

Research Article

Generalized Superposition Modulation and Iterative Demodulation: A Capacity Investigation

Christian Schlegel, Marat Burnashev, and Dmitri Truhachev

Department of Computing Science, University of Alberta, Edmonton, AB, Canada T6G 2E8

Correspondence should be addressed to Christian Schlegel, schlegel@ualberta.ca

Received 1 March 2010; Revised 30 June 2010; Accepted 12 August 2010

Academic Editor: Peter Hoher

Copyright © 2010 Christian Schlegel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modulation with correlated signal waveforms is considered. Such correlation arises naturally in a number of modern communications systems and channels, for example, in code-division multiple-access (CDMA) and multiple-antenna systems. Data entering the channel in parallel streams either naturally or via inverse multiplexing is transmitted redundantly by adding additional signal waveforms populating the same original time-frequency space, thus not requiring additional bandwidth or power. The transmitted data is spread over a frame of N signaling intervals by random permutations. The receiver combines symbol likelihood values, calculates estimated signals and iteratively cancels mutual interference. For a random choice of the signal waveforms, it is shown that the capacity of the expanded waveform set is nondecreasing and achieves the capacity of the Gaussian multiple access channel as its upper limit when the number of waveforms becomes large. Furthermore, it is proven that the iterative demodulator proposed here can achieve a fraction of 0.995 or better of the channel capacity irrespective of the number of transmitted data streams. It is also shown that the complexity of this iterative demodulator grows only linearly with the number of data streams.

1. Introduction

Modulation is the process of injectively mapping elements of a discrete set, called the messages, onto functions of time, called the signals, for the purpose of information transmission. The signals form a (finite-dimensional) Hilbert space, called the signal space. Geometric representations of the signals are often called signal constellations. Basic modulation methods prefer the use of orthogonal bases of the signal space as the signals themselves, since demodulation can be accomplished by projection onto these bases. For example, equidistant m -ary pulse-amplitude modulation (PAM) uses discrete amplitudes on each basis [1].

Signals experience distortion during transmission which is modeled probabilistically, mainly due to the addition of noise. The received signals are therefore no longer identical with the transmitted signals. The demodulation problem is that of mapping a received signal back to a message such that the probability of the demodulated message not equalling

the original transmitted message is minimized. Under the assumption of additive white Gaussian noise, picking the message whose signal is closest to the received signal using the natural Euclidean distance metric is optimal (if noise is correlated, for example, a generalized metric needs to be used) [1]. This is referred to as maximum-likelihood (ML) decoding since it minimizes the message error probability.

However, ML decoding quickly becomes practically infeasible by the “curse of combinatorics,” and other methods are needed to be considered. Shannon [2] showed that every transmission channel has a maximum possible transmission rate which it can support, called the Shannon capacity, and that there exist coding and decoding methods which can operate to within ϵ of this capacity at arbitrarily low error rates. Shannon’s nonconstructive proofs did not require ML decoding, opening the door to possibly low-complexity capacity-achieving signaling methods. Unfortunately, to achieve capacity requires continuous input alphabets which is highly impractical. Discrete modulations,

mapped onto orthogonal bases, such as PAM, cannot achieve the Shannon capacity on the Gaussian channel. Certain high-dimensional discrete constellations, such as lattices, have been reported to achieve capacity, but in many ways their regular (discrete) structure is lost in the process [3].

In this paper we pursue another approach, abandoning the use of orthogonal bases as signals. In many practical situations the signals utilized are correlated, either by design, or by the effects that occur during transmission. An example of the former is (random) code-division multiple-access (CDMA) [4], and an example of the latter is multiple-antenna transmission (MIMO) [5]. In both cases the signals are densely correlated-which makes efficient demodulation extremely difficult. If the correlation pattern is sparse, that is, if any given signal waveform interferes with only a few other (neighboring) signal waveforms, sequence detection algorithms like the Viterbi algorithm can be used efficiently. A number of modulation methods based on superposition of individual data streams have been proposed (see [6–8]). When a number of independent signals add up in the channel, they can sometimes be decoded sequentially. Onion-peeling decoding starts from the largest power signal, decodes it treating the rest of the signals as noise and subtracts the result from the composite received signal. The decoding then continues analogously with the second strongest signal to the weakest. A number of methods based on successive decoding have been proposed and studied for various types of signals (data streams), including binary [9]. Channel capacity can be approached in the case when powers and rates of the signals follow specific precise arrangements, which is, however, challenging to accomplish in practice.

In this paper we assume a random correlation among the signals by postulating that these signals correspond to random vectors in signal space. The CDMA and MIMO channels are practical examples of such random channels [5, 10, 11]. Transmission relies on repeating the symbols of a message with random delays. Each time the symbol is modulated onto a new signal. While this increases the number of signals utilized, it allows for a very efficient iterative demodulation method to be used. This iterative demodulator forms the first stage of a two-stage receiver, where the second stage is a conventional forward error control (FEC) decoder for individual (binary) data streams. That is, the iterative first stage efficiently separates the correlated data streams. Specific adaptations of generalized modulation have recently been proposed for both CDMA [12] and MIMO channels [13].

Our contributions in this paper are two-fold. First we show that using random signals incurs no capacity loss, and furthermore, that regular-spaced PAM-type modulation on these random signals can achieve the Shannon capacity. We then discuss transmission using redundant signaling and an iterative demodulation method for which we show that it can operate close to the Shannon capacity over the entire range of operational interest. In showing this, we will only assume that we have capacity-achieving binary error control codes available, a very reasonable assumption given the current state-of-the-art in error control coding [14].

2. Modulation

Generally, a discrete data stream \mathbf{d} is mapped onto signals from a finite set of such signals according to some mapping rule. In the ubiquitous pulse-amplitude (PAM) modulation, a discrete amplitude x_r for each value of d_r from \mathbf{d} is first selected then used to multiply the r th signal. Most commonly, one of 2^B amplitude levels is selected for each B -bit data symbol. In the case of 8-PAM, for example, with $B = 3$, the discrete equispaced amplitudes shown in Figure 1 are used on each signal. This signal constellation can be interpreted as the superposition of three simple binary constellations, where Bit 1 has 4 times the power of Bit 0, and Bit 2 has 16 times its power.

In general, any properly labeled 2^B -PAM modulation can be written as the superposition of B binary antipodal amplitudes, that is,

$$v = \sum_{j=0}^{B-1} 2^j b_j, \quad (1)$$

where $b_j \in \{-1, 1\}$. If an entire sequence \mathbf{v} of 2^B -ary PAM symbols is considered, it may be viewed as the superposition of B binary modulated data streams with powers $P_0, 4P_0, 16P_0, \dots, 4^{B-1}P_0$ on binary data streams which make up the PAM symbol sequence \mathbf{v} .

This viewpoint is quite productive in that it suggests a capacity-achieving demodulation method for large constellations based on cancellation. Consider the case where the highest-power uncanceled data stream considers all lower power data streams as noise. Its maximum rate is then given by the mutual information

$$C_j = I(b_j; y | b_{j+1}, \dots, b_{B-1}), \quad (2)$$

where y is the output of the channel. As long as the rate on the j th binary data stream $R_j < C_j$, it can be correctly decoded using a binary Shannon-capacity-achieving code. (While no class of binary codes with nonexponential decoding complexity exist which can provably achieve the capacity on a binary-input channel, codes which can achieve this capacity “practically” with “implementable” complexities have recently emerged from intense research. The most popular representatives are turbo codes and low-density parity-check codes. Both utilize iterative message passing decoding algorithms [14].) By virtue of (1), knowledge of b_{j+1}, \dots, b_{B-1} implies that these data streams can be canceled from the received signal, and C_j is the capacity of the j th binary data stream. This thought model leads to a successive decoding and canceling method which can achieve the mutual information rate

$$C_{\text{symmetric}} = I(v; y) = \sum_{j=0}^{B-1} C_j \quad (3)$$

by the chain rule of mutual information. $C_{\text{symmetric}}$ is of course not equal to the capacity of the additive white Gaussian noise channel $y = v + n$, since the input distribution of v is uniform, rather than being Gaussian distributed as

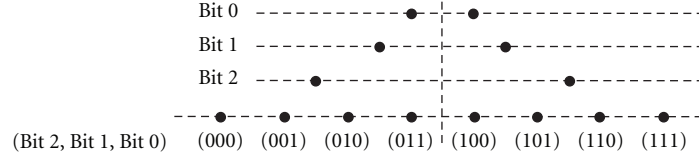


FIGURE 1: Pulse-amplitude modulation as superposition of binary antipodal modulation with geometric power distribution.

required to achieve the channel capacity. In fact, $C_{\text{symmetric}}$ loses the so-called shaping gain of 1.52 dB with respect to the capacity of the Gaussian channel [14].

3. Main Result

In this paper, we propose a generalized PAM modulation method which operates with random signals, rather than the orthogonal bases implied by the discussion in the previous section. We present a two-stage demodulator/decoder which remedies the difficulties of the onion-peeling method discussed above. Specifically, the demodulator consists of an iterative demodulator which operates in parallel and achieves a signal-to-noise ratio (SNR) improvement on each of the binary data streams. These are then decoded using external binary error control codes. The latency issue is basically confined to that of the parallel demodulator and that of the follow-up error control decoder. The iterative demodulator is based on cancellation. This means that its complexity, and that of the entire decoder scales linearly with the number of data streams.

Our main result is that we will prove that such an iterative demodulator/decoder can achieve a cumulative data rate R_d per dimension such that

$$R_d \geq 0.995C_{\text{GMAC}} - 0.54 \left[\frac{\text{bits}}{\text{dimension}} \right], \quad (4)$$

where C_{GMAC} is the Shannon capacity of the Gaussian multiple-access channel. That is, our system can achieve a fraction of 0.995 of channels information theoretic capacity, irrespective of system size.

In order for the iterative demodulator to function, we require that the number of signals in the signal space is increased, but not the power or spectral resources.

4. System Model

4.1. Signaling. We are considering communication of multiple data streams \mathbf{d}_k using random signals \mathbf{s}_k of dimension N . If N is sufficiently large, the number of useful signals is arbitrarily large. A set of K data symbols $d_{k,l}$ from the data streams \mathbf{d}_k is transmitted at each time interval l . There are basically two ways to do this. Conventionally each symbol $d_{k,l}$ is directly modulated onto an individual signal $\mathbf{s}_{k,l}$. In this paper, however, we propose an alternative where we duplicate each symbol $d_{k,l}$ M -fold. These duplicates are then modulated onto separate signals $\mathbf{s}_{k,\pi_{k,m}(l)}$ at M random time intervals within a certain signal block, where $\pi_{k,m}(l)$ is the random location within the block where the m th

copy of symbol $d_{k,l}$ is located. The function $\pi_{k,m}(l)$ is a permutation function with inverse $\pi_{k,m}^{-1}(l)$. Even though we have increased the number of signals by a factor M , scaling the power with M , and requiring that the signal set $\{\mathbf{s}_{k,\pi_{k,m}(l)}\}$ occupies the original N -dimensional signal space, this will not affect total power or the total spectrum utilization. (Another form of modulation based on randomly correlated signals called ‘‘partitioned transmission’’ has been recently proposed in [15]. Partition signalling creates redundancy and sparseness in transmitted data by partitioning K existing N -dimensional signal waveforms and permuting the resulting partitions. Generalized modulation relies on populating the signals space with additional N -dimensional signal waveforms. The latter gives an opportunity to create the required level M of redundancy independently of signal dimensionality N . Further, near capacity operation with generalized modulation does not require $N \rightarrow \infty$.) A diagram of this modulator is given in Figure 2.

We make the convenient, but in no way necessary assumption that the channel is block-synchronous, that is, that the signal waveforms at time interval l interfere only within that time interval, and that there is no correlation of signal waveforms between time intervals. With this we can write the channel in the linear matrix form

$$\mathbf{y}_l = \mathbf{S}_l \mathbf{W}^{1/2} \mathbf{x}_l + \eta_l, \quad (5)$$

where the $N \times KM$ matrix \mathbf{S}_l contains the signal vectors as columns. The capacity per dimension of this channel is well known [10] and is given by

$$C_s = \frac{1}{2N} \log \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{S} \mathbf{W} \mathbf{S}^T \right), \quad (6)$$

where

$$\mathbf{W} = \text{diag} \left(\frac{P_1}{M}, \frac{P_1}{M}, \dots, \frac{P_1}{M}, \frac{P_2}{M}, \dots, \frac{P_K}{M} \right) \quad (7)$$

is a $KM \times KM$ diagonal matrix with the powers used for transmission of the different signal vectors. We now assume that the signals $\mathbf{s}_{k,l}$ are chosen randomly from the signal space (the individual components $s_{k,l,n}$, $n = 1, \dots, N$, of signals $\mathbf{s}_{k,l}$ can, for example, be selected randomly out of the set $\{-1/\sqrt{N}, 1/\sqrt{N}\}$ picking each entry with probability 1/2. However, other random selections satisfying (8) are also possible) such that the mutual pairwise expected correlation between signals is

$$\mathbb{E} \left[\mathbf{s}_{j,l}^* \mathbf{s}_{j',l} \right]^2 = \frac{1}{N}; \quad j \neq j'. \quad (8)$$

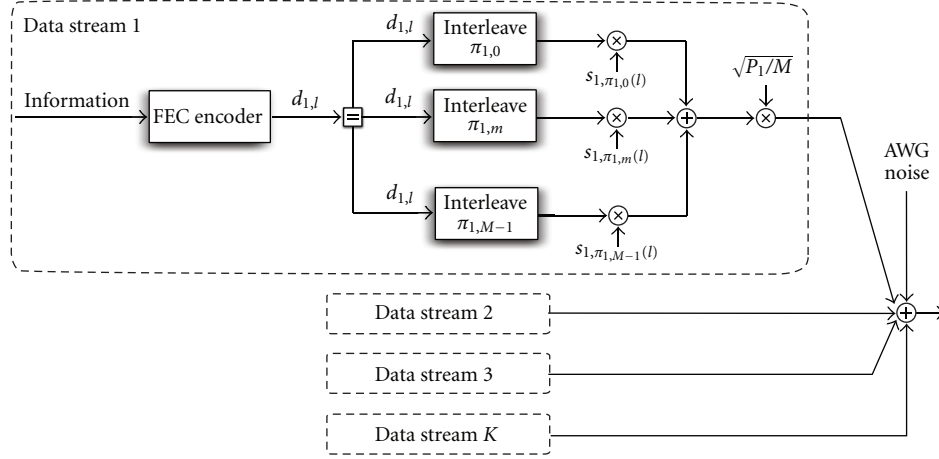


FIGURE 2: Modulator with superimposed binary data streams.

This model captures among others the random code-division multiple-access channel and the isotropic multiple-antenna channel model.

The capacity $C_{\bar{S}}$ of this random vector channel is given by the expectation over \mathbf{S} in (6). Using Jensen's inequality

$$C_{\bar{S}} \leq \mathbb{E}_S \left[\frac{1}{2N} \log \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{S} \mathbf{S}^T \right) \right] \quad (9)$$

with equality if and only if $\mathbf{W} = \mathbf{I}$ (see [16]). That is to say that an equal distribution of powers over the signals maximizes the capacity of the random vector channel (5).

We now investigate the information theoretic impact of increasing the signal population as proposed by the M -fold duplication. The following lemma addresses this issue.

Lemma 1. *Keeping the transmit power $\text{tr}(\mathbf{W}) = P$ constant, the capacity $C_{\bar{S}}$ as a function of K and N approaches the capacity of the Gaussian multiple-access channel in the limit, that is,*

$$C_{\bar{S}} \rightarrow C_{\text{GMAC}} = \frac{1}{2} \log \left(1 + \frac{\sum_{k=1}^K P_k}{N\sigma^2} \right). \quad (10)$$

It approaches this limit from below as $M, K \rightarrow \infty$, that is, $C_{\bar{S}} < C_{\text{GMAC}}$ for all $K/N < \infty$.

Proof. See Appendix A. \square

Lemma 1 reveals useful information in several ways. Firstly, it guarantees that the signaling strategy presented above, that is, the addition of extra random signal waveforms, incurs no capacity loss, and secondly, in the limit, arbitrary power assignments become capacity achieving, not only the equal power assignment.

4.2. Demodulation. The first stage of the demodulation process starts with matched filtering of the received signal with respect to each transmitted signal waveform in each time

interval. Given the received signal embedded in Gaussian noise as

$$\mathbf{y}_r = \sum_{k=1}^K \sum_{m=0}^{M-1} \sqrt{\frac{P_k}{M}} d_{k,\pi_{k,m}^{-1}(r)} \mathbf{s}_{k,\pi_{k,m}^{-1}(r)} + \mathbf{n}_r, \quad (11)$$

these matched filter outputs are given by

$$z_{k,l} = \mathbf{s}_{k,l}^* \cdot \mathbf{y}_r = \sum_{k'=1}^K \sum_{m'=0}^{M-1} \sqrt{\frac{P_{k'}}{M}} d_{k',\pi_{k',m'}^{-1}(r)} \mathbf{s}_{k,l}^* \cdot \mathbf{s}_{k',\pi_{k',m'}^{-1}(r)} + n_r, \quad (12)$$

$$l = \pi_{k,m}^{-1}(r), \quad l' = \pi_{k',m'}^{-1}(r),$$

where n_r is the sampled noise of variance σ^2 , and $\mathbf{s}_{k,l}^* \cdot \mathbf{s}_{k',l'} = \rho_{m,m',l,l'}^{k,k'}$ is the correlation value between the target signal and a given interfering signal. The matched filter outputs in (12) consist of $d_{k,l}$ and an interference and noise term, which is given by

$$I_{k,m,r} = \sum_{\substack{k'=1 \\ k' \neq k}}^K \sum_{m'=0}^{M-1} \sqrt{\frac{P_{k'}}{M}} \rho_{m,m',l,l'}^{k,k'} d_{k',\pi_{k',m'}^{-1}(r)} + n_r. \quad (13)$$

At this point the graphical illustration shown in Figure 3 may prove helpful, which shows how the different symbols and signals combine to generate the sequence of received signal vectors \mathbf{y}_r . Note that in the interference equation (13) above, self-interference is not included. Apart from unnecessarily complicating the notation, this self-interference is negligible as shown below. Furthermore, in many cases it is not difficult to ensure that the signal vectors used for the different signals from a user k impinging on channel symbol \mathbf{y}_r are orthogonal, that is,

$$\mathbf{s}_{k,\pi_{k,m}^{-1}(r)}^* \cdot \mathbf{s}_{k,\pi_{k,m'}^{-1}(r)} = 0, \quad \forall k, \quad (14)$$

and cause no self-interference. In [12], for example, different time intervals are used for the duplicate signals to accomplish this. The graphical representation reveals the similarity with graph-based error control codes, in particular with fountain

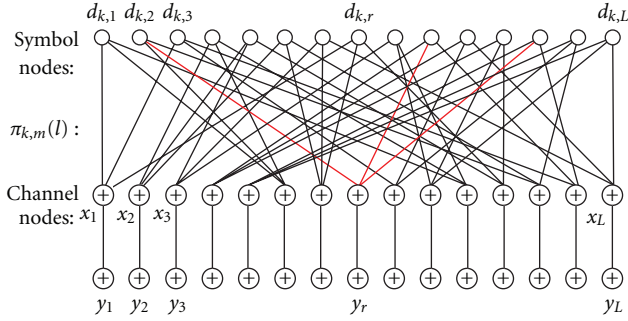


FIGURE 3: Signaling arrangement for the k th data stream. Symbols aggregating on the same channel node may use orthogonal signal waveforms. The actual received signal vectors \mathbf{y}_r are the superposition of K such data streams, causing the channel nodes to have large message degrees.

codes [17]. Consequently, we will explore a demodulation method based on message passing.

Iterative demodulation follows the following message-passing principle. At the channel nodes, updated matched filter output signals are computed at each iteration by subtracting interference to generate

$$z_{k,l,m}^{(i)} = \mathbf{s}_{k,l}^* \cdot \left(\mathbf{y}_r - \sum_{\substack{k'=1 \\ (k' \neq k)}}^K \sum_{m'=0}^{M-1} \sqrt{\frac{P_{k'}}{M}} \tilde{d}_{k',l,m'}^{(i-1)} \cdot \mathbf{s}_{k',l} \right), \quad (15)$$

where $\tilde{d}_{k',l,m'}$ is a soft symbol estimate of the m' th copy of $d_{k,l}$. Note that, following the extrinsic principle, the m different estimates for the same symbol are not necessarily identical (see below). These soft symbol estimates, in turn, are computed at the symbol nodes from the M matched filter signals for each copy of $d_{k,l}$. While $d_{k,l}$ can, in general, be any complex integer, we will concentrate on the basic binary case where $d_{k,l} \in \{-1, 1\}$. We will show later how to build larger modulation alphabets from this basic binary case using the binary decomposition of PAM signals.

In the binary case, the soft symbols are calculated as

$$\tilde{d}_{k,l,m}^{(i)} = \tanh \left(\sum_{\substack{m'=0 \\ (m' \neq m)}}^{M-1} \sqrt{\frac{P_k}{M}} \frac{z_{k,l,m'}^{(i)}}{\sigma_{k,i}^{(i)}} \right) \quad (16)$$

which is the optimal *local* minimum-variance estimate of $d_{k,l}$ given that interference and noise combined form a Gaussian random variable with power $\sigma_{k,i}^2$. The variance of the symbol estimates (16) will be required in the analysis in Section 5.

Defining this variance at iteration i as $\sigma_{d,k,i}^2 = E|d_k - \tilde{d}_{k,m}^{(i)}|^2$, and assuming that correlation between interference experienced by different replicas of the same symbol is negligible due to sufficiently large interleaving, it can be calculated adapting the development in [18] for CDMA as

$$\sigma_{d,k,i}^2 = E \left(1 - \tanh(\mu + \sqrt{\mu} \xi) \right)^2 = g(\mu), \quad \forall i, \quad (17)$$

where $\xi \sim \mathcal{N}(0, 1)$ and $\mu = (M-1)P_k/(M\sigma_i^2)$ from (16), and $\sigma_{k,i}^2 = \sigma_i^2$, for all k . The function $g(\mu)$ has no closed form, but the following bounds are quite tight [19]:

$$g(\mu) \leq 1(1 + \mu); \quad \mu < 1, \quad (18)$$

$$g(\mu) \leq \pi Q(\sqrt{\mu}); \quad \mu \geq 1, \quad (19)$$

where $Q(\cdot)$ is the complementary error function. The final output signal after I iterations is $z_{k,l}^{(I)} = \sum_{m=0}^{M-1} z_{k,l,m}^{(I)}$, which is passed to binary error control decoders for data stream k . The final signal-to-noise/interference ratio (SINR) of $z_{k,l}^{(I)}$ is what primarily matters for the error performance of these error control decoders.

After I detection iterations of the first stage the data is passed to the second stage of demodulation. The second stage of the reception is the error control decoding which is executed for each of the K data streams individually. SINR for data stream k equals P_k/σ_i^2 and it can be argued that the residual noise and interference is Gaussian [15]. Ultimately the information rate (i.e., the rate of the error control code) of stream k should satisfy

$$R_k \leq C_{\text{BIAWGN}} \left(\frac{P_k}{\sigma_i^2} \right) \quad (20)$$

for error-free decoding at the second stage. Here by $C_{\text{BIAWGN}}(x)$ we denote the capacity of the binary-input real-valued output AWGN channel with SNR x .

5. Generalized Modulation

The discussion above treated the case of binary modulation on the different signal waveforms, however, as illustrated in Section 2, we can create the regular-spaced PAM modulations with geometrically scaled binary modulations using powers

$$P_0 4^b, \quad 0 \leq b \leq B-1. \quad (21)$$

We assume that there are K_b data streams with powers $P_0 4^b$. Thus, the total number of streams equals $K = \sum_{b=0}^{B-1} K_b$.

Assuming large enough interleavers, the evolution of the interference in this iterative demodulator can be captured with a standard density evolution analysis. Since the average correlation between signal waveforms $E[(\rho_{m,m',l,l'}^{k,k'})^2] = 1/N$ (see (8)), the interference and noise on stream k is given by

$$\sigma_{k,i}^2 = \frac{1}{N} \sum_{\substack{k'=1 \\ (k' \neq k)}}^K P_{k'} \sigma_{d,k',i-1}^2 + \sigma^2 \leq \frac{1}{N} \sum_{k=1}^K P_k \sigma_{d,k,i-1}^2 + \sigma^2 = \sigma_i^2 \quad (22)$$

which is common to all streams. The upper bound in (22) contains the self-interference term for $k' = k$, which, however, becomes negligible as K and M grow. Using (17) in (22) and the PAM power distribution we obtain

$$\frac{\sigma_i^2}{P_0} = \sum_{b=0}^{B-1} K_b \frac{4^{b-1}}{N} g \left(\frac{M-1}{M} \frac{4^{b-1}}{\sigma_{i-1}^2/P_0} \right) + \frac{\sigma^2}{P_0}. \quad (23)$$

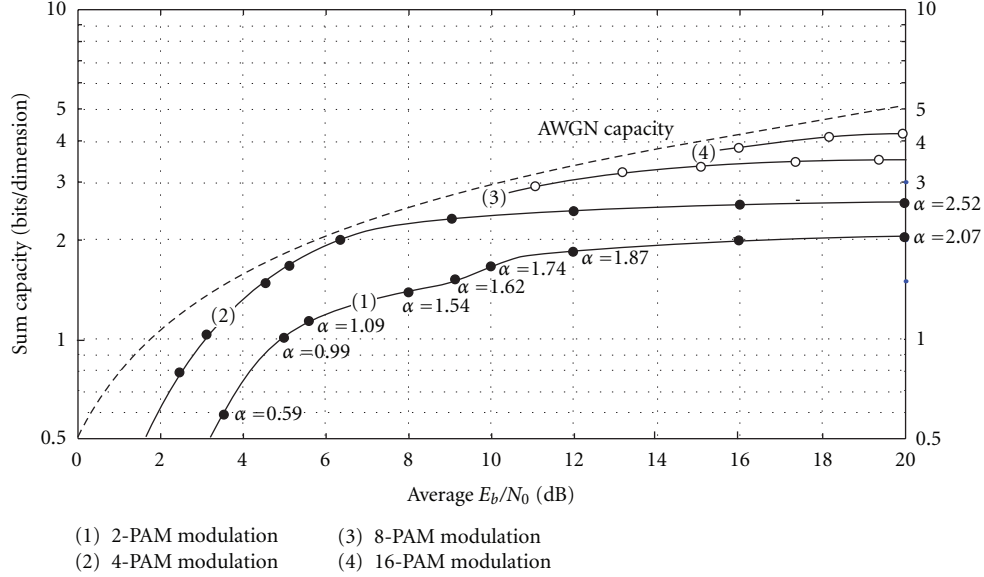


FIGURE 4: Achievable spectral efficiencies using iterative demodulation of various PAM constellations.

The next theorem proves that generalized PAM modulation used with the two-stage demodulation described above can closely approach the channel capacity.

Theorem 1. Consider generalized PAM modulation (21) with B levels and $K_b = 0.995N$ data streams per level for $b = 0, 1, \dots, B-1$ giving a total number of streams $K = 0.995BN$. One assumes that each data stream is encoded with a binary error control code which is capacity achieving on the binary-input AWGN channel, that is,

$$R_k = C_{\text{BIAWGN}} \left(\frac{P_k}{\sigma_\infty^2} \right), \quad (24)$$

and let $M \rightarrow \infty$. Then the resulting spectral efficiency per dimension

$$R_d = \frac{1}{N} \sum_{k=1}^K R_k \geq 0.995 C_{\text{GMAC}} - 0.54. \quad (25)$$

Proof. See Appendix B. \square

We note that for $B_1 < B_2$ the corresponding capacity approaching power profiles $P_0, P_04, \dots, P_04^{B_1-1}$ and $P_0, P_04, \dots, P_04^{B_2-1}$ coincide for $b \leq B_1 - 1$. The importance of these results is that new data streams can always be added without affecting decodability of the existing streams.

The gap between achieved spectral efficiency and the channel capacity can also be introduced in terms of average E_b/N_0 , instead of data stream power profile. Average E_b/N_0 for the power profile used in Theorem 1 can be upper bounded as

$$\frac{E_b}{N_0} = \frac{1}{2R_d} \eta \sum_{b=0}^{B-1} \gamma_0 4^b \leq \frac{2\eta(4^B - 1)}{3\eta(B - 1 + 0.6706)} \quad (26)$$

from (B.14), and therefore the corresponding capacity of AWGN channel $C(E_b/N_0)$, using

$$\frac{E_b}{N_0} = \frac{2^{2C(E_b/N_0)} - 1}{2C(E_b/N_0)} \quad (27)$$

can be upper bounded as

$$C(E_b/N_0) \leq B + 0.76. \quad (28)$$

As a result we obtain

$$\eta C \left(\frac{E_b}{N_0} \right) - R_d \left(\frac{E_b}{N_0} \right) \leq 1.09. \quad (29)$$

In Figure 4 we plot the achievable spectral efficiencies for the proposed generalized PAM modulation (21) for $B = 1, 2, 3, 4$ levels and assume ideal posterror control decoding with rates satisfying (24). Such performance can be closely approached with appropriate standard error control codes, which are very well developed for the binary case [20, 21]. We assume that $K_b = \alpha N/B$, for $b = 0, 1, 2, \dots, B-1$, where parameter $\alpha \in (0, \infty)$. Each curve corresponds to fixed B and plotted as a function of average E_b/N_0 which is in turn the function of α . We can observe that spectral efficiency of generalized PAM modulation exceeds the capacity of the same PAM modulation using orthogonal waveforms. This is because the number of allowable correlated signal waveforms K exceeds the number of available orthogonal dimensions N . This advantage is most noticeable for 2-PAM, where the maximum achievable capacity of 2.08 is more than twice the number of orthogonal dimensions. For higher PAM constellations, the capacity per level $\alpha_b = K_b/N = \alpha/B \rightarrow 1$ rapidly from above. Note that for $\alpha = \eta = 0.995$ the gap between the performance curves and the capacity curve satisfies (29). Specifically, point $\alpha = \eta$ for $B = 1$ gives $E_b/N_0 = 4.72$ dB, for $B = 2$ gives $E_b/N_0 = 7.74$ dB, for $B = 3$ gives $E_b/N_0 = 11.94$ dB, and for $B = 4$ gives $E_b/N_0 = 16.63$ dB.

6. Conclusions

We have presented and analyzed a two-stage iterative demodulation methodology for generalized PAM constellations using correlated random signals rather than the usual orthogonal bases. The method operates by introducing redundant duplicate copies of the data symbols modulated onto extra signals. An exponential power distribution, inherently present in PAM modulations, allows this two-stage iterative demodulator to achieve 99.5% of the Shannon capacity using binary capacity-achieving error control codes for each data stream. This generalized PAM modulation format was shown to approach the channel capacity over a wide range of operating SNRs, and can exceed the capacity of traditional PAM constellations on orthogonal signals.

Appendices

A. Proof of Lemma 1

Decompose the argument of (6) as

$$\begin{aligned} \det\left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{SWS}^T\right) &= \det\left(\left(1 + \frac{\sum_{k=1}^K P_k}{N\sigma^2}\right) \mathbf{I} + \mathbf{B}\right) \\ &= \left(1 + \frac{\sum_{k=1}^K P_k}{N\sigma^2}\right)^N \det(\mathbf{I} + \mathbf{F}), \end{aligned} \quad (\text{A.1})$$

where \mathbf{F} is the matrix of the off-diagonal elements with

$$F_{ij} = \kappa \sum_{m=1}^M \sum_{k=1}^K \frac{P_k}{M} b_{m,k}, \quad (\text{A.2})$$

where $\kappa = (\sigma^2 + \sum_{k=1}^K P_k)^{-1}$ and $b_{m,k} \in \{\pm 1\}$ with uniform probabilities. The entries B_{ij} have zero mean and variance

$$\text{var}(F_{ij}) = \frac{1}{M^2} \frac{\sum_{k=1}^K P_k^2}{\left(\sum_{k=1}^K P_k\right)^2}, \quad (\text{A.3})$$

which vanishes as (i) $M \rightarrow \infty$ or (ii) $K \rightarrow \infty$. Condition (ii), however, requires the Lindeberg condition to hold on the set $\{P_k\}$.

Using (i), or (ii), the elements in \mathbf{F} are sufficiently small to apply Jacobi's formula, that is,

$$\det(\mathbf{I} + \mathbf{F}) = (1 - \text{tr}(\mathbf{F})) + O(F_{ij}^2). \quad (\text{A.4})$$

Since $\text{tr}(\mathbf{F}) = 0$, and the second moment of F_{ij} vanishes, the limit value of the Lemma is proven.

Using Hadamard's inequality it is straightforward to show that

$$C_{\bar{\zeta}} < C_{\text{GMAC}}, \quad (\text{A.5})$$

and the limit is approached from below. While $\det(\mathbf{I} + \mathbf{F}) \rightarrow 1$ in probability, $C_{\bar{\zeta}} \rightarrow C_{\text{GMAC}}$ almost surely.

B. Proof of Theorem 1

Let us define $\gamma = P_0/\sigma_i^2$, $\gamma' = P_0/\sigma^2$, and $\gamma_\infty = P_0/\sigma_\infty^2$. Consider $B = \infty$ here for simplicity and define $\eta = 0.995$. Convergence defined by (23) (for σ_i^2 , $i = 0, 1, \dots$) follows from

$$1 > \sum_{b=0}^{\infty} \eta \gamma 4^b g(\gamma 4^b) + \frac{\gamma}{\gamma'}, \quad \text{for } \gamma \in (0, \gamma_\infty], \quad (\text{B.1})$$

and $K_b = \eta N$. Success of the demodulation stage happens if γ_∞ is close to γ' . This means that the interstream interference is canceled almost entirely. Let us choose a somewhat arbitrary lowest power P_0 such that $\gamma' = 4$ and prove that $\gamma_\infty > 1.79$.

Let us define the functions

$$t(x) = xg(x), \quad (\text{B.2})$$

$$f(\gamma) = \sum_{b=0}^{\infty} \eta t(\gamma 4^b) + \frac{\gamma}{\gamma'} = \sum_{b=0}^{\infty} \eta t(\gamma 4^b) + \frac{\gamma}{4}. \quad (\text{B.3})$$

The function $t(x)$ monotonically increases for $0 < x < x_0$ and monotonically decreases for $x > x_0$, where $x_0 \approx 1.508$. To find an upper bound on $f(\cdot)$, we consider the terms $t(\gamma 4^b)$ for very small and very large arguments separately, that is,

$$\begin{aligned} f(\gamma) &= \eta \sum_{b=0}^{\infty} t(\gamma 4^b) + \frac{\gamma}{4} \\ &= \eta \sum_{b \text{ s.t. } \gamma 4^b < A_1} t(\gamma 4^b) + \eta \sum_{b \text{ s.t. } \gamma 4^b > A_2} t(\gamma 4^b) \\ &\quad + \eta \sum_{b \text{ s.t. } A_1 \leq \gamma 4^b \leq A_2} t(\gamma 4^b) + \frac{\gamma}{4}. \end{aligned} \quad (\text{B.4})$$

Using the fact that $g(x) \leq 1$ for any x we obtain for any $b_1 \geq 0$

$$\sum_{b=0}^{b_1} t(\gamma 4^b) \leq \gamma \sum_{b=0}^{b_1} 4^b = \frac{\gamma(4^{b_1+1} - 1)}{3} < \frac{\gamma 4^{b_1+1}}{3}. \quad (\text{B.5})$$

From (19) we obtain

$$\begin{aligned} \sum_{b=b_2}^{\infty} t(\gamma 4^b) &\leq \gamma \pi \sum_{b=b_2}^{\infty} 4^b Q(2^b \sqrt{\gamma}) \\ &\leq \sqrt{\frac{\gamma \pi}{2}} \sum_{b=b_2}^{\infty} 2^b e^{-\gamma 4^b/2} \leq \sqrt{\frac{\gamma \pi}{2}} \frac{2^{b_2} e^{-\gamma 4^{b_2}/2}}{(1 - e^{-3\gamma 4^{b_2}/2})} \end{aligned} \quad (\text{B.6})$$

for b_2 such that $\gamma 4^{b_2} > 1$. The last inequality in (B.6) is computed by upper bounding the sum by the geometrical progression using the inequality

$$\frac{t(\gamma 4^{b_2+1})}{t(\gamma 4^{b_2})} = 2e^{-3\gamma 4^{b_2}/2} \leq e^{-3\gamma 4^{b_2}/2}. \quad (\text{B.7})$$

Using (B.5) we get for any A_1

$$\sum_{b \text{ s.t. } \gamma 4^b < A_1} t(\gamma 4^b) < \frac{4A_1}{3}. \quad (\text{B.8})$$

Analogously, (B.6) gives

$$\sum_{b \text{ s.t. } \gamma 4^b > A_2} t(\gamma 4^b) \leq \sqrt{\frac{\pi A_2}{2}} \frac{e^{-A_2/2}}{(1 - 2e^{-3A_2/2})}. \quad (\text{B.9})$$

By choosing $A_1 = 0.00003$ and $A_4 = 24$, we compute numeric values of the bounds (B.8) and (B.9) for the tails as

$$\sum_{b \text{ s.t. } \gamma 4^b < A_1} t(\gamma 4^b) < 4 \cdot 10^{-5}, \quad (\text{B.10})$$

$$\sum_{b \text{ s.t. } \gamma 4^b > A_2} t(\gamma 4^b) < 4 \cdot 10^{-5}. \quad (\text{B.11})$$

Let us define

$$\bar{f}_{10}(\gamma) = 8 \cdot 10^{-5} + \eta \sum_{b=0}^9 t(\gamma 4^b) + \frac{\gamma}{4}. \quad (\text{B.12})$$

It follows from (B.4), (B.10), and (B.11) that for any γ

$$f(\gamma) < \bar{f}_{10}(\gamma). \quad (\text{B.13})$$

We also notice that it is enough to consider $\gamma \in [A_1, \gamma']$. Numerical calculation shows that the only root $\bar{\gamma}$ of $\bar{f}_{10}(\gamma) - 1$ on the interval $\gamma \in [A_1, 4]$ equals 1.79374. Thus, $\gamma_\infty \geq \bar{\gamma} = 1.79374$ due to (B.13).

We calculate the spectral efficiency (or sum-rate per dimension) as follows:

$$\begin{aligned} R_d &= \sum_{b=0}^{B-1} \eta C_{\text{BIAWGN}}(\gamma_\infty 4^b) \\ &= \eta(0.6859 + 0.9835 + B - 2 - \epsilon) \geq \eta(B - 1 + 0.6706), \end{aligned} \quad (\text{B.14})$$

where we use a bound from [15]

$$1 - C_{\text{BIAWGN}}(x) \leq \frac{2\pi^{3/2}}{\ln 2(\pi^2 - 8)} e^{-1/2x} < 10e^{-1/2x} \quad (\text{B.15})$$

to upper bound ϵ as

$$\begin{aligned} \epsilon &= \sum_{b=2}^{B-1} \eta \left(1 - C_{\text{BIAWGN}}(\gamma_\infty 4^b)\right) \\ &\leq \sum_{b=2}^{B-1} 10e^{-(\gamma_\infty 4^b)/2} \leq 10^{-5}, \quad \text{for any } B. \end{aligned} \quad (\text{B.16})$$

The capacity of the additive Gaussian channel corresponding to power profile (21) with $K_b = \eta N$ streams per level can be calculated as follows

$$C_{\text{GMAC}} = \frac{1}{2} \log_2 \left(1 + \eta \sum_{b=0}^{B-1} \gamma_0 4^b \right) = \frac{1}{2} \log_2 \left(1 + 4\eta \sum_{b=0}^{B-1} 4^b \right), \quad (\text{B.17})$$

$$= \frac{1}{2} \log_2 \left(\frac{1 + 4\eta(4^B - 1)}{3} \right) \leq B - 1 + 1.21. \quad (\text{B.18})$$

Combining (B.18) and (B.14), we obtain (25).

References

- [1] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.
- [2] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [3] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, 1997.
- [4] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, New York, NY, USA, 1995.
- [5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [6] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 6, pp. 995–1004, 2009.
- [7] T. Wo and P. Hoeher, "Superposition mapping with application to bit-interleaved coded modulation," in *Proceedings of the 8th International ITG Conference on Source and Channel Coding (SCC '10)*, January 2010.
- [8] T. Wo and P. Hoeher, "Iterative processing for superposition mapping," This Special Issue, 2010.
- [9] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '97)*, p. 374, Ulm, Germany, July 1997.
- [10] C. Schlegel and A. Grant, *Coordinated Multiple User Communications*, Springer, Berlin, Germany, 2006.
- [11] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [12] C. Schlegel, "CDMA with partitioned spreading," *IEEE Communications Letters*, vol. 11, no. 12, pp. 913–915, 2007.
- [13] Z. Bagley, C. Schlegel, D. Truhachev, and L. Krzymien, "Partitioned-mapping for variable rank MIMO channels," in *Proceedings of the Allerton Conference*, September 2006.
- [14] C. Schlegel and L. Perez, *Trellis and Turbo Coding*, IEEE/Wiley, New York, NY, USA, 2004.
- [15] D. Truhachev, C. Schlegel, and L. Krzymien, "A two-stage capacity-achieving demodulation/decoding method for random matrix channels," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 136–146, 2009.
- [16] E. Biglieri and G. Taricco, *Transmission and Reception with Multiple Antennas: Theoretical Foundations*, Now Publishers, Hanover, Mass, USA, 2004.
- [17] J. W. Byers, M. Luby, and M. Mitzenmacher, "A digital fountain approach to asynchronous reliable multicast," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, pp. 1528–1540, 2002.
- [18] C. Schlegel, Z. Shi, and M. Burnashev, "Optimal power/rate allocation and code selection for iterative joint detection of coded random CDMA," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4286–4294, 2006.
- [19] M. Burnashev, C. Schlegel, W. Krzymien, and Z. Shi, "Characteristics analysis of successive interference cancellation methods," *Problemy Peredachi Informatsii*, vol. 40, no. 4, pp. 297–317, 2004.

- [20] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, Mass, USA, 1963.
- [21] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 9, pp. 1261–1271, 1996.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

