Faster-than-Nyquist Coding: The Merits of a Regime Change*

Uri Erez and Gregory W. Wornell
Dept. EECS, MIT, Cambridge, MA
{uri,gww}@allegro.mit.edu

Mitchell D. Trott
HP Labs, Palo Alto, CA
mitchell.trott@hp.com

Abstract

Rateless codes are good codes of infinite length that have the property that prefixes of such codes are themselves good codes. This makes them attractive for applications in which the channel quality is uncertain, where systems transmit as much of a codeword as necessary for decoding to be possible. While low complexity rateless codes are known to exist for the erasure channel, this paper shows they can also be constructed for any Gaussian channel.

We consider two classes of such codes. The first class employs a structure whereby the transmission is block-structured, and is applicable when the time at which decoding will begin is known to the transmitter. In the first block, the bits to be sent are divided into several groups, each of which is binary encoded and the results are superimposed to form a layered code. In subsequent blocks, the binary codewords from the first block are simply repeated, but with a random dither. The associated decoder structure employs successive cancellation together with maximal ratio combining. An efficient recursion is developed for the power allocation in each block to ensure the rateless property. When the time at which decoding will begin is not known, we develop a variant on this approach whereby the layering is accomplished by faster-than-Nyquist signaling and where the successive cancellation is implemented by a block-structured decision feedback equalizer that is used in conjunction with an interleaver. This architecture leads to the necessary symmetric power allocation.

Both approaches require very low complexity, and can be used to come within any desired fraction of capacity on an unknown Gaussian channel by choosing a good binary "base" code of sufficiently low rate. We quantify the tradeoffs, which reveal, for example, that to achieve 90% of capacity requires a code of rate roughly 1/7.

The design of effective "rateless" codes has received renewed strong interest in the coding community, motivated by a number of emerging applications. Such codes have a long history, and have gone by various names over time, among them incremental redundancy codes, rate-compatible punctured codes, H-ARQ type II codes, flexible rate codes, and static broadcast codes. The focus of this work is on the design of such codes for the additive white Gaussian noise (AWGN) channel.

From a purely information theoretic perspective, the problem of variable rate transmission is by now well understood; see, e.g., [7] for a comprehensive treatment. Indeed, for classes of channels having one maximizing input distribution, a codebook drawn independently and identically distributed (iid) at random according to the capacity-achieving input distribution will be good

*This work was supported in part by the HP/MIT Alliance and an MIT Deshpande Center Ignition Grant.

with high probability, when truncated to (a finite number of) different lengths. From a coding perspective however, we want codes that are capacity approaching[1] while still allowing for low-complexity encoding and decoding. A remarkable example of such codes for *erasure* channels are the recent Raptor codes [6], which build on the LT codes of Luby.

Surprisingly little is known about what is possible beyond the realm of erasure channels. Recent work on the application of Raptor codes to binary-input symmetric-output channels are [5, 1]. In these works, the performance of Raptor codes was studied when applied to a binary-input AWGN channel (among other channels) where the degree distribution is optimized to this class of channels. It is shown that no distribution allows Raptor codes to approach the capacity of this class of channels simultaneously (at different SNRs). Beyond this, there is the problem that the use of binary codes in itself precludes achieving the capacity of the original AWGN channel: from a practical standpoint, binary signaling may be "nearly" capacity achieving only at low SNR.

As we will develop, good rateless codes for the AWGN are possible, and can exploit the fact that at very low SNR, a trivial means to obtain a variable rate code is by means of repetition. While this is true for a much broader class of channels [7], for the AWGN channel there is a very simple way to combine the received repeated blocks: maximal ratio combining (MRC). Ultimately, we obtain our rateless codes by combining this idea with dithering and a superposition coding strategy to obtain many low SNR channels from one higher SNR one. The resulting scheme has flexible rate but has however a specified time origin. We then develop a Faster-than-Nyquist (FTN) signaling scheme that allows for a time-invariant structure.

1 Superposition coding

Consider a power constrained AWGN channel $y = x + z$ where $z \sim \mathcal{N}(0, N)$ and $E\{x^2\} \leq P$. The capacity[2] of the channel is $\frac{1}{2}\log(1 + \text{SNR})$, where $\text{SNR} = P/N$, and is achieved taking $x \sim \mathcal{N}(0, P)$.

In superposition coding, the transmitted signal x is the sum of independent signals, x_i, so that $x = \sum_{l=1}^{L} x_l$. If we take x_l to be Gaussian, we obtain a capacity achieving distribution for the AWGN channel. That is, we allot powers P_l with $P = \sum_{l=1}^{L} P_l$. We may write the capacity as $C(P) = \sum_{l=1}^{L} C_l$, where

$$C_l = \frac{1}{2}\log\left(1 + \frac{P_l}{\sum_{k<l} P_k + N}\right). \tag{1}$$

Superposition coding with interference cancellation may be interpreted as a decomposition of the channel into L *ordered* parallel channels where channel l views the code for channels $k < l$ as additive interference but suffers from no interference from channels "farther downstream" $k > l$. One coding approach in this case is that of successive cancellation at the decoder. Thus, we may use a (capacity achieving) codebook $\mathcal{C}_l$, $l = 1, \ldots, L$, for each channel, i.e., $|\mathcal{C}_l| = e^{nC_l}$. The encoder chooses a codeword $\mathbf{x}_l \in \mathcal{C}_l$ from each codebook. The decoder may start by decoding

[1] We use the term "capacity approaching" loosely to mean practical codes that allow to approach capacity "closely".

[2] In this paper, all logarithms are natural and the unit of information is nats unless stated otherwise.

channel $l = L$. At this stage, the signals from all the other channels are treated as noise so the SNR is $P_L / \left(\sum_{k=1}^{L-1} P_k + N\right)$ and we may achieve a rate of C_L on this channel. Next, we subtract off the decoded word $\hat{\mathbf{x}}_L$ from the received output $\mathbf{y}$, leaving us with an effective SNR of $P_{L-1} / \left(\sum_{k=1}^{L-2} P_k + N\right)$. Comparing with (1), we see that message $\mathbf{x}_{L-1}$ may now be successfully decoded with high probability. We continue in this "onion peeling" approach down to layer one.

2 Thou shalt not repeat??

When a (Gaussian) codeword is repeated r times over an AWGN channel, the resulting mutual information per symbol is

$$I_{\text{rep}}(\text{SNR}) = \frac{1}{2r} \log(1 + r \cdot \text{SNR}).$$

On the other hand, when transmitting r independent (Gaussian) codewords, the mutual information is

$$I_{\text{ind}}(\text{SNR}) = \frac{1}{2} \log(1 + \text{SNR}).$$

As the SNR decreases we have

$$\lim_{\text{SNR} \to 0} \frac{I_{\text{rep}}(\text{SNR})}{I_{\text{ind}}(\text{SNR})} = 1.$$

Therefore, the loss due to repetition vanishes as the SNR goes to zero.

Say we want to transmit over an AWGN channel with unknown SNR but we have some upper bound on the SNR, i.e., we know that $\text{SNR} < \text{SNR}^*$. A natural approach thus for obtaining a rateless code would be to use a large number of layers so that each subchannel is in the low SNR regime. Let

$$C^* = \frac{1}{2} \log(1 + \text{SNR}^*). \tag{2}$$

Assigning equal rates to the subchannels, each subchannel has a capacity of C^*/L. Denote the number of collected blocks by r. Define $\text{SNR}(r)$ by

$$\frac{1}{2} \log(1 + \text{SNR}(r)) = \frac{C^*}{r}, \tag{3}$$

and let $N(r)$ be the corresponding noise power, i.e., $N(r) = P/(e^{2C^*/r} - 1)$. Note that $\text{SNR}(1) = \text{SNR}^*$.

Denote the power allocated to layer l in block i by $P_l(i)$. Also let $\text{SNR}_{l,i}(r)$ denote the SNR of this layer assuming the actual SNR is $\text{SNR}(r)$, i.e.,

$$\text{SNR}_{l,i}(r) = \frac{P_l(i)}{\sum_{k<l} P_k(i) + N(r)}. \tag{4}$$

The corresponding capacity of the layer is $C_{l,i}(r) = \frac{1}{2} \log\left(1 + \text{SNR}_{l,i}(r)\right)$. We next find the power allocation $P_k(i)$ such that for $\text{SNR} = \text{SNR}(r)$, if we collect the first r blocks, each layer will have a capacity of C^*/Lr per symbol. That is, for $l = 1, \ldots, L$, and for every r, we want to have

$$\sum_{i=1}^{r} C_{l,i}(r) = \frac{C^*}{L}. \tag{5}$$

Note that since $\sum_{i=1}^{L} P_l(i) = P$, for any r and i, we have

$$\sum_{l=1}^{L} C_{l,i}(r) = \frac{C^*}{r} = C(r), \tag{6}$$

regardless of the power allocation. It follows that (5) may be solved by recursion on the block number i. Assume (5) is satisfied for $r-1$. The additional rate layer l needs is

$$\Delta_l(r) = \frac{C^*}{L} - \sum_{i=1}^{r-1} C_{l,i}(r). \tag{7}$$

But it follows from (6) that $\sum_{l=1}^{L} \Delta_l(r) = C^*/r = \frac{1}{2}(1 + \mathrm{SNR}\,(r))$. Thus, we need simply assign powers $P_l(r)$ corresponding to $\Delta_l(r)$. Explicitly, the recursion is

$$\frac{1}{2} \log \left(1 + \frac{P_l(r)}{[P - \sum_{k>l} P_k(r)] - P_l(r) + N(r)} \right) = \Delta_l(r), \tag{8}$$

where we begin with $r = 1$, proceed from layer L to layer one, and then go on to $r = 2$, and so forth. Note that the only indeterminate in (8) is $P_l(r)$.

As we can ensure that every layer is at sufficiently low SNR, a naive approach to obtain a rateless code would be to repeat the same L codewords from block to block, scaling the codeword so as to have power $P_l(i)$, and then use MRC at the receiver. This is obviously flawed, as (without power scaling) it amounts to repetition of the block, which cannot be efficient in a mutual information sense at high SNR. The snag, from the point of view of an individual layer, is that while the Gaussian noise is independent from block to block, the interference is not and is combined *coherently*. In the next section we show how to circumvent this problem.

2.1 Dithered repetition transmission

Let $\mathbf{x}_l$ be taken from an iid unit variance Gaussian codebook of size $e^{nC^*/L}$ and define $\mathbf{x}_l(i) = \sqrt{P_l(i)} \cdot \mathbf{x}_l$. Let $d_l(r)$, $l = 1, \ldots, L$, be vectors of ± 1s drawn iid Bernoulli 1/2, known to both transmitter and receiver. The transmitter sends at block i

$$\mathbf{x}(i) = \sum_{l=1}^{L} \mathbf{x}_l(i) \odot \mathbf{d}_l(i) \tag{9}$$

where $\odot$ denotes component-wise multiplication.

The received i-th block is $\mathbf{y}(i) = \mathbf{x}(i) + \mathbf{z}(i)$. Let $\alpha_l(i) = \text{SNR}_{l,i}(r)/\sum_{k=1}^{r} \text{SNR}_{k,i}(r)$. For each layer $l = L, \ldots, 1$, the receiver forms the MRC estimate

$$\mathbf{y}_l = \sum_{i=1}^{r} \alpha_l(i) \cdot \mathbf{d}_l(i) \odot \frac{\mathbf{y}(i) - \sum_{k>l} \sqrt{P_k(i)}\hat{\mathbf{x}}_k \odot \mathbf{d}_k(i)}{\sqrt{P_l(i)}}, \tag{10}$$

where the $\hat{\mathbf{x}}_k$ are the previously decoded codewords. Assuming $\hat{\mathbf{x}}_k = \mathbf{x}_k$, we have

$$\mathbf{y}_l = \mathbf{x}_l + \sum_{i=1}^{r} \alpha_l(i) \cdot \mathbf{z}_l(i) \triangleq \mathbf{x}_l + \mathbf{z}_l, \tag{11}$$

where

$$\mathbf{z}_l(i) = \frac{1}{\sqrt{P_l(i)}} \cdot \left(\sum_{k<l} \mathbf{x}_k(i) \odot \mathbf{d}_l(i) \odot \mathbf{d}_k(i) + \mathbf{z}(i) \right). \tag{12}$$

Thus, $\mathbf{z}_l$ is an iid random vector and the resulting SNR is $\text{SNR}_l = \sum_{i=1}^{r} \text{SNR}_{l,i}(r)$. The receiver decodes $\hat{\mathbf{x}}_l$ from $\mathbf{y}_l$.

By the convexity of the logarithm function and since $\mathbf{z}_l$ is not Gaussian, it follows that the accumulated SNR in each layer, i.e., the SNR in channel (11) satisfies

$$\frac{1}{2}\text{SNR}_l \geq \frac{C^*}{L}. \tag{13}$$

Therefore, the achievable rate per layer of the coding scheme is lower bounded by

$$R \geq \frac{1}{2} \log \left(1 + \frac{2C^*}{L} \right). \tag{14}$$

Thus, by choosing L sufficiently large, we may approach capacity arbitrarily closely. The fraction of capacity attained, which we refer to as the efficiency of the scheme, satisfies

$$\text{efficiency} = \frac{L \cdot R}{C^*} \geq \frac{2R}{e^{2R} - 1} \geq 1 - R. \tag{15}$$

See Figure 1. In practice, as we will choose a sufficient number of layers such that the SNR per layer is low, we could use a binary code instead of a Gaussian codebook.

3 Faster-than-Nyquist coding scheme

We have obtained a variable-rate transmission scheme. However, due to the time-varying power allocation, the scheme has an absolute time origin. In effect, while the number of received blocks may vary, the receiver has to start listening to the first transmitted block. We next describe a time-invariant layering scheme based on FTN signaling.

For simplicity, we consider a baseband bandlimited AWGN channel model,

$$y(t) = x(t) + z(t), \tag{16}$$

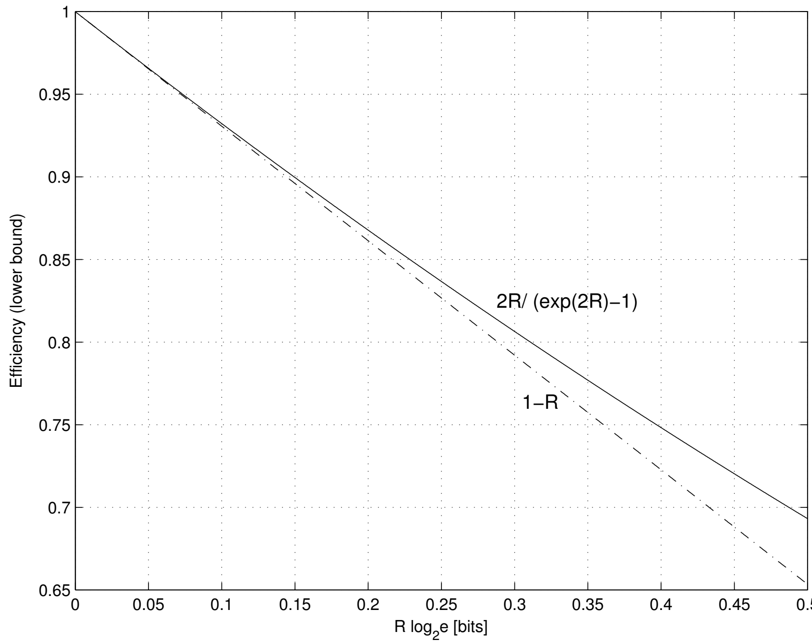

Figure 1: Efficiency of dithered repetition scheme.

where we assume an ideal channel of bandwidth W, and $n(t)$ is AWGN noise with one-sided power spectral density N_0. The input signal is subject to a power constraint $E\{x^2(t)\} \leq P$. The capacity of this channel is

$$C_{[\text{nats/s}]} = W \log\left(1 + \frac{P}{N_0 W}\right). \tag{17}$$

We assume PAM modulation so that

$$x(t) = \sum_l x_l \cdot p(t - l \cdot T), \tag{18}$$

where T is the symbol duration. The sampling rate is taken to be equal to the signaling rate. At the output of the sampled matched filter (MF) we have a discrete-time channel,

$$y_l = x_l * h_l + z_l, \tag{19}$$

where $h(t) = p(t) * p(-t)$, $H(e^{j2\pi f}) = \mathcal{F}\{h(t)\} = \frac{1}{T}\sum_l \left|P\left(f + \frac{l}{T}\right)\right|^2$, $h_l = h(l \cdot T)$, and $S_{zz}(e^{j2\pi f}) = \frac{N_0}{2} H(e^{j2\pi f})$. Denote the Nyquist signaling rate by $T^* = 1/2W$ and the "over-signaling" ratio by $\gamma = T^*/T$. We assume a Nyquist pulse $p(t)$ so that

$$\frac{1}{T^*}\sum_l \left|P\left(f + \frac{l}{T^*}\right)\right|^2 = 1. \tag{20}$$

In the following, for sake of simplicity, we assume that $h(t)$ is a sinc function so there is no capacity loss due to excess bandwidth. The discrete-time channel response is thus $h_l = \sin(\pi l \gamma)/\pi l \gamma$. If we signal at the Nyquist rate, i.e., if $\gamma = 1$, we have $h_l = \delta_l$. Taking $\gamma > 1$ necessarily introduces intersymbol interference (ISI) which in effect creates the "layering".

3.1 MMSE-DFE equalizer induced channel

In Section 2 we used successive decoding and stripping of layers. The analogous receiver structure for an ISI Gaussian noise channel is that of unbiased MMSE decision-feedback (DFE) decoding.

Consider an unbiased MMSE-DFE receiver with feedforward filter $\mathrm{FFE}(z)$ (see [4]). The resulting channel after the FFE filtering is

$$v_l = y_l * \mathrm{FFE}_l = x_l * g_l + w_l, \tag{21}$$

where $G(z) = H(z)\mathrm{FFE}(z)$ is the resulting impulse response and $W(z) = Z(z)\mathrm{FFE}(z)$ is the filtered noise. We denote by $G(z)^+$ the Z-transform of the *causal* part of g_l. We note that $g_0 = 1$ since the equalizer is unbiased. We define the precursor and postcursor interference suffered by x_l as

$$z_l^{\mathrm{pre}} = \sum_{k=-\infty}^{-1} g_k x_{l-k} \tag{22}$$

and

$$z_l^{\mathrm{post}} = \sum_{k=1}^{\infty} g_k x_{l-k} \tag{23}$$

Assuming error-free decisions (the ideal DFE assumption), the feedback filter eliminates the postcursor ISI and thus as the input to the slicer we have

$$\begin{aligned} \tilde{v}_l &= x_l + z_l^{\mathrm{pre}} + w_l && (24)\\ &= x_l + \tilde{w}_l, && (25) \end{aligned}$$

where $\tilde{w}_l$ is the combined filtered Gaussian noise w_l and the precursor interference.

The ideal DFE condition cannot hold on a symbol-wise basis but can be justified when decisions are made on coded blocks; see [3, 2]. We briefly describe such a system as proposed by Guess and Varanasi. We need to restrict attention to impulse responses $G(z)$ having a finite support[3]. Thus we assume that K is the smallest integer [4] such that $g_k = 0$ for all $|k| > K$. Let $\mathbf{x}_l = 0$ for $l = 0, 1, \ldots, K-1$, where the vectors are of length n, the components are denoted by x_l^i, where $i = 0, \ldots, n-1$. Also let $\mathbf{x}_K, \mathbf{x}_{K+1}, \ldots, \mathbf{x}_{K+M-1}$ be codewords (possibly from the same codebook). Thus, the total number of transmitted blocks is $L = K + M$. In order for postcursor cancellation to be possible, an interleaver of depth L is used. The transmitter sends the interleaved sequence $\tilde{x}_i$ where

$$\tilde{x}_i = x_{\lfloor i/n \rfloor}^{i \bmod n} \quad \text{for} \quad i = 0, \ldots, nL-1. \tag{26}$$

Thus, the transmitted stream is

$$\underbrace{0,\ldots,0}_{K}, x_K^0, x_{K+1}^0, \ldots, x_{K+M-1}^0, \underbrace{0,\ldots,0}_{K}, x_K^1, x_{K+1}^1, \ldots, x_{K+M-1}^1, \ldots, \underbrace{0,\ldots,0}_{K}, x_K^{n-1}, x_{K+1}^{n-1}, \ldots, x_{K+M-1}^{n-1}$$

[3] In practice this would mean that we would approximate the ideal FFE filter with one that guarantees that this constraint is met or ignore the tail of the postcursor from that point, leaving some residual ISI as noise.

[4] As we shall see in Section 4, this length is of great practical importance.

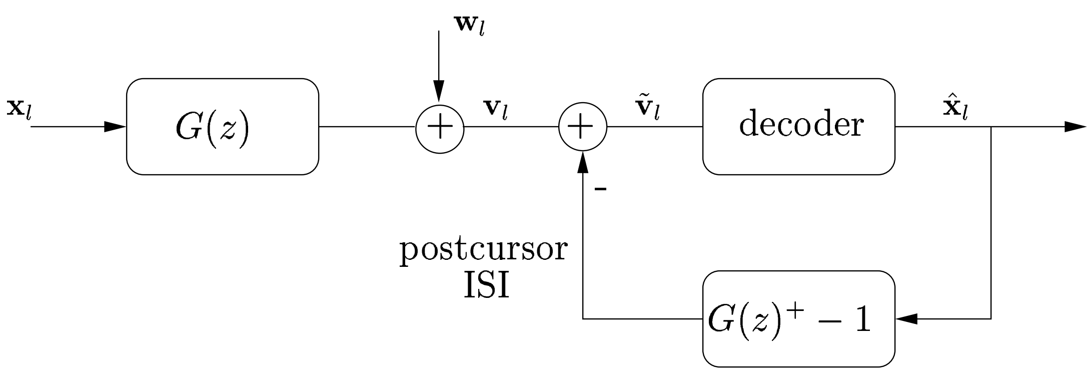

Figure 2: Block DFE structure.

The effect of the interleaving is that the impulse response now effectively operates on blocks rather than symbols, in the sense that we have

$$\mathbf{v}_l = \sum_k g_k \mathbf{x}_{l-k} + \mathbf{w}_l. \tag{27}$$

We define the (block) precursor and postcursor interference suffered by $\mathbf{x}_l$ as $\mathbf{z}_l^{\text{pre}} = \sum_{k=-\infty}^{-1} g_k \mathbf{x}_{l-k}$ and $\mathbf{z}_l^{\text{post}} = \sum_{k=1}^{\infty} g_k \mathbf{x}_{l-k}$. We may now recover equations (25) in block form and obtain

$$\begin{aligned} \tilde{\mathbf{v}}_l &= \mathbf{x}_l + \mathbf{z}_l^{\text{pre}} + \mathbf{w}_l & (28) \\ &= \mathbf{x}_l + \tilde{\mathbf{w}}_l, & (29) \end{aligned}$$

where the postcursor is now canceled out after *decoding* of the corresponding codewords. The resulting DFE coding structure is shown in Figure 2.

Assume that $\text{FFE}(z)$ is an unbiased MMSE-DFE front end filter and that the input vectors $\mathbf{x}_k$, $k \neq l$ are iid Gaussian. The channel (29) at time l is an AWGN channel over the n-dimensional l-th block, having the same capacity as the original channel (19); see, e.g., [3, 2]. Note that there is an overhead of K wasted codewords. Thus, we need to take $M \gg K$ to approach capacity. The capacity is independent of the signaling rate $1/T$. We therefore have $\frac{1}{2T} \log\left(1 + \text{SNR}_{\text{MMSE-DFE-U}}\right) = W \log\left(1 + \frac{P}{WN_0}\right)$ or

$$\text{SNR}_{\text{MMSE-DFE-U}} = \left(1 + \frac{P}{WN_0}\right)^{T/T^*} - 1, \tag{30}$$

since $W = 1/2T^*$. Thus, as the signaling rate $1/T$ grows, the SNR per sample decreases.

3.2 FTN dithered repetition scheme

We may use fast signaling to drive the discrete-time SNR to the low SNR regime. Denote a target SNR per sample by SNR^* similarly to Section 2.1. We may use a Gaussian codebook designed to approach the AWGN capacity of $\frac{1}{2}\log(1+\text{SNR}^*)$, and transmit M codeword blocks as described above. We then repeat a dithered version of these same M codewords over and over again. The receiver in turn "listens" and collects r copies, applies FFE filtering to each block, and then combines (averages) them. The number of collected blocks is chosen such that such that the accumulated SNR satisfies,

$$r \cdot \text{SNR}_{\text{MMSE-DFE-U}} \geq \text{SNR}^*. \tag{31}$$

It then applies the block MMSE-DFE decoding structure to obtain the codewords $\mathbf{x}_K, \ldots, \mathbf{x}_{K+M-1}$. Note that as the front-end filter $\text{FFE}(z)$ depends on the SNR and thus will vary with the number of received blocks. Consequently, the impulse response $G(z)$ also implicitly depends on the SNR.

We use dithering as in Section 2.1 to prevent the correlation of the precursors. Let $\mathbf{d}_l(i)$ $l = 0, \ldots, L-1$ be binary ± 1 valued vectors, drawn iid with equal probability. We transmit at "time" (repetition/copy) i the vectors

$$\mathbf{x}_0 \odot \mathbf{d}_0(i), \mathbf{x}_1 \odot \mathbf{d}_1(i), \ldots, \mathbf{x}_{L-1} \odot \mathbf{d}_{L-1}(i).$$

At the receiver, assuming $\mathbf{x}_k$ were successfully decoded for $k < l$, we combine the signals as

$$\bar{\mathbf{v}}_l = \frac{1}{r}\sum_{i=1}^{r} \mathbf{d}_l(i) \cdot \tilde{\mathbf{v}}_l(i) \tag{32}$$

$$= \mathbf{x}_l + \frac{1}{r}\sum_{i=1}^{r}\sum_{k=-\infty}^{-1} g_k \mathbf{d}_l(i) \odot \mathbf{d}_{l-k}(i) \odot \mathbf{x}_{l-k} + \frac{1}{r}\sum_{i=1}^{r}\mathbf{w}_l(i) \tag{33}$$

$$= \mathbf{x}_l + \sum_{k=-\infty}^{-1} g_k \left(\frac{1}{r}\sum_{i=1}^{r} \mathbf{x}_{l-k} \odot \mathbf{d}_{l-k}(i) \odot \mathbf{d}_l(i)\right) + \frac{1}{r}\sum_{i=1}^{r}\mathbf{w}_l(i). \tag{34}$$

Note that $\mathbf{x}_{l-k} \odot \mathbf{d}_{l-k}(i) \odot \mathbf{d}_l(i)$, $i = 1, \ldots, r$, are uncorrelated. Thus, the variance of the precursor is reduced by a factor of r, just like the Gaussian noise. The SNR increases therefore by a factor of r as desired so that (31) is satisfied. As in Section 2.1, the precursor noise is not quite Gaussian (which is beneficial). Thus, we may successfully decode $\mathbf{x}_l$ from $\bar{\mathbf{v}}_l$ with high probability. In practice, we could replace the Gaussian codebook with a binary one.

4 Discussion and Remarks

We presented two rateless coding schemes. The first uses dithered repetition and time-varying power allocation and results in a rateless code having a fixed time origin. We then proposed an FTN coding scheme that allows reception to begin at an arbitrary point in time.

When considering the merits of the FTN coding scheme, there are a few practical issues to be addressed. As we saw, the total number of coded blocks is $M + K$. In the K first blocks we send no information and so these blocks are "wasted" in terms of transmission rate. Thus we need M to be much larger than K. However, choosing M very large also has a drawback. For a given total transmission length $(M+K)n$, the length of the individual codewords n is inversely proportional to M. This is undesirable from a code design perspective as one needs a long blocklength to approach error free transmission close to capacity. Thus, given K, choosing the interleaver depth is non trivial.

Determining K, the length of the ISI that is taken into account, also requires consideration. The discrete-time impulse response $G(z)$ is dependent on the SNR and thus K should be large enough so as to leave only a small tail of $G(z)$ as residual ISI for all SNRs in which the system should operate. Choosing K very large however will be disadvantageous as discussed above.

Finally, the signaling rate should be addressed. To approach capacity the SNR per layer must be sufficiently low and hence fast signaling is beneficial. For instance, to achieve ninety percent of capacity, we need an effective capacity-approaching code of rate $1/7$, see Figure 1. While this operating point is within the current state-of-the-art, coding at lower rates poses a challenging code design problem.

It is interesting to compare the merits of FTN coding with that of conventional multilevel signaling. In multilevel signaling, the cardinality of the signal set is matched to the SNR. A mapping is then applied from bits to constellation points. A design problem oftentimes is the rate allocation for the individual layers. In FTN coding, a binary code is used in a more direct and symmetrical manner, while the "multilevel nature" of the signal is reflected in the induced ISI. Thus, in FTN coding, it is sufficient to design a single low rate code which can be highly optimized to a prescribed SNR. The actual transmission rate is then determined by fixing the signaling rate accordingly. A downside of the symmetry of FTN coding, however, is that all bits are coded, in contrast to standard multilevel approaches. The gap to capacity is determined, for FTN signaling, by the effectiveness of the underlying binary code. Thus, if the binary code achieves a certain *fraction* of capacity, so will the overall rate, while the *gap*-to-capacity grows with the SNR. Capacity approaching FTN coding further suffers from the need for interleaving.

References

[1] O. Etesami, , M. Molkaraie, and A. Shokrollahi, *Rateless codes on symmetric channels*, Int. Symp. Inform. Theory (ISIT2004), Chicago, Illinois, June, 2004, p. 38.

[2] G. D. Forney, Jr., *Shannon meets Wiener II: On MMSE estimation in successive decoding schemes.*, 42st Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois, Oct. 2004.

[3] T. Guess and M. K. Varanasi, *An information-theoretic framework for deriving canonical decision-feedback receivers in gaussian channels*, IEEE Trans. Information Theory (to appear).

[4] E. A. Lee and D. G. Messerschmitt, *Digital communication*, Kluwer, 2nd ed., 1994.

[5] R. Palanki and J.S. Yedidia, *Rateless codes on noisy channels*, Int. Symp. Inform. Theory (ISIT2004), Chicago, Illinois, June, 2004, p. 37.

[6] A. Shokrollahi, *Raptor codes*, (preprint, 2002).

[7] N. Shulman, *Universal channel coding*, Ph.D. thesis, Tel-Aviv University, 2004.