Broadcasting in Massive MIMO Using OSTBC with Reduced Dimension

Marcus Karlsson, Emil Björnson and Erik G. Larsson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Broadcasting in Massive MIMO Using OSTBC with Reduced Dimension

Marcus Karlsson, Emil Björnson and Erik G. Larsson

Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden

Email: {marcus.karlsson, emil.bjornson, erik.g.larsson}@liu.se.

Abstract—An analysis of broadcasting in massive MIMO (multiple-input and multiple-output) systems with a limited coherence interval is presented. When broadcasting common information, such as control signals, the base station does not have channel state information to the terminals. We propose that the base station broadcasts this common information using a low dimensional orthogonal space-time block code (OSTBC). This code is mapped onto the large antenna array with the use of a dimension reducing matrix, effectively "shrinking" the channel. The terminal can estimate the effective channel and decode the information, even when the coherence interval is short compared to the number of base station antennas. Different OSTBCs are compared in terms of outage capacity in practical scenarios using estimated CSI. In particular, the trade-off between diversity and rate, when little or no time/frequency diversity is available, is investigated.

I. Introduction

Massive MIMO (multiple-input and multiple-output) is a contender for 5G wireless networks technology, where each base station (BS) is equipped with a large number of antennas, in the order of hundreds, serving many users in the same time-frequency resource utilizing the large number of degrees of freedom to perform spatial beamforming. To reap the benefits of the massive array, to perform beamforming, the BS requires channel state information (CSI), which is acquired through uplink pilot transmission from the terminals.

Much previous work on massive MIMO has focused on precoders/detectors, channel estimation, pilot allocation or resource allocation, to improve the spectral efficiency of the payload transmission. However, little attention has been given to broadcasting channels, where the BS transmits common information to all users in the cell, including the ones which are not currently being served. One example is the transmission of control information—including system information, paging information and random access information—which is necessary for the system operation. In a massive MIMO system, the BS will have to use these broadcast channels to tell the inactive terminals to send pilots. This will probably be a very short message, but still needs to be conveyed reliably. Reliability is key here, since failure to detect causes delays and false detection may cause unnecessary pilot contamination [1].

For the terminal to be aware of everything happening in the cell, the terminal needs to continuously decode the broadcasted information, even when it is idle. This calls for a low complexity detection method to save energy.

When transmitting short packets, the whole packet may only span one channel realization, consequently, time or frequency diversity may not be available. Thus, in order to convey a message reliably over a fading channel when little or no time/frequency diversity is available, extra spatial diversity from the coded transmission is desirable. In light of this, as well as the need for low complexity detectors, this paper considers broadcasting information using orthogonal space-time block codes (OSTBCs). An OSTBC offers spatial diversity on top of the time/frequency diversity offered by the channel and is very simple to decode.

In general, when transmitting longer packets, the diversity offered by a space-time code might not be as useful as when transmitting short packets, for a number of reasons. Firstly, in a low mobility scenario, the coherence interval is long, so estimating the channel very accurately at the BS, through feedback, is feasible. Secondly, in high mobility scenarios, or if the channel is highly disperse, the channel offers enough time/frequency diversity so the spatial diversity from the code is redundant [2]. However, when short packets that require low error probability are conveyed, space-time diversity may be beneficial [2]. Moreover, looking at the outage performance is more relevant in this scenario than the ergodic data rate that assumes coding over many fading realizations.

With a limited coherence interval and a large number of antennas at the BS, it is infeasible to send a pilot block from the BS in order for the terminals to learn the complete channels [3], [4]. In [3] we formulated this problem and suggested a simple solution by spatial repetition of a lower dimensional code over the antenna array. However, no further analysis was conducted. The paper [4] follows a similar idea and optimizes the mapping from the lower dimensional code to the antennas to make the transmitted signal have constant envelope and the BS to emit the signals isotropically. The main effect of this mapping, however, is that it decreases the channel dimension.

This paper investigates how to choose the mapping from the low dimensional code to the antenna array and analyzes the performance of different OSTBCs in several practical scenarios.

II. System Model

We consider downlink transmission, without CSI at the BS, in a single cell system. The BS is equipped with M_0 antennas and serves a number of single antenna terminals.

Furthermore, the terminals have high mobility, leading to a limited coherence interval. We let τ_c denote the length of the coherence interval, measured in transmission symbols.

As all terminals will receive the same information, we consider a single arbitrary terminal within the cell. If τ is the block length, where $\tau \leq \tau_c$, the $1 \times \tau$ received signal vector at the terminal can be expressed as

$$\boldsymbol{y}^{\mathrm{H}} = \boldsymbol{g}^{\mathrm{H}} \boldsymbol{X}_0 + \boldsymbol{w}^{\mathrm{H}}, \qquad (1)$$

where $\boldsymbol{g}^H \in \mathbb{C}^{1 \times M_0}$, $\boldsymbol{X}_0 \in \mathbb{C}^{M_0 \times \tau}$ and $\boldsymbol{w}^H \in \mathbb{C}^{1 \times \tau}$ are the channel from the BS to the terminal, the transmitted symbol block and white Gaussian noise, respectively. $\boldsymbol{X}_0$ in (1) can represent both a pilot block, in case of channel estimation, and a data block, in the case of data transfer.

For channel estimation, the BS transmits an orthogonal pilot block, from which the terminal applies a maximum-likelihood (ML) or a minimum mean squared error (MMSE) estimator [5]. However, if τ_p ($\tau_p < \tau_c$) is the number of symbols spent on pilots and $\tau_p < M_0$, the pilot block $\boldsymbol{X}_{0p} \in \mathbb{C}^{M_0 \times \tau_p}$ is tall. This means that the ML estimator cannot be used since $(\boldsymbol{X}_{0p}\boldsymbol{X}_{0p}^{\mathrm{H}})^{-1}$ does not exist. While the MMSE can be used, the channel estimate will be confined to the column space of $\boldsymbol{X}_{0p}$. If the coherence interval is limited, but $\tau_c > M_0$, both of these estimation techniques can be used, but the pilot block may eat up a large part of the coherence interval, leaving little room for payload data.

A. The Dimension Reducing Matrix

To still be able to estimate the channel at the terminal, even when the coherence interval is limited, the BS can use a fixed matrix $\boldsymbol{\Phi} \in \mathbb{C}^{M_0 \times M}$, to effectively reduce the channel dimension. Here, M is referred to as the number of *virtual antennas* (VAs). We call $\boldsymbol{\Phi}$ a dimension-reducing (DR) matrix, since it reduces channel dimension from M_0 to M elements.

The BS uses the DR matrix to map a code of smaller dimensions onto the large antenna array. Let $\boldsymbol{X} \in \mathbb{C}^{M \times \tau}$ be a low dimensional code ($M < M_0$). This smaller matrix is mapped onto the array using $\boldsymbol{\Phi}$ as

$$\boldsymbol{X}_0 = \boldsymbol{\Phi}\boldsymbol{X}. \qquad (2)$$

By inserting (2) into (1) we get

$$\boldsymbol{y}^{\mathrm{H}} = \boldsymbol{g}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{w}^{\mathrm{H}} = \boldsymbol{h}^{\mathrm{H}}\boldsymbol{X} + \boldsymbol{w}^{\mathrm{H}}, \qquad (3)$$

where $\boldsymbol{h} \triangleq \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{g} \in \mathbb{C}^{M \times 1}$ is referred to as the *effective channel*. Now, if the DR matrix is chosen in such a way that $M < \tau_p$, then the pilot block $\boldsymbol{X}_p \in \mathbb{C}^{M \times \tau_p}$ is a wide matrix, and standard estimation techniques can be used to estimate all dimensions of the effective channel. The DR matrix basically spreads the low dimensional code matrix over the BS antennas. In order to use this channel estimate, all subsequent data will have to be sent using the same DR matrix. One very simple example of $\boldsymbol{\Phi}$—if we assume $M_0 = Mr$, where r is an integer—is when the symbols are repeated cyclically over the

antennas, giving

$$\boldsymbol{X}_0 = \frac{1}{\sqrt{r}} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X} \\ \vdots \\ \boldsymbol{X} \end{bmatrix} \text{ and } \boldsymbol{\Phi} = \frac{1}{\sqrt{r}} \begin{bmatrix} \boldsymbol{I}_M \\ \boldsymbol{I}_M \\ \vdots \\ \boldsymbol{I}_M \end{bmatrix}.$$

We choose $\boldsymbol{\Phi}$ to be semi-unitary, i.e., $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} = \boldsymbol{I}_M$ in order to make $\mathrm{tr}(\boldsymbol{X}_0\boldsymbol{X}_0^{\mathrm{H}}) = \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}})$. Hence, the output power does not depend on the size of the BS array, but solely on the underlying, lower dimensional code.

Any choice of $\boldsymbol{\Phi}$ confines the effective channel to the subspace spanned by the columns of $\boldsymbol{\Phi}^{\mathrm{H}}$. That is, the BS implicitly beamforms into this subspace. For channels not in this subspace (channels in the approximate nullspace of $\boldsymbol{\Phi}^{\mathrm{H}}$), the channel gains will be small. If the true channel stems from an i.i.d. Rayleigh fading distribution, however, this does not matter because any randomly chosen (semi-unitary) DR matrix will, statistically, be as good as any other, when looking at many channel realizations. This is so since an i.i.d. Rayleigh fading channel is statistically invariant to rotations. If $\boldsymbol{g} \sim \mathcal{CN}(0, \beta\boldsymbol{I}_{M_0})$, then $\boldsymbol{h} \sim \mathcal{CN}(0, \beta\boldsymbol{I}_M)$, so the DR matrix simply makes the effective channel smaller.

An alternative to choosing a random DR matrix, is to optimize the DR matrix with respect to a particular metric. In [4], the authors optimize this DR matrix to have constant envelope and to make the BS emit energy isotropically. Here, however, we let the DR matrix be semi-unitary and analyze how to choose the dimensions of this matrix in different scenarios.

However, if the channel *does* have a structure some choices of the DR matrix will be better than others. To avoid beam-forming in the wrong direction for an extended period of time when the BS has no knowledge of the structure, the BS can use a set of different DR matrices, and alternate between these. If there is a statistical structure that the BS is aware of, the BS can exploit this based on the dominating eigenspaces, similar to the pilot optimization performed in [6].

In the remainder of this article, the word "channel" refers to the smaller effective channel, $\boldsymbol{h}$ with M elements, unless otherwise explicitly stated.

B. Channel Estimation

For the terminal to learn the channel, the BS transmits a pilot block $\boldsymbol{X}_p \in \mathbb{C}^{M \times \tau_p}$, known a priori to the terminal. The terminal then computes the MMSE estimate of the channel [5], i.e.,

$$\hat{\boldsymbol{h}} = \boldsymbol{C}_h\boldsymbol{X}_p(\boldsymbol{I} + \boldsymbol{X}_p^{\mathrm{H}}\boldsymbol{C}_h\boldsymbol{X}_p)^{-1}\boldsymbol{y}, \qquad (4)$$

where it is assumed that the temporal noise correlation is identity. Furthermore, we let the channel covariance matrix

$$\boldsymbol{C}_h \triangleq \mathbb{E}[\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}] = \beta\boldsymbol{I}_M,$$

so (4) can be simplified to

$$\begin{aligned} \hat{\boldsymbol{h}} &= \beta\boldsymbol{X}_p(\boldsymbol{I} + \beta\boldsymbol{X}_p^{\mathrm{H}}\boldsymbol{X}_p)^{-1}\boldsymbol{y} \\ &= (\boldsymbol{I}\beta^{-1} + \boldsymbol{X}_p\boldsymbol{X}_p^{\mathrm{H}})^{-1}\boldsymbol{X}_p\boldsymbol{y}, \end{aligned} \qquad (5)$$

where β is a parameter depending on the physical position of the terminal and is assumed to be constant for many coherence intervals. Moreover, the pilot matrix is orthogonal and satisfies

$$X_p X_p^H = \frac{\rho_p \tau_p}{M} I,$$

where ρ_p is the normalized transmit power used for pilots. The channel estimate has zero mean and the covariance

$$\mathbb{E}[\hat{h}\hat{h}^H] \triangleq C_{\hat{h}} = \frac{\beta^2 \rho_p \tau_p}{M + \beta \rho_p \tau_p} I. \quad (6)$$

In addition, the channel estimation errors, $\tilde{h} \triangleq h - \hat{h}$, are statistically independent, have zero mean and covariance $C_{\tilde{h}} = C_h - C_{\hat{h}}$.

Remark: Choosing a larger number M of VAs while spending the same amount of energy on pilots makes the channel estimate worse in the sense that the error variance for each element, $\frac{M\beta}{M + \beta \rho_p \tau_p}$, increases with M.

III. ACHIEVABLE RATES

We derive an achievable rate for when an arbitrary OSTBC is used. Let $X \in \mathbb{C}^{M \times \tau}$ be an OSTBC of the form

$$X = \sum_{n=1}^{n_s} (A_n s_n^R + i B_n s_n^I), \quad (7)$$

where i is the imaginary unit and $\{s_n\}$ ($s_n = s_n^R + i s_n^I$) are the symbols transmitted by the BS. Furthermore, the set of matrices $\{A_n, B_n\}$ is an *amicable orthogonal design*. These matrices define the code and are used, together with the symbols, to build up the codeword X. Moreover, the rate of the code is n_s/τ symbols per channel use.

To decode the data symbols, the terminal treats the estimated channel in (4) as the true channel. Hence, in the following calculations, all expectations are taken conditioned on $\hat{h}$. The estimate of the real (or imaginary) part of symbol s_n is given by multiplying the received signal with $A_n^H \hat{h}$ (or $-i B_n^H \hat{h}$) from the right, followed by taking the real part, i.e., the estimate of the real part of s_n is[1] [7, Theorem 7.3]

$$\begin{aligned}
\hat{s}_n^R &= \Re\left(y^H A_n^H \hat{h}\right) = \Re\left(h^H X A_k^H \hat{h} + w^H A_n^H \hat{h}\right) \\
&= \Re\left(\hat{h}^H X A_n^H \hat{h} - \tilde{h}^H X A_n^H \hat{h} + w^H A_n^H \hat{h}\right) \\
&= ||\hat{h}||^2 s_n^R - \Re\left(e_1^n\right) + \Re\left(e_2^n\right),
\end{aligned} \quad (8)$$

where we in the last equality use the fact that

$$\Re\left(v^H A_k A_n^H v\right) = 0, \ k \neq n \text{ and } \Re\left(iv^H B_k A_n^H v\right) = 0, \ \forall k,$$

for any complex vector v. Moreover, in (8) we have defined the error terms $e_1^n \triangleq \tilde{h}^H X A_n^H \hat{h}$ and $e_2^n \triangleq w^H A_n^H \hat{h}$. The first error term is the error associated with the channel estimation error, while the second error term stems from the white Gaussian noise.

In order to find an achievable rate, the signal-to-noise ratio (SNR) of each decoded symbol is calculated using the conventional approach from [8] to compute a lower bound on mutual

information. The two error terms in (8) are uncorrelated, so the error variance is the sum of the variance of the errors. With

$$\begin{aligned}
\mathbb{E}\left[|e_1^n|^2|\hat{h}\right] &= \mathbb{E}\left[\hat{h}^H A_n X^H \tilde{h} \tilde{h}^H X A_n^H \hat{h}|\hat{h}\right] \\
&= \hat{h}^H A_n \mathbb{E}\left[X^H C_{\tilde{h}} X\right] A_n^H \hat{h},
\end{aligned} \quad (9)$$

and

$$\mathbb{E}\left[|e_2^n|^2|\hat{h}\right] = \mathbb{E}\left[\hat{h}^H A_n w w^H A_n^H \hat{h}|\hat{h}\right] = ||\hat{h}||^2 \quad (10)$$

the SNR for the complex[2] symbol s_n is

$$\text{SNR}_n = \frac{||\hat{h}||^4 \mathbb{E}[|s_n|^2]}{||\hat{h}||^2 + \hat{h}^H A_n \mathbb{E}\left[X^H C_{\tilde{h}} X\right] A_n^H \hat{h}}. \quad (11)$$

Note that the SNR is not necessarily the same for all received symbols for one specific channel estimate. As a special case we consider square OSTBC[3], where $M = \tau$, which implies

$$\mathbb{E}\left[X X^H\right] = \mathbb{E}\left[X^H X\right] = \sum_{n=1}^{n_s} \mathbb{E}\left[|s_n|^2\right] I_M = \frac{\rho_d \tau}{M} I_M,$$

where ρ_d is the normalized transmit power used to transmit data. With this we can simplify the SNR expression in (11) to

$$\text{SNR} = ||\hat{h}||^2 \frac{\rho_d \tau}{M n_s} \bigg/ \left(1 + \frac{\beta \rho_d \tau}{M + \beta \rho_p \tau_p}\right). \quad (12)$$

IV. PERFORMANCE METRICS

In the case of little or no time/frequency diversity and tight latency constraints, the outage capacity is a meaningful performance metric. Furthermore, in this setting, the outage capacity is a good approximation of the maximal achievable rate $R^*(n, \epsilon)$ for blocklength n and error probability ϵ [10].

In this paper we consider short messages, of finite block length and therefore a finite number of *diversity branches* [11] or channel realizations. For an OSTBC with rate n_s/τ, spanning L diversity branches, resulting in SNRs equal to SNR^l, $l = 1, \ldots, L$, the outage probability is

$$p_{\text{out}}(R) = \Pr\left\{\frac{(\tau_c - \tau_p)n_s}{L \tau_c \tau} \sum_{l=1}^{L} \log_2(1 + \text{SNR}^l) < R\right\}, \quad (13)$$

where $(\tau_c - \tau_p)/\tau_c$ is the fraction of the coherence interval spent on payload data. The ϵ-outage rate of a code is defined as the maximum rate, R, for which $p_{\text{out}}(R) < \epsilon$. Note that if we let $L \to \infty$, the outage rate coincides with the ergodic rate of the code.

The number of diversity branches, L, is a measure of how much time/frequency diversity the channel provides. If L is large, the added diversity from the space-time codes is somewhat superfluous and it would probably be better to prioritize high rate instead of high diversity, when choosing code. However, if L is small, the added diversity from the space-time code might be very beneficial. In the following section, we quantify these intuitive claims.

---

[1]For the imaginary part, the calculations are almost identical.

[2]Note that taking the real part of the above errors effectively divides the variance in 2. However, when the variances of both the real and imaginary part are added, the 2 cancels out.

[3]The famous Alamouti [9] scheme is one example of a square OSTBC.

V. Rate versus Diversity

To see the trade-off between rate and diversity, four different OSTBCs are considered, summarized in Table I. The first code is the Alamouti code, and the other three can be found in [7].[4]

TABLE I
Summary of the different codes investigated.

code id	M	τ	n_s	Rate (n_s/τ)
ala	2	2	2	1
343	3	4	3	3/4
443	4	4	3	3/4
884	8	8	4	1/2

The scenario considered is of a single resource-block pair [12] in LTE consisting of 12 subcarriers and 14 OFDM symbols, adding up to a coherence interval of $\tau_c = 168$ symbols. 1 ms (the length of 14 OFDM symbols) approximately corresponds to the coherence time experienced at highway speed, when using a 2 GHz carrier.

Since the terminal works with estimated CSI, part of the coherence interval will be spent on pilots, and channel estimation for codes with more VAs will consume a larger part of the coherence interval. Here we assume that $\tau_p = M$, i.e. as few symbols as possible are spent on the pilot sequence. However, the normalized transmit power of the BS is the same. This means that a large code will spend more energy in total on pilots, so that the channel estimate is of similar quality, no matter what code we pick. The total normalized power spent on the τ_c symbols is $\rho\tau_c$. We assume that the normalized powers in the training phase and data phase are equal to ρ.

A. The Cost of More Virtual Antennas

When is more VAs beneficial, and at what point does the cost of estimating more channel coefficients outweigh the gain from the extra diversity? To shed some light on this, we compare two codes with the same rate, but different number of virtual antennas, i.e., 343 and 443 from Table I.

Figure 1 shows the ergodic capacity and the outage rate for $\epsilon = 10^{-5}$ for codes 343 and 443, using estimated CSI. As can be seen, the number of VAs has little impact on the ergodic capacity ($L \to \infty$). We see that even with more training (4 symbols compared to 3), the code with more spatial diversity has a little advantage. However, if equal amount of energy would be spent on pilots, the code with fewer VAs would be better. The reason is that when the same amount of energy is spent on pilots, there is less energy spent on each estimated channel coefficient for the code with more spatial diversity.

Looking at the other extreme, when no time/frequency diversity is available ($L = 1$), the curves shows that the benefit of having extra diversity from the additional VA is greater than the cost of having to estimate one more channel coefficient. This means that square codes ($M = \tau$) are preferred, when there is very little time/frequency diversity.
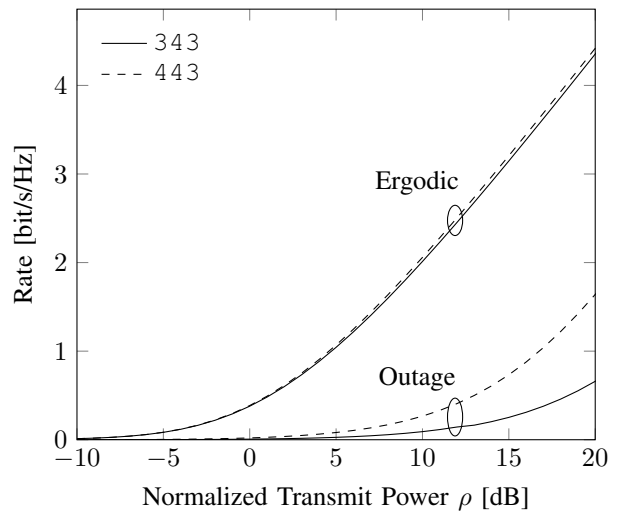
[4]Investigating larger codes is part of future work.

Fig. 1. Comparison of the ergodic and outage ($\epsilon = 10^{-5}$) rate of a square OSTBC and a rectangular OSTBC with the same pre-log factor. The two codes are almost equal when looking at ergodic rate, but the code with more spatial diversity performs better in a low diversity scenario.

B. Low Diversity Scenario

In a low diversity scenario ($L = 1$) we concluded in Section V-A that square codes are preferred to rectangular codes. Hence, here we compare the outage rate of the three square codes from Table I. To see how each of these codes perform, we use the SNR expression in (12). For the ergodic rates, the pre-log factor is the most important thing, which basically means that the ala code ($M = 2$) is the best choice, since it is the only code in Table I with rate 1.

Looking at Figure 2, more antennas are beneficial at low normalized transmit power. The increase in diversity more than compensates for the loss in rate given by the pre-log factor. However, at high normalized transmit power, the pre-log factor n_s/τ determines the slope of the rate curve. This means that the codes with the largest pre-log factor will dominate the others. In other words the ala code is optimal in this sense, for high normalized transmit powers.

As shown in Figure 2, the best choice of the number of VAs varies, depending on the rate of communication. For example, if 1 bit/s/Hz is sufficient, the 884 code is the best, since we can communicate with an error probability less than 10^{-5} using less power than the other codes. As the rate demand increases, so will the preference of rate over diversity when choosing which code to use. The 443 code is the best if the rate demand is around 7 bit/s/Hz, while the rate demand has to exceed 12 bit/s/Hz for the ala code to be the preferred choice.

A looser reliability constraint favors codes with higher rates. If we would loosen the reliability constraint, to $\epsilon = 10^{-3}$, for example (not shown due to lack of space), the curves would shift to the left. Other than the higher rates, we would see that the intersections of the curves are moved even further to the left. For example, if communication with 3 bit/s/Hz is required, then the 884 code is the best choice for the case of error probability less than 10^{-5}, while the 443 code is better if the error probability is less than 10^{-3}.
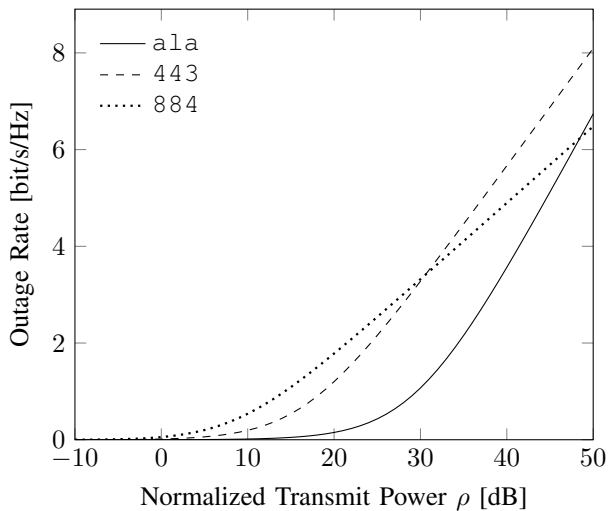
Fig. 2. Outage rates for the three square codes in Table I with error probability 10^{-5}. $L = 1$ diversity branch and packet length 168 symbols, giving a coherence interval of $\tau_c = 168/1 = 168$ symbols.
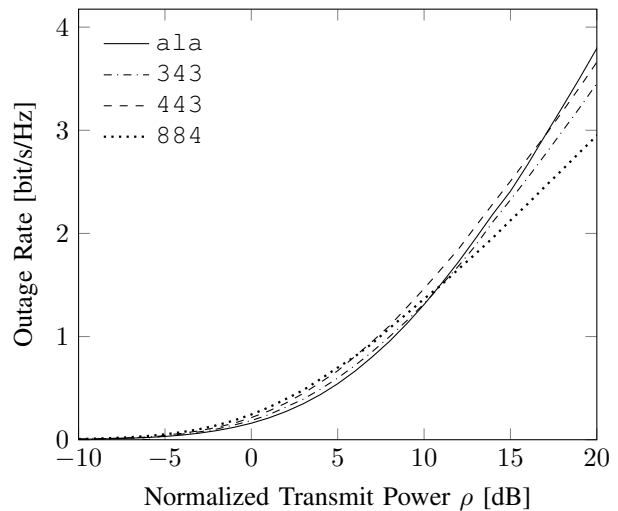

Fig. 3. Outage rates for the four codes in Table I with error probability 10^{-5}. $L = 12$ diversity branches and packet length 168 symbols, giving a coherence interval of $\tau_c = 168/12 = 14$ symbols.

C. Intermediate Diversity Scenario

We now consider the case when we have a stringent delay constraint, but are allowed to spread out the subcarriers over a larger bandwidth. If the subcarriers are far enough apart, it is reasonable to assume that each one sees a different channel realization. Hence, here we have $L = 12$ diversity branches. Here we have more frequency diversity than in the previous case, but still not as much to be able to say that the ergodic rate is a good metric.

Since we now average over 12 realizations, the outage capacity will improve, but at the cost of training. Since all 12 channels need to be estimated by the terminal, the BS needs to send pilots on each of the diversity branches. Each branch is now only 14 symbols long, which means that training takes up a significant part of the coherence interval.

Comparing Figure 3 with Figure 2 we see that, because of the increased cost of training, codes with more diversity are not as dominant, even for very low rates. Adding more diversity branches quickly reaches the point of diminishing returns for this scenario. The rates are also significantly lower, as a large part of the coherence interval is used for pilots.

VI. Conclusion

When broadcasting common information in a massive MIMO system, the information needs to be conveyed from the BS to terminals for which no CSI is available at the BS. We derived an achievable rate for an arbitrary OSTBC and analyzed the performance of different codes when the packet length is short and reliability is important. As seen in the numerical analysis, the optimal choice of the number of virtual antennas (or dimension of the code) varies with reliability and rate requirements. The numerical results indicate that, for moderate rates in a low diversity scenario, the high diversity code is preferred. The Alamouti code performs well if there is enough time and/or frequency diversity, but falls behind codes

with more spatial diversity when time/frequency diversity is limited. In a low diversity scenario, where the message only sees one channel realization additional diversity offered by the code seems necessary, especially when the reliability constraint is tightened.

References

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[2] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 186–197, January 2010.

[3] M. Karlsson and E. Larsson, "On the operation of massive MIMO with and without transmitter csi," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014, pp. 1–5.

[4] X. Meng, X.-G. Xia, and X. Gao, "Constant-envelope omni-directional transmission with diversity in massive MIMO systems," in *2014 IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 3784–3789.

[5] M. Biguesh and A. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP J. Appl. Signal Process.*, no. 9, pp. 1330–1339, 2004.

[6] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, 2010.

[7] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. New York, NY, USA: Cambridge University Press, 2003.

[8] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.

[9] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct 1998.

[10] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.

[11] J. Östman, W. Yang, G. Durisi, and T. Koch, "Diversity versus multiplexing at finite blocklength," in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, Aug 2014, pp. 702–706.

[12] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*. Kidlington, Oxford, UK: Academic Press, 2013.