
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Karpuk, David; Hollanti, Camilla; Barreal, Amaro

Node repair for distributed storage systems over fading channels

Published in:

Proceedings of 2014 International Symposium on Information Theory and Its Applications, ISITA 2014

Published: 08/12/2014

Document Version

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

Karpuk, D., Hollanti, C., & Barreal, A. (2014). Node repair for distributed storage systems over fading channels. In *Proceedings of 2014 International Symposium on Information Theory and Its Applications, ISITA 2014* (pp. 383-387). Article 6979870 IEEE.

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Node Repair for Distributed Storage Systems over Fading Channels

David Karpuk, Camilla Hollanti, Amaro Barreal

Dept. of Mathematics and Systems Analysis

Aalto University

P.O. Box 11100

FI-00076 Aalto, Finland

email: {david.karpuk, camilla.hollanti, amaro.barreal}@aalto.fi

Abstract—Distributed storage systems and associated storage codes can efficiently store a large amount of data while ensuring that data is retrievable in case of node failure. The study of such systems, particularly the design of storage codes over finite fields, assumes that the physical channel through which the nodes communicate is error-free. This is not always the case, for example, in a wireless storage system.

We study the probability that a subpacket is repaired incorrectly during node repair in a distributed storage system, in which the nodes communicate over an AWGN or Rayleigh fading channels. The asymptotic probability (as SNR increases) that a node is repaired incorrectly is shown to be completely determined by the repair locality of the DSS and the symbol error rate of the wireless channel. Lastly, we propose some design criteria for physical layer coding in this scenario, and use it to compute optimally rotated QAM constellations for use in wireless distributed storage systems.

Index Terms—distributed storage systems, repair locality, locally repairable codes, Rayleigh fading channels, rotation codes

I. INTRODUCTION

In a distributed storage system (DSS), data is distributed among several storage nodes in a network, in such a way that the data is made robust against node failures by introducing redundancy. There are many invariants of a DSS which measure its effectiveness. Among the most important are repair bandwidth, storage, and *repair locality*, which is the number of nodes a newcomer node needs to contact to repair the system. The importance of repair locality was discovered independently in [1], [2], [3] and linear codes over finite fields, known as *locally repairable codes* in [4] and *self-homomorphic codes* in [2] were designed with these criteria in mind. Repair locality is particularly important for cloud storage applications, in which the disk I/O (which is proportional to the repair locality) appears to be the main judge of performance [4], [1].

Analyses and proposed codes for DSSs have almost exclusively concerned the logical layer, that is, the bits storing the data and their encoding and decoding. Almost no attention has been paid to the physical layer over which the nodes communicate, and communication over this layer is usually assumed error-free. However, the physical reality is that this channel can be noisy, for example in wireless storage systems [5] or wireless sensor networks [6], [7]. In [5], the author

describes an efficient way to partially download the contents of a wireless storage network, and in [6] and [7], the authors apply network coding principles to data storage in wireless sensor networks. Regenerating codes have been developed to correct errors in DSSs, but operate exclusively on the logical layer, as in [8]. More closely related with the present work is [9], in which the authors present protocols for physical layer storage transmission encoding.

In this article, we present an analysis of the probability, P_{sub} , that a node subpacket is reconstructed incorrectly during node repair, for DSSs where communication between nodes takes place over potentially noisy channels. While this is explained in more detail in Section III, we should mention now that by “subpacket” we mean the smallest unit into which the DSS divides the contents of a node. In summary, the main contributions of this article are

- Theorem 1, which describes the asymptotic behavior of P_{sub} as $\text{SNR} \rightarrow \infty$ in terms of the repair locality of the DSS and the symbol error rate of the wireless channel,
- simulation results for AWGN and Rayleigh fading channels demonstrating how P_{sub} varies with repair locality and the underlying finite field, and the accuracy of our approximation to P_{sub} , and
- design criteria presented in Section VI for coding over Rayleigh fading channels in a wireless storage system, which are then used to find optimal rotations of QAM constellations which minimize P_{sub} for a given repair locality, finite field size, and SNR.

In an asymptotic sense, our design criteria match those of traditional physical layer coding over Rayleigh fading channels (as in [10]), and thus it is likely that invariants such as the minimum product distance can also be useful in designing codes for the storage scenario.

II. BASIC SETUP

To establish the types of DSSs we consider, let us briefly describe a standard setup, considered for example in [11]. Many examples of distributed storage codes fit the following description, including the locally repairable codes of [4] and [1], and the self-repairing homomorphic codes of [2]. Consider a distributed storage system housing a file which is divided into k pieces, distributed amongst n nodes using an (n, k) -MDS

(Maximum Distance Separable) code over \mathbf{F}_q . This encoded data at a particular node is often subdivided into subpackets. The MDS property guarantees that a data collector need only contact any k of the n nodes to collect the data. We denote by ω_i a subpacket of the encoded data stored in the i th node, and we assume $\omega_i \in \mathbf{F}_q$ for some $q = 2^m$.

Suppose now that one of the n nodes containing the subpacket ω fails, and that an incoming node has to contact r helper nodes to repair the system. The data ω can be expressed as an \mathbf{F}_q -linear combination of certain subpackets of the helper nodes. That is, $\omega = \sum_{i=1}^r \alpha_i \omega_i$ where the $\alpha_i \in \mathbf{F}_q$ are determined by the particular storage code and may depend on the particular subpacket and node which need repair.

The $\omega_1, \dots, \omega_r$ are to be transmitted to the newcomer via some noisy channel, for example through an AWGN or Rayleigh fading channel. To do this, we incorporate a bijective lift function $L : \mathbf{F}_q \rightarrow \mathcal{C}$ which modulates the contents of the helper node into information symbols contained in a finite constellation \mathcal{C} of size q , intended for transmission over the noisy channel. We define $x_i := L(\omega_i)$ for all i .

While many types of modulation are possible, our lift function L first represents $\omega \in \mathbf{F}_q$ as a bit string of length $m = \log_2(q)$ by choosing a basis for \mathbf{F}_q as a vector space over \mathbf{F}_2 . Then L assigns to ω an element of \mathcal{C} , a q -QAM constellation, via the Gray labeling. We will assume that all subpackets ω are uniformly distributed on \mathbf{F}_q , which guarantees that the error rates we are interested in studying are independent of the choice of basis.

The r helper nodes transmit some information symbols $x_1, \dots, x_r \in \mathcal{C}$ to the newcomer node through a noisy channel, which the newcomer decodes to get ML-estimates $\hat{x}_1, \dots, \hat{x}_r \in \mathcal{C}$. It then computes

$$\hat{\omega}_i := L^{-1}(\hat{x}_i) \quad \text{for } i = 1, \dots, r, \text{ and } \hat{\omega} := \sum_{i=1}^r \alpha_i \hat{\omega}_i. \quad (1)$$

The newcomer repairs the subpacket successfully if $\hat{\omega} = \omega$. We assume that time or frequency division is employed, which allows the newcomer to receive each subpacket at a different time or frequency instant.

For a fixed channel model and modulation scheme, we define P_s to be the corresponding symbol error rate. That is, if x is some information symbol transmitted over a wireless channel and \hat{x} is some ML-estimate computed by the decoder, then $P_s = P(\hat{x} \neq x)$. The quantity P_s of course depends on the modulation scheme and the channel. However, we will consider only AWGN and Rayleigh fading channels in our examples, and only use q -QAM modulation, though many other modulations schemes are of course possible.

III. LOCALLY REPAIRABLE CODES

Our motivating example of a storage code is the locally repairable codes of [4]. To briefly summarize their construction, a file \mathbf{w} is represented as $\mathbf{w} = [\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(r+1)}]$, where each $\mathbf{w}^{(i)} = (\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_n^{(i)}) \in \mathbf{F}_q^n$ with $(r+1)n$. We then define $\mathbf{s} = \sum_i \mathbf{w}^{(i)}$. For our analysis it suffices to consider

$r+1 = n$, that is, using the vocabulary of [4], we only consider one ‘‘repair group’’. When $r+1 = n$, the $\mathbf{w}^{(i)}$ and \mathbf{s} are then distributed among the n nodes in the following way:

| node 1 | node 2 | \dots | node r | node $r+1$ |
|----------------------|--------------------------|----------|--------------------------|--------------------------|
| $\mathbf{w}_1^{(1)}$ | $\mathbf{w}_2^{(1)}$ | \dots | $\mathbf{w}_r^{(1)}$ | $\mathbf{w}_{r+1}^{(1)}$ |
| $\mathbf{w}_1^{(2)}$ | $\mathbf{w}_3^{(2)}$ | \dots | $\mathbf{w}_{r+1}^{(2)}$ | $\mathbf{w}_1^{(2)}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| $\mathbf{w}_r^{(r)}$ | $\mathbf{w}_{r+1}^{(r)}$ | \dots | $\mathbf{w}_{r-2}^{(r)}$ | $\mathbf{w}_{r-1}^{(r)}$ |
| \mathbf{s}_{r+1} | \mathbf{s}_1 | \dots | \mathbf{s}_{r-1} | \mathbf{s}_r |

(2)

If node i is lost, a newcomer node can recover subpacket $\mathbf{w}_j^{(i)} \in \mathbf{F}_q$ by XORing together subpackets from all of the r remaining nodes. In our notation, a subpacket ω_i can refer to any of the $\mathbf{w}_j^{(i)}$, or to the coordinate of \mathbf{s} stored in the i th node; the analysis is independent of the particular subpacket.

The construction of [4] uses an (n, k) -MDS code as an outer code, so that the $\mathbf{w}^{(i)}$ are actually themselves encoded data. Since we are interested only in whether or not a node is repaired correctly, we will ignore the outer coding, while remarking that the MDS code allows one to potentially reconstruct the file correctly despite faulty reconstruction of individual subpackets.

IV. SUBPACKET ERROR PROBABILITY

Our main benchmark for judging performance will be the probability $P_{sub} = P(\hat{\omega} \neq \omega)$ that a subpacket is reconstructed incorrectly during node repair. Let us write $\hat{\omega}_i = \omega_i + \tilde{\omega}_i$ for all $i = 1, \dots, r$, so that $\tilde{\omega}_i$ is the ‘‘error’’ introduced by the transmission of the data of helper node i over the corresponding noisy channel. Then, clearly,

$$P_{sub} = P(\hat{\omega} \neq \omega) = P\left(\sum_{i=1}^r \alpha_i \tilde{\omega}_i \neq 0\right) \quad (3)$$

or in other words, the subpacket is repaired correctly when the linear combination of the errors is zero. For the locally repairable codes of [4], we have $\alpha_i = 1$ for all i . The following theorem completely describes the asymptotic behavior of P_{sub} as $\text{SNR} \rightarrow \infty$, in terms of repair locality and the symbol error rate of the wireless channel.

Theorem 1: Consider a distributed storage system with repair locality r , where a subpacket ω is reconstructed via an \mathbf{F}_q -linear combination of subpackets from the helper nodes: $\omega = \sum_{i=1}^r \alpha_i \omega_i$, for $\alpha_i \in \mathbf{F}_q$. Then as $\text{SNR} \rightarrow \infty$ we have

$$P_{sub} \sim r P_s (1 - P_s)^{r-1} \sim r P_s \quad (4)$$

where P_s is the symbol error rate of the channel used by the nodes.

Proof: Suppose that during node repair we have some collection i_1, \dots, i_j of j out of the r helper nodes which transmit their data incorrectly, that is, that $\hat{\omega}_{i_1} \neq \omega_{i_1}, \dots, \hat{\omega}_{i_j} \neq \omega_{i_j}$. Let

$$c(i_1, \dots, i_j) = \begin{cases} 1 & \text{if } \alpha_{i_1} \tilde{\omega}_{i_1} + \dots + \alpha_{i_j} \tilde{\omega}_{i_j} \neq 0 \\ 0 & \text{if } \alpha_{i_1} \tilde{\omega}_{i_1} + \dots + \alpha_{i_j} \tilde{\omega}_{i_j} = 0 \end{cases} \quad (5)$$

so that the quantity $c(i_1, \dots, i_j)$ tells us whether or not the error terms cancel each other out. Then for this fixed j -tuple, the quantity $P(c(i_1, \dots, i_j) = 1)$ measures the probability that the error terms will not cancel, over all possible collections of error terms $\tilde{\omega}_{i_1}, \dots, \tilde{\omega}_{i_j}$.

Summing over all possible j -tuples of helper nodes and then over all j , we count all of the ways that the subpacket can be reconstructed incorrectly to arrive at

$$P_{sub} = \sum_{j=1}^r \sum_{i_1 \neq \dots \neq i_j} P_s^j (1 - P_s)^{r-j} P(c(i_1, \dots, i_j) = 1) \quad (6)$$

We claim that

$$rP_s(1 - P_s)^{r-1} \leq P_{sub} \leq \sum_{k=1}^r \binom{r}{k} P_s^k (1 - P_s)^{r-k} \quad (7)$$

To arrive at the upper bound, plug the trivial inequality $P(c(i_1, \dots, i_j) = 1) \leq 1$ into the above equation, and note that there are $\binom{r}{k}$ possible k -tuples of helper nodes.

Now suppose that $j = 1$, that is, that we are only interested in situations where the data from exactly one node, say node i_1 , is transmitted incorrectly. Then $\tilde{\omega}_{i_1} \neq 0$, and $\alpha_{i_1} \neq 0$ by assumption. Therefore $c(i_1) \neq 0$, that is, there can be no cancellation of error terms when there is only one error. Combining this fact with the trivial lower bound $0 \leq P(c(i_1, \dots, i_j) = 1)$ for $j \geq 2$, one arrives at the lower bound.

We now claim that the bounds in (7) are asymptotically tight. That is, as $\text{SNR} \rightarrow \infty$, we claim that

$$P_{sub} \sim rP_s(1 - P_s)^{r-1} \text{ and } P_{sub} \sim \sum_{j=1}^r \binom{r}{j} P_s^j (1 - P_s)^{r-j}$$

To see this, note that by (7) it suffices to show that

$$\frac{\sum_{j=1}^r \binom{r}{j} P_s^j (1 - P_s)^{r-j}}{rP_s(1 - P_s)^{r-1}} \rightarrow 1 \quad (8)$$

as $\text{SNR} \rightarrow \infty$. To do this we treat the sum term-by-term. For a fixed j , we have

$$\frac{\binom{r}{j} P_s^j (1 - P_s)^{r-j}}{rP_s(1 - P_s)^{r-1}} = \frac{1}{r} \binom{r}{j} P_s^{j-1} (1 - P_s)^{1-j} \quad (9)$$

which equals 1 if $j = 1$, and for $j \geq 2$ clearly approaches 0 as $P_s \rightarrow 0$, i.e. as $\text{SNR} \rightarrow \infty$.

The claim that $rP_s(1 - P_s)^{r-1} \sim rP_s$ as $\text{SNR} \rightarrow \infty$ follows easily by expanding out $(1 - P_s)^{r-1}$ and noting that the dominant term is rP_s . ■

V. REPAIR OVER FADING CHANNELS

Node repair over a Rayleigh fading channel involves two equations for each helper node $i = 1, \dots, r$ (see [10]):

$$y_{i_j} = h_{i_j} x_{i_j} + z_{i_j}, \quad \text{for } j = 1, 2 \quad (10)$$

where $x_i = x_{i_1} + \sqrt{-1}x_{i_2}$ is an information symbol from a q -QAM constellation \mathcal{C} sent by the i th helper node, $h_{i_j}, j = 1, 2$ are i.i.d. Rayleigh random variables satisfying $\mathbf{E}(h_{i_j})^2 = 1$, $z_{i_j}, j = 1, 2$ are i.i.d. zero-mean Gaussian random variables with variance $N_0/2$ representing the noise between helper

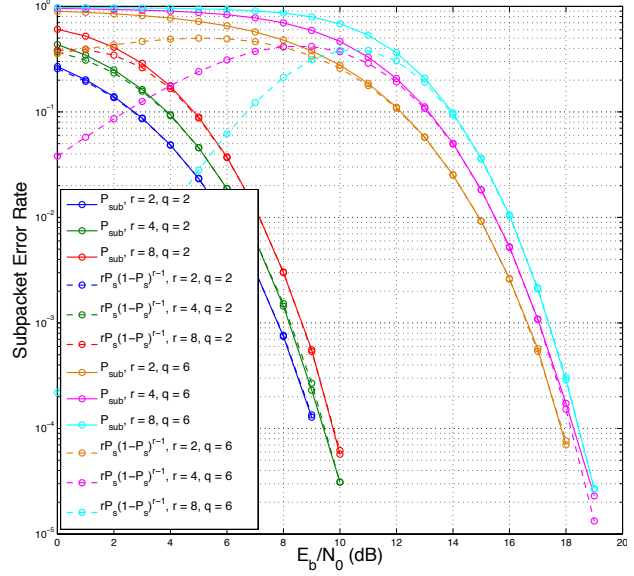


Fig. 1. Plots of P_{sub} and $rP_s(1 - P_s)^{r-1}$, for the locally repairable codes of [4] and AWGN channel, using the finite fields \mathbf{F}_4 and \mathbf{F}_{64} , i.e. 4-QAM and 64-QAM.

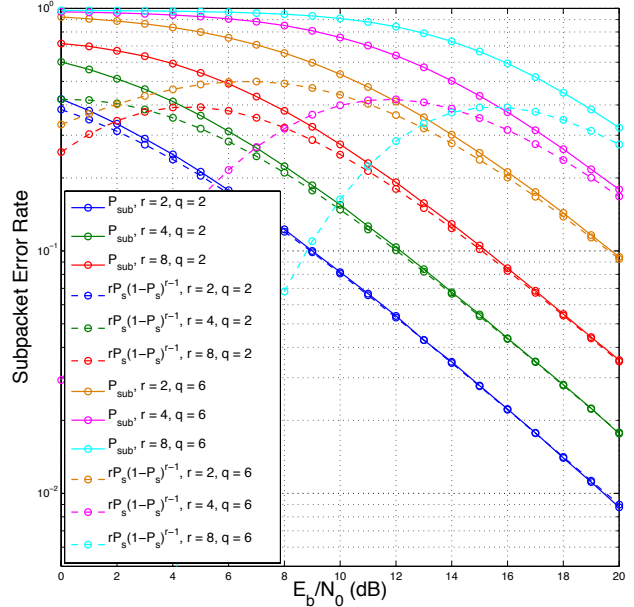


Fig. 2. Plots of P_{sub} and $rP_s(1 - P_s)^{r-1}$, for the locally repairable codes of [4] and Rayleigh fading channel, using the finite fields \mathbf{F}_4 and \mathbf{F}_{64} , i.e. 4-QAM and 64-QAM.

node i and the newcomer, and $y_{i_j}, j = 1, 2$ are received by the incoming node from helper node i . If we instead set $h_{i_j} = 1$ for all i, j , we are in the case of an AWGN channel.

The incoming node decodes by computing an ML-estimate \hat{x}_i for each $i = 1, \dots, r$, and reconstructs $\hat{\omega}$ from the \hat{x}_i as described in the previous sections. One defines the *average energy per bit* E_b of the constellation \mathcal{C} by

$$E_b = \frac{1}{\log_2(q)} \sum_{x \in \mathcal{C}} \frac{\|x\|^2}{q} \quad (11)$$

and often plots error rates as a function of the ratio E_b/N_0 .

To empirically test the accuracy of the approximation

$$P_{\text{sub}} \approx rP_s(1 - P_s)^{r-1} \quad (12)$$

simulations were carried out for the AWGN and Rayleigh fading channels. We generated r bit strings $\omega_1, \dots, \omega_r$ each of length $\log_2(q)$ uniformly at random and computed their bitwise XOR $\omega = \sum_{i=1}^r \omega_i$. We then computed $L(\omega_i) = x_i$ using the Gray labeling, passed x_i through either an AWGN or Rayleigh fading channel, decoded to get \hat{x}_i , and computed $\hat{\omega}$ from these. We counted the number of subpacket errors, that is, the number of times $\hat{\omega} \neq \omega$ out of 10^6 trials at each level of E_b/N_0 .

In Fig. 1 and Fig. 2 we plot the results of our simulations for AWGN and Rayleigh fading channels. One can see from the plots how the subpacket error rate increases with both the size of the underlying finite field and the repair locality. For the AWGN channel, one can see that our approximation $P_{\text{sub}} \approx rP_s(1 - P_s)^{r-1}$ is quite sharp once $P_s < 10^{-1}$. For the Rayleigh fading channel one observes similar behavior for 4-QAM modulation and all repair localities, and it is reasonable to extrapolate from the current data that for 64-QAM modulation the approximation is quite accurate once $P_s < 10^{-1}$.

We omit plots describing the accuracy of the approximation $P_{\text{sub}} \approx rP_s$, since the ratio $rP_s(1 - P_s)^{r-1}/(rP_s) = (1 - P_s)^{r-1}$ as $\text{SNR} \rightarrow \infty$ is easily calculable, and therefore the behavior of the approximation $rP_s(1 - P_s)^{r-1} \approx rP_s$ is trivial to understand.

VI. ROTATION CODES FOR WIRELESS STORAGE SYSTEMS

Given a distributed storage system with repair locality r , we can use the estimates provided by Theorem 1 to minimize P_{sub} . Specifically, suppose we are using q -QAM modulation. One can rotate (counter-clockwise) a q -QAM constellation by θ radians to improve P_s , and a natural question is whether or not rotations can similarly improve P_{sub} .

Let \mathcal{C} denote a q -QAM constellation, and suppose that we denote by \mathcal{C}_θ the constellation \mathcal{C} rotated by θ radians. Let $P_s(\theta)$ and $P_{\text{sub}}(\theta)$ denote the corresponding symbol and subpacket error rates for the constellation \mathcal{C}_θ , for a fixed finite value of SNR. Choosing an optimal rotation demands that we solve

$$\arg \min_{\theta} P_{\text{sub}}(\theta), \quad \theta \in [0, \pi/2] \quad (13)$$

and Theorem 1 suggests that we minimize either of:

$$g_1(\theta) = rP_s(\theta)(1 - P_s(\theta))^{r-1} \quad (14)$$

$$g_2(\theta) = rP_s(\theta) \quad (15)$$

Minimizing $g_2(\theta)$ is of course equivalent to minimizing $P_s(\theta)$, and there is a plethora of literature concerning optimal rotations of QAM constellations. It is not clear for finite SNR that the same θ minimizes $g_1(\theta)$, and since $rP_s(1 - P_s)^{r-1}$ is a more accurate estimate of P_{sub} , it is not obvious that the same rotations which are optimal for traditional QAM modulation should be optimal for the storage scenario.

Let \mathcal{C} denote a q -QAM constellation. It is well-known (see [12]) that the pairwise error probability for the above channel can be upper-bounded by

$$P(x \rightarrow y) \leq \frac{1}{2} \sum_{x \neq y \in \mathcal{C}} s(x, y)$$

where

$$s(x, y) := \prod_{i=1}^2 \frac{1}{1 + \frac{E_s}{4N_0} |x_i - y_i|^2}.$$

In fact, if we assume that \mathcal{C} has been rotated so that it is full-diversity, there is a lower bound on the pairwise error probability (see Theorem 1 of [13]), given by

$$P(x \rightarrow y) \geq P_L(x, y)$$

where

$$P_L(x, y) := \frac{1}{4} \left(\sum_{i=1}^2 \frac{1}{(1 + \delta)^i} \right) s(x, y)$$

and

$$\delta := \max_{i=1,2} \sqrt{\frac{\frac{E_s}{4N_0} |x_i - y_i|^2}{1 + \frac{E_s}{4N_0} |x_i - y_i|^2}}.$$

It follows that $P_s \geq \min_{x,y \in \mathcal{C}} P_L(x, y)$ and hence that

$$\begin{aligned} rP_s(1 - P_s)^{r-1} &\leq \frac{r}{2} \left(\sum_{x \neq y \in \mathcal{C}} s(x, y) \right) \\ &\times \left(1 - \min_{x \neq y \in \mathcal{C}} P_L(x, y) \right)^{r-1}. \end{aligned}$$

We can now succinctly state our two proposed design criteria as follows. For a fixed r , q , and N_0 , we propose selecting the rotation $\theta \in [0, \pi/2]$ which minimizes

$$f_1(\theta) = \left(\sum_{x \neq y \in \mathcal{C}_\theta} s(x, y) \right) \left(1 - \min_{x \neq y \in \mathcal{C}_\theta} P_L(x, y) \right)^{r-1} \quad (16)$$

or the simpler

$$f_2(\theta) = \sum_{x \neq y \in \mathcal{C}_\theta} s(x, y) \quad (17)$$

Of course, Theorem 1 concerns asymptotic estimates of P_{sub} , so even though the proposed design criteria are for finite levels of SNR, we should expect rotations obtained by minimizing f_1 and f_2 to perform best for large SNR.

To test the effectiveness of our proposed rotations, we plot in Fig. 3 and Fig. 4 the subpacket error rate as a function of E_b/N_0 , for unrotated constellations, and for constellations rotated by the values of θ which minimize f_1 and f_2 , for the given level of SNR. Minimization of f_1 and f_2 was done by brute force, which is feasible since only one parameter determines a rotation in two-dimensional space.

It is clear from the plots that rotated constellations improve the subpacket error rate of a wireless storage system, the same

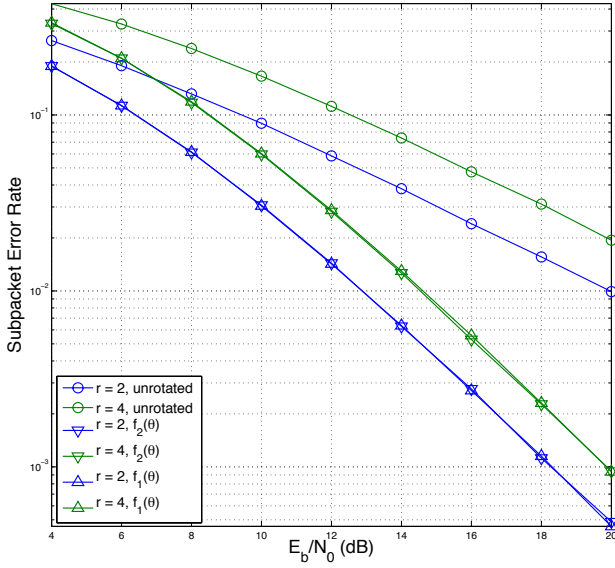


Fig. 3. Probability of subpacket reconstruction error for the locally repairable codes of [4] and Rayleigh fading channel, using rotation codes for 4-QAM constellations (i.e. with storage code over \mathbf{F}_{16}) optimized according to $f_1(\theta)$ and $f_2(\theta)$

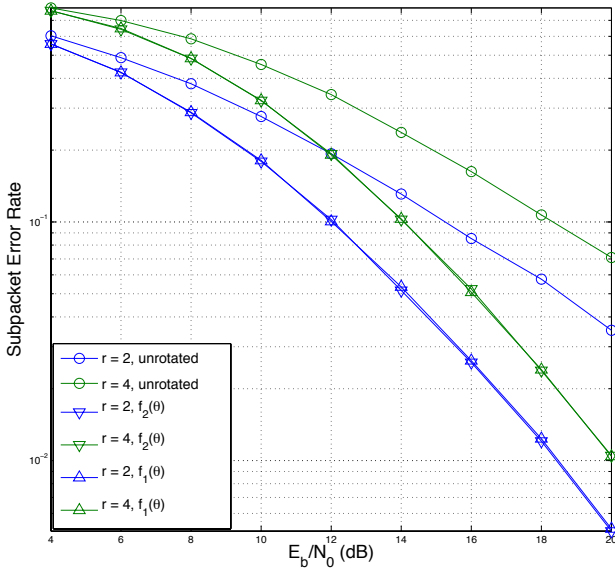


Fig. 4. Probability of subpacket reconstruction error for the locally repairable codes of [4] and Rayleigh fading channel, using rotation codes for 16-QAM constellations (i.e. with storage code over \mathbf{F}_{16}) optimized according to $f_1(\theta)$ and $f_2(\theta)$

way that they reduce P_s for traditional q -QAM modulation. Furthermore, it is also clear that the rotations produced by minimizing either objective function, namely either f_1 or f_2 , have almost exactly the same performance. In fact, nearly all of the rotations obtained by minimizing f_1 and f_2 were within 10^{-2} radians of the rotation for 4-QAM which optimizes the minimum product distance. Nevertheless, we feel it is important to establish unique design criteria for physical layer coding for wireless storage systems, because it is not obvious

that the minimum product distance will continue to be good design criteria for higher dimensional lattices or wireless MIMO storage systems.

VII. CONCLUSIONS AND FUTURE WORK

In this article we described the probability that subpackets are repaired incorrectly in distributed storage systems, in which nodes communicate over AWGN and Rayleigh fading channels, in terms of the repair locality of the distributed storage system and the symbol error rate of the channel. We used this description to establish design criteria for physical layer coding for wireless storage systems, and provided simulation results to show that our design criteria results in more accurate node repair. Further topics which deserve attention include data collection over noisy channels and generalizing our results to MIMO wireless storage systems.

VIII. ACKNOWLEDGEMENTS

The first author is supported by a grant from the Magnus Ehrnrooth Foundation and Academy of Finland grant #268364.

REFERENCES

- [1] D.S. Papailiopoulos, J. Luo, A.G. Dimakis, and J. Li C. Huang, "Simple regenerating codes: Network coding for cloud storage", in *IEEE INFOCOM Proceedings*, 2012.
- [2] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems", in *IEEE INFOCOM Proceedings*, 2011.
- [3] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols", *IEEE Transactions on Information Theory*, vol. 58, no. 11, November 2012.
- [4] D.S. Papailiopoulos and A.G. Dimakis, "Locally repairable codes", in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.
- [5] C. Gong, "On partial downloading for wireless distributed storage", *IEEE Trans. on Signal Processing*, vol. 60, pp. 3278–3288, June 2012.
- [6] W. Lei, Y. Yuwang, Z. Wei, and L. Wei, "Network coding for energy-efficient distributed storage system in wireless sensor networks", April 2013, preprint available <http://arxiv.org/abs/1304.1705>.
- [7] N. Wang and J. Lin, "Network coding for distributed data storage and continuous collection in wireless sensor networks", in *4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, 2008.
- [8] K. Ramchandran P.V. Kumar K.V. Rashmi, N. Shah, "Regenerating codes for errors and erasures in distributed storage", in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.
- [9] C. Hollanti, D. Karpuk, A. Barreal, and H.-F. (Francis) Lu, "Space-time storage codes for wireless distributed storage systems", 2014, Global Wireless Summit, to appear.
- [10] F. Oggier and E. Viterbo, "Algebraic number theory and code design for rayleigh fading channels", *Commun. Inf. Theory*, vol. 1, no. 3, pp. 333–416, 2004.
- [11] F. Oggier and A. Datta, "Coding techniques for repairability in networked distributed storage systems", *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 4, pp. 383–466, 2012.
- [12] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both rayleigh fading and gaussian channels", *IEEE Transactions on Information Theory*, vol. 12, no. 2, March 1996.
- [13] S.B. Slimane and T. Le-Ngoc, "Tight bounds on the error probability of coded modulation schemes in rayleigh fading channels", *IEEE Trans. on Vehicular Technology*, vol. 44, pp. 121–130, February 1995.