# Rateless Space-Time Coding

Uri Erez and G. W. Wornell
EECS, MIT, Cambridge, MA.
Email: {uri,gww}@allegro.mit.edu.

Mitchell D. Trott
HP Labs, Palo Alto, CA
Email: 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. In particular, rateless codes are potentially attractive for wireless communication.**

**In a recent work, a rateless coding scheme was proposed for the AWGN channel, based on layering, repetition and random dithering.**

**We extend this scheme to multiple-input single-output (MISO) Gaussian channels. We show that the rate loss associated with orthogonal design space-time codes may be alleviated by layering and dithering, very similar to the rateless approach for the AWGN channel. We then combine the two schemes and arrive at a close-to-capacity rateless code for MISO channels. The required complexity depends on the fraction of capacity that is targeted, is linear in the capacity of the channel and does not depend on the number of transmit antennas. Furthermore, the coding scheme uses only one base AWGN code.**

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 MISO Gaussian channels.

From a purely information theoretic perspective, the problem of variable rate transmission is by now well understood; see, e.g., [5] for a comprehensive treatment. Indeed, for classes of channels having one maximizing input distribution, a codebook drawn independently and identically distributed (i.i.d. ) at random according to the capacity-achieving input distribution will be good 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 [4], 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 [3], [2]. 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.

In a recent work [1], it was shown that 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 [5], for the AWGN channel there is a very simple way to combine the received repeated blocks: maximal ratio combining (MRC). Ultimately, the rateless codes are obtained combining this idea with dithering and a superposition coding strategy to obtain many low SNR channels from one higher SNR one.

In recent years there has been great progress in finding effective coding schemes for transmission over MISO channels, often referred to as space-time coding. An attractive approach is transmission using orthogonal design space-time block codes (OSTBC) [6].

Orthogonal Design Space-Time codes, using only linear pre/post-processing, convert the MISO channel into an AWGN channel. However (except for the $2 \times 1$ case), this comes at the cost of a reduced symbol rate (bandwidth loss). It turns out that dithered layering offers a way around this loss. In fact, the same approach may be used to allow for improved space-time coding and for rateless transmission. Moreover, the two goals may be achieved simultaneously.

## I. Rateless Transmission over AWGN channels

When a (Gaussian) codeword $\mathbf{x}$ (of length $n$) is repeated $r$ times over an AWGN channel, the resulting mutual information per symbol is

$$I_{\text{rep}}(\text{SNR}) = \frac{1}{r \cdot n} I(\mathbf{x}; \mathbf{x}+\mathbf{z}_1, \mathbf{x}+\mathbf{z}_2, \ldots, \mathbf{x}+\mathbf{z}_r) \quad (1)$$

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

On the other hand, when transmitting $r$ independent (Gaussian) codewords, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$, the mutual information is

$$I_{\text{ind}}(\text{SNR}) = \frac{1}{n} I(\mathbf{x}_1; \mathbf{x}_1 + \mathbf{z}_1) \quad (3)$$

$$= \frac{1}{2} \log(1 + \text{SNR}). \quad (4)$$

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

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., it is known 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

$$C^* = \frac{1}{2}\log(1+\text{SNR}^*). \tag{5}$$

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

$$\frac{1}{2}\log(1+\text{SNR}(r)) = \frac{C^*}{r}, \tag{6}$$

and let $N(r)$ be the corresponding noise power, i.e., $N(r) = P/(e^{2C^*/r}-1)$. Note that $\text{SNR}(1) = \text{SNR}^*$. Thus, if $\text{SNR} = \text{SNR}(r)$, then in order to send a message, transmission needs to be $r$ times longer than when $\text{SNR} = \text{SNR}^*$.

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.,

$$\text{SNR}_{l,i}(r) = \frac{P_l(i)}{\sum_{k<l}P_k(i)+N(r)}. \tag{7}$$

The corresponding capacity of the layer is

$$C_{l,i}(r) = \frac{1}{2}\log\left(1+\text{SNR}_{l,i}(r)\right). \tag{8}$$

### A. Time-varying power allocation

We next show that there is a power allocation $P_l(i)$ 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. This is formalized in the following

*Lemma 1:* There exists a power allocation $P_l(i)$, $l = 1,\dots,L$, $i = 1,\dots,\infty$, such that for every $r$,

$$\sum_{i=1}^{r} C_{l,i}(r) = \frac{C^*}{L}. \tag{9}$$

Specifically, the power allocation may be solved by the following recursion on $r$. For $r = 1,\dots,\infty$ do:

1) Update noise level:

$$N(r) = P/(e^{2C^*/r}-1). \tag{10}$$

2) Necessary incremental rate:
For $l = 1,\dots,L$, compute

$$\Delta_l(r) = \frac{C^*}{L} - \sum_{i=1}^{r-1} C_{l,i}(r). \tag{11}$$

3) Power allocation:
For $l = L,\dots,1$, assign the power

$$P_l(i) = \left(N(r) + \sum_{k=l+1}^{L} P_k(i)\right)\cdot\left(e^{2\Delta_l(r)}-1\right). \tag{12}$$

*Proof:* The proof is by induction on $r$. We have for any $l = 1,\dots,L$,

$$\sum_{i=1}^{r-1} C_{l,i}(r-1) = \frac{C^*}{L}. \tag{13}$$

From the definition of $C_{l,i}(r)$ it follows that $c_{l,i}(r)$ is monotonically decreasing in $r$. It follows that $\Delta_l(r)$ is positive for any $l$. Note that since $\sum_{i=1}^{L} P_l(i) = P$, for any $r$ and $i$, we have

$$\sum_{l=1}^{L} C_{l,i}(r) = \frac{C^*}{r} = C(r), \tag{14}$$

regardless of the power allocation. It follows that

$$\begin{aligned}\sum_{l=1}^{L}\Delta_l(r) &= C^* - \sum_{i=1}^{r-1}\sum_{l=1}^{L} C_{l,i}(r) &(15)\\ &= C(r). &(16)\end{aligned}$$

Note that for a Gaussian multiple access channel with $L$ users (layers) with sum power $P$, any rate vector $(R_1,\dots,R_L)$ such that $R_i \geq 0$ for all $i$ and $\sum_{i=1}^{L} R_i = C(P)$ is achievable. Thus, it follows form (14) and (16) that (12) will indeed yield a solution with nonnegative powers $P_l(r)$. ■

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.

### B. Dithered repetition transmission

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

$$\mathbf{x}(i) = \sum_{l=1}^{L}\mathbf{x}_l(i)\odot\mathbf{d}_l(i) \tag{17}$$

where $\odot$ denotes component-wise multiplication.

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

$$\mathbf{y}_l = \sum_{i=1}^{r}\alpha_l(i)\cdot\mathbf{d}_l(i)\odot\frac{\mathbf{y}(i)-\sum_{k>l}\sqrt{P_k(i)}\hat{\mathbf{x}}_k\odot\mathbf{d}_k(i)}{\sqrt{P_l(i)}}, \tag{18}$$

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

$$\mathbf{y}_l = \mathbf{x}_l + \sum_{i=1}^{r} \alpha_l(i) \cdot \mathbf{z}_l(i) \triangleq \mathbf{x}_l + \mathbf{z}_l, \tag{19}$$

where

$$\mathbf{z}_l(i) = \frac{1}{\sqrt{P_l(i)}} \cdot \left( \sum_{k<l} \mathbf{x}_k(i) \odot \mathbf{d}_l(i) \odot \mathbf{d}_k(i) + \mathbf{z}(i) \right). \tag{20}$$

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

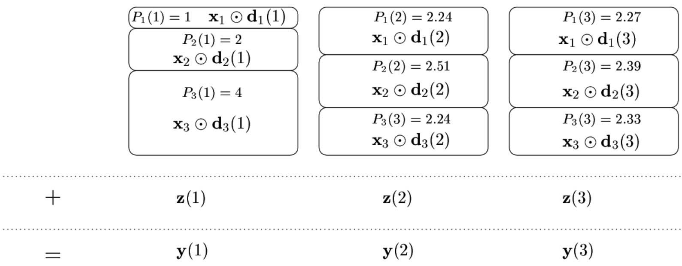


Fig. 1. Example: Dithered repetition scheme with time-varying power allocation, where $P = 7$ and $C^* = 1.5$bits.

By (8), (9) and since $x \geq \log(1+x)$,

$$\sum_{i=1}^{r} \frac{1}{2}\mathrm{SNR}_{l,i}(r) \geq \sum_{i=1}^{r} \frac{1}{2}\log(1+\mathrm{SNR}_{l,i}(r)) = \frac{C^*}{L}. \tag{21}$$

It follows that the accumulated SNR in each layer, i.e., the SNR in channel (19) satisfies

$$\mathrm{SNR}_l \geq \frac{2C^*}{L}. \tag{22}$$

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

$$R \geq \frac{1}{2}\log\left(1 + \frac{2C^*}{L}\right). \tag{23}$$

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

$$\text{efficiency} = \frac{L \cdot R}{C^*} \geq \frac{2R}{e^{2R}-1} \geq 1 - R. \tag{24}$$

This implies for instance that to obtain 90% of capacity requires a code of rate roughly $1/7$. We note that this efficiency bound holds uniformly, regardless of the number of incremental redundancy blocks or the number of layers. It requires only that the block length of the base code be long enough that mutual information is a reasonable indicator of performance. In practice, when the number of layers is sufficiently large such that the SNR per layer is low, a binary code may be used instead of a Gaussian codebook.

## II. Capacity Approaching Codes for MISO Channels

We have seen a method to obtain rateless codes for an AWGN channel. A potential application of such codes could be for transmission over a wireless channel with a single antenna at the transmitter as well as at the receiver. A natural question that arises is whether one can derive similar coding techniques also for multiple antenna systems. In the sequel we show that this may indeed be achieved for MISO channels.[2] In particular, in this section we begin by showing how to approach an arbitrarily high fraction of the capacity of a MISO channel using a good low-SNR AWGN code. Then in Section III we show how to combine these results with the rateless coding results of Section I to obtain rateless codes for the MISO channel.

A power-constrained Gaussian MISO system with $M$ transmit antennas is described by,

$$y = \sum_{i=1}^{M} h_i x_i + z, \tag{25}$$

where $h_i \in \mathcal{C}$ and $z \sim \mathcal{CN}(0, N)$, and the input satisfies the power constraint $E\|\mathbf{x}\|^2 \leq P$. We consider a coherent static channel where the channel gains $h_i$ are constant and are known to the receiver but not to the transmitter. The white-input capacity of the channel is

$$C = \frac{1}{2}\log\left(1 + \frac{\|\mathbf{h}\|^2}{M}\mathrm{SNR}\right), \tag{26}$$

where $\mathrm{SNR} = P/N$, and where capacity is measured in nats per *real* dimension.

We would like to first reduce the MISO channel into a scalar channel by pre/post-processing, allowing then to use the rateless coding techniques developed in the previous section for the scalar AWGN channel.

We first address the first goal. Specifically, we wish to obtain a (simple/efficient) method to convert the channel (25) into an AWGN channel of the form,

$$y = \frac{\|\mathbf{h}\|}{\sqrt{M}} x + z, \tag{27}$$

where $\mathbf{h}$, $x$ and $z$ are as above.

The simplest way to convert the MISO channel into an AWGN channel that depends only on $\|\mathbf{h}\|$ is by using repetition. That is one may use $M$ symbol durations to send a single symbol, at each time instant sending it over a different antenna. At the receiver we form the average of the matched-filtered output of the received $M$ symbols. The resulting channel is,

$$\begin{aligned} y &= \frac{1}{\|\mathbf{h}\|}\sum_{i=1}^{M} h_i^* \left(h_i x + z_i\right) && (28) \\ &= \|\mathbf{h}\| x + \frac{1}{\|\mathbf{h}\|}\sum_{i=1}^{M} h_i^* z_i && (29) \\ &= \|\mathbf{h}\| x + z, && (30) \end{aligned}$$

[2]We note that rateless transmission over SIMO channels is straightforward.

where $\mathbf{z} \sim \mathcal{CN}(0, N)$. Thus, repetition boosts the SNR by a factor of $M$, at the cost of an $M$-fold reduction in bandwidth. The corresponding capacity (normalized per real transmitted dimension) is,

$$C^{\text{rep}} = \frac{1}{2M} \log(1 + \|\mathbf{h}\|^2 \text{SNR}). \tag{31}$$

Comparing (31) and (26), it is clear that repetition performs very poorly (trading bandwidth for SNR) at high SNR. Nonetheless, at low SNR, the loss vanishes. This is very similar to the behavior of repetition coding over an AWGN channel as observed in Section I. Repetition is very inefficient unless the SNR is extremely low. We choose to employ Orthogonal Design Space-Time block codes to overcome this. OSTB coding, much like repetition coding, suffers from a reduction in rate. Nevertheless, with an OSTBC, this "bandwidth loss" is bounded above by a factor of two, for any number of transmit antennas. This loss is then alleviated by using layering.

We illustrate the scheme for the case of a $4 \times 1$ MISO channel, noting that it may be generalized. A rate $1/2$ OSTBC is used in the example as rate $1/2$ OSTBCs exist for any number of transmit antennas. Transmission is done in groups of eight channel symbols over which four information symbols $x_1, x_2, x_3, x_4$ are sent. The transmitter sends the followings,

$$X = \begin{pmatrix} x_1 & -x_2 & -x_3 & -x_4 & x_1^* & -x_2^* & -x_3^* & -x_4^* \\ x_2 & x_1 & x_4 & -x_3 & x_2^* & x_1^* & x_4^* & -x_3^* \\ x_3 & -x_4 & x_1 & x_2 & x_3^* & -x_4^* & x_1^* & x_2^* \\ x_4 & x_3 & -x_2 & x_1 & x_4^* & x_3^* & -x_2^* & x_1^* \end{pmatrix}^T \tag{32}$$

The receiver gets,

$$\mathbf{y} = \mathbf{h}^T X + \mathbf{z} \tag{33}$$

Equation (32) may be rewritten as,

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5^* \\ y_6^* \\ y_7^* \\ y_8^* \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 \\ h_2 & -h_1 & h_4 & -h_3 \\ h_3 & -h_4 & -h_1 & h_2 \\ h_4 & h_3 & -h_2 & -h_1 \\ h_1^* & h_2^* & h_3^* & h_4^* \\ h_2^* & -h_1^* & h_4^* & -h_3^* \\ h_3^* & -h_4^* & -h_1^* & h_2^* \\ h_4^* & h_3^* & -h_2^* & -h_1^* \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5^* \\ z_6^* \\ z_7^* \\ z_8^* \end{pmatrix}. \tag{34}$$

Redefining $\mathbf{y}$ to be the vector on the left hand side of (34), $H$ the matrix and $\mathbf{z}$ the noise vector (with the last four entries conjugated), we have

$$\mathbf{y} = H\mathbf{x} + \mathbf{z}, \tag{35}$$

where $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$. Note that the columns of $H$ are orthogonal and we have

$$H^* H = \begin{pmatrix} 2\|\mathbf{h}\|^2 & 0 & 0 & 0 \\ 0 & 2\|\mathbf{h}\|^2 & 0 & 0 \\ 0 & 0 & 2\|\mathbf{h}\|^2 & 0 \\ 0 & 0 & 0 & 2\|\mathbf{h}\|^2 \end{pmatrix}. \tag{36}$$

The receiver uses a matched filter to obtain (again redefining $\mathbf{y}$)

$$\mathbf{y} = \frac{1}{\sqrt{2}\|\mathbf{h}\|} H^* \left( H\mathbf{x} + \mathbf{z} \right). \tag{37}$$

Therefore, the OSTBC scheme converts the MISO channel into four parallel AWGN channels

$$y_i = \sqrt{2}\|\mathbf{h}\| x_i + z_i, \quad i = 1, 2, 3, 4, \tag{38}$$

using *eight* symbol durations, and where with abuse of notation $z_i$ denotes the resulting noise (which is a linear combination of the $z_j$ in (34)). Note that $z_i \sim \mathcal{CN}(0, N)$ and is i.i.d. . This is analogous to repetition with a factor of two.

We next introduce dithering. Let $D_{\text{col}}$ and $D_{\text{row}}$ be two $8 \times 4$ dither matrices,

$$D_{\text{row}} = \begin{pmatrix} d^1 & d^2 & d^3 & d^4 & d^1 * & d^{2*} & d^{3*} & d^{4*} \\ d^2 & d^1 & d^4 & d^3 & d^{2*} & d^{1*} & d^{4*} & d^{3*} \\ d^3 & d^4 & d^1 & d^2 & d^{3*} & d^{4*} & d^{1*} & d^{2*} \\ d^4 & d^3 & d^2 & d^1 & d^{4*} & d^{3*} & d^{2*} & d^{1*} \end{pmatrix}^T \tag{39}$$

$$D_{\text{col}} = \begin{pmatrix} d_1 & d_2 & d_3 & d_4 & d_5^* & d_6^* & d_7^* & d_8^* \\ d_1 & d_2 & d_3 & d_4 & d_5^* & d_6^* & d_7^* & d_8^* \\ d_1 & d_2 & d_3 & d_4 & d_5^* & d_6^* & d_7^* & d_8^* \\ d_1 & d_2 & d_3 & d_4 & d_5^* & d_6^* & d_7^* & d_8^* \end{pmatrix}^T \tag{40}$$

where the $d_i$ and $d^j$ are all random and independent phases. Let $D = D_{\text{row}} \odot D_{\text{col}}$. We send, $X \odot D$ over the channel. The receiver signal is,

$$\mathbf{y} = H^d \mathbf{x} + \mathbf{z}, \tag{41}$$

where

$$H^d = \begin{pmatrix} h_1 \cdot d_1 \cdot d^1 & h_2 \cdot d_1 \cdot d^2 & h_3 \cdot d_1 \cdot d^3 & h_4 \cdot d_1 \cdot d^4 \\ h_2 \cdot d_2 \cdot d^1 & -h_1 \cdot d_2 \cdot d^2 & h_4 \cdot d_2 \cdot d^3 & -h_3 \cdot d_2 \cdot d^4 \\ h_3 \cdot d_3 \cdot d^1 & -h_4 \cdot d_3 \cdot d^2 & -h_1 \cdot d_3 \cdot d^3 & h_2 \cdot d_3 \cdot d^4 \\ h_4 \cdot d_4 \cdot d^1 & h_3 \cdot d_4 \cdot d^2 & -h_2 \cdot d_4 \cdot d^3 & -h_1 \cdot d_4 \cdot d^4 \\ h_1^* \cdot d_5 \cdot d^1 & h_2^* \cdot d_5 \cdot d^2 & h_3^* \cdot d_5 \cdot d^3 & h_4^* \cdot d_5 \cdot d^4 \\ h_2^* \cdot d_6 \cdot d^1 & -h_1^* \cdot d_6 \cdot d^2 & h_4^* \cdot d_6 \cdot d^3 & -h_3^* \cdot d_6 \cdot d^4 \\ h_3^* \cdot d_7 \cdot d^1 & -h_4^* \cdot d_7 \cdot d^2 & -h_1^* \cdot d_7 \cdot d^3 & h_2^* \cdot d_7 \cdot d^4 \\ h_4^* \cdot d_8 \cdot d^1 & h_3^* \cdot d_8 \cdot d^2 & -h_2^* \cdot d_8 \cdot d^3 & -h_1^* \cdot d_8 \cdot d^4 \end{pmatrix}$$

Note that we still have $H^{d^*} H^d = 2\|\mathbf{h}\|^2 \cdot I$.

Consider now a dithered layered transmission, where we send

$$X = X_1 \odot D_1 + \cdots + X_L \odot D_L, \tag{42}$$

and where the matrices $X_l$ are generated from the information vectors $\mathbf{x}_l = (x_{l,1}, x_{l,2}, x_{l,3}, x_{l,4})$ according to (32) and where the matrices $D_l$ are generated according to (39) and (40) independently. We allocate power $P_l$ to layer $l$ so that $E\|\mathbf{x}_l\|^2 \leq P_l$ where $\sum_{l=1}^{L} P_l = P$. As we distribute the power evenly among the antennas, every symbol in layer $l$ will have an average power $P_l/4$. The received signal (after a front end conjugation of the last four symbols) is

$$\mathbf{y} = H_1^d \cdot \mathbf{x}_1 + \cdots + H_L^d \cdot \mathbf{x}_L + \mathbf{z}. \tag{43}$$

Note that while the four symbols are being transmitted through orthogonal effective channels, this is not the case between different layers.

We first decode layer $L$. The receiver uses a matched filter to obtain,

$$\begin{aligned} \hat{\mathbf{x}}_L &= \frac{1}{\sqrt{2}\|\mathbf{h}\|}(H_L^d)^*\mathbf{y} && (44)\\ &= \sqrt{2}\|\mathbf{h}\|\mathbf{x}_L && (45)\\ &+ \frac{1}{\sqrt{2}\|\mathbf{h}\|}\sum_{l=1}^{L-1}(H_L^d)^* H_l^d \cdot \mathbf{x}_l && (46)\\ &+ \frac{1}{\sqrt{2}\|\mathbf{h}\|}(H_L^d)^*\mathbf{z} && (47)\\ &= \sqrt{2}\|\mathbf{h}\|\mathbf{x}_L + \mathbf{z}^{\text{int}} + \mathbf{z}^{\text{G}}. && (48) \end{aligned}$$

It is not hard to see that for $i = 1, \ldots 4$,

$$\begin{aligned} E\left[|z_i^{\text{int}}|^2\right] &= \frac{1}{2\|\mathbf{h}\|^2}\sum_{l=1}^{L} P_l/4 \cdot 2\sum_{i,j}|h_i|^2|h_j|^2 && (49)\\ &= \|\mathbf{h}\|^2\sum_{l=1}^{L-1} P_l/4 && (50) \end{aligned}$$

This is due to the fact that the magnitudes of the entries of the matrix $H^d$ form a matrix such that both rows and columns are permutations of each other, and that due to the dithering all the interference components add *non-coherently*. Equation (50) is key as without dithering the power of the interference would be twice as large. We also have for $i = 1, \ldots 4$,

$$E\left[|z_i^{\text{G}}|^2\right] = N. \qquad (51)$$

Thus, the capacity of the $L$-th layer satisfies

$$R_L \geq \frac{1}{2}\cdot\frac{1}{2}\log\left(1 + 2\cdot\frac{\|\mathbf{h}\|^2 P_L/4}{\|\mathbf{h}\|^2\sum_{j=1}^{L-1}P_j/4 + N}\right), \qquad (52)$$

where the first factor of half is due to the bandwidth reduction. The inequality is actually strict since the noise is not Gaussian due to the interference component.

We then "strip" $\mathbf{x}_L$ from the received vector and proceed to decode layer $L-1$ and so on. Thus, the capacity for layer $l$ satisfies,

$$R_l \geq \frac{1}{4}\log\left(1 + 2\cdot\frac{\frac{\|\mathbf{h}\|^2}{4}P_l}{\frac{\|\mathbf{h}\|^2}{4}\sum_{j=1}^{l-1}P_j + N}\right). \qquad (53)$$

With the same power allocation the capacity of the original MISO channel is

$$\begin{aligned} C &= \sum_{l=1}^{L} C_l && (54)\\ &= \sum_{l=1}^{L}\frac{1}{2}\log\left(1 + \frac{\frac{\|\mathbf{h}\|^2}{4}P_l}{\frac{\|\mathbf{h}\|^2}{4}\sum_{j=1}^{l-1}P_j + N}\right) && (55)\\ &= \frac{1}{2}\log\left(1 + \frac{\|\mathbf{h}\|^2}{4}\frac{P}{N}\right) && (56) \end{aligned}$$

We may allocate the power so that every layer (subject to the decoder order) has the same SNR and correspondingly the same lower bound on the achievable rate, i.e., so that $R_l = R$. Define the efficiency of the scheme by

$$\text{efficiency} = \frac{RL}{C}. \qquad (57)$$

Equations (53) and (55) may be solved for $C_l$ as a function of $R_l$, yielding

$$\text{efficiency} \geq \frac{R}{\frac{1}{2}\log\left(\frac{e^{4R}+1}{2}\right)}. \qquad (58)$$

This implies for example that to obtain 86% of capacity requires a code of rate roughly $1/6$. The number of layers needed is a function of the capacity of the channel and the targeted efficiency but not of the number of antennas.

## III. Rateless Space-Time Codes

The layered OSTBC scheme may be combined with rateless coding (as described in Section I). As the MISO channel has been reduced into a scalar one, we may simply concatenate the two schemes. That is, given a rateless code, we group together four consecutive symbols of each layer and send them via dithered OSTBC as described in (42). The resulting efficiency may be lower bounded by combining (24) and (58). In effect, we substitute $C_l = \frac{1}{2}\log\left(\frac{e^{4R}+1}{2}\right)$ for $R$ in (24) and obtain,

$$\text{efficiency} \geq \frac{2\log\left(\frac{e^{4R}+1}{2}\right)}{e^{4R}-1}. \qquad (59)$$

## References


[1] U. Erez, G. Wornell, and M. Trott, *Coding for Faster-than-Nyquist signaling: The merits of a regime change*, 42nd Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois, Sept. 29-Oct. 1, 2004.

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

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

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

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

[6] V. Tarokh, N. Sheshadri, and A. R. Calderbank, *Space-time block codes form orthogonal designs*, IEEE Trans. Information Theory **IT-45** (July 1999), 1456–1467.