Proposition 11: Algorithm PERIODIC-MF-TRIES computes from the automata $\mathcal{A}_k$ accepting $\mathrm{Fact}^{(k)}(y)$ the set of tries accepting the minimal periodic forbidden words of y.

Proof: Again, the proof is an extension of the correctness proof of the computation of the minimal forbidden words of a word from the factor automaton of y (see [10, Sec. 6.5, pp. 182] or [7]). □

Proposition 12: Algorithm PERIODIC SUFFIX AUTOMATON followed by algorithm PERIODIC-MF-TRIES runs in time $O(|y| \times T \times \log|A|)$.

Proof: The complexity is straightforward. □

REFERENCES

[1] J. Moon and B. Brickner, "Maximum transition run codes for data storage," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3992–3994, Sep. 1996.

[2] A. Wijngaarden and K. S. Immink, "Maximum run-length limited codes with error control properties," *IEEE J. Select. Areas Commun.*, vol. 19, no. 4, pp. 602–611, Apr. 2001.

[3] J. C. de Souza, B. H. Marcus, R. New, and B. A. Wilson, "Constrained systems with unconstrained positions," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 866–879, Apr. 2002.

[4] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. Pless and W. Huffman, Eds. Amsterdam, The Netherlands: North Holland, 1998, vol. II, ch. 20, pp. 1635–1764.

[5] B. E. Moision and P. H. Siegel, "Periodic-finite-type shift spaces," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 65.

[6] M.-P. Béal, M. Crochemore, F. Mignosi, A. Restivo, and M. Sciortino, "Computing forbidden words of regular languages," *Fundamenta Informaticae*, vol. 56, no. 1,2, pp. 121–135, 2003.

[7] M. Crochemore, F. Mignosi, and A. Restivo, "Automata and forbidden words," *Inf. Process. Lett.*, vol. 67, pp. 111–117, 1998.

[8] M. Crochemore, F. Mignosi, A. Restivo, and S. Salemi, "Data compression using antidictionaries," *Proc. IEEE (Special Issue on Lossless Data Compression)*, vol. 88, no. 11, pp. 1756–1768, Nov. 2000.

[9] D. A. Lind and B. H. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[10] M. Crochemore, C. Hancart, and T. Lecroq, *Algorithmique du Texte*. Paris, France: Vuibert, 2001.

[11] A. Blumer, J. Blumer, A. Ehrenfeucht, D. Haussler, and R. McConnell, "Linear size finite automata for the set of all subwords of a word : An outline of results," *Bull. Europ. Assoc. Theoret. Comput. Sci.*, no. 21, pp. 12–20, 1983.

[12] ——, "Building a complete inverted file for a set of text files in linear time," in *Proc. 16th ACM Symp. Theory of Computing*, Washington, DC, Apr./May 1984, pp. 349–351.

Achieving the Welch Bound With Difference Sets

Pengfei Xia, *Student Member, IEEE*, Shengli Zhou, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Consider a codebook containing N unit-norm complex vectors in a K-dimensional space. In a number of applications, the codebook that minimizes the maximal cross-correlation amplitude ($I_{\max}$) is often desirable. Relying on tools from combinatorial number theory, we construct analytically optimal codebooks meeting, in certain cases, the Welch lower bound. When analytical constructions are not available, we develop an efficient numerical search method based on a generalized Lloyd algorithm, which leads to considerable improvement on the achieved $I_{\max}$ over existing alternatives. We also derive a composite lower bound on the minimum achievable $I_{\max}$ that is effective for any codebook size N.**

***Index Terms*—Difference sets, generalized Lloyd algorithm, Grassmannian line packing, Welch bound.**

I. INTRODUCTION

Consider a complex (N, K) codebook that is a collection of N unit-norm complex vectors in a K-dimensional vector space. One problem that often arises is to minimize the codebook's maximal cross-correlation amplitude $I_{\max}$. For multiple-antenna transmit beamforming based on limited-rate feedback, minimizing $I_{\max}$ among codewords (beamforming vectors) approximately optimizes various performance metrics including outage probability [19], average signal-to-noise ratio (SNR) [16], and symbol error probability [31]. Minimizing $I_{\max}$ in the context of unitary space–time modulations is equivalent to minimizing the block error probability [9]. For multiple description coding over erasure channels, minimizing $I_{\max}$ of the finite frames leads to minimal reconstruction error [26].

Finding the optimal codebook with minimal $I_{\max}$ is very difficult in general, both "analytically and numerically" [16], [23]. The challenge, on the other hand, bears close connection with many other problems in different areas. One equivalent problem is line packing in the Grassmannian manifold, where one seeks N lines in the K-dimensional space so that the maximum chordal distance between any two lines is minimized [2]. In frame theory, such a codebook with $I_{\max}$ minimized is known as a Grassmannian frame. Other closely related problems include the design of equi-angular line sets, the design of antipodal spherical codes, spherical t-designs, and characterization of strongly regular graphs [26].

Because analytical construction of the optimal codebook is possible only in very special cases [2], [23], numerical search algorithms are often sought to obtain near-optimal codebooks. Aside from finding the optimal codebook, lower-bounding the achievable $I_{\max}$ is also important. The Welch lower bound [28] on $I_{\max}$ is particularly useful for

Manuscript received September 28, 2004; revised January 31, 2005. This correspondence was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The material in this correspondence was presented in part at the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, March 19–23, 2005.

P. Xia and G. B. Giannakis are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: pfxia@ece.umn.edu; georgios@ece.umn.edu).

S. Zhou is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: shengli@engr.uconn.edu).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2005.846411

relatively small values of N, but becomes quite loose for large N: it is known that the Welch bound is no longer achievable when $N > K^2$ in the complex case, or, when $N > K(K+1)/2$ in the real case [3].

Optimal real codebooks have been studied in [2] and [24], while an extensive list of putatively optimal real codebooks can be found in [25]. In this correspondence, we focus on designing optimal complex codebooks, and our contributions are as follows.

- Relying on tools from combinatorial design theory, we derive analytical constructions of optimal codebooks. An additional feature of the so-constructed optimal codebooks is that all codebook entries have equal amplitudes.
- When analytical constructions do not exist, we develop a modified Lloyd search algorithm, which improves considerably upon the achieved $I_{\max}$ of existing designs [9], [16].
- We finally develop a composite lower bound on the achievable $I_{\max}$ that is effective for any N, as verified by our searched codebooks through the modified Lloyd search algorithm.

A. Notation

Bold-face upper and lower case letters denote matrices and column vectors, respectively; $\boldsymbol{I}_K$ is the $K \times K$ identity matrix; $\mathbb{E}$ is the ensemble average operator; $(\cdot)^*$ and $(\cdot)^{\mathcal{H}}$ denote conjugate and Hermitian transpose, respectively; $j = \sqrt{-1}$ is the imaginary unit; $\setminus$ is the set minus operator with $A \setminus B := \{x : x \in A \text{ and } x \notin B\}$; $\mathbb{Z}_+$ denotes the set of positive integers; $\mathbb{Z}_N := \{0, 1, \ldots, N-1\}$ denotes the set of integers modulo N, while addition, subtraction, and multiplication over $\mathbb{Z}_N$ are all performed modulo N; $\mathbb{Z}_N^* := \mathbb{Z}_N \setminus \{0\}$; and $\mathbb{F}_N$ is the finite field of order N.

II. Problem Formulation and Context

Without loss of generality, we consider a codebook comprising N codewords $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$, with every codeword $\boldsymbol{w}_\ell$ being a unit norm $K \times 1$ complex vector; and define the $K \times N$ matrix $\boldsymbol{W} := [\boldsymbol{w}_1 \ \ldots \ \boldsymbol{w}_N]$ to represent this (N, K) codebook. We emphasize that throughout this correspondence, every codebook entry can take complex values. The root-mean-square (rms) cross correlation and the maximum cross-correlation amplitudes of such a codebook are defined as

$$I_{\text{rms}}(\boldsymbol{W}) := \sqrt{\frac{1}{N(N-1)} \sum_{\ell=1}^{N} \sum_{\ell' \neq \ell}^{N} |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell'}|^2} \tag{1}$$

$$I_{\max}(\boldsymbol{W}) := \max_{1 \leq \ell < \ell' \leq N} |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell'}|. \tag{2}$$

The following results are well known [18], [28].

Lemma 1 (Welch Lower Bound): For any codebook $\boldsymbol{W}$ with $N \geq K$

$$I_{\text{rms}}(\boldsymbol{W}) \geq \sqrt{\frac{N-K}{(N-1)K}} \tag{3}$$

with equality if and only if $\sum_{\ell=1}^{N} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^{\mathcal{H}} = (N/K)\boldsymbol{I}_K$. Also,

$$I_{\max}(\boldsymbol{W}) \geq \sqrt{\frac{N-K}{(N-1)K}} \tag{4}$$

with equality if and only if

$$|\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell'}| = \sqrt{\frac{N-K}{(N-1)K}}, \qquad \forall \ell \neq \ell'. \tag{5}$$

If equality holds in (3), the codebook $\boldsymbol{W}$ meets the Welch bound on I_{rms} with equality, and is generally referred to as a Welch-bound-equality (WBE) codebook. In frame theory, WBE codebooks are also known as uniform tight frames. A codebook $\boldsymbol{W}$ that satisfies (4) as an equality, and thus meets the Welch bound on $I_{\max}$, is referred to as a maximum-Welch-bound-equality (MWBE) codebook. Throughout the correspondence, the term "Welch bound" will be used to represent the bound on $I_{\max}$ (4), unless explicitly specified. Apparently, if a codebook $\boldsymbol{W}$ is MWBE, it automatically solves the problem

$$\min_{\boldsymbol{W}} \max_{\ell \neq \ell'} |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell'}| \tag{6}$$

which is of interest in this correspondence and in many applications.

The following results are available for codebooks with complex or real entries.[1]

- An MWBE codebook is WBE [23], but not *vice versa*. This can be easily verified from definition of the WBE codebook (2) and the necessary and sufficient conditions of the MWBE codebook (5).
- WBE codebooks are "almost trivially easy" to find [23], while minimizing $I_{\max}$ is notoriously difficult in general, both "analytically and numerically" [16], [23]. Analytic constructions of MWBE codebooks are very limited. Two known examples are
  i) simplex signaling for $(N, N-1)$ codebooks [21];
  ii) construction based on conference matrices [2], [26], when $N = 2K = 2^{d+1}$ with $d \in \mathbb{Z}_+$, yielding complex codebooks; and when $N = 2K = p^d + 1$ with p a prime number and $d \in \mathbb{Z}_+$, resulting into real codebooks.

III. Novel MWBE Codebook Construction

In this section, we derive analytical constructions of complex MWBE codebooks, by exploiting tools from combinatorial design theory. Interestingly, every entry of the so-constructed codebooks will turn out to have the same amplitude.

A. Systematic Construction Based on Fast Fourier Transform (FFT) Matrices

Recalling that every MWBE codebook is WBE [23], we can first restrict ourselves to a finite/infinite collection of WBE codebooks, and try to find the desired MWBE codebook (if possible) within this pool of codebooks. Such a strategy would generally simplify the design process, as WBE codebooks are relatively easy to construct and often possess certain favorable structures. We are particularly interested in one special class of WBE codebooks with the following highly restricted structure [9]:

$$\boldsymbol{W}(\boldsymbol{u}) = \frac{1}{\sqrt{K}} \begin{bmatrix} 1 & e^{j\frac{2\pi}{N}u_1} & \ldots & e^{j\frac{2\pi}{N}u_1(N-1)} \\ 1 & e^{j\frac{2\pi}{N}u_2} & \ldots & e^{j\frac{2\pi}{N}u_2(N-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{j\frac{2\pi}{N}u_K} & \ldots & e^{j\frac{2\pi}{N}u_K(N-1)} \end{bmatrix} \tag{7}$$

where $u_k \in \mathbb{Z}_N, \forall 1 \leq k \leq K$ and $u_k \neq u_\ell, \forall k \neq \ell$. Without loss of generality, it is convenient to let every column index range between 0 and $N-1$. Before normalization, $\boldsymbol{W}(\boldsymbol{u})$ consists of K distinct rows chosen from an $N \times N$ FFT matrix, while $\boldsymbol{u} := \{u_1, \ldots, u_K\}$ denotes the set of the selected row indices. Notice that every entry of this codebook is constrained to have the same amplitude $1/\sqrt{K}$. This is desirable in many cases, e.g., in equal gain transmit beamforming for multiple-antenna wireless communications [20], and equal gain unitary space–time modulations [9]. Interestingly, for every so-constructed codebook the codeword cross correlation

$$\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell'} = \frac{1}{K} \sum_{k=1}^{K} e^{j\frac{2\pi}{N}u_k(\ell'-\ell)} \tag{8}$$

[1]In this correspondence, we only study the design of complex codebooks without alphabet constraints. Designing binary WBE/MWBE codebooks is a different problem and can be difficult in general; see, e.g., [4], [10]–[12]. The results and properties for general complex codebooks do not necessarily carry over to the binary, or the finite alphabet cases.

is determined only by the difference $(\ell' - \ell) \bmod N$. Therefore, by constructing codebooks as in (7), $\boldsymbol{W}(\boldsymbol{u})$ exhibits a circulant correlation structure [9].

The codebook $\boldsymbol{W}(\boldsymbol{u})$ is now determined by only K parameters $\{u_1, \ldots, u_K\}$. For every possible $\boldsymbol{u}$, the codebook $\boldsymbol{W}(\boldsymbol{u})$ is WBE by construction; but it may generally be far from being MWBE. The choice of the row indices $\boldsymbol{u}$ vastly influences $I_{\max}$ of the so-constructed codebook; see [9, Figs. 2 and 3] for an illustrative example. Theoretically, one can search exhaustively over all possible $\binom{N}{K}$ row indices, in an attempt to find an MWBE codebook from this pool of codebooks. However, as $\binom{N}{K}$ increases explosively with N, a randomly incomplete search is often employed instead of the exhaustive search. This is precisely the codebook search algorithm developed in [9].

The codebook in (7) may be too restrictive to yield MWBE codebooks, due to the additional constraints of equal amplitude for every entry and circulant correlation structure. Nevertheless, we will show next that such a highly structured approach still enables analytic MWBE codebook construction for certain (N, K) pairs. The core questions are then: when can we construct MWBE codebooks with structure as in (7)? and how can we achieve this without numerical search?

B. Codebook Constructions Based on Difference Sets

Based on the necessary and sufficient conditions in (5), the MWBE codebook design problem now boils down to

$$\text{find } \boldsymbol{u}, \text{ s. t. } f(1) = \cdots = f(N-1) = \sqrt{\frac{N-K}{(N-1)K}} \tag{9}$$

where

$$f(m) := |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{w}_{\ell-m}| = \frac{1}{K}\left|\sum_{k=1}^{K} e^{j2\pi \frac{m}{N} u_k}\right| \tag{10}$$

with $f(0) \equiv 1$ and $f(m) \equiv f(N-m)$.

We start by investigating two trivial cases: $K = N$ and $K = N-1$. For $K = N$, selecting all the N rows from the FFT matrix forms a trivial (N, N) MWBE codebook; and, therefore, $\boldsymbol{u} = \mathbb{Z}_N$ is an optimal solution to (9) when $K = N$.

When $K = N-1$, let us select arbitrarily $N-1$ rows from the FFT matrix by excluding the ℓth row. It directly follows that $\forall m \in \mathbb{Z}_N^*$, and $\forall \ell \in \mathbb{Z}_N$

$$f(m) = \frac{1}{K}\left|\sum_{k=0, k\neq\ell}^{N-1} e^{j\frac{2\pi}{N}km}\right| = \frac{1}{K} = \sqrt{\frac{N-K}{(N-1)K}}. \tag{11}$$

Therefore, choosing any $N-1$ rows from the FFT matrix forms an $(N, N-1)$ MWBE codebook, which corresponds to a simplex design of N signals in an $(N-1)$-dimensional space. Without loss of generality, we choose $\boldsymbol{u} = \mathbb{Z}_N^*$ as an optimal solution to (9) when $K = N-1$.

The optimal row indices $\boldsymbol{u}$ in the aforementioned two cases share a subtle resemblance that we will clarify after we introduce this definition.

Definition: A subset $\boldsymbol{u} = \{u_1, \ldots, u_K\}$ of $\mathbb{Z}_N$ is called an (N, K, λ) *difference set* if the $K(K-1)$ differences

$$(u_k - u_\ell) \bmod N, \qquad k \neq \ell$$

take all possible nonzero values $1, 2, \ldots, N-1$, with each value exactly λ times.

It can be readily verified that in the two special cases we identified as MWBE, $\boldsymbol{u} = \mathbb{Z}_N$ is an (N, N, N) difference set, and $\boldsymbol{u} = \mathbb{Z}_N^*$ is an $(N, N-1, N-2)$ difference set. Notice that the three parameters (N, K, λ) are not independent by definition, as they satisfy

$$\lambda(N-1) \equiv K(K-1). \tag{12}$$

With the tool of difference sets, we have the following general results.

Theorem 1: The (N, K) codebook $\boldsymbol{W}(\boldsymbol{u})$ as in (7) is MWBE if and only if $\boldsymbol{u}$ is an (N, K, λ) difference set.

Proof: From the definition in (10), $\forall m \in \mathbb{Z}_N$, we have

$$\begin{aligned} f_m^2 &= \frac{1}{K^2}\left(\sum_{i=1}^{K} e^{j2\pi m u_i/N}\right)\left(\sum_{k=1}^{K} e^{-j2\pi m u_k/N}\right) \\ &= \frac{N-K}{(N-1)K} + \frac{K-1}{(N-1)K} \\ &\quad + \frac{1}{K^2}\sum_{i=1}^{K}\sum_{k=1,k\neq i}^{N} e^{j2\pi m(u_i-u_k)/N}. \end{aligned} \tag{13}$$

For all possible $(u_i - u_k) \bmod N$, where $k \neq i$, let a_ℓ denote the number of occurrences of $\ell, \forall \ell = 1, \ldots, N-1$. Thus, the last term in (13) can be written as

$$\frac{1}{K^2}\sum_{i=1}^{K}\sum_{k\neq i} e^{j2\pi m(u_i-u_k)/N} = \frac{1}{K^2}\sum_{\ell=1}^{N-1} a_\ell e^{j2\pi m\ell/N}. \tag{14}$$

Hence, we can rewrite (13) as

$$K^2 \cdot \left(f_m^2 - \frac{N-K}{(N-1)K}\right) = \frac{K(K-1)}{(N-1)} + \sum_{\ell=1}^{N-1} a_\ell e^{j2\pi m\ell/N}. \tag{15}$$

With the definitions of

$$x[m] := K^2 \cdot \left(f_m^2 - \frac{N-K}{(N-1)K}\right)$$

and

$$a_0 := \frac{K(K-1)}{(N-1)}$$

we have

$$x[m] = \sum_{\ell=0}^{N-1} a_\ell e^{j2\pi m\ell/N}, \qquad \forall m \in \mathbb{Z}_N \tag{16}$$

i.e., $\{x[m]\}_{m=0}^{N-1}$ and $\{a_\ell\}_{\ell=0}^{N-1}$ form a discrete Fourier transform pair. Therefore,

$$x[0] = \frac{K(K-1)N}{(N-1)},\ x[1] = \cdots = x[N-1] = 0 \tag{17}$$

leads to

$$a_0 = a_1 = \cdots = a_{N-1} = \frac{K(K-1)}{(N-1)} \tag{18}$$

and *vice versa*, as the Fourier transform of a delta function is a constant. According to (5), the codebook $\boldsymbol{W}(\boldsymbol{u})$ is MWBE if and only if (17) is satisfied, while by definition, $\boldsymbol{u}$ is a difference set if and only if (18) holds true. Therefore, the codebook $\boldsymbol{W}(\boldsymbol{u})$ is MWBE if and only if $\boldsymbol{u}$ is a difference set. □

Difference sets have been well studied in the combinatorial design theory [5], [8], [27] and are known to exist for certain pairs of parameters (N, K), while the search for new difference sets is still under way. For a comprehensive repository of difference sets, please refer to [14]. We list in the following several families of nontrivial MWBE codebook designs based on difference sets [1], [8]. For illustration purpose, we also list several nontrivial codebook examples in Table I.

TABLE I
SOME NONTRIVIAL MWBE CODEBOOK CONSTRUCTIONS BASED ON DIFFERENCE SETS

N	K	FFT row indices $\mathbf{u}$ (subset of $\mathbb{Z}_N$)	Family
7	3	{1, 2, 4}	Quadratic
7	4	{0, 3, 5, 6}	Property 1
13	4	{0, 1, 3, 9}	Singer
11	5	{1, 3, 4, 5, 9}	Quadratic
21	5	{3, 6, 7, 12, 14}	Singer
11	6	{0, 2, 6, 7, 8, 10}	Property 1
31	6	{1, 5, 11, 24, 25, 27}	Singer
15	7	{0, 1, 2, 4, 5, 8, 10}	Twin-primes
15	8	{3, 6, 7, 9, 11, 12, 13, 14}	Property 1
57	8	{1, 6, 7, 9, 19, 38, 42, 49}	Singer
13	9	{2, 4, 5, 6, 7, 8, 10, 11, 12}	Property 1
37	9	{1, 7, 9, 10, 12, 16, 26, 33, 34}	Quartic
73	9	{1, 2, 4, 8, 16, 32, 37, 55, 64}	Octic
40	13	{0, 1, 3, 5, 9, 15, 22, 25, 26, 27, 34, 35, 38}	Singer

Family 1—MWBE Codebooks Based on Singer Difference Sets: Let $q = p^r$ be a power of a prime, $d \geq 2$ be a positive integer, α be a generator of the multiplicative group of $\mathbb{F}_{q^{d+1}}$

$$\mathrm{trace}(\omega) = \sum_{i=0}^{d} \omega^{q^i}$$

be the trace function from $\mathbb{F}_{q^{d+1}}$ to $\mathbb{F}_q$, and

$$N = \frac{q^{d+1}-1}{q-1}, \qquad K = \frac{q^d-1}{q-1}, \qquad \lambda = \frac{q^{d-1}-1}{q-1}. \tag{19}$$

Then $\boldsymbol{u} = \{t : 0 \leq t < N, \mathrm{trace}(\alpha^t) = 0\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Family 2—MWBE Codebooks Based on Quadratic Difference Sets: Let $q = p^r = 3 \bmod 4$ be a power of a prime and

$$N = q, \;\; K = \frac{q-1}{2}, \;\; \lambda = \frac{q-3}{4}. \tag{20}$$

Then $\boldsymbol{u} = \{t^2 : t \in \mathbb{Z}_N^*\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook. Lemma 2 is thus only a special case of this family when $r = 1$.

Family 3—MWBE Codebooks Based on Quartic Difference Sets: Let $p = 4a^2 + 1$ be a prime with a odd, and

$$N = p, \;\; K = \frac{p-1}{4}, \;\; \lambda = \frac{p-5}{16}. \tag{21}$$

Then $\boldsymbol{u} = \{t^4 : t \in \mathbb{Z}_N^*\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Similarly, let $p = 4a^2 + 9$ be a prime with a odd, and

$$N = p, \;\; K = \frac{p+3}{4}, \;\; \lambda = \frac{p+3}{16}. \tag{22}$$

Then $\boldsymbol{u} = \{t^4 : t \in \mathbb{Z}_N\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Family 4—MWBE Codebooks Based on Octic Difference Sets: Let $p = 8a^2 + 1 = 64b^2 + 9$ be a prime with a, b odd, and

$$N = p, \;\; K = a^2, \;\; \lambda = b^2. \tag{23}$$

Then $\boldsymbol{u} = \{t^8 : t \in \mathbb{Z}_N^*\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Similarly, let $p = 8a^2 + 49 = 64b^2 + 441$ be a prime with a odd, b even, and

$$N = p, \;\; K = a^2 + 6, \;\; \lambda = b^2 + 7. \tag{24}$$

Then $\boldsymbol{u} = \{t^4 : t \in \mathbb{Z}_N\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Family 5—MWBE Codebooks Based on Twin-Primes Difference Sets: Let p and $q = p + 2$ be a pair of twin primes, g be a common primitive root of both p and q, and

$$N = pq, \;\; K = \frac{pq-1}{2}, \;\; \lambda = \frac{pq-3}{4}. \tag{25}$$

Then $\boldsymbol{u} = \{1, g, g^2 \bmod N, \ldots, g^{(p^2-3)/2} \bmod N, 0, q, \ldots, (p-1)q\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

Similarly, let p, q be a pair of twin primes such that $(p-1, q-1) = 4$ with $d = (p-1)(q-1)/4$, g be a common primitive root of both p and q, and

$$N = pq, \;\; K = \frac{pq-1}{4}, \;\; \lambda = \frac{pq-5}{16}. \tag{26}$$

Then $\boldsymbol{u} = \{1, g, g^2 \bmod N, \ldots, g^{d-1} \bmod N, 0, q, 2q, \ldots, (p-1)q\}$ is an (N, K, λ) difference set, and $\boldsymbol{W}(\boldsymbol{u})$ forms an (N, K) MWBE codebook.

We next present a property of the MWBE codebook constructions based on (7).

Property 1: Let $\boldsymbol{u} = \{u_1, u_2, \ldots, u_K\}$ be the indices of the chosen K rows, and $\bar{\boldsymbol{u}} = \mathbb{Z}_N \setminus \boldsymbol{u}$ be the indices of the remaining $N - K$ rows. If $\boldsymbol{W}(\boldsymbol{u})$ is an (N, K) MWBE codebook, then $\boldsymbol{W}(\bar{\boldsymbol{u}})$ is an $(N, N-K)$ MWBE codebook.

Codebooks constructed from difference sets readily satisfy such a property, because for every (N, K, λ) difference set $\boldsymbol{u}$, the complement $\bar{\boldsymbol{u}}$ is also a difference set with parameters $(N, N-K, N-2K+\lambda)$ [1].

In the pursuit of equiangular vector sets, only the special case of $d = 2$ in family 1 has been reported in [13], while the general concept of difference sets has not been recognized therein. In other areas, difference sets have been used in designing binary code-division multiple-access (CDMA) signature sets [4].

IV. MODIFIED LLOYD SEARCH ALGORITHM AND COMPOSITE LOWER BOUND ON $I_{\max}$

When analytical constructions do not exist, we employ a numerical search method based on generalized Lloyd algorithm, which will turn out to improve considerably the achieved $I_{\max}$ of existing designs. We will further develop in this section a composite lower bound on the achievable $I_{\max}$, that is effective for any codebook size N.

A. Modified Lloyd Search (MLS) Algorithm

When codebook construction is not possible analytically, computationally complex numerical search has to be employed to find a near-optimal solution to (6). A brute-force computer search is used in [16], by choosing the codebook with the smallest $I_{\max}$ from a large group of randomly generated codebooks (with arbitrary alphabet). Alternatively, by realizing the fact that all MWBE codebooks have to be WBE in the first place, one can perform either an exhaustive or an incomplete search for an MWBE solution from a finite/infinite collection of WBE codebooks, as in [9] based on FFT matrices.

Seeking a more efficient numerical search, we consider a sphere vector quantizer, where the N codeword vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ and the random source input $\boldsymbol{g}$ are all $K \times 1$ complex vectors constrained on

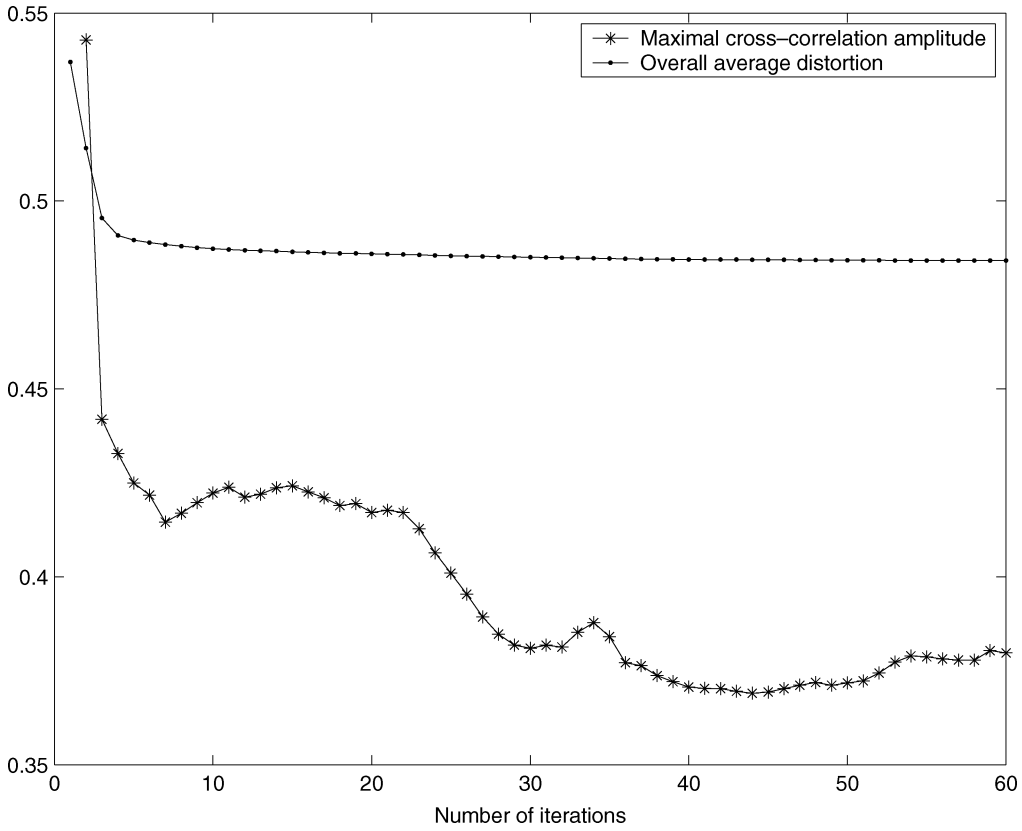

Fig. 1. Different convergence behaviors of $I_{\max}$ and the average distortion.

TABLE II
COMPARISON OF NUMERICAL SEARCH ALGORITHMS

N	K	Lloyd search	Brute-force search [17]	WBE codebook search [9]	Welch bound
8	2	0.82161	0.84152	0.92388	0.65465
16	3	0.67658	0.80793	0.78973	0.53748
16	4	0.45139	0.75252	0.58171	0.44721
64	4	0.74468	0.79731	0.79731	0.48795

the complex unit hypersphere Ω^K; and in particular, $\boldsymbol{g}$ is uniformly distributed on Ω^K (see [30] for extensions to correlated channels). Define a distortion metric between $\boldsymbol{w}_\ell$ and $\boldsymbol{g}$ as

$$d(\boldsymbol{g}, \boldsymbol{w}_\ell) := 1 - |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{g}|^2.$$

A "good" sphere vector quantizer shall find a codebook $\boldsymbol{W} := [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N]$ to minimize the average distortion; i.e., it will solve the optimization problem

$$\min_{\boldsymbol{W}} \ \mathbb{E}_{\boldsymbol{g}} \left[1 - |\boldsymbol{w}_*^{\mathcal{H}}(\boldsymbol{g}) \cdot \boldsymbol{g}|^2 \right], \quad \text{where} \quad \boldsymbol{w}_*(\boldsymbol{g}) = \arg \min_{\boldsymbol{w}_\ell \in \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N\}} 1 - |\boldsymbol{w}_\ell^{\mathcal{H}} \boldsymbol{g}|^2. \quad (27)$$

It has been shown in [19] and [16] that the optimal quantizer codebook minimizing the overall average distortion, serves as a good candidate toward minimizing $I_{\max}$.

The vector quantizer in (27) can be obtained through the generalized Lloyd algorithm [6]. The use of generalized Lloyd algorithm in codebook designs first appeared in [20] for the $K = 2$ case, and the extension to the general case is detailed in [30]. Normally, the generalized Lloyd algorithm is initialized with a random codebook $\boldsymbol{W}^{(0)}$, and every iteration of the algorithm produces a new codebook $\boldsymbol{W}^{(n)}$ with average distortion no larger than the previous iteration, as shown in Fig. 1. Also depicted therein is a typical refinement of $I_{\max}(\boldsymbol{W}^{(n)})$, which in general does not converge as the iteration proceeds. Normally, $I_{\max}$ keeps decreasing during the early iterations, but may fluctuate slightly afterwards. We, therefore, take as candidates all interim quantizer codebooks at the output of every Lloyd iteration. The one with sthe mallest $I_{\max}$ should provide a near-optimal solution to (6). To further reduce $I_{\max}$, the numerical search can be carried out many times with randomly different initializations. We summarize in the following the MLS algorithm for finding near-optimal MWBE codebooks.

Modified Lloyd search algorithm:

1. Randomly choose a large sample of source input realizations following a uniform distribution on Ω^K.
2. Perform the generalized Lloyd algorithm for a number of iterations as in [20], [30].
3. Record all interim quantizer codebooks at the output of every Lloyd iteration, and pick the one with the smallest $I_{\max}$.
4. Repeat steps 1–3 until a good codebook is found.

For several pairs of N and K, Table II lists the minimal $I_{\max}$ obtained by different search algorithms, with results of the brute-force search method taken directly from [17]. The codebooks generated by the MLS algorithm are available in [29]. While the search algorithms in [9] and [16] often exhibit relatively smaller computational complexity, the MLS algorithm generally produces better codebooks with considerably smaller $I_{\max}$. Since the codebook search is performed off-line, and is done once and for all, computational complexity should not be an issue as long as the overall offline cost is manageable.

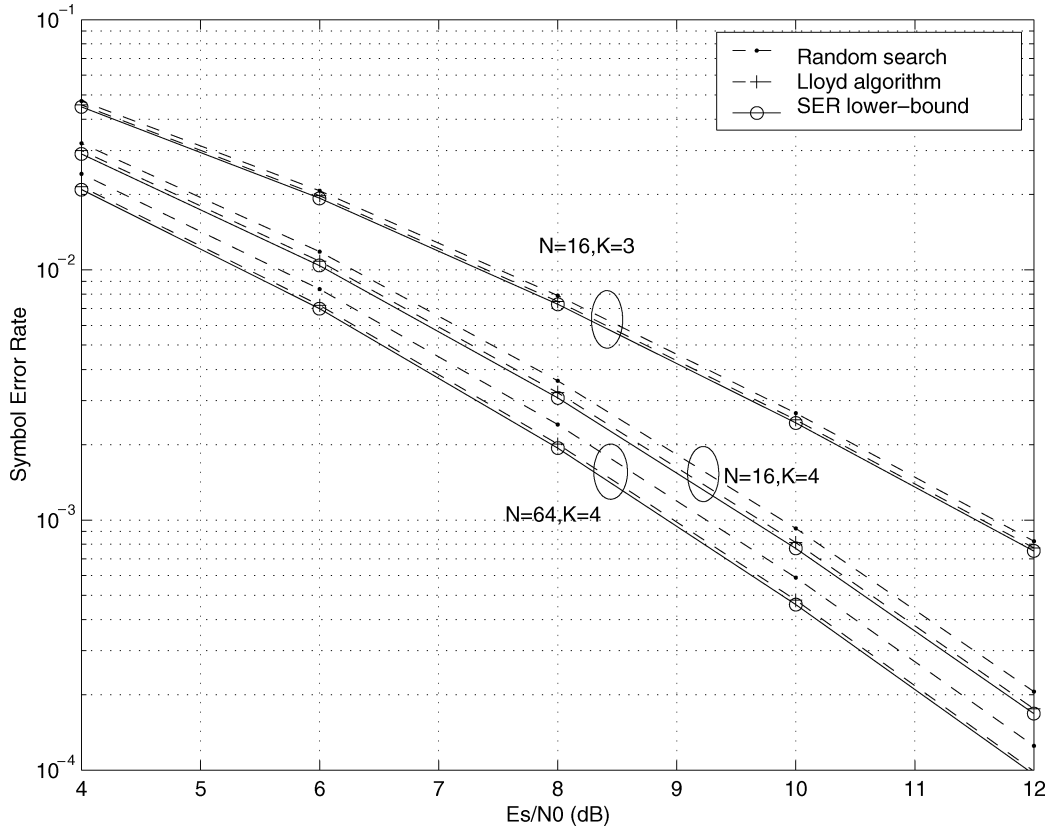

Fig. 2. SER performance of a transmit beamforming system with codebooks obtained using random search [17] and Lloyd algorithm.

The symbol error rate (SER) performance of a finite-rate transmit-beamforming system with codebooks from [17] has been tested in [31] against the SER lower bound derived therein which is applicable to any beamforming codebook. In [31], we have observed a small gap between the actual SER and the lower bound. Using the improved codebooks in Table II by the MLS algorithm, the actual SER improves considerably, and is almost indistinguishable from the SER lower bound, as depicted in Fig. 2 for a quaternary phase-shift keying (QPSK) constellation. This confirms that: i) the SER lower bound is accurate even for small (N, K) pairs; and ii) the MLS algorithm generates very good codebooks.

B. Composite Lower Bound on $I_{\max}$

The Welch lower bound on $I_{\max}$ is very useful for relatively small values of N, but becomes quite loose for large N. This can also bewitnessed from Table II, where the best $I_{\max}$ found through numerical search algorithms is often far away from the Welch bound. We are therefore motivated to look for a tighter bound on $I_{\max}$ that are useful for such cases.

Theorem 2: For an (N, K) codebook, $I_{\max}$ is lower-bounded by

$$I_{\max} \geq \max\left(\sqrt{\frac{N-K}{(N-1)K}},\ 1-2N^{-\frac{1}{K-1}}\right). \tag{28}$$

Proof: Apparently, half of this composite bound is due to the Welch bound. The other half

$$I_{\max} \geq 1-2N^{-\frac{1}{K-1}} \tag{29}$$

is mainly due to [19], although not explicitly recognized as a bound on $I_{\max}$ therein. In fact, by plugging [19, eq. (58)] into the second equality of [19, eq. (55)] and realizing the fact that any probability is always less than or equal to 1, we can easily obtain (29). □

Also based on results from [19], another bound becomes available [16, Theorem 2]

$$I_{\max} \geq \left(1-4N^{-1/(K-1)}\right)^{1/2} \tag{30}$$

which is always inferior to the new bound in (29).

For $K = 2$ and different values of N, Fig. 3 plots various bounds on $I_{\max}$ together with the minimum achieved $I_{\max}$ through analytical constructions or by using the MLS algorithm. As N increases, the Welch bound becomes increasingly looser while the new bound (29) gets increasingly tighter. The composite bound in (28) is effective throughout the range of N. Fig. 3 also demonstrates the efficiency of the MLS algorithm, as most codebooks obtained by numerical search yield $I_{\max}$ close to the bound. Also notice that bound (30) stays always below bound (29).

In the literature of Grassmannian line packing, the minimal chordal distance is defined as

$$\delta(\boldsymbol{W}) := \min_{1\leq \ell<\ell'\leq N} \sqrt{1-|\boldsymbol{w}_\ell^{\mathcal{H}}\boldsymbol{w}_{\ell'}|^2}. \tag{31}$$

As $\delta(\boldsymbol{W}) = \sqrt{1-I_{\max}^2}$, we infer from Theorem 2 the following result.

Corollary 1: The minimal chordal distance δ of Grassmannian line packing is upper-bounded by

$$\delta \leq \min\left(\sqrt{\frac{N(K-1)}{K(N-1)}},\ 2\sqrt{N^{-\frac{1}{K-1}}-N^{-\frac{2}{K-1}}}\right). \tag{32}$$

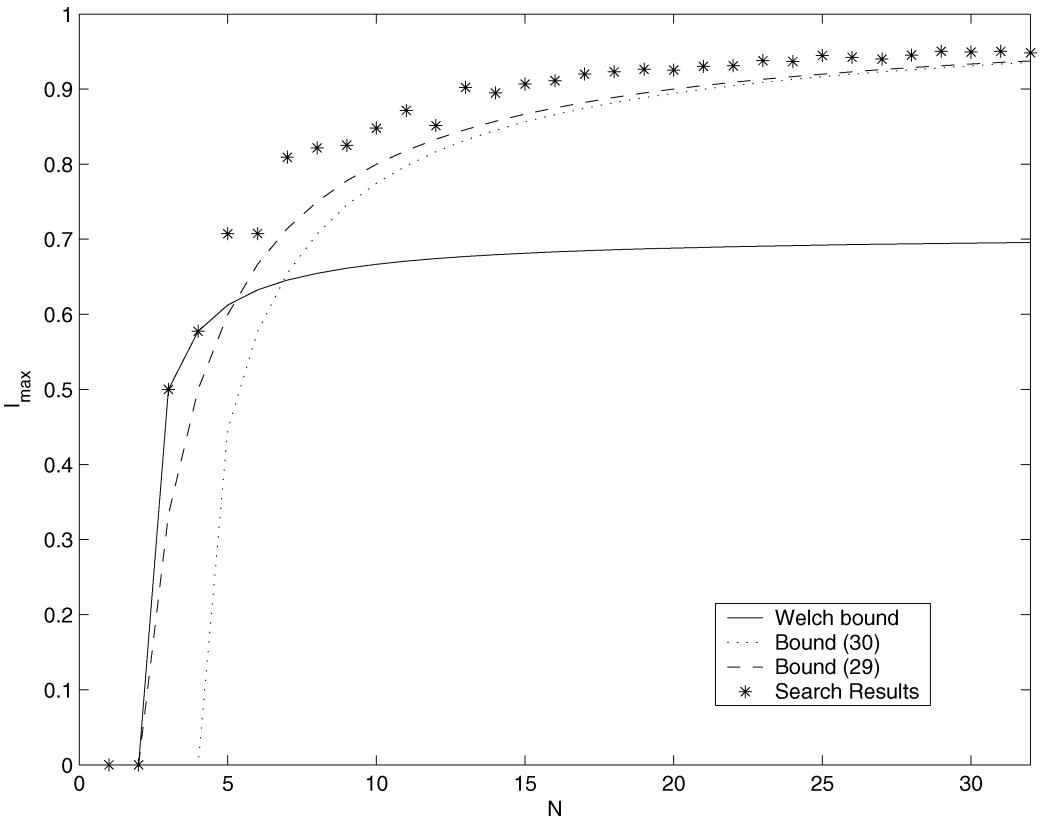

Fig. 3. Lower bounds and minimum achieved values of $I_{\max}$, for $K = 2$.

V. Conclusion

We investigated the problem of minimizing the maximum cross-correlation amplitude $I_{\max}$ of a complex (N, K) codebook, which is of interest in many applications. Relying on the notion of difference sets, we derived analytical constructions of optimal codebooks in certain cases. When analytical constructions are not available, we developed a modified Lloyd search (MLS) algorithm, which leads to considerable improvement of the achieved $I_{\max}$ over existing designs. We also developed a composite lower bound on the achievable $I_{\max}$ that is effective for any N. The analytical and numerical design methods are directly applicable to solving other closely related problems, such as Grassmannian line packing and the design of equi-angular line sets.

References

[1] C. J. Colbourn and J. H. Dinitz, *The CRC Handbook of Combinatorial Designs*. Boca Raton, FL: CRC, 1996.

[2] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[3] P. Delsarte, J. M. Goethals, and J. J. Seidel, "Bounds for systems of lines and Jacobi polynomials," *Philips Res. Repts.*, vol. 30, no. 3, pp. 91–105, 1975.

[4] C. Ding, M. Golin, and T. Klove, "Meeting the Welch and Karystinos-Pados bounds on DS-CDMA binary signature sets," *Des., Codes Cryptogr.*, vol. 30, no. 1, pp. 73–84, Aug. 2003.

[5] J. H. Dinitz and D. R. Stinson, *Contemporary Design Theory: A Collection of Surveys*. New York: Wiley, 1992.

[6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic, 1992.

[7] V. K. Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 203–233, 2001.

[8] M. Hall, *Combinatorial Theory*. London, U.K.: Blaisdell, 1967.

[9] B. M. Hochwald, T. L. Marzetta, T. L. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[10] V. Ipatov, "On the Karystinos-Pados bounds and optimal binary DS-CDMA signature ensembles," *IEEE Commun. Lett.*, vol. 8, no. 2, pp. 81–83, Feb. 2004.

[11] G. N. Karystinos and D. A. Pados, "New bounds on the total squared correlation and optimum design of DS-CDMA binary signature sets," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 48–51, Jan. 2003.

[12] ——, "The maximum squared correlation, sum capacity, and total asymptotic efficiency of minimum total-squared-correlation binary signature sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 348–355, Jan. 2005.

[13] H. König, "Cubature formulas on spheres," in *Advances in Multivariate Approximation*. New York: Wiley, 1999, pp. 201–211.

[14] La Jolla Difference Set Repository [Online]. Available: http://www.ccr-west.org/diffsets/diff_sets/

[15] P. W. H. Lemmens and J. J. Seidel, "Equiangular lines," *J. Algebra*, vol. 24, pp. 494–512, 1973.

[16] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[17] D. J. Love. Grassmannian Subspace Packing. [Online]. Available: http://dynamo.ecn.purdue.edu/ djlove/packings/one_d/

[18] J. L. Massey and T. Mittlelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II: Methods in Communication, Security and Computer Sciences*. Heiderlberg/New York: Springer-Verlag, 1993, pp. 63–78.

[19] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. A. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[20] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[21] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[22] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.

[23] D. V. Sarwate, "Meeting the Welch bound with equality," in *Sequences and their Applications*. New York: Springer-Verlag, 1999, pp. 79–102.

[24] P. W. Shor and N. J. A. Sloane, "A family of optimal packings in Grassmannian manifolds," *J. Alg. Combin.*, no. 7, pp. 157–163, 1998.

[25] N. J. A. Sloane. Packings in Grassmannian Spaces. [Online]. Available: http://www.research.att.com/~njas/grass/index.html

[26] T. Strohmer and R. W. Heath Jr., "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmonic Anal.*, vol. 14, no. 3, pp. 257–275, 2003.

[27] D. R. Stinson, *Combinatorial Designs: Constructions and Analysis*. New York: Springer-Verlag, 2004.

[28] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 397–399, May 1974.

[29] P. Xia. Codebook Constructions Using the Lloyd Algorithm. [Online]. Available: http://www.ece.umn.edu/users/pfxia/codebook/lloyd

[30] P. Xia and G. B. Giannakis, "Design and analysis of transmit beamforming based on limited-rate feedback," in *Proc. IEEE Vehicular Technology Conf.*, Los Angeles, CA, Sep. 26–29, 2004.

[31] S. Zhou, Z. Wang, and G. B. Giannakis, "Performance analysis of transmit-beamforming with finite-rate feedback," *IEEE Trans. Commun.*. Also available [Online] at http://spincom.ece.umn.edu/papers04/twc05szhou.pdf., to be published.