Achieving Capacity on Non-Binary Channels with Generalized Reed–Muller Codes

Galen Reeves*† and Henry D. Pfister*‡
Departments of Electrical and Computer Engineering*, Statistical Science†, and Mathematics‡
Duke University

***Abstract*—Recently, the authors showed that Reed–Muller (RM) codes achieve capacity on binary memoryless symmetric (BMS) channels with respect to bit error rate. This paper extends that work by showing that RM codes defined on non-binary fields, known as generalized RM codes, achieve capacity on sufficiently symmetric non-binary channels with respect to symbol error rate. The new proof also simplifies the previous approach (for BMS channels) in a variety of ways that may be of independent interest.**

***Index Terms*—Channel Capacity, Group Codes, EXIT Area Theorem, Reed-Muller Code, Strong Data-Processing Inequality**

I. Introduction

Generalized Reed–Muller (GRM) codes were introduced by Kasami, Lin, and Peterson in 1968 [1] as the natural generalization of binary Reed–Muller (RM) codes [2], [3] to non-binary alphabets. GRM codes are closely related to other interesting code families including Reed–Solomon codes [4], multiplicity codes [5], and lifted codes [6]. These families remain interesting subjects of research due to their connections with topics such as local decodability and list decoding.

In 2016, it was established that sequences of RM codes can achieve capacity on the binary erasure channel (BEC) [7], [8]. This was followed by some extensions and related work [9]–[11]. A nice tutorial overview of RM codes and results until 2020 is provided by [12]. Then, in 2021, the authors showed that RM codes achieve capacity on binary memoryless symmetric (BMS) channels with respect to bit error rate [13].

The main result of this paper is the following theorem. We note that all terminology will be defined in later sections.

Theorem 1. *Consider a memoryless channel W with capacity C whose input alphabet is $\mathcal{X} = \mathbb{F}_q$ and let G be the symmetry group of the channel. Suppose one of the following holds:*

(i) G contains the affine group over $\mathbb{F}_q$;
(ii) q is prime, G contains the additive group of $\mathbb{F}_q$, and the smallest principal inertia component of W (for the uniform input distribution) is strictly positive.

Then, for every sequence of GRM codes over $\mathbb{F}_q$ with strictly increasing blocklength and rate converging to $R \in [0, C)$, the symbol-error rate (SER) under symbol-MAP decoding converges to zero.

This research was supported in part by NSF Grants 2106213, 2212437, and 1750362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Corollary 2. *If q is prime and the channel symmetry group is transitive (i.e., the channel is symmetric) but it does not contain the additive group of $\mathbb{F}_q$, then the input alphabet can be relabeled so that the conclusion of Theorem 1 still holds.*

Corollary 3. *Consider a memoryless channel W with input alphabet $\mathcal{X} = \mathbb{F}_q$ and let I_s be the mutual information between its input and output with a uniform input distribution. Consider block-coded transmission using group symmetrization over the affine group of $\mathbb{F}_q$ (i.e., each channel use is modulated by a random affine map known at the receiver). Then, for every sequence of GRM codes over $\mathbb{F}_q$ with strictly increasing blocklength and rate converging to $R \in [0, I_s)$, the SER under symbol-MAP decoding converges to zero.*

For the purposes of our analysis, there are two finite-input channels of interest. The first is the memoryless channel W (with capacity C) over which the codeword $\boldsymbol{X}$ is transmitted and the output $\boldsymbol{Y}$ is received. Our goal is to show that code rates strictly less than C can be achieved with vanishing symbol error rate. This involves analyzing a second channel from X_0 to $Y_{\sim 0}$, which we call the *coset* channel due to the group structure in the code.

The high-level idea of our proof is to first show that, as the blocklength increases, the sequence of coset channels converges to a *deterministic* channel (i.e., a channel for which a minimal sufficient statistic is a non-random function of its input). For binary inputs, the only deterministic channels are the perfect channel and uninformative channel but, for non-binary inputs, there are other possibilities. In the remainder of the proof, we use channel symmetry and the area theorem to argue that this limiting channel must be the perfect channel whenever the rate of the code is strictly less than capacity.

The two conditions appearing in Theorem 1 have different implications for the symmetry group of the implied coset channel. Case (i) implies doubly transitive symmetry of the coset channel, which simplifies much of the analysis. Case (ii) implies transitive symmetry of the coset channel, and in this case, we need the assumption that q is prime and some additional arguments to rule out the possibility that the coset channel converges to a deterministic limit that is neither perfect nor uninformative.

Comparison with prior work: The proof for binary RM codes on BMS channels [13] is based on the convergence of the power series expansion of the binary entropy function around the uninformative point. This fails for the non-binary

case because the analogous power series converges only on a small subset of the domain. Instead, an approach inspired by strong data-processing inequalities (Lemma 6) is used here to bound mean-squared error in terms of mutual information.

In [13], the "influences" of two subsets (of channel outputs) on the conditional mean estimator are bounded separately using two different arguments. In this paper, all influences are bounded using a single simpler argument (Lemma 15) that uses extrinsic information transfer (EXIT) functions [14] rather than generalized EXIT functions [15].

Our analysis of channels builds on the framework developed by Blackwell [16] and Le Cam [17]. Our introduction and analysis of the overlap matrix is related to principal inertial components (PICs) [18], [19] and strong data-processing inequalities [20]–[22]. A key innovation in this work is Lemma 15, which combines these ideas with a differential analysis of channels enabled by the EXIT area theorem.

Notation: The real numbers and are denoted by $\mathbb{R}$, the natural numbers are denoted by $\mathbb{N} \coloneqq \{1, 2, \ldots\}$, and $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$. For $N \in \mathbb{N}_0$, a range of natural numbers is denoted by $[N] \coloneqq \{0, 1, \ldots, N-1\}$. Let Δ_q denote the probability simplex on q elements and $\boldsymbol{e}_i \in \mathbb{R}_q$ be the i-th standard basis vector. We use $\mathbb{F}_q$ to denote the Galois field with q elements. For a set $\mathcal{X}$, the N-element vector $\boldsymbol{x} \in \mathcal{X}^N$ is denoted by boldface and is indexed from 0 so that $\boldsymbol{x} = (x_0, , \ldots, x_{N-1})$. For an M-element index set $A = \{a_0, a_1, \ldots, a_{M-1}\} \subseteq [N]$ with $a_0 < a_1 < \cdots < a_{M-1}$, we define the subvector $x_A = (x_{a_0}, x_{a_1}, \ldots, x_{a_{M-1}}) \in \mathcal{X}^M$ without using boldface. A single random variable is denoted by a capital letter (e.g., X, Y, Z). Vectors of random variables are denoted by boldface capital letters (e.g., $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$). All unspecified logarithms (i.e., $\log$'s) are taken base-q and thus expressions involving entropy and mutual information are reported in *qits*.

All proofs appear in the online version of this paper [23].

II. Channels

We assume throughout that $q \in \mathbb{N}$ with $q \geq 2$ and $\mathcal{X} = \{0, 1, \ldots, q-1\}$. The symmetric group $\mathbb{S}_q$ is the set of bijective functions (i.e., permutations) mapping $\mathcal{X}$ to $\mathcal{X}$ with the group operation given by composition. A permutation group $G \subseteq \mathbb{S}_q$ is *transitive* if, for each $x, x' \in \mathcal{X}$, there exists a permutation in G the maps x to x'. Likewise, it is *doubly transitive* if, for any $x_1, x_2, x_1', x_2' \in \mathcal{X}$ with $x_1 \neq x_2$ and $x_1' \neq x_2'$, there exists a permutation in G that maps x_k to x_k' for $k = 1, 2$.

We define the action of the symmetric group $\mathbb{S}_q$ on $\mathbb{R}^q$ (and the probability simplex $\Delta_q \subseteq \mathbb{R}^q$) according to

$$\sigma(v_0, \ldots, v_{q-1}) = (v_{\sigma(0)}, \ldots, v_{\sigma(q-1)})$$

for every $\sigma \in \mathbb{S}_q$ and $(v_0, \ldots, v_{q-1}) \in \mathbb{R}^q$. This operation is extended to a probability measure P on $\mathbb{R}^q$ via the pushforward measure σP defined by

$$(\sigma P)(B) = P(\sigma^{-1} B), \quad \forall B \in \mathcal{B}$$

where $\sigma^{-1} B = \{\sigma^{-1} p \,|\, p \in B\}$.

A. Finite-Input Channels

A q-ary input channel W is a conditional distribution mapping from an input $x \in \mathcal{X}$ to a probability measure $W(\cdot \,|\, x)$ on a measurable space $(\mathcal{Y}, \mathcal{A})$. When convenient, we use the compact notation $W_x(\cdot) = W(\cdot \,|\, x)$.

Following the approach of Blackwell [16], we introduce a standard version of the channel whose output alphabet is the probability simplex. For a channel W, the *canonical map* $\phi \colon \mathcal{Y} \to \Delta_q$ is defined by

$$\phi(y) \coloneqq \big(\phi_0(y), \phi_1(y), \ldots, \phi_{q-1}(y)\big),$$

where $\phi_x(y) \coloneqq (dW_x / d\bar{W})(y)$ is the Radon-Nikodym derivative of W_x with respect to the reference measure

$$\bar{W}(A) \coloneqq \sum_{x \in \mathcal{X}} W(A \,|\, x), \quad A \in \mathcal{A}.$$

The canonical map can also be viewed as the posterior pmf of with respect to a uniform prior distribution, i.e., $\phi_x(y)$ is the probability that the input is x given the output is y.

Composing the channel W with its canonical map produces a new channel, W^{s}, on the output space $(\Delta_q, \mathcal{B})$ satisfying

$$W^{\mathrm{s}}(B \,|\, x) = W(\phi^{-1} B \,|\, x), \quad \forall x \in \mathcal{X}, \forall B \in \mathcal{B}.$$

Because the canonical map is a sufficient statistic for the channel input, the mapping from W to W^{s} preserves the relevant properties of the channel, such as its capacity and minimum error probability.

Following Blackwell's definition of a standard experiment [16], we call a channel *standard* if its canonical map is the identity map, and we refer to W^{s} as the standard channel associated with W. Furthermore, we will call two channels *Blackwell equivalent* if they have the same standard channel.

B. Channel Symmetry

In communication theory, the term symmetric channel is used to refer to a variety of related (but distinct) symmetry conditions [24], [25]. This paper uses the following definition due to its compatibility with Blackwell equivalence.

Definition 1 (Channel Symmetry Group)**.** The symmetry group G of a q-ary channel W is the permutation group

$$G = \{\sigma \in \mathbb{S}_q \,|\, \forall x \in \mathcal{X}, \forall B \in \mathcal{B}, W^{\mathrm{s}}(\sigma B \,|\, \sigma x) = W^{\mathrm{s}}(B \,|\, x)\},$$

where W^{s} is the standard channel associated with W. In other words, G is the group of all permutations σ such that the distribution of the canonical map $\phi(Y)$ under input σx is equal to the distribution of $\sigma\psi(Y)$ under input x.

A channel is called *symmetric* if its symmetry group is transitive. Its well known that this condition is sufficient to ensure that the capacity of the channel is achieved by the uniform input distribution [24]. More generally, for any decision-theoretic problem whose loss function has the same symmetries as the channel, the uniform input distribution maximizes the expected loss [26, Chapter 6].

A slightly stronger notion of symmetry occurs when the symmetry group of a channel is associated with a group

structure on the input alphabet. Let $(\mathcal{X}, \boxplus)$ be a group with binary operation denoted by $\boxplus$, and assume without loss of generality that 0 is the identity element. Each $x \in \mathcal{X}$ defines a permutation $\sigma_x \in \mathbb{S}_q$ according to $\sigma_x x' = x \boxplus x'$ for every $x' \in \mathcal{X}$. By Cayley's theorem, the group $(\mathcal{X}, \boxplus)$ is isomorphic to the permutation group $H \coloneqq \{\sigma_x \,|\, x \in \mathcal{X}\}$, which we will refer to as the permutation representation of $(\mathcal{X}, \boxplus)$. The channel W is called *group symmetric* if its symmetry group contains H as a subgroup.

While there exist channels that are symmetric but not group symmetric, this can occur only if q is not a prime power.

Lemma 4. *If a channel W is symmetric with q equal to a prime power, then it is group symmetric.*

C. Group Symmetrization

There is a simple (and commonly used) process that can equip any channel with any desired symmetry. Moreover, if the channel capacity is achieved by a uniform input distribution, then this process does not change the capacity. Let W be a channel with input alphabet $\mathcal{X}$ and let H be any subgroup of $\mathbb{S}_q$. Now, define a new channel W' that, for the input $x \in \mathcal{X}$, chooses a uniform random element of $\sigma \in H$ and transmits σx through W. Then, the output of W' is defined to be (Y, σ) where $Y \in \mathcal{Y}$ is the output of W. We call this operation group symmetrization and the symmetry group of the resulting channel must contain H as a subgroup. The standard channel W'^{s} satisfies

$$W'^{\mathrm{s}}(B \,|\, x) = \frac{1}{|H|} \sum_{\sigma \in H} W^{\mathrm{s}}(\sigma^{-1} B \,|\, x).$$

If H is transitive, then the group symmetrization operation implies that the effective input distribution seen by the original channel W is uniform. So, if one is content to include group symmetrization in the system, then any desired channel symmetry can be engineered while still achieving the rate I_s equal to the mutual information between the channel input and output under a uniform input distribution.

D. Overlap Matrix

We define *overlap matrix* $Q \in \mathbb{R}^{q \times q}$ associated with a q-ary channel W according to

$$Q_{x,x'} \coloneqq \int \phi_x(y)\, \phi_{x'}(y) \bar{W}(dy).$$

This matrix is symmetric, positive semidefinite, and doubly stochastic. Thus, its eigenvalues are real positive numbers that satisfy $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{q-1} \geq 0$. It can be verified that the eigenvalues of index $x \geq 1$ correspond to the principal inertia components (PICs) of the channel with respect to the uniform input distribution [18], [19]. The smallest PIC λ_{q-1}, which plays a prominent role in our analysis, has been considered previously in the context of perfect privacy [19].

For random variables (X, Y), the *symbol error rate* (SER) is defined by

$$\mathrm{SER}(X \,|\, Y) \coloneqq \min_{X - Y - \hat{X}} \mathbb{P}\Big[X \neq \hat{X}\Big],$$

where the minimum is over all Markov chains $X - Y - \hat{X}$.

Lemma 5. *For any input-output pair (X, Y) through a q-ary channel with overlap matrix Q, we have*

$$\mathrm{SER}(X \,|\, Y) \leq 1 - \mathrm{Tr}(\mathrm{diag}(p) Q),$$

where $p \in \Delta_q$ is the prior pmf of X.

The following lemma will also us to connect MMSE estimation error with conditional mutual information.

Lemma 6. *For any Markov chain $S - T - X - Y$ where Y is an observation of X through q-ary channel, we have,*

$$\begin{aligned} &I(X; Y \,|\, S) - I(X; Y \,|\, T) \\ &\qquad \geq \frac{\lambda_{q-1}^2}{2 \ln q} \mathbb{E}\big[\|\mathbb{E}[\boldsymbol{e}_X \,|\, T] - \mathbb{E}[\boldsymbol{e}_X \,|\, S]\|^2\big], \end{aligned}$$

where λ_{q-1} is the minimal eigenvalue of the overlap matrix

For a q-ary channel W with canonical map ϕ, we define the squared-error discrepancy

$$\delta \coloneqq \mathbb{E}\big[\|\phi(Y) - \mathbb{E}[\phi(Y) \,|\, X]\|^2\big]$$

where Y is the output for a uniformly distributed input X. Note that δ is zero if and only if the output of the standard channel is determined uniquely by the input (i.e., the channel is deterministic). This discrepancy can also be expressed in terms of the overlap matrix Q or the PICs:

$$\delta = \frac{1}{q} \mathrm{Tr}(Q) - \frac{1}{q} \mathrm{Tr}(Q^2) = \frac{1}{q} \sum_{x=1}^{q-1} \lambda_x (1 - \lambda_x).$$

If δ is close to zero, then the PICs are clustered near the boundaries 0 and 1 of the unit interval. The next result gives sufficient conditions under which the PICs are all close to the same boundary point.

Lemma 7. *Consider a q-ary channel with symmetry group G and squared-error discrepancy δ. Suppose one of the following holds:*

(i) G is doubly transitive;
(ii) G is transitive, q is prime, and $4q^2\delta < 1$.

Then, the overlap matrix Q satisfies

$$\min_{b \in \{1, q\}} |\,\mathrm{Tr}(Q) - b| \leq 2q\delta.$$

III. Codes

A. Group Codes

In coding theory, the term group code is used to refer to a few related (but distinct) mathematical objects. These include binary codes closed under modulo-2 addition [25], sets of points real space generated by a group of orthogonal transformations applied to a single point [27], and codes whose codewords are elements of a group ring [28]. The group structure of binary codes was recognized early and exploited in [25], [29]. Later, similar ideas were developed for non-binary codes and channels [27], [30]. In this paper, we use the following definition.

Definition 2 (Group code)**.** Let $(\mathcal{X}, \boxplus)$ be a group where $0 \in \mathcal{X}$ is the identity. A set $\mathcal{C} \subseteq \mathcal{X}^N$ is called a *group code over* $(\mathcal{X}, \boxplus)$ if the set $\mathcal{C}$ forms a group with respect to the binary operation $\boldsymbol{x} \boxplus \boldsymbol{x}' = (x_0 \boxplus x'_0, x_1 \boxplus x'_1, \ldots, x_{N-1} \boxplus x'_{N-1})$.

Definition 3 (Matched to channel)**.** A group code over $(\mathcal{X}, \boxplus)$ is matched to channel W if the symmetry group of W contains the permutation representation of $(\mathcal{X}, \boxplus)$ as a subgroup.

For a group code matched to a channel, the code and channel together have a property which is akin to the geometric uniformity defined by Forney [30]. Similar ideas were explored more recently for group codes over integer rings [31]. In particular, one gets the uniform error property where the error rate of the optimal decoder is independent of the transmitted codeword.

B. The Coset Channel

For this section, assume that $\mathcal{C} \subseteq \mathcal{X}^N$ is a group code over $(\mathcal{X}, \boxplus)$ and that for all $x \in \mathcal{X}$, there exists $\boldsymbol{c} \in \mathcal{C}$ such that $c_0 = x$. Under these assumptions, the cosets of the subgroup $\{\boldsymbol{c} \in \mathcal{C} \,|\, c_0 = 0\}$ partition the code $\mathcal{C}$ into q sets of equal size. We define the *coset channel* to be the channel from X_0 to $Y_{\sim 0}$ defined as follows:

1) Given the input $x_0 \in \mathcal{X}$, a codeword $\boldsymbol{X}$ is drawn uniformly from the coset $\{\boldsymbol{c} \in \mathcal{C} \,|\, c_0 = x_0\}$.
2) The output $Y_{\sim 0}$ is a memoryless observation of $X_{\sim 0}$ through the channel W.

From now on, we will use V to denote the coset channel and ψ to denote its canonical map. Composing V with ψ gives the standard coset channel V^{s}.

Lemma 8. *If the code is matched to the channel W and, for all $u \in \mathcal{X}$, there exists $\mathbf{c} \in \mathcal{C}$ with $c_0 = u$, then the coset channel V is group symmetric. Also, the output of the standard coset channel does not depend on which coset element is chosen, i.e., for an arbitrarily distributed input X_0, the output $\psi(Y_{\sim 0})$ is conditionally independent of $X_{\sim 0}$ given X_0.*

If the code has additional symmetries that are matched to the channel, then the symmetry group of the coset channel may be larger. For the following result, let G be the permutation group of the channel W, let H be the permutation representation of the group $(\mathcal{X}, \boxplus)$, and let the group of homogeneous alphabet relabelings that preserve the code be given by

$$F \coloneqq \big\{\sigma \in \mathbb{S}_q \,|\, \forall \boldsymbol{c} \in \mathcal{C}, \big(\sigma(c_0), \sigma(c_1), \ldots, \sigma(c_{N-1})\big) \in \mathcal{C}\big\}.$$

Since both F and G are subgroups of $\mathbb{S}_q$, their intersection $F' \coloneqq F \cap G$ is also a subgroup of $\mathbb{S}_q$.

Lemma 9. *If the code is matched to the channel W, then the symmetry group of the coset channel V contains the group $\langle F', H\rangle$ generated by F' and H.*

For the case where q is not prime, our proof technique requires that the coset channel has doubly transitive symmetry. In view of Lemma 9, a sufficient condition for this can be readily verified when q is a prime power and $\mathcal{C}$ is a linear code over $\mathcal{X} = \mathbb{F}_q$ (and hence a group code with respect to the additive group of $\mathbb{F}_q$). Furthermore, if the code contains the all ones codeword (i.e., $(1, \ldots, 1) \in \mathcal{C}$) then its group of homogeneous alphabet relabelings contains the affine group over $\mathbb{F}_q$, which is defined by $A_q \coloneqq \{\sigma_{a,b} \in \mathbb{S}_q \,|\, a \in \mathbb{F}_q \setminus \{0\}, b \in \mathbb{F}_q\}$ where $\sigma_{a,b}(x) = (a \cdot x) + b$ uses $\mathbb{F}_q$ addition and multiplication.

Lemma 10. *If $\mathcal{C} \subseteq \mathbb{F}_q^N$ is a linear code that contains the all ones codeword and the symmetry group of the channel W contains the affine group A_q, then the code is matched to the channel W and the symmetry group of the coset channel V is doubly transitive.*

Lastly, we recall that the permutation automorphism group of the code of a $\mathcal{C} \subseteq \mathcal{X}^N$ is the group of permutations $\pi \in \mathbb{S}_N$ such that $(c_{\pi(0)}, c_{\pi(1)}, \ldots, c_{\pi(N-1)}) \in \mathcal{C}$ for all $c \in \mathcal{C}$. If this group is transitive then the coset channel defined for symbol position 0 is Blackwell equivalent to the coset channel for any other position $i \in [N]$.

C. Generalized Reed–Muller Codes and Puncturing

The natural generalization of binary RM codes to non-binary alphabets was introduced by Kasami et al. in 1968 and dubbed Generalized Reed–Muller (GRM) codes [1]. The GRM code $\mathrm{RM}_q(r, m) \subseteq \mathbb{F}_q^N$ is a length $N = q^m$ linear code over $\mathbb{F}_q$. Like binary RM codes, GRM codes can be defined as polynomial evaluation codes.

The rate of $\mathrm{RM}_q(r, m)$, denoted by $R_q(r, m)$, is computed using the base-q logarithm. Thus, it equals the number of $\mathbb{F}_q$ information symbols (i.e., the dimension of the code) divided by the number of $\mathbb{F}_q$ codeword symbols.

Lemma 11. *For integers $q \geq 2$, $0 \leq r \leq m(q-1)$, and $0 \leq k < m - r$, the rates of $\mathrm{RM}_q(r, m-k)$ and $\mathrm{RM}_q(r, m)$ satisfy*

$$R_q(r, m-k) - R_q(r, m) \leq \frac{4k}{\sqrt{m-k}}.$$

Definition 4 (Punctured Code)**.** For a code $\mathcal{C} \subseteq \mathbb{F}_q^N$ and index set $I \subseteq [N]$, the punctured code formed by the symbol positions indexed by I is given by $\mathcal{C}_I \coloneqq \{\boldsymbol{c}_I \in \mathbb{F}_q^{|I|} \mid \boldsymbol{c} \in \mathcal{C}\}$.

Remark 1. Although one may also consider a puncturing operation that includes reordering of code symbols, this is not needed for our results. For our definition, the symbols are kept in the same order but their indices are renumbered.

Lemma 12 (GRM Puncturing)**.** *If one punctures the code $\mathcal{C} = \mathrm{RM}_q(r, m)$ by keeping only the first q^{m-k} symbol positions (i.e., giving $\mathcal{C}_I$ with $I = [q^{m-k}]$), then $\mathcal{C}_I = \mathrm{RM}_q(r, m-k)$. Moreover, puncturing a uniform random codeword from $\mathcal{C}$ gives a uniform random codeword from $\mathcal{C}_I$.*

IV. Main Results

A. SER of the Coset Channel

This section gives bounds on the SER of the coset channel (defined in Section III-B) under the following conditions:

Condition 1 (Code)**.** The input $\boldsymbol{X}$ is distributed uniformly of the codewords of a q-ary group code $\mathcal{C}$ that has code rate

R. The code has a transitive permutation automorphism group and, for each $x \in \mathcal{X}$, there exists $\boldsymbol{c} \in \mathcal{C}$ with $c_0 = x$.

Condition 2 (Channel)**.** The output $\boldsymbol{Y}$ is an observation of the input through a symmetric memoryless channel W that is matched to the code and has capacity C.

Under these conditions, X_0 is uniformly distributed, the coset channel V is group symmetric, and the SER of the coset channel is an upper bound on the maximal SER of the code:

$$\max_{i\in[N]} \mathrm{SER}(X_i \,|\, \boldsymbol{Y}) = \mathrm{SER}(X_0 \,|\, \boldsymbol{Y}) \leq \mathrm{SER}(X_0 \,|\, Y_{\sim 0})$$

For the purposes of analysis, we introduce a degraded family of channels that interpolates between W and an uninformative channel. Specifically, we define W_t to be the composition of W and an erasure channel with erasure probability $t \in [0, 1]$. The implied coset channel is denoted by V_t and its squared error discrepancy is given by

$$\delta(t) \coloneqq \mathbb{E}\big[\|\Psi(t) - \mathbb{E}[\Psi(t) \,|\, X_0]\|^2\big], \quad 0 \leq t \leq 1$$

where $\Psi(t) \coloneqq \psi_t(Y_{\sim 0}(t)) = \mathbb{E}[\boldsymbol{e}_{X_0} \,|\, Y_{\sim 0}(t)]$. Here, the second expression for $\Psi(t)$ holds because X_0 is uniformly distributed, and thus the canonical map is equal to the posterior pmf. Finally, the *average discrepancy* is defined to be

$$\delta_{\mathrm{avg}} \coloneqq \int_0^1 \delta(t)\, dt.$$

We begin with a lower bound on the overlap matrix of the coset channel. In combination with Lemma 5, this bound shows that the SER of the coset channel is strictly less than the trivial upper bound $1 - 1/q$ whenever the code rate R is strictly less than the capacity of the channel W_t.

Lemma 13. *Assume that Conditions 1 and 2 hold with $R < C$. For all $0 \leq t < 1 - R/C$, the overlap matrix $Q(t)$ of the coset channel V_t satisfies*

$$\mathrm{Tr}(Q(t)) \geq q^{C - R/(1-t)}.$$

Next, we combine Lemma 13 with the constraints on the overlap matrix in Lemma 7 to provide a stronger bound on the SER in terms of the average discrepancy.

Lemma 14. *Assume that Conditions 1 and 2 hold with $R < C$ and the average discrepancy satisfies*

$$\delta_{\mathrm{avg}} \leq \frac{(1 - R/C)(q^{\frac{1}{2}(C-R)} - 1)}{q}.$$

Further, suppose that one of the following conditions holds:

(i) the symmetry group of V is doubly transitive, or

(ii) the symmetry group of V is transitive, q is prime, and

$$\delta_{\mathrm{avg}} < \frac{1 - R/C}{8q^2}.$$

Then, the SER satisfies

$$\mathrm{SER}(X_0 \,|\, Y_{\sim 0}) \leq \frac{4\delta_{\mathrm{avg}}}{1 - R/C}.$$

Finally, we provide a link between the average discrepancy and the entropy rates of subsets of the code. The following result is obtained by combining a decomposition of the squared error discrepancy, via the Efron-Stein-Steele inequality, with Lemma 6 and the EXIT area theorem.

Lemma 15. *Assume Conditions 1 and 2 hold. Furthermore, assume that overlap matrix of W has minimal eigenvalue $\lambda_{\min} > 0$. Let $\mathcal{B}$ be a collection of subsets of $[N]$ such that*

(i) $\bigcap_{B\in\mathcal{B}} B = \{0\}$

(ii) For each $B \in \mathcal{B}$, the punctured code $\mathcal{C}_B$ has a transitive permutation automorphism group.

Then, we have

$$\delta_{\mathrm{avg}} \leq \frac{2\ln q}{\lambda_{\min}^2} \sum_{B\in\mathcal{B}} \left(\frac{H(X_B)}{|B|} - R \right).$$

B. Proof of Theorem 1

We begin by verifying the conditions used in Section IV-A. Since the GRM code is a linear code over $\mathbb{F}_q$, it is automatically a group code under the additive group of $\mathbb{F}_q$. Also, is is well known that GRM codes have doubly transitive permutation automorphism groups [1]. For a linear code, each code position either takes all possible values or is always 0. Thus, for all $x \in \mathcal{X}$, there is a $\boldsymbol{c} \in \mathcal{C}$ such that $c_0 = x$ because otherwise $\mathcal{C}$ would only contain the all zero codeword due to transitive symmetry and have rate zero. Together, these results imply Condition 1 is satisfied.

Next, we note that cases *(i)* and *(ii)* of Theorem 1 both require that G contains the additive group of $\mathbb{F}_q$. This implies that the channel W is group symmetric and the code is matched to the channel. Thus, Condition 2 is also satisfied.

Under these conditions, we can apply Lemma 8 to see that the coset channel is is group symmetric and, for case *(ii)*, we can apply Lemma 10 to see that it is doubly transitive. Having verified the assumptions of Lemma 14, we conclude that for any $\epsilon \in (0, 1]$ there exists $\delta^* > 0$, such that if $R_q(r, m) \leq (1 - \epsilon)C$ and $\delta_{\mathrm{avg}} \leq \delta^*$, then

$$\mathrm{SER}(X_0 \,|\, Y_{\sim 0}) \leq \frac{4\delta_{\mathrm{avg}}}{1 - R/C}.$$

The final step of the proof is to show that δ_{avg} converges to zero as $m \to \infty$. If $C = 0$, the theorem is vacuous, so we assume $C > 0$. If the channel symmetry group is doubly transitive, then $C > 0$ implies $\lambda_{\min} > 0$. Otherwise, q is prime and $\lambda_{\min} > 0$ by assumption. The desired convergence is established by the following result, which is obtained by combining the rate difference property of GRM codes in Lemma 11 with the generic bound in Lemma 15.

Lemma 16. *For a GRM code $\mathrm{RM}_q(r, m)$ with $m \geq q^2$ on a channel W whose overlap matrix has minimal eigenvalue $\lambda_{\min} > 0$, we find that*

$$\delta_{\mathrm{avg}} \leq \frac{2\ln q}{\lambda_{\min}^2} \left(\frac{7 + 3\log_q m}{\sqrt{m}} \right) = O\left(\frac{\ln m}{\sqrt{m}} \right).$$

We note that proofs delegated to the extended version [23] also utilize the following additional references [32]–[35].

REFERENCES

[1] T. Kasami, S. Lin, and W. W. Peterson, "New generalizations of the Reed-Muller codes–I: Primitive codes," *IEEE Trans. Inform. Theory*, vol. 14, pp. 189–199, Mar 1968.

[2] D. Muller, "Application of Boolean algebra to switching circuit design and to error detection," *IRE Tran. on Electronic Computers*, vol. EC-3, pp. 6–12, Sept 1954.

[3] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *IRE Tran. on Information Theory*, vol. 4, pp. 38–49, September 1954.

[4] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Indust. Math.*, vol. 8, no. 2, pp. 300–304, 1960.

[5] S. Kopparty, S. Saraf, and S. Yekhanin, "High-rate codes with sublinear-time decoding," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, pp. 167–176, 2011.

[6] A. Guo, S. Kopparty, and M. Sudan, "New affine-invariant codes from lifting," in *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pp. 529–540, 2013.

[7] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, 2016.

[8] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels," *IEEE Trans. Inform. Theory*, vol. 63, no. 7, pp. 4298–4316, 2017.

[9] O. Sberlo and A. Shpilka, "On the performance of Reed-Muller codes with respect to random errors and erasures," in *Proc. of the Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 1357–1376, SIAM, 2020.

[10] E. Abbe and M. Ye, "Reed-Muller codes polarize," *IEEE Trans. Inform. Theory*, vol. 66, no. 12, pp. 7311–7332, 2020.

[11] J. Hązła, A. Samorodnitsky, and O. Sberlo, "On codes decoding a constant fraction of errors on the BSC," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, pp. 1479–1488, 2021.

[12] E. Abbe, A. Shpilka, and M. Ye, "Reed–Muller codes: Theory and algorithms," *IEEE Trans. Inform. Theory*, vol. 67, no. 6, pp. 3251–3277, 2020.

[13] G. Reeves and H. D. Pfister, "Reed-Muller codes achieve capacity on BMS channels." [Online]. Available: https://arxiv.org/abs/2110.14631, 2021.

[14] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2657–2674, Nov. 2004.

[15] C. Méasson, A. Montanari, T. J. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4793–4821, Nov. 2009.

[16] D. Blackwell, "Equivalent comparisons of experiments," *The Annals of Mathematical Statistics*, vol. 24, no. 2, pp. 265–272, 1953.

[17] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*. Springer, 1986.

[18] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[19] F. d. P. Calmon, A. Makhdoumi, M. Médard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5011–5038, 2017.

[20] A. Makur and Y. Polyanskiy, "Comparison of channels: Criteria for domination by a symmetric channel," *IEEE Trans. Inform. Theory*, vol. 64, no. 8, pp. 5704–5725, 2018.

[21] M. Raginsky, "Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 1220–1224, 2011.

[22] M. Raginsky, "Logarithmic Sobolev inequalities and strong data processing theorems for discrete channels," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 419–423, 2013.

[23] G. Reeves and H. D. Pfister, "Achieving capacity on non-binary channels with generalized Reed–Muller codes." arXiv online preprint, 2023.

[24] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2576–2595, 2013.

[25] D. Slepian, "A class of binary signaling alphabets," *The Bell Syst. Techn. J.*, vol. 35, no. 1, pp. 203–234, 1956.

[26] M. L. Eaton, "Group invariance applications in statistics," *Regional Conference Series in Probability and Statistics*, 1989.

[27] D. Slepian, "Group codes for the Gaussian channel," *The Bell Syst. Techn. J.*, vol. 47, no. 4, pp. 575–602, 1968.

[28] S. Berman, "On the theory of group codes," *Cybernetics*, vol. 3, no. 1, pp. 25–31, 1967.

[29] A. Fontaine and W. Peterson, "Group code equivalence and optimum codes," *IRE Trans. Inform. Theory*, vol. 5, no. 5, pp. 60–70, 1959.

[30] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.

[31] G. Como and F. Fagnani, "The capacity of finite abelian group codes over symmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2037–2054, 2009.

[32] I. G. Shevtsova, "On the absolute constants in the Berry–Esseen inequality and its structural and nonuniform improvements," *Informatika i Ee Primeneniya [Informatics and its Applications]*, vol. 7, no. 1, pp. 124–125, 2013.

[33] J. S. Milne, "Group theory (v4.00)," 2021. Available at www.jmilne.org/math/.

[34] M. Artin, *Algebra*. Pearson Education, 2011.

[35] C. Méasson, A. Montanari, and R. L. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inform. Theory*, vol. 54, pp. 5277–5307, Dec. 2008.