Input-output distance properties of good linear codes

Hajir Roozbehani and Yury Polyanskiy
MIT
Email: {hajir,yp}@mit.edu

***Abstract*—Consider a linear code defined as a mapping between vector spaces of dimensions k and n. Let β^* denote the minimal (relative) weight among all images of input vectors of full Hamming weight k. Operationally, β^* characterizes the threshold for adversarial (erasure) noise beyond which decoder is guaranteed to produce estimate of k-input with 100% symbol error rate (SER). This paper studies the relation between β^* and δ, the minimum distance of the code, which gives the threshold for 0% SER. An optimal tradeoff between β^* and δ is obtained (over large alphabets) and all linear codes achieving $\beta^* = 1$ are classified: they are repetition-like. More generally, a design criteria is proposed for codes with favorable graceful degradation properties.**

As an example, it is shown that in an overdetermined system of n homogeneous linear equations in k variables (over a field) it is always possible to satisfy some $k-1$ equations with non-zero assignments to every unknown, provided that any subset of k equations is linearly independent. This statement is true if and only if $n \geq 2k-1$.

I. Introduction

A mapping of k symbols to n symbols is said to have the (α, β)-property if it sends any two strings of (Hamming) distance more than αk to two strings of (Hamming) distance more than βn. This property was first introduced in [1] and is relevant to the Combinatorial Joint Source Channel Coding problem [2]–[4].

Definition 1. *A map $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is said to be (α, β) if*

$$|x - y| > \alpha k \implies |f(x) - f(y)| > \beta n,$$

where $|\cdot|$ denotes the Hamming weight.

For a linear map f, we define[1]

$$\beta(\alpha) := \inf_x \{ \frac{|f(x)|}{n} \mid |x| > \alpha k \} - \frac{1}{n}$$

and

$$\beta^* := \beta(1 - \frac{1}{k}). \tag{1}$$

Recall that the (relative) minimum distance of f is $\delta := \inf_{x \neq y} \frac{|f(x)-f(y)|}{n}$. Note that $\delta = \beta(0) + \frac{1}{n}$.

The (α, β)-property is relevant to the problems of graceful degradation and partial data recovery. One often encodes a message x by a map f to build tolerance against external noise. For instance, one may map x to $f(x)$ and save the outcome on a storage device. Then noise may act by erasing some of stored bits in an adversarial manner. One then observes the non-erased bits and provides an estimate $\hat{x}$ for x. With the above notation, a map can fully recover the input from $\beta(0)n$ erasures. As the number of erasures exceeds $\beta(0)n$, it is desired that x be recovered with good fidelity, that is, we want $|x - \hat{x}|$ to be as small as possible. In general, $\beta(\alpha)n$ erasures on the output can cause at most αk distortions in the input. Indeed if we let $\hat{x}$ be an arbitrary point in the pre-image, it is guaranteed that $|x - \hat{x}| \leq \alpha k$. In some cases, this arbitrary point is the best estimate available for the input. For instance, if q is large and f is linear, then one cannot find an estimate with provably lower error[2]. Thus it makes sense to think of $1 - \alpha$ as the quality of estimation in recovering x against adversarial noise with intensity β. In this sense, β^* can be thought of as a measure for the ability of the code to partially recover the input in the presence of strong erasure noise.

A related concept is that of unequal protection (UEP) codes [5]–[8]. A code with minimum distance d is said to have the UEP if, for some fixed i, it can always recover the i-th input bit from more than d erasures. In this sense, the UEP codes are often said to have the graceful degradation property. A map with the (α, β)-property does not necessarily provide this type of biased protection. If d erasures occur there is no guarantee that any specific bit can always be fully recovered. However, more can be said about the joint estimates. For instance, if a code has $\beta(1/k) > \beta(0)$ and exactly d erasures occur, then the symbol error rate (SER) on estimating m bits from d erasures can be shown to be at most $\frac{1}{m}$. In other words, the (α, β)-property does not provide unequal protection for any specific bit but it can still ensure graceful degradation of overall SER as the noise level exceeds the error correction capabilities of the code depending on how fast β increases with α.

It is a classic problem in coding theory to find maps with large $\beta(0)$. It is thus useful to have estimates on how large $\beta(0)$ can be. The answer to this question is not yet known unless the alphabet size is large, though various upper bounds on $\beta(0)$ exist (c.f. [9]). The recent work has extended this problem to finding estimates on $\beta(\alpha)$ [1], [10], [11]. Again the exact answer is known only when the alphabet size is large. It was shown in [10] that $\beta(\alpha) \leq 1 - \frac{1-\alpha}{\rho}$, where $\rho := \frac{n}{k}$ and

This work was supported in part by the National Science Foundation award under grant agreement CCF-17-17842 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, and a grant from Skoltech–MIT Joint Next Generation Program (NGP).

[1]The term $-\frac{1}{n}$ appears due to the strictness of inequalities in Definition 1.

[2]When $q > n$, the Chebyshev radius of a linear subspace of $\mathbb{F}_q^n$ is equal to its diameter.

equality can be achieved if $q \geq n$. In this paper, we focus on a different problem.

The above discussion motivates the need for a code with large minimum distance and monotonically increasing $\beta(\alpha)$. Such a code can fully recover the input when the number of erasures is less than its minimum distance, and as the number of erasures exceeds its minimum distance, it can offer some partial recovery guarantees. It turns out, however, that there is a tradeoff between full and partial recovery. In the (α, β)-spectrum, we can fix one point, namely, the minimum distance (or equivalently $\beta(0)$), and ask how large $\beta(\alpha)$ can be at some other point? We give some results in this direction for linear codes. In particular, we show in Section II that there is a tradeoff between the minimum distance δ of a linear code and its β^* (see (1) for definition). We characterize the optimal tradeoff between δ and β^* (over large alphabets). We further show that optimal codes have some input-output limitations: they must send some input vectors with large weight to codewords with minimal weight δ. A priori, the (α, β)-property asks for the mapping of dissimilar messages to be also dissimilar and as such is a relaxation of the locality sensitive hashing (LSH) property [1]. Our results show, however, that at least in the case of linear codes there is a stronger connection between the two in the sense that if a code sends dissimilar messages to dissimilar codewords, it must also send some similar messages to similar codewords (see Theorem 3).

The case of stochastic noise is also of interest. In our scenario above, we may let the noise act by randomly erasing some output coordinates. Again, it is desirable to find codes with good estimation quality that degrades gracefully as the noise level increases. These codes are to be contrasted with capacity achieving codes, which offer small error below a certain noise level and large error after that.

It is hard to design a code with a prescribed SER behavior. However, since the (α, β)-property is more accessible at the design stage, one can use it as a guideline to develop codes with good SER properties. In section III, we provide some techniques to design small binary codes with good (α, β)-profiles using a generalization of MacWilliams identity. We observe that such codes have lower bit error rate (BER) than the Hadamard code of the same length and dimension when used in transmission over BSC for a wide range of channel parameters.

II. Main results

In this section we give a converse bound on β^* (see (1) for definition) as a function of δ for linear codes. Our bound is alphabet independent, and can be tight (over large alphabets). We first prove a bound for MDS codes and then prove a general result for all linear codes. We prove some further (α, β)-limitations of the codes that achieve the bound. In particular, we show that if a code with positive distance achieves the bound, then there exists some x with relatively large weight for which $|f(x)| = \delta n$.

Throughout this section, we assume familiarity with the concept of a geometric (α, β)-system associated to a linear map. We refer the reader to [1] for details. We briefly mention here that the (α, β)-property of f is determined by its image as well as a choice of an embedding. If we write $f(x) = xG$, then we can think of the columns of G as elements of the projective space $\mathbf{P}^{k-1}$, which we will call β-points, while the projective images of the k standard basis vectors are called α-points. The non-zero codewords[3] (up to scaling) are in one-to-one correspondence with hyperplanes in $\mathbf{P}^{k-1}$. The weight of a codeword is the number of β-points that do not lie on the hyperplane. Likewise, the weight of the associated message is the number of α-points that do not lie on the hyper-plane. In this language, for example, $1 - \beta^*$ is the largest fraction of β-points through which we can pass a hyperplane avoiding all α-points. We denote the set of α, β-points, respectively, by $\Gamma_\alpha, \Gamma_\beta$. We also define the sets of α-only points $\Gamma_{\alpha\setminus\beta} := \Gamma_\alpha \setminus \Gamma_\beta$, and β-only points $\Gamma_{\beta\setminus\alpha} := \Gamma_\beta \setminus \Gamma_\alpha$.

We remark that the minimum distance of a map is a property of its image and, hence, depends only upon the configuration of its β-points. On the other hand, β^* is a property of both the image and the embedding and as such depends on the arrangements of both α-points and β-points. The bounds in this section are thus to be interpreted as follows: fixing a property (the minimum distance) of the image, bound β^* for all possible embeddings, i.e., any configuration of α-points.

A. MDS codes

Here we show that, when $n \geq 2k - 1$, linear MDS codes have $\beta^* = 1 - \frac{1}{\rho}$, where $\rho := \frac{n}{k}$. We remark that this result can be seen as a generalization of Theorems 8 and 9 in [10].

Theorem 1. *Suppose the image of $f : \mathbb{F}^k \to \mathbb{F}^n$ is a linear MDS code. If $n \geq 2k-1$ then there exists $x \in \mathbb{F}^k$ with $|x| = k$ such that $|f(x)| \leq n - k + 1$. This implies that $\beta^* = 1 - \frac{1}{\rho}$.*

We omit the proof of this theorem but mention that the proof technique is essentially the same as what appears in the proof of Theorem 3.

Remark 1. *Consider solving a system of linear equations $y = xG$ where G is a $k \times n$ matrix with Kruskal rank k (i.e., any k columns of G span a k-dimensional space). It is possible to find x with $|x| = k$ that satisfies some $k-1$ of the constraints. In other words, there exists a full-weight vector in the left null space of some $k \times (k-1)$ sub-matrix of G.*

For MDS codes of length $n \leq 2k - 2$ we have:

Theorem 2. *Suppose the image of $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is a linear MDS code with $q > k$. If $k + 1 \leq n < 2k - 1$, there exists $x \in \mathbb{F}_q^k$ with $|x| \geq k - sk$ for all $0 \leq sk \leq 2k - n - 1$ such that $|f(x)| \leq k - sk$. In other words, $\beta(\alpha) \leq \frac{1-s}{\rho}$ for all $\alpha < 1 - s$.*

Again, the proof relies on the same technique as that used in the proof of Theorem 3 and is omitted due to space constraints.

[3] By a non-zero codeword we mean the image of a non-zero message.

Remark 2. *The bound $\beta(\alpha) \leq \frac{1-s}{\rho}$ is tight and is achieved if $\Gamma_\alpha \subset \Gamma_\beta$, i.e., if the code is systematic. In fact, this result, combined with Theorem 1, can be used to characterize which $k \times (k-1)$ sub-matrices of G have full-weight elements in their left null space (as discussed in Remark 1) over large alphabets: if $G = [I|A]$ has full Kruskal rank, then a $k \times (k-1)$ sub-matrix of G contains a full weight element in its left null space if and only if it is a submatrix of A. This follows from Theorems 1 and 2 and the fact that a shortened MDS code is still an MDS code.*

B. Linear codes

In this section we show that there is a tradeoff between β^* and δ, the relative minimum distance of a linear code. We recall that $\rho = \frac{n}{k}$.

Theorem 3. *Let $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ be a linear code of relative minimum distance δ with $q > k$. Then there exists $x \in \mathbb{F}_q^k$ with $|x| = k$ such that $|f(x)| \leq \frac{n}{2}(1 + \sqrt{1 - \frac{4\delta}{\rho(1-\frac{1}{n})^2} + \frac{4}{n(1-\frac{1}{n})^2}}) + 1$. In other words,*

$$\beta^* \leq \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4\delta}{\rho}}$$

More generally, for $\alpha < 1$ we asymptotically have

$$\beta(\alpha) \leq 1 - \frac{1 + \frac{1}{\rho} - \frac{\alpha}{\rho}}{2}\Big(1 - \sqrt{1 - \frac{4(1 - \alpha(1-\delta))}{\rho(1 + \frac{1}{\rho} - \frac{\alpha}{\rho})^2}}\Big)$$

Furthermore, if $|f(x)| \geq n - t$ for all x with $|x| = k$ and some $t < k$, then there exists x with $|x| \geq t$ such that $|f(x)| \leq (n-t)\frac{t+1}{k}$. In other words, for all $\alpha < \rho(1-\beta^*)$ we have $\beta(\alpha) \leq \rho\beta^*(1-\beta^*)$. In particular, if a code achieves the above bound on β^*, then for all $\alpha < \frac{\rho}{2}(1 - \sqrt{1 - \frac{4\delta}{\rho}})$ we have $\beta(\alpha) = \delta$.

Proof. Consider two sets A of m arbitrary points and B of l points inside $\mathbf{P}^r_{\mathbb{F}_q}$ with the property that any hyperplane contains at most $l(1-\delta)$ fraction of the points in B. Consider successive projections of A, B from the points in (the image of) B that are not in (the image of) A. Note that, to project from a point p, we draw a line from p to every point (except for p) in A, B, and map that point to the intersection of the line with $\mathbf{P}^r$. Suppose that after s projections we can no longer find any B-point to further project from. We say that a B-point is lost in projection if its image is not defined (i.e., it lies on the point from which we project). Let λ be the number of points in $B \backslash A$ that are lost in the projections after t steps. Suppose the image of A contains m' unique points $p_1, \cdots, p_{m'}$ inside $\mathbf{P}^{r'}$ where $r' := r - t$. The image of B contains $l - \lambda$ points counted with multiplicities. Let b_i be the number of points in B that get mapped to p_i in the image of A. We may assume that $b_1 \geq b_2 \geq ... \geq b_{m'}$. On average, there are $c = \frac{l-\lambda}{m'} \geq \frac{l-\lambda}{m}$ points of B lying on top of a point in A. If we pick a hyperplane that passes through $p_1, \cdots, p_{r'}$ inside $\mathbf{P}^{r'}_{\mathbb{F}_q}$, it must contain at least $r'\frac{l-\lambda}{m}$ points in the image of B. We can lift this hyperplane back to $\mathbf{P}^r_{\mathbb{F}_q}$ to get a hyperplane containing at least $\lambda + \frac{(l-\lambda)(r-t)}{m}$ points in B. The assumption on B requires that

$$\delta l \leq l - \left((l-\lambda)\frac{(r-t)}{m} + \lambda\right) \tag{2}$$

Using $\lambda \geq t$, we can write this as

$$t \geq \frac{\delta l m}{l-t} - m + r \tag{3}$$

This implies:

$$t \geq \frac{r + l - m}{2}\left(1 - \sqrt{1 - \frac{4l(-m(1-\delta) + r)}{(r + l - m)^2}}\right) \tag{4}$$

If $q > m$, there exists a hyperplane inside $\mathbf{P}^{r'}_{\mathbb{F}_q}$ that contains no point in the image A' of A. To see this, note that there are $\frac{q^{r'}-1}{q-1}$ hyperplanes inside $\mathbf{P}^{r'}_{\mathbb{F}_q}$. For a fixed point $p \in A'$, there are $\frac{q^{r'-1}-1}{q-1}$ hyperplanes that pass through p. By the union bound, if $\frac{q^{r'}-1}{q-1} > m\frac{q^{r'-1}-1}{q-1}$, there must exist a hyperplane that passes through no point of A'. Setting $l := n, m := k, r := k-1$, we get

$$\beta^* \leq 1 - \frac{t}{n} \leq \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4\delta}{\rho}} \tag{5}$$

as desired. We can remove a point p from A and apply the above argument to $A \backslash \{p\}$. If $\alpha < 1 - \frac{s}{k}$, then removing s points from A gives

$$\beta(\alpha) \leq 1 - \frac{1 + \frac{1-\alpha}{\rho}}{2}[1 - \sqrt{1 - \frac{4(1 - \alpha(1-\delta))}{\rho(1 + \frac{1}{\rho} - \frac{\alpha}{\rho})^2}}] \tag{6}$$

Now suppose that $f(x) \geq n - t$ for all x with $|x| = k$. Then the above sequence of projections must stop after t steps. Applying the same argument as above will prove the second part. ∎

Remark 3. *This result shows that there is a tradeoff between the "smoothness" of a code and its ability to correct errors. The tradeoff stems from the fact that smoothness requires local structures (c.f. [12]), and these in turn cannot spread messages too far out.*

Remark 4. *This result strengthens the connection between the (α, β)-property and the locality sensitive hashing (LSH) property. A priori, the (α, β)-property is only a relaxation of the LSH condition (see [1]), in the sense that a map that is good in the (α, β)-sense sends far away messages to far away codewords. This result suggests that such map must send some nearby messages to nearby codewords as well.*

Remark 5. *For MDS codes the above result gives that $\beta^* \leq 1 - \frac{1}{\rho}$ for $\rho \geq 2$ and $\beta^* \leq \frac{1}{\rho}$ for $\rho \leq 2$, which agrees with Theorems 1,2. The repetition code can asymptotically achieve $\delta = 0$ and $\beta^* = 1$. Thus the bound is tight at the two extreme points $\delta = 0, \delta = 1 - \frac{1}{\rho}$. The bound can be achieved at other values of δ as well.*

Remark 6. *It follows from the above proof that any linear code achieving $\beta^* = 1$ in the asymptotic regime (as $k \to \infty$) must be repetition-like, that is almost all columns of the generator matrix must have weight 1. Indeed the depth of the above projection sequence can be at most $o(k)$ for any such*

code. The authors present in [13] a family of non-linear codes that can achieve $\beta^* = 1$.

Remark 7. *It is asked in [1] what codes can (asymptotically) achieve $\alpha = \beta$ when ρ is not an integer. It follows from our proof that such codes, if they exist, cannot be linear (over large alphabets). Indeed one can check that there are no repetition-like codes achieving $\alpha = \beta$ for non-integral ρ and any linear code achieving $\beta^* = 1$ is repetition-like as discussed above.*

Problem 1. *The bound of Theorem 3 can be tight when the alphabet size is large. It is a (hard) open problem to improve the bound over small alphabets.*

III. Designing linear maps with good (α, β)-properties

Let C be the graph of a linear map inside $\mathbb{F}_2^k \times \mathbb{F}_2^n$. Given $u \in \mathbb{F}_2^{n+k}$, write it as (u_α, u_β), where u_α contains the first k coordinates. Define

$$g(u) = x^{|u_\alpha|} y^{|u_\beta|}$$

where $|\cdot|$ denotes the Hamming weight. Let

$$W_C(x, y) = \sum_{u_\alpha, u_\beta \in C} g(u)$$

be the bi-weight enumerator of the map. Then

Proposition 1 ([9])**.** *The following (MacWilliams) identity holds:*

$$W_{C^\perp}(x, y) = \frac{1}{|C|}(1+x)^k (1+y)^n W_C\left(\frac{1-x}{1+x}, \frac{1-y}{1+y}\right)$$

If we define, $P_l(k, x)$ to be the Krawchouk polynomial

$$P_l(k, x) = \sum_{s=0}^{i} (-1)^s \binom{i}{s} \binom{k-x}{l-s} \tag{7}$$

and $P_{lm}(i, j) = P_l(k, i) P_m(n, j)$, then the above identities give that

$$A'_{lm} = \frac{1}{2^n} \sum_{i,j=0}^{k,n} P_{lm}(i, j) A_{ij} \tag{8}$$

is non-negative. Here A_{ij} (resp. A'_{ij})) denotes the number of codewords of bi-weight i, j in C (resp. $C^\perp$). Then any linear map with (α, β)-property must satisfy the following set of constraints:

$$\begin{aligned} &\sum_{i,j=0}^{k,n} P_{lm}(i, j) A_{ij} \geq 0, \quad \forall l \leq n-k, m \leq n \\ &A_{00} = 1, \quad A_{ij} \geq 0, \quad \sum_j A_{ij} = \binom{k}{i}, \quad \sum_{j \leq \beta(\frac{i}{k})n} A_{ij} = 0; \end{aligned} \tag{9}$$

To bound the size of a candidate code one can vary k and check the feasibility of the above set of linear constraints.

We consider two notions of (α, β)-optimality. One of them requires one to compare (α, β)-properties of codes with different dimension. In such settings, it becomes useful to have an absolute version of $\beta(\alpha)$. We define

$$A_i^*(f) := \inf\{|f(x) - f(y)| : |x - y| \geq i\} \tag{10}$$

Definition 2 (Weakly optimal maps)**.** *A code $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is said to be weakly (α, β)-optimal if there does not exist $f' : \mathbb{F}_q^{k+1} \to \mathbb{F}_q^n$ such that*

$$A_i^*(f') \geq A_i^*(f) \quad \forall i \leq k$$

In other words, a code f is weakly optimal if no code with larger dimension can achieve the same or better $A_i^*(f)$'s.

Definition 3 (Strongly optimal maps)**.** *A code $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is said to be strongly (α, β)-optimal if it is not dominated by any other code, i.e., there does not exist an code $f' : \mathbb{F}_q^k \to \mathbb{F}_q^n$ such that*

$$A_i^*(f') \geq A_i^*(f) \quad \forall i \leq k$$

where at least one inequality is strict.

The examples below show that weak optimality is indeed strictly weaker than strong optimality. For the reverse direction, we have the following result:

Proposition 2. *A strongly optimal map is weakly optimal.*

In other words, if there exist a larger code that achieves the same (α, β)-profile as f, then f cannot be strongly optimal. Before we present the proof we remark that the analogous statement for minimum distance is false. Indeed, a code maybe optimal in the sense of minimum distance, yet, there may exists a larger code that achieves the same minimum distance. For instance, Tanner [14] constructed a binary [12,4,6]-code. The linear programming (LP) bound rules out the existence of a [12,3,7]-code. Thus any [12,3,6]-subcode of the Tanner code is still optimal in the sense of minimum distance. In general, one can expect such codes to exists over any field where the singleton bound is not tight. Over such fields, the existence of an $[n, k+1, d]$-code need not imply the existence of an $[n, k, d+1]$-code. However, the above proposition states that the existence of an $[n, k+1]$-code implies the existence an $[n, k]$-code with improved A_i^*'s.

Proof (of Proposition 2). Suppose a strongly optimal $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ is weakly dominated by $f' : \mathbb{F}_q^{k+1} \to \mathbb{F}_q^n$. Take the 1st coordinate and select the most common symbol among the codewords of f'. Take all the codewords of f' that start with this common symbol and remove the rest of the codewords. Now shorten the code by removing the first coordinate. This gives an $(n-1, k)$-subcode of f' with the same A_i^*'s as f. Now define an extension of f' as follows: $f''(x) := (f'(x), x_1)$ where x_1 is the first input coordinate. Clearly, all messages x, x' with $d(x, x') = k$ are sent to codewords that have distance $|f''(x) - f''(x')| = 1 + |f(x) - f(x')|$. This violates strong optimality of f. ∎

A. A weakly optimal quasi-cyclic code

Here we present a code that is optimal in the weak sense. Let $C_\beta \subset \mathbb{F}_2^7$ be the [7,4]-cyclic code generated by the primitive polynomial $x^3 + x + 1$. Similarly, C_{β^3} denotes the code generated by the primitive polynomial of β^3 (which is $x^3 + x^2 + 1$). Consider the code

$$C = \{(x, y) | x \in C_\beta, y \in C_{\beta^3}\} \tag{11}$$

The code has a minimum distance of 6. One can check that after applying a linear transform $x \to x + x^2$, the resulting spectrum contains the following (α, β) pairs:

$A_1^* = A_2^* = 6, A_3^* = 8, A_4^* = 10$ (see (10) for the definition of A_i^*), with the following generating matrix:

$$G = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{12}$$

Under these (α, β)-constraints, the LP in (9) becomes infeasible for $k = 5$. This implies that no $[14, 5]$-code exists with the same (or better) (α, β)-properties. We note that relaxing any of the (α, β)-constraints in the linear program will render the LP feasible with $k = 5$. This code is optimal in the weak sense but not in the strong sense as the construction below shows. We extend this code by appending the column $c := [0, 1, 0, 1]'$ to its generating matrix so it has comparable length with the Hadamard code. It becomes a $[15, 4]$-code with the following generating matrix:

$$G_e = \left[\begin{array}{c|c} G & c \end{array} \right] \tag{13}$$

After the extension, the code contains the following (α, β)-pairs:$A_1^* = 6, A_2^* = 7, A_3^* = 9, A_4^* = 11$.

B. A strongly optimal linear code

We ask if there exists a $[14, 4]$-code that dominates the quasi-cyclic code of (12). The LP in (9) is infeasible if we set $A_2^* = 7$ while keeping the rest of A_i^*'s from above unchanged. However, one can ask if there exists a code with the following profile $A_1^* = 6, A_2^* = 6, A_3^* = 9, A_4^* = 12$. The space of $[14, 4]$-linear codes is too big to search over. The LP in (9) can help reduce the size of the search space by severely restricting A_{1j}'s. With the above A_i^*'s, it turns out that the LP is infeasible when $A_{16} < 3$. This means that such a linear code can exists only if at least three of the rows in its generating matrix have weight 6. We can now efficiently search over the space of linear codes with $A_{16} = 3$ after taking out the symmetries. Here is the generator matrix of a code that was found using computer search over the reduced search space:

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{14}$$

The LP and some mild extra work suffice to prove that this code is optimal in strong sense. We also extend this code by adding a column $c = [1, 0, 0, 0]'$ to its generating matrix to make it have the same length as the Hadamard code:

$$G_e = \left[\begin{array}{c|c} G & c \end{array} \right] \tag{15}$$

The corresponding BER profile when used in communication over BSC is shown in Fig.1. It can be seen that for a wide range of channel parameters p the code of (15) outperforms both the quasi-cyclic code of (13) and the Hadamard code. We note that the $[15, 4, 8]$ Hadamard code is also optimal in the strong sense, as is the shortened $[14, 4, 7]$ Hadamard code. It has larger distance but lower β^* than the other two codes and thus serves as a reasonable basis for BER comparisons. While the BER differences may seem marginal, we expect to see more significant improvements for larger codes.

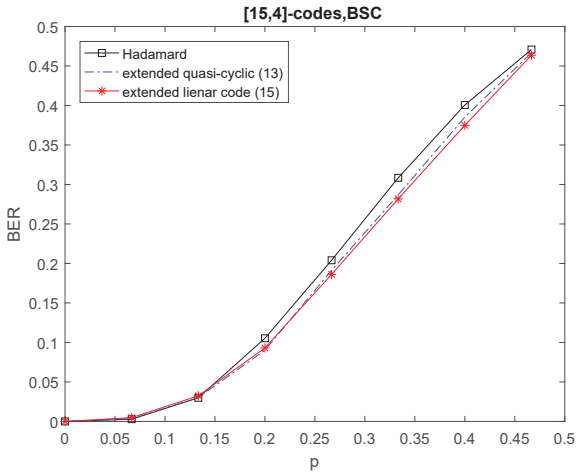

Fig. 1: The BER profiles under bitMAP decoding for: 1) the [15,4] Hadamard code 2) the [15,4] extended quasi-cyclic code of (13) 3) the extended linear code of (15).

REFERENCES

[1] Y. Polyanskiy, "On metric properties of maps between hamming spaces and related graph homomorphisms," *arXiv preprint arXiv:1503.02779*, 2015.

[2] Y. Kochman, A. Mazumdar, and Y. Polyanskiy, "The adversarial joint source-channel problem," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2112–2116.

[3] A. J. Young and Y. Polyanskiy, "Converse and duality results for combinatorial source-channel coding in binary hamming spaces," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 261–265.

[4] Y. Kochman, A. Mazumdar, and Y. Polyanskiy, "Results on combinatorial joint source-channel coding," in *Information Theory Workshop (ITW), 2012 IEEE*. IEEE, 2012, pp. 10–14.

[5] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 600–607, 1967.

[6] S. I. Bross and S. Litsyn, "Improved upper bounds for codes with unequal error protection," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3329–3333, 2006.

[7] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE journal on selected areas in communications*, vol. 18, no. 6, pp. 819–828, 2000.

[8] N. Rahnavard and F. Fekri, "Unequal error protection using low-density parity-check codes," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*. IEEE, 2004, p. 449.

[9] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*. Elsevier, 1977.

[10] A. Mazumdar, Y. Polyanskiy, A. S. Rawat, and H. Roozbehani, "Distance preserving maps and combinatorial joint source-channel coding for large alphabets," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 3067–3071.

[11] Y. Polyanskiy and A. Samorodnitsky, "Improved log-sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube," *arXiv preprint arXiv:1606.07491*, 2016.

[12] I. Benjamini, D. Ellis, E. Friedgut, N. Keller, and A. Sen, "Juntas in the ℓ^1-grid and lipschitz maps between discrete tori," *Random Structures & Algorithms*, vol. 49, no. 2, pp. 253–279, 2016.

[13] H. Roozbehani and Y. Polyanskiy, "Input-output distance properties of good linear codes," in *Information Science and Systems (CISS), 2018 Annual Conference on*. IEEE, 2018.

[14] R. M. Tanner, "A transform theory for a class of group-invariant codes," *IEEE transactions on information theory*, vol. 34, no. 4, pp. 725–775, 1988.