Energy-minimizing error-correcting codes

Henry Cohn and Yufei Zhao

***Abstract*—We study a discrete model of repelling particles, and we show using linear programming bounds that many familiar families of error-correcting codes minimize a broad class of potential energies when compared with all other codes of the same size and block length. Examples of these universally optimal codes include Hamming, Golay, and Reed-Solomon codes, among many others, and this helps explain their robustness as the channel model varies. Universal optimality of these codes is equivalent to minimality of their binomial moments, which has been proved in many cases by Ashikhmin and Barg. We highlight connections with mathematical physics and the analogy between these results and previous work by Cohn and Kumar in the continuous setting, and we develop a framework for optimizing the linear programming bounds. Furthermore, we show that if these bounds prove a code is universally optimal, then the code remains universally optimal even if one codeword is removed.**

***Index Terms*—Error correction codes, Combinatorial mathematics.**

I. Introduction

Analogies between discrete and continuous packing problems have long played a key role in coding theory. In this paper, we extend these analogies to encompass discrete models of physics, by showing that certain classical codes are ground states of natural physics models. In fact, they are ground states of many different models simultaneously. We call this phenomenon *universal optimality*, motivated by [7].

As we will explain after Lemma 4, a code is universally optimal if and only if all the binomial moments of its distance distribution are minimal. This problem has been studied by Ashikhmin and Barg [1], with a very different combinatorial motivation (namely, counting pairs of codewords in subcodes with restricted support), and they gave some important examples, such as Hamming, Golay, and Reed-Solomon codes. Thus, universal optimality is not a new property. However, the physics motivation appears to be new, and we provide new proof techniques. We also prove strong structural results about these codes, including our most surprising theorem: if the linear programming bounds prove a code is universally optimal, then it remains universally optimal if any single codeword is removed.

Let $\mathbb{F}_q$ denote an alphabet with q elements, and let $|x-y|$ denote the Hamming distance between words $x, y \in \mathbb{F}_q^n$. Of course, this notation suggests that $\mathbb{F}_q$ is a finite field, but we will make no use of the field structure.

We view $\mathbb{F}_q^n$ as a discrete model of the universe, and we envision a code in $\mathbb{F}_q^n$ as specifying the locations of some particles. To separate these particles from each other, we will let them repel each other. Specifically, we will choose a pairwise potential function between the particles, and then we will study the *ground states* of this system, i.e., the particle arrangements that minimize the total energy.

Given a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ and a function $f\colon \{1, 2, \dots, n\} \to \mathbb{R}$, the *potential energy* of $\mathcal{C}$ with respect to the *potential function* f is defined to be

$$E_f(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{\substack{x,y \in \mathcal{C} \\ x \neq y}} f\left(|x-y|\right).$$

The normalization factor of $1/|\mathcal{C}|$ is convenient but not essential.

Repulsive forces correspond to decreasing potential functions, and we wish the repulsion to grow stronger as the points grow closer together. The completely monotonic functions extend these properties in a particularly compelling way. Let Δ be the finite difference operator, defined by $\Delta f(n) = f(n+1) - f(n)$. A function $f\colon \{a, a+1, \dots, b\} \to \mathbb{R}$ is *completely monotonic* if its iterated differences alternate in sign via $(-1)^k \Delta^k f \geq 0$; more precisely, $(-1)^k \Delta^k f(i) \geq 0$ whenever $k \geq 0$ and $a \leq i \leq b-k$.

For example, the inverse power laws $f(r) = r^{-\alpha}$ with $\alpha > 0$ are completely monotonic. To see why, note that their derivatives obviously alternate in sign, and then the mean value theorem implies that the same is true for finite differences.

Definition 1. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is *universally optimal* if

$$E_f(\mathcal{C}) \leq E_f(\mathcal{C}')$$

for every $\mathcal{C}' \subseteq \mathbb{F}_q^n$ with $|\mathcal{C}'| = |\mathcal{C}|$ and all completely monotonic $f\colon \{1, \dots, n\} \to \mathbb{R}$.

Every universally optimal code $\mathcal{C}$ maximizes the minimal distance between codewords given its size $|\mathcal{C}|$, because it minimizes the energy under $r \mapsto r^{-\alpha}$ as $\alpha \to \infty$. However, universal optimality is a far stronger condition than that.

The definition of universal optimality is analogous to that of Cohn and Kumar [7] in the continuous setting. They studied particle arrangements in spheres or projective spaces and showed that many beautiful configurations are universally optimal, including the icosahedron, the E_8 root system, and the minimal vectors in the Leech lattice. More generally, universal optimality helps explain the occurrence of certain remarkable symmetry groups in discrete mathematics and physics [6].

Bouman, Draisma, and van Leeuwaarden [5] have independently studied energy minimization models on toric grids under the Lee metric. Their main theorem implies universal optimality for certain checkerboard arrangements of particles filling half of the grid, but they do not investigate other codes.

Universally optimal codes have robust energy minimization properties, which translate into good performance according

Henry Cohn is with Microsoft Research New England, One Memorial Drive, Cambridge, MA 02142, USA (e-mail: cohn@microsoft.com).

Yufei Zhao is with the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: yufeiz@mit.edu).

Zhao was supported by an internship at Microsoft Research New England.

to a broad range of measures. For example, they minimize the probability of an undetected error under the q-ary symmetric channel, provided that each symbol is more likely to remain the same than to become any other fixed symbol. (See Section V of [1].)

For another application, consider maximum-likelihood decoding for a binary-input discrete memoryless channel. The exact error probability for decoding is subtle, but for relatively low-rate codes it is frequently estimated using a union bound (see Theorem 7.5 in [17, p. 153]). This bound shows that the error probability for a random codeword from a code $\mathcal{C}$ is at most $E_f(\mathcal{C})$ with $f(r) = \gamma^r$, where γ is the Bhattacharyya parameter. Because $\gamma \le 1$, the potential function f is completely monotonic, and thus a universally optimal code must minimize this upper bound for the decoding error. It does not necessarily minimize the true decoding error [14], but minimizing a useful upper bound is nearly as good.

Optimality is by no means limited to this particular union bound. For example, the same holds true for the AWGN channel with antipodal signaling. (Verifying complete monotonicity for the potential function requires a brief inductive proof, but it is not difficult.) This explains the observations of Ferrari and Chugg [12], who used linear programming bounds to verify that certain Hamming and Golay codes minimize this bound for a wide range of signal-to-noise ratios. Our results prove that this always works and show how to generalize it to other codes.

We will prove that all the codes listed in Table I are universally optimal. (See the longer version arXiv:1212.1913v1 of this paper for a review of the definitions of these codes, as well as other background and discussion removed for lack of space.) For the Hamming, Hadamard, Golay, MDS, and Nordstrom-Robinson codes, universal optimality is a theorem of Ashikhmin and Barg [1], as mentioned above.

Universally optimal codes are common for short block lengths, but they become increasingly rare for long block lengths. Brute force searches show that there is a unique universally optimal binary code of size N and block length n (up to translation and permutation of the coordinates) whenever $n \le 4$ and $1 \le N \le 2^n$. For $n = 5$, such a code exists if and only if $N \notin \{9, 12, 13, 14, 18, 19, 20, 23\}$, and it is unique except when $N = 5$ or $N = 27$, in which case there are two isomorphism classes (see Section VII for an explanation of the $N \leftrightarrow 32 - N$ symmetry). Thus, a universal optimum need not exist or be unique if it does exist.

Our main technical tool for bounding energy is the linear program developed by Delsarte [9], which was originally used to bound the size of codes given their minimum distance and was applied to energy minimization and related problems by Yudin [20] and by Ashikhmin, Barg, and Litsyn [1], [2]. We will call a code *LP universally optimal* if its universal optimality follows from these bounds, as occurs for all the cases in Table I.

Our most surprising theorem is that LP universally optimal codes continue to minimize energy even after we remove a single codeword. We know of no continuous analogue of this property. Furthermore, such codes are distance regular (for each distance, every codeword has the same number of codewords at that distance).

Theorem 2. *Every LP universally optimal code is distance regular, and it remains universally optimal when any single codeword is removed.*

Removing a codeword yields a universal optimum, but the resulting code will generally not be LP universally optimal. Thus, this process cannot be iterated.

II. Linear programming bounds

We begin by formulating the linear programming bound for energy minimization. Suppose $\mathcal{C} \subseteq \mathbb{F}_q^n$. The Delsarte inequalities constrain the *distance distribution* $(A_0, A_1, \dots, A_n)$ of $\mathcal{C}$, where

$$A_i = \frac{1}{|\mathcal{C}|} \left|\left\{(x, y) \in \mathcal{C}^2 : |x - y| = i\right\}\right| \quad \text{for } i = 0, 1, \dots, n.$$

Specifically, let K_k denote the k-th Krawtchouk polynomial, defined by

$$\begin{aligned} K_k(x) &= K_k(x; n, q) \\ &= \sum_{j=0}^{k} (-1)^j (q-1)^{k-j} \binom{x}{j} \binom{n-x}{k-j}. \end{aligned} \tag{1}$$

Krawtchouk polynomials are orthogonal polynomials with respect to the binomial distribution $\mathrm{Binom}(n; (q-1)/q)$. In other words,

$$\frac{1}{q^n} \sum_{i=0}^{n} \binom{n}{i} (q-1)^i K_j(i) K_k(i) = \binom{n}{j} (q-1)^j \delta_{jk}. \tag{2}$$

The Delsarte inequalities are

$$\sum_{i=0}^{n} A_i K_j(i) \ge 0$$

for $j = 0, 1, \dots, n$ (see Theorem 3 in [11]). Thus, the following linear program in the variables $A_0, A_1, \dots, A_n$ gives a lower bound for $E_f(\mathcal{C})$ when $|\mathcal{C}| = N$:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} A_i f(i) \\ \text{subject to} \quad & \sum_{i=0}^{n} A_i K_j(i) \ge 0 \qquad \text{for } j = 1, 2, \dots, n, \\ & A_0 + A_1 + \cdots + A_n = N, \\ & A_0 = 1, \\ & A_i \ge 0 \qquad \text{for } i = 1, 2, \dots, n. \end{aligned} \tag{3}$$

Definition 3. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is *LP universally optimal* if its distance distribution $(A_0, \dots, A_n)$ optimizes (3) for every completely monotonic potential function f.

Ashikhmin and Barg [1] call LP universally optimal codes *extremal codes*.

For any fixed code, checking whether it is universally optimal or LP universally optimal is a finite problem, since we can write down a basis for the cone of completely monotonic functions as follows.

TABLE I
LP UNIVERSALLY OPTIMAL CODES.

We write $a \to b$ to mean $a, a+1, a+2, \ldots, b$ and $a \xrightarrow{2} b$ to mean $a, a+2, a+4, \ldots, b$. The justification numbers in square brackets tell which lemmas or propositions imply LP universal optimality; when there is no such number, the proof is by directly solving the linear programs.

Name [justification]	q	n	N	Support $\subseteq$	Dual support $\subseteq$
Binary Hamming [9, 24]	2	2^r-1	2^{2^r-r-1}	$\{3 \to n-3, n\}$	$\left\{\frac{n+1}{2}\right\}$
– extended [9, 22]	2	2^r	2^{2^r-r-1}	$\left\{4 \xrightarrow{2} n-4\right\}$	$\left\{\frac{n}{2}, n\right\}$
– even subcode [9, 22]	2	2^r-1	2^{2^r-r-2}	$\left\{4 \xrightarrow{2} n-3\right\}$	$\left\{\frac{n-1}{2}, \frac{n+1}{2}, n\right\}$
– shortened [9, 22]	2	2^r-2	2^{2^r-r-2}	$\{3 \to n-2\}$	$\left\{\frac{n}{2}, \frac{n}{2}+1\right\}$
– 2× shortened [9, 25]	2	2^r-3	2^{2^r-r-3}	$\{3 \to n-1\}$	$\left\{\frac{n-1}{2}, \frac{n+1}{2}, \frac{n+3}{2}\right\}$
– punctured [9, 24]	2	2^r-2	2^{2^r-r-1}	$\{2 \to n-2, n\}$	$\left\{\frac{n}{2}+1\right\}$
q-ary Hamming [9, 24]	q	$\frac{q^r-1}{q-1}$	q^{n-r}	$\{3 \to n\}$	$\left\{q^{r-1}\right\}$
– shortened [9, 24]	q	$\frac{q^r-q}{q-1}$	q^{n-r}	$\{3 \to n\}$	$\left\{q^{r-1}-1, q^{r-1}\right\}$
– punctured [9, 24]	q	$\frac{q^r-q}{q-1}$	q^{n-r+1}	$\{2 \to n\}$	$\left\{q^{r-1}\right\}$
Simplex (1-design) [24]	q	n	N	$\{a\}$	$\{2 \to n\}$
– punctured [24]	q	$n-1$	N	$\{a-1, a\}$	$\{2 \to n-1\}$
Hadamard [22]	2	$4k$	$2n$	$\left\{\frac{n}{2}, n\right\}$	$\left\{4 \xrightarrow{2} n-4\right\}$
– punctured [22]	2	$4k-1$	$2n+2$	$\left\{\frac{n-1}{2}, \frac{n+1}{2}, n\right\}$	$\left\{4 \xrightarrow{2} n-3\right\}$
Conference [22]	2	$4k+1$	$2n+2$	$\left\{\frac{n-1}{2}, \frac{n+1}{2}, n\right\}$	$\left\{4 \xrightarrow{2} n-1\right\}$
Binary Golay [9, 23]	2	23	2^{12}	{7,8,11,12,15,16,23}	{8,12,16}
– extended [23]	2	24	2^{12}	{8,12,16,24}	{8,12,16,24}
– punctured [9, 23]	2	22	2^{12}	{6→16,22}\{9,13}	{8,12,16}
– shortened	2	22	2^{11}	{7,8,11,12,15,16}	{7,8,11,12,15,16}
– 2× shortened	2	21	2^{10}	{7,8,11,12,15,16}	{6→16}\{9,13}
– punctured and 2× shortened	2	20	2^{10}	{6→16}\{9,13}	{6→16}\{9,13}
Ternary Golay [9, 23]	3	11	3^6	{5,6,8,9,11}	{6,9}
– extended [23]	3	12	3^6	{6,9,12}	{6,9,12}
– shortened	3	10	3^5	{5,6,8,9}	{5,6,8,9}
– 2× shortened [22]	3	9	3^4	{5,6,8,9}	{4→9}
– 3× shortened [22]	3	8	3^3	{5,6,8}	{3→8}
– 4× shortened [24]	3	7	3^2	{5,6}	{2→7}
– punctured [9, 23]	3	10	3^6	{4→10}	{6,9}
– 2× punctured [9, 23]	3	9	3^6	{3→9}	{6,9}
– 3× punctured [9, 24]	3	8	3^6	{2→8}	{6}
MDS [16, 21]	q	n	q^{n-d+1}	$\{d \to n\}$	$\{n-d+2 \to n\}$
Ovoid ($q > 2$) [22]	q	q^2+1	q^4	$\left\{q^2-q, q^2\right\}$	$\{4 \to n\}$
– shortened [22]	q	q^2	q^3	$\left\{q^2-q, q^2\right\}$	$\{3 \to n\}$
– 2× shortened [24]	q	q^2-1	q^2	$\left\{q^2-q\right\}$	$\{2 \to n\}$
– punctured [22]	q	q^2	q^4	$\{q^2-q-1, q^2-q, q^2-1, q^2\}$	$\{4 \to n\}$
Nordstrom-Robinson	2	16	256	{6,8,10,16}	{6,8,10,16}
– punctured	2	15	256	{5→10,15}	{6,8,10}
– shortened	2	15	128	{6,8,10}	{5→10,15}
– 2× shortened	2	14	56	{6,8,10}	{4→10,14}

Lemma 4. *The complete monotonic functions on* $\{0, 1, \ldots, n\}$ *are the nonnegative span of the* fundamental potential functions $f_0, f_1, \ldots, f_n$ *defined by* $f_j(x) = \binom{n-x}{j}$.

The potential energy with respect to f_j is exactly the j-th *binomial moment* of the distance distribution, as defined by Ashikhmin and Barg [1]. Thus, Lemma 4 shows that a code is universally optimal if and only if its binomial moments are minimal, so the results of [1] can be restated in terms of universal optimality.

Lemma 4 includes 0 in the domain of f, which will be notationally convenient in Section III, but we can always extend f from $\{1, 2, \ldots, n\}$ to $\{0, 1, \ldots, n\}$ by setting $f(0)$ to be a sufficiently large value that complete monotonicity continues to hold.

We say that a function $f\colon \{a, a+1, \ldots, b\} \to \mathbb{R}$ is *absolutely monotonic* if all its finite differences are nonnegative; i.e., $\Delta^k f(i) \geq 0$ whenever $k \geq 0$ and $a \leq i \leq b-k$.

Proof of Lemma 4. By changing x to $n-x$, it suffices to prove that the functions $g_j(x) = \binom{x}{j}$ span the cone of absolutely monotonic functions. Indeed, $\Delta^r g_j(x) = g_{j-r}(x)$ for $r \leq j$, and $\Delta^r g_j(x) = 0$ for $r > j$, so each g_j is absolutely monotonic. Conversely, every function $g\colon \{0, 1, \ldots, n\} \to \mathbb{R}$

satisfies

$$g(x) = \sum_{j=0}^{n} \binom{x}{j} \Delta^j g(0) = \sum_{j=0}^{n} g_j(x) \Delta^j g(0)$$

by the discrete calculus analogue of the Taylor series expansion. If g is absolutely monotonic, then $\Delta^j g(0) \geq 0$ for all j, as desired. □

Thus, checking whether a code of block length n is LP universally optimal amounts to solving n linear programs (the f_0 case is trivial). However, checking whether a code is universally optimal seems far more difficult.

Linear programming duality transforms (3) into its dual as follows. Here $c_0, \ldots, c_n$ are the dual variables, and the equality conditions follow from complementary slackness.

Proposition 5. *Suppose $f\colon \{1, \ldots, n\} \to \mathbb{R}$ is any function, $h\colon \{0, 1, \ldots, n\} \to \mathbb{R}$ satisfies*

$$h(i) \leq f(i) \qquad \text{for } i = 1, 2, \ldots, n,$$

and there exist $c_0, c_1, \ldots, c_n$ with $c_j \geq 0$ for $j \geq 1$ such that

$$h(i) = \sum_{j=0}^{n} c_j K_j(i) \qquad \text{for } i = 0, 1, \ldots, n.$$

Then every code $\mathcal{C} \subseteq \mathbb{F}_q^n$ with $|\mathcal{C}| = N$ has f-potential energy at least $Nc_0 - h(0)$. Furthermore, equality holds if and only if $h(i) = f(i)$ for all $i > 0$ satisfying $A_i > 0$ and $c_j = 0$ for all $j > 0$ satisfying $A_j^\perp > 0$, where (A_i) is the distance distribution of $\mathcal{C}$ and $(A_j^\perp)$ is the dual distance distribution *defined by*

$$A_j^\perp = \frac{1}{|\mathcal{C}|} \sum_{i=0}^{n} A_i K_j(i). \tag{4}$$

III. Quasicodes and duality

In this section we show that LP universal optimality is preserved under the duality operation expressed by (4). Curiously, this symmetry seems to have no analogue in the continuous setting of [7].

We use the term *quasicode* for a feasible point in the Delsarte linear program, equipped with a duality operator called the MacWilliams transform [16, p. 137].

Definition 6. A *quasicode* $\mathbf{a}$ of length n and size N over $\mathbb{F}_q$ is a real column vector $(A_0, A_1, \ldots, A_n)$ satisfying the constraints of the linear program (3). In other words,

$$\mathbf{a} \geq 0, \qquad K\mathbf{a} \geq 0, \qquad \sum_{i=0}^{n} A_i = N, \qquad \text{and} \qquad A_0 = 1.$$

Here K stands for the matrix $(K_i(j))_{0 \leq i,j \leq n}$, and $\mathbf{a} \geq 0$ means that all coordinates of $\mathbf{a}$ are nonnegative. We write $|\mathbf{a}|$ for the *size* N of the quasicode. Based on (4), the *dual* of $\mathbf{a}$ is defined to be the quasicode

$$\mathbf{a}^\perp = \frac{1}{|\mathbf{a}|} K\mathbf{a}.$$

To see that $\mathbf{a}^\perp$ is a quasicode, we can use the identity $K^2 = q^n I$ (see (11), (12), and (17) in [11]). (The reason is that K is the radial Fourier transform and $K^2 = q^n I$ is Fourier inversion.) It follows that $\left|\mathbf{a}^\perp\right| |\mathbf{a}| = q^n$ and $\mathbf{a}^{\perp\perp} = \mathbf{a}$. For every code $\mathcal{C} \subseteq \mathbb{F}_q^n$, its distance distribution is a quasicode $\mathbf{a}$ with $|\mathbf{a}| = |\mathcal{C}|$. Furthermore, if $\mathcal{C}$ is a linear code, then its dual linear code $\mathcal{C}^\perp$ has distance distribution $\mathbf{a}^\perp$.

We say that $\mathbf{a}$ is a *t-design* if its dual $\mathbf{a}^\perp$ satisfies $A_j^\perp = 0$ for $1 \leq j \leq t$. Using Krawtchouk polynomials as a basis for polynomials of degree at most t, one can check that $\mathbf{a}$ is an t-design if and only if every polynomial f of degree at most t satisfies

$$\frac{1}{N} \sum_{i=0}^{n} A_i f(i) = \frac{1}{q^n} \sum_{i=0}^{n} \binom{n}{i} (q-1)^i f(i). \tag{5}$$

Given a potential function $f\colon \{0, 1, \ldots, n\} \to \mathbb{R}$, let $\mathbf{f}$ be the column vector $(f(0), f(1), \ldots, f(n))$. Minimizing the f-potential energy of a quasicode $\mathbf{a}$ amounts to minimizing the inner product

$$\mathbf{f}^t \mathbf{a} = \sum_{i=0}^{n} f(i) A_i.$$

This quantity differs from the earlier definition of energy by including $f(0)$, but it does not affect the notion of universal optimality since $A_0 = 1$, independently of $\mathbf{a}$.

Definition 7. A quasicode $\mathbf{a}$ of length n over $\mathbb{F}_q$ *minimizes f-potential energy* if $\mathbf{f}^t \mathbf{a} \leq \mathbf{f}^t \mathbf{b}$ for every quasicode $\mathbf{b}$ of length n over $\mathbb{F}_q$ with $|\mathbf{a}| = |\mathbf{b}|$. It is a *universally optimal quasicode* if it minimizes f-energy for every completely monotonic f.

Note that a code is LP universally optimal if and only if its distance distribution is a universally optimal quasicode. Universally optimal quasicodes often exist in low dimensions; for example, they exist for all $n \leq 11$ and $1 \leq N \leq 2^n$. Nevertheless, they do not always exist. For example, there are no universally optimal quasicodes for $n = 12$, $q = 2$, and $24 < N < 40$.

Given a quasicode $(A_0, \ldots, A_n)$ with dual $(A_0^\perp, \ldots, A_n^\perp)$, we call $\{i > 0 : A_i \neq 0\}$ the *support* of the quasicode, and $\left\{i > 0 : A_i^\perp \neq 0\right\}$ the *dual support* of the quasicode. Of course we apply the same definitions to actual codes.

Proposition 5 also applies to quasicodes, because its proof used only the Delsarte inequalities. Note that the conditions for equality do not take into account the actual values of the quasicode, but only its support and dual support:

Proposition 8. *Whether a quasicode is universally optimal depends only on its length, size, support, and dual support.*

A universally optimal quasicode is uniquely determined by its length and size if it exists, because the energies with respect to the $n + 1$ fundamental potential functions put $n + 1$ constraints on the quasicode, which are linearly independent because there is one potential function of each degree. Furthermore, if $\mathbf{a}$ is a universally optimal quasicode and $\mathbf{b}$ is another quasicode of the same length and size whose support and dual support are respectively contained in those of $\mathbf{a}$, then $\mathbf{b}$ is also universally optimal, since the same h that works for $\mathbf{a}$ also works for $\mathbf{b}$. Thus, $\mathbf{b} = \mathbf{a}$.

Proposition 9. *Let $\mathbf{a}$ be a quasicode. Then $\mathbf{a}$ is universally optimal if and only if its dual $\mathbf{a}^\perp$ is.*

For example, the dual of an LP universally optimal linear code is LP universally optimal. We first show that complete monotonicity is preserved under duality.

Lemma 10. *If* $\mathbf{f}$ *represents a completely monotonic function, then so does* $K^t\mathbf{f}$.

Proof. Recall from Lemma 4 that the functions $f_j(x) = \binom{n-x}{j}$ form a basis for the cone of completely monotonic functions. Let $\mathbf{f}_j$ denote the column vector corresponding to f_j. To see that K^t leaves the cone of completely monotonic functions invariant, we will use the identity

$$K^t\mathbf{f}_j = q^{n-j}\mathbf{f}_{n-j}. \tag{6}$$

Note that it can be rewritten as

$$\sum_{k=0}^{n} \binom{n-k}{j} K_k(i) = q^{n-j}\binom{n-i}{n-j}. \tag{7}$$

We use the following generating function for Krawtchouk polynomials [16, p. 151]:

$$\sum_{k=0}^{n} K_k(i)z^k = (1+(q-1)z)^{n-i}(1-z)^i.$$

By setting $z = (1+w)^{-1}$ we can rewrite it as

$$\sum_{k=0}^{n} K_k(i)(w+1)^{n-k} = (w+q)^{n-i}w^i.$$

Then (7) follows from comparing the coefficients of w^j in the above formula. □

Proof of Proposition 9. Since the duality operator is an involution, it suffices to prove that if $\mathbf{a}$ is universally optimal, then so is $\mathbf{a}^\perp$. Every quasicode can be written as $\mathbf{b}^\perp$ for some quasicode $\mathbf{b}$. So it suffices to show that $\mathbf{f}^t\mathbf{a}^\perp \le \mathbf{f}^t\mathbf{b}^\perp$ for every completely monotonic potential $\mathbf{f}$ whenever $|\mathbf{a}| = |\mathbf{b}|$. By Lemma 10, $K^t\mathbf{f}$ is also completely monotonic, and by the universal optimality of $\mathbf{a}$ we have

$$\begin{aligned}|\mathbf{a}|\,\mathbf{f}^t\mathbf{a}^\perp = \mathbf{f}^tK\mathbf{a} &= (K^t\mathbf{f})^t\mathbf{a}\\ &\le (K^t\mathbf{f})^t\mathbf{b} = \mathbf{f}^tK\mathbf{b} = |\mathbf{b}|\,\mathbf{f}^t\mathbf{b}^\perp.\end{aligned}$$

Therefore $\mathbf{a}^\perp$ is universally optimal. □

IV. Constructing dual solutions

To construct auxiliary functions h for use in Proposition 5, we will use polynomial interpolation. In this section, we first review the theory of positive definite functions, and then we prove inequalities on the values of interpolating polynomials.

A. Positive definite functions

For every function $h\colon \{0,1,\dots,n\} \to \mathbb{R}$, we can find $c_0, c_1, \dots, c_n$ such that

$$h(i) = \sum_{j=0}^{n} c_jK_j(i) \qquad \text{for } i = 0,1,\dots,n. \tag{8}$$

Specifically, if $\mathbf{h}$ is the column vector with entries $(h(i))_{0\le i\le n}$, then $\mathbf{h}^t = \mathbf{c}^tK$, so that $q^n\mathbf{c}^t = \mathbf{h}^tK$ as $K^2 = q^nI$. Call c_j the *Krawtchouk coefficients* of h. For any $0 \le s \le n$, the Krawtchouk polynomials $K_0, K_1, \dots, K_s$ span the polynomials of degree at most s, so if h is given by a polynomial of degree s, then $c_j = 0$ for $j > s$.

Now we consider the requirement $c_j \ge 0$ from Proposition 5.

Definition 11. A function $h\colon \{0,1,\dots,n\} \to \mathbb{R}$ is *positive definite* if its Krawtchouk coefficients are nonnegative.

Such functions are called "positive definite" because they are the functions for which $(h(|x-y|))_{x,y\in\mathbb{F}_q^n}$ is a positive semidefinite matrix (see Theorem 2 in [11]).

Proposition 5 does not actually require $c_0 \ge 0$. However, there seems to be little harm in assuming it. Doing so allows us to use properties of positive definite functions such as the following standard lemma, which follows from (24) in [11].

Lemma 12. *The product of two positive definite functions is positive definite.*

Lemma 13. *The function* $h(x) = a - x$ *is positive definite iff* $a \ge (q-1)n/q$.

Proof. This assertion follows immediately from

$$K_1(x) = (q-1)n - qx. \qquad \square$$

Corollary 14. *If* $a_1, a_2, \dots, a_s \ge (q-1)n/q$, *then* $h(x) = (a_1 - x)(a_2 - x)\cdots(a_s - x)$ *is positive definite.*

Lemma 15. *Let* $\mathbf{a} = (A_0, \dots, A_n)$ *be a quasicode whose support consists of* $a_1 < a_2 < \cdots < a_s$ *and suppose that* $\mathbf{a}$ *is a* $(2s-1)$*-design. Then* $h(x) = (a_1 - x)(a_2 - x)\cdots(a_s - x)$ *is positive definite.*

Proof. Let c_j be as in (8). Since h is a degree s polynomial, $c_j = 0$ for $j > s$. To show that $c_s > 0$, all we need to check is that the leading coefficient of K_s has sign $(-1)^s$, which is in fact true for each term in (1). Now, for $j \le s-1$, using the fact that $\mathbf{a}$ is a $(2s-1)$-design and $h\cdot K_j$ is a polynomial of degree at most $2s-1$, we have by the orthogonality (2) of the Krawtchouk polynomials and (5) that

$$\begin{aligned}(q-1)^j\binom{n}{j}c_j &= q^{-n}\sum_{i=0}^{n}\binom{n}{i}(q-1)^ih(i)K_j(i)\\ &= \frac{1}{N}\sum_{i=0}^{n}A_ih(i)K_j(i).\end{aligned}$$

The right side is nonnegative since $A_ih(i) = 0$ for $i \ge 1$ (because h vanishes on the support of $\mathbf{a}$) and $A_0h(0)K_j(0) \ge 0$. □

Lemma 16. *For* $0 \le j \le n$, *the function* $h(x) = (n-j+1-x)(n-j+2-x)\cdots(n-x)$ *is positive definite.*

Proof. We have $h(x) = j!f_j(x)$, where f_j is the fundamental potential function from Lemma 10. So $\mathbf{c}^t = q^{-n}\mathbf{h}^tK = q^{-n}j!\mathbf{f}_j^tK = q^{-j}j!\mathbf{f}_{n-j}^t \ge 0$ by (6). □

B. Polynomial interpolation

We begin with an analogue of Rolle's theorem.

Lemma 17. *Let $a < b$ be integers. If*

$$g\colon \{a, a+1, \dots, b+1\} \to \mathbb{R}$$

satisfies $g(a)g(a+1) \le 0$ and $g(b)g(b+1) \le 0$, then $\Delta g(c)\Delta g(c+1) \le 0$ for some $a \le c < b$.

Proof. Without loss of generality, we assume $g(a) \le 0$ and $g(a+1) \ge 0$. Since at least one of $g(b) \le 0$ and $g(b+1) \le 0$ is true, the sequence $g(a+1), g(a+2), \dots, g(b+1)$ cannot be strictly increasing. If c is the smallest integer such that $g(c+1) \ge g(c+2)$, then $\Delta g(c) > 0$ and $\Delta g(c+1) \le 0$, as desired. □

Lemma 18. *Let $a_1 < a_2 < \cdots < a_r$ be integers. If a function*

$$g\colon \{a_1, a_1+1, a_1+2, \dots, a_r+1\} \to \mathbb{R}$$

satisfies $g(a_i)g(a_{i+1}) \le 0$ for $i = 1, 2, \dots, r$, then there is some integer c such that $a_1 \le c \le a_r - r + 1$ and $\Delta^{r-1} g(c)\Delta^{r-1} g(c+1) \le 0$.

Proof. This follows from repeatedly applying Lemma 17. □

Lemma 19. *Let $f\colon \{0, 1, \dots, n\} \to \mathbb{R}$ be completely monotonic, let $a_1, \dots, a_r \in \{0, 1, \dots, n\}$ be distinct, and let p be the unique polynomial of degree less than r such that $p(a_i) = f(a_i)$ for $i = 1, 2, \dots, r$. Then*

$$(f(x) - p(x)) \prod_{i=1}^{r} (a_i - x) \ge 0 \tag{9}$$

for all $x = 0, 1, \dots, n$, and p has the expansion

$$p(x) = \sum_{j=0}^{r-1} c_j \prod_{i=1}^{j} (a_i - x) \tag{10}$$

with $c_0, \dots, c_{r-1} \ge 0$.

Proof. For (9), suppose $x \notin \{a_1, \dots, a_r\}$, since otherwise the inequality is trivial, and define $g\colon \{0, 1, \dots, n\} \to \mathbb{R}$ by

$$g(t) = f(t) - p(t) - A(t - a_1)(t - a_2) \cdots (t - a_r) \tag{11}$$

with the constant A chosen so that $g(x) = 0$; in other words,

$$A = \frac{f(x) - p(x)}{\prod_{i=1}^{r} (x - a_i)}.$$

We have $g(a_i) = 0$ for $i = 1, 2, \dots, r$ as well as $g(x) = 0$, so Lemma 18 implies that there is some integer c such that $\Delta^r g(c)\Delta^r g(c+1) \le 0$. Thus, $(-1)^r \Delta^r g(c') \le 0$ for either $c' = c$ or $c' = c + 1$. Now, (11) implies

$$\Delta^r g(c') = \Delta^r f(c') - Ar!$$

and we have $(-1)^r \Delta^r f(c') \ge 0$ by complete monotonicity, so $(-1)^r A \ge 0$. Therefore,

$$(f(x) - p(x)) \prod_{i=1}^{r} (a_i - x) = (-1)^r A \prod_{i=1}^{r} (x - a_i)^2 \ge 0.$$

For (10), we solve for $c_0, \dots, c_{r-1}$ successively starting with $c_0 = p(a_1) = f(a_1) \ge 0$. Now for each ℓ, the polynomial p_ℓ defined by

$$p_\ell(x) = \sum_{j=0}^{\ell-1} c_j \prod_{i=1}^{j} (a_i - x)$$

is the unique polynomial of degree less than ℓ satisfying $p_\ell(a_i) = f(a_i)$ for $i = 1, 2, \dots, \ell$. Applying (9) to p_ℓ, we find that for $1 \le \ell \le r - 1$,

$$\begin{aligned} 0 &\le (f(a_{\ell+1}) - p_\ell(a_{\ell+1})) \prod_{i=1}^{\ell} (a_i - a_{\ell+1}) \\ &= (p_{\ell+1}(a_{\ell+1}) - p_\ell(a_{\ell+1})) \prod_{i=1}^{\ell} (a_i - a_{\ell+1}) \\ &= c_\ell \prod_{i=1}^{\ell} (a_i - a_{\ell+1})^2. \end{aligned}$$

It follows that $c_\ell \ge 0$, as desired. □

V. Criteria for universal optimality

We now use the inequalities from Section IV to construct auxiliary functions for use in Proposition 5. These results are applied in Table I as indicated by the lemma or proposition numbers in square brackets in each line of the table. For the lines without references, we must resort to solving linear programs directly.

Recall that we do not include zero in the support of a quasicode.

Definition 20. Given a quasicode $\mathbf{a}$ of length n over $\mathbb{F}_q$, a *pair covering* is a subset $T \subseteq \{1, 2, \dots, n\}$ with elements $b_1 < b_2 < \cdots < b_t$ containing the support of $\mathbf{a}$ and such that $b_{2i-1} + 1 = b_{2i}$ whenever $2i \le t$, while $b_t = n$ if t is odd.

Proposition 21. *Let $\mathbf{a}$ be a quasicode of length n and T a pair covering of $\mathbf{a}$ with elements $b_1 < b_2 < \cdots < b_t$. Then $\mathbf{a}$ is universally optimal if the following two hypotheses are satisfied:*

(a) *The quasicode $\mathbf{a}$ is a $(t-1)$-design.*
(b) *For $1 \le j \le t-1$, the function $q_j(x) = \prod_{i=0}^{j-1} (b_{t-i} - x)$ is positive definite.*

We conjecture that condition (a) alone suffices.

Proof. Let $f\colon \{0, 1, \dots, n\} \to \mathbb{R}$ be completely monotonic, and let h be the unique polynomial of degree less than t such that $h(x) = f(x)$ for all $x \in T$. We will show that h satisfies the hypotheses of Proposition 5 and the conditions for equality.

For the inequality $f(x) \ge h(x)$, we apply (9) with $a_i = b_{t+1-i}$ and use the fact that $\prod_{i=1}^{t} (b_i - x) \ge 0$ for all x because T is a pair covering.

To show that h is positive definite, we write

$$h(x) = \sum_{j=0}^{t-1} c_j \prod_{i=0}^{j-1} (b_{t-i} - x)$$

with $c_j \ge 0$ by (10) and $\prod_{i=0}^{j-1} (b_{t-i} - x)$ being positive definite by hypothesis (b).

All that remains is to check the complementary slackness conditions. Because $h(x) = f(x)$ for all $x \in T$, they are equal on the support of $\mathbf{a}$. Because $\mathbf{a}$ is a $(t-1)$-design, the dual support is contained in $\{t, \dots, n\}$ and hence the Krawtchouk coefficients of h vanish on the dual support. Thus, $\mathbf{a}$ is universally optimal. □

Now we discuss two special cases in which part (b) of Proposition 21 is easy to verify using our results from Section IV-A, as well as two elementary cases that do not fit into the framework of Proposition 21.

Proposition 22. *Let* $\mathbf{a}$ *be a quasicode of length* n *over* $\mathbb{F}_q$*, and let* T *be a pair covering of* $\mathbf{a}$ *with* $|T| = t$*. If* $\mathbf{a}$ *is a* $(t-1)$*-design and at most one element of* T *is less than* $(q-1)n/q$*, then* $\mathbf{a}$ *is universally optimal.*

Proof. We only need to check condition (b) of Proposition 21. Since at most one element of T is less than $(q-1)n/q$, namely b_1, the product $\prod_{i=0}^{j-1}(b_{t-i} - x)$ is positive definite by Corollary 14 for $1 \le j \le t-1$. □

Proposition 23. *Let* $\mathbf{a}$ *be a quasicode of length* n *over* $\mathbb{F}_q$*. Suppose that* $\mathbf{a}$ *has* s *support elements and is a* $(2s-1)$*-design. Furthermore, suppose that every two elements in the support differ by at least* 2*, and at most one element of the support is less than* $(q-1)n/q$*. Then* $\mathbf{a}$ *is LP universally optimal.*

Proof. We shall construct a pair covering that satisfies the conditions of Proposition 21. Suppose that nonzero elements of the support are $a_1 < a_2 < \cdots < a_s$, so that $a_i \ge (q-1)n/q$ for all $i \ge 2$. If $a_s < n$, then set

$$T = \{a_1 - 1, a_1\} \cup \{a_2, a_2 + 1\} \cup \{a_3, a_3 + 1\} \cup \cdots \cup \{a_s, a_s + 1\}$$

and if $a_s = n$, then set

$$T = \{a_1 - 1, a_1\} \cup \{a_2, a_2 + 1\} \cup \{a_3, a_3 + 1\} \cup \cdots \cup \{a_s\}.$$

By construction, T is a pair covering. Let $t = |T|$. When $a_s < n$, we have $t = 2s$, and when $a_s = n$, we have $t = 2s - 1$. So $\mathbf{a}$ is always a $(t-1)$-design and condition (a) of Proposition 21 is satisfied.

Now we check condition (b) of Proposition 21. In the $a_s < n$ case, the partial product of an initial segment of

$$(a_s + 1 - x)(a_s - x) \cdots (a_2 + 1 - x)(a_2 - x)$$

is positive definite by Corollary 14 since $a_j \ge (q-1)n/q$ for $j \ge 2$. Furthermore

$$(a_s + 1 - x)(a_s - x) \cdots (a_2 + 1 - x)(a_2 - x)(a_1 - x)$$

is positive definite:

$$(a_s - x)(a_{s-1} - x) \cdots (a_1 - x)$$

is positive definite by Lemma 15 and

$$(a_s + 1 - x)(a_{s-1} + 1 - x) \cdots (a_2 + 1 - x)$$

is positive definite by Corollary 14, and so their product is positive definite by Lemma 12. This completes the $a_s < n$ case. The $a_s = n$ case is nearly identical. Thus, condition (b) of Proposition 21 is satisfied, and $\mathbf{a}$ is universally optimal. □

Proposition 24. *Let* $\mathbf{a}$ *be a quasicode. Suppose that* $\mathbf{a}$ *is a* 1*-design, whose support consists of a single integer or two consecutive integers. Then* $\mathbf{a}$ *is universally optimal.*

Sketch of proof. We use a linear auxiliary function that agrees with the potential function on the support, and on a neighboring point if the support has size one. □

Proposition 25. *Let* $\mathbf{a}$ *be a binary quasicode of length* n*. Suppose* $\mathbf{a}$ *is supported at* $\{0, a-1, a, a+1\}$ *where* a *is odd, while its dual* $\mathbf{a}^\perp$ *satisfies* $A_1^\perp = A_n^\perp = 0$*. Then* $\mathbf{a}$ *is universally optimal.*

Sketch of proof. For a potential function f, one can check that the auxiliary function

$$f(a-1) + \frac{1}{2}(f(a-1) - f(a+1))(a - 1 - x) + \frac{1}{4}(f(a-1) - 2f(a) + f(a+1))(K_n(x) - 1)$$

works as $h(x)$ in Proposition 5. □

VI. Removing a codeword from a code

In this section, we show that removing a single codeword from an LP universally optimal code always yields a universally optimal code. This surprising fact will follow from a strengthening of the Delsarte linear program due to Ashikhmin and Simonis [3]. It can fail without LP universal optimality: in $\mathbb{F}_2^2$, the three-point code $\{(0,0), (0,1), (1,1)\}$ is universally optimal, but $\{(0,0), (0,1)\}$ is not.

Proposition 26 (Ashikhmin and Simonis [3])**.** *Let* $\mathcal{C}$ *be a code of length* n *over an alphabet of size* q *and such that* q *does not divide* $|\mathcal{C}|$*, and let* $(A_0, \ldots, A_n)$ *be its distance distribution. Then for* $0 \le j \le n$*,*

$$|\mathcal{C}| \sum_{i=0}^{n} A_i K_j(i) \ge (q-1)^j \binom{n}{j}.$$

See arXiv:1212.1913v1 for a streamlined variant of the proof from [3].

Lemma 27. *Let* f *be any potential function. If the Delsarte linear program proves that a code* $\mathcal{C} \subseteq \mathbb{F}_q^n$ *minimizes* f*-potential energy, then either* $|\mathcal{C}|$ *is a multiple of* q *or* f *is minimized at all the distances between pairs of distinct codewords in* $\mathcal{C}$*.*

Proof. Suppose $|\mathcal{C}|$ is not a multiple of q. Proposition 26 shows that the dual distance distribution of $\mathcal{C}$ is strictly positive, and thus the auxiliary function in Proposition 5 must be constant. Then the conclusion follows, because the auxiliary function is less than or equal to f everywhere and equal on the support of $\mathcal{C}$. □

Corollary 28. *Every LP universally optimal code in* $\mathbb{F}_q^n$ *has size a multiple of* q *unless all pairs of distinct points in the code are at distance* n*.*

In the latter case, there can be at most q points in the code.

Proposition 29. *Let* $\mathcal{C} \subseteq \mathbb{F}_q^n$ *be a code and let* $f\colon \{1, 2, \ldots, n\} \to \mathbb{R}$ *be any function (not necessarily completely monotonic) such that the Delsarte linear programming bounds prove* $\mathcal{C}$ *minimizes* f*-potential energy. Let* $c \in \mathcal{C}$*. Then* $E_f(\mathcal{C} \setminus \{c\}) \le E_f(\mathcal{C}')$ *for every code* $\mathcal{C}' \subseteq \mathbb{F}_q^n$ *with* $|\mathcal{C}'| = |\mathcal{C}| - 1$*.*

Proof. By Lemma 27 we may assume that $|\mathcal{C}|$ is a multiple of q, because the other case in the lemma is trivial. Let $N = |\mathcal{C}|$, let $(1, A_1, \ldots, A_n)$ be the distance distribution of $\mathcal{C}$, and

let $(1, B_1, \ldots, B_n)$ be the expected distance distribution after removing a random codeword from $\mathcal{C}$. Given $(x, y) \in \mathcal{C}^2$ with $x \neq y$, the probability that neither will be removed is $(N-2)/N$. Thus, $B_0 = 1$ while for $i \geq 1$,

$$\begin{aligned} B_i &= \frac{1}{N-1} \cdot \frac{N-2}{N} \left|\left\{(x,y) \in \mathcal{C}^2 : |x-y| = i\right\}\right| \\ &= \frac{N-2}{N-1} A_i. \end{aligned}$$

Under this relationship between A_i and B_i, the Delsarte inequalities

$$\sum_{i=0}^{n} A_i K_j(i) \geq 0$$

simply say

$$K_j(0) + \frac{N-1}{N-2} \sum_{i=1}^{n} B_i K_j(i) \geq 0$$

and hence are equivalent to

$$(N-1) \sum_{i=0}^{n} B_i K_j(i) \geq K_j(0) = (q-1)^j \binom{n}{j}.$$

The key observation underlying the proof is that these inequalities on (B_i) are exactly the Ashikhmin-Simonis inequalities from Proposition 26; i.e., (A_i) satisfies the Delsarte inequalities if and only if (B_i) satisfies the Ashikhmin-Simonis inequalities.

Thus, our hypothesis that (A_i) minimizes the f-potential energy

$$\sum_{i=1}^{n} A_i f(i)$$

among nonnegative vectors subject to the Delsarte inequalities, $A_0 = 1$, and $\sum_i A_i = N$ implies that (B_i) minimizes the expected energy

$$\sum_{i=1}^{n} B_i f(i) = \frac{N-2}{N-1} \sum_{i=1}^{n} A_i f(i)$$

subject to the Ashikhmin-Simonis inequalities, $B_0 = 1$, and $\sum_i B_i = N - 1$.

The Ashikhmin-Simonis inequalities apply to all codes of size $N-1$, because N is a multiple of q and hence $N-1$ is not. This means no code of size $N-1$ in $\mathbb{F}_q^n$ can have lower f-potential energy than the expected energy after removing a random codeword from $\mathcal{C}$. Removing different codewords might yield non-isomorphic codes, but by linearity of expectation they must all have the same energy, since none of them can have lower energy than the average. It follows that for every $c \in \mathcal{C}$, the code $\mathcal{C} \setminus \{c\}$ minimizes f-potential energy among all codes of size $|\mathcal{C}| - 1$. □

In particular, by letting f vary over all completely monotonic functions, we see that if $\mathcal{C}$ is LP universally optimal, then $\mathcal{C} \setminus \{c\}$ is universally optimal for all $c \in \mathcal{C}$. All of these codes $\mathcal{C} \setminus \{c\}$ must have the same distance distribution, since they have the same energy for all completely monotonic potential functions, which span the space of all potential functions. Thus $\mathcal{C}$ must be distance regular.

This completes the proof of Theorem 2. We find the result quite surprising, and the role of the Ashikhmin-Simonis inequalities in the proof is mysterious. It is natural to look for other proofs of these inequalities. There is a much simpler proof for binary codes (Theorem 5 in [4]), which we have been able to generalize to alphabets of prime power order but no further. The elegant proof of the Delsarte inequalities in [19] can also be adapted to give a proof of the Ashikhmin-Simonis inequalities, but in fact there is an error in [19]: equation (13″) is incorrect and the map σ is not well defined for a general alphabet. When the alphabet has prime power order, the proof works, but we see no way to salvage it in general.

VII. Further questions and generalizations

In the introduction we mentioned a $N \leftrightarrow 32 - N$ symmetry for codes of size N in $\mathbb{F}_2^5$. The unoccupied locations in a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ can be viewed as antiparticles, which are subject to exactly the same forces as the original particles:

$$\begin{aligned} (q^n - |\mathcal{C}|)\, E_f(\mathbb{F}_q^n \setminus \mathcal{C}) = |\mathcal{C}|\, E_f(\mathcal{C}) \\ + (q^n - 2|\mathcal{C}|) \sum_{k=1}^{n} \binom{n}{k} (q-1)^k f(k) \end{aligned}$$

by a simple inclusion-exclusion argument (see also Section 1.3.4 of [15] for an essentially equivalent lemma). Thus, $\mathbb{F}_q^n \setminus \mathcal{C}$ is universally optimal if and only if $\mathcal{C}$ is.

Linear programming bounds do not respect this antiparticle symmetry. For codes of size greater than $q^n/2$ in $\mathbb{F}_q^n$, passing to the complement can strengthen the Delsarte bounds, while for codes of size at most $q^n/2$ one can show that this yields no improvement. Of course, few important codes have size greater than $q^n/2$.

Beyond linear programming bounds and antiparticle symmetry, are there systematic techniques that could be applied? Semidefinite programming bounds [18], [13] are the most powerful approach known to proving coding theory bounds. They have been applied to potential energy minimization in projective space [8], but we have not investigated them in $\mathbb{F}_q^n$.

Many of our results generalize straightforwardly to metric and cometric association schemes, i.e., distance-regular graphs under the graph metric with the "Q-polynomial" property [11]. There are several noteworthy omissions, namely the theory of duality (including the definition of the dual quasicode and Proposition 9) and the results of Section VI. However, the results of Sections IV and V all generalize if $(q-1)n/q$ is replaced with the average distance between a pair of randomly selected points in the graph, with the exception of Lemma 16 (which is needed only for MDS codes) and Proposition 25. The proofs are essentially identical.

We have not attempted to compile an exhaustive list of examples for this more general theory, along the lines of Table I, but there are several interesting applications. For example, consider the Johnson space of binary vectors of length n and weight w. Every projective plane of order q yields an $S(2, q+1, q^2+q+1)$ Steiner system and thus a configuration of q^2+q+1 points in the Johnson space with parameters $(n, w) = (q^2+q+1, q+1)$. This configuration

is a simplex and a 2-design, so it is universally optimal. The $S(5, 8, 24)$ Steiner system is a somewhat deeper example.

The role of duality in association scheme theory is well understood (see Section 2.6 in [10]), and it does not generalize to arbitrary metric and cometric association schemes. However, the results of Section VI are far more mysterious, and we have no idea how far they might generalize. In particular, we have no conceptual explanation for why Proposition 26 turns out to be exactly what we require to analyze removing one point from an LP universal optimum. Any progress on generalizing either the results or the proof techniques to other association schemes would be exciting.

ACKNOWLEDGMENTS

We thank Alexei Ashikhmin, Alexander Barg, and the anonymous referee for providing valuable feedback and suggestions.

REFERENCES

[1] A. Ashikhmin and A. Barg, "Binomial moments of the distance distribution: bounds and applications," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 438–452, Mar. 1999, doi:10.1109/18.748994.

[2] A. Ashikhmin, A. Barg, and S. Litsyn, "Estimates of the distance distribution of codes and designs," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1050–1061, Mar. 2001, doi:10.1109/18.915662.

[3] A. Ashikhmin and J. Simonis, "On the Delsarte inequalities," *Linear Algebra Applicat.*, vol. 269, no. 1–3, pp. 197–217, Jan. 1998, doi:10.1016/S0024-3795(97)00065-7.

[4] M. R. Best, A. E. Brouwer, F. J. MacWilliams, A. M. Odlyzko, and N. J. A. Sloane, "Bounds for binary codes of length less than 25," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 81–93, Jan. 1978, doi:10.1109/TIT.1978.1055827.

[5] N. Bouman, J. Draisma, and J. van Leeuwaarden, "Energy minimization of repelling particles on a toric grid," *SIAM J. Discrete Math.*, vol. 27, no. 3, pp. 1295–1312, 2013, doi:10.1137/120869067.

[6] H. Cohn, "Order and disorder in energy minimization," in *Proceedings of the International Congress of Mathematicians*, vol. IV. New Delhi, India: Hindustan Book Agency, 2010, pp. 2416–2443, doi:10.1142/9789814324359_0152.

[7] H. Cohn and A. Kumar, "Universally optimal distribution of points on spheres," *J. Amer. Math. Soc.*, vol. 20, no. 1, pp. 99–148, Jan. 2007, doi:10.1090/S0894-0347-06-00546-7.

[8] H. Cohn and J. Woo, "Three-point bounds for energy minimization," *J. Amer. Math. Soc.*, vol. 25, no. 4, pp. 929–958, Oct. 2012, doi:10.1090/S0894-0347-2012-00737-1.

[9] P. Delsarte, "Bounds for unrestricted codes, by linear programming," *Philips Research Reports*, vol. 27, pp. 272–289, 1972.

[10] ——, "An algebraic approach to the association schemes of coding theory," *Philips Research Reports Suppl.*, vol. 10, 1973.

[11] P. Delsarte and V. I. Levenshtein, "Association schemes and coding theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2477–2504, Oct. 1998, doi:10.1109/18.720545.

[12] G. Ferrari and K. M. Chugg, "Linear programming-based optimization of the distance spectrum of linear block codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1794–1800, Jul. 2003, doi:10.1109/TIT.2003.813483.

[13] D. Gijswijt, A. Schrijver, and H. Tanaka, "New upper bounds for nonbinary codes based on the Terwilliger algebra and semidefinite programming," *J. Combinatorial Theory Series A*, vol. 113, no. 8, pp. 1719–1731, Nov. 2006, doi:10.1016/j.jcta.2006.03.010.

[14] T. Helleseth, T. Kløve, and V. I. Levenshtein, "The simplex codes and other even-weight binary linear codes for error correction," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2818–2823, Nov. 2004, doi:10.1109/TIT.2004.836708.

[15] T. Kløve, *Codes for Error Detection*. Hackensack, NJ: World Scientific Publishing Co., 2007, doi:10.1142/9789812770516.

[16] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, the Netherlands: North-Holland Publishing Co., 1977.

[17] R. J. McEliece, *The Theory of Information and Coding*, student edition. Cambridge, U.K.: Cambridge University Press, 2004, doi:10.1017/CBO9780511819896.

[18] A. Schrijver, "New code upper bounds from the Terwilliger algebra and semidefinite programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2859–2866, Aug. 2005, doi:10.1109/TIT.2005.851748.

[19] J. Simonis and C. de Vroedt, "A simple proof of the Delsarte inequalities," *Designs, Codes and Cryptography*, vol. 1, no. 1, pp. 77–82, 1991, doi:10.1007/BF00123961.

[20] V. A. Yudin, "The minimum of potential energy of a system of point charges," *Discrete Math. Applicat.*, vol. 3, no. 1, pp. 75–81, 1993, doi:10.1515/dma.1993.3.1.75.

Henry Cohn is a principal researcher at Microsoft Research New England and adjunct professor of mathematics at MIT. His received his Ph.D. in mathematics from Harvard in 2000, under the supervision of Noam Elkies. His research is in discrete mathematics, with connections to physics and computer science.

Yufei Zhao is a Ph.D. student in the Department of Mathematics at the Massachusetts Institute of Technology. He received his B.Sc. from MIT in 2010 and M.A.St. from the University of Cambridge in 2011. His research interests include extremal and probabilistic combinatorics, as well as their applications to other subjects such as number theory, probability, and geometry. He was a recipient of the 2013 Microsoft Research PhD Fellowship.