Media-based Modulation: A New Approach to Wireless Transmission

Amir K. Khandani
E&CE Department, University of Waterloo, Waterloo, ON, Canada

Abstract

It is shown that embedding part or all of the information in the (intentional) variations of the transmission media (end-to-end channel) can offer significant performance gains vs. traditional SISO, SIMO and MIMO systems, at the same time with a lower complexity. This is in contrast to the traditional wireless systems where the information is entirely embedded in the variations of a source prior to the antenna to propagate via the channel to the destination. In particular, it is shown that using a single transmit antenna and K receive antennas; significant savings in energy with respect to a $K \times K$ traditional MIMO are achieved. Similar energy savings are possible in SISO, and SIMO setups.

I. Introduction

This work introduces the advantages of varying the end-to-end channel, according to the input data (see Fig. 1), in a wireless communications system with multi-path fading. Such information bearing channel variations are detected at the receiver end, resulting in an equivalent modulation scheme with Additive White Gaussian Noise (AWGN). In other words, the carrier is modulated after leaving the transmit antenna(s) by changing the Radio Frequency (RF) properties of the environment close to the transmitter, but external to the transmit antenna(s). This method of modulating the wave after leaving the transmit antenna(s), coined in this article as Media-Based Modulation (MBM), offers several advantages vs. traditional methods in which RF carrier is modulated prior to leaving the transmit antenna(s). This article refers to the traditional methods as Source-Based Modulation (SBM), in contrast to MBM.

MBM can be realized by changing RF properties, namely permittivity, and/or permeability and/or resistivity, of the propagation environment close to the transmitter. It is well known that permittivity, permeability and resistivity appear in the Maxwell equations and affect the corresponding solution. This in turn affects the end-to-end channel, and consequently the magnitude and phase of the received signal will vary. Note that in a rich scattering environment, a small perturbation in the environment close to the transmitter will be augmented by many random reflections in the propagation path, resulting in an overall independent end-to-end channel realization. If there are $M = 2^{R_m}$ choices available for such channel perturbations, the corresponding received constellation will consist of $M = 2^{R_m}$ points. It is obviously possible, and indeed desirable, to combine MBM and SBM by modulating the carrier partly prior to leaving the transmitter and partly afterwards. If an MBM of 2^{R_m} options is used together with an SBM of 2^{R_s} options, the overall constellation transmits $R_s + R_m$ bits which are channel coded to achieve reliable transmission over the underlying discrete input AWGN channel.

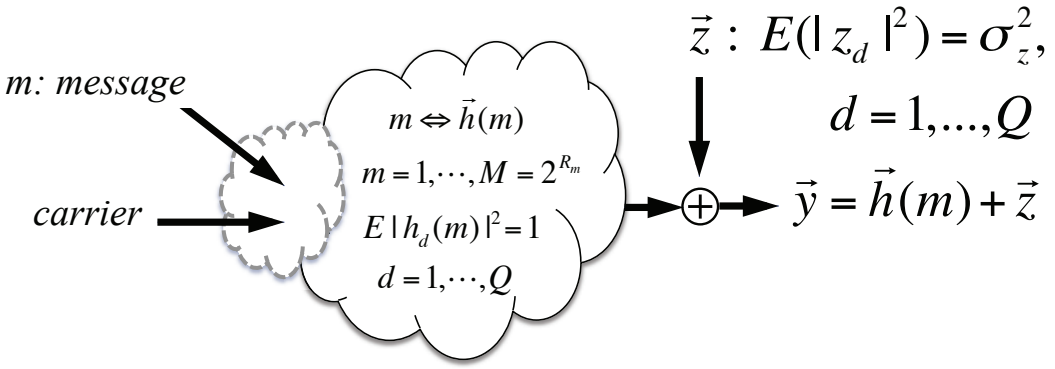


Fig. 1: System block diagram.

The basic idea of embedding information in varying a wireless channel is not new, and indeed precedes modern systems. Polybius came up with a system of alphabetical smoke signals around 150 BC. MachZehnder modulators, widely used for signaling over fiber, modifies the light beam after leaving the laser. However, due to the lack of multi-path in single mode fibers, the advantages discussed here in the context of wireless do not apply. The use of tunable parasitic elements/objects external to antenna for the purpose of beam-forming is widely studied and practically realized in various forms. However, the objective in classical beam-forming is "to focus/steer the energy beam, which does not realize any of the advantages discussed here (when data is modulated by tuning such external parasitic elements/objects). There have been some recent works on embedding data in antenna beam-patters. Alrabadi et al [1] [2] discuss embedding phase information in orthogonal antenna patterns. However, this is motivated by reducing the number of transmit chains and no other benefits are discussed. Bains [3] discusses using parasitic elements for data modulation, and shows limited gains due to energy saving. However, the main features associated with such a setup (e.g., additive properties of information over multiple receive antennas) and methods to realize them, which are the sources of reported improvements in the current article, are not discussed. This article establishes the benefits of MBM, and methods to realize them. A similar analogy exists in the development of Multiple-Input Multiple-Output (MIMO) antenna systems, in the sense that the use of multiple antennas for beam forming was known, but the main advantages of MIMO in terms of spatial diversity and/or data multiplexing, and methods to realize them, were established in late 90s (see references [4] [5] [6] and references therein).

Next, pros and cons of MBM vs. SBM are discussed.

II. Advantages and Disadvantages of MBM vs. SBM

A first advantage of MBM is in the increase of the number of received constellation points without increasing energy. To improve spectral efficiency, SBM alone should rely on using larger values for R_s, resulting in an exponential increase in transmit energy. In contrast, R_m can be increased without directly affecting the transmit energy. Overall, R_s and R_m are selected to achieve the desired rate with the minimum transmit energy. Increase of R_s can be, for example, achieved by modulating the carrier phase with $\pm\pi$ (changing the sign) and $\pm\pi/2$ (modulating over I or Q) to achieve a symmetrical constellation.

A second advantage of MBM is in its inherent diversity in dealing with slow fading. As the constellation points in MBM correspond to different channel realizations, unlike SBM, deep fades do not have a persisting effect. In other words, good and bad channel conditions contribute to a single constellation and span the entire constellation space. As a result, the spacing between constellation points is determined by relative values of different (low and high) complex channel gains (relative to each other). We refer to this feature as the Constellation Diversity. As the constellation size increases, this feature essentially converts a static multi-path fading channel into a non-fading AWGN channel with effective signal energy equal to the received energy averaged over fading statistics. This feature, which occurs even in a SISO-MBM, is inherent and does not involve any tradeoffs. This is in contrast to the MIMO-SBM, in which diversity over a static fading channel can be improved only at the cost of a reduction in rate (multiplexing gain) [7].

A third advantage of MBM arises when the receiver has multiple antennas, namely in SIMO-MBM. In a $1 \times K$ SIMO-SBM, the vector received over K receive antennas spans a single complex dimension. Consequently, the effect of using multiple receive antennas is limited to energy saving which can be realized through maximum ratio combining. In contrast, for a $1 \times K$ SIMO-MBM, the received vector spans all the K complex receive dimensions. As a result, the spacing between received constellation points enjoys a scaling with the transmit energy similar to a SISO-SBM with K times the bandwidth, or similar to a $K \times K$ MIMO-SBM. This feature mimics the multiplexing gain feature of MIMO [7].

A fourth advantage of MBM is in the independence of the noise components over multiple receiving antennas. Note that in both SBM and MBM, assuming transmitter uses uncorrelated code-books of equal energy over its transmit antennas, the statistics of the energy received per receive antenna does not depend on the number of receive antennas. This means, assuming K receive antennas and a fading channel with a mean gain of one, E units of transmit energy results in receiving KE units of energy on the average. In the following, this feature is referred to as "K times energy harvesting. Note that $1 \times K$ SIMO-MBM enjoys K times energy harvesting, similar to that of $K \times K$ MIMO-SBM, and $1 \times K$ SIMO-SBM. The main promise of a $K \times K$ MIMO-SBM is in providing an effect similar to that of K parallel complex channels, with K times energy harvesting. However, the performance of $K \times K$ MIMO-SBM falls short of such K parallel channels as the MIMO channel matrix is typically non-orthogonal, or equivalently, noise components over information bearing dimensions are dependent. This shortcoming is resolved in $1 \times K$ SIMO-MBM. In other words, one unit of transmit energy in a $1 \times K$ SIMO-MBM results in receiving, on the average, one unit of energy per receive antenna, while unlike $1 \times K$ SIMO-SBM, the energy received over each antenna constructs a new constellation. In other words, the received energy forms a constellation that spans the entire K complex receive dimensions. This feature mimics the information scaling of a $K \times K$ MIMO-SBM with multiplexing gain of K. This is in contrast to a $1 \times K$ SIMO-SBM in which the received energy spans a single complex dimension, allowing merely energy saving through receive beam forming (multiplexing gain is limited to one, regardless of K).

A fifth advantage of MBM is in the possibility of energy saving through selecting a subset of channel configurations, which results in a better overall performance for the given energy and spectral efficiency. This feature is similar to the so-called multi-user diversity gain in network scheduling.

A first disadvantage of MBM is that the arrangement of the constellation points is random and constellation points are used with equal probability, while in SBM, constellation points can be uniformly arranged, e.g. QAM (Quadrature Amplitude Modulation) constellation, and can be used with non-uniform probabilities to realize some *shaping gain*. As it will be shown later, the degradation due to the random placement of the constellation points, and also due to using the points with equal probabilities, will be negligible as the constellation size increases.

A second disadvantage of MBM is that the transmitter is generally, although not necessarily, assumed to be oblivious to the details of the constellation, falling into the class of transmission with outage. As will be shown later, the degradation due to the lack of transmit adaptation to channel will become negligible as the constellation size increases.

A third disadvantage of MBM is that the system is Linear, Time Varying (LTV), while SBM is Linear, Time Invariant (LTI). Unlike LTI, LTV systems can potentially expand the spectrum. Such a time varying feature also contradicts the functionality of the traditional channel equalization techniques. On the other hand, as the LTV nature of the system is due to the random selection (with equal probabilities) of one out of M options for an underlying LTI channel, the power spectrum observed at any given receive point will be equal to the average of the power spectrums of the M underlying LTI channels, times the input power spectrum. As wireless channel can transmit a wide range of frequencies, the underlying LTI systems will have a wide spectrum. As a result, the overall power spectrum follows the shape of the input spectrum (RF carrier is spectrally shaped prior to transmission).

Another potential problem arises as traditional SBM systems exploit the LTI property and rely on some form of equalization to compensate the effect of the Inter Symbol Interference (ISI). Due to ISI, the energy associated with a single time symbol is spread over its neighbors. This means the signal over a single dimension at the input is spread over L dimensions at the output, where L denotes the length of the channel impulse response. In the case of SBM, the resulting L dimensional output vector spans a single dimension, and equalization procedure should (ideally) accumulate the received energy corresponding to any given input signal (spread over these L dimensions) into a single decision variable. In contrast to SBM, in MBM, the L dimensional vector at the channel output corresponding to a single transmission spans an L dimensional space, increasing the information bearing capability of any given time transmission by a factor of L. As a result, a transmission in MBM can be followed by L zeros to flush out the channel memory and this is achieved without sacrificing the effective dimensionality of the overall signal space (see Fig. 2). An alterative would be to reduce the gap between successive transmissions to less than L and apply sequence detection in time to account for the ISI. This can be achieved using a state diagram (evolving in time) with states corresponding to previous constellation points, or a quantized version of them to reduce the size of the state space. A more detailed analysis should include the effect of the channel impulse response for MBM, and the specifics of the equalization method for SBM, and is beyond the scope of this article.

As mentioned above, a natural approach to deal with channel memory in MBM is to rely on time domain methods, including sequence detection in time. However, this problem has a much simpler solution in multi-user setups, and in particular in networks based on Orthogonal Frequency Division Multiple Access

(OFDMA). In OFDMA, different nodes use different tones of an OFDM system, which are all synchronized within the OFDM cyclic prefix. In this case, the channel at the transmitter side of each link can be varied from OFDM symbol to OFDM symbol (to embed information). This means the channel is varied at the beginning of each OFDM symbol, while it remains the same throughout the time duration of any given OFDM symbol. As the channel remains static over any given OFDM symbol, OFDM structure remains intact. As a result, each link can rely on a simple single tap equalizer instead of sequence detection in time. The cost is the waste of bandwidth in sending the cyclic prefix, as is the case in any OFDM system. However, similar to traditional OFDMA, this waste is shared among all users. The length of the cyclic prefix is determined by the channel memory and consequently the waste remains the same, regardless of the number of OFDM tones (relative waste can be reduced by increasing the number of tones). Such a setup can be used as long as transmitters are separated in space, i.e., in the uplink (separate transmitters use different OFDM tones to send to a common receiver), and in parallel interfering links (each link uses a different OFDM tone).

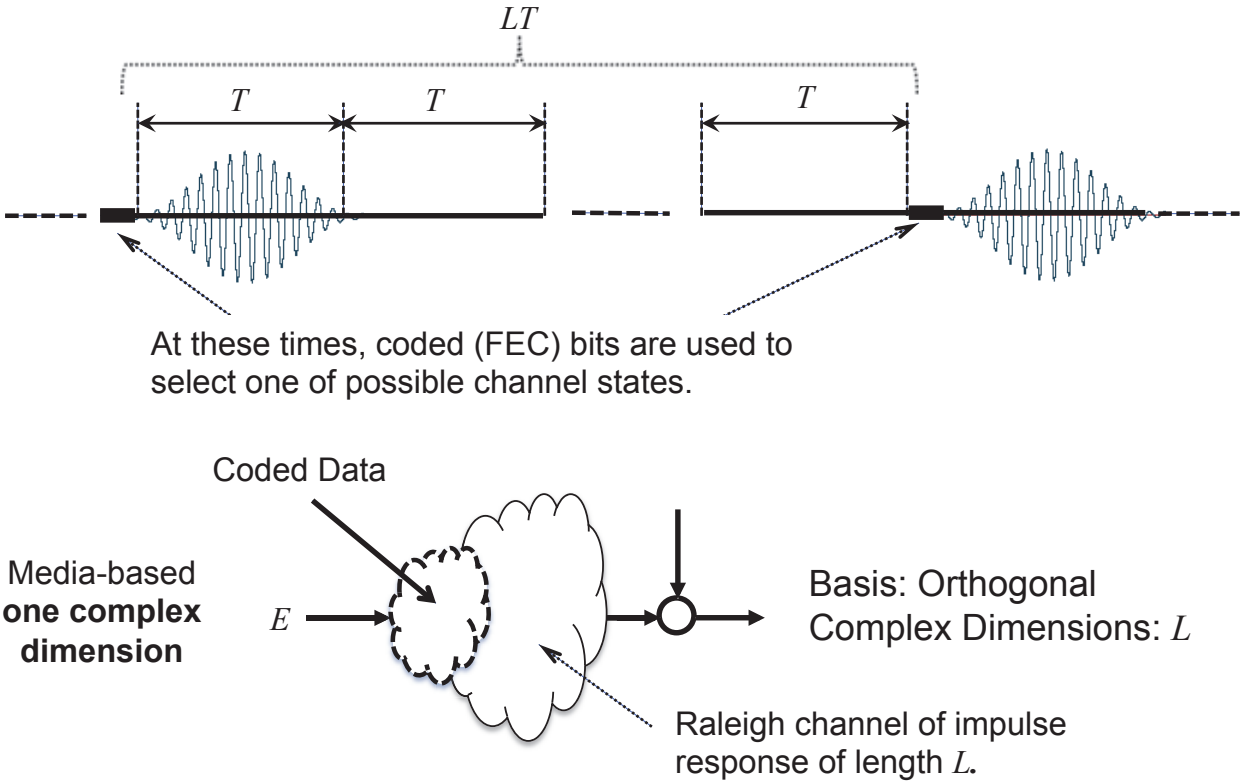


Fig. 2: Spectral shaping and impulse response.

III. System Model and Relative Merits

Figure 1 shows the setup of a $1 \times K = Q/2$ SIMO-MBM. For the sake of simplicity, the concept is explained by focusing only on the MBM part of the transmission, and the combination with SBM is straightforward. There are $M = 2^{R_m}$ messages, indexed by $m = 1, , M$, which select one of the M

channel realizations corresponding to channel realizations $\vec{h}(m)$ with components $h_d(m)$ for $d = 1, , Q$, and $m = 1, , M$. Due to the normalization of fading, we have $E|h_d(m)|^2 = 1$ where E denotes statistical averaging. AWGN vector $\vec{z}$ has independent identically distributed (i.i.d) components z_d, $d = 1, , Q$, where $E|z_d|^2 = \sigma_z^2$.

In this setup, although the transmitter selects the channel realization $\vec{h}(m)$ to transmit message m, transmitter is oblivious to the details of $\vec{h}(m)$. On the other hand, the receiver knows $\vec{h}(m)$ and can perform signal detection, if the mutual information is sufficient. Receiver training is achieved by sending a set of pilots over different channel realizations, enabling the receiver to measure $h_d(m)$, $\forall d$, $\forall m$. For a Raleigh fading channel (rich scattering), $h_d(m)$ are i.i.d Gaussian, which is in accordance with the optimality of Gaussian random coding over AWGN channels. However, as the transmitter is oblivious to the details of $\vec{h}(m)$, an outage may occur. A similar outage phenomenon exists in the case of SBM over a static fading channel which is handled by increasing the transmit energy, or by exploiting spatial diversity offered by using multiple antennas. However, in the case of SBM, this compensation necessitates a significant increase in the transmit energy, or loss in the rate by exploiting spatial diversity. Due to the inherent diversity of MBM, the issue of outage in slow fading channels will be much less problematic compared to SBM. We have,

$$I(\vec{y}; m) = I(\vec{y}; \vec{h}(m)),$$

which, noting the AWGN channel, results in

$$I(\vec{y}; m) = H(\vec{y}) - H(\vec{z}) = H(\vec{y}) - K \log_2(2\pi e \sigma_z^2).$$

Although the maximum rate of such a transmission scheme is limited to $\log_2 M$, as is the case in any channel with a discrete constellation of size M, the rate achievable in a proper operating point prior to the saturation can be significantly higher than its SBM counterpart.

Next, MBM and SBM are compared in terms of the slope of rate vs. energy at low *SNR*.

IV. Slope of Rate at low SNR, and Effective Dimensionality

Recall that MBM and SBM are preferably combined, in which case due to symmetrical phase modulation in the underlying SBM portion, the resulting signal set will be symmetrical. Among other benfits, this results in a constellation with zero mean. Assuming a constellation with zero mean, for low values of Signal to Noise Ratio (*SNR*), we have (see appendix A):

$$\lim_{\epsilon \to 0} \frac{I(\epsilon)}{\epsilon} = \frac{Q G_2}{2}$$

and,

$$E\left[\lim_{\epsilon \to 0} \frac{I(\epsilon)}{\epsilon}\right] = \frac{Q^2}{2}$$

where $I(\epsilon)$ is the mutual information per real dimension (for σ^2 normalized to one) as a function of a small increase in RF energy, namely ϵ, starting from $\epsilon = 0$ (zero RF energy), G_2 is the sample second moment of the M constellation points, $Q = 2K$ is the number of real dimensions, and $E[.]$ denotes statistical expectation. Although G_2 is a random variable, its variance approaches zero with $1/M$. This is

in contrast to the case of a $K \times K$ MIMO-SBM for which the scaling of rate vs. SNR at low SNR is at best (i.e., assuming feedback and water filling) limited to the largest eigenvalue of the channel matrix. This means for low SNR values, a $K \times K$ MIMO-SBM is essentially a two-dimensional channel (one complex dimension) with an energy gain corresponding to the largest eigenvalue. For a Raleigh fading channel, these are eigenvalues of a $K \times K$ random Wishart matrix[1] for which the expected value of the largest eigenvalue, although being increasing with K, is limited to 4 which is approached as $K \to \infty$ [8]. In the case of a $K \times K$ MIMO-SBM, as SNR increases, water filling results in occupying more of the available dimensions. See Fig. 3.

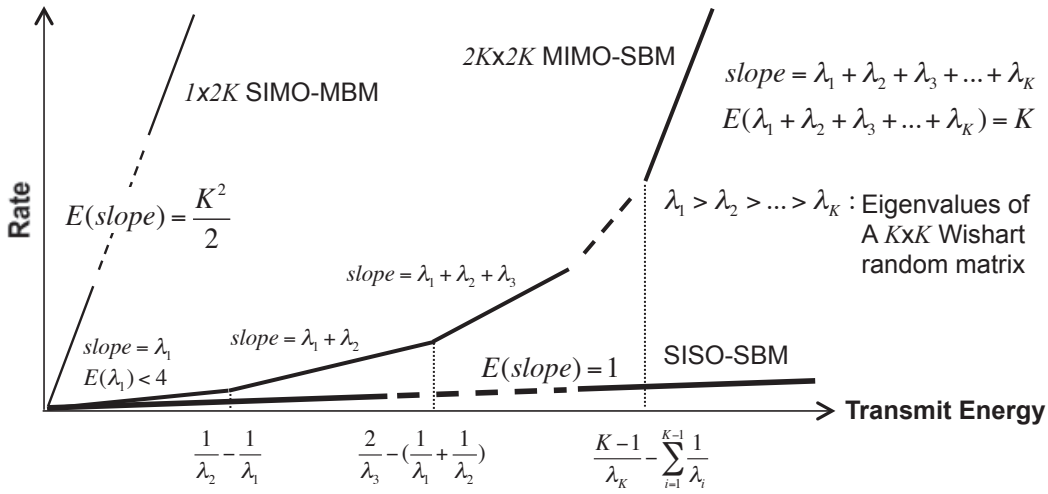


Fig. 3: Effective dimensionality of legacy MIMO vs. media-based Modulation.

In comparing MBM vs. SBM, to have a fair comparison, one should ideally compare the outage capacity of the two systems. However, a major benefit of MBM vs. SBM is that it changes the statistical behavior of the end-to-end channel, and consequently that of the mutual information, and in particular reduces its variance. This is due to the inherent diversity of MBM as each constellation point corresponds to a different channel realization and consequently good and bad channel conditions contribute to any single transmission (received constellation). This effect, which will be more pronounced at higher transmission rates, essentially converts a static fading channel into an Additive White Gaussian Noise (AWGN) channel where the SNR is determined by the received energy averaged over fading statistics. This is unlike MIMO-SBM where diversity over a static fading channel can be improved only at the cost of a reduction in spectral efficiency. In addition, such a fair comparison depends on the statistics of fading, and is further complicated by the interplay between rate and diversity order in MIMO-SBM. For these reasons, although being to the disadvantage of SIMO-MBM, the relative merits of MBM vs. SBM are studied in two different scenarios. A first scenario, which focuses on the effect of the diversity inherent to MBM, primarily relies on SISO links and compares SISO-MBM vs. SISO-SBM in terms of the outage capacity. A second scenario, which focuses on the other features of MBM, compares SIMO-MBM vs. MIMO-SBM in terms of Ergodic capacity.

[1]Eigenvalues are in pairs of equal magnitudes corresponding to quadrature components.

Next, the gain due to inherent constellation diversity in MBM is discussed. It is shown that the mutual information across an AWGN channel with MBM tends to the capacity of the underlying AWGN channel as $(1/M)^{1/K}$, for $M \to \infty$.

V. Gain due to Inherent Constellation Diversity

Assuming Raleigh fading, the components of the constellation points are random and follow an i.i.d. Gaussian distribution. This is in agreement with random coding over an AWGN channel, however, the constellation structure in MBM remains the same over subsequent transmissions instead of having independent realizations as is required in random coding. This causes some loss in the achievable rate as compared to the capacity of the underlying AWGN channel, namely AWGN with an energy gain equal to the statistical average of fading. The achievable rate is a random variable (depends on the specific constellation), fluctuating around the capacity of the underlying AWGN channel. As the number of constellation points increases, the variance of this random variable decreases, and the achievable rate will eventually tend to the capacity of the underlying AWGN channel.

It is difficult to compute the mutual information across the channel shown in Fig. 1. The reason is that for a given realization of the discrete input constellation, although the components of the constellation are i.i.d. Gaussian (Raleigh fading), as the constellation structure remains the same, the distribution of the channel output will not have a Gaussian distribution. In the following, a method is presented to compute, on the average, the loss in the capacity as compared to a traditional AWGN channel for which a Gaussian random code-book with i.i.d. Gaussian elements is used. Averaging is performed with respect to all possible discrete random code-books, each corresponding to a different realization of the underlying M points constellation.

Let us normalize the statistical average of the Raleigh fading to one, power of the AWGN to σ_z^2, i.e. $N(0, \sigma_z^2)$, and consider two ensembles of random codes.

Ensemble I: A realization of the M-points constellation; i.e. a fixed set of M points with i.i.d Gaussian components, its extension with equal probability for the M constellation points (i.e., cartesian product of the given constellation), and a random code selecting a subset of cardinality $\mathscr{C}$ in this extension. This is equivalent to using an i.i.d uniform distribution for random coding over the given M-points constellation.

Ensemble II: An ensemble of cardinality $\mathscr{C}$ with i.i.d Gaussian components of variance one, i.e., $N(0, 1)$, over time and spatial dimensions.

Consider the collection of such random codes, each corresponding to a different realization of the M points constellation. To average the rate over different realizations of the M points constellation, which is equivalent to time domain average over different code-books of Ensemble I, let us consider the cartesian product of all such code-books. Let us also consider a similar cartesian product for ensembe II. These cartesian products are called "composite ensembles" and their elements are called "composite code-words". Capital vector notation in Fig. 4(a) corresponds to such composite code-words, with the corresponding per code-book components shown in Fig. 4(b). Let us select a codebook from each of these two composite ensembles, and for each code-word in the composite codebook from Ensemble I, find the code-word in the composite codebook from Ensemble II that is at the minimum square distance to it. This is equivalent to quantization using minimum mean square error criterion. Let us denote the corresponding vector of

quantization error by $\vec{N}_q$, i.e. $\vec{N}_q = \hat{\vec{X}} - \vec{X}$, and its per code-book components by $\vec{n}_q$, as shown in Fig. 4(a)(b). This setup allows us to tackle the bottleneck in providing a closed form solution for the mutual informaition, as $\vec{Y}$, and consequenclty $\hat{\vec{Y}}$, in Fig. 4 are composed of i.i.d. Gaussian elements. Two points need to be considered:

- Quantization noise, namely $\vec{N}_q$, will be dependent on $\vec{X}$.
- Mapping from $\vec{X} \rightarrow \vec{Y}$ is not one-to-one, because several $\vec{X}$ may be quantized to the same $\vec{Y}$. As a result, elements of $\vec{Y}$ occur with non-equal probabilities, even when elements of $\vec{X}$ have equal probabilities. Although elements of $\vec{Y}$ occur with non-equal probabilities, it does not affect the hardening effect in the underlying Gaussian channel and $\hat{\vec{Y}}$ will have a uniform distribution over a spherical region at the channel output. This means $\hat{\vec{Y}}$ satisfies the conditions for sphere hardening which is the basis for achieving error free transmission (see proof of AWGN capacity in [10]). This guarantees error free transmission if the mutual information across the concatenated channel, i.e., from $\vec{X}$ to $\hat{\vec{Y}}$, is sufficient. Indeed, including the non-equal probability of points would enable using Maximum A Posteriori (MAP) decoding instead of Maximum Likelihood (ML), which would only reduce the probability of error.

Noting Ergodicity, time domain average over the setup in Fig. 4(a) is replaced with statistical average over the setup in Fig. 4(b). The following steps are taken to simplify the computation of the statistical average over Fig. 4(b):

1) The dependency among the components of the quantization noise is ignored.
2) The combined additive noise, namely $\vec{n}_q + \vec{z}$, is replaced with an i.i.d Gaussian noise, with a variance conditioned on the constellation point, $\vec{h}$.

These two steps ignore the dependency among the components of the quantization noise, and replace it with an additive Gaussian noise with a variance conditioned on the input $\vec{h}$. As a result of these two steps, the entropy of the total additive noise (Gaussian noise, plus quantization noise) is replaced by an upper-bound (memory is ignored and distribution is assumed to be i.i.d. Gaussian). Note that, as the distribution of $\vec{Y}$ and $\hat{\vec{Y}}$ is i.i.d. Gaussian prior to (regardless of) these two steps, entropy of the channel output in Fig. 4 will be that of a Gaussian, and these two steps result in a lower bound on the mutual information (entropy of the additive noise is upper-bounded without affecting the entropy of the output).

As long as the resulting lower bound on the mutual information between input $\vec{X}$ and output $\hat{\vec{Y}}$ is sufficient, reliable decoding of $\vec{X}$ would be possible. This imposes a limit on the cardinality of the input messages, i.e.,

$$\log_2(\mathscr{C}) \leq I(\vec{h}; \hat{\vec{y}}). \tag{1}$$

Appendix B shows that, as $M \rightarrow \infty$,

$$\sigma^2_{n_q|\vec{h}} \simeq \left(\frac{1}{M}\right)^{2/Q} \rightarrow 0, \text{ as } M \rightarrow \infty. \tag{2}$$

As a result, the time average of the capacity of setup in Fig. 4(a), which is equal to the statistical average of the capacity of setup in Fig. 4(b), tend to the capacity of a $K \times K$ AWGN channel with identity channel matrix (this is the same as the capacity of K parallel AWGN channels with K times energy harvesting).

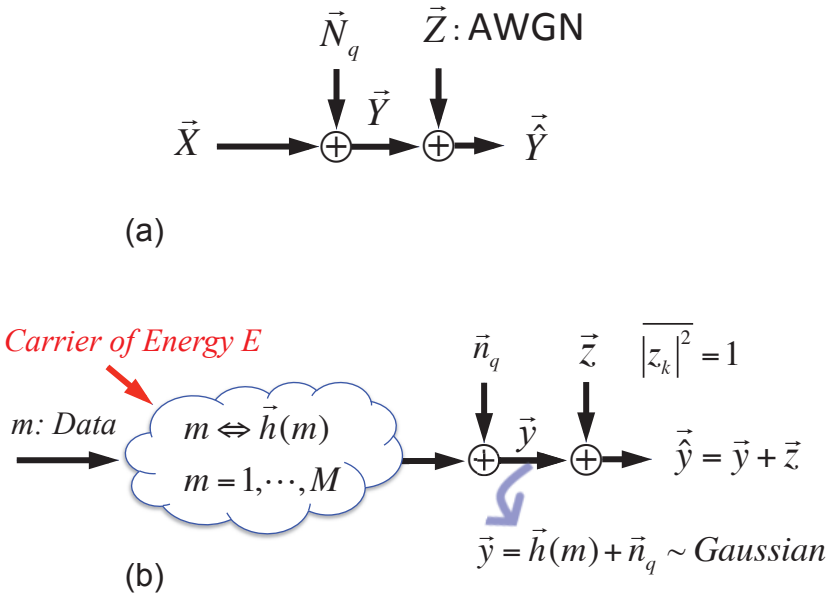


Fig. 4: (a) Block diagram for the composite ensemble of code-books, and (b) their per code-book components.

VI. Time Sharing, Shaping Gain and Selection Gain

The following question, although intuitively obvious, may arise. *Given a set of constellation points and noise power, can time sharing between two energy values improve the mutual information?* This question can be addressed in two cases, namely: (1) *enhanced time-sharing*, where some additional information is embedded in the time-sharing coefficients, and (2) simple time-sharing. It is straightforward to see that *enhanced time-sharing* is equivalent to embedding some information in the so-called source-code book introduced earlier, and can indeed improve the mutual information. This is similar to legacy setups. In the case of *simple time-sharing*, first it is straightforward to show that in media-based modulation, similar to legacy systems, mutual information over AWGN channel will be function of the "ratio of signal power to the noise power, i.e., *SNR*, and the curve of the mutual information vs. energy will be a monotonically increasing concave function and as a result simple time-sharing between energy values does not increase the mutual information.

In legacy constellations, using the constellation points with equal probability results in the loss of shaping gain. However, in a media-based modulation, constellation points will asymptotically have a Gaussian distribution (assuming Raleigh fading), which in turn means the shaping gain will be inherently realized in full. In addition, in a non-adaptive setup, which is the more attractive scenario for the use of media-based modulation, all the constellation points are independently generated with the same distribution while the transmitter is oblivious to the actual realization. As a result, it follows that the constellation points should be used with equal probability. Motivated by these observations, and to simplify the encoding procedure in adaptive setups, we assume the constellation points in a media-based constellation are used with equal probability.

If the constellation points are restricted to have equal probability, selecting a subset of them can increase

the mutual information. The optimum solution, i.e., selecting the subset of points that maximizes the mutual information, may be too complex. There are two relevant observations: 1) as the SNR increases a larger subset of constellation points should be used, and 2) the slope of the mutual information vs. energy at low SNR is determined by the sample second moments of the selected subset (see appendix, and note that assuming a symmetrical constellation, the sample first moment will be zero). These observations motivate selecting the subset of points with highest energy, which maximizes the slope of the rate at zero SNR. Figure 13 shows an example for the gains achievable through such selection, including a comparison between *optimum selection* and *selection with the highest sample second moment* which is seen to perform very close to optimum. As very low SNR, the slope of the mutual information vs. energy will scale with the maximum norm of the constellation points. Appendix B contains mathematical derivations regarding this gain, which analogous to the multi-user diversity gain (scheduling gain) which is known to scale with logarithm of the cardinality.

VII. Equalization

In legacy systems, where channel is fixed and linear, optimum equalization over an AWGN channel entails orthogonalization of the channel (using the channel input eigen-functions) and performing water filling according to the eigenvalues. This converts the channel into a set of parallel sub-channels with gains equal to the respective eigenvalues [9]. For a band-limited frequency flat channel of gain G, the corresponding eigenvalues will be all equal to G. In this case, the optimum receiver will be a matched filter with a sinc-function impulse response. In applying media-based modulation over such a channel, the value of G, which represents the constellation point, varies from symbol-to-symbol according to the input data. This means for all channel configurations (constellation points) the optimum receiver is the same sinc-function. As a result, a simple matched filter receiver will be optimum regardless of the particular constellation point. This desirable feature (frequency flat channel), which simplifies the task of equalization in media-based modulation, is a valid model for narrow-band channels. This makes Orthogonal Frequency Division Multiple Access (OFDMA), where narrow-band orthogonal sub-channels (tones) are allocated to *separate (in space) transmitters*, is a proper candidate to realize media-based modulation with a simple channel equalizer. Note that the only requirement is that the transmitting nodes, each occupying a small number of adjacent narrow band tones, be sufficiently separated in space. Examples include: i) interfering links (separate transmitter/receiver pairs) operating over non-overlapping narrow-band sub-channels (a group of adjacent tones), and ii) uplink transmission where distant transmitters use non-overlapping sub-channels to communicate to a common receiver. It should be added that in OFDM/OFDMA transmission, the cyclic prefix added in time results in a waste of bandwidth and possibly waste of transmit energy. Indeed, if the OFDM is composed of N tones and a cyclic prefix of length CP, the receiver structure extracts N time samples from the received $N+CP$ samples for the Fast Fourier Transform (FFT). In this case, if the if the length of the channel impulse response is less than CP, the receiver will have flexibility where to select the N FFT samples, resulting in several versions of the received signal with dependent noise. The loss is energy due to the transmission of the cyclic prefix can be partially compensated by coherent combining over such replicas of the transmitted signal. All these arguments are valid for the case of using OFDMA for

media-based modulation. Note that in using MBM with OFDMA, the channel is kept the same throughput each OFDM symbol, and is changed from OFDM symbol to OFDM symbol.

Next we discuss how to reduce the waste of bandwidth in the case of point-to-point OFDM.

A. OFDM with repetition coding across tones

First of all, if the channel is kept constant during each OFDM symbol (changed from OFDM symbol to OFDM symbol), transmitter can modulate data in different tones using conventional SBM, with the particular channel impulse response selected by the MBM part modulating all the tones. In this case, the SBM over different tones can act as redundancy where the constellation points in different tones are scrambled to increase the minimum distance. Such a setup will perform better than simple repeation coding with maximum ratio combining. This essentially recovers the loss in bandwidth, by using the additional transmissions as channel coding redundancy.

B. OFDM with up-sampling

Advantages of OFDM come at the price of a loss in bandwidth and energy efficiency due to the transmission of cyclic prefix. A solution for using OFDM with MBM is to reduce the number of tones. However, as the number of OFDM tones reduces, this waste will relatively increase. A partial solution to this problem is to use a larger sampling rate and modulating a small number of tones in the bandwidth of interest, while the rest of the band is left empty to avoid interference to units operating over adjacent frequency bands.

On the other hand, if the channel is wideband, the entire channel impulse response varies according to the input data and above simple equalization is not optimum any longer. For such cases (single carrier transmission over wide-band channels), receivers of legacy systems rely on methods such as channel inversion, decision feedback or sequence estimation using probability propagation. Among these, channel inversion is not applicable to the case of media-based modulation, as the channel is changing with time. In this case, one can still rely on sequence estimation. This can be achieved, for example, by using a state diagram to capture the collective effect of the sequence of impulse responses on consecutive time samples, and thereby track the sequence of transmitted symbols. In particular, methods based on iterative decoding using probability propagation between such a soft output sequence estimator and a soft output channel decoder can be applied.

VIII. Training and Tracking

Legacy modulation schemes are sensitive to and suffer from non-idealities such as: (1) I/Q imbalance, (2) non-linear amplification, and (3) limited PA efficiency. Most of these shortcomings disappear in the case of media-based modulation. On the other hand, in media-based modulation, one should rely on a training phase with a complexity that grows linearly with the number of constellation points. As mentioned earlier, two bits of information can be embedded in the carrier sign change and exchanging the role of I and Q. This feature can be used to reduce the number of training symbols by a factor of 2^{2K}, where K is the

number of receive antennas. The small variations (in time) in the structure of the received constellation can be tracked and adjusted using decision feedback, in conjunction with occasional retransmission of training phase. In this case, it will be sufficient to retransmit only a subset of constellation points in subsequent phases of training and rely on interpolation to estimate the changes in the rest of constellation points.

IX. Numerical Results

Figure 4 primarily shows the effect of the constellation diversity and captures the relative performance of MBM vs. SBM in terms of outage behavior. Figure 5 shows the relative performance of MBM vs. SBM in terms of Ergodic capacity for SBM. This captures the effect of noise independence over receive dimensions for MBM. Figure 9 shows capacity of a 256 QAM in comparison with a random MBM constellation with 256 points. As seen in Fig. 9, the mutual information of a constellation with 256 points is relatively close to that of 256QAM constellation, with minor fluctuations (see the example shown in the sub-figure for a particular SNR of 15dB). Sub-figure in Fig. 9 shows that subject to about one dB of energy over-budgeting (as compared to AWGN), the SISO-MBM will have a reliable performance, but the SISO-SBM requires 30dB to 50dB over-budgeting of SNR relative to the benchmark corresponding to AWGN channel. Figur 13 shows the effect of selecting a subset of a set of random points to maximize the mutual information subject to equal probability for the selected points (selection gain). As seen, a simple rule based on selecting a subset of points with highest sample energy achieves a performance close to the best selection.

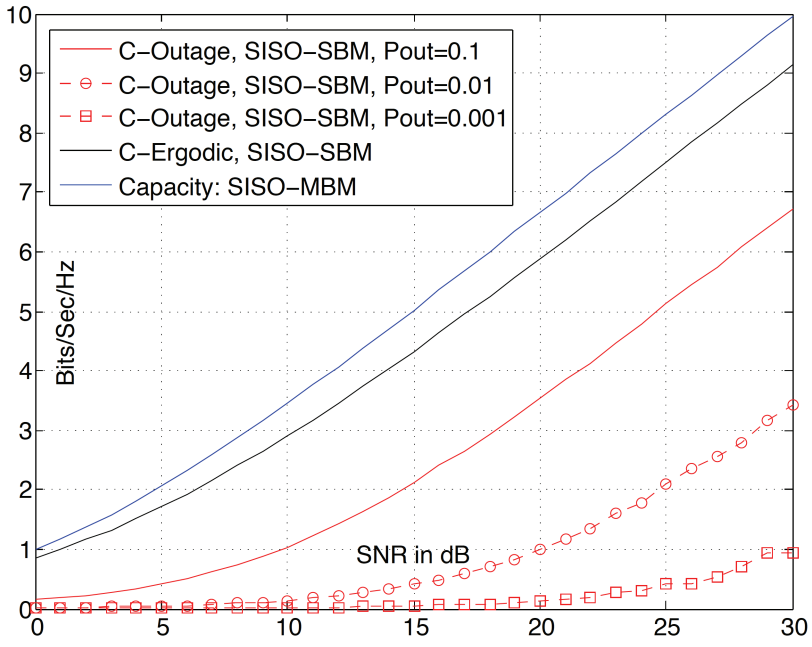


Fig. 5: Outage capacity of SISO-SMB vs. SISO-MBM.

X. Candidates for RF Channel Perturbation

Traditional RF beam forming schemes aim at concentrating energy in certain directions to increase received SNR. On the other hand, in MBM, the aim is to cause random variations in the received signal,

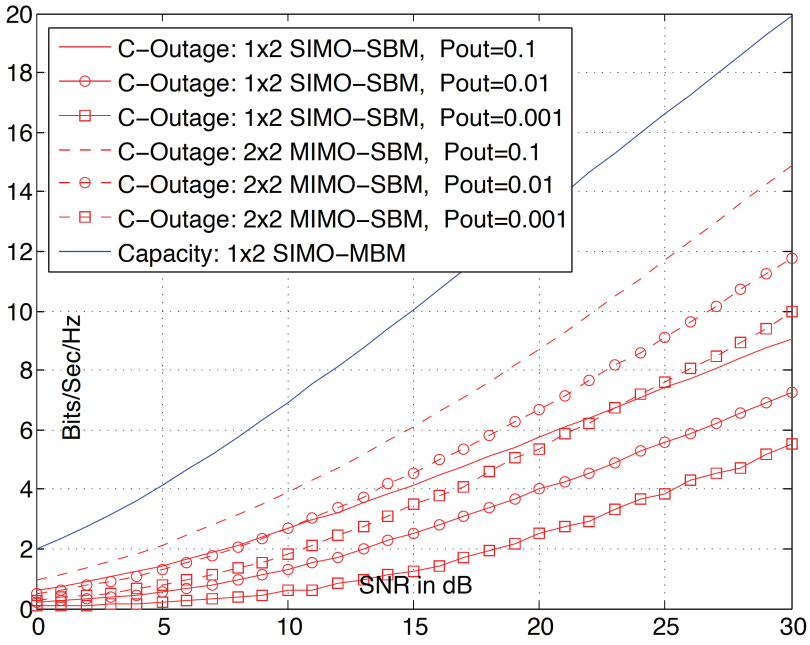


Fig. 6: Outage capacity of SIMO-SMB vs. SIMO-MBM for two antennas.

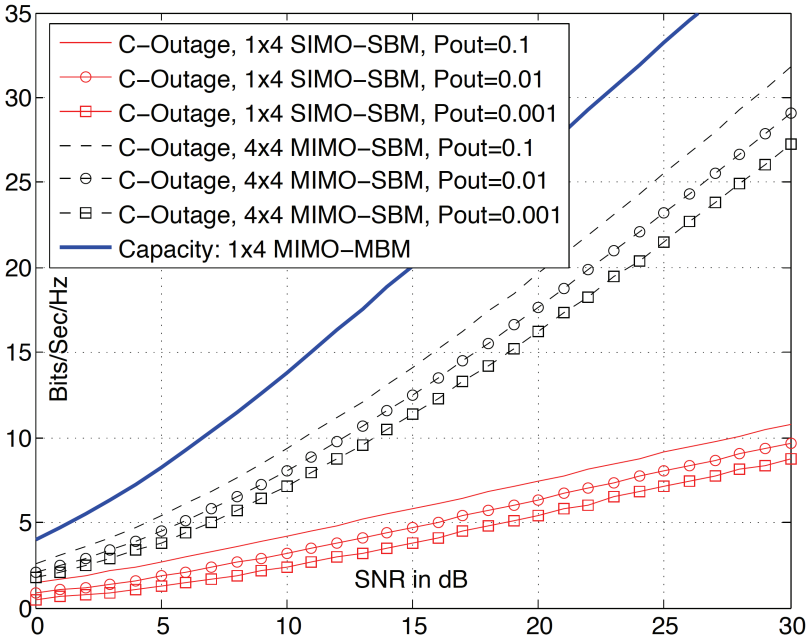


Fig. 7: Outage capacity of SIMO-SMB vs. SIMO-MBM for four antennas.

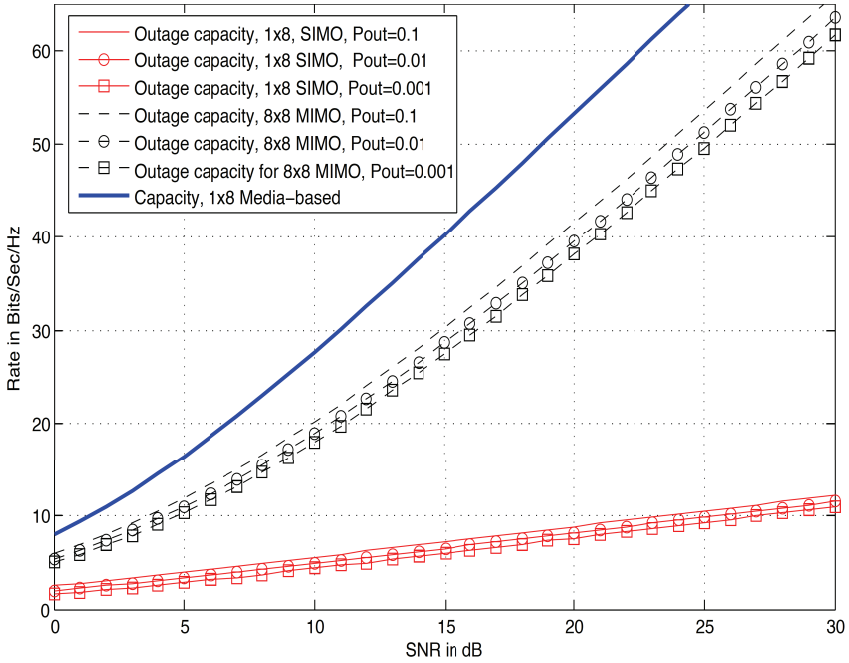


Fig. 8: Outage capacity of SIMO-SMB vs. SIMO-MBM for four antennas.

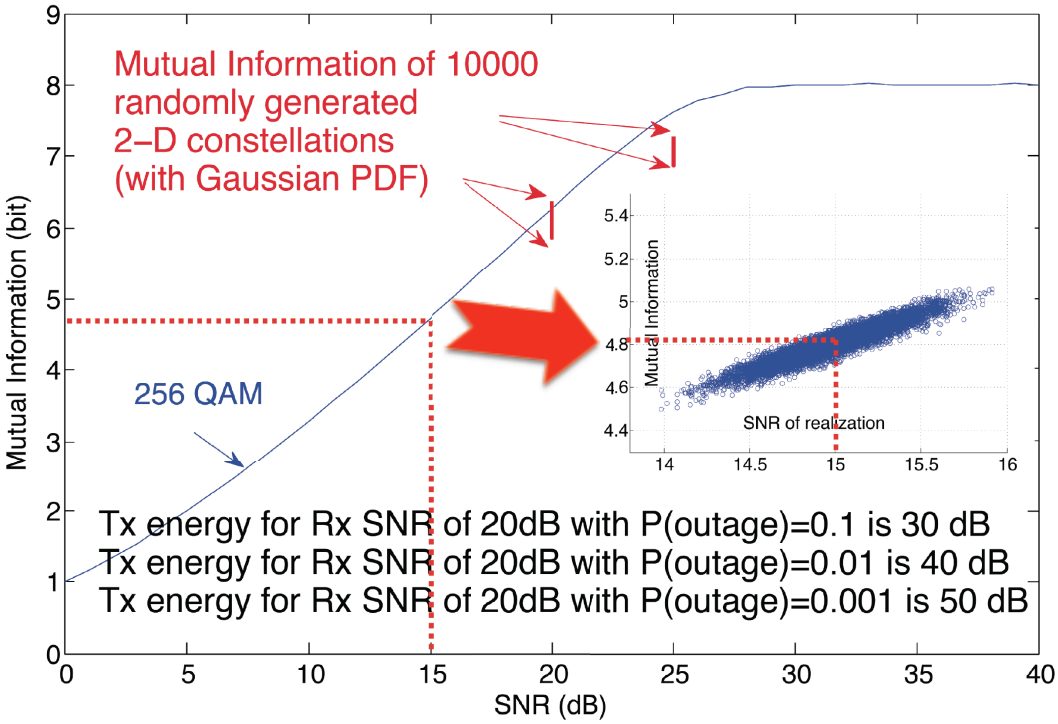


Fig. 9: Random SISO-MBM constellation vs. regular Quadrature Amplitude Modulation (QAM).

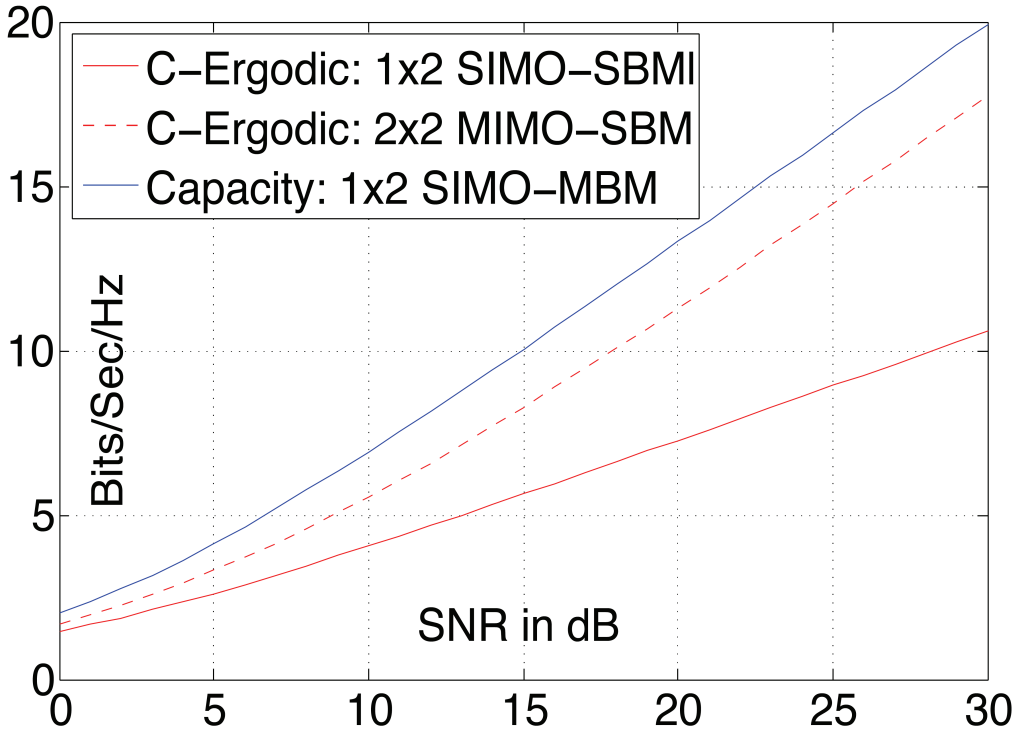


Fig. 10: Ergodic capacity of SIMO-SBM vs. SIMO-MBM for two antennas.

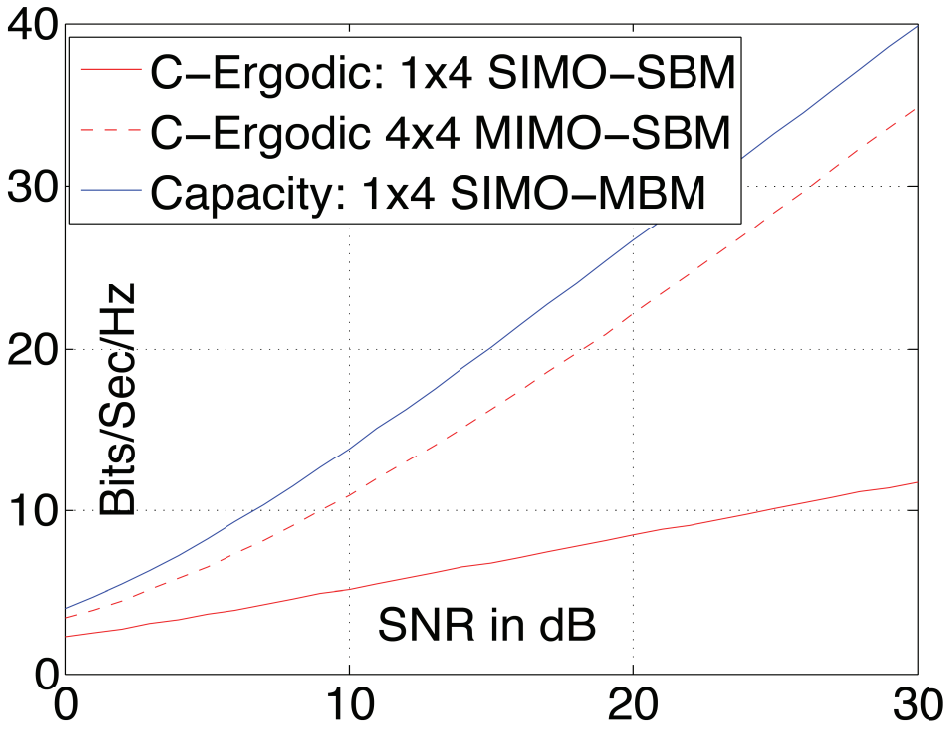


Fig. 11: Ergodic capacity of SIMO-SBM vs. SIMO-MBM for four antennas.

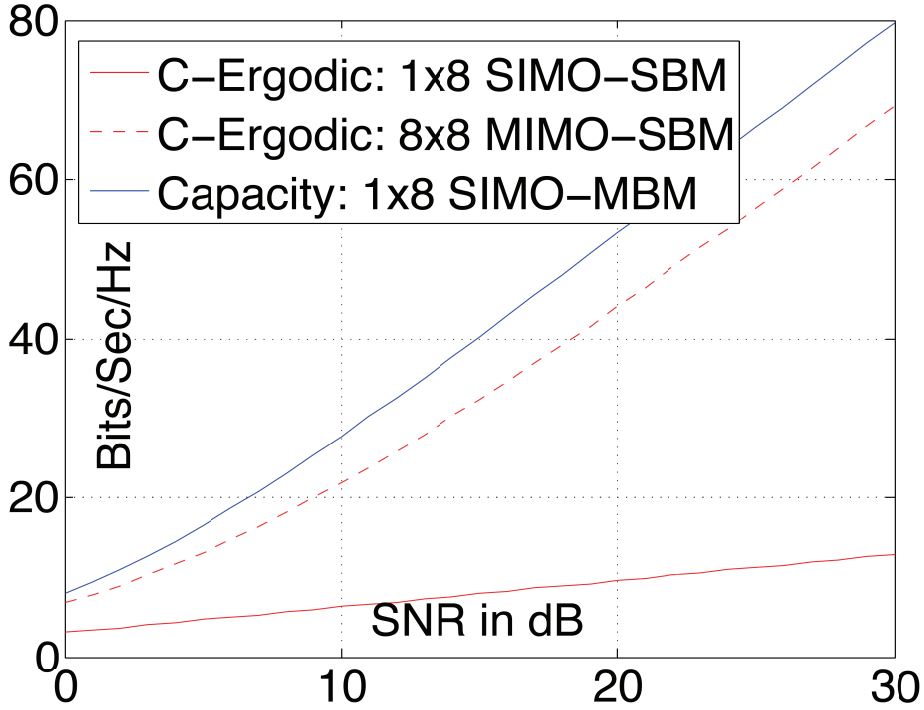


Fig. 12: Ergodic capacity of SIMO-SBM vs. SIMO-MBM for eight antennas.

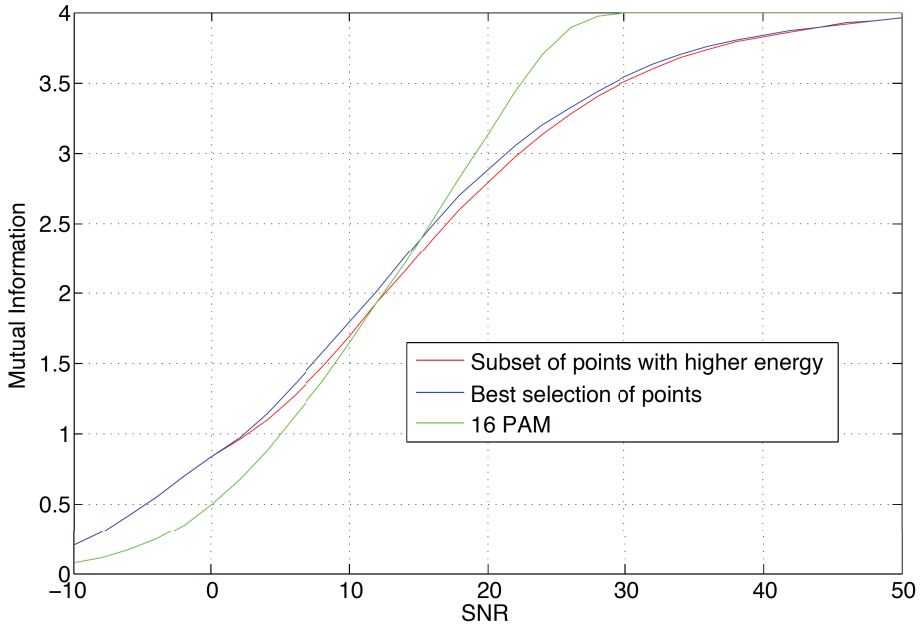


Fig. 13: Selection gain in a 256 points constellation.

without the need to know and/or control the imposed variations, neither to focus the energy. This objective is easier to realize as compared to traditional RF beam forming. In spite of these differences, many of the techniques developed for RF beam forming are applicable to MBM, including: Methods for changing property of a wave-guide by surface plasma generated through light sources, or leaky wave antenna (based on a waveguide with tunable surface leakage). Creating surface plasma in an external to antenna parasitic object, e.g. using light intensity to change plasma depth causing a tunable impendence surface. Tunable impedance surface as an external to antenna parasitic element, e.g. by changing the permeability of ferrite via a current-carrying coil) changing the permittivity of ferroelectric material via a bias voltage, using meta-material as a parasitic object with changeable refraction index. Figures 141516 show examples for the practical realization of MBM. Figure 17 shows more details for oractical realization of an RF mirror using a periodic switched structure. Figure 17 shows examples of corresponding antenna patterns (simulated using HFSS). Figure 18 shows examples for the realization of MBM constellation. Figure 20 shows the indoor simulation environment for the MBM constellation in Fig. 19. Simulation are performed using HFSS (Fig. 18) and Remcom Wireless Insite (Figs. 20 to 22).

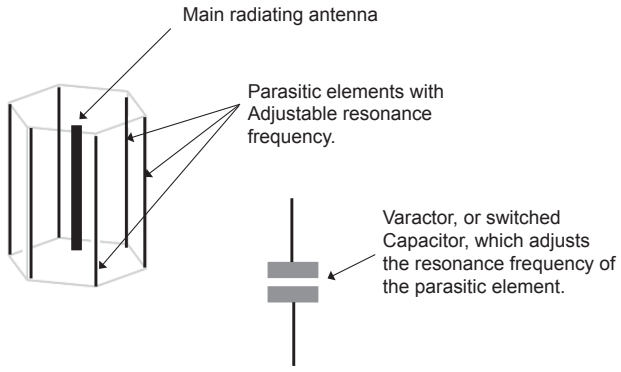


Fig. 14: Example of a cylindrical structure with tunable parasitic elements surrounding a main transmit antenna.

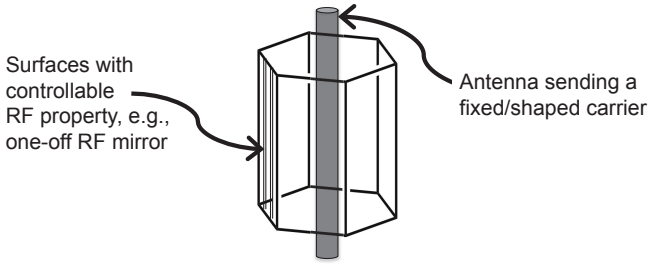


Fig. 15: Example of a cylindrical structure with on-off RF mirrors surrounding a main transmit antenna.

Acknowledgment: Author is thankful to H. Ebrahimzad, M. Salehi, and E. Seifi and for many useful inputs and various computer simulations.

REFERENCES

[1] O. N. Alrabadi, A. Kalis, C. B. Papadias, R. Prasad, "Aerial modulation for high order PSK transmission schemes," *1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic*

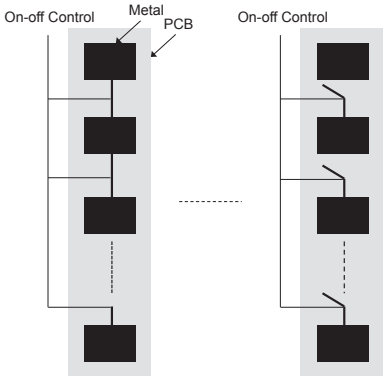


Fig. 16: Example of periodic structure for the walls of the cylindrical structure surrounding a main transmit antennas. Switches on a given wall, can be all or partially switched to change the channel. Off walls reflect the wave and form a cavity internally which results in several reflection for the RF signal prior to exiting the cylinder. This adds to the richness (independence) of the various channel configurations.

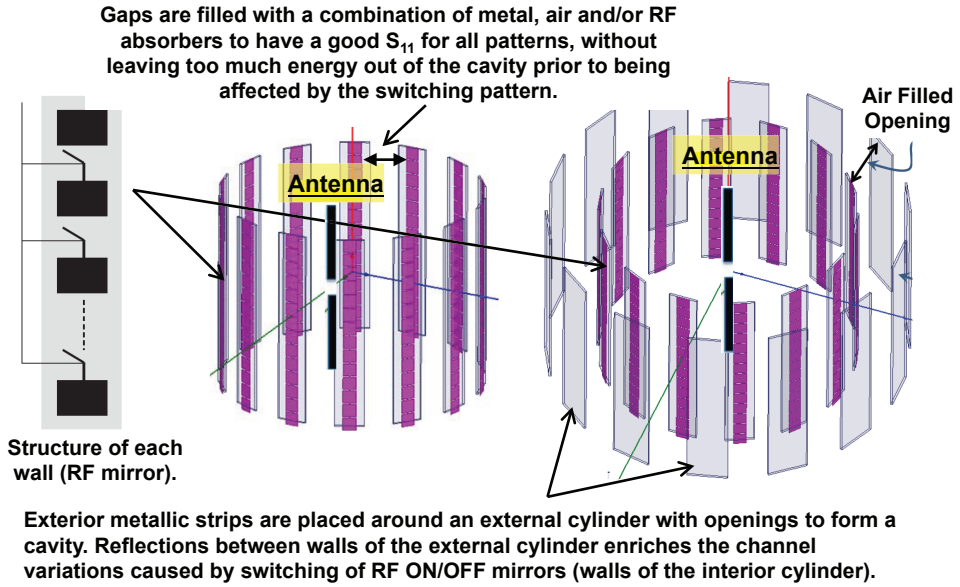


Fig. 17: Details for the example of a cylindrical structure with RF mirrors (switched periodic structure) surrounding a main transmit antenna.

Systems Technology, VITAE 2009.

[2] O. N. Alrabadi, A. Kalis, C. B. Papadias, R. Prasad,"A universal encoding scheme for MIMO transmission using a single active element for PSK modulation schemes", *IEEE Transactions on Wireless Communications*, Volume:8 , Issue: 10, Page(s):5133 5142

[3] R. Bains, "On the Usage of Parasitic Antenna Elements in Wireless Communication Systems', PhD Thesis

[4] G. J. Foschini, "Layered spacetime architecture for wireless communication in a fading environment when using multiple antennas, *Bell Labs Syst. Tech. J.*, vol. 1, p. 4159, 1996.

[5] E. Telatar "Capacity of Multi-antenna Gaussian Channels", *European Trans. on Telecomm*10 (6): 585595.

[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Spacetime codes for high data rate wireless communication: Performance analysis and code construction", *IEEE Trans. on Information Theory* 44 (2): 744765

[7] L. Zheng and D. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels, *IEEE Transactions*

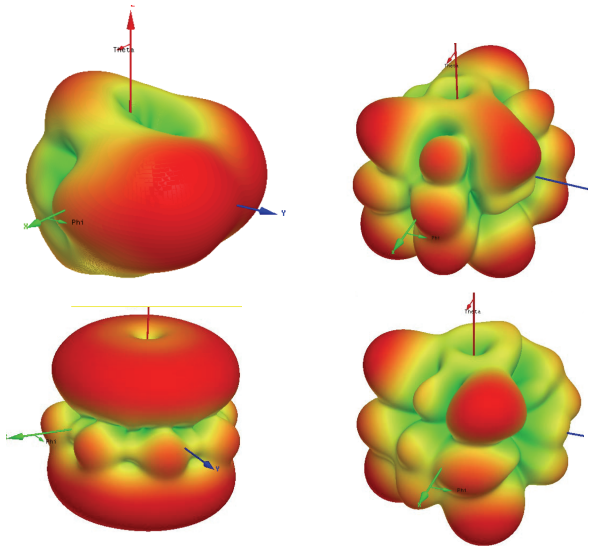


Fig. 18: Examples of antenna patterns corresponding to the periodic switched structure in Fig. 17.

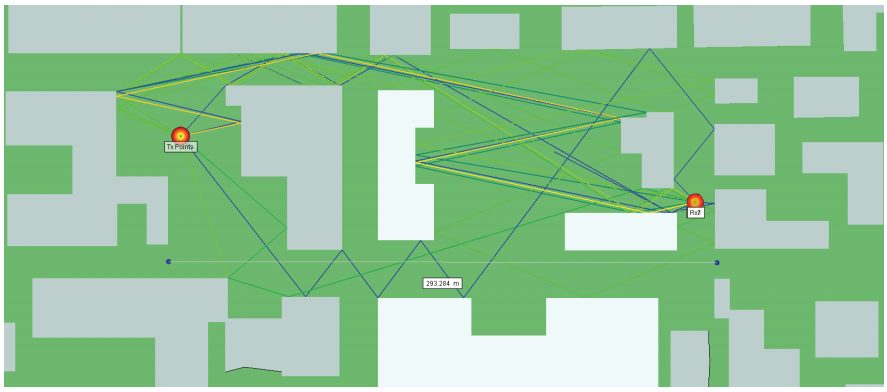


Fig. 19: Propagation environments corresponding to the MBM constellation points shown in Fig. 20

.

on Information Theory, vol. 49, May 2003, pp. 1073-96

[8] Edelman, Rao. Random matrix theory. Acta # 14:233297, 2005.

[9] *Information Theory and Reliable Communication*, Robert G. Gallager, Wiley, 1968

[10] *Principles of communication engineering,* John M. Wozencraft, Irwin M. Jacobs Wiley, 1965

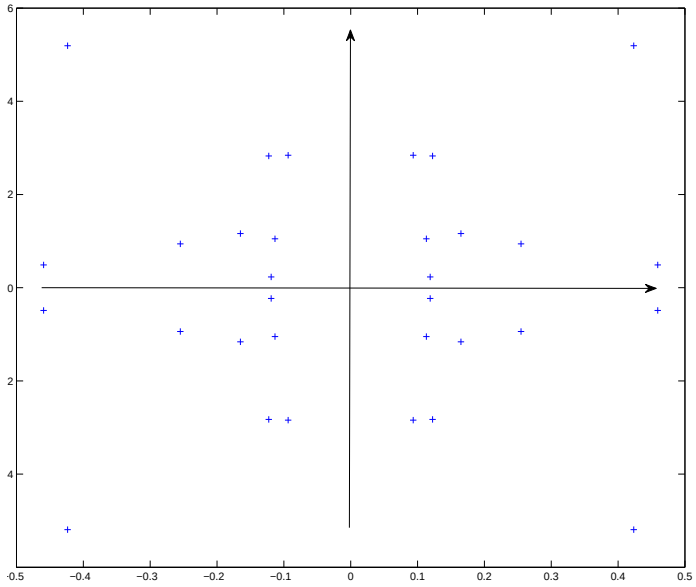


Fig. 20: Examples of constellation points corresponding to the periodic switched structure in Fig. 17 and propagation environments in Fig. 20 (constellation is made symmterical).

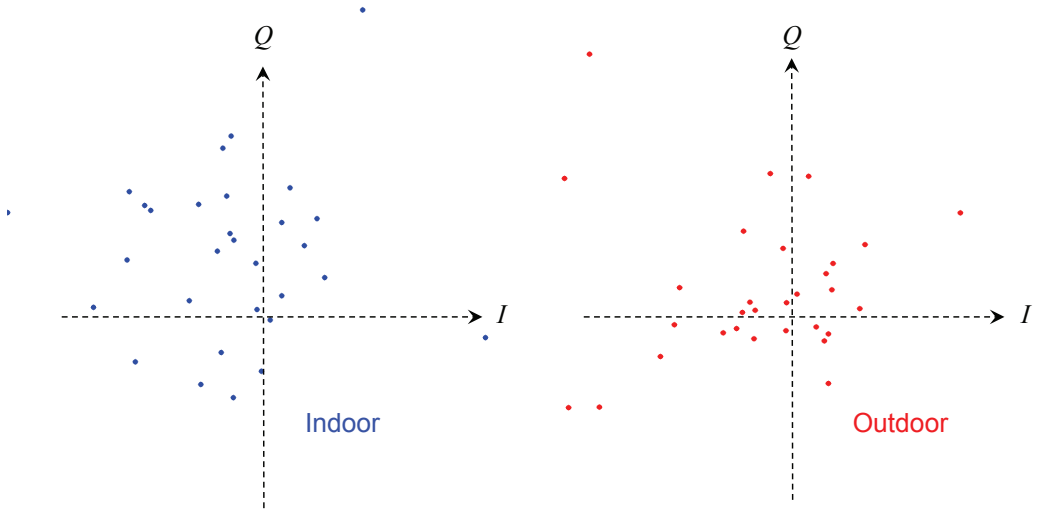
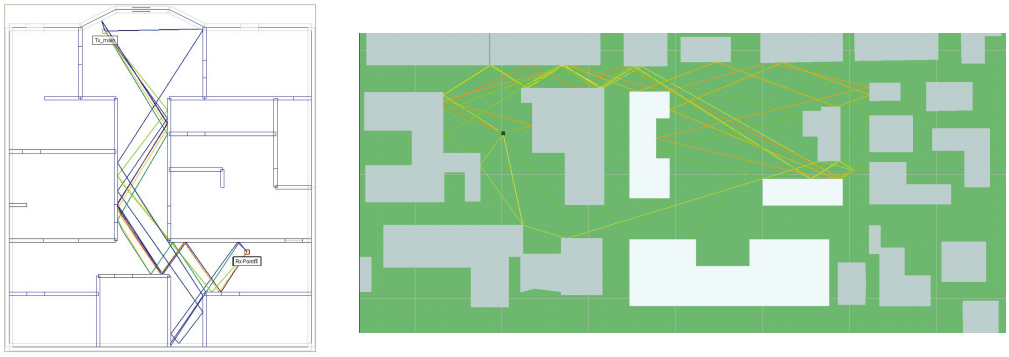


Fig. 21: Two other examples of constellation points corresponding to the periodic switched structure in Fig. 17 (raw constellation points).



Indoor (residential building with dry-walls) Outdoor Model (down-town Ottawa)

Fig. 22: Indoor/Outdoor Propagation environments corresponding to the MBM constellation points shown in Fig. 21

.

APPENDIX A
SLOPE OF RATE VS. SNR AT $SNR \to 0$

Let us consider a single antenna transmitting a signal at power $\epsilon = \alpha/Q$, i.e., total receive power over $Q = 2K$ receive dimensions is α. We have,

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{2\pi}\right)^{(Q/2)} \exp\left(\frac{|\vec{y}-\sqrt{\alpha}\vec{h}_i|^2}{2}\right)$$

On the other hand,

$$\begin{aligned}
&\left(\frac{1}{2\pi}\right)^{(Q/2)} exp\left(\frac{|\vec{y}-\sqrt{\alpha}\vec{h}_i|^2}{2}\right) = \left(\frac{1}{2\pi}\right)^{(Q/2)} \exp\left(-\frac{||\vec{y}||^2}{2}\right)\exp\left(\sqrt{\alpha}\vec{h}_i.\vec{y}\right)\exp\left(-\frac{\alpha||\vec{h}_i||^2}{2}\right)\\
\simeq\ &\left(\frac{1}{2\pi}\right)^{(Q/2)} exp\left(-\frac{||\vec{y}||^2}{2}\right)\left(1+\sqrt{\alpha}\vec{h}_i.\vec{y}+\alpha(\vec{h}_i.\vec{y})^2\right)\left(1-\frac{\alpha||\vec{h}_i||^2}{2}\right)\\
\simeq\ &N_Y(0,I_Q)\left(1+\sqrt{\alpha}\sum_{k=1}^{Q}h_{ik}y_k+\frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q}h_{ik_1}h_{ik_2}y_{k_1}y_{k_2}\right)\left(1-\frac{\alpha||\vec{h}_i||^2}{2}\right).
\end{aligned}$$

where,

$$N_Y(0,I_Q) = \left(\frac{1}{2\pi}\right)^{(Q/2)} exp\left(-\frac{||\vec{y}||^2}{2}\right).$$

Using G_1 and G_2 to represent the first and the second sample moments of the set of M constellation points, i.e.,

$$G_1 \equiv \frac{1}{M}\sum_{i=1}^{M}\vec{h}_i \quad \text{and} \quad G_2 \equiv \frac{1}{M}\sum_{i=1}^{M}\vec{h}_i\vec{h}_i^t$$

we obtain,

$$f_{\vec{Y}}(\vec{y}) \simeq N_Y(0,I_Q)\left(1-\frac{\alpha}{2}\sum_{k=1}^{Q}G_2(k,k)+\sqrt{\alpha}\sum_{k=1}^{Q}G_1(k)y_k+\frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q}G_2(k_1,k_2)y_{k_1}y_{k_2}\right)$$

and

$$\log f_{\vec{Y}}(\vec{y}) \simeq \log N_Y(0, I_Q) +$$
$$\left(-\frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k) + \sqrt{\alpha}\sum_{k=1}^{Q} G_1(k)Y_k + \frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q} G_2(k_1,k_2)y_{k_1}y_{k_2} - \frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q} G_1(k_1)G_1(k_2)Y_{k_1}Y_{k_2} \right)$$

Thus

$$\log f_{\vec{Y}}(\vec{y}) \simeq -\frac{Q}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{k=1}^{Q} y_k^2\right) +$$
$$\left(-\frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k) + \sqrt{\alpha}\sum_{k=1}^{Q} G_1(k)y_k + \frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q} G_2(k_1,k_2)y_{k_1}y_{k_2} - \frac{\alpha}{2}\sum_{k_1=1}^{Q}\sum_{k_2=1}^{Q} G_1(k_1)G_1(k_2)y_{k_1}y_{k_2} \right)$$

and

$$h(\vec{y}) = -\int f_{\vec{Y}}(\vec{y})\log f_{\vec{Y}}(\vec{y})d\vec{y} \simeq$$
$$\frac{Q}{2}(\log 2\pi e)\left(1 - \frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k)\right) + \frac{\alpha}{2}\left(\sum_{k=1}^{Q} G_2(k,k)\right)\left(1 - \frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k)\right)$$
$$-\frac{\alpha}{2}\left(\sum_{k=1}^{Q} G_2(k,k)\right)\left(1 - \frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k)\right)$$
$$+\frac{\alpha}{2}\left(\sum_{k=1}^{Q} G_1^2(k)\right)\left(1 - \frac{\alpha}{2}\sum_{k=1}^{Q} G_2(k,k)\right) - \alpha\sum_{k=1}^{Q} G_1^2(k) + \frac{\alpha Q}{4}\log 2\pi\sum_{k=1}^{Q} G_2(k,k) + \frac{\alpha}{4}(Q+2)\sum_{k=1}^{Q} G_2(k,k)$$

This result in,

$$I(\vec{h},\vec{y}) = h(\vec{y}) - h(\vec{z}) \simeq \frac{\alpha}{2}\left(\sum_{k=1}^{Q} G_2(k,k) - \sum_{k=1}^{Q} G_1^2(k)\right)$$

If we spend $\epsilon = \alpha/Q$ unit of power in transmitter then we have $\alpha = \epsilon Q$ unit of power in receiver. Thus:

$$I(\vec{h},\vec{y}) \simeq \frac{Q\epsilon}{2}\left(\sum_{k=1}^{Q} G_2(k,k) - \sum_{k=1}^{Q} G_1^2(k)\right)$$

$$I^{'} = \lim_{\epsilon \longrightarrow 0}\frac{I(\epsilon)}{\epsilon} \simeq \frac{Q}{2}\left(\sum_{k=1}^{Q} G_2(k,k) - \sum_{k=1}^{Q} G_1^2(k)\right)$$

APPENDIX B
MUTUAL INFORMATION OF A FINITE RANDOM CONSTELLATION WITH I.I.D. GAUSSIAN POINTS

Consider the random coding scheme in Fig. 4, which selects the constellation points, $\vec{h}(m)$, $m = 1, \ldots, M$, with equal probability. We have,

$$\vec{y} = \vec{h} + \vec{z}. \tag{3}$$

Let us consider,

$$\vec{\hat{y}} = \vec{h} + \vec{n}_q + \vec{z} \tag{4}$$

where $\vec{h}+\vec{n}_q$, and consequently, $\vec{\hat{y}}$, has a Gaussian distribution. Using data processing theorem [9], we have:

$$I(\vec{h};\vec{y}) \geq I(\vec{h};\vec{\hat{y}}) \tag{5}$$

$$\begin{aligned} I(\vec{h};\vec{\hat{y}}) &= h(\vec{\hat{y}}) - h(\vec{\hat{y}} \mid \vec{h}) \\ &= h(\vec{\hat{y}}) - h(\vec{n}_q + \vec{z} \mid \vec{h}) \\ &= \frac{1}{2}\log(2\pi e\sigma^2_{\hat{\vec{y}}}) - h(\vec{n}_q + \vec{z} \mid \vec{h}) \end{aligned} \tag{6}$$

To obtain a computable lower bound on the capacity, the entropy of the combined additive noise, i.e., $\vec{n}_q+\vec{z}$, is upper-bounded by the entropy of an i.i.d. Gaussian noise of the same conditional variance. This results in,

$$E_{\vec{h}}\left[I(\vec{h};\vec{\hat{y}})|\vec{h}\right] \geq \frac{Q}{2}\log\left(2\pi e\sigma^2_{\hat{y}}\right) - QE_{\vec{h}}\left[\frac{1}{2}\log\left(2\pi e\sigma_z^2 + 2\pi e\sigma^2_{n_q}\right) \mid \vec{h}\right] \tag{7}$$

where $\sigma^2_{\hat{y}} = 1+\sigma_z^2$ (signal power is normalized to one) and $\sigma^2_{n_q}$ is the variance of the quantization noise per dimension. Inequality 7 holds since entropy of $\vec{n}_q+\vec{z}$ is bounded by the entropy of an i.i.d. Gaussian vector of the same variance. Noting convexity of the $\log$ function, we can further extend the chain of inequalities using,

$$E_{\vec{h}}\left[\frac{1}{2}\log\left(2\pi e\sigma_z^2 + 2\pi e\sigma^2_{n_q}\right) \mid \vec{h}\right] \geq E_{\vec{h}}\left[\frac{1}{2}\log\left(2\pi e\sigma_z^2 + 2\pi e\sigma^2_{n_q|\vec{h}}\right)\right]. \tag{8}$$

For a given $\vec{h}$, let us use the notaion $\vec{g}$ to refer to the Gaussian vector at the minimum distance to $\vec{h}$, i.e., $\vec{h}$ is quantized to $\vec{g}$. We have,

$$\sigma^2_{n_q|\vec{h}} = \frac{1}{Q}E\left[\vec{n}_q^2 \mid \vec{h}\right] = \frac{1}{Q}\int\limits_{\vec{g}\in\mathcal{R}^Q} ||\vec{h}-\vec{g}||^2 f_{\vec{G}|\vec{H}}(\vec{g} \mid \vec{h})d\vec{g} = \frac{1}{Q}\int\limits_{\vec{g}\in\mathcal{R}^Q} ||\vec{h}-\vec{g}||^2 \frac{f_{\vec{G},\vec{H}}(\vec{g},\vec{h})}{f_{\vec{H}}(\vec{h})}d\vec{g}. \tag{9}$$

where $f_{\vec{H}}(\vec{.})$, $f_{\vec{G},\vec{H}}(\vec{.},\vec{.})$ and $f_{\vec{G}|\vec{H}}(\vec{.}|\vec{.})$ are the marginal, conditional and joint distribution of $\vec{h}$ and $\vec{g}$, respectively. Distributions $f_{\vec{H}}(\vec{h})$ and $f_{\vec{G}}(\vec{g})$ will be subsequently replaced with $\mathbf{N}(\vec{h})$ and $\mathbf{N}(\vec{g})$ to denote i.i.d. Gaussian.

Note that, by using the time average of mutual information in setup of Fig. 4(a), which is replaced with the statistical average over the setup of Fig. 4(b), and using step above, power of quantization noise is averaged over all possible realizations of the constellation. This results in simplifying the solution, by avergaing over different realizations of the M-points constellation.

For the sake of clarity, let us explicitly include the index of the selected constellation point as a random variable, resulting in,

$$\sigma^2_{n_q|\vec{h}} = \frac{1}{Q}\int\limits_{\vec{g}\in\mathcal{R}^Q} ||\vec{h}-\vec{g}||^2 \frac{\sum_{i=1}^{M} f_{\vec{G},\vec{H},I}(\vec{g},\vec{h},i)}{f_{\vec{H}}(\vec{h})}d\vec{g} = \frac{1}{Q}\int\limits_{\vec{g}\in\mathcal{R}^Q} ||\vec{h}-\vec{g}||^2 \frac{M f_{\vec{G},\vec{H},I}(\vec{g},\vec{h},\vec{i})}{f_{\vec{H}}(\vec{h})}d\vec{g} \tag{10}$$

where index $\check{i} \in \{1, ..., M\}$ is used to reflect that the corresponding terms are all equal to each other, $f_{\vec{G},\vec{H},I}(\vec{g}, \vec{h}, i)$ is the joint distrubtuion of the event: ith message is mapped to the constellation point $\vec{h}$, which is qunantized to $\vec{g}$.

Note that the event,

$$\{\vec{G} = \vec{g}, \vec{H} = \vec{h}, I = \check{i}\} \tag{11}$$

is equivalent to:

$$\{\vec{G} = \vec{g}, \vec{h}(\check{i}) = \vec{h},\ \vec{h}(j) \notin \mathcal{S}^Q(\vec{g}, ||\vec{h} - \vec{g}||),\ \forall j \neq \check{i}\}, \tag{12}$$

where,

$$\mathcal{S}^Q(\vec{o}, r) = \{\vec{x} \in \mathcal{R}^Q : ||\vec{x} - \vec{o}|| \leq r\} \tag{13}$$

is a sphere of radius r centered at $\vec{o}$. Note that 12 captures the operation of quantization. Using $\mathbf{N}(\vec{.})$ to show the joint density of a Gaussian vector with zero mean and unit variance, we can rewrite (10) as follows:

$$\frac{M}{Q} \int\limits_{\vec{g} \in \mathcal{R}^Q} ||\vec{h} - \vec{g}||^2 \frac{\mathbf{N}(\vec{h})\mathbf{N}(\vec{g}) \left(1 - P(\vec{g}, \vec{h})\right)^{M-1}}{\mathbf{N}(\vec{h})} d\vec{g} \tag{14}$$

where $P(\vec{g}, \vec{h})$ is the probability of event $\{\vec{h} \in \mathcal{S}^Q(\vec{g}, ||\vec{h} - \vec{g}||)\}$ which is the probability that a Gaussian point falls in a Q-dimensional hyper-sphere, $\mathcal{S}^Q$, centered around $\vec{g}$ with radius $r = ||\vec{h} - \vec{g}||$. This means,

$$P(\vec{g}, \vec{h}) = \int\limits_{\vec{x} \in \mathcal{S}^Q(\vec{g}, ||\vec{h} - \vec{g}||)} \mathbf{N}(\vec{x}) d\vec{x} \tag{15}$$

Applying the inequality $\ln(A) \leq A - 1,\ \forall A > 0$ to 14, results in

$$\begin{aligned} &\frac{M}{Q} \int\limits_{\vec{g} \in \mathcal{R}^Q} ||\vec{h} - \vec{g}||^2 \mathbf{N}(\vec{g}) \left(1 - P(\vec{g}, \vec{h})\right)^{M-1} d\vec{g} \leq \\ &\frac{M}{Q} \int\limits_{\vec{g} \in \mathcal{R}^Q} ||\vec{h} - \vec{g}||^2 \mathbf{N}(\vec{g}) \exp\left[-(M-1)P(\vec{g}, \vec{h})\right] d\vec{g} \end{aligned} \tag{16}$$

The leading asymptotic behaviour of the integral in 16 is obtained by following Laplace method. Consider integrals of general form:

$$S(M) = \int_a^b \psi(x) \exp\left[M\phi(x)\right] dx \tag{17}$$

If the real continuous function $\phi(x)$ has its maximum in the interval $a \leq x \leq b$ at an intermediate point $x = c$, then it is only the immediate neighbour of $x = c$ that contributes to asymptotic expansion of $S(M)$. In our case,

$$S(M) = \frac{M}{Q} \int\limits_{\vec{g} \in \mathcal{R}^Q} \mathbf{N}(\vec{g}) ||\vec{h} - \vec{g}||^2 \exp\left[-(M-1)P(\vec{g}, \vec{h})\right] d\vec{g}.$$

On the other hand, $\phi(\,.\,) = -P(\vec{g},\vec{h})$ is maximized at $\vec{g}=\vec{h}$. For this reason, for large values of M, the leading asymptotic behavior of the intergral is governed by values of $\vec{g}$ within a small sphere, say of radious r_0, around $\vec{h}$. To find the first order term of the asymptotic behaviour, we may approximate the integral as,

$$S(M) \simeq \frac{M}{Q} \int\limits_{\vec{g}\in\mathcal{S}^Q(\vec{h},r_0)} \mathbf{N}(\vec{g})||\vec{h}-\vec{g}||^2 \exp\left[-(M-1)P(\vec{g},\vec{h})\right] d\vec{g} \tag{18}$$

where r_0 is small enough such that $P(\vec{g},\vec{h})$, i.e., the integration of Gaussian distribution over hyper-sphere $\mathcal{S}^Q(\vec{h},r_0)$, can be approximated with $\mathbf{N}(\vec{h})$ times the corresponding volume. On the other hand, due to spherical symmetry of multivariate Gaussian distribution, $\sigma^2_{n_q|\vec{h}}$ only depends on norm $||\vec{h}||$, hence it only suffices to compute the conditional variance for points of the form $\vec{h}=\frac{||\vec{h}||}{\sqrt{Q}}(1,...,1)_Q$, where $(1,...,1)_Q$ is the all-one vector of size Q. Therefore,

$$\begin{aligned} S(M) \simeq & \frac{M}{Q} \int\limits_{\vec{g}\in\mathcal{R}^Q} \mathbf{N}(\vec{h}) \left[\left(g_1-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2 + ... + \left(g_Q-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2\right] \\ & \exp\left\{-(M-1)\mathbf{N}(\vec{h})\frac{\pi^{Q/2}}{\Gamma(Q/2+1)}\left[\left(g_1-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2 ... + \left(g_Q-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2\right]^{Q/2}\right\} dg_1...dg_Q \end{aligned}$$

Hence,

$$\begin{aligned} & \lim_{M\to\infty} \frac{M}{Q} \int\limits_{\vec{g}\in\mathcal{R}^Q} \mathbf{N}(\vec{g})||\vec{h}-\vec{g}||^2 \exp\left[-(M-1)P(\vec{g},\vec{h})\right] d\vec{g} \\ \simeq & \frac{M}{Q} \int\limits_{\vec{g}\in\mathcal{R}^Q} \mathbf{N}(\vec{h}) \left[\left(g_1-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2 + ... + \left(g_Q-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2\right] \\ & \exp\left\{-(M-1)\mathbf{N}(\vec{h})\frac{\pi^{Q/2}}{\Gamma(Q/2+1)}\left[\left(g_1-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2 ... + \left(g_Q-\frac{||\vec{h}||}{\sqrt{Q}}\right)^2\right]^{Q/2}\right\} dg_1...dg_Q \end{aligned} \tag{19}$$

The integral in (19) can be computed using change of variables:

$$\begin{aligned} & \frac{M}{Q} \int\limits_{r\in\mathcal{R}} \mathbf{N}(\vec{h}) r^2 \exp\left[-(M-1)\mathbf{N}(\vec{h})\frac{\pi^{Q/2}}{\Gamma(Q/2+1)}r^Q\right] \frac{\pi^{Q/2}}{\Gamma(Q/2+1)} dr^Q \\ & = \frac{M}{Q} \int\limits_{s\in\mathcal{R}} \mathbf{N}(\vec{h}) \left[\frac{s\Gamma(Q/2+1)}{\pi^{Q/2}}\right]^{2/Q} \exp\left[-(M-1)\mathbf{N}(\vec{h})s\right] ds \\ & = \frac{2\Gamma(2/Q+1)}{Q} \left[\frac{\Gamma(Q/2+1)}{M}\right]^{2/Q} \exp\left(\frac{||\vec{h}||^2}{Q}\right). \end{aligned} \tag{20}$$

Finally, we obtain:

$$\sigma^2_{n_q|\vec{h}} \simeq \frac{2\Gamma(2/Q+1)}{Q} \left[\frac{\Gamma(Q/2+1)}{M}\right]^{2/Q} \exp\left(\frac{||\vec{h}||^2}{Q}\right). \tag{21}$$

This means

$$\sigma^2_{n_q|\vec{h}} \simeq \left(\frac{1}{M}\right)^{2/Q} \to 0, \quad as \quad M\to\infty.$$

Figure (23) shows an example for the accuracy of the expressions computed above.

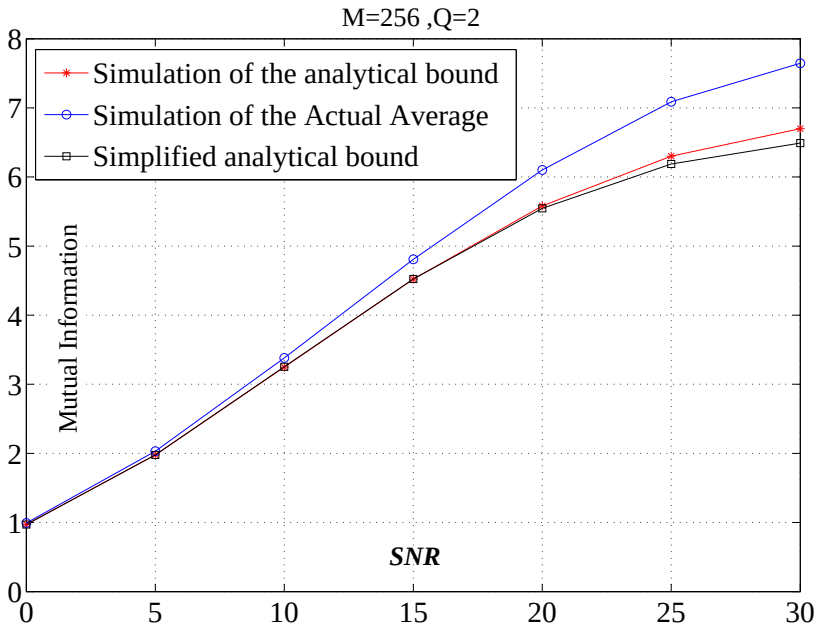


Fig. 23: An example for the accuracy of the capacity exressions presented in Appendix B. Curve specified as "Simulation of the analytical bound" correspond to using monte-carlo simulaiton for cimpute the integral in (9), and the curve specified as "Simplified analytical bound" corresponds to using (21).

APPENDIX C
MAXIMUM SQUARE NORM, AND SELECTION GAIN AT LOW *SNR* VALUES

Let $X_{1:M}, X_{2:M}, ..., X_{M:M}$ be order statistics obtained from samples $X_1, ..., X_M$. It is well known that for any continuous Cumulative Distribution Function $F(x)$, the s'th order statistic of M random variables,

$$X_{s:M} = F^{-1}(U_{s:M}) \tag{22}$$

where $U_{s:M}$ is the s'th order statistic of standard uniform distribution. By expanding $F^{-1}(U_{s:M})$ in Taylor expansion around $E[U_{s:M}] = \frac{s}{M+1}$ and then taking expected values of both side, the following expansion for expected value of the s'th order statistic of M random variables is obtained:

$$\begin{aligned} E[X_{s:M}] \simeq\ & F^{-1}(p_s) + \frac{p_s q_s}{2(M+2)} D^2(p_s) + \\ & \frac{p_s q_s}{(M+2)^2} \left[1/3(q_s - p_s) D^3(p_s) + 1/8(q_s p_s) D^4(p_s) \right] \end{aligned} \tag{23}$$

where $D^i = (F^{-1})^i$ represents the i'th order derivative of F^{-1} and $p_s = 1 - q_s = \frac{s}{M+1}$. For a large sample the expected value of s'th order statistic can be approximated by the first term in (23),

$$E[X_{s:M}] \simeq F^{-1}\left(\frac{s}{M+1}\right) \tag{24}$$

In case of M points generated i.i.d according to Gaussian distribution, the square norm has a chi-squared distribution and cumulative distribution function can be expressed in terms of lower incomplete gamma function as,

$$F(||\vec{h}||^2; Q) = 1 - \frac{\Gamma(Q/2, ||\vec{h}||^2/2)}{\Gamma(Q/2)} \tag{25}$$

where Q is the number of dimensions. For large enough $||\vec{h}||^2$, equation 25 can be approximated as,

$$F(||\vec{h}||^2; Q) \simeq 1 - \frac{(||\vec{h}||^2)^{Q/2-1} e^{-||\vec{h}||^2}}{\Gamma(Q/2)} \tag{26}$$

Hence, for s close enough to M, the quantile function which is the inverse of CDF is written in terms of Lambert W-function, W_{-1}. Consequently, expected of value of s'th order statistic for square norm of Gaussian random variables is approximated as follows:

$$\begin{aligned} E[||\vec{h}||^2{}_{s:M}] &\simeq -2(Q/2-1) W_{-1}\left(\frac{-1}{Q/2-1}\left[\Gamma(Q/2)\left(1-\frac{s}{M+1}\right)\right]^{\frac{1}{Q/2-1}}\right) \\ &\simeq -2(Q/2-1)\left(\ln\left[\frac{-1}{Q/2-1}\left(\Gamma(Q/2)\left[1-\frac{s}{M+1}\right]\right)^{\frac{1}{Q/2-1}}\right]\right) \\ &- \ln\left(-\ln\left[\frac{-1}{Q/2-1}\left(\Gamma(Q/2)\left[1-\frac{s}{M+1}\right]\right)^{\frac{1}{Q/2-1}}.\right]\right) \end{aligned} \tag{27}$$

Therefore,

$$\begin{aligned} E[||\vec{h}||^2{}_{s:M}] &\simeq 2\left[(Q/2-1)\ln(Q/2-1) + \ln\left(\frac{M+1}{\Gamma(Q/2)(M+1-s)}\right)\right] \\ &+ (Q/2-1)\ln\left(-\ln\left(\frac{-1}{Q/2-1}\left[\Gamma(Q/2)(1-\frac{s}{M+1})\right]^{\frac{1}{Q/2-1}}\right)\right) \end{aligned} \tag{28}$$

The leading asymptotic behaviour for the expected value of the maximum square norm of M independent Gaussian random can be obtained by replacing s by M in 28:

$$\begin{aligned} &E\left[\max\left(||\vec{h}(1)||^2, ..., ||\vec{h}(M)||^2\right)\right] = E[||\vec{h}||^2_{M:M}] \simeq \\ &2\left((Q/2-1)\ln(Q/2-1) + \ln\left[\frac{M+1}{\Gamma(Q/2)}\right]\right) + (Q/2-1)\ln\left(-\ln\left[-\frac{\Gamma(Q/2)}{(M+1)(Q/2-1)}\right]^{\frac{1}{Q/2-1}}\right) \end{aligned} \tag{29}$$

Finally,

$$\begin{aligned} &E\left[\max\left(||\vec{h}(1)||^2, ..., ||\vec{h}(M)||^2\right)\right] = E[||\vec{h}||^2_{M:M}] \simeq \\ &2\left((Q/2-1)\ln(Q/2-1) + \ln\left[\frac{M+1}{\Gamma(Q/2)}\right]\right) + (Q/2-1)\ln\left(\ln\left[\frac{(M+1)(Q/2-1)}{\Gamma(Q/2)}\right]^{\frac{1}{Q/2-1}}\right) \end{aligned} \tag{30}$$

Ignoring the $\ln\ln(.)$ term, we obtain:

$$E\left[\max\left(||\vec{h}(1)||^2, ..., ||\vec{h}(M)||^2\right)\right] \simeq H(Q) + \ln(M+1),$$

where

$$H(Q) = 2\left[(Q/2-1)\ln(Q/2-1) - \ln\Gamma(Q/2)\right].$$

Table (I) shows the values of the offset term, $H(Q)$, for different values of Q. We have $H(Q) \simeq K = 2Q$ for large K.

$K = 2Q$	2	4	8	16	32	64
$H(Q)$	0	1.5	5	12.7	28.4	60

TABLE I: Example for the values of $H(Q)$ as a function of $K = 2Q$, K is the number of receive antennas.

Figure (24) shows examples for the accuracy of the expressions computed above.

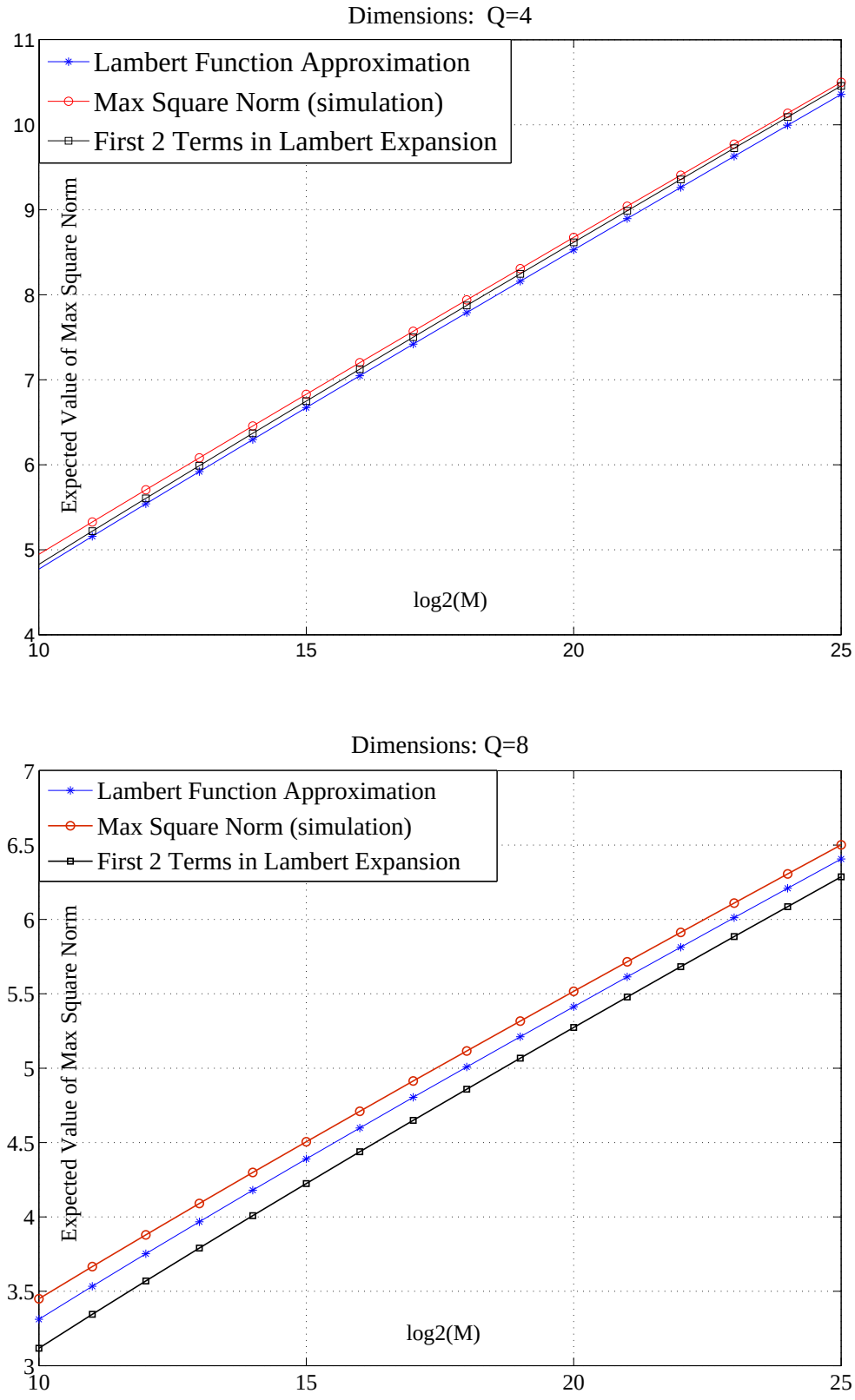


Fig. 24: An example for the accuracy of the capacity exressions presented in Appendix B.