Using zero-rate feedback on binary additive channels with individual noise sequences

Krishnan Eswaran, Anand D. Sarwate, Anant Sahai, and Michael Gastpar

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720, USA
Email: {keswaran, asarwate, sahai, gastpar}@eecs.berkeley.edu

***Abstract*— Recently, Shayevitz and Feder introduced an individual sequence formulation of channel coding under model uncertainty and an elegant coding strategy that adapts Horstein's scheme to this setting to achieve the empirical capacity of the channel. Their scheme requires both full-rate output feedback and common randomness. We present a strategy in the style of Hybrid ARQ that requires no output feedback by using common randomness and zero-rate active feedback. This strategy still asymptotically achieves the empirical capacity.**

I. Introduction

Models for communication in time-varying environments often assume that the channel dynamics can be modeled, for example, as a finite-state machine or Markov chain. However, such models may not be appropriate in many situations. For example, in shared unlicensed frequency bands, a communication system may be ignorant of the coding and modulation schemes adopted by other systems, and these other systems may occupy the band intermittently. What design principles should a communications system employ to be robust to such uncertainty about its environment while exploiting as much rate as possible?

As a concrete example, consider a channel that takes binary inputs and produces binary outputs, where the output is produced by potentially flipping some bits of the channel input. If channel noise does not depend on the channel input symbol, then a convenient way to express the output $\mathbf{y} \in \{0,1\}^N$ is

$$\mathbf{y} = \mathbf{x} \oplus \mathbf{z}, \tag{1}$$

where $\mathbf{x} \in \{0,1\}^N$ is the channel input, $\mathbf{z} \in \{0,1\}^N$ is the noise sequence, and addition is carred out modulo-2. Let p be the empirical fraction of 1's in $\mathbf{z}$.

A simple model for $\mathbf{z}$ is the binary symmetric channel (BSC), in which $\mathbf{z}$ is i.i.d. with $\mathbb{P}(z_j = 1) = p^*$. The capacity for such channels is $1 - h(p^*)$. Another model is the arbitrarily varying channel (AVC) [1] in which $p \le p^*$, where p^* is known in advance and the noise sequence may be chosen adversarially with respect to the coding strategy. Codes for the AVC are designed to target the worst-case capacity $1 - h(p^*)$. When the noise is not adversarial and $p \ll p^*$, these coding strategies are not opportunistic and do not exploit the potential capacity of the channel.

In order to behave opportunistically, the encoder and decoder need to be able to develop a protocol to adjust the rate of transmission. In order to do this, there must be a feedback link from the decoder to the encoder. Feedback is a necessary component for any coding strategy attempting to achieve $R \approx 1 - h(p)$ without prior knowledge of p. Without feedback the encoder and decoder must choose a rate R in advance. This rate will be larger than $1 - h(p)$ for some p and by the strong converse to the noisy channel coding theorem, the decoding error will tend to 1. Full output feedback could allow the encoder to track the decoder's progress and thereby control the decoding time. When the channel is less noisy the decoder can terminate earlier, and when the channel is noisier it will take longer.

Shayevitz and Feder recognized that significant gains could come with feedback, and recently proposed a channel model and novel coding strategy. In their model, the encoder has an infinite supply of bits to transmit and the noise sequence $\mathbf{z}$ is fixed in advance but is otherwise arbitrary. We can think of their coding strategy as using full output feedback and common randomness to achieve the rate $1 - h(p)$ for all p and $\mathbf{z}$ [2]. They call this rate the *empirical capacity* of the channel, and we will continue to use this term. In particular, given a blocklength of N, the decoder chooses to decode the first NR bits of the message, where R is not fixed in advance. Their coding scheme guarantees that the NR decoded bits are equal to the first NR bits of the message with high probability for all noise sequences $\mathbf{z}$ as the blocklength $N \to \infty$.

In the model proposed by Shayevitz and Feder [2], the encoder has access to (passive) output feedback from the decoder that allows the encoder to provide control and estimation information in a set of training sequences that can be selected via common randomness. Because the information provided in the feedback is so rich, they can create a scheme that is horizon-free and can be terminated at any time. The limited common randomness required for training sequence positions can also be determined via active feedback. However, in the wireless communication systems motivated earlier, the encoder typically does not have a way to access the full output feedback of the decoder.

In light of this, we consider only limited active feedback from the decoder to encoder along with common randomness.

To do this, we adapt the feedback-reducing block/chunk strategies used earlier in the context of reliability functions [3], [4] and most specifically in [5]. They are in turn inspired by Hybrid ARQ [6]. The flavor of our algorithm is different – in our scheme the decoder uses the feedback link to terminate rounds that are too noisy but otherwise attempts to correct the error in less noisy rounds. By doing away with the output feedback, we lose some of the simplicity of the scheme in [2], but we show that a similar performance can still be obtained with almost negligible feedback.

In Section II, the coding strategy is described in detail along with a statement of the main result. For this strategy, Section III bounds the error probability, rate loss, and sets parameters to complete the proof of Theorem 1. Section IV concludes the paper with some discussion and future directions.

II. The coding strategy

We divide the blocklength N into *chunks* of length $b = b(N)$. The encoder attempts to send $k = k(N)$ bits over several chunks comprising a *round*. Prior to transmission, the decoder and encoder use common randomness to choose $n_1 = n_1(N)$ *training positions* T_n for chunk n. The encoder and decoder will also choose a random codebook for each round. In a round, the encoder divides the codebook into segments of length $b-n_1$ and transmits the n-th segment over the $b-n_1$ non-training positions in chunk n. Dependence of the parameters on N will be suppressed throughout.

The decoder uses the outputs in T_n to estimate the empirical noise distribution in chunk n. After each chunk the decoder will either (a) decide to decode the k bits and tell the encoder to terminate the round, (b) decide that the empirical noise is too bad and tell the encoder to terminate the round and start over, or (c) decide that it cannot decode yet and tell the encoder to send another chunk.

A more formal description of the scheme follows, and an illustration is provided in Figure 1. For each round, the following steps are repeated for each chunk:

1) The encoder sends chunk n of length b. For $j \in T_n$, it sends $x_j = 0$ and for $j \notin T_n$ the bits $\{x_j : j \notin T_i\}$ are the next $b-n_1$ entries in the code for the current round.
2) The decoder estimates $p^{(n)}$, the empirical frequency of 1's outside of T_n, and $\bar{p}^{(n)}$, the average over chunks in the round:
$$q^{(n)} = \frac{1}{n_1} \sum_{j \in T_n} z_j \qquad \bar{q}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} q^{(i)} \ , \tag{2}$$
where $c_n = \{b(n-1)+1, \ldots, bn\}$ and
$$p^{(n)} = \frac{1}{b} \sum_{j \in c_n} z_j \qquad \bar{p}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} p^{(i)} \ . \tag{3}$$
3) The decoder makes a decision based on $\bar{q}^{(n)}$ and n:
   a) if
$$\bar{q}^{(n)} \in \left[\frac{1}{2} - \tau(N), \frac{1}{2} + \tau(N)\right] \tag{4}$$
   then the decoder feeds back "BAD NOISE."
   b) if
$$\frac{k}{(b-n_1)\cdot n} < 1 - h\left(\bar{q}^{(n)}\right) - \epsilon_1(N) \tag{5}$$
   then the decoder decodes and feeds back "DECODED."
   c) otherwise the decoder feeds back "KEEP GOING" and goes to 1).

Our main result states that this strategy asymptotically approaches the empirical capacity with vanishing error probability.

Theorem 1: When used over a binary channel whose noise sequence has pN ones, with probability $1-\delta(N)$ the algorithm below achieves a rate $1 - h_b(p) - \beta(N)$ for $p \notin [1/2 - \tau(N), 1/2 + \tau(N)]$, where $\delta(N) \to 0$, $\beta(N) \to 0$, $\tau(N) \to 0$ as $N \to \infty$.

III. Analysis

The analysis of our strategy, carried out below, consists of two parts. In the first part we show that the training positions provide a good estimate of the empirical noise frequency (Lemma 2) and that the condition in (5) is sufficient to decode the k bits for a round with small probability of error (Lemma 3). The second part of the analysis shows that the loss in rate from our scheme becomes negligible as the blocklength increases. The rate loss within a round is small (Lemma 5). The overall rate loss across rounds is also small (Lemma 6). We show that all of the bounds can be satisfied by setting the parameters at the end of this section.

A. Error analysis

We will declare an error if one of following two events occurs:

1) (E_1) We declare an error if
$$\left|\bar{q}^{(M)} - \bar{p}^{(M)}\right| > \epsilon_2(N). \tag{6}$$
2) (E_2) We will have an error if we decode to the wrong codeword.

We must first find a bound on the length of a round in chunks. Let
$$M = \inf_{n>0} \left\{ \bar{q}^{(n)} \in \left[\frac{1}{2} - \tau, \frac{1}{2} + \tau\right] \right. \\ \left. \text{or } \frac{k}{(b-n_1)n} < 1 - h\left(\bar{q}^{(n)}\right) - \epsilon_1 \right\} \tag{7}$$
We now argue that M cannot be too large.

Lemma 1 (Bounds on M): We have $M \leq M^*$, where
$$M^* := \left\lceil \frac{k}{(b-n_1)\cdot\left(1 - h(\frac{1}{2}-\tau) - \epsilon_1\right)} \right\rceil \ . \tag{8}$$
If the decoder attempted to decode, then $M \geq M_*$, where
$$M_* = \left\lceil \frac{k}{b-n_1} \right\rceil \ . \tag{9}$$

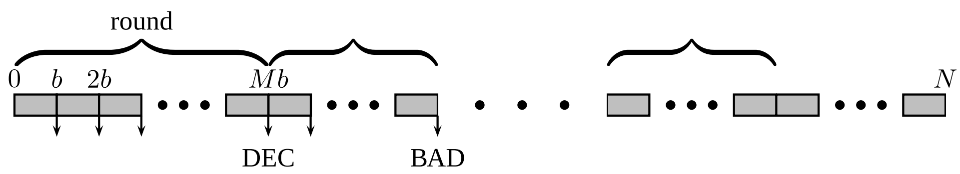

Fig. 1. After each chunk of length b feedback can be sent. Rounds end by decoding a message or declaring the noise to be bad.

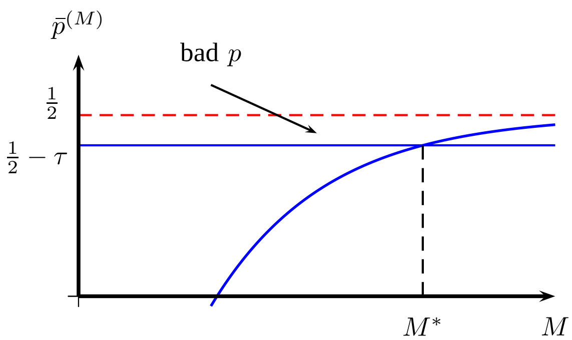

Fig. 2. Curve illustrating why M is finite.

Proof: The argument is illustrated in Figure 2. We must simply find the point where the curve defined by (5) intersects the "BAD NOISE" threshold. This gives the bound in (8). The lower bound is trivial from the definition in (5). ■

Lemma 2 (Bound on E_1): We have

$$\mathbb{P}(E_1) \leq \frac{2N}{b} \exp\left(-2n_1 \cdot \epsilon_2^2\right) \ . \tag{10}$$

Proof: We will first prove that our estimate of $\bar{p}^{(n)}$ improves with n. We can view each $q^{(i)}$ as an independent random variable with mean $p^{(i)}$. By Hoeffding's inequality [7]:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} q^{(i)} - \frac{1}{n}\sum_{i=1}^{n} p^{(i)}\right| > \epsilon_2\right) \leq 2\exp(-2 \cdot n \cdot \epsilon_2^2) \ . \tag{11}$$

Because we can terminate a round after any number of chunks, it is sufficient to bound the probability of E_1 only for a single chunk.

Another result of Hoeffding [7, Theorem 4] states that the exponential inequalities for sampling with replacement hold for sampling without replacement as well, so $q^{(i)}$ converges to $p^{(i)}$ and

$$\mathbb{P}\left(\left|q^{(i)} - p^{(i)}\right| > \epsilon_2\right) \leq 2\exp\left(-2 \cdot n_1 \cdot \epsilon_2^2\right) \ . \tag{12}$$

Taking a union bound over all N/b chunks, we obtain (10). ■

Lemma 3 (Bound on E_2): Assume $\tau > \epsilon_2$ and $\bar{q}^{(M)} \notin [1/2 - \tau, 1/2 + \tau]$, and let $\epsilon_3(N) = h(\frac{b}{b-n_1}\epsilon_2)$ Then

$$\mathbb{P}(E_2 | E_1^c) \leq \exp\left(-k\left(\frac{(\epsilon_1 - \epsilon_3 - M^*/k)^2}{8/e^2 + 4(\ln 2)^2}\right)\right) \ . \tag{13}$$

Proof: Let us define the empirical noise in the non-training positions by

$$\tilde{p}^{(M)} = \frac{1}{M(b-n_1)} \sum_{n=1}^{M} \sum_{j \notin T_i} z_j \ . \tag{14}$$

Then $M(b-n_1)\tilde{p}^{(M)} + Mn_1\bar{q}^{(M)} = M\bar{p}^{(M)}$. Then E_1^c implies

$$\left|M(b-n_1)\bar{q}^{(M)} - \sum_{n=1}^{M} \sum_{j \notin T_i} z_j\right| < Mb\epsilon_2 \ . \tag{15}$$

So

$$\left|\bar{q}^{(M)} - \tilde{p}^{(M)}\right| < \frac{b}{b-n_1}\epsilon_2 \ . \tag{16}$$

Thus under E_1^c we know that the empirical noise frequency in the non-training positions is also close to $\bar{q}^{(M)}$.

Since after every chunk the decoder must decide whether to decode, the actual rates that can be realized in a round fall in the discrete set $\{k/M(b-n_1) : M_* \leq M \leq M^*\}$. Under E_1^c we have from Lemma 7 that $|h(\bar{q}^{(M)}) - h(\tilde{p}^{(M)})| < \epsilon_3$. From the definition of the decoding rule in (5) we have

$$\begin{aligned} \frac{k}{M(b-n_1)} &< 1 - h(\bar{q}^{(M)}) - \epsilon_1 \\ &\leq 1 - h(\tilde{p}^{(M)}) - (\epsilon_1 - \epsilon_3) \ . \end{aligned} \tag{17}$$

Thus our code will be operating at a gap at least $(\epsilon_1 - \epsilon_3)$ from the empirical capacity.

Since the capacity-achieving input distribution for any binary symmetric channel is Bernoulli(1/2), the encoder and decoder will choose an iid random codebook of blocklength $M^*(b-n_1)$ with 2^{k+M^*} messages from that distribution using common randomness. Because M is not known in advance, we must guarantee that the codebook and its truncations to blocklengths $M(b-n_1)$ perform well, for $M_* \leq M \leq M^*$. From an exercise in Gallager [8, Exercise 5.23, p.539], we have that the error averaged over the ensemble and messages is upper bounded by

$$\exp\left(-M(b-n_1)\left(\frac{(C - \frac{k+M^*}{M(b-n_1)})^2}{8/e^2 + 4(\ln 2)^2}\right)\right) \ . \tag{18}$$

To turn this into a bound on maximum error, we remove the half of the messages with highest error probability. We must guarantee that the error probability is small for all $M \in \{M_*, \ldots, M^*\}$, so it is sufficient to perform M^* such thinning operations to obtain a codebook with 2^k messages.

The loosest bound in (18) is for $M(b-n_1)=k$ in the exponent and $C-k/M(b-n_1)=(\epsilon_1-\epsilon_3)$, so

$$\mathbb{P}\left(E_2|E_1^c\right)\leq \exp\left(-k\left(\frac{(\epsilon_1-\epsilon_3-M^*/k)^2}{8/e^2+4(\ln 2)^2}\right)\right) \ . \quad (19)$$

Under the assumptions $\tau>\epsilon_1$, $\bar{q}^{(M)}\notin[1/2-\tau,1/2+\tau]$, and E_1^c, we know that $\bar{p}^{(M)}$ and $\bar{q}^{(M)}$ must lie on the same side of $1/2$. Therefore our decoding rule can be a simple minimum-distance (if $\bar{q}^{(M)}<1/2$) or maximum-distance (if $\bar{q}^{(M)}>1/2$) decoding rule, which is as good as maximum-likelihood for these channels. ■

We denote the overall error by $\delta=\delta(N)=\mathbb{P}(E_1)+\mathbb{P}(E_2|E_1^c)$. The bounds in (10) and (13) provide an upper bound on δ in terms of the other parameters.

B. Rate analysis

There two types of rate loss – loss within a round and loss for the final uncompleted round.

Lemma 4 (Rate loss for uncompleted round): The fraction of channel uses lost $\gamma(N)$ due to a nonterminating final round is upper bound by

$$\gamma(N)\leq\frac{M^*b}{N} \ . \quad (20)$$

Proof: The proof follows immediately from Lemma 1 since no round can be larger than M^* chunks. ■

Lemma 5 (Rate loss within a round): Suppose we are under the event E_1^c. Then for a round with an empirical noise frequency $\bar{p}^{(M)}$ ones in the noise sequence, the rate is at least

$$\frac{k}{(b-n_1)M}\geq 1-h(\bar{p}^{(M)})-\beta_0(N) \ , \quad (21)$$

where $\beta_0(N)=\max\{h(M_*^{-1})+\epsilon_3+\epsilon_1+\frac{M_*}{(M_*-1)^2},h(1/2-\tau-\epsilon_2)\}$.

Proof: We experience rate loss in both rounds that terminate due to "BAD NOISE" and ones in which we decode. In the "BAD NOISE" rounds, the rate is obviously 0. Under E_1^c, this can only happen when $\bar{p}^{(M)}\in[\frac{1}{2}-\tau-\epsilon_2,\frac{1}{2}+\tau+\epsilon_2]$, so we have the lower bound

$$\frac{k}{(b-n_1)M}\geq 1-h(\bar{p}^{(M)})-h(1/2-\tau-\epsilon_2) \ . \quad (22)$$

In other rounds, our loss comes from ϵ_1 and deviations in $\bar{q}^{(M)}$. While satisfying (5) gives us an upper bound on the rate, we want a lower bound of the form given in (21). To get this, we will use the fact that (5) is not satisfied in chunk $M-1$, so

$$\frac{k}{(b-n_1)(M-1)}\geq 1-h(\bar{p}^{(M-1)})-\epsilon_1-\epsilon_3 \ . \quad (23)$$

To get the bound, we must bound the change in rate from $M-1$ to M and the change in $\bar{p}^{(M-1)}$ to $\bar{p}^{(M)}$ to get the bound. From Lemmas 9 and 1, we have

$$\frac{k}{(b-n_1)M}\geq\frac{k}{(b-n_1)(M-1)}-\frac{M_*}{(M_*-1)^2} \ . \quad (24)$$

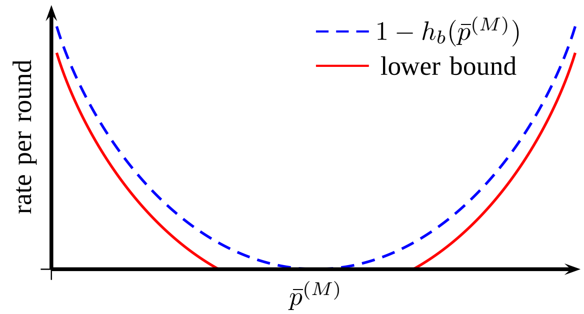

Fig. 3. Example of convex lower bound.

Combining Lemmas 8 and 1 with equations (23) and (24), we get

$$\begin{aligned}\frac{k}{(b-n_1)M}\geq & 1-h(\bar{p}^{(M)})-h(M_*^{-1})\\ & -\epsilon_3-\epsilon_1-\frac{M_*}{(M_*-1)^2} \ . \end{aligned}\quad (25)$$

Combining the lower bounds (22) and (25), we get (21). ■

An example of the bound above is given in Figure 3.

Lemma 6 (Overall rate loss): Under E_1^c, we have

$$\frac{k}{bM}\geq 1-h(p)-\beta(N) \ , \quad (26)$$

where

$$\beta(N)=\beta_0+h(\frac{M^*b}{N})+\frac{M^*b}{N}+\frac{n_1}{b} \ .$$

Proof: One can express p as a weighted sum of the $\bar{p}^{(M)}$ from each of the completed rounds and a penalty for the nonterminating round of no more than $\frac{M^*b}{N}$ by Lemma 4, where the weights correspond exactly to the fraction of chunks used for that round. Using the same weights on the empirical rates from each round gives us gives us the rate over the entire block N. Lemmas 4 and 5 along with the convexity of the $1-h(p)$ in p gives

$$\begin{aligned}\frac{k}{bM}\geq & 1-h(p)-\beta_0(N)-h\left(\frac{M^*b}{N}\right)\\ & -\frac{M^*b}{N}-\frac{n_1}{b} \ . \end{aligned}$$

■

C. Setting the parameters

The algorithm as described has a large number of parameters which must satisfy various asymptotic conditions if we are to approach $1-h_b(p)$. By way of an example, let us set

$k(N)$	$\Theta(N^{1/2})$
$b(N)$	$\Theta(N^{1/4})$
$n_1(N)$	$\Theta(N^{1/8})$
$\epsilon_1(N)$	$\Theta(N^{-1/16})$
$\epsilon_2(N)$	$\Theta(N^{-1/32})$
$\tau(N)$	$\Theta(N^{-1/8})$

Since $h_b(\epsilon) = O(\epsilon \log \epsilon^{-1})$, these parameters give $\epsilon_3 = O(N^{-1/32} \log N)$ so $\epsilon_1 - \epsilon_3 > 0$. Therefore the errors in (10) and (13) both converge to 0. Furthermore, the rate loss $\beta(N) \to 0$ by Lemma 6. This is sufficient to complete the proof of Theorem 1.

IV. Discussion

In the algorithm, we assume that the encoder and decoder use $(N \log 3)/b(N)$ bits of active feedback to send the decisions after each chunk as well as common randomness to choose the training positions T_i and the random codebooks for each round. The decoder could use active feedback to inform the encoder of the $n_1(N)$ training positions in chunk i before it is sent. Over the $N/b(N)$ chunks this would require an additional $N(n_1(N)/b(N)) \log N$ bits, which is sublinear in N for the example in the previous section.

A more difficult issue is the randomness used to generate the codebooks for each round. We can use a single codebook and choose a random permutation of the entries in each round, which would require $O(M^*(b-n_1)\log(M^*(b-n_1)))$ bits of common randomness per round, or $O(N \log N)$ bits overall. One way of reducing this randomness would be to prove the existence of nested *list-decodable codes* and use active feedback to provide a small secret key that can disambiguate the list, as in [9], [10]. Such an extension will appear in the journal version of this paper.

As in the model and algorithm considered by Shayevitz and Feder, our algorithm does not rely heavily on the assumption of binary inputs and noise. The major feature needed is that the parameters of the channel can be estimated at the decoder using training sequences. Channels in which the capacity is achievable with a single input distribution fall into this category. Feedback may also be useful to change the codebook when there is not one universal capacity-achieving input distribution but there may be nontrivial rate loss since the mutual information is concave with respect to the channel input distribution.

We could also consider a class of noise models whose p varies in a piecewise-constant fashion. One could view this as a kind of "block fading" model for the binary additive channel. These models may be related to the on-line estimation problems studied by Kozat and Singer [11]. It may be possible to modify our algorithm to adapt the value of k by trying to learn the coherence time of the channel. In the sense of competitive optimality, our competitor class could be coding algorithms that know the coherence intervals exactly.

As in the work of Shayevitz and Feder [2], the algorithm described in this paper exploits the *local* variations in the empirical noise distribution $\bar{p}^{(M)}$ on the order of the round length. This in turn implies that our scheme could achieve rates higher than $1 - h_b(p)$ for noise sequences which have local variation on that order. By decreasing k we can exploit finer variations at the expense of the error performance. Future work will include finer analysis of the scheme to make explicit these gains from local variability.

Acknowledgments

We thank Ofer Shayevitz and Meir Feder for providing a preprint of their paper after their presentation at the Kailath Colloquium [12]. This work grew out of a presentation of that work for UC Berkeley's advanced information theory course EE290S. Special thanks go to the other students in that class for helpful discussions. The work of A.D. Sarwate and M. Gastpar was supported in part by the National Science Foundation under award CCF-0347298. The work of K. Eswaran, A. Sahai, and M. Gastpar was supported in part by the National Science Foundation under award CNS-0326503.

Appendix

We state three simple technical lemmas. The proofs can be found in [13].

Lemma 7: If

$$|p_1 - p_2| \leq \epsilon, \tag{27}$$

then

$$|h_b(p_1) - h_b(p_2)| \leq h_b(\epsilon). \tag{28}$$

Lemma 8:

$$\left| h_b(\bar{p}^{(n-1)}) - h_b(\bar{p}^{(n)}) \right| \leq h_b\left(n^{-1}\right) . \tag{29}$$

Lemma 9: For $n \geq n_0$,

$$\frac{1}{n} \geq \frac{1}{n-1} - \frac{1}{(n_0-1)^2}. \tag{30}$$

References

[1] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 27–34, 1988.

[2] O. Shayevitz and M. Feder, "Achieving the empirical capacity using feedback part I: memoryless additive models," submitted to IEEE Transactions on Information Theory. [Online]. Available: http://www.eng.tau.ac.il/˜ofersha/empirical_capacity_part1.pdf

[3] A. Sahai, "Why block length and delay are not the same thing," *IEEE Transactions on Information Theory*, Submitted 2006. [Online]. Available: http://www.eecs.berkeley.edu/˜sahai/Papers/FocusingBound.pdf

[4] A. Sahai and S. Draper, "Beating the Burnashev bound using noisy feedback," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2006.

[5] A. Sahai, "Balancing forward and feedback error correction," in *Proc. of the Information Theory and Applications Workshop*, Jan. 2007.

[6] E. Soljanin, "Hybrid ARQ in wireless networks," in *DIMACS Workshop on Network Information Theory*, March 2003.

[7] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 1, pp. 13–30, March 1963.

[8] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.

[9] M. Langberg, "Private codes or Succinct random codes that are (almost) perfect," in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004)*, Rome, Italy, 2004.

[10] A. Sarwate and M. Gastpar, "Channels with nosy "noise"," in *Proceedings of the 2007 IEEE International Symposium on Information Theory*, Nice, France, 2007.

[11] S. Kozat and A. Singer, "Universal Switching Linear Least Squares Prediction," in *Proc. of the 2006 Information Theory and its Applications Workshop*. La Jolla, CA: UCSD, February 2006.

[12] M. Feder, "Achieving the empirical capacity of individual noise channels using feedback," 2006 Kailath Symposium, July 2006.

[13] K. Eswaran, A. Sarwate, A. Sahai, and M. Gastpar, "Binary additive channels with individual noise sequences and limited active feedback," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2007-5, January 8 2007. [Online]. Available: http://sipc.eecs.berkeley.edu/