Communicating Using Feedback over a Binary Channel with Arbitrary Noise Sequence

Ofer Shayevitz and Meir Feder
Tel Aviv University, Dept. of EE-Systems
Tel Aviv 69978, Israel
Email: {ofersha,meir}@eng.tau.ac.il

***Abstract*— Communications over a binary channel with an additive (modulo 2) individual noise sequence and a full causal feedback link is explored. A randomized sequential transmission scheme that adapts its rate to the individual noise realization is presented. The decoding rate is analyzed for a special case, and shown to asymptotically approach $1 - h_b(p_{emp})$ with a vanishing probability of error w.r.t. the scheme's randomization, where $h_b(p_{emp})$ is the empirical entropy of the noise sequence. Therefore, while the classical capacity of this channel is zero, information may be reliably transmitted over the channel by not committing to a rate in advance, but rather decoding at a rate dictated by the realized noise sequence.**

I. Introduction

The capacity of a channel is classically defined as the supremum of all rates for which communication with arbitrarily low probability of error can be guaranteed in advance. However, when a feedback link between the receiver and the transmitter exist, one does not necessarily have to commit to a rate prior to transmission, and transmission can take place using some sequential scheme at a variable rate determined by the specific realization of the channel, thus the better the channel realization the higher the rate of transmission. When the channel law is known such an approach cannot achieve average rates exceeding those attainable by constant rate feedback schemes, but it may have a better error exponent and a lower complexity. Several sequential transmission schemes possessing those merits were proposed for the binary symmetric channel (BSC) [1][2], the Gaussian additive noise channel [3], discrete memoryless channels [4] and finite-state channels (FSC) [5].

When the channel law is unknown to some degree, the variable rate sequential schemes become even more attractive. In [6], universal variable rate transmission schemes for compound BSC and Z-channel with feedback were described, and shown to attain any fraction of the realized channel capacity and achieve the Burnashev error exponent. In [7] it was shown that for compound FSC with feedback, it is possible to transmit at a rate approaching the mutual information of the realized channel for any input distribution, by iteratively compressing the noise sequence via Lempel-Ziv coding.

The goal of this paper is to investigate an extreme case of channel law ignorance and explore the possible merits of a feedback link in that scenario. We will consider a binary-input binary-output channel where a bit either retains its value or is inverted in a deterministic but unknown fashion, independent of the channel's input. The channel is thus essentially a modulo-2 additive noise channel with an individual noise sequence. It is also assumed that a causal feedback link between the receiver and the transmitter exists. This setting is also equivalent to an arbitrarily varying channel (AVC) [8] with feedback and two states, one a clean channel and the other an inverting channel. It is well known that this channel's classical capacity is zero [9]. This result stems from the fact that the worst sequences in this setting are half zeros half ones. However, it may be possible to take advantage of situations where the noise sequence has a different composition, at the cost of not committing to a rate in advance.

This paper introduces a communication scheme that utilizes the empirical statistics of the noise sequence gathered by the transmitter to facilitate reliable communications. The paper further provides initial analysis indicating that the scheme's rate may approach the empirical capacity of the realized channel given by $1 - h_b(p_{emp})$, where p_{emp} is the empirical distribution of the noise sequence, and $h_b(\cdot)$ is the binary entropy function.

II. The Transmission Scheme

The channel model considered in this paper is given by

$$y_k = x_k + z_k \tag{1}$$

where k is the time index, x_k and y_k are the input and output sequences and z_k is the noise sequence, which is assumed to be an individual sequence. The + sign stands for modulo-2 addition. We assume that a noiseless feedback link exists, so that the transmitter at time k knows the received sequence up to time $k-1$.

We shall use a transmission scheme based on the one proposed by Horstein [1] for the BSC. In that scheme, the (infinite) message bitstream is first represented as a point $\theta \in [0, 1)$, which is referred to as the *message point*. A zero or one is transmitted according to whether the message point θ is to the left or to the right of the interval's midpoint. The receiver then calculates the a-posteriori probability measure of the message point position over the interval, given the received bit, and finds the median point which is a point with equal probability to the left and to the right of it. The next transmitted bit is zero or one according to whether the message point is to the left or to the right of the median, and so on. A bit is decoded whenever its corresponding interval has accumulated a probability greater than $1 - p_e$, where p_e is some prescribed probability of bit error. This scheme asymptotically achieves the capacity of the BSC [1]. In this work we generalize the scheme to the case where the noise sequence is an individual sequence.

Our modified scheme is now described. Since we assume no statistical model for the noise sequence, we let the crossover probability used by the receiver to calculate the so called a-posteriori probability measure, to vary with time. Specifically, denoting by $f_k(\theta)$ and a_k the a-posteriori probability measure at the receiver and the median point at time k (after y_k was received) respectively, we set $f_0(\theta)$ to be uniform over the interval $\theta \in [0, 1)$ and set $a_0 = \frac{1}{2}$. We then have

$$f_{k+1}(\theta) = \begin{cases} 2(p_k y_{k+1} + q_k(1 - y_{k+1}))f_k(\theta) & \theta < a_k \\ 2(q_k y_{k+1} + p_k(1 - y_{k+1}))f_k(\theta) & \theta > a_k \end{cases} \tag{2}$$

where $p_k, q_k \stackrel{\triangle}{=} 1 - p_k$ are determined after receiving y_k, and are thus generally a function of y_1^k. Later in the paper we suggest a method for updating p_k that can be implemented simultaneously at both ends. For the BSC with crossover probability p, fixing $p_k = p$ results in the Horstein scheme as a special case. We refer to $f_k(\theta)$ as a weight density function rather than a probability measure, since in the individual sequence setting the latter does not have the regular meaning. We assume that decision is made only at the end of transmission, after n bits have been received. When transmission ends, we look for the smallest interval of the form $[m2^{-\ell}, (m + 1)2^{-\ell}]$ that has a total weight larger than $1 - \epsilon$ for some decoding threshold $0 < \epsilon \leq \frac{1}{2}$. We then decode the ℓ bits representing that interval.

III. Decoding Rates

Let n_0 and n_1 be the number of zeros and ones respectively in the observed noise sequence of length n and assume throughout this paper, without loss of generality, that $n_0 > n_1$. The ratio $p_{emp} = \frac{n_1}{n}$ will be referred to as the *empirical distribution* of the noise sequence. It was shown [10] that when the maximal value of p_{emp} is known in advance to both the transmitter and the receiver, the achievable rate R is upper bounded by the hamming bound $1 - h_b(p_{emp})$, and a straight line through $(R, p_{emp}) = (0, \frac{1}{3})$ tangent to the hamming bound. This implies, for one, that even in this much simpler setting, error free communication at positive rates is impossible whenever $p_{emp} > \frac{1}{3}$ is allowed.

However, for a BSC with a crossover probability equal to p_{emp} it is well known that a rate of $1 - h_b(p_{emp})$ is attainable. The source of this rate penalty lies in the fact that for almost every message point there exists a small set of admissible bad noise sequences that cause the receiver to err. Nevertheless, it is strongly believed that this gap can be closed, since the analysis in [10] does not take into account transmission strategies where errors are deliberately inserted by the transmitter. The idea is that since the transmitter knows the message point, it knows the bad noise sequences that correspond to it, and as the number of these bad sequences becomes relatively negligible with the block size increased, the transmitter should be able to avoid them with high probability by inserting a small number errors in a randomized manner.

The analysis below first considers the case of constant p_k agreed by the transmitter and receiver. We find the decoding rate attainable when the message point is $\theta = 0$. From [10] we know that this rate cannot be attained for general θ, but we nevertheless conjecture that it can be attained using proper randomization. Next we turn to the main goal of the paper. We start by considering the case where p_k varies according to a sequential estimate (e.g. Laplace estimate) of the empirical noise distribution. Clearly in this case the transmission scheme is not feasible as the receiver cannot know that estimate. In addition we derive the decoding rate for $\theta = 0$, but here the extension to any θ seems to be straightforward. We then observe that the decoding rate obtained by the above scheme that consecutively updates p_k, can be obtained even when p_k is updated once every $b = o(\frac{n}{\log n})$ channel uses. Following this, we introduce a *feasible* randomized scheme where the update information is communicated to the receiver at random locations, allowing the scheme to obtain the optimal rate of $1-h_b(p_{emp})$, with an error probability, w.r.t the randomization, that tends to zero.

A. Constant Crossover Probability

We first study the case where p_k is constant, agreed upon by the transmitter and receiver before transmission starts. The following Theorem provides a lower bound on the number of successfully decoded bits (for $\theta = 0$) in this case, together with the corresponding asymptotic decoding rate.

Theorem 1: For $p_k = p$ and any empirical distribution $p_{emp} = \frac{n_1}{n}$ of the noise sequence, the number of successfully decoded bits is lower bounded by

$$L \geq \left\lfloor n + n_0 \log q + n_1 \log p + 1 - \frac{\log 2p}{\log 2q} \log \epsilon \right\rfloor^{+} \tag{3}$$

which correspond to an asymptotic rate of

$$R \geq |\ 1 - h_b(p_{emp}) - D(p_{emp} \,\|\, p)\ |^{+} \tag{4}$$

Moreover, the bound (3) corresponds to the number of most significant bits in the message point representation successfully decoded.

Proof: We assume without loss of generality that $p < \frac{1}{2}$. Let w_k be the weight of the interval $I_\ell \stackrel{\triangle}{=} [0, 2^{-\ell}]$ at time k (after y_k was received). The weight w_{k+1} is given by:

$$w_{k+1} = \begin{cases} 2qw_k & z_{k+1} = 0\,, a_k \notin I_\ell \\ 2pw_k + (1 - 2p) & z_{k+1} = 0\,, a_k \in I_\ell \\ 2pw_k & z_{k+1} = 1\,, a_k \notin I_\ell \\ 2qw_k + (1 - 2q) & z_{k+1} = 1\,, a_k \in I_\ell \end{cases}$$

The accumulated weight of I_ℓ can be calculated using this recursive relation. Notice that the median $a_k \in I_\ell$ if and only if $w_k > \frac{1}{2}$, and so for a fixed z_{k+1}, the weight w_{k+1} is monotonically increasing as a function of w_k.

We shall first show that given any noise sequence with the desired composition, changing the order of a zero followed by a one cannot increase the final weight w_n. Assume that $z_k = 0$ and $z_{k+1} = 1$. To prove our claim, it is sufficient to show that exchanging the order of z_k and z_{k+1} does not increase the weight w_{k+1}, as this results in a non-increasing w_n as well. To this end, we go through the possible cases: First, assume that the median is outside the interval for both the original z_k and z_{k+1}, and so the weight gain introduced is $4pq$. Since we assume $p < \frac{1}{2}$, the exchange keeps the median outside the interval, and so the gain remains $4pq$. Second, assume the median a_{k-1} is outside the interval, but a_k is inside it. In this case we have

$$w_{k+1} = 4q^2 w_{k-1} + (1-2q) \tag{5}$$

Exchanging the order, we have that both medians are outside the interval and so $w_{k+1} = 4pqw_{k-1}$ which is smaller than (5), provided that $w_{k-1} > \frac{1}{4q}$. But this is always fulfilled, since assuming a_k is inside the interval implies $2qw_{k-1} > \frac{1}{2}$. Finally, assume that both medians are inside the interval. In this case we have

$$w_{k+1} = 2q(2pw_{k-1} + (1-2p)) + (1-2q) \tag{6}$$

Exchanging the order, we either have that both medians remain inside the interval which does not affect w_{k+1}, or that a_k is outside the interval that gives

$$w_{k+1} = 2p(2pw_{k-1} + (1-2p)) \tag{7}$$

which is smaller than (6), provided that $w_{k-1} > 1 - \frac{1}{4p}$. But this is always fulfilled, since assuming a_k is inside the interval implies $2pw_{k-1} + (1-2p) > \frac{1}{2}$.

We can therefore conclude that the final weight w_n is lower bounded by the weight of a noise sequence of the same composition, having all-ones followed by all zeros, which we denote by $\hat{w}_n$. For that sequence we have $\hat{w}_{n_1} = 2^{-\ell}(2p)^{n_1}$, and as long as the medians do not fall inside the interval, we can also write

$$\hat{w}_{n_1+m} = 2^{-\ell}(2p)^{n_1}(2q)^m \tag{8}$$

The maximal $m = \hat{m}$ for which (8) still holds can be found by comparing $\hat{w}_{n_1+m}$ with $\frac{1}{2}$, which results in

$$\hat{m} = \frac{\ell - 1 - n_1 \log 2p}{\log 2q} \tag{9}$$

so the final weight of the worst sequence is given by

$$\begin{aligned} \hat{w}_n &= 2^{-\ell}(2p)^{n_1}(2q)^{\hat{m}}(2p)^{n_0-\hat{m}} + (1-2p)\sum_{k=0}^{n_0-\hat{m}-1}(2p)^k \\ &= 2^{-\ell}(2p)^{n-\hat{m}}(2q)^{\hat{m}} + 1 - (2p)^{n_0-\hat{m}} \geq 1 - (2p)^{n_0-\hat{m}} \end{aligned}$$

and the final weight for any noise sequence of the same composition is therefore lower bounded by

$$w_n \geq 1 - (2p)^{n_0-\hat{m}} \tag{10}$$

Now, if the lower bound above is greater than $1-\epsilon$, then the decoded interval is guaranteed to lie within the interval I_ℓ. Comparing (10) to $1-\epsilon$ and substituting (9) we get that

$$\ell \leq n_0 \log 2q + n_1 \log 2p + 1 - \frac{\log 2q}{\log 2p}\log\epsilon \tag{11}$$

The maximal ℓ for which this holds is therefore a lower bound on the number of correctly decoded bits, which gives us the desired result. The decoded interval always resides within I_ℓ for any ℓ satisfying the above, thus the bits corresponding to the lower bound are the most significant bits in the message point representation. ∎

Although it is believed that the rates (4) can be attained for all message points by introducing randomization at the transmitter, it is also obvious that the constant crossover probability scheme is not suitable for the individual noise sequence scenario, as it fails to track the empirical statistics of the noise sequence.

B. *Consecutively Varying Crossover Probability*

We now turn to the case where the crossover probabilities are allowed to vary with time. At first, we shall assume that p_k is set to be the Laplace probability estimator:

$$p_k = \frac{\sum_{j=1}^{k} z_j + 1}{k+2} \tag{12}$$

Although this assignment is not possible in practice, since it requires the receiver to be familiar with the past noise sequence, it is still insightful to analyze this case, and a more practical approach will be developed in the sequel, based on this analysis. For the rest of the paper, we set the decoding threshold to be $\epsilon = \frac{1}{2}$, as this simplifies the exposition and does not affect our results in the individual sequence setting, since the noise composition will always be known accurately enough to the receiver at the end of transmission.

Theorem 2: For any empirical distribution $p_{emp} = \frac{n_1}{n}$ of the noise sequence, the number of successfully decoded bits using the Laplace estimator is lower bounded by

$$L \geq n + n_0 \log\frac{n_0}{n} + n_1 \log\frac{n_1}{n} - \delta_1(n) \triangleq L_0 \tag{13}$$

which corresponds to a decoding rate of

$$R \geq 1 - h_b(p_{emp}) - \xi_1(n) \tag{14}$$

where

$$\delta_1(n) = O(\log n)\,, \quad \xi_1(n) = O(\frac{\log n}{n}) \tag{15}$$

Furthermore, L_0 is computable at the receiver and thus the lower bound rate can be attained with zero probability of error.

Proof: Again we limit our discussion to $\theta = 0$, though in this case generalization to an arbitrary θ seems to follow through. It can be shown, by arguments in the spirit of those used in the proof of Theorem 1, that the noise sequence resulting in the minimal final weight of the interval I_ℓ using the Laplace estimator, is the sequence

$$z_1^n = (\overbrace{1,0,1,0,\ldots,1,0}^{2n_1},\overbrace{0,0,\ldots,0}^{n0-n1}) \tag{16}$$

For this sequence the last median point is also the one closest to the message point, and so the interval $[0, a_n]$ has a constant weight density. It is easy to show that $a_n < \frac{1}{2}$, and therefore since $\epsilon = \frac{1}{2}$, the decoded interval is always contained in the interval I_ℓ, where ℓ is the largest such that $2^{-\ell} > a_n$ is satisfied. Thus, the number of bits guaranteed to be correctly decoded is bounded by $L \geq \lfloor -\log a_n \rfloor$. To complete the proof, it remains to find a_n.

It is can be verified that the weight density in the interval $[0, a_n]$ is equal to the probability assigned to the noise sequence by the Laplace estimator, multiplied by 2^n. It is also a known fact that this probability is uniquely determined by the composition of the sequence. Therefore, in order to find a_n it is sufficient to find the probability assigned to any noise sequence with the same composition. Specifically, we shall consider the sequence of all zeros followed by all ones.

The probability assigned by the Laplace estimator to that sequence is given by

$$\begin{array}{lll} q_k = & \frac{k+1}{k+2} & k = 0, \dots n_0 - 1 \\ p_k = & \frac{k-n_0+1}{k+2} & k = n_0, \dots n-1 \end{array}$$

and thus the log-probability assigned to the sequence is

$$\begin{aligned} \log(p(z_1^n)) &= \sum_{k=0}^{n_0-1} \log \frac{(k+1)}{k+2} + \sum_{k=n_0}^{n-1} \log \frac{(k-n_0+1)}{k+2} \\ &= \log \frac{1}{n_0+1} + \log \frac{n_1!(n_0+1)!}{(n+1)!} \\ &\geq \log \left(\sqrt{\frac{2\pi pq}{n}} (np)^{np} (nq)^{nq} n^{-n} \right) - \log(n+1) \\ &= np \log p + nq \log q - \frac{1}{2} \log \left(\frac{n(n+1)^2}{2\pi pq} \right) \qquad (17) \end{aligned}$$

where subscripts were omitted for clarity, by denoting the empirical distribution p_{emp} as p and defining $q \triangleq 1 - p$. The inequality transition makes use of the Stirling bound.

Since a_n is the last median, it satisfies $a_n \cdot 2^n p(z_1^n) = \frac{1}{2}$. Combining this with (17), the number of bits guaranteed to be correctly decoded is bounded by

$$L \geq n + np \log p + nq \log q - \frac{1}{2} \log \left(\frac{n(n+1)^2}{2\pi pq} \right) \triangleq L_0 \quad (18)$$

and by substituting $p = \frac{n_1}{n}, q = \frac{n_0}{n}$ we have the desired result, where the rate $R = L/n$ is bounded correspondingly. Using the (impractical) Laplace estimator, the receiver knows the noise sequence composition at the end of transmission, and can always correctly decode precisely L_0 bits, thus avoiding from decoding extra bits that may be erroneous, and attaining the lower bound rate with zero probability of error. ∎

C. Block Varying Crossover Probability

Since using the Laplace estimator amounts to assuming that the receiver sequentially knows the noise sequence, we shall use a block estimator instead, meaning that the transition probability estimator is updated once per some channel uses, and retains its value otherwise. As this still requires some knowledge of the noise sequence at the receiver, we assume for now that an auxiliary channel from the transmitter to the receiver exists, and that the number of ones that appeared in the noise sequence since the last update is reliably relayed to the receiver. We analyze the penalty in decoding rates caused by a slower update, and find the optimal balance between the rate of the auxiliary channel and the loss in rate of the data channel. We assume that during transmission, the estimator used by the receiver is no more than b channel uses old, and refer to b as the *maximal block length* of our scheme.

Theorem 3: For any empirical distribution $p_{emp} = \frac{n_1}{n}$ of the noise sequence, the number of successfully decoded bits using the block estimator is lower bounded by

$$L \geq n + n_0 \log \frac{n_0}{n} + n_1 \log \frac{n_1}{n} - \delta_2(b, n) \triangleq L_1 \qquad (19)$$

which corresponds to a rate of

$$R \geq 1 - h_b(p_{emp}) - \xi_2(b, n) \qquad (20)$$

where

$$\delta_2(b, n) = O(b \log n), \quad \xi_2(b, n) = O\left(\frac{b \log n}{n}\right) \qquad (21)$$

and b is the maximal block length. Furthermore, L_1 is computable at the receiver and thus the lower bound rate can be attained with zero probability of error.

Proof: Denote by n_k^0, n_k^1 the position within the noise sequence of the k-th zero and k-th one, respectively. The difference between the log-probability $\log p(z_1^k)$ assigned to the noise sequence by the Laplace estimator and $\log p_b(z_1^k)$ assigned by our block estimator can be bounded as follows:

$$\begin{aligned} &\log p_b(z_1^k) - \log p(z_1^k) \ \geq \\ \geq\ & C + \sum_{k=b+1}^{n_0} \log \left(\frac{k-b}{n_k^0+1} \cdot \frac{n_k^0+1}{k} \right) + \\ & + \sum_{k=b+1}^{n_1} \log \left(\frac{k-b}{n_k^1+1} \cdot \frac{n_k^1+1}{k} \right) \\ \geq\ & C + \sum_{k=1}^{n} \log \left(1 - \frac{b}{k} \right) \\ \geq\ & C - b \sum_{k=1}^{n} \frac{1}{k} \ \geq C - b(1 + \log n) \cdot ln2 \qquad (22) \end{aligned}$$

where C stems from the estimation difference over the first b bits, and can be bounded by $C \geq b \log \left(\frac{b+1}{2b} \right) \geq -b$.

It can be shown that the sequence attaining the minimal final weight of the interval I_ℓ for the Laplace estimator, also asymptotically attains the minimal weight for the block estimator. Thus, combining (22) with (17), and using again the fact that the receiver knows the composition at the end of transmission as in Theorem 2, we get the desired result. ∎

It is therefore seen that for a given p_{emp}, if $\frac{b \log n}{n} \to 0$ as n grows, any decoding rate below the first order empirical capacity of the channel can be attained with zero probability of error. The auxiliary channel transfers $\log b$ bits every b data channel uses, and so has a rate of $\frac{\log b}{b}$. The optimal trade off between rate penalty due to block estimation on the one hand and auxiliary rate on the other is attained for $b = O(\sqrt{n})$.

We are now ready to present the main result of the paper.

Theorem 4: For the channel (1) with feedback, and for any empirical distribution $p_{emp} = \frac{n_1}{n}$ of the noise sequence z_k, a decoding rate of

$$R \geq 1 - h_b(p_{emp}) - \xi_3(n) \tag{23}$$

can be attained by a randomized sequential transmission scheme with a probability of error that tends to zero, where

$$\xi_3(n) = O(\frac{\log n}{\sqrt[4]{n}}) \tag{24}$$

Sketch of the proof: The auxiliary channel is embedded into the data channel, using a constant crossover probability scheme. This is possible since the rate of the auxiliary channel can be made arbitrarily small. Divide the n channel uses into $\sqrt{n}$ blocks of equal length $\sqrt{n}$, and set some p_a to be used as the constant crossover probability used by the receiver to decode the embedded auxiliary channel. In each block, we dedicate a number of bits in the order of $\frac{\log n}{1-h_b(p_a)}$ to the auxiliary channel. Since choosing these bits in advance may generally result in a very bad channel, we randomize the positions of the bits in each block. To create a common randomness between the transmitter and the receiver we sent, at the beginning of each block, a number of bits in the order of $\frac{\log^2 n}{1-h_b(p_a)}$, describing random positions for the auxiliary channel bits in the block. We then make use of the feedback to learn the positions as appearing at the receiver. Since now the auxiliary channel uses random positions inside each block, it appears, for a large n, as a BSC with crossover probability close to the empirical distribution of the noise sequence in the block. Thus as long as this crossover probability takes its values in the interval $[0, p_a]$, the auxiliary channel data may be decoded at the end of the block, with arbitrarily low error probability.

To handle the case where the crossover probability lies in $[p_a, \frac{1}{2}]$, we add a cyclic redundancy check (CRC) to the auxiliary bits at every block, so this situation is identified with high probability. When that happens, the last block is disregarded and the scheme continues with the weights and the probability estimation of the last correctly received block. The situation where the empirical distribution is in the range of $[\frac{1}{2}, 1]$ is handled by creating a second auxiliary channel using a crossover probability of $1 - p_a$, and with a CRC in a manner similar to the described above. The CRC in both channels are also used to identify which of the channels is to be relied on in each block.

To assure convergence to the empirical capacity for any value of the empirical noise distribution p_{emp}, we now let p_a vary with the block length so that $p_a \rightarrow \frac{1}{2}$ as n grows. The way in which p_a tends to $\frac{1}{2}$ should be carefully set. The rate penalty due to the auxiliary channel is in the order of $\frac{\log^2 n}{(1-h_b(p_a))\sqrt{n}}$. Using Jensen's inequality, the rate penalty due to disregarding the blocks in which the empirical distribution is worse than p_a can be bounded by $1-h_b(p_a)$. Therefore, the penalties can be balanced by setting $1 - h_b(p_a) = O(\frac{\log n}{\sqrt[4]{n}})$, and we have the desired convergence rate. ■

The penalty may be reduced to $O(\sqrt{\frac{\log n}{\sqrt{n}}})$ by means of creating the common randomness for the auxiliary channels using the received data from the previous block in lieu of the randomly generated positions at the beginning of each block. Additionally, it may also be possible to modify this horizon-dependent scheme so to become horizon-free, by using an arithmetically growing block size so that for every n the effective block size will be in the order of $\sqrt{n}$.

IV. Summary and Future Research

A sequential transmission scheme for a binary channel with an individual noise sequence and a full causal feedback was introduced. The bit decoding rates achievable for the zero message point were determined, and were shown to approach $1 - h_b(p_{emp})$, i.e., the BSC capacity associated with a crossover probability p_{emp}. To complete the analysis, the extension to any message point is currently explored, and seems to follow through.

By employing a universal predictor at the receiver, it may be possible to cancel some of the effect of the noise sequence, so that its empirical distribution will approach its predictability $\pi(\boldsymbol{z})$ [11]. That way, the rates achievable on this channel may well approach $1 - h_b(\pi(\boldsymbol{z}))$. It may be even more interesting to generalize our scheme to consider higher order empirical statistics of the noise sequence so that the empirical Markovian capacity of a given order is approached, at the expense of a slower convergence. Hopefully, increasing the Markov order with time in an appropriate manner will enable us to approach a rate of $1 - \rho(\boldsymbol{z})$, where $\rho(\boldsymbol{z})$ is the Markovian or even the finite-state compressibility of the noise sequence.

References

[1] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Info. Theory*, pp. 136–143, July 1963.

[2] J. P. M. Schalkwijk and K. A. Post, "On the error probability for a class of binary recursive feedback strategies," *IEEE Trans. Info. Theory*, vol. IT-19, pp. 498–511, 1973.

[3] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–i: No bandwidth constraint," *IEEE Trans. Info. Theory*, vol. IT-12, pp. 172 – 182, 1966.

[4] M. Horstein, "Sequential transmission of digital information with feedback," *Technical report 375, Research Laboratory of Electronics, MIT*, 1960.

[5] J. M. Ooi and G. W. Wornell, "Fast iterative coding techniques for feedback channels," *IEEE Trans. Info. Theory*, vol. IT-44, pp. 2960–2976, 1998.

[6] A. Tchamkerten and I. E. Telatar, "Variable length coding over unknown channels," *submitted to the IEEE Trans. Info. Theory*, 2005.

[7] J. M. Ooi, "A framework for low-complexity communication over channels with feedback," *Doctoral Thesis*, 1998.

[8] D. Blackwell, L. Breiman, and A.J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Stat.*, vol. 31, pp. 558–567, 1960.

[9] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Info. Theory*, pp. 181–193, Mar. 1988.

[10] E. R. Berlekamp, "Block coding with noiseless feedback," *doctoral dissertation, MIT*, 1964.

[11] M. Feder, N. Merhav, and M. Gutman, "Universal prediction of individual sequences," *IEEE Trans. Info. Theory*, pp. 1258–1270, July 1992.