Bounds on the Reliability of a Typewriter Channel

Marco Dalai
University of Brescia
marco.dalai@unibs.it

Yury Polyanskiy
Massachusetts Institute of Technology
yp@mit.edu

***Abstract*—We give new bounds on the reliability function of a typewriter channel with 5 inputs and crossover probability $1/2$. The lower bound is more of theoretical than practical importance; it improves very marginally the expurgated bound, providing a counterexample to a conjecture on its tightness by Shannon, Gallager and Berlekamp which does not need the construction of algebraic-geometric codes previously used by Katsman, Tsfasman and Vlăduţ. The upper bound is derived by using an adaptation of the linear programming bound and it is essentially useful as a low-rate anchor for the straight line bound.**

I. Introduction

Consider the typewriter channel $W:\mathbb{Z}_5\to\mathbb{Z}_5$ with five inputs [1, Fig. 2] and crossover probability $1/2$. This channel has a great historical importance [1], [2]. In this paper we study the problem of bounding its reliability function $E(R)$ defined by

$$E(R)=\limsup_{n\to\infty}\frac{1}{n}\log\frac{1}{\mathsf{P}_\mathrm{e}(\lceil 2^{nR}\rceil,n)}$$

where $\mathsf{P}_\mathrm{e}(M,n)$ is the smallest possible probability of error[1] of codes with M codewords of length n. Here and below, rates are in bits and all logarithms are taken to base 2.

The interval of interest is $C_0<R<C$, where $C_0=\log\sqrt{5}$ is the zero-error capacity and $C=\log(5/2)$ is the ordinary capacity, since $E(R)=+\infty$ for $R\leq C_0$ and $E(R)=0$ for $R\geq C$. All equations below should be interpreted as restricted to this interval. To the best of our knowledge, the best known bounds in the literature date back to [4], [5], [2] and reduce to the following (see Section II for a detailed discussion of these bounds).

Proposition 1 (Random/expurgated bound [4]): We have $E(R)\geq E_{\text{r/ex}}(R)$ where

$$E_{\text{r/ex}}(R)=\log(5/2)-R. \tag{1}$$

Proposition 2 (Straight line bound [5]): We have $E(R)\leq E_{\text{sl}}(R)$ where

$$E_{\text{sl}}(R)=(\log(\sqrt{5}/2))^{-1}(\log(5/2)-R)\,. \tag{2}$$

Let $H_q(t)$ be the q-ary entropy function defined as

$$H_q(t)=t\log(q-1)-t\log t-(1-t)\log(1-t)\,.$$

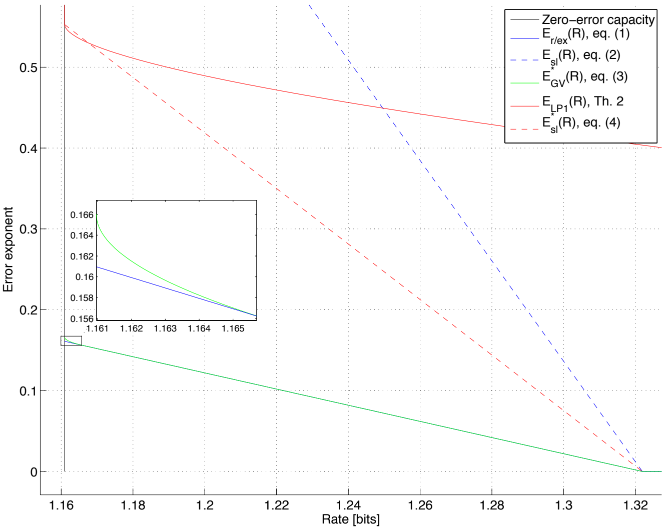

Fig. 1. Bounds for a 5-input typewriter channel with cross-over probability $1/2$.

In this paper we prove the following bounds (see Figure 1):

Theorem 1: We have $E(R)\geq E^*_{\text{GV}}(R)$ where

$$E^*_{\text{GV}}(R)=\begin{cases}\left(\frac{4}{3}-H_2(1/3)\right)\delta(R) & \log\sqrt{5}\leq R\leq R^*\\ E_{\text{r/ex}}(R) & \text{otherwise}\end{cases}, \tag{3}$$

$$R^*=\log(5)-\frac{1}{2}H_2(1/4)-\frac{3}{4},$$

and $\delta(R)$ is the solution of the equation

$$R=\log(5)-2\delta+\frac{1}{2}H_2(2\delta).$$

Theorem 2: We have $E(R)\leq E_{\text{LP1}}(R)$ where $E_{\text{LP1}}(\cdot)$ is defined implicitly by its inverse function

$$E^{-1}_{\text{LP1}}(E)=\log\sqrt{5}+R_{\text{LP1}}\left(\sqrt{5},E\right)$$

where

$$R_{\text{LP1}}(q,\delta)=H_q\left(\frac{(q-1)-(q-2)\delta-2\sqrt{(q-1)\delta(1-\delta)}}{q}\right)$$

is the linear programming bound for codes in a q-ary Hamming space.

Since $E_{\text{LP1}}(\log\sqrt{5})=(1-1/\sqrt{5})$, we obtain the following improvement of the straight line bound[2].

The research was supported by the NSF grant CCF-13-18620 and NSF Center for Science of Information (CSoI) under grant agreement CCF-09-39370. The work was mainly completed while the authors were visiting the Simons Institute for the Theory of Computing at UC Berkeley, whose support is gratefully acknowledged.

[1] In particular, the definition of $E(R)$ does not depend on whether we use maximal or average probability of error over codewords, see [3].

[2] Anchoring the straight line bound to E_{LP1} at $R=\log\sqrt{5}$ is only very marginally suboptimal, as is seen from Figure 1.

Corollary 1: $E(R)\leq E^*_{\text{sl}}(R)$ where

$$E^*_{\text{sl}}(R)=\left(1-\frac{1}{\sqrt{5}}\right)E_{\text{sl}}(R)\,. \qquad (4)$$

Remark 1: While Theorem 1 is expecially derived for the channel with five inputs, Theorem 2 can be extended to any odd number of inputs, and the proof in Section IV is given in this more general form.

II. Discussion of Propositions 1 and 2

In this section we prove that the best previously known bounds on $E(R)$ are those given by Propositions 1 and 2.

We first prove that the bound given in Proposition 1 corresponds to the best possible one which can be derived from Gallager's random coding and expurgated bounds [4], even when computed on blocks of arbitrary lengths. For this particular channel, since the capacity equals the cutoff rate, it suffices to consider the expurgated bound.

Let $g_n:\mathbb{Z}_5^n\times\mathbb{Z}_5^n\to\mathbb{R}$ be the function defined by

$$g_n(\boldsymbol{x}_1,\boldsymbol{x}_2)=\sum_{\boldsymbol{y}\in\mathbb{Z}_5^n}\sqrt{\boldsymbol{W}(\boldsymbol{y}|\boldsymbol{x}_1)\boldsymbol{W}(\boldsymbol{y}|\boldsymbol{x}_2)},$$

where $\boldsymbol{W}$ is the n-fold memoryless extension of W, and define

$$Q^n(\rho,P^n)=\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2}P^n(\boldsymbol{x}_1)P^n(\boldsymbol{x}_2)g_n(\boldsymbol{x}_1,\boldsymbol{x}_2)^{1/\rho}\,,$$

$$E^n_{\text{x}}(\rho)=-\frac{\rho}{n}\log\min_{P^n}Q^n(\rho,P^n)\,.$$

Gallager's bound [4] can be stated as $E(R)\geq E^n_{\text{ex}}(R)$, where

$$E^n_{\text{ex}}(R)=\sup_{\rho\geq1}\left[E^n_{\text{x}}(\rho)-\rho R\right]. \qquad (5)$$

This bound holds for any n and hence it makes sense to consider the optimal bound $\sup_n E^n_{\text{ex}}(R)$. Since the function $nE^n_{\text{ex}}(R)$ is super-additive in n, by Fekete's lemma we have

$$\sup_n E^n_{\text{ex}}(R)=\lim_{n\to\infty}E^n_{\text{ex}}(R).$$

We thus focus on the limit $E^\infty_{\text{ex}}(R)$, which gives the best bound $E(R)\geq E^\infty_{\text{ex}}(R)$. Computing $E^n_{\text{ex}}(R)$ for a general channel is prohibitive even for small values of n. However, for the considered typewriter channel, we can determine $E^\infty_{\text{ex}}(R)$. To the best of our knowledge the following proposition has not been reported before in the literature.

Proposition 3: For the considered channel, we have $E^\infty_{\text{ex}}(R)=E^2_{\text{ex}}(R)=E_{\text{r/ex}}(R)$.

Proof: Consider first the minimization of the quadratic form $Q^n(\rho,P^n)$ and note that the $5^n\times5^n$ matrix with elements $g_n(\boldsymbol{x}_1,\boldsymbol{x}_2)^{1/\rho}$, call it $g_n^{\odot\frac{1}{\rho}}$, is the n-fold Kronecker power of the 5×5 matrix

$$g_1^{\odot\frac{1}{\rho}}=\begin{pmatrix}1&\alpha&0&0&\alpha\\ \alpha&1&\alpha&0&0\\ 0&\alpha&1&\alpha&0\\ 0&0&\alpha&1&\alpha\\ \alpha&0&0&\alpha&1\end{pmatrix},\quad \alpha=2^{-\frac{1}{\rho}}\,.$$

As observed by Jelinek [6], if $g_1^{\odot\frac{1}{\rho}}$ is a positive semidefinite matrix, so is $g_n^{\odot\frac{1}{\rho}}$. In that case, the quadratic form defining $Q^n(\rho,P^n)$ for any n is a convex function of P^n. Jelinek showed that it is minimized by an i.i.d. distribution $P^n=P\times P\cdots\times P$, where P is optimal for $n=1$, and the achieved minimum is just the n-th power of the minimum achieved for $n=1$. Thus, if the matrix with elements $g_1(x_1,x_2)^{1/\rho}$ is positive semidefinite, then $E^n_{\text{x}}(\rho)=E^1_{\text{x}}(\rho)$. Furthermore, in this case the convexity of the quadratic form and the symmetry of $g_1^{\odot\frac{1}{\rho}}$ imply that the uniform distribution is optimal. Hence, by direct computation,

$$E^n_{\text{x}}(\rho)=-\rho\log\left(\frac{1}{5}+\frac{2^{1-1/\rho}}{5}\right) \qquad (6)$$

whenever $g_1^{\odot\frac{1}{\rho}}$ is positive semidefinite. Since its eigenvalues are $\lambda_k=1+2^{1-1/\rho}\cos(2\pi k/5)$, $k=0,\ldots,4$, the matrix is positive semidefinite for $\rho\leq\bar{\rho}=\log2/\log(2\cos(\pi/5))\approx1.4404$.

For $\rho>\bar{\rho}$, $g_1^{\odot\frac{1}{\rho}}$ is not positive semidefinite, and the minimization of $Q^n(\rho,P^n)$ is in general difficult to study. We prove that $E^n_{\text{x}}(\rho)\leq\rho\log(5)/2$ and then show that $E^2_{\text{x}}(\rho)\geq\rho\log(5)/2$, which implies $E^\infty_{\text{x}}(\rho)=E^2_{\text{x}}(\rho)=\rho\log(5)/2$. To prove this, observe that the minimum of $Q^n(\rho,P^n)$ is non-decreasing in ρ, and hence for $\rho>\bar{\rho}$

$$\begin{aligned}E^n_{\text{x}}(\rho)&\leq-\frac{\rho}{n}\log\min_{P^n}Q^n(\bar{\rho},P^n)\\&=\frac{\rho}{\bar{\rho}}E^n_{\text{x}}(\bar{\rho})\\&=\rho\log(5)/2\,,\end{aligned}$$

where the last step is obtained using the definition of $\bar{\rho}$ and equation (6). To prove that $E^2_{\text{x}}(\rho)\geq\rho\log(5)/2$, simply evaluate the function $Q^2(\rho,P^n)$ when P^2 is the indicator function of Shanon's zero-error code of length two for the pentagon.

So, we have finally proved that

$$E^\infty_{\text{x}}(\rho)=E^2_{\text{x}}(\rho)=\begin{cases}-\rho\log\left(\frac{1}{5}+\frac{2^{1-1/\rho}}{5}\right) & \text{if } \rho\leq\bar{\rho}\\ \rho\log(5)/2 & \text{if } \rho>\bar{\rho}\end{cases}.$$

Explicit computation of $E^2_{\text{ex}}(R)$ then reveals that the supremum over $\rho\geq1$ is achieved by $\rho=1$ if $R\geq\log(5)/2$ and as $\rho\to\infty$ if $R<\log(5)/2$, proving $E^2_{\text{ex}}(R)=E_{\text{r/ex}}(R)$. ■

Note in particular that, for $R\geq\log(5)/2$, $E^2_{\text{ex}}(R)$ coincides with the simple random coding bound [4], while it is infinite for $R<\log(5)/2$ as implied by the known existence of zero-error codes at those rates [1].

Proposition 3 combined with [4] implies the proof of Proposition 1. Since, to the best of out knowledge, no previous improvement of the bound $E^\infty_{\text{ex}}(R)$ was known for this channel, it also implies that $E_{\text{r/ex}}(R)$ is the best bound on $E(R)$ deducible from the known results in the literature.

Remark 2: The quantity $E^\infty_{\text{ex}}(R)$ was conjectured[3] to equal the true reliability function in [3]. This conjecture was disproved by Katsman, Tsfasman and Vlăduţ [7] using algebraic-geometric codes which beat the Gilbert-Varshamov bound. To

[3] Citing from [3]: "The authors would all tend to conjecture [...] As yet there is little concrete evidence for this conjecture."

the best of our knowledge, no other disproof is known in the literature. Theorem 1 proves that $E(R) > E_{\text{ex}}^{\infty}(R)$ for the considered channel and hence it offers a second disproof of the conjecture, which only uses an elementary extension of the Gilbert-Varshamov procedure carefully tuned for the particular case at hand.

We finally comment the bound stated in Proposition 2 showing how it is derived and why it is not trivial to improve it. For the particular channel considered, the sphere packing bound is essentially trivial; it states that $E(R) \leq 0$ above capacity, that is $R{>}\log(5/2)$, while $E(R){\leq}\infty$ below capacity. On the other hand, Lovász' proof of the zero-error capacity implies that $E(R)$ is finite for $R{>}C_0{=}\log(5)/2$. Finding good upper bounds on $E(R)$ in the range $\log(5)/2 < R < \log(5/2)$ appears to be a difficult problem. To the best of our knowledge the most effective bound to date is obtained as follows. For $R{>}\log(5)/2$ at least two codewords are confusable and, from the point of view of these two codewords, the channel is like a binary erasure channel. Hence, in the extreme case when the two codewords are confusable but they differ in all positions, the probability of error is just the probability that all symbols are erased, that is 2^{-n}. This implies that $E(R) \leq \log 2$ for $R{>}\log(5)/2$. Using the straight line bound [3] we deduce the result of Proposition 2.

We observe that we have considered the optimistic condition where all confusable pairs of codewords differ in all possible positions. This may look too optimistic, but we point out that any sequence is confusable with 2^n other sequences which differ in all single position from the considered one. So, it is not obvious at all how we can improve the bound by reasoning along this line. We would need to prove that for $R{>}\log(5)/2$ there are sequences which are both confusable and differ only in a fraction $\delta < 1$ of the positions. This is precisely what we will do using a linear programming approach in Section IV.

III. Proof of Theorem 1

We upper bound the error probability for a random code by using a standard union bound on the probability of confusion among single pairs of codewords. The code is built using a Gilbert-Varshamov-like procedure, though we exploit carefully the properties of the channel to add some structure to the random code and obtain better results than just picking random independent codewords.

Consider a code with M codewords $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$. Let us first consider the probability $P(\boldsymbol{x}_j|\boldsymbol{x}_i)$ that a codeword $\boldsymbol{x}_i \in \mathbb{Z}_5^n$ sent through the channel is incorrectly decoded as a second codeword $\boldsymbol{x}_j \in \mathbb{Z}_5^n$. This is possible only if the two codewords are confusable, which means that their coordinates are all pairwise confusable. As for the problem of discriminating among these two codewords, the channel is equivalent to an erasure channel with erasure probability $1/2$. So, the sequence $\boldsymbol{x}_i$ can be incorrectly decoded as sequence $\boldsymbol{x}_j$ only if all differences are erased, which happens with probability $2^{-d_{\mathrm{H}}(\boldsymbol{x}_i,\boldsymbol{x}_j)}$. So, we have

$$P(\boldsymbol{x}_j|\boldsymbol{x}_i) \leq \begin{cases} 2^{-d_{\mathrm{H}}(\boldsymbol{x}_i,\boldsymbol{x}_j)} & \boldsymbol{x}_i,\boldsymbol{x}_j \text{ confusable} \\ 0 & \boldsymbol{x}_i,\boldsymbol{x}_j \text{ not confusable} \end{cases}$$

where d_{H} is the Hamming distance. We can rewrite this in a simpler form if we introduce a notion of distance $d{:}\mathbb{Z}_5 \times \mathbb{Z}_5 \to \{0,1,\infty\}$

$$d(x_1,x_2) = \begin{cases} 0 & x_1 = x_2 \\ 1 & x_1 - x_2 = \pm 1 \\ \infty & x_1 - x_2 \neq \pm 1 \end{cases}$$

and then extend it additively to sequences in $\mathbb{Z}_5^n$

$$d(\boldsymbol{x}_1,\boldsymbol{x}_2) = \sum_k d(x_{1,k},x_{2,k}).$$

Using this definition we can rewrite

$$P(\boldsymbol{x}_j|\boldsymbol{x}_i) \leq 2^{-d(\boldsymbol{x}_i,\boldsymbol{x}_j)}.$$

The average probability of error can then be bounded using the union bound as

$$\begin{aligned} \mathsf{P}_{\mathrm{e}} &\leq \frac{1}{M}\sum_{i \neq j} 2^{-d(\boldsymbol{x}_i,\boldsymbol{x}_j)} \\ &= \sum_{z=0}^{n} A_z 2^{-z}, \end{aligned}$$

where A_z is the spectrum of the code

$$A_z = \frac{1}{M}\left|\{(i,j): d(\boldsymbol{x}_i,\boldsymbol{x}_j) = z\}\right|.$$

Consider now linear codes of length $n' = 2n$ with $(n+k) \times 2n$ generator matrix of the form

$$G^+ = \begin{pmatrix} I_n & 2I_n \\ 0 & G \end{pmatrix}$$

where I_n is the $n \times n$ identity matrix and G is a $k \times n$ matrix. We will study the family of random codes obtained when G is a random matrix with uniform independent entries in $\mathbb{Z}_5$. Note that this corresponds to taking 5^k randomly shifted versions of the n-fold cartesian power of Shannon's zero-error code of length 2 [1]. Since we focus on linear codes, the A_z's take the simpler form $A_z = |\{i : w(\boldsymbol{x}_i) = z\}|$, where we set $w(\boldsymbol{x}_i) = d(\boldsymbol{x}_i,\boldsymbol{0})$, the weight of the codeword (and similarly $w(x) = d(x,0)$).

We now proceed to the study of A_t. We can decompose any information sequence $\boldsymbol{u} \in \mathbb{Z}_5^{n+k}$ in two parts, $\boldsymbol{u} = (\boldsymbol{u}_1,\boldsymbol{u}_2)$, with $\boldsymbol{u}_1 \in \mathbb{Z}_5^n$ and $\boldsymbol{u}_2 \in \mathbb{Z}_5^k$. The associated codeword $\boldsymbol{v} = \boldsymbol{u}G^+$ can be correspondingly decomposed in two parts $\boldsymbol{v} = (\boldsymbol{v}_1,\boldsymbol{v}_2)$ with $\boldsymbol{v}_1 = \boldsymbol{u}_1$ and $\boldsymbol{v}_2 = 2\boldsymbol{u}_1 + \boldsymbol{u}_2 G$. Call $\boldsymbol{\nu} = \boldsymbol{u}_2 G$. We now relate the weight $w(\boldsymbol{v})$ to the *Hamming weight* $w_{\mathrm{H}}(\boldsymbol{\nu})$ and to the form of $\boldsymbol{u}_1$. Note in particular that we can write

$$w(\boldsymbol{v}) = \sum_{i=1}^{n} w((v_{1,i},v_{2,i}))$$

and that

$$(v_{1,i},v_{2,i}) = u_{1,i}(1,2) + (0,\nu_i).$$

Note first that $w(\boldsymbol{v}) = \infty$ if $u_{1,i} = \pm 2$ for some i. So, for the study of A_z we need only consider the cases $u_{1,i} \in \{0,\pm 1\}$. Consider first the case when $\nu_i = 0$. If $u_{1,i} = 0$ then $w(v_{1,i}) = w(v_{2,i}) = 0$ while if $u_{1,i} = \pm 1$ then $w(v_{2,i})$ is infinite. So, if

$\nu_i{=}0$ one choice of $u_{1,i}$ gives no contribution to $w(\boldsymbol{v})$ while all other choices lead to $w(\boldsymbol{v}){=}\infty$, and hence give no contribution to A_z for any finite z. Consider then the case of a component $\nu_i \neq 0$. It is not too difficult to check that one choice of $u_{1,i}$ in $\{0,\pm 1\}$ gives $w(v_{1,i}) = w(v_{2,i}) = 1$, one gives $w(v_{1,i}) = 1$ and $w(v_{2,i}) = 0$ or vice-versa, and the remaining one gives $w(v_{2,i}) = \infty$. So, if $\nu_i \neq 0$ one choice of $u_{1,i}$ contributes 1 to $w(\boldsymbol{v})$, one choice of $u_{1,i}$ contributes 2, while all other choices lead to $w(\boldsymbol{v}) = \infty$, and hence give no contribution to A_z for any finite z.

So, for a fixed $\boldsymbol{\nu}$ of Hamming weight d, and for a fixed $t{\in}\{1{,}2{,}\ldots{,}d\}$, there are $\binom{d}{t}$ vectors $\boldsymbol{u}_1$ which give codewords $\boldsymbol{v}$ of weight $2t+(d-t){=}d{+}t$. If B_d is the number of sequences $\boldsymbol{u}_2$ which lead to a $\boldsymbol{\nu}$ of Hamming weight d, then we have

$$\mathsf{P}_{\mathrm{e}} \leq \sum_{d=1}^{n} \sum_{t=1}^{d} B_d \binom{d}{t} 2^{-(d+t)} . \tag{7}$$

But B_d is now simply the spectrum of the linear code with generator matrix G, and it is known from the Gilbert-Varshamov procedure that as we let n and k grow to infinity with ratio $k/n \to r$, matrices G exist for which

$$B_{\delta n} = \begin{cases} 0 & \text{if } \delta < \delta_{GV}(r) \\ 5^{n(r-1)} \binom{n}{\delta n} 4^{n\delta}(1+o(1)) & \text{if } \delta \geq \delta_{GV}(r) \end{cases}$$

where $\delta_{GV}(r)$ is the Gilbert-Varshamov bound at rate r determined implicitly by the relation

$$r \log(5) = \log(5) - H_2(\delta_{GV}(r)) + 2\delta_{GV}(r) .$$

Defining $\delta{=}d/n$, $\tau{=}t/d$ and $r{=}k/n$, the probability of error is bounded to the first order in the exponent by the largest term in the sum (7) as

$$\begin{aligned} \frac{1}{n} \log \mathsf{P}_{\mathrm{e}} \leq \max_{\delta \geq \delta_{GV}(r), \tau \in [0,1]} & [\log(5)(r-1) + H_2(\delta) + 2\delta \\ & + \delta H_2(\tau) - (\delta + \delta \cdot \tau)] + o(1) . \end{aligned}$$

The maximum over τ is obtained by maximizing $H_2(\tau) - \tau$, which gives $\tau = 1/3$, independently of δ. So, we are left with the maximization

$$\max_{\delta \geq \delta_{GV}(r)} [\log(5)(r-1) + H_2(\delta) + \delta(h(1/3) + 2/3)] .$$

The argument is increasing for $\delta \leq 3/4$, where it achieves the maximum value $\log(5)(r-1)+2$, and decreasing for $\delta > 3/4$. So, the maximizing δ is $\delta = 3/4$ if $3/4 \geq \delta_{GV}(r)$ and $\delta_{GV}(r)$ otherwise. Combining these facts, after some computation we find

$$\frac{1}{n} \log \mathsf{P}_{\mathrm{e}} \leq \begin{cases} (\log(5)(r-1)+2) + o(1), & \delta_{GV}(r) \leq 3/4 \\ \delta_{GV}(r)(H_2(1/3) - 4/3) + o(1), & \delta_{GV}(r) > 3/4 . \end{cases}$$

Considering that the block length is $2n$ and the rate of the global code is $R{=}\log(5)(1{+}r)/2$, after some simple algebraic manipulations we obtain the claimed bound.

IV. Proof of Theorem 2

The upper bound we derive here is based on bounding the maximum probability of error over all codewords $\mathsf{P}_{\mathrm{e,max}}$, which in turn we bound in terms of the minimum distance of codes for the distance measure introduced in the previous section. Note in particular that we have

$$\mathsf{P}_{\mathrm{e,max}} \geq \max_{i \neq j} \frac{1}{2} \cdot 2^{-d(\boldsymbol{x}_i, \boldsymbol{x}_j)} .$$

Indeed, if there is no pair of confusable codewords, then the inequality is trivial, while if codewords i and j are confusable, any (possibly randomized) decoder will decode in error with probability at least $1/2$ either when codeword i or codeword j is sent. So, we can bound the reliability as

$$E(R) \leq \min_{i \neq j} \frac{1}{n} d(\boldsymbol{x}_i, \boldsymbol{x}_j)(1+o(1)) . \tag{8}$$

The rest of this section is devoted to bounding the minimum distance. In particular we prove that codes for which

$$\min_{i \neq j} \frac{1}{n} d(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq \delta$$

have rate R upper bounded as

$$R \leq \frac{1}{2} \log 5 + R_{\mathrm{LP1}}(\sqrt{5}, \delta)(1+o(1)) . \tag{9}$$

Note that Theorem 2 follows from equations (8)-(9).

Our bound is based on θ functions and Delsarte's linear programming bound [8], but it is easier to describe it in terms of Fourier transforms. We set here $q = 5$ and give a proof in terms of q which also holds for any other odd larger value.

For any $f : \mathbb{Z}_q^n \to \mathbb{C}$ we define its Fourier transform as

$$\hat{f}(\boldsymbol{\omega}) = \sum_{\boldsymbol{x} \in \mathbb{Z}_q^n} f(\boldsymbol{x}) e^{\frac{2\pi i}{q} <\boldsymbol{\omega}, \boldsymbol{x}>}, \quad \boldsymbol{\omega} \in \mathbb{Z}_q^n$$

where the non-degenerate $\mathbb{Z}_q$-valued bilinear form is defined as usual

$$<\boldsymbol{x}, \boldsymbol{y}> \overset{\triangle}{=} \sum_{k=1}^{n} x_k y_k .$$

We also define the inner product as follows

$$(f, g) \overset{\triangle}{=} q^{-n} \sum_{\boldsymbol{x} \in \mathbb{Z}_q^n} \bar{f}(\boldsymbol{x}) g(\boldsymbol{x}) .$$

The starting point is a known rephrasing of linear programming bound. Let $\mathcal{C}$ be a code with minimum distance at least d. Let f be such that $f(\boldsymbol{x}) \leq 0$ if $d(\boldsymbol{x}, \mathbf{0}) = w(\boldsymbol{x}) \geq d$, $\hat{f} \geq 0$ and $\hat{f}(\mathbf{0}) > 0$. Then, consider the Plancherel identity

$$(f * 1_{\mathcal{C}}, 1_{\mathcal{C}}) = q^{-n} (\hat{f} \cdot \widehat{1_{\mathcal{C}}}, \widehat{1_{\mathcal{C}}}) ,$$

where 1_A is the indicator function of a set A. Upper bounding the left hand side by $|\mathcal{C}| f(\mathbf{0})$ and lower bounding the right hand side by the zero-frequency term $q^{-n} \hat{f}(\mathbf{0}) |\mathcal{C}|^2$, one gets

$$|\mathcal{C}| \leq \min q^n \frac{f(\mathbf{0})}{\hat{f}(\mathbf{0})} . \tag{10}$$

The proof of our theorem is based on a choice f which combines Lovász' assignment used to obtain his bound on the zero-error capacity with the one used in [9] to obtain bounds on the minimum distance of codes in Hamming spaces.

Observe first that Lovász assignment can be written in one dimension ($n=1$) as

$$g_1(x)=1_0(x)+\varphi 1_{\pm 1}(x),\quad x\in\mathbb{Z}_q\,,$$

where $\varphi=(2\cos(\pi/q))^{-1}$. This gives

$$\widehat{g_1}(\omega)=1+2\varphi\cos(2\pi\omega/q),\quad \omega\in\mathbb{Z}_q.$$

Correspondingly, define the n-dimensional assignment

$$g(\boldsymbol{x})=\prod_{j=1}^{n} g_1(x_j),\quad \hat{g}(\boldsymbol{\omega})=\prod_{j=1}^{n}\widehat{g_1}(\omega_j),\quad \boldsymbol{x},\boldsymbol{\omega}\in\mathbb{Z}_q^n.$$

Note that $\widehat{g_1}\geq 0$ and, additionally, $\widehat{g_1}(\omega)=0$ for $\omega=\pm c$, with $c=(q-1)/2$. So, $\hat{g}\geq 0$, with $g(\boldsymbol{\omega})=0$ if $\boldsymbol{\omega}$ contains any $\pm c$ entry. Since $g(\boldsymbol{x})=0$ for $\boldsymbol{x}\notin\{0,\pm1\}^n$, g satisfies all the properties required for f in the case $d=\infty$, and when used in place of f in (10) it gives Lovász' bound

$$\begin{aligned}|\mathcal{C}|&\leq q^n\frac{g(\mathbf{0})}{\hat{g}(\mathbf{0})}\\&=q^n\left(\frac{\cos(\pi/q)}{1+\cos(\pi/q)}\right)^n\end{aligned}$$

for codes of infinite minimum distance.

For the case of finite $d\leq n$, we build a function f of the form $f(\boldsymbol{x})=g(\boldsymbol{x})h(\boldsymbol{x})$, for an appropriate $h(\boldsymbol{x})$. In particular, since $g(\boldsymbol{x})$ is non-negative and already takes care of setting $f(\boldsymbol{x})$ to zero if $x\notin\{0,\pm1\}^n$, it suffices to choose h such that $h(\boldsymbol{x})\leq 0$ whenever $\boldsymbol{x}\in\{0,\pm1\}^n$ contains at least d entries with value ±1. We restrict attention to h such that $\hat{h}\geq 0$, so that $\hat{f}=q^{-n}\hat{g}*\hat{h}\geq 0$. In particular, we consider functions h whose Fourier transform is constant on each of the following "spheres" in $\mathbb{Z}_q^n$

$$S_\ell^c=\{\boldsymbol{\omega}:|\{i:\omega_i=\pm c\}|=\ell,|\{i:\omega_i=0\}|=n-\ell\},\quad \ell=0,\ldots,n,$$

and zero outside. This choice is motivated by the fact, observed before, that $\hat{g}_1(\pm c)=0$. Restricting $\hat{h}$ to be null out of these spheres simplifies the problem considerably. We thus define

$$\hat{h}(\boldsymbol{\omega})=\sum_{\ell=0}^{n}\hat{h}_\ell 1_{S_\ell^c}(\boldsymbol{\omega})\,,\quad h(\boldsymbol{x})=q^{-n}\sum_{\ell=0}^{n}\hat{h}_\ell\widehat{1_{S_\ell^c}}(\boldsymbol{x})\,,\tag{11}$$

where $\hat{h}_\ell\geq 0$ and $\hat{h}_0>0$ will be optimized later. Since $\hat{g}(\boldsymbol{\omega})=0$, $\boldsymbol{\omega}\in S_\ell\,,\ell>0$, setting $f(\boldsymbol{x})=g(\boldsymbol{x})h(\boldsymbol{x})$ gives $\hat{f}(\mathbf{0})=q^{-n}(\hat{g}*\hat{h})(\mathbf{0})=q^{-n}\hat{g}(\mathbf{0})\hat{h}_0$. So, the bound (10) becomes

$$|\mathcal{C}|\leq\left(q^n\frac{g(\mathbf{0})}{\hat{g}(\mathbf{0})}\right)\left(q^n\frac{h(\mathbf{0})}{\hat{h}_0}\right).$$

The first term above is precisely Lovász bound and corresponds, for $q=5$, to the $\frac{1}{2}\log(5)$ term in the right hand side of (9). We now show that the second term corresponds to the linear programming bound of an imaginary "Hamming scheme" with a special non-integer alphabet size $q'=1+\cos(\pi/q)^{-1}$. To do this, define analogously to S_ℓ^c the spheres

$$S_u^1=\{\boldsymbol{x}:|\{i:x_i=\pm1\}|=u,\,|\{i:x_i=0\}|=n-u\}\,.$$

Our constraint is that $h(\boldsymbol{x})\leq 0$ if $\boldsymbol{x}\in S_u^1$, $u\geq d$. Direct computation shows that for $\boldsymbol{x}\in S_u^1$,

$$\begin{aligned}\widehat{1_{S_\ell^c}}(\boldsymbol{x})&=\sum_{j=0}^{\ell}\binom{u}{j}\binom{n-u}{\ell-j}(-1)^j2^\ell(\cos(\pi/q))^j\,,\qquad(\boldsymbol{x}\in S_u^1)\\&=(2\cos(\pi/q))^\ell K_\ell(u;q'),\qquad(q'=1+\cos(\pi/q)^{-1}),\end{aligned}$$

where $K_\ell(u;q')$ is a Krawtchouck polynomial of degree ℓ and parameter q' in the variable u. We can thus define

$$\Lambda(u)=h(\boldsymbol{x})\,,\boldsymbol{x}\in S_u^1\,,\qquad \lambda_\ell=q^{-n}(2\cos(\pi/q))^\ell\cdot\hat{h}_\ell\,,$$

and write

$$q^n\frac{h(\mathbf{0})}{\hat{h}_0}=\frac{\Lambda(0)}{\lambda_0}\,,\tag{12}$$

where the conditions on h can be restated as

$$\begin{aligned}\Lambda(u)&=\sum_{\ell=0}^{n}\lambda_\ell K_\ell(u;q')\,,\quad u=0,\ldots,n,\\\lambda_\ell&\geq 0,\quad \ell\geq 0,\\\Lambda(u)&\leq 0,\quad u\geq d.\end{aligned}$$

So, the minimization of (12) is reduced to the standard linear programming problem for the Hamming space, though with a non-integer parameter q'. Since the construction of the polynomial used in [9] and [10] can be applied verbatim for non-integer values of q' (see also [11] for the position of the roots of $K_\ell(u;q')$), the claimed bound follows.

References

[1] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.

[2] L. Lovász, "On the Shannon Capacity of a Graph," *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[3] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[4] R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, 1965.

[5] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. II," *Information and Control*, vol. 10, pp. 522–552, 1967.

[6] F. Jelinek, "Evaluation of Expurgated Error Bounds," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, 1968.

[7] G. L. Katsman, M. A. Tsfasman, and S. G. Vlăduţ, "Spectra of linear codes and error probability of decoding," in *Coding Theory and Algebraic Geometry*, ser. Lecture Notes in Mathematics, 1992, vol. 1518, pp. 82–98.

[8] P. Delsarte, "An Algebraic Approach to the Association Schemes of Coding Theory," *Philips Res. Rep.*, vol. 10, 1973.

[9] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *Information Theory, IEEE Transactions on*, vol. 23, no. 2, pp. 157 – 166, mar 1977.

[10] M. Aaltonen, "A new upper bound on nonbinary block codes," *Discrete Mathematics*, vol. 83, no. 2, pp. 139–160, 1990.

[11] M. E. H. Ismail and P. Simeonov, "Strong Asymptotics for Krawtchouk Polynomials," *J. Comput. Appl. Math.*, vol. 100, no. 2, pp. 121–144, 1998.