The Gaussian Channel with Noisy Feedback: Near-Capacity Performance via Simple Interaction

Assaf Ben-Yishai and Ofer Shayevitz

Abstract—Consider a pair of terminals connected by two independent additive white Gaussian noise channels, and limited by individual power constraints. The first terminal would like to reliably send information to the second terminal, within a given error probability. We construct an explicit interactive scheme consisting of only (non-linear) scalar operations, by endowing the Schalkwijk-Kailath noiseless feedback scheme with modulo-arithmetic. Our scheme achieves a communication rate close to the Shannon limit, in a small number of rounds. For example, for an error probability of $10^{-6}$, if the Signal to Noise Ratio (SNR) of the feedback channel exceeds the SNR of the forward channel by 20dB, our scheme operates 0.8dB from the Shannon limit with only 19 rounds of interaction. In comparison, attaining the same performance using state of the art Forward Error Correction (FEC) codes requires two orders of magnitude increase in delay and complexity. On the other extreme, a minimal delay uncoded system with the same error probability is bounded away by 9dB from the Shannon limit.

I. INTRODUCTION

Feedback cannot improve the capacity of point-to-point memoryless channels [1]. Nevertheless, noiseless feedback can significantly simplify the transmission schemes and improve the error probability performance, see e.g. [2]–[5]. These elegant schemes fail however in the presence of arbitrarily small feedback noise, rendering them grossly impractical. This fact has been initially observed in [3] for the Additive White Gaussian Noise (AWGN) channel, and further strengthened in [6]. A handful of works have tackled the problem of noisy feedback as means for improving error performance, see e.g. [7]–[10]. However, these works attain their superior error performance at the cost of a significant increase in complexity w.r.t. their noiseless feedback counterparts. There appears to be no simple scheme (in the spirit of [3]–[5]) that is robust to feedback noise known hitherto.

Our work is therefore motivated by the following question: Does the simplicity of the infeasible noiseless feedback schemes extend itself to the more realistic noisy feedback setup, while still offering near-optimal performance? While the answer to this question appears to be negative if one insists on approaching capacity in the usual sense (vanishing error probability in the limit of large delay), we answer it here in the affirmative under a fixed (but small) error probability criterion. Specifically, we consider the following setup: Two Terminals A and B are connected by pair of independent AWGN channels, and are limited by individual power constraints. The channel from Terminal A (resp. B) to Terminal B (resp. A) is referred to as the feedforward (resp. feedback) channel. Terminal A wishes to send bits to Terminal B, within a given bit error probability. The figure-of-merit we look at is the capacity gap, which is the amount of excess SNR required by our scheme over the minimal possible SNR for an optimal Shannon scheme (of unbounded complexity), achieving the same bit rate and bit error probability. For this setup, we introduce and analyze a simple interactive scheme, that can operate near capacity. Our construction is based on an interactive representation of the Schalkwijk-Kailath (S-K) noiseless feedback scheme [3], endowed with modulo-arithmetic. Specifically, our scheme is founded on the following observations (further elaborated in the sequel):

1) The capacity gap (in dB) attained by the S-K scheme (for noiseless feedback) is inversely proportional to the number of iterations, and hence capacity is approached in a small number of rounds.

2) The S-K scheme can be interpreted as follows (see also Subsection III-B). Terminal A encodes and sends its message via Pulse Amplitude Modulation (PAM), and in subsequent rounds, sends a scaled version of the estimation error of Terminal B (which is computable due to noiseless feedback), thereby exponentially decreasing the variance of the total estimation error. This scheme can operate using only passive feedback. Alternatively, Terminal B could clearly employ active feedback by always transmitting its current estimate of the message, rather then its observations. This simple tweak is meaningless in the noiseless feedback case, yet turns out to be essential when feedback is noisy.

3) Suppose we use the S-K scheme when noise is present in the feedback channel. In each round, Terminal B knows the sum of the estimation error and the PAM message, whereas Terminal A knows the PAM message only. Describing the estimation error to Terminal A over the feedback channel is therefore a joint source-channel coding problem with side information at the receiver. Exploiting the side information could potentially yield a markedly better description of the estimation error. One simple way to reap this gain is by employing modulo-arithmetic in the spirit of Tomlinson-Harashmia precoding [11], [12].

4) Following the above joint source-channel coding procedure, the estimation error of Terminal B becomes known at Terminal A, up to some excess additive noise induced by the noisy feedback. Due to the modulo-linearity of the operations, this excess noise can be effectively pushed into the forward channel.

In a nutshell, our scheme operates as follows. Terminal A
encodes and sends its message using PAM. In subsequent rounds, Terminal B computes its best linear estimate of the message, and feeds back a scaled version of that estimate, modulo a fixed interval. In turn, Terminal A employs a suitable modulo computation and obtains the estimate error, corrupted by excess additive noise. This quantity is then properly scaled and sent over the feedforward channel to Terminal B. After a fixed number of rounds, Terminal B decodes the message via a simple minimum distance rule. Loosely speaking, the scheme’s error probability is dictated by the events of a modulo aliasing in one of the rounds, and the event where the remaining estimation noise in the last round exceeds half the minimum distance of the PAM constellation. The maximal number of rounds is limited by the need to control the modulo-aliasing errors.

The resulting capacity gap (Theorem I) consists of four terms: 1) An “S-K term” that is inversely proportional to the number of rounds; 2) A “concatenated channel” term, that corresponds to the decrease in SNR incurred by trivially concatenating the forward and feedback channels, and is (roughly) inversely proportional to the excess SNR of the feedback channel over the feedforward channel; 3) a “modulo-aliasing” term that stems from the error floor imposed by employing the modulo operation, and is (roughly) inversely proportional to the SNR of the feedback channel; and 4) An auxiliary term that is (roughly) inversely proportional to the SNR of the feedforward channel.

As an example, for a bit error probability of 10^{-6}, if the SNR of the feedback channel exceeds the SNR of the feedforward channel by 20dB (resp. 10dB), our scheme operates at a capacity gap of 0.8dB (resp. 3.5dB), with only 19 (resp. 11) rounds of interaction. This should be juxtaposed against two reference systems: On the one hand, state-of-the-art FEC codes attaining the same capacity gap and error probability, require roughly a two orders-of-magnitude increase in delay and complexity. On the other hand, the capacity gap attained by a minimal delay uncoded system with the same error probability, is at least 9dB.

The rest of the paper is organized as follows. The problem setup is introduced in Section III. Necessary background including the capacity gap of uncoded PAM and an active feedback representation of the S-K scheme are given in Section III. Our new scheme is described in Section IV, and its performance is discussed in Section V. A detailed analysis of the scheme is provided in Section VI. Some numerical results and figures are given in Section VII. Implementation issues and the applicability of our scheme to real world scenarios are treated in Section VIII. A discussion of the results and their context appears in Section IX.

II. Setup

In the sequel, we use the following notations. For any number \( x > 0 \), we write \( x_{\text{dB}} \triangleq 10 \log_{10}(x) \) to denote the value of \( x \) in decibels. The Gaussian Q-function is \( Q(x) \triangleq (2\pi)^{-\frac{1}{4}} \int_{x}^{\infty} \exp(-u^2/2) \, du \), and \( Q^{-1}(\cdot) \) is its functional inverse. We use the vector notation \( x^n \triangleq (x_1, \ldots, x_n) \). We write
We assume that Terminal A (resp. Terminal B) is subject to a power constraint \( P \) (resp. \( \tilde{P} \)), namely
\[
\sum_{n=1}^{N} \mathbf{E}(X_n^2) \leq N \cdot P,
\]
(5)
\[
\sum_{n=1}^{N} \mathbf{E}(\tilde{X}_n^2) \leq N \cdot \tilde{P}.
\]
(6)
We denote the feedforward (resp. feedback) SNR by \( \text{SNR}^f \equiv \frac{P}{\sigma^2} \) (resp. \( \text{SNR}^l \equiv \frac{\tilde{P}}{\sigma^2} \)). The ratio between the feedback SNR and the feedforward SNR is denoted by \( \Delta \text{SNR} \equiv \frac{\text{SNR}^f}{\text{SNR}^l} \). Throughout this work, we assume that the feedback channel has excess SNR over the feedforward channel, i.e., \( \Delta \text{SNR} > 1 \).

An interactive scheme \((\varphi, \tilde{\varphi})\) is associated with a rate \( R \equiv \frac{\log N}{\text{SNR}} \) and an error probability \( p_e \), which is the probability that Terminal B errs in decoding the message \( W \) at time \( N \), under the optimal decision rule.

The capacity gap \( \Gamma \) attained by the scheme is defined as follows. Recall that the Shannon capacity of the AWGN channel implies that the maximal rate achievable by any scheme (of unbounded complexity/delay, with or without feedback) under vanishing error probability, is given by
\[
C = \frac{1}{2} \log(1 + \text{SNR}).
\]
(7)
Conversely, the minimal SNR required to attain a rate \( R \) is \( 2^{2R} - 1 \). The capacity gap is the excess SNR required by the scheme, i.e.,
\[
\Gamma(\varphi, \tilde{\varphi}) = \Gamma \equiv \frac{\text{SNR}}{2^{2R} - 1}.
\]
(8)
Note that if a nonzero bit/symbol error probability is allowed, then one can achieve rates exceeding the Shannon capacity \( C \), and this effect should in principle be accounted for, to make the definition of the capacity gap fair. However, for small error probabilities the associated correction factor (given by the inverse of the corresponding rate distortion function) becomes negligible, and we therefore ignore it in the sequel.

III. PRELIMINARIES

In this section, we describe the three building blocks underlying our interactive scheme. First, we discuss the performance of PAM uncoded transmission, and the associated capacity gap. We then provide a simple active (noiseless) feedback interpretation of the S-K scheme, and derive the associated decay of the capacity gap as a function of the number of interaction rounds. Lastly, we briefly recall the notations and properties of modulo-arithmetic to be used in our scheme.

A. Uncoded PAM

PAM is a simple and commonly used modulation scheme, where \( 2^R \) symbols are mapped (one-to-one) to the set \( \{ \pm 1\eta, \pm 3\eta, \ldots, \pm (2^R - 1)\eta \} \). Canonically, the normalization factor \( \eta \) is set so that the overall mean square of the constellation (assuming equiprobable symbols) is unity. A straightforward calculation yields \( \eta = \sqrt{3/(2^{2R} - 1)} \). In the general case where the mean square of the constellation is constrained to be \( P \), \( \eta \) is replaced with \( \eta\sqrt{P} \).

It is easy to show that for an AWGN channel with zero mean noise of variance \( \sigma^2 \) and average input power constraint \( P \), the probability of error incurred by the optimal detector is bounded by the probability that the noise exceed half the minimal distance of the PAM constellation, i.e.,
\[
p_e < 2Q\left( \frac{\sqrt{P}\eta}{\sigma} \right) = 2Q\left( \frac{3\text{SNR}}{2^R - 1} \right).
\]
(9)
Fixing the error probability \( p_e \) and solving the inequality \( \text{Q} \) for \( R \) yields:
\[
R > \frac{1}{2} \log \left( 1 + \frac{\text{SNR}}{\Gamma} \right),
\]
(10)
where
\[
\Gamma(0, p_e) = \frac{1}{3} \left( Q^{-1} \left( \frac{p_e}{2} \right) \right)^2.
\]
(11)
Comparing (10) and (7), we see that PAM signaling with error probability \( p_e \) admits a capacity gap of \( \Gamma(0, p_e) \). For a typical value of \( p_e = 10^{-6} \), the capacity gap of uncoded PAM is \( \Gamma(0.04) = 9 \text{dB} \).

Finally, we assume as usual that bits are mapped to PAM constellation symbols via Gray labeling. The associated bit error probability can thus be bounded by
\[
p_b < \frac{2}{R} \text{Q} \left( \frac{\sqrt{P}\eta}{\sigma} \right) + 2Q \left( 3\sqrt{P}\eta/\sigma \right) \approx \frac{p_e}{R}. \]
(12)
where the approximation is becomes tight for small \( p_e \) due to the strong decay of the Q-function.

B. The S-K Scheme with Active Feedback

Consider the setting of communication over the AWGN with noiseless feedback, i.e., where \( \sigma^2 = 0 \). The S-K scheme with active feedback is described as follows. First, Terminal A maps the message \( W \) to a PAM constellation point \( \Theta \). In the first round, it sends a scaled version of \( \Theta \) satisfying the power constraint \( P \). In subsequent rounds, Terminal B maintains an estimate \( \hat{\Theta}_n \) of \( \Theta \) given all the observation it has, and feeds it back to Terminal A. In turn, Terminal A computes the estimation error \( \varepsilon_n \equiv \hat{\Theta}_n - \Theta \), and sends a properly scaled version of it to Terminal B. Formulating the above description:

(A) Initialization:
Terminal A: Map the message \( W \) to a PAM point \( \Theta \).
Terminal A \( \Rightarrow \) Terminal B:
• Send \( X_1 = \sqrt{P}\Theta \)
• Receive \( Y_1 = X_1 + Z_1 \)

Terminal B: Initialize the \( \Theta \) estimate\(^\dagger\) to \( \hat{\Theta}_1 = \frac{Y_1}{\sqrt{P}} \).

(B) Iteration:
Terminal B \( \Rightarrow \) Terminal A:
• Send the current \( \Theta \) estimate: \( \hat{X}_n = \hat{\Theta}_n \)
• Receive \( \hat{Y}_n = \hat{X}_n \)

Terminal A: Compute the estimation error \( \varepsilon_n = \hat{Y}_n - \Theta \).

\(^\dagger\)Note that this is the minimum variance unbiased estimate of \( \Theta \).
Terminal A ⇒ Terminal B:
- Send the scaled estimation error $X_{n+1} = \alpha_n \tilde{\varepsilon}_n$, where $\alpha_n = \frac{1}{\sqrt{\sigma_n^2 + \beta_n^2}}$ so that the input power constraint holds, and where $\sigma_n^2 \overset{\text{def}}{=} E[\varepsilon_n^2]$.
- Receive $Y_n = X_n + Z_n$

Terminal B: Update the $\Theta$ estimate $\hat{\Theta}_{n+1} = \hat{\Theta}_n - \tilde{\varepsilon}_n$, where

$$\tilde{\varepsilon}_n = \beta_{n+1} Y_{n+1}$$

is the Minimum Mean Square Error (MMSE) estimate of $\varepsilon_n$, thus

$$\beta_{n+1} = \frac{\sigma_n^2}{\sigma_n^2 + \beta_n^2} = \frac{1}{1 + \text{SNR}}.$$  

(C) Decoding:
At time $N$ the receiver decodes the message using a minimum distance decoder for $E_{N}$ w.r.t. the PAM constellation.

To calculate the error probability and rate attained by the S-K scheme, we note that $\varepsilon_{n+1} = \varepsilon_n - \tilde{\varepsilon}_n$. Computing the corresponding variance by plugging in the optimal values of $\alpha_n, \beta_n, Y_n$ yields:

$$\text{SNR}_N = \frac{\text{SNR} \cdot (1 + \text{SNR})^N - 1}{\text{SNR} \cdot (1 + \text{SNR})^N - 1}.$$  

Plugging $\text{SNR}_N$ into (2) and bounding the Q-function by $Q(x) < \frac{1}{2} \exp(-\frac{1}{2} x^2)$ gives:

$$p_e < \frac{1}{2} \exp \left( \frac{3 \text{SNR} \cdot (1 + \text{SNR})^N - 1}{2^{2N\text{SNR}} - 1} \right).$$  

Plugging in the AWGN capacity (7) and removing the “−1” term, we obtain:

$$p_e < \frac{1}{2} \exp \left( \frac{3 \text{SNR} \cdot (1 + \text{SNR})^N \cdot (1 + \text{SNR})^{N-1}}{2^{2N\text{SNR}} - 1} \right).$$  

which is the well-known doubly exponential decay of the error probability of the S-K scheme.

Let us now provide an alternative interpretation of the S-K scheme performance, in terms of the capacity gap attained after a finite number of rounds. Plugging $\text{SNR}_N$ in (10) yields:

$$R > \frac{1}{2N} \log \left( \frac{1 + \text{SNR} \cdot (1 + \text{SNR})^N - 1}{\Gamma} \right).$$  

Plugging the resulting $R$ in the definition of the capacity gap (8) and using the inequality: $-\ln(1 - x) \leq \frac{x}{\sqrt{1 - x}}$ for $x < 1$ yields:

$$\Gamma_{\text{dB}}(p_e, N) < \frac{\Gamma_{0,\text{dB}}(p_e)}{N} + \frac{10/\ln 10}{\text{SNR} \cdot \Gamma_{0,\text{dB}}(p_e) - 1},$$  

which reduces to the following approximation for $\text{SNR} \gg 1$:

$$\Gamma_{\text{dB}}(p_e, N) \approx \frac{\Gamma_{0,\text{dB}}(p_e)}{N}.$$  

This phenomenon is depicted by the dashed curve in Fig. 4.

C. Modulo Arithmetics
We briefly overview basic notations and properties of modulo-arithmetic. For a given $d > 0$, define the modulo function

$$M[x] \overset{\text{def}}{=} x - d \cdot \text{round} \left( \frac{x}{d} \right).$$

where the round($\cdot$) operator returns nearest integer to its argument. The following properties are easily verified:

(i) $M[x] \in [-\frac{d}{2}, \frac{d}{2})$. Then $\text{M}[x + V]$ is uniformly distributed over $[-\frac{d}{2}, \frac{d}{2})$ for any $x \in \mathbb{R}$.

(ii) if $\tilde{d}_1 + \tilde{d}_2 \in [-\frac{d}{2}, \frac{d}{2})$, then

$$M[M[x + \tilde{d}_1] + \tilde{d}_2 - x] = d_1 + d_2.$$  

otherwise, a modulo-aliasing error term of $kd \neq 0$ is added to the right-hand-side (23), for some integer $k$.

(iii) Let $V \sim \text{Uniform}([-\frac{d}{2}, \frac{d}{2}))$. Then $M[x + V]$ is uniformly distributed over $[-\frac{d}{2}, \frac{d}{2})$ for any $x \in \mathbb{R}$.

(iv) Therefore, $E[M[x + V^2]]^2 = \frac{d^2}{12}$.  

IV. THE PROPOSED SCHEME
In what follows we assume that the terminals share a common random i.i.d sequence $\{V_{\alpha}\}_{\alpha=1}^{\infty}$, where $V_{\alpha} \sim \text{Uniform}([-\frac{d}{2}, \frac{d}{2})$. Furthermore, we set $d = \sqrt{12P}$ which guarantees that $E[M[x + V_{\alpha}]]^2 = P$ for any $x \in \mathbb{R}$. Recall that the estimation of the PAM point at Terminal B and time instance $n$ is denoted by $\hat{\Theta}_n$, and the associated estimation error by $\varepsilon_n \overset{\text{def}}{=} \hat{\Theta}_n - \Theta$.

Our scheme is described below.

(A) Initialization:
Terminal A: Map the message $W$ to a PAM point $\Theta$.  

Terminal A ⇒ Terminal B:
- Send $X_1 = \sqrt{P} \Theta$
- Receive $Y_1 = X_1 + Z_1$

Terminal B: Initialize the $\Theta$ estimate $\hat{\Theta}_1$ to $\hat{\Theta}_1 = \frac{Y_1}{\sqrt{P}}$.

(B) Iteration:
Terminal B ⇒ Terminal A:
- Given the $\Theta$ estimate $\hat{\Theta}_n$, compute and send $\tilde{X}_n = M[\gamma_n \hat{\Theta}_n + V_n]$  

Receive $\tilde{Y}_n = \tilde{X}_n + \tilde{Z}_n$

Terminal A: Extract a noisy scaled version of estimation error $\varepsilon_n$:

$$\varepsilon_n = M[\tilde{Y}_n - \gamma_n \Theta - V_n]$$  

Note that $\varepsilon_n = \gamma_n \varepsilon_n + \tilde{Z}_n$, unless a modulo-aliasing error occurs.

Terminal A ⇒ Terminal B:
- Send a scaled version of $\varepsilon_n$: $X_{n+1} = \alpha \varepsilon_n$, where $\alpha$ is set so that to meet the input power constraint $P$ (computed later).
- Receive $Y_n = X_n + Z_n$

2We arbitrarily define round $(k + \frac{1}{2}) = k + 1$ for every integer $k$.  

3Note that this is the minimum variance unbiased estimate of $\Theta$.  

**Terminal B:** Update the $\Theta$ estimate \( \hat{\Theta}_{n+1} = \hat{\Theta}_n - \varepsilon_n \), where

\[
\varepsilon_n = \beta_n Y_{n+1}
\]

(26) is the MMSE estimate of $\varepsilon_n$. The optimal selection of $\beta_n$ is described in the sequel.

(C) **Decoding:**

At time $N$ the receiver decodes the message using a minimum distance decoder for $\hat{\Theta}_N$ w.r.t. the PAM constellation.

**V. MAIN RESULT**

Recall the capacity gap function $\Gamma_0(\cdot)$ of uncoded PAM given in (11). Fix some target error probability $p_e$. Define:

\[
\lambda \overset{\text{def}}{=} 3 \left( \frac{p_e}{4N} \right)^{1/2}, \\
\Psi_1 \overset{\text{def}}{=} 1 + (\lambda \cdot \Delta \text{SNR})^{-1}, \\
\Psi_2 \overset{\text{def}}{=} \frac{1}{1 - (\lambda \cdot \text{SNR})^{-1}}, \\
\Psi_3 \overset{\text{def}}{=} \frac{10}{10} \ln 10 \left( \Psi_1^{\frac{1}{N}} \Psi_2^{\frac{1}{N}} \Gamma_0^{-\frac{1}{N}} \left( \frac{p_e}{1-N} \right) - 1 \right)
\]

(27)

**Theorem 1:** For a proper choice of parameters, the interactive communication scheme described in Section [V] achieves in $N$ rounds an error probability $p_e$ and a capacity gap $\Gamma_{DB}^*$ satisfying:

\[
\Gamma^*_{dB}(p_e, N) < \frac{1}{N} \Gamma_0_{dB}(\frac{p_e}{1}) + \frac{N - 1}{N} (\Psi_1_{dB} + \Psi_2_{dB} + \Psi_3)
\]

(28)

We prove this theorem is Section [VI]

**Remark 2:** $\lambda$ is a factor that encapsulates the cost of controlling the modulo-aliasing error, as seen below. It decreases with a decreasing $p_e$.

**Remark 3:** $\Psi_1$ is a penalty term roughly corresponding to the decrease in SNR incurred by trivially concatenating the forward and feedback channels. To see this, consider the **concatenated channel** from $X_n$ to $\tilde{Y}_n$, where Terminal B performs simple linear scaling to meet the power constraint $\tilde{P}$, i.e. $\tilde{X}_n = \sqrt{\frac{\tilde{P}}{N+2}} Y_n$. The SNR of this channel is

\[
\text{SNR}_{\text{concatenated}} = \frac{\text{SNR} \cdot \text{SNR}}{\text{SNR} + \text{SNR} + 1}
\]

(29) hence, the associated SNR loss w.r.t. the feedforward channel is

\[
\frac{\text{SNR}}{\text{SNR}_{\text{concatenated}}} = 1 + \frac{1}{\Delta \text{SNR}} + \frac{1}{\text{SNR}} \approx 1 + \frac{1}{\Delta \text{SNR}}.
\]

(30)

This latter expression is very similar to $\Psi_1$, with the exception of the additional $\lambda$ factor. Hence, loosely speaking, $\Psi_1$ encapsulates the inherent loss due to essentially employing a feedback scheme over the concatenated channel, together with a feedback power reduction by the amount of $\lambda$ used to avoid modulo-aliasing errors. This loss vanishes for a fixed $p_e$ as $\Delta \text{SNR}$ increases. However, if $\Delta \text{SNR}$ is fixed, this term does not vanish in the limit of high SNR.

**Remark 4:** $\Psi_2$ can be interpreted as a penalty term stemming from the modulo-aliasing error endemic to the system, due to the presence of feedback noise in the modulo operations at Terminal A. For a fixed SNR, the maximal value of $\lambda$ supported by our scheme is given by $\text{SNR}^{-1}$, which in turn dictates the minimal error probability that can be attained. Due to this error floor, our scheme cannot achieve any rate in the usual sense. The loss of SNR incurred by $\Psi_2$ vanishes for any fixed error probability $p_e$ as SNR increases.

**Remark 5:** $\Psi_3$ is an additional penalty term (already in logarithmic scale), that result from the fact that we consider the capacity gap in terms of SNR ratios, whereas the explicit term arising from the capacity formula is related to $\log(1 + \text{SNR})$ rather than $\log(\text{SNR})$. Note that $\Psi_3 = O(\text{SNR}^{-1})$.

**Corollary 1 (High SNR behavior):** Let $\Delta \text{SNR}$ and $p_e$ be fixed. The capacity gap attained by our scheme for SNR large
based on the actual distribution of $\psi_1, \psi_2$ the upper bound can be derived via a simple coupling argument.

The first term is roughly the capacity gap of the S-K scheme with noiseless feedback. The second term pertains to the SNR loss w.r.t. a concatenated channel as well as modulo-aliasing errors, as discussed in Remark 3.

Remark 6: Note that there is a “low SNR” regime (related also to the target error probability or to $\Delta \text{SNR}$), where the loss terms $\psi_{1, \text{dB}} + \psi_{2, \text{dB}}$ are larger than say $\Gamma_{0, \text{dB}}(p_e)$. In that case, setting $N = 1$, namely using an uncoded system with no interaction, is the optimal choice of parameters for our scheme. As we shall see however, for many practical values of SNR, SNR and $p_e$, interaction results in significant gains.

VI. PROOF OF MAIN RESULT

In Subsection III-B we analyzed the error probability of S-K with noiseless feedback, relying on the fact that all the noises are jointly Gaussian, including the noise $\varepsilon_N$ experienced by the PAM decoder. To that end, we were able to directly use the error probability analysis of simple PAM over AWGN discussed in Subsection II-A.

In the noisy feedback case however, the non-linearity induced by modulo operations at both terminals induce a non-Gaussian distribution of $\varepsilon_N$. An analysis of the decoding error based on the actual distribution of $\varepsilon_N$ is very involved. Yet, an upper bound can be derived via a simple coupling argument described below.

Recall that Terminal A computes $\bar{\varepsilon}_n$, a noisy scaled version of the estimation error of Terminal B, via a modulo operation $\mod{d}$. For any $n \in \{1, \ldots, N - 1\}$ we define $E_n$ as the event where this computation results in a modulo-aliasing error, i.e.,

$$E_n = \{ \gamma_n \varepsilon_n + \mod{d} \notin \left[ -\frac{d}{2}, \frac{d}{2} \right] \}.$$  

Furthermore, we define $E_N$ as the PAM decoding error event:

$$E_N = \{ \varepsilon_N \notin \left[ -\frac{d_{\text{min}}}{2}, \frac{d_{\text{min}}}{2} \right] \},$$  

where $d_{\text{min}}$ is the PAM constellation minimal distance. As mentioned above, the distribution of $\varepsilon_N$ is not Gaussian due to the nonlinearity introduced by the modulo operations. To circumvent this, we consider the following upper bound for the error probability:

$$p_e < \Pr \left( \bigcup_{n=1}^{N} E_n \right).$$  

(34)

The inequality stems from the fact that a modulo-aliasing error does not necessarily cause a PAM decoding error.

To proceed, we define the coupled system as a system that is fed by the same message and experiences the (sample-path) exact same noises, with the only difference being that no modulo operations are implemented at neither of the terminals. Clearly, the coupled system violates the power constraint at

the error probability:

The inequality stems from the fact that a modulo-aliasing error does not necessarily cause a PAM decoding error.

To proceed, we define the coupled system as a system that is fed by the same message and experiences the (sample-path) exact same noises, with the only difference being that no modulo operations are implemented at neither of the terminals. Clearly, the coupled system violates the power constraint at

enough, can be approximated by

$$\Gamma_{\text{dB}}(p_e, N) \approx \frac{1}{\Delta} \Gamma_{0, \text{dB}}(\frac{d}{2}) + \frac{N-1}{\Delta} [1 + \frac{1}{\lambda \Delta \text{SNR}}] \text{dB}. \quad (31)$$

Remark 7: $\Lambda$ defined in (27) is a special case of the above, where $p_m$ is set to be $\frac{1}{2N}$. Note again that it must hold that $\lambda \leq 2$. From (39) it stems that:

$$\alpha = \sqrt{\frac{P}{\lambda P}}.$$  

(42)

Terminal B, however, given the message $W$, all the random variables in the coupled system are jointly Gaussian, and in particular, the estimation errors $\varepsilon_n$ in that system are Gaussian for $n = 1, \ldots, N$. Moreover, it is easy to see that the estimation errors are sample-path identical between the original system and the coupled system until the first modulo-aliasing error occurs. To be precise:

Lemma 1: Let $\bar{\Pr}$ denote the probability operator in for the coupled process. Then for any $N > 1$:

$$\bar{\Pr} \left( \bigcup_{n=1}^{N} E_n \right) = \bar{\Pr} \left( \bigcup_{n=1}^{N} E_n \right).$$  

(35)

Proof:

$$\Pr \left( \bigcup_{n=1}^{N} E_n \right) = \Pr(E_1) + \sum_{n=2}^{N-1} \Pr \left( \bigcap_{i=1}^{n-1} (E_i^C) \bigcap E_n \right).$$  

(36)

Moreover, for any $i \in \{2, \ldots, N\}$

$$\Pr \left( \bigcap_{i=1}^{n} (E_i^C) \bigcap E_n \right) = \bar{\Pr} \left( \bigcap_{i=1}^{n-1} (E_i^C) \bigcap E_n \right).$$  

(37)

and trivially $\Pr(E_1) = \bar{\Pr}(E_1)$. Combining the above with (34) and applying the union bound in the coupled system, we obtain

$$p_e \leq \sum_{n=1}^{N} \Pr (E_n).$$  

(38)

Calculating the above probabilities now involves only scalar Gaussian densities, which significantly simplifies the analysis.
solving the optimization problem yields:
\[ E_{\text{opt}} \]

The optimal estimate in the coupled system, in which the function of the number of interaction rounds and unequal determination of the modulo-aliasing error probability. This factor corresponds to \( n_{\text{opt}} \), \( \Psi_3 \) in Theorem 1 where \( \Psi_3 \) is a reminder term obtained by pedestrian manipulations and the inequity \( -\ln(1-x) \leq \frac{x}{1-x} \) for \( x < 1 \). Specifically, the result in Theorem 1 was obtained for the specific choice \( p_m = \frac{p_n}{2N} \) of modulo-aliasing error. In general, reducing \( p_m \) decreases \( \lambda \) which in turn decreases \( SNR_N \), and hence increases the second addend on the right-hand-side of (45), resulting in a trade-off that could potentially be further optimized.

VII. NUMERICAL RESULTS

The behavior of the capacity gap for our scheme as a function of the number of interaction rounds and \( \Delta SNR \) is depicted in Fig. 3 and Fig. 4 for “high SNR” and “low SNR” setups. In both figures we plotted the capacity gap, for a target rate \( R \) and a target error probability \( p_e = 10^{-6} \), where the SNR corresponding to \( R \) was found by numeric search on (19), and the capacity gap calculated by definition (5). We can see that the higher \( \Delta SNR \), the smaller the capacity gap, where \( \Delta SNR = 30 \)dB is close to noiseless feedback. The points marked \( n_{\text{opt}} \) are those for which the capacity gap is less than 0.2dB above the minimal value attained. In Fig. 3 \( R = 1 \), and can see that \( \Delta SNR = 10 \)dB reduces the capacity gap to 4.2dB in 12 iterations, and \( \Delta SNR = 20 \)dB reduces the capacity gap to 1.1dB in 22 iterations. In Fig. 4 \( R = 4 \) and for \( \Delta SNR = 10 \)dB the capacity gap to 3.5dB in 11 iterations, and \( \Delta SNR = 20 \)dB reduces the capacity gap to 0.8dB in 19 iterations. Observing (31) we can see that for high SNR the result is only a function of \( \Delta SNR \), thus does not depend on the target rate or the base SNR.

VIII. NOTES ON IMPLEMENTATION

The scheme described in this paper is simple and practical, as opposed to its noiseless feedback counterparts. This provides impetus for further discussing implementation related
aspects. The following conditions should be met for our results to carry merit: 1) Information asymmetry: Terminal A has substantially more information to convey than Terminal B; 2) SNR asymmetry: The SNR of the feedforward channel is lower than the SNR of the feedback channel. This can happen due to differences in power constraints and/or path losses; 3) Complexity/delay constraints: There are severe complexity or delay constraint at Terminal A; 4) Two-way signaling: Our scheme assumes sample-wise feedback. The communication system should therefore be full duplex where both terminals have virtually the same signaling rate. This implies that both terminals share the same bandwidth, but cannot divide it according to their information rates. This condition can be inherent to the system, or alternatively can be partially mitigated by standard power–bandwidth tradeoffs.

The use of a very large PAM constellations, whose size is exponential in the product of rate and interaction rounds, seemingly requires extremely low noise and distortion at the digital and analog circuits in Terminal A, which may appear to impose a major implementation obstacle. Fortunately, this is not the case. The full resolution implied by the size of the PAM constellation is by construction confined only to original message Θ and the final estimate ΘN; the transmitted and received signals in the course of interaction, can be safely quantized at a resolution determined only by the channel noise (and not the final estimation noise), as in commonplace communication systems. Figuratively speaking, the source bits are revealed along the interaction process, where every interaction step reveals a certain amount of source bits, determined by the channel SNR. This desirable property has also been confirmed in simulations.

Another important implementation issue is sensitivity to model assumptions. We have successfully verified the robustness of the proposed scheme in several reasonable scenarios including correlative noise, excess quantization noise, and multiplicative channel estimation noise. The universality of the scheme and its performance for a wider range of models remains to be further investigated.

IX. DISCUSSION

Note that so far we have limited our discussion to the PAM symbol error rate pe. The bit-error rate is in fact lower, since an error in PAM decoding affects only a single bit with high probability (12), assuming Gray labeling. However, note that the modulo-aliasing error will typically result in many erroneous bits, and hence optimizing the bit error rate does not yield a major improvement over its upper bound pe. Further fine-tuning of the scheme can be obtained by non-uniform power allocation over interaction rounds in both Terminal A and B; in particular, we note that Terminal B is silent in the last round, which can be trivially leveraged. We also note in passing that our scheme can be used in conjunction with FEC as an outer code, to achieve other power/delay/complexity/error probability tradeoffs.

We note again that for any choice of SNR and ΔSNR, the error probability attained by our scheme cannot be made to vanish with the number interaction rounds while maintaining a non-zero rate, as in the noiseless feedback S-K scheme case. The reason is that (40) implies a minimal attainable error probability dictated by the modulo-aliasing incurred by feedback noise. Equivalently, one cannot get arbitrarily close to capacity for a given target error probability; the reason is that while increasing the number of iterations would increase SNR, and reduce the PAM decoding error term in (45), it would also increase the modulo-aliasing error term in (45).

Hence, our scheme is not capacity achieving in the usual sense. However, it can get close to capacity in the sense of reducing the capacity gap using a very short block length, typically N ≈ 20 in the examples presented. To the best of our knowledge, FEC schemes require a block length typically larger by two order of magnitudes to reach the same gap at the same error probability. Consequently, the encoding delay of our scheme is also markedly lower than that of competing FEC schemes. Alternatively, compared to a minimal delay uncoded system under the same error probability, our scheme operates at a much lower capacity gap for a wide regime of settings, and hence can be significantly more power efficient.

Another important issue is that of encoding and decoding complexity. Our proposed scheme applies only a two multiplications and one modulo operation at each terminal in each interaction round. This is markedly lower than the encoding/decoding complexity of FEC, even if non-optimal methods such as iterative decoding are employed.

REFERENCES

[1] C. E. Shannon, “The zero-error capacity of a noisy channel,” *IEEE Trans. Inf. Theory*, vol. IT-2, pp. 8–19, Sep 1956.
[2] J. P. M. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback part I: No bandwidth constraint,” *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 172–182, Apr 1966.
[3] J. P. M. Schalkwijk, “A coding scheme for additive noise channels with feedback part II: Band-limited signals,” *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 183–189, Apr 1966.
[4] M. Horstein, “Sequential transmission using noiseless feedback,” *IEEE Trans. Inf. Theory*, vol. IT-9, pp. 136–143, Jul 1963.
[5] O. Shayevitz and M. Feder, “Optimal feedback communication via posterior matching,” *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar 2011.
[6] Y.-H. Kim, A. Lapidoth, and T. Weissman, “On the reliability of Gaussian channels with noisy feedback,” in *Proc. 41st Allerton Conf. Communication, Control Computing*, Sep. 2006, pp. 364–371.
[7] Z. Chance and D. J. Love, “Concatenated coding for the AWGN channel with noisy feedback,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 6633–6649, Oct. 2011.
[8] A. Sato and H. Yamamoto, “Error exponents of discrete memoryless channels and awgn channels with noisy feedback,” in *ISITA*, 2010, pp. 452–457.
[9] M. V. Burnashev and H. Yamamoto, “On reliability function of BSC with noisy feedback,” *Problems of Information Transmission*, vol. 46, pp. 2–23, 2010.
[10] M. V. Burnashev and H. Yamamoto, “Noisy feedback improves the gaussian channel reliability function,” in *ISIT*, 2014, pp. 2554–2558.
[11] M. Tomlinson, “New automatic equalizer employing modulo arithmetic,” *Electronics Letters*, vol. 7, no. 5, pp. 138–139, 1971.
[12] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference,” *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, 1972.