (32).

IV. Conclusion

The error exponent achieved by the VLC scheme presented here for the one-way AWGN channel with active AWGN feedback with a more restricted AS power constraint turns out to be equivalent to a known result achieved under fixed block length coding using the more relaxed EXP power constraint, which is four times that achieved without feedback for $|\mathcal{W}| = 2$. Interestingly, we have shown that synchronization even over the noisy channel can be handled using a round-robin and anytime-like coding scheme that depends on a parameter L, the number of message stacks that are interleaved. Our results demonstrate that an appropriate choice of L guarantees that the probability of error is dominated by the event of incorrect messages decoding in the forward channel. We note that our scheme does *not* require a better feedback channel than forward channel – any noisy feedback channel may be accommodated by a larger choice of L so as to satisfy (32). We also note that the generalization on the probability of erasure decoding provided in [7] may be used to extend Theorem 1 to any finite $|\mathcal{W}|$ as follows:

$$E_{\text{vl}}\left(|\mathcal{W}|, P, \sigma^2, P_{\text{FB}}, \sigma^2_{\text{FB}}, \Delta\right) \geq \left(\frac{|\mathcal{W}|}{|\mathcal{W}|-1}\right)\frac{P}{\sigma^2}. \quad (33)$$

References

[1] M. V. Burnashev, “Data transmission over a discrete channel with feedback: Random transmission time,” *Problemy peredachi informatsii*, vol. 12, no. 4, pp. 10–30, 1976.

[2] H. Yamamoto and K. Itoh, “Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.),” *IEEE Transactions on Information Theory*, vol. 25, no. 6, pp. 729–733, November 1979.

[3] G. Forney, “Exponential error bounds for erasure, list, and decision feedback schemes,” *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[4] S. C. Draper and A. Sahai, “Variable-length channel coding with noisy feedback,” *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 355–370, Jun. 2008.

[5] A. Sato and H. Yamamoto, “Error exponents of discrete memoryless channels and awgn channels with noisy feedback,” in *2010 International Symposium On Information Theory and Its Applications*, Oct 2010, pp. 452–457.

[6] Y. H. Kim, A. Lapidoth, and T. Weissman, “Error Exponents for the Gaussian Channel With Active Noisy Feedback,” *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1223–1236, March 2011.

[7] K. Palacio-Baus and N. Devroye, “Achievable Error Exponents of the One-Way and Two-Way AWGN Channels,” *Submitted to: IEEE Transactions on Information Theory*, December 2018.

[8] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *The Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

[9] M. S. Pinsker, “The probability of error in block transmission in a memoryless Gaussian channel with feedback,” *Problemy Peredachi Informatsii*, vol. 4, no. 4, pp. 3–19, 1968.

[10] Y. H. Kim, A. Lapidoth, and T. Weissman, “The Gaussian Channel with Noisy Feedback,” in *Proc. IEEE Int. Symp. Inf. Theory*, June 2007, pp. 1416–1420.

[11] Y. Xiang and Y. H. Kim, “Gaussian Channel with Noisy Feedback and Peak Energy Constraint,” *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4746–4756, Aug. 2013.

[12] M. V. Burnashev and H. Yamamoto, “On the reliability function for a noisy feedback Gaussian channel: Zero rate,” *Problems of Information Transmission*, vol. 48, no. 3, pp. 199–216, 2012.

[13] ——, “Noisy feedback improves the Gaussian channel reliability function,” in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 2554–2558.