Spectral Distribution of Product of Pseudorandom Matrices Formed From Binary Block Codes

Behtash Babadi, *Student Member, IEEE*, and Vahid Tarokh, *Fellow, IEEE*

Dedicated to The Memory of Thomas M. Cover

***Abstract*—Let $\mathbf{A} \in \{-1,1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b \times n}$ be two matrices whose rows are drawn i.i.d. from the codewords of the binary codes $\mathcal{C}^a$ and $\mathcal{C}^b$ of length n and dual distances d'^a and d'^b, respectively, under the mapping $0 \mapsto 1$ and $1 \mapsto -1$. It is proven that as $n \to \infty$ with $y_a := n/N_a \in (0,\infty)$ and $y_b := n/N_b \in (0,\infty)$ fixed, the empirical spectral distribution of the matrix $\mathbf{A}\mathbf{B}^*/\sqrt{N_a N_b}$ resembles a universal distribution (closely related to the distribution function of the free multiplicative convolution of two members of the Marchenko–Pastur family of densities) in the sense of the Lévy distance, if the asymptotic dual distances of the underlying binary codes are large enough. Moreover, an explicit upper bound on the Lévy distance of the two distributions in terms of y_a, y_b, d'^a, and d'^b is given. Under mild conditions, the upper bound is strengthened to the Kolmogorov distance of the underlying distributions. Numerical studies on the empirical spectral distribution of the product of random matrices from BCH and Gold codes are provided, which verify the validity of this result.**

***Index Terms*—Binary block codes, free probability theory, Lévy distance, Marchenko–Pastur law, pseudorandom matrices, random matrix theory.**

I. Introduction

THE elegant theory of random matrices has attracted a considerable amount of attention in recent years. Random matrix theory studies the emergence of deterministic collective behavior from a large collection of random elements in the domain of matrices. Born in theoretical physics, and nurtured by mathematicians, this theory has also found its home in several other disciplines of science such as economics [9] and communication theory [16]. Most of the important results of random matrix theory rely on the fact that the underlying matrix elements have an i.i.d. structure. In this paper, however, we study the spectral behavior of certain pseudorandom matrices, and relate their spectral behavior to their fully random counterparts.

More explicitly, let $\mathcal{C}$ be an $(n, \mathcal{M}, d)$ binary block code of length n, size $\mathcal{M}$, and minimum Hamming distance d over $\mathrm{GF}(2)^n$. Let $\mathbf{c}_i := (c_{i1}, c_{i2}, \ldots, c_{in})$ be a codeword in $\mathcal{C}$. We say that an $N \times n$ random matrix $\mathbf{\Phi}$ is *based on* a binary block code $\mathcal{C}$, if for a randomly i.i.d. drawn set of codewords of $\mathcal{C}$, $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$, we have $\Phi_{ij} = (-1)^{c_{ij}}$, for all $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n$. Let $\mathbf{A} \in \{-1,1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b \times n}$ be two random matrices based on the binary codes $\mathcal{C}^a$ and $\mathcal{C}^b$ of length n, respectively. We study the empirical spectral distribution of the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ and show that if the dual distances of the underlying codes are sufficiently large, the asymptotic empirical spectral distribution resembles a deterministic universal distribution. This universal distribution pertains to the case where the elements of $\mathbf{A}$ and $\mathbf{B}$ are drawn i.i.d. from the set $\{-1, 1\}$, and is closely related to the free multiplicative convolution of two of the members of the Marchenko–Pastur family of densities [15], [18]. We upper bound the Lévy distance of the asymptotic empirical spectral distribution of the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ to the aforementioned universal distribution as a function of $y_a := n/N_a$, $y_b := n/N_b$, and the dual distances of $\mathcal{C}^a$ and $\mathcal{C}^b$. This result is also strengthened to an upper bound on the Kolmogorov distance of the underlying distributions under mild conditions. Numerical experiments on BCH and Gold codes are provided, which confirm the theoretical result of this paper. Although independently interesting from the viewpoint of random matrix theory, the result of this paper suggests a criterion for evaluating the cross randomness of two set of binary codes or sequences.

The outline of this paper follows next. In Section II, we introduce the notation and state the main theorems of the paper followed by a discussion of the main results. The detailed proofs of the main theorems are presented in Section III. Finally, numerical experiments in Section IV conclude the paper.

II. Main Result

A. Results and Notation From Coding Theory

Before presenting the main result, we introduce the notation and state some preliminary definitions.

A $(n, \mathcal{M}, d)$ binary code $\mathcal{C}$ is defined as a set of $\mathcal{M}$ binary n-tuples such that any two such n-tuples differ in at least d places, with d being the largest number with this property. The Hamming weight of an n-tuple $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathrm{GF}(2)^n$, denoted by $\mathrm{wt}(\mathbf{u})$, is defined as the number of nonzero elements of $\mathbf{u}$.

Manuscript received October 26, 2011; revised September 26, 2012; accepted October 02, 2012. This work was supported in part by a fellowship from the John Simon Guggenheim Memorial Foundation. The material in this paper was presented in part at the 2011 Allerton Conference on Communication, Control, and Computing, Monticello, IL [3].

B. Babadi is with the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: behtash@nmr.mgh.harvard.edu).

V. Tarokh is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: vahid@seas.harvard.edu).

Communicated by A. Moustakas, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2223812

Consider the group algebra over $\mathrm{GF}(2)^n$, in which the code $\mathcal{C}$ is represented by the element

$$\mathcal{C} := \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} c_{\mathbf{u}} t^{\mathbf{u}} \tag{1}$$

where $t^{\mathbf{u}} := t_1^{u_1} t_2^{u_2} \cdots t_n^{u_n}$, and

$$c_{\mathbf{u}} := \begin{cases} 1, & \text{if } \mathbf{u} \in \mathcal{C} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

For all $\mathbf{u}, \mathbf{v} \in \mathrm{GF}(2)^n$, the product of $t^{\mathbf{u}}$ and $t^{\mathbf{v}}$ is defined as

$$t^{\mathbf{u}} t^{\mathbf{v}} := t^{\mathbf{u}\oplus\mathbf{v}} = t_1^{u_1\oplus v_1} t_2^{u_2\oplus v_2} \cdots t_n^{u_n\oplus v_n} \tag{3}$$

where $\oplus$ denotes the binary addition. For a binary n-tuple $\mathbf{u} \in \mathrm{GF}(2)^n$, let $\chi_{\mathbf{u}}$ be the character mapping

$$\chi_{\mathbf{u}}(t^{\mathbf{v}}) = (-1)^{\mathbf{u}\cdot\mathbf{v}} \tag{4}$$

with $\mathbf{u}\cdot\mathbf{v} := \sum_i u_i v_i \bmod 2$, for all $\mathbf{v} \in \mathrm{GF}(2)^n$. Suppose that for a code $\mathcal{C}$, corresponding to the element of the group algebra given by (1), we have

$$\mathcal{M} = \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} c_{\mathbf{u}} \neq 0. \tag{5}$$

Now, consider

$$\mathcal{D} := \frac{1}{\mathcal{M}}\mathcal{C}^2 = \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} d_{\mathbf{u}} t^{\mathbf{u}}. \tag{6}$$

The *distance distribution* of the code $\mathcal{C}$ is defined as the set $\{B_0, B_1, \ldots, B_n\}$, where

$$B_i := \sum_{\mathrm{wt}(\mathbf{u})=i} d_{\mathbf{u}}. \tag{7}$$

The *transformed* distance distribution of the code $\mathcal{C}$ under the character mapping is given by the set $\{B'_0, \ldots, B'_n\}$, where

$$B'_j := \frac{1}{\mathcal{M}} \sum_{\mathrm{wt}(\mathbf{u})=j} \chi_{\mathbf{u}}(\mathcal{D}). \tag{8}$$

Finally, the *dual distance* of the code $\mathcal{C}$ is defined as d' such that $B'_i = 0$ for $1 \leqslant i \leqslant d'-1$ and $B'_{d'} \neq 0$ [7].

B. Results and Notation From Random Matrix Theory

Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of the matrix $\mathbf{X} \in \mathbb{R}^{N\times N}$, and let

$$\mu_{\mathbf{X}} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \tag{9}$$

denote the spectral measure of $\mathbf{X}$, where δ_x denotes the Dirac measure. Let $M_{\mathbf{X}}(x)$ denote the distribution function associated with the measure $\mu_{\mathbf{X}}$. Finally, let $m_{\mathbf{X}}^{(\ell)}$ denote the ℓth moment of $\mu_{\mathbf{X}}$. The Stieltjes transform of the density μ is defined as

$$s_\mu(z) := \int \frac{1}{z-x}\mu(dx) \tag{10}$$

for $\{z \in \mathbb{C} | \Im\{z\} \geqslant 0\}$. In particular, we denote by $M_{\mathsf{MP}}(x; y)$ the distribution corresponding to the Marchenko–Pastur measure $\mu_{\mathsf{MP}}(x; y)$ whose density is given by

$$\frac{d\mu_{\mathsf{MP}}(x;y)}{dx} := \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\,1_{(a\leqslant x\leqslant b)} \tag{11}$$

with $a = (1-\sqrt{y})^2$ and $b = (1+\sqrt{y})^2$. It can be shown that the Stieltjes transform of the Marchenko–Pastur density, denoted by $s_{\mathsf{MP}}(z; y)$, satisfies the following quadratic equation [8]:

$$yzs_{\mathsf{MP}}^2(z;y) - (1-y-z)s_{\mathsf{MP}}(z;y) + 1 = 0. \tag{12}$$

Let $\mathbf{A} \in \{-1,1\}^{N_a\times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b\times n}$. Let $y_a := \frac{n}{N_a} \in (0,\infty)$ and $y_b := \frac{n}{N_b} \in (0,\infty)$ be fixed numbers. Consider the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ given by $\mathscr{G} := \frac{1}{N_a N_b}\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*$. If the elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ are i.i.d. drawn from the set $\{-1,1\}$, it can be shown that the spectral measure of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ tends to a deterministic limit almost surely (see [4, Th. 1.1]). In order to identify this measure, we first consider the closely related measure corresponding to the asymptotic empirical spectral distribution of the matrix $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$. Clearly, the asymptotic spectral densities of the matrices $\frac{1}{N_a}\mathbf{A}^*\mathbf{A}$ and $\frac{1}{N_a}\mathbf{B}^*\mathbf{B}$ are given by $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$, respectively. Free probability theory implies that the asymptotic spectral density of the product $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is given by

$$\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \mu_{\mathsf{MP}}(x; y_a) \boxtimes \mu_{\mathsf{MP}}(x; y_b) \tag{13}$$

where $\mu_1 \boxtimes \mu_2$ denotes the free multiplicative convolution of the densities μ_1 and μ_2 [15], [18]. We also denote the distribution function and Stieltjes transform of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ by $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, respectively. Using the polynomial method for algebraic random matrices [11], one can directly characterize $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, without the need for evaluating the free multiplicative convolution, as given by the following lemma.

Lemma 2.1: The Stieltjes transform of the free multiplicative convolution of two densities from the Marchenko–Pastur family with parameters y_a and y_b, denoted by $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, satisfies the following cubic equation:

$$\begin{aligned} &y_a y_b z^2 \overline{s}^3_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) + z(2y_a y_b - y_a - y_b)\overline{s}^2_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) \\ &- (z - (1-y_a)(1-y_b))\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) - 1 = 0. \end{aligned} \tag{14}$$

Proof: The proof is based on the main result of [14] which states that the Stieltjes transform of the asymptotic spectral distribution of $\frac{1}{N}\mathbf{X}^*\mathbf{X}\mathbf{T}$ (denoted by $s(z)$) for $\mathbf{X} \in \mathbb{R}^{n\times N}$ having i.i.d. unit variance elements and $\mathbf{T} \in \mathbb{R}^{n\times n}$ being a random nonnegative definite Hermitian with asymptotic spectral distribution $H(x)$, and $n/N \to y$, satisfies

$$s(z) = \int \frac{1}{x\big(1-y-yzs(z)\big)-z} dH(x). \tag{15}$$

Specializing (15) to the case of $N = N_a$, $\mathbf{X} = \mathbf{A}$, and $\mathbf{T} = \frac{1}{N_b}\mathbf{B}^*\mathbf{B}$ and noting that the Stieltjes transform of $H(x)$ in this case satisfies (12), it is not hard to show that $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ corresponds to an algebraic random matrix [11] and satisfies (12) with y, $s_{\mathsf{MP}}(z; y)$ and z replaced by y_a, $(1 - y_b - y_b z\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b))\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ and $z/(1 - y_b - y_b z\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b))$, respectively. The latter establishes the result of the lemma. ■

The solution to (14) can be obtained using the standard cubic function root formula. Since the coefficient of the leading term of $\overline{s}_{\boxtimes^2\mathsf{MP}}$ in the aforementioned equation, $y_a y_b z^2$, has roots at $z = 0$, the nonatomic portion of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}$ is given by the positive imaginary part of $\overline{s}_{\boxtimes^2\mathsf{MP}}$ with a scaling of $1/\pi$ [11]. The support of the nonatomic portion of the density is the range of x where the discriminant of the cubic equation is negative, and hence, the cubic equation has two complex conjugate roots. A closed-form solution for the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}$ can be explicitly obtained from (14), but is omitted for brevity. Moreover, (14) is a powerful tool for numerical computation of the density.

The following lemma characterizes the possible point mass of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ at $x = 0$.

Lemma 2.2: The density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ has a point mass of $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$ at $x = 0$.

Proof: Note that the possible point mass of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ at $x = 0$ is given by the coefficient of $-1/z$ in the Puiseux expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ in terms of the powers of z and $1/z$ [11]. Let κ denote the coefficient of $-1/z$ in the Puiseux expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$. By inspecting the expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ from (14), we get

$$y_a y_b \kappa^3 - (2y_a y_b - y_a - y_b)\kappa^2 + (1 - y_a)(1 - y_b)\kappa = 0 \quad (16)$$

which has the roots 0, $1 - 1/y_a$, and $1 - 1/y_b$. Hence, the point mass can be identified as one of these roots. Now, suppose that $\mathbf{A} \in \{-1, 1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1, 1\}^{N_b \times n}$ are random matrices with elements drawn i.i.d. from $\{-1, 1\}$. From elementary linear algebra, the rank of $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is no greater than $\min\{n, N_a, N_b\}$. Hence, the point mass of the asymptotic density of $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is no smaller than $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$. The latter fact and the possible values of the point mass obtained from the Puiseux expansion imply that the point mass is indeed equal to $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$. ■

Now, let $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ denote the asymptotic spectral density of the matrix $\frac{1}{N_a N_b}\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*$. By an elementary linear algebraic argument, one can show that $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is related to $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ by the following transformation [14]:

$$\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \big(1 - y_a\big)\delta_0 + y_a\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (17)$$

Similarly, the distribution function corresponding to $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is denoted by $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and is given by

$$M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \big(1 - y_a\big) + y_a\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (18)$$

The Lévy distance between two distribution functions $M_1(x)$ and $M_2(x)$ is defined as

$$\begin{aligned}&\mathscr{L}\big(M_1(x), M_2(x)\big)\\ &:= \inf\Big\{\epsilon > 0 \Big| M_1(x - \epsilon) - \epsilon \leqslant M_2(x) \leqslant M_1(x + \epsilon) + \epsilon,\\ &\qquad \forall x \in \mathbb{R}\Big\}. \end{aligned} \quad (19)$$

The Kolmogorov distance between two distribution functions $M_1(x)$ and $M_2(x)$ is defined as

$$\sup_x \Big|M_1(x) - M_2(x)\Big|. \quad (20)$$

C. Main Result

The main theorem of this paper is the following:

Theorem 2.3: Let $\{\mathcal{C}_n^a\}_{n=1}^\infty$ and $\{\mathcal{C}_n^b\}_{n=1}^\infty$ be two sequences of binary block codes of length n. Let $d_n'^a$ and $d_n'^b$ denote the dual distances of $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $\{N_a^n\}_{n=1}^\infty$ and $\{N_b^n\}_{n=1}^\infty$ be two sequences such that $n/N_a^n \to y_a \in (0, \infty)$ and $n/N_b^n \to y_b \in (0, \infty)$, as $n \to \infty$. Let $\big\{\mathbf{A}_n \in \{-1, 1\}^{N_a^n \times n}\big\}_{n=1}^\infty$ and $\big\{\mathbf{B}_n \in \{-1, 1\}^{N_b^n \times n}\big\}_{n=1}^\infty$ be two sequences of random matrices, where $\mathbf{A}_n$ and $\mathbf{B}_n$ are based on $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $M_{\mathscr{G}_n}(x)$ denote the spectral distribution function of the Gram matrix of $\frac{1}{\sqrt{N_a^n N_b^n}}\mathbf{A}_n\mathbf{B}_n^*$ and $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ denote the distribution of the free multiplicative convolution of the Marchenko–Pastur densities $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$. Let $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := (1 - y_a) + y_a\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Let r_n be the greatest even integer less than or equal to $\min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$, and $r := \liminf_n r_n$. Then, we have

$$\limsup_n \mathscr{L}\big(M_{\mathscr{G}_n}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\big) \leqslant \mathscr{A}\frac{1}{r} + \mathscr{B}\frac{\ln r}{r}$$

almost surely, where $\mathscr{A}$ and $\mathscr{B}$ are only functions of y_a, y_b, and r (explicitly given in this paper), and are bounded in r.

The statement of Theorem 2.3 can be strengthened as follows.

Theorem 2.4: Let y_a, y_b, $M_{\mathscr{G}_n}(x)$, $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, and r be as in Theorem 2.3. Suppose that $y_a \in (0, 1) \cup (1, \infty)$ and $y_b \in (0, 1) \cup (1, \infty)$. Then, we have

$$\limsup_n \Big|M_{\mathscr{G}_n}(x) - M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\Big| \leqslant \mathscr{C}\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

for all x almost surely, where $\mathscr{C}$ is a function of y_a, y_b, and r (implicitly given in this paper), and is bounded in r.

D. Discussion of the Main Result

Theorems 2.3 and 2.4 state that the empirical spectral distribution of the Gram matrix of the random matrix $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$, with $\mathbf{A}$ and $\mathbf{B}$ based on binary block codes $\mathcal{C}^a$ and $\mathcal{C}^b$, respectively, resembles the universal empirical spectral distribution $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ in the sense of Lévy and Kolmogorov distance, respectively, as $n \to \infty$, provided that the dual distances of the codes $\mathcal{C}^a$ and $\mathcal{C}^b$ are large enough.

Previously, the authors proved the following result, which is stated as follows for completeness [2].

Theorem 2.5 (Main Theorem of [2]): Consider a sequence of $[n, k_n, d_n]$ binary linear block codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$. Let $d_n^{\perp}$ denote the dual distance of $\mathcal{C}_n$. Let $\mathbf{\Phi}_{\mathcal{C}_n}$ be a $p \times n$ random matrix based on $\mathcal{C}_n$, $\mathcal{G}_{\mathcal{C}_n}$ denote the Gram matrix of $\frac{1}{\sqrt{n}}\mathbf{\Phi}_{\mathcal{C}_n}$, and $M_{\mathcal{C}_n}(x)$ denote the empirical spectral distribution of $\mathcal{G}_{\mathcal{C}_n}$. Finally, let r_n be the greatest even integer less than or equal to $[(d_n^{\perp}-1)/2]$, and let $r := \liminf_n r_n$. Then, as $n \to \infty$ with $y := p/n \in (0,1)$ fixed, we have

$$\limsup_n \left| M_{\mathcal{C}_n}(x) - M_{\mathsf{MP}}(x) \right| \leqslant c(y,r)\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

almost surely for all z, where $c(y,r)$ is a bounded function of r (explicitly given in [2]).

The aforementioned theorem states that the asymptotic empirical spectral distribution of a random matrix from a binary linear code $\mathcal{C}$ resembles the Marchenko–Pastur distribution in the sense of Kolmogorov distance, provided that the dual distance of $\mathcal{C}$ is large enough [2]. The proof of the main theorem of [2] is simpler to those of this paper; however, the result of [2] is stronger than Theorem 2.3, since the Lévy distance is weaker than the Kolmogorov distance. Theorem 2.4, however, provides an upper bound on the Kolmogorov distance of the underlying distribution, under the more restrictive assumption that $y_a, y_b \neq 1$. As will be shown in the proof of Theorem 2.3, this assumption implies boundedness of the density $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Moreover, the functionality of the upper bound of Theorem 2.4 with respect to y_a and y_b is implicit, due to the complicated analytic expression of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, whereas the upper bound of Theorem 2.3 is given explicitly as a function of y_a, y_b, and r.

Furthermore, the proof of Theorem 2.3 (and similarly Theorem 2.4) of this paper offers more difficulty in identifying and bounding the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, which are more involved than those of $\mu_{\mathsf{MP}}(x; y)$. Finally, the results of this paper hold for binary (possibly nonlinear) codes in general, and more importantly, the constraint of $y \in (0,1)$ in [2] is strengthened to $y_a, y_b \in (0, \infty)$ in this paper.

III. Proof of the Main Results

We present a proof along the lines of the proof of the main result of [2]. We therefore need to establish a number of lemmas. First, we state the following lemma, known as the Zolotarev's inequality, for bounding the Lévy distance of two distributions.

Lemma 3.1 (Zolotarev's Inequality): Let $M_1(x)$ and $M_2(x)$ be two distribution functions, with characteristic functions $m_1(t)$ and $m_2(t)$, respectively. Then, for all $T > 1.3$, we have

$$\mathscr{L}\big(M_1(x), M_2(x)\big) < \frac{1}{\pi}\int_0^T \Big| m_1(t) - m_2(t) \Big| \frac{dt}{t} + 2e\frac{\ln T}{T}.$$

Proof: The proof uses smoothing techniques to bound the Lévy distance of the two distributions, and is given in [20]. ■

Similarly, when the density of $M_2(x)$ is bounded, we have the following lemma for bounding the absolute distance of $M_1(x)$ and $M_2(x)$.

Lemma 3.2: Let $M_1(x)$ be a probability distribution with vanishing expectation and characteristic function $m_1(t)$. Suppose that $M_1(x) - M_2(x)$ vanishes at $\pm\infty$ and that $M_2(x)$ has a derivative $\mu_2(x)$ such that $|\mu_2(x)| \leqslant A$. Finally, suppose that $\mu_2(x)$ has a continuously differentiable Fourier transform $m_2(t)$ such that $m_2(0) = 1$ and $m_2'(0) = 0$. Then, for all z and $T > 0$, we have

$$|M_1(x) - M_2(x)| \leqslant \frac{1}{\pi}\int_{-T}^{T} \left| \frac{m_1(t) - m_2(t)}{t} \right| dt + \frac{24A}{\pi T}. \quad (21)$$

Proof: The proof uses smoothing techniques to cope with the fact that the underlying density of $M_1(x)$ may not enjoy sufficient smoothness properties. The detailed proof can be found in [5] and is thus omitted for brevity. ■

The following lemma establishes an important combinatorial property of a binary code $\mathcal{C}$.

Lemma 3.3 [7]: Any set of $r \leqslant d' - 1$ columns of the code $\mathcal{C}$ contains each binary r-tuple exactly $\mathcal{M}/2^r$ times, and d' is the largest number with this property.

Proof: By the definition of d' and the properties of the character mapping, we have $\chi_{\mathbf{u}}(\mathcal{C}) = 0$ for all $\mathbf{u}$ with $\mathrm{wt}(\mathbf{u}) = 1, 2, \ldots, d'-1$. For $\mathrm{wt}(\mathbf{u}) = 1$, this implies that each component of the codewords takes the values of 0 and 1 a total of $\mathcal{M}/2$ times each. For $\mathrm{wt}(\mathbf{u}) = 2$, this implies that any set of two components of the codewords takes the combinations $00, 01, 10, 11$ a total of $\mathcal{M}/4$ times each, etc. Hence, any set of $d'-1$ components of the codewords takes all the possible $(d'-1)$-tuples a total of $\mathcal{M}/2^{(d'-1)}$ times each. Since $B_{d'}' \neq 0$, there must be a codeword $\mathbf{u}$ with $\mathrm{wt}(\mathbf{u}) = d'$ such that $\chi_{\mathbf{u}}(\mathcal{C}) \neq 0$. ■

The following lemma establishes the almost sure convergence of the first $d' := \liminf_n \min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$ moments of $\mu_{\mathscr{G}_n}$ to those of $\mu_{\boxtimes^2\mathsf{MP}}$, as $n \to \infty$.

Lemma 3.4: Consider the sequences $\{N_a^n\}_{n=1}^{\infty}$ and $\{N_b^n\}_{n=1}^{\infty}$, such that $n/N_a^n \to y_a \in (0,\infty)$ and $n/N_b^n \to y_b \in (0,\infty)$, as $n \to \infty$. Consider two sequence of binary block codes $\{\mathcal{C}_n^a\}_{n=1}^{\infty}$ and $\{\mathcal{C}_n^b\}_{n=1}^{\infty}$, where $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$ are of length n and dual distances $d_n'^a$ and $d_n'^b$, respectively. Let $\{\mathbf{A}_n \in \{-1,1\}^{N_a^n \times n}\}_{n=1}^{\infty}$ and $\{\mathbf{B}_n \in \{-1,1\}^{N_b^n \times n}\}_{n=1}^{\infty}$ be two sequence of random matrices, such that $\mathbf{A}_n$ and $\mathbf{B}_n$ are based on the binary codes $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $\mu_{\mathscr{G}_n}$ denote the spectral measure of the Gram matrix of $\frac{1}{\sqrt{N_a^n N_b^n}}\mathbf{A}_n\mathbf{B}_n^*$. Let

$$m_{\mathscr{G}_n}^{(\ell)} := \int x^{\ell}\mu_{\mathscr{G}_n}(dx) \quad (22)$$

$$m_{\boxtimes^2\mathsf{MP}}^{(\ell)}(y_a, y_b) := \int x^{\ell}\mu_{\boxtimes^2\mathsf{MP}}(dx; y_a, y_b) \quad (23)$$

and

$$m_{\mathsf{MP}}^{(\ell)}(y) := \int z^{\ell}\mu_{\mathsf{MP}}(dx; y) = \sum_{i=0}^{\ell-1} \frac{y^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i} \quad (24)$$

be the ℓth moment of the spectral measures $\mu_{\mathscr{G}_n}$, $\mu_{\boxtimes^2\mathsf{MP}}$, and μ_{MP}, respectively. Then, for

$$1 \leq \ell \leq d' := \liminf_n \min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$$

we have

$$m^{(\ell)}_{\mathscr{G}_n} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) = \sum_{i=1}^{\ell} y_a^{\ell-i} \sum_{\substack{k_1+k_2+\cdots+k_i=\ell-i+1 \\ k_1+2k_2+\cdots+ik_i=\ell}} \frac{\ell!}{i!} \prod_{j=1}^{i} \frac{m^{(j)}_{\mathsf{MP}}{}^{k_j}(y_b)}{k_j!}$$

almost surely.

Proof: Let $P_{\mathscr{G}_n}$ be the probability measure induced by the i.i.d. selection of N_a^n and N_b^n codewords from $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. For notational convenience, we drop the dependence on n, where there is no ambiguity. From an application of the Borel–Cantelli Lemma (see, e.g., [1] or [17]), it is enough to show

$$\mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \tag{25}$$

and

$$\mathbb{E}_{P_{\mathscr{G}}}\left\{\left|m^{(\ell)}_{\mathscr{G}} - \mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\}\right|^4\right\} = \mathcal{O}\left(\frac{1}{n^2}\right) \tag{26}$$

where $\mathbb{E}_{P_{\mathscr{G}}}$ is the expectation with respect to $P_{\mathscr{G}}$.

Thus, we need to prove that the average of the first d' moments of the measure $\mu_{\mathscr{G}}$ coincides with those of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ almost surely and that the fourth central moment of these moments drops as $1/n^2$. In what follows, we drop the subscript n for notational convenience. The ℓth moment of $\mu_{\mathscr{G}}$ can be written as

$$\begin{aligned}
\mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\} &= \mathbb{E}_{P_{\mathscr{G}}}\left\{\int x^\ell \mu_{\mathscr{G}}(dx)\right\} = \mathbb{E}_{P_{\mathscr{G}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\lambda_i^\ell\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\mathbb{E}_{P_{\mathscr{G}}}\left\{\mathrm{Tr}\left\{(\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*)^\ell\right\}\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\mathbb{E}_{P_{\mathscr{G}}}\left\{\mathrm{Tr}\left\{(\mathbf{A}^*\mathbf{A}\mathbf{B}^*\mathbf{B})^\ell\right\}\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\sum_{\mathcal{I},\mathcal{J}}\mathbb{E}_{P_{\mathscr{G}}}\left\{(-1)^{s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}}\right\}
\end{aligned}$$

where

$$\mathcal{I}^a := \{i_t^a\}_{t=1}^{\ell} \in \{1, 2, \ldots, N_a\}^\ell \tag{27}$$
$$\mathcal{J}^a := \{j_t^a\}_{t=1}^{\ell} \in \{1, 2, \ldots, n\}^\ell \tag{28}$$
$$\mathcal{I}^b := \{i_t^b\}_{t=1}^{\ell} \in \{1, 2, \ldots, N_b\}^\ell \tag{29}$$
$$\mathcal{J}^b := \{j_t^b\}_{t=1}^{\ell} \in \{1, 2, \ldots, n\}^\ell \tag{30}$$

and

$$s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b} := c_{i_1^a j_1^a} \oplus c_{i_2^a j_1^a} \oplus c_{i_1^b j_1^b} \oplus c_{i_2^b j_1^b} \oplus \cdots \oplus c_{i_\ell^b j_\ell^b} \oplus c_{i_1^a j_\ell^b}$$

with $\oplus$ denoting the binary addition and the summation running over all $\mathcal{I}^a \in \{1, 2, \ldots, N_a\}^\ell$, $\mathcal{I}^b \in \{1, 2, \ldots, N_b\}^\ell$, and $\mathcal{J}^a, \mathcal{J}^b \in \{1, 2, \ldots, n\}^\ell$. Note that $s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}$ corresponds to a directed cycle of length 4ℓ on a complete quadripartite graph $G = (X^a \cup Y^a \cup X^b \cup Y^b, E)$ with $X^a := \{1, 2, \ldots, N_a\}$, $X^b := \{1, 2, \ldots, N_b\}$, and $Y^a = Y^b := \{1, 2, \ldots, n\}$, where, for instance, $c_{i^a j^a}$ corresponds to an edge from node $i^a \in \mathcal{I}^a$ to node $j^a \in \mathcal{J}^a$.

Generalizing the proofs of Wigner [19] and Marchenko and Pastur [8], Bai *et al.* [4] have elegantly counted the number of such directed cycles, whose skeleton is a double tree and therefore have a nonzero contribution to the moments in the limit, which is given by

$$pN_a^\ell N_b^\ell \sum_{i=1}^{\ell} y_a^{\ell-i} \sum_{\substack{k_1+k_2+\cdots+k_i=\ell-i+1 \\ k_1+2k_2+\cdots+ik_i=\ell}} \frac{\ell!}{i!} \prod_{j=1}^{i} \frac{m^{(j)}_{\mathsf{MP}}{}^{k_j}(y_b)}{k_j!}.$$

In addition, for $\ell \leqslant \min\left\{[(d_n'^a - 1)/4, (d_n'^b - 1)/4]\right\} < d_n'^a, d_n'^b$, from Lemma 3.3, the term $(-1)^{s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}}$ has a vanishing expectation. Hence, the first statement follows. We note that it is also possible to obtain the moments in the framework of free probability theory, by directly working with the free multiplicative convolution of $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$ [15], [18]. However, the explicit computation of the moments using the free multiplicative convolution is more complicated, and we have thus chosen to use the direct combinatorial result of Bai *et al.* [4].

Proving the second statement is similar to the case of Wigner [1]. The number of distinct elements c_{ij}^a and c_{ij}^b appearing in cycles of length 8ℓ pertaining to the fourth moment is at most $4\min\left\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\right\} < d_n'^a, d_n'^b$. Hence, by Lemma 3.3, the only cycles contributing to the fourth moment are those with a double tree skeleton (where each element appears at least twice), whose contribution drops as $\mathcal{O}(1/n^2)$ [4]. This completes the proof of the Lemma. ■

Finally, the following lemma gives an upper bound on the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$.

Lemma 3.5: There exist constants $C(y_a, y_b)$ and $\gamma(y_a, y_b)$ such that for all $\ell \geqslant 1$, the ℓth moment of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, denoted by $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$, is upper bounded as follows:

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant C(y_a, y_b)\gamma^\ell(y_a, y_b).$$

Proof: First, note that by forming the moment generating function of $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$ [4], the ℓth moment is the coefficient of z^ℓ in

$$\frac{1}{y_a(\ell+1)}\left[1 + y_a\left(\sum_{k=1}^{\infty} m^{(k)}_{\mathsf{MP}}(y_b) z^k\right)\right]^{\ell+1}. \tag{31}$$

Similarly, by an application of Theorem 1.1 of [4], the ℓth moment of the Marchenko–Pastur distribution $m^{(\ell)}_{\mathsf{MP}}(y)$ is the coefficient of z^ℓ in

$$\frac{1}{y(\ell+1)}\left[1 + y\left(\sum_{k=1}^{\infty} z^k\right)\right]^{\ell+1} = \frac{1}{y(\ell+1)}\left[1 + y\frac{z}{1-z}\right]^{\ell+1} \tag{32}$$

which can be explicitly computed as

$$m^{(\ell)}_{\mathsf{MP}}(y) = \sum_{i=0}^{\ell-1} \frac{y^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i}. \tag{33}$$

Suppose that there exist constants $\alpha(y)$ and $c(y)$, such that

$$m^{(\ell)}_{\mathsf{MP}}(y) \leqslant c(y)\alpha^\ell(y) \tag{34}$$

for all $\ell \geqslant 1$. Then, from (31), the ℓth moment $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$ is upper bounded by the coefficient of z^ℓ in

$$\frac{1}{y_a(\ell+1)}\left[1+y_a\left(\sum_{k=1}^{\infty} c(y_b)\alpha(y_b)^k z^k\right)\right]^{\ell+1} = \frac{1}{y_a(\ell+1)}\left[1+y_a c(y_b)\frac{\alpha(y_b)z}{1-\alpha(y_b)z}\right]^{\ell+1}. \quad (35)$$

By comparing (35) and (32), and using (33), we get the following upper bound for the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$:

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant c(y_b)\alpha(y_b)^\ell \sum_{i=0}^{\ell-1} \frac{\big(y_a c(y_b)\big)^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i} \leqslant c(y_b)c\big(y_a c(y_b)\big)\Big(\alpha(y_b)\alpha\big(y_a c(y_b)\big)\Big)^\ell. \quad (36)$$

Thus, we only need to show that the constants $\alpha(y)$ and $c(y)$ exist. For simplicity, suppose that $y < 1$. Then, the ratio of the ith and $(i-1)$th terms of the summand of (33) is the following:

$$y\frac{(\ell-i+1)(\ell-i)}{i(i+1)}. \quad (37)$$

The largest integer i for which this ratio stays greater or equal to 1, namely i^*, corresponds to the maximum summand in the aforementioned summation. Solving for i^* yields

$$i^* = \left\lfloor \frac{\sqrt{(2\ell y+y+1)^2+4\ell(\ell+1)y(1-y)}-(2\ell y+y+1)}{2(1-y)} \right\rfloor. \quad (38)$$

Hence, the ℓth moment can be lower/upper bounded as

$$m^{(\ell)}_{\mathsf{MP}}(y) \geqslant \frac{y^{i^*}}{i^*+1}\binom{\ell}{i^*}\binom{\ell-1}{i^*} \quad (39)$$

and

$$m^{(\ell)}_{\mathsf{MP}}(y) \leqslant \frac{\ell y^{i^*}}{i^*+1}\binom{\ell}{i^*}\binom{\ell-1}{i^*}. \quad (40)$$

Recall that $m^{(1)}_{\mathsf{MP}}(y) = 1$, so in the following, we assume that $\ell \geqslant 2$. We can use the following version of the Stirling's bounds on $n!$:

$$\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\exp\left(\frac{1}{12n+1}\right) \leqslant n! \leqslant \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\exp\left(\frac{1}{12n}\right)$$

in order to simplify the upper bound on $m^{(\ell)}_{\mathsf{MP}}(y)$. Letting $t_\ell := i^*/\ell$, we get

$$\begin{aligned} m^{(\ell)}_{\mathsf{MP}}(y) &\leqslant \frac{\ell^2}{2\pi i^*(\ell-i^*)}\Big(y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \times \exp\left(\frac{1}{6\ell}-\frac{1}{6(1-t_\ell)\ell+1}-\frac{1}{6t_\ell\ell+1}\right) \\ &< \frac{\ell^2}{2\pi(\ell-1)}\Big(y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \quad (41) \\ &< \frac{1}{\pi}\Big(2y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \quad (42) \end{aligned}$$

where $H(x) := -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function, and the last inequality follows from $\ell < 2^\ell$ for all $\ell > 1$. Also, for a given value of y, one can find the value of ℓ from (38) which maximizes the expression $y^{t_\ell}4^{H(t_\ell)}$. Note that for $y < 1$, we have $y^{t_\ell}4^{H(t_\ell)} < 4$. Hence, by maximizing over ℓ, one can find the constants $\pi \leqslant \alpha(y) < 4$ and $c(y) = 1/\pi$ that satisfy the upper bound. The proof for $y > 1$ is very similar and is thus omitted for brevity. In general, for $y \in (0, \infty)$, one can show that there exists a constant $\alpha(y)$ such that $\pi < \alpha(y) \leqslant 4\max\{y, 1\}$ and $c(y) = 1/\pi$, satisfying the upper bound on the moments of the Marchenko–Pastur measure $\mu_{\mathsf{MP}}(x; y)$. The upper bound given in (36) implies that there exist constants $\gamma(y_a, y_b)$, $C(y_a, y_b)$ such that

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant C(y_a, y_b)\gamma^\ell(y_a, y_b).$$

A loose estimate of these parameters is $C(y_a, y_b) = 1/\pi^2$ and $\gamma(y_a, y_b) = 16\max\{1, y_a\}\max\{1, y_b\}$, which clearly satisfies the upper bound of the lemma. This proves the statement of the lemma. ■

Proof of Theorem 2.3: In what follows, we may drop the dependence on n, y_a, and y_b for notational convenience. Let d'_e be the greatest even integer less than or equal to $d' = \min\{[(d'^a - 1)/4], [(d'^b - 1)/4]\}$ and let r be an integer such that $r \leqslant d'_e$ for all $n > n_0$, for some n_0.

Let $m_{\mathscr{G}}(t)$ and $m_{\boxtimes^2\mathsf{MP}}(t)$ denote the characteristic functions of the distributions $M_{\mathscr{G}}(x)$ and $M_{\boxtimes^2\mathsf{MP}}(x)$, respectively. Using the inequality

$$\left|\exp(it) - 1 - \frac{it}{1!} - \cdots - \frac{(it)^{r-1}}{(r-1)!}\right| \leqslant \frac{|t|^r}{r!} \quad (43)$$

which can easily be verified by induction on r, one can obtain the following bound on the tail of the characteristic function $m_{\mathscr{G}}(t)$:

$$\left|m_{\mathscr{G}}(t) - \sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!}\right| \leqslant m^{(r)}_{\mathscr{G}}\frac{|t|^r}{r!}. \quad (44)$$

From Lemma 3.4, we know that $m^{(\ell)}_{\mathscr{G}} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}$ almost surely as $n \to \infty$, for $\ell = 1, 2, \ldots, r$. Hence

$$\limsup_n \left|m_{\mathscr{G}}(t) - \sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!}\right| \leqslant m^{(r)}_{\boxtimes^2\mathsf{MP}}\frac{|t|^r}{r!}. \quad (45)$$

We have

$$\begin{aligned} &\limsup_n \frac{1}{\pi}\int_0^T \Big|m_{\mathscr{G}}(t) - m_{\boxtimes^2\mathsf{MP}}(t)\Big|\frac{dt}{t} \\ &\leqslant \limsup_n \frac{1}{\pi}\int_0^T \frac{1}{t}\left|\sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!} + \sum_{\ell=r}^{\infty} m^{(\ell)}_{\boxtimes^2\mathsf{MP}}\frac{(it)^\ell}{\ell!} - m_{\boxtimes^2\mathsf{MP}}(t)\right| dt \\ &\quad + \frac{2}{\pi}\int_0^T m^{(r)}_{\boxtimes^2\mathsf{MP}}\frac{|t|^{r-1}}{r!}dt. \quad (46) \end{aligned}$$

It is straightforward to show that

$$f(t) := \frac{1}{t}\left|\sum_{\ell=0}^{r-1} m_{\mathscr{G}}^{(\ell)}\frac{(it)^\ell}{\ell!}+\sum_{\ell=r}^{\infty} m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\frac{(it)^\ell}{\ell!}-m_{\boxtimes^2\mathsf{MP}}(t)\right| \leqslant \frac{1}{t}\sum_{\ell=1}^{r-1}\left|m_{\mathscr{G}}^{(\ell)}-m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\right|\frac{|t|^{\ell-1}}{\ell!} =: g(t). \quad (47)$$

Moreover, we have

$$\int_0^T g(t)dt < \left(\sum_{\ell=1}^{r-1}\left|m_{\mathscr{G}}^{(\ell)}-m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\right|\right)e^T. \quad (48)$$

From lemma 3.4, and by an application of the Markov's inequality, we have

$$\mathbb{P}\left(\left|m_{\mathscr{G}}^{(\ell)}-\mathbb{E}_{P_{\mathsf{G}}}\{m_{\mathscr{G}}^{(\ell)}\}\right| > \mathcal{O}\left(\frac{1}{\sqrt[8]{n}}\right)\right) \leqslant \mathcal{O}\left(\frac{1}{n^{3/2}}\right) \quad (49)$$

for all $\ell = 1, 2, \ldots, d'_e$, where d'_e is the greatest even integer less than or equal to $d' = \min\{[(d'^a - 1)/4], [(d'^b - 1)/4]\}$. Hence, we have

$$\left|m_{\mathscr{G}}^{(\ell)}-\mathbb{E}_{P_{\mathsf{G}}}\{m_{\mathscr{G}}^{(\ell)}\}\right| \leqslant \mathcal{O}\left(\frac{1}{\sqrt[8]{n}}\right) \quad (50)$$

for some large n, almost surely. Therefore, $\int_0^T g(t)dt$ is bounded almost surely. Now, applying the Fatou's lemma [12, p. 23] to the nonnegative sequence $g(t) - f(t)$ yields

$$\limsup_n \int_0^T f(t)dt \leqslant \int_0^T \limsup_n f(t)dt = 0 \quad (51)$$

almost surely. Hence, from (46) and the Zolotarev's inequality given in Lemma 3.1, we will have the following asymptotic bound:

$$\limsup_n \mathscr{L}\big(M_{\mathscr{G}_n}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\big) \leqslant \frac{2}{\pi r}m_{\boxtimes^2\mathsf{MP}}^{(r)}\frac{T^r}{r!}+2e\frac{\ln T}{T}. \quad (52)$$

Utilizing the upper bound of Lemma 3.5 on $m_{\boxtimes^2\mathsf{MP}}^{(r)}$ and optimizing the bound of (52) with respect to T yield

$$\frac{C(y_a, y_b)\gamma^r(y_a, y_b)}{\pi^2 r!}T^{r+1} = e(\ln T - 1) \quad (53)$$

which is a nonalgebraic equation in T. An approximate solution can be obtained by taking

$$T = \mathscr{D}\left(\frac{r}{e\gamma(y_a, y_b)}\right) \quad (54)$$

where

$$\mathscr{D} := \left\{\frac{\gamma(y_a, y_b)}{C(y_a, y_b)}\pi^2 e^2\frac{\big(\ln r - 1\big)}{r}\right\}^{\frac{1}{r+1}}. \quad (55)$$

With this choice of T, the asymptotic bound becomes

$$\limsup_n \mathscr{L}\left(M_{\mathscr{G}}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right) \leqslant \mathscr{A}\frac{1}{r} + \mathscr{B}\frac{\ln r}{r}$$

where

$$\mathscr{A} = \frac{2C(y_a, y_b)}{\pi^2}\mathscr{D}^r + \frac{2e^2\gamma(y_a, y_b)\big(\ln \mathscr{D} - \ln\gamma(y_a, y_b) - 1\big)}{\mathscr{D}}$$

and

$$\mathscr{B} := \frac{2e^2\gamma(y_a, y_b)}{\mathscr{D}}. \quad (56)$$

In particular, if we let $r = \liminf_n r_n$, the aforementioned bound holds as $n \to \infty$. This proves the statement of Theorem 2.3.

Proof of Theorem 2.4: The proof is identical to the proof of Theorem 2.3, up to the application of Zolotarev's inequality given in Lemma 3.1.

Recall from Lemma 2.2 that the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ has a point mass of $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$ at $x = 0$. Hence, denoting the nonatomic portion of this density by $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, the full density is given by

$$\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \Big(1 - \min\big\{1, 1/y_a, 1/y_b\big\}\Big)\delta_0 + \underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (57)$$

Hence, by taking only the nonatomic part $\underline{\mu}_{\boxtimes^2\mathsf{MP}}$ with appropriate renormalization, one gets the density

$$\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}}\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b).$$

Now, consider $\mathbf{A} \in \{-1, 1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1, 1\}^{N_b \times n}$ based on the linear codes $\mathcal{C}^a$ and $\mathcal{C}^b$, respectively. Similarly, since the rank of the matrix $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is at most $\min\{N_a, N_b, n\}$, the asymptotic density has a point mass of at least $\big(1 - \min\{1, \frac{1}{y_a}, \frac{1}{y_b}\}\big)$ at $x = 0$. Hence, by removing $\big(1-\min\{1, \frac{1}{y_a}, \frac{1}{y_b}\}\big)$ from its point mass, and appropriate renormalization, one gets the density

$$\widehat{\mu}_{\mathscr{G}}(x) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}}\left(\mu_{\mathscr{G}}(x) - \Big(1 - \min\big\{1, \frac{1}{y_a}, \frac{1}{y_b}\big\}\Big)\delta_0\right).$$

Let the distribution functions corresponding to $\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and $\widehat{\mu}_{\mathscr{G}}(x)$ be denoted by $\widehat{M}_{\mathscr{G}}(x)$ and $\widehat{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, respectively. It is easy to see that

$$\left|M_{\mathscr{G}}(x) - M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right| = \min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}\left|\widehat{M}_{\mathscr{G}}(x) - \widehat{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right|. \quad (58)$$

Since $\mu_{\mathsf{MP}}(x; y)$ is supported on the interval $\big[(1-\sqrt{y})^2, (1+\sqrt{y})^2\big]$, by basic linear algebra, one can show that the support of $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is contained in the interval $\big[(1 - \sqrt{y_a})^2(1 - \sqrt{y_b})^2, (1+\sqrt{y_a})^2(1+\sqrt{y_b})^2\big]$. Also, by the cubic function root formula, since the coefficient of the leading term of (14) has its root at $z = 0$, the density has only a pole at $x = 0$ [11]. Hence, if $x = 0$ is not contained in the support of $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, it is clearly bounded. The condition of $y_a, y_b \neq 1$ implies that the lower bound on the support given by $(1 - \sqrt{y_a})^2(1 - \sqrt{y_b})^2$ is positive. Hence, $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and consequently the density

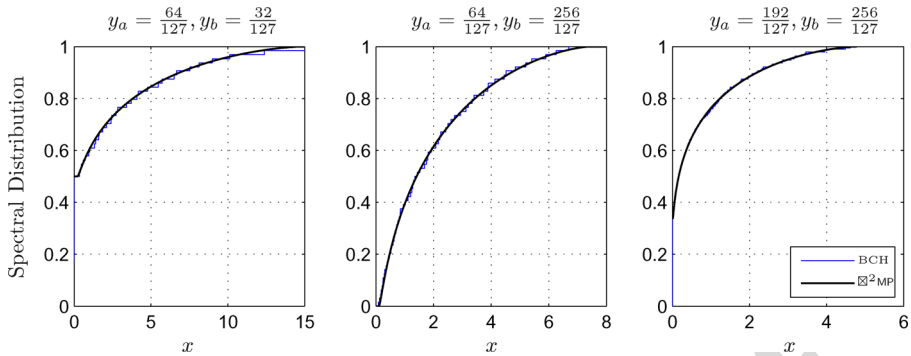

Fig. 1. Empirical spectral distribution of the product of two random matrices based on the $[127, 99, 9]$ and $[127, 92, 11]$ BCH codes versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$.

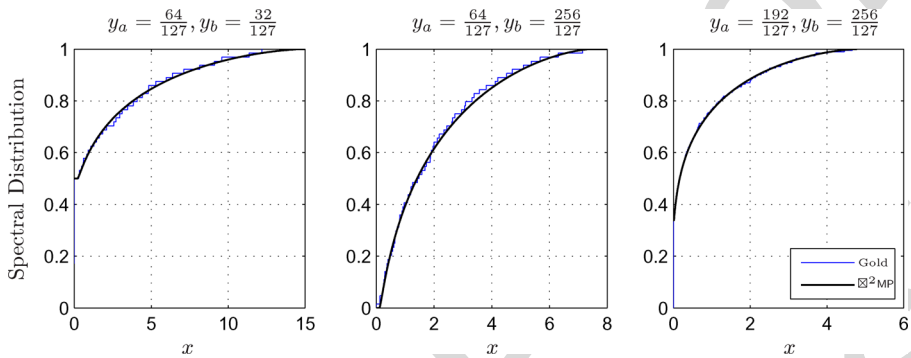

Fig. 2. Empirical spectral distribution of the product of two random matrices based on the $[127, 14, 56]$ Gold code versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$.

$\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ are bounded, and therefore, Lemma 3.2 can be used (instead of the Zolotarev's inequality) in order to bound the Kolmogorov distance of the two distributions $\widehat{M}_{\mathscr{G}}$ and $\widehat{M}_{\boxtimes^2\mathsf{MP}}$. The resulting upper bound is similar to that of the main theorem of [2], and is repeated here for completeness. Let

$$\tau(y_a, y_b) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}} \sup_x \underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) < \infty.$$

Then, we have the following almost sure bound for all x:

$$\limsup_n \left| M_{\mathscr{G}}(x) - M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) \right| \leqslant \mathscr{C}\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

where

$$\mathscr{C} := \frac{24\left\{e \min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\} \tau(y_a, y_b) \gamma(y_a, y_b)\right\}^{\frac{r}{r+1}}}{\pi\left\{\frac{6}{C(y_a, y_b) r}\right\}^{\frac{1}{r+1}}} \tag{59}$$

and $\gamma(y_a, y_b)$ and $C(y_a, y_b)$ are the same as in Lemma 3.5. Hence, the statement of Theorem 2.4 follows.

IV. Numerical Experiments

A. BCH Codes

As for the first set of experiments, we consider random matrices from binary BCH codes. It is known that certain BCH codes have large dual distances [7]. In particular, the Carlitz–Uchiyama bound [7] implies that a binary BCH code of length $n := 2^m - 1$ and designed distance $2t + 1$ with $2t - 1 < 2^{\lceil m/2 \rceil} + 1$ has a dual distance of at least $2^{m-1} - (t-1)2^{m/2}$. For instance, for the $[127, 99, 9]$, the Carlitz–Uchiyama bound implies $d' \geqslant 31$. BCH codes are known for their empirically random-like properties, and, for instance, have been successfully used in the design of near-capacity achieving turbo block codes [10]. Fig. 1 shows the empirical spectral distribution of the product of two random matrices based on the $[127, 99, 9]$ and $[127, 92, 11]$ BCH codes versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$. In order to obtain the analytical curves, $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is first obtained numerically from (14), by the cubic function root formula and using (57). Then, $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is obtained by numerical integration of $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Finally, $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is obtained from the relation given in (18). As predicted by the theorems of this paper, the spectral distribution of the product of two random matrices from BCH codes resembles $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$.

TABLE I
Distance Distribution of the Gold Code

i	B_i
$2^{m-1} - 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} + 2^{(m-3)/2})$
2^{m-1}	$(2^m - 1)(2^{m-1} + 1)$
$2^{m-1} + 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} - 2^{(m-3)/2})$

B. Gold Sequences

Gold sequences are a class of pseudorandom sequences which can be obtained by XOR-ing the shifted versions of two shift register sequences generated by two distinct primitive polynomials [6]. Let $h_1(x)$ and $h_2(x)$ be two primitive polynomials of degree m over $\mathrm{GF}(2)$, such that $h_1(\alpha) = 0$ and $h_2(\alpha^\delta) = 0$ for some integer δ. Suppose that $m \neq 0 \bmod 4$. If $\delta = 2^h + 1$ or $\delta = 2^{2h} - 2^h + 1$, and m/e is odd, with $e := \gcd(m, h)$, then the two polynomials $h_1(x)$ and $h_2(x)$ are denoted by the preferred pair of polynomials. Let $\mathbf{u}$ and $\mathbf{v}$ denote two shift register sequences of length $2^m - 1$, corresponding to the preferred pair of polynomials $h_1(x)$ and $h_2(x)$, respectively. Then, the set of Gold sequences $\mathcal{G}(\mathbf{u}, \mathbf{v})$ is defined as

$$\mathcal{G}(\mathbf{u}, \mathbf{v}) := \{\Re^a \mathbf{u}, \Re^b \mathbf{v}, \text{ or } \Re^a \mathbf{u} \oplus \Re^b \mathbf{v} | 0 \leqslant a, b \leqslant 2^m - 2\}$$

where $\oplus$ and $\Re$ denote the binary XOR and cyclic shift operators, respectively. The set $\mathcal{G}(\mathbf{u}, \mathbf{v})$ consists of $2^{2m} - 1$ binary sequences of length $2^m - 1$, with desirable cross-correlation properties [6].

The set $\mathcal{G}(\mathbf{u}, \mathbf{v}) \cup \{\mathbf{0}\}$ is a binary block code denoted by the Gold code. The dual of the Gold code, for m odd, is the $[2^m - 1, 2^m - 2m - 1, 5]$ double-error-correcting BCH code [7]. Hence, the dual distance of the gold code for odd m is 5, independent of its length. For m odd, the distance distribution of the Gold code is given in Table I.

Fig. 2 shows the empirical spectral distribution of the product of two random matrices based on the $[127, 14, 56]$ Gold code versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$. Although the dual distance of the Gold code is $d' = 5$, the spectral distribution of the product of two random matrices from the Gold code curiously resembles $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. This observation suggests that the condition of large dual distance is only a sufficient condition for the results of this paper to hold. The problems of strengthening this sufficient condition or finding necessary conditions remain open.

ACKNOWLEDGMENT

The authors would like to thank N. Raj Rao for insightful suggestions, Roland Speicher and Horng-Tzer Yau for instructive comments, and Olgica Milenkovic and Alexander Barg for useful discussions.

REFERENCES

[1] G. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge, U.K.: Cambridge Univ. Press, 2009, to be published.

[2] B. Babadi and V. Tarokh, "Spectral distribution of random matrices from binary linear block codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3955–3962, Jun. 2011.

[3] B. Babadi and V. Tarokh, "Spectral distribution of the product of two random matrices based on binary block codes," in *Proc. 49th Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 27–30, 2011, pp. 917–919.

[4] Z. D. Bai, B. Miao, and B. Jin, "On limit theorem for the eigenvalues of product of two random matrices," *J. Multivariate Anal.*, vol. 98, pp. 76–101, 2007.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1991, vol. 2.

[6] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 154–156, Jan. 1968.

[7] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1988.

[8] V. A. Marchenko and L. A. Pastur, "The distribution of eigenvalues in certain sets of random matrices," *Math. USSR-Sbornik*, vol. 72, pp. 507–536, 1967.

[9] S. Pafka, M. Potters, and I. Kondor, Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization 2004 [Online]. Available: arxiv: cond-mat/0402573

[10] R. M. Pyndiah, "Near-optimum decoding of product codes: Block turbo codes," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1003–1010, Aug. 1998.

[11] N. Raj Rao and A. Edelman, "The polynomial method for random matrices," *Found. Comput. Math.*, vol. 8, no. 6, pp. 649–702, 2008.

[12] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[13] V. M. Sidel'nikov, "Weight spectrum of binary Bose-Chaudhuri-Hoquinghem codes," *Problems Inf. Transmiss.*, vol. 7, pp. 11–17, 1971.

[14] J. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[15] R. Speicher, Combinatorial aspects of free probability theory. Goettingen, Germany, 2005 [Online]. Available: http://www.math.uni-sb.de/ag/speicher/speicher_publ_surveys.html

[16] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA: Now Publishers Inc., 2004, Foundations and Trends in Communications and Information Theory.

[17] B. Valkó, Lecture notes on random matrices 2009 [Online]. Available: http://www.math.wisc.edu/\;valko/courses/833/833.html

[18] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*. Providence, RI: Amer. Math. Soc., 2002, CRM Monographs.

[19] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Ann. Math.*, vol. 62, no. 3, pp. 548–564, Nov. 1955.

[20] V. M. Zolotarev, "Estimates of the difference between distributions in the Lévy metric," *Collect. Articles. Part I, Trudy Mat. Inst. Steklov.*, vol. 112, pp. 224–231, 1971.

Behtash Babadi (S'08) received the Ph.D. and M.Sc. degrees in engineering sciences from Harvard University, Cambridge, MA, USA in 2011 and 2008, respectively, and the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran in 2006. He is currently a post-doctoral fellow at the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology, Cambridge, MA, USA and the Department of Anesthesia, Critical Care and Pain Medicine at Massachusetts General Hospital, Boston, MA, USA and a research associate at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His research interests include biological signal processing, adaptive signal processing, information theory, and compressed sensing.

Vahid Tarokh (M'97–SM'02–F'09) received the M.Sc. degree from the University of Windsor, Windsor, ON, Canada, in 1992, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, in 1995. He was with AT&T Labs-Research until August 2000, where he was Head of the Department of Wireless Communications and Signal Processing. He then joined the Department of Electrical Engineering and Computer Sciences (EECS) at the Massachusetts Institute of Technology, Cambridge, as an Associate Professor. In 2002, he joined Harvard University as a Professor and Senior Fellow. His recent research interest includes non-linear information theory and communications, environmentally friendly networks, non-linear signal processing, Tomography, Capsule Endoscopy, Body Area Networks, EEG Signal Analysis and Brain Computer Interaction. His recent awards include a 2011 Guggenheim fellowship in Applied Mathematics, the 2012 IEEE COMSOC Cognitive Networks Technical Committee Publications Award, and the 2013 IEEE Eric E. Sumner Award. He holds two honorary degrees.

Spectral Distribution of Product of Pseudorandom Matrices Formed From Binary Block Codes

Behtash Babadi, *Student Member, IEEE*, and Vahid Tarokh, *Fellow, IEEE*

Dedicated to The Memory of Thomas M. Cover

***Abstract*—Let $\mathbf{A} \in \{-1,1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b \times n}$ be two matrices whose rows are drawn i.i.d. from the codewords of the binary codes $\mathcal{C}^a$ and $\mathcal{C}^b$ of length n and dual distances d'^a and d'^b, respectively, under the mapping $0 \mapsto 1$ and $1 \mapsto -1$. It is proven that as $n \to \infty$ with $y_a := n/N_a \in (0,\infty)$ and $y_b := n/N_b \in (0,\infty)$ fixed, the empirical spectral distribution of the matrix $\mathbf{A}\mathbf{B}^*/\sqrt{N_a N_b}$ resembles a universal distribution (closely related to the distribution function of the free multiplicative convolution of two members of the Marchenko–Pastur family of densities) in the sense of the Lévy distance, if the asymptotic dual distances of the underlying binary codes are large enough. Moreover, an explicit upper bound on the Lévy distance of the two distributions in terms of y_a, y_b, d'^a, and d'^b is given. Under mild conditions, the upper bound is strengthened to the Kolmogorov distance of the underlying distributions. Numerical studies on the empirical spectral distribution of the product of random matrices from BCH and Gold codes are provided, which verify the validity of this result.**

***Index Terms*—Binary block codes, free probability theory, Lévy distance, Marchenko–Pastur law, pseudorandom matrices, random matrix theory.**

I. Introduction

THE elegant theory of random matrices has attracted a considerable amount of attention in recent years. Random matrix theory studies the emergence of deterministic collective behavior from a large collection of random elements in the domain of matrices. Born in theoretical physics, and nurtured by mathematicians, this theory has also found its home in several other disciplines of science such as economics [9] and communication theory [16]. Most of the important results of random matrix theory rely on the fact that the underlying matrix elements have an i.i.d. structure. In this paper, however, we study the spectral behavior of certain pseudorandom matrices, and relate their spectral behavior to their fully random counterparts.

More explicitly, let $\mathcal{C}$ be an $(n, \mathcal{M}, d)$ binary block code of length n, size $\mathcal{M}$, and minimum Hamming distance d over $\mathrm{GF}(2)^n$. Let $\mathbf{c}_i := (c_{i1}, c_{i2}, \ldots, c_{in})$ be a codeword in $\mathcal{C}$. We say that an $N \times n$ random matrix $\mathbf{\Phi}$ is *based on* a binary block code $\mathcal{C}$, if for a randomly i.i.d. drawn set of codewords of $\mathcal{C}$, $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$, we have $\Phi_{ij} = (-1)^{c_{ij}}$, for all $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n$. Let $\mathbf{A} \in \{-1,1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b \times n}$ be two random matrices based on the binary codes $\mathcal{C}^a$ and $\mathcal{C}^b$ of length n, respectively. We study the empirical spectral distribution of the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ and show that if the dual distances of the underlying codes are sufficiently large, the asymptotic empirical spectral distribution resembles a deterministic universal distribution. This universal distribution pertains to the case where the elements of $\mathbf{A}$ and $\mathbf{B}$ are drawn i.i.d. from the set $\{-1,1\}$, and is closely related to the free multiplicative convolution of two of the members of the Marchenko–Pastur family of densities [15], [18]. We upper bound the Lévy distance of the asymptotic empirical spectral distribution of the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ to the aforementioned universal distribution as a function of $y_a := n/N_a$, $y_b := n/N_b$, and the dual distances of $\mathcal{C}^a$ and $\mathcal{C}^b$. This result is also strengthened to an upper bound on the Kolmogorov distance of the underlying distributions under mild conditions. Numerical experiments on BCH and Gold codes are provided, which confirm the theoretical result of this paper. Although independently interesting from the viewpoint of random matrix theory, the result of this paper suggests a criterion for evaluating the cross randomness of two set of binary codes or sequences.

The outline of this paper follows next. In Section II, we introduce the notation and state the main theorems of the paper followed by a discussion of the main results. The detailed proofs of the main theorems are presented in Section III. Finally, numerical experiments in Section IV conclude the paper.

II. Main Result

A. Results and Notation From Coding Theory

Before presenting the main result, we introduce the notation and state some preliminary definitions.

A $(n, \mathcal{M}, d)$ binary code $\mathcal{C}$ is defined as a set of $\mathcal{M}$ binary n-tuples such that any two such n-tuples differ in at least d places, with d being the largest number with this property. The Hamming weight of an n-tuple $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathrm{GF}(2)^n$, denoted by $\mathrm{wt}(\mathbf{u})$, is defined as the number of nonzero elements of $\mathbf{u}$.

Manuscript received October 26, 2011; revised September 26, 2012; accepted October 02, 2012. This work was supported in part by a fellowship from the John Simon Guggenheim Memorial Foundation. The material in this paper was presented in part at the 2011 Allerton Conference on Communication, Control, and Computing, Monticello, IL [3].

B. Babadi is with the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: behtash@nmr.mgh.harvard.edu).

V. Tarokh is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: vahid@seas.harvard.edu).

Communicated by A. Moustakas, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2223812

Consider the group algebra over $\mathrm{GF}(2)^n$, in which the code $\mathcal{C}$ is represented by the element

$$\mathcal{C} := \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} c_{\mathbf{u}} t^{\mathbf{u}} \tag{1}$$

where $t^{\mathbf{u}} := t_1^{u_1} t_2^{u_2} \cdots t_n^{u_n}$, and

$$c_{\mathbf{u}} := \begin{cases} 1, & \text{if } \mathbf{u} \in \mathcal{C} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

For all $\mathbf{u}, \mathbf{v} \in \mathrm{GF}(2)^n$, the product of $t^{\mathbf{u}}$ and $t^{\mathbf{v}}$ is defined as

$$t^{\mathbf{u}} t^{\mathbf{v}} := t^{\mathbf{u}\oplus\mathbf{v}} = t_1^{u_1\oplus v_1} t_2^{u_2\oplus v_2} \cdots t_n^{u_n\oplus v_n} \tag{3}$$

where $\oplus$ denotes the binary addition. For a binary n-tuple $\mathbf{u} \in \mathrm{GF}(2)^n$, let $\chi_{\mathbf{u}}$ be the character mapping

$$\chi_{\mathbf{u}}(t^{\mathbf{v}}) = (-1)^{\mathbf{u}\cdot\mathbf{v}} \tag{4}$$

with $\mathbf{u}\cdot\mathbf{v} := \sum_i u_i v_i \bmod 2$, for all $\mathbf{v} \in \mathrm{GF}(2)^n$. Suppose that for a code $\mathcal{C}$, corresponding to the element of the group algebra given by (1), we have

$$\mathcal{M} = \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} c_{\mathbf{u}} \neq 0. \tag{5}$$

Now, consider

$$\mathcal{D} := \frac{1}{\mathcal{M}}\mathcal{C}^2 = \sum_{\mathbf{u}\in\mathrm{GF}(2)^n} d_{\mathbf{u}} t^{\mathbf{u}}. \tag{6}$$

The *distance distribution* of the code $\mathcal{C}$ is defined as the set $\{B_0, B_1, \ldots, B_n\}$, where

$$B_i := \sum_{\mathrm{wt}(\mathbf{u})=i} d_{\mathbf{u}}. \tag{7}$$

The *transformed* distance distribution of the code $\mathcal{C}$ under the character mapping is given by the set $\{B'_0, \ldots, B'_n\}$, where

$$B'_j := \frac{1}{\mathcal{M}} \sum_{\mathrm{wt}(\mathbf{u})=j} \chi_{\mathbf{u}}(\mathcal{D}). \tag{8}$$

Finally, the *dual distance* of the code $\mathcal{C}$ is defined as d' such that $B'_i = 0$ for $1 \leqslant i \leqslant d'-1$ and $B'_{d'} \neq 0$ [7].

B. Results and Notation From Random Matrix Theory

Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of the matrix $\mathbf{X} \in \mathbb{R}^{N\times N}$, and let

$$\mu_{\mathbf{X}} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \tag{9}$$

denote the spectral measure of $\mathbf{X}$, where δ_x denotes the Dirac measure. Let $M_{\mathbf{X}}(x)$ denote the distribution function associated with the measure $\mu_{\mathbf{X}}$. Finally, let $m_{\mathbf{X}}^{(\ell)}$ denote the ℓth moment of $\mu_{\mathbf{X}}$. The Stieltjes transform of the density μ is defined as

$$s_\mu(z) := \int \frac{1}{z-x}\mu(dx) \tag{10}$$

for $\{z \in \mathbb{C} | \Im\{z\} \geqslant 0\}$. In particular, we denote by $M_{\mathsf{MP}}(x; y)$ the distribution corresponding to the Marchenko–Pastur measure $\mu_{\mathsf{MP}}(x; y)$ whose density is given by

$$\frac{d\mu_{\mathsf{MP}}(x;y)}{dx} := \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\,1_{(a\leqslant x\leqslant b)} \tag{11}$$

with $a = (1-\sqrt{y})^2$ and $b = (1+\sqrt{y})^2$. It can be shown that the Stieltjes transform of the Marchenko–Pastur density, denoted by $s_{\mathsf{MP}}(z; y)$, satisfies the following quadratic equation [8]:

$$yzs^2_{\mathsf{MP}}(z;y) - (1-y-z)s_{\mathsf{MP}}(z;y) + 1 = 0. \tag{12}$$

Let $\mathbf{A} \in \{-1,1\}^{N_a\times n}$ and $\mathbf{B} \in \{-1,1\}^{N_b\times n}$. Let $y_a := \frac{n}{N_a} \in (0,\infty)$ and $y_b := \frac{n}{N_b} \in (0,\infty)$ be fixed numbers. Consider the Gram matrix of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ given by $\mathscr{G} := \frac{1}{N_a N_b}\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*$. If the elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ are i.i.d. drawn from the set $\{-1,1\}$, it can be shown that the spectral measure of $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$ tends to a deterministic limit almost surely (see [4, Th. 1.1]). In order to identify this measure, we first consider the closely related measure corresponding to the asymptotic empirical spectral distribution of the matrix $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$. Clearly, the asymptotic spectral densities of the matrices $\frac{1}{N_a}\mathbf{A}^*\mathbf{A}$ and $\frac{1}{N_a}\mathbf{B}^*\mathbf{B}$ are given by $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$, respectively. Free probability theory implies that the asymptotic spectral density of the product $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is given by

$$\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \mu_{\mathsf{MP}}(x; y_a) \boxtimes \mu_{\mathsf{MP}}(x; y_b) \tag{13}$$

where $\mu_1 \boxtimes \mu_2$ denotes the free multiplicative convolution of the densities μ_1 and μ_2 [15], [18]. We also denote the distribution function and Stieltjes transform of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ by $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, respectively. Using the polynomial method for algebraic random matrices [11], one can directly characterize $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, without the need for evaluating the free multiplicative convolution, as given by the following lemma.

Lemma 2.1: The Stieltjes transform of the free multiplicative convolution of two densities from the Marchenko–Pastur family with parameters y_a and y_b, denoted by $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$, satisfies the following cubic equation:

$$\begin{aligned} &y_a y_b z^2 \overline{s}^3_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) + z(2y_a y_b - y_a - y_b)\overline{s}^2_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) \\ &- (z - (1-y_a)(1-y_b))\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b) - 1 = 0. \end{aligned} \tag{14}$$

Proof: The proof is based on the main result of [14] which states that the Stieltjes transform of the asymptotic spectral distribution of $\frac{1}{N}\mathbf{X}^*\mathbf{X}\mathbf{T}$ (denoted by $s(z)$) for $\mathbf{X} \in \mathbb{R}^{n\times N}$ having i.i.d. unit variance elements and $\mathbf{T} \in \mathbb{R}^{n\times n}$ being a random nonnegative definite Hermitian with asymptotic spectral distribution $H(x)$, and $n/N \to y$, satisfies

$$s(z) = \int \frac{1}{x\big(1-y-yzs(z)\big) - z} dH(x). \tag{15}$$

Specializing (15) to the case of $N = N_a$, $\mathbf{X} = \mathbf{A}$, and $\mathbf{T} = \frac{1}{N_b}\mathbf{B}^*\mathbf{B}$ and noting that the Stieltjes transform of $H(x)$ in this case satisfies (12), it is not hard to show that $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ corresponds to an algebraic random matrix [11] and satisfies (12) with y, $s_{\mathsf{MP}}(z; y)$ and z replaced by y_a, $(1 - y_b - y_b z \overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b))\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ and $z/(1 - y_b - y_b z \overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b))$, respectively. The latter establishes the result of the lemma. ■

The solution to (14) can be obtained using the standard cubic function root formula. Since the coefficient of the leading term of $\overline{s}_{\boxtimes^2\mathsf{MP}}$ in the aforementioned equation, $y_a y_b z^2$, has roots at $z = 0$, the nonatomic portion of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}$ is given by the positive imaginary part of $\overline{s}_{\boxtimes^2\mathsf{MP}}$ with a scaling of $1/\pi$ [11]. The support of the nonatomic portion of the density is the range of x where the discriminant of the cubic equation is negative, and hence, the cubic equation has two complex conjugate roots. A closed-form solution for the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}$ can be explicitly obtained from (14), but is omitted for brevity. Moreover, (14) is a powerful tool for numerical computation of the density.

The following lemma characterizes the possible point mass of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ at $x = 0$.

Lemma 2.2: The density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ has a point mass of $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$ at $x = 0$.

Proof: Note that the possible point mass of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ at $x = 0$ is given by the coefficient of $-1/z$ in the Puiseux expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ in terms of the powers of z and $1/z$ [11]. Let κ denote the coefficient of $-1/z$ in the Puiseux expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$. By inspecting the expansion of $\overline{s}_{\boxtimes^2\mathsf{MP}}(z; y_a, y_b)$ from (14), we get

$$y_a y_b \kappa^3 - (2y_a y_b - y_a - y_b)\kappa^2 + (1 - y_a)(1 - y_b)\kappa = 0 \quad (16)$$

which has the roots 0, $1 - 1/y_a$, and $1 - 1/y_b$. Hence, the point mass can be identified as one of these roots. Now, suppose that $\mathbf{A} \in \{-1, 1\}^{N_a \times n}$ and $\mathbf{B} \in \{-1, 1\}^{N_b \times n}$ are random matrices with elements drawn i.i.d. from $\{-1, 1\}$. From elementary linear algebra, the rank of $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is no greater than $\min\{n, N_a, N_b\}$. Hence, the point mass of the asymptotic density of $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is no smaller than $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$. The latter fact and the possible values of the point mass obtained from the Puiseux expansion imply that the point mass is indeed equal to $\big(1 - \min\{1, 1/y_a, 1/y_b\}\big)$. ■

Now, let $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ denote the asymptotic spectral density of the matrix $\frac{1}{N_a N_b}\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*$. By an elementary linear algebraic argument, one can show that $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is related to $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ by the following transformation [14]:

$$\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \big(1 - y_a\big)\delta_0 + y_a \overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (17)$$

Similarly, the distribution function corresponding to $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is denoted by $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and is given by

$$M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \big(1 - y_a\big) + y_a \overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (18)$$

The Lévy distance between two distribution functions $M_1(x)$ and $M_2(x)$ is defined as

$$\begin{aligned}&\mathscr{L}\big(M_1(x), M_2(x)\big)\\ &:= \inf\Big\{\epsilon > 0 \Big| M_1(x - \epsilon) - \epsilon \leqslant M_2(x) \leqslant M_1(x + \epsilon) + \epsilon,\\ &\qquad \forall x \in \mathbb{R}\Big\}. \end{aligned} \quad (19)$$

The Kolmogorov distance between two distribution functions $M_1(x)$ and $M_2(x)$ is defined as

$$\sup_x \Big|M_1(x) - M_2(x)\Big|. \quad (20)$$

C. Main Result

The main theorem of this paper is the following:

Theorem 2.3: Let $\{\mathcal{C}_n^a\}_{n=1}^\infty$ and $\{\mathcal{C}_n^b\}_{n=1}^\infty$ be two sequences of binary block codes of length n. Let $d_n'^a$ and $d_n'^b$ denote the dual distances of $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $\{N_a^n\}_{n=1}^\infty$ and $\{N_b^n\}_{n=1}^\infty$ be two sequences such that $n/N_a^n \to y_a \in (0, \infty)$ and $n/N_b^n \to y_b \in (0, \infty)$, as $n \to \infty$. Let $\big\{\mathbf{A}_n \in \{-1, 1\}^{N_a^n \times n}\big\}_{n=1}^\infty$ and $\big\{\mathbf{B}_n \in \{-1, 1\}^{N_b^n \times n}\big\}_{n=1}^\infty$ be two sequences of random matrices, where $\mathbf{A}_n$ and $\mathbf{B}_n$ are based on $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $M_{\mathscr{G}_n}(x)$ denote the spectral distribution function of the Gram matrix of $\frac{1}{\sqrt{N_a^n N_b^n}}\mathbf{A}_n\mathbf{B}_n^*$ and $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ denote the distribution of the free multiplicative convolution of the Marchenko–Pastur densities $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$. Let $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := (1 - y_a) + y_a \overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Let r_n be the greatest even integer less than or equal to $\min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$, and $r := \liminf_n r_n$. Then, we have

$$\limsup_n \mathscr{L}\big(M_{\mathscr{G}_n}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\big) \leqslant \mathscr{A}\frac{1}{r} + \mathscr{B}\frac{\ln r}{r}$$

almost surely, where $\mathscr{A}$ and $\mathscr{B}$ are only functions of y_a, y_b, and r (explicitly given in this paper), and are bounded in r.

The statement of Theorem 2.3 can be strengthened as follows.

Theorem 2.4: Let y_a, y_b, $M_{\mathscr{G}_n}(x)$, $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, and r be as in Theorem 2.3. Suppose that $y_a \in (0, 1) \cup (1, \infty)$ and $y_b \in (0, 1) \cup (1, \infty)$. Then, we have

$$\limsup_n \Big|M_{\mathscr{G}_n}(x) - M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\Big| \leqslant \mathscr{C}\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

for all x almost surely, where $\mathscr{C}$ is a function of y_a, y_b, and r (implicitly given in this paper), and is bounded in r.

D. Discussion of the Main Result

Theorems 2.3 and 2.4 state that the empirical spectral distribution of the Gram matrix of the random matrix $\frac{1}{\sqrt{N_a N_b}}\mathbf{A}\mathbf{B}^*$, with $\mathbf{A}$ and $\mathbf{B}$ based on binary block codes $\mathcal{C}^a$ and $\mathcal{C}^b$, respectively, resembles the universal empirical spectral distribution $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ in the sense of Lévy and Kolmogorov distance, respectively, as $n \to \infty$, provided that the dual distances of the codes $\mathcal{C}^a$ and $\mathcal{C}^b$ are large enough.

Previously, the authors proved the following result, which is stated as follows for completeness [2].

Theorem 2.5 (Main Theorem of [2]): Consider a sequence of $[n, k_n, d_n]$ binary linear block codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$. Let $d_n^{\perp}$ denote the dual distance of $\mathcal{C}_n$. Let $\mathbf{\Phi}_{\mathcal{C}_n}$ be a $p \times n$ random matrix based on $\mathcal{C}_n$, $\mathcal{G}_{\mathcal{C}_n}$ denote the Gram matrix of $\frac{1}{\sqrt{n}}\mathbf{\Phi}_{\mathcal{C}_n}$, and $M_{\mathcal{C}_n}(x)$ denote the empirical spectral distribution of $\mathcal{G}_{\mathcal{C}_n}$. Finally, let r_n be the greatest even integer less than or equal to $[(d_n^{\perp}-1)/2]$, and let $r := \liminf_n r_n$. Then, as $n \to \infty$ with $y := p/n \in (0,1)$ fixed, we have

$$\limsup_n \left| M_{\mathcal{C}_n}(x) - M_{\mathsf{MP}}(x) \right| \leqslant c(y,r)\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

almost surely for all z, where $c(y,r)$ is a bounded function of r (explicitly given in [2]).

The aforementioned theorem states that the asymptotic empirical spectral distribution of a random matrix from a binary linear code $\mathcal{C}$ resembles the Marchenko–Pastur distribution in the sense of Kolmogorov distance, provided that the dual distance of $\mathcal{C}$ is large enough [2]. The proof of the main theorem of [2] is simpler to those of this paper; however, the result of [2] is stronger than Theorem 2.3, since the Lévy distance is weaker than the Kolmogorov distance. Theorem 2.4, however, provides an upper bound on the Kolmogorov distance of the underlying distribution, under the more restrictive assumption that $y_a, y_b \neq 1$. As will be shown in the proof of Theorem 2.3, this assumption implies boundedness of the density $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Moreover, the functionality of the upper bound of Theorem 2.4 with respect to y_a and y_b is implicit, due to the complicated analytic expression of the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, whereas the upper bound of Theorem 2.3 is given explicitly as a function of y_a, y_b, and r.

Furthermore, the proof of Theorem 2.3 (and similarly Theorem 2.4) of this paper offers more difficulty in identifying and bounding the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, which are more involved than those of $\mu_{\mathsf{MP}}(x; y)$. Finally, the results of this paper hold for binary (possibly nonlinear) codes in general, and more importantly, the constraint of $y \in (0,1)$ in [2] is strengthened to $y_a, y_b \in (0, \infty)$ in this paper.

III. Proof of the Main Results

We present a proof along the lines of the proof of the main result of [2]. We therefore need to establish a number of lemmas. First, we state the following lemma, known as the Zolotarev's inequality, for bounding the Lévy distance of two distributions.

Lemma 3.1 (Zolotarev's Inequality): Let $M_1(x)$ and $M_2(x)$ be two distribution functions, with characteristic functions $m_1(t)$ and $m_2(t)$, respectively. Then, for all $T > 1.3$, we have

$$\mathscr{L}\big(M_1(x), M_2(x)\big) < \frac{1}{\pi}\int_0^T \Big| m_1(t) - m_2(t) \Big| \frac{dt}{t} + 2e\frac{\ln T}{T}.$$

Proof: The proof uses smoothing techniques to bound the Lévy distance of the two distributions, and is given in [20]. ■

Similarly, when the density of $M_2(x)$ is bounded, we have the following lemma for bounding the absolute distance of $M_1(x)$ and $M_2(x)$.

Lemma 3.2: Let $M_1(x)$ be a probability distribution with vanishing expectation and characteristic function $m_1(t)$. Suppose that $M_1(x) - M_2(x)$ vanishes at $\pm\infty$ and that $M_2(x)$ has a derivative $\mu_2(x)$ such that $|\mu_2(x)| \leqslant A$. Finally, suppose that $\mu_2(x)$ has a continuously differentiable Fourier transform $m_2(t)$ such that $m_2(0) = 1$ and $m_2'(0) = 0$. Then, for all z and $T > 0$, we have

$$|M_1(x) - M_2(x)| \leqslant \frac{1}{\pi}\int_{-T}^{T} \left| \frac{m_1(t) - m_2(t)}{t} \right| dt + \frac{24A}{\pi T}. \quad (21)$$

Proof: The proof uses smoothing techniques to cope with the fact that the underlying density of $M_1(x)$ may not enjoy sufficient smoothness properties. The detailed proof can be found in [5] and is thus omitted for brevity. ■

The following lemma establishes an important combinatorial property of a binary code $\mathcal{C}$.

Lemma 3.3 [7]: Any set of $r \leqslant d' - 1$ columns of the code $\mathcal{C}$ contains each binary r-tuple exactly $\mathcal{M}/2^r$ times, and d' is the largest number with this property.

Proof: By the definition of d' and the properties of the character mapping, we have $\chi_{\mathbf{u}}(\mathcal{C}) = 0$ for all $\mathbf{u}$ with $\mathrm{wt}(\mathbf{u}) = 1, 2, \ldots, d'-1$. For $\mathrm{wt}(\mathbf{u}) = 1$, this implies that each component of the codewords takes the values of 0 and 1 a total of $\mathcal{M}/2$ times each. For $\mathrm{wt}(\mathbf{u}) = 2$, this implies that any set of two components of the codewords takes the combinations $00, 01, 10, 11$ a total of $\mathcal{M}/4$ times each, etc. Hence, any set of $d'-1$ components of the codewords takes all the possible $(d'-1)$-tuples a total of $\mathcal{M}/2^{(d'-1)}$ times each. Since $B'_{d'} \neq 0$, there must be a codeword $\mathbf{u}$ with $\mathrm{wt}(\mathbf{u}) = d'$ such that $\chi_{\mathbf{u}}(\mathcal{C}) \neq 0$. ■

The following lemma establishes the almost sure convergence of the first $d' := \liminf_n \min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$ moments of $\mu_{\mathscr{G}_n}$ to those of $\mu_{\boxtimes^2\mathsf{MP}}$, as $n \to \infty$.

Lemma 3.4: Consider the sequences $\{N_a^n\}_{n=1}^{\infty}$ and $\{N_b^n\}_{n=1}^{\infty}$, such that $n/N_a^n \to y_a \in (0,\infty)$ and $n/N_b^n \to y_b \in (0,\infty)$, as $n \to \infty$. Consider two sequence of binary block codes $\{\mathcal{C}_n^a\}_{n=1}^{\infty}$ and $\{\mathcal{C}_n^b\}_{n=1}^{\infty}$, where $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$ are of length n and dual distances $d_n'^a$ and $d_n'^b$, respectively. Let $\{\mathbf{A}_n \in \{-1,1\}^{N_a^n \times n}\}_{n=1}^{\infty}$ and $\{\mathbf{B}_n \in \{-1,1\}^{N_b^n \times n}\}_{n=1}^{\infty}$ be two sequence of random matrices, such that $\mathbf{A}_n$ and $\mathbf{B}_n$ are based on the binary codes $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. Let $\mu_{\mathscr{G}_n}$ denote the spectral measure of the Gram matrix of $\frac{1}{\sqrt{N_a^n N_b^n}}\mathbf{A}_n\mathbf{B}_n^*$. Let

$$m_{\mathscr{G}_n}^{(\ell)} := \int x^{\ell}\mu_{\mathscr{G}_n}(dx) \quad (22)$$

$$m_{\boxtimes^2\mathsf{MP}}^{(\ell)}(y_a, y_b) := \int x^{\ell}\mu_{\boxtimes^2\mathsf{MP}}(dx; y_a, y_b) \quad (23)$$

and

$$m_{\mathsf{MP}}^{(\ell)}(y) := \int z^{\ell}\mu_{\mathsf{MP}}(dx; y) = \sum_{i=0}^{\ell-1} \frac{y^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i} \quad (24)$$

be the ℓth moment of the spectral measures $\mu_{\mathscr{G}_n}$, $\mu_{\boxtimes^2\mathsf{MP}}$, and μ_{MP}, respectively. Then, for

$$1 \leq \ell \leq d' := \liminf_n \min\big\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\big\}$$

we have

$$m^{(\ell)}_{\mathscr{G}_n} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) = \sum_{i=1}^{\ell} y_a^{\ell-i} \sum_{\substack{k_1+k_2+\cdots+k_i=\ell-i+1 \\ k_1+2k_2+\cdots+ik_i=\ell}} \frac{\ell!}{i!} \prod_{j=1}^{i} \frac{{m^{(j)}_{\mathsf{MP}}}^{k_j}(y_b)}{k_j!}$$

almost surely.

Proof: Let $P_{\mathscr{G}_n}$ be the probability measure induced by the i.i.d. selection of N_a^n and N_b^n codewords from $\mathcal{C}_n^a$ and $\mathcal{C}_n^b$, respectively. For notational convenience, we drop the dependence on n, where there is no ambiguity. From an application of the Borel–Cantelli Lemma (see, e.g., [1] or [17]), it is enough to show

$$\mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \tag{25}$$

and

$$\mathbb{E}_{P_{\mathscr{G}}}\left\{\left|m^{(\ell)}_{\mathscr{G}} - \mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\}\right|^4\right\} = \mathcal{O}\left(\frac{1}{n^2}\right) \tag{26}$$

where $\mathbb{E}_{P_{\mathscr{G}}}$ is the expectation with respect to $P_{\mathscr{G}}$.

Thus, we need to prove that the average of the first d' moments of the measure $\mu_{\mathscr{G}}$ coincides with those of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ almost surely and that the fourth central moment of these moments drops as $1/n^2$. In what follows, we drop the subscript n for notational convenience. The ℓth moment of $\mu_{\mathscr{G}}$ can be written as

$$\begin{aligned}
\mathbb{E}_{P_{\mathscr{G}}}\left\{m^{(\ell)}_{\mathscr{G}}\right\} &= \mathbb{E}_{P_{\mathscr{G}}}\left\{\int x^\ell \mu_{\mathscr{G}}(dx)\right\} = \mathbb{E}_{P_{\mathscr{G}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\lambda_i^\ell\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\mathbb{E}_{P_{\mathscr{G}}}\left\{\operatorname{Tr}\left\{(\mathbf{A}\mathbf{B}^*\mathbf{B}\mathbf{A}^*)^\ell\right\}\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\mathbb{E}_{P_{\mathscr{G}}}\left\{\operatorname{Tr}\left\{(\mathbf{A}^*\mathbf{A}\mathbf{B}^*\mathbf{B})^\ell\right\}\right\} \\
&= \frac{1}{pN_a^\ell N_b^\ell}\sum_{\mathcal{I},\mathcal{J}}\mathbb{E}_{P_{\mathscr{G}}}\left\{(-1)^{s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}}\right\}
\end{aligned}$$

where

$$\mathcal{I}^a := \{i_t^a\}_{t=1}^{\ell} \in \{1, 2, \ldots, N_a\}^\ell \tag{27}$$
$$\mathcal{J}^a := \{j_t^a\}_{t=1}^{\ell} \in \{1, 2, \ldots, n\}^\ell \tag{28}$$
$$\mathcal{I}^b := \{i_t^b\}_{t=1}^{\ell} \in \{1, 2, \ldots, N_b\}^\ell \tag{29}$$
$$\mathcal{J}^b := \{j_t^b\}_{t=1}^{\ell} \in \{1, 2, \ldots, n\}^\ell \tag{30}$$

and

$$s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b} := c_{i_1^a j_1^a} \oplus c_{i_2^a j_1^a} \oplus c_{i_1^b j_1^b} \oplus c_{i_2^b j_1^b} \oplus \cdots \oplus c_{i_\ell^b j_\ell^b} \oplus c_{i_1^a j_\ell^b}$$

with $\oplus$ denoting the binary addition and the summation running over all $\mathcal{I}^a \in \{1, 2, \ldots, N_a\}^\ell$, $\mathcal{I}^b \in \{1, 2, \ldots, N_b\}^\ell$, and $\mathcal{J}^a, \mathcal{J}^b \in \{1, 2, \ldots, n\}^\ell$. Note that $s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}$ corresponds to a directed cycle of length 4ℓ on a complete quadripartite graph $G = (X^a \cup Y^a \cup X^b \cup Y^b, E)$ with $X^a := \{1, 2, \ldots, N_a\}$, $X^b := \{1, 2, \ldots, N_b\}$, and $Y^a = Y^b := \{1, 2, \ldots, n\}$, where, for instance, $c_{i^a j^a}$ corresponds to an edge from node $i^a \in \mathcal{I}^a$ to node $j^a \in \mathcal{J}^a$.

Generalizing the proofs of Wigner [19] and Marchenko and Pastur [8], Bai *et al.* [4] have elegantly counted the number of such directed cycles, whose skeleton is a double tree and therefore have a nonzero contribution to the moments in the limit, which is given by

$$pN_a^\ell N_b^\ell \sum_{i=1}^{\ell} y_a^{\ell-i} \sum_{\substack{k_1+k_2+\cdots+k_i=\ell-i+1 \\ k_1+2k_2+\cdots+ik_i=\ell}} \frac{\ell!}{i!} \prod_{j=1}^{i} \frac{{m^{(j)}_{\mathsf{MP}}}^{k_j}(y_b)}{k_j!}.$$

In addition, for $\ell \leqslant \min\left\{[(d_n'^a - 1)/4, (d_n'^b - 1)/4]\right\} < d_n'^a, d_n'^b$, from Lemma 3.3, the term $(-1)^{s_{\mathcal{I}^a,\mathcal{I}^b,\mathcal{J}^a,\mathcal{J}^b}}$ has a vanishing expectation. Hence, the first statement follows. We note that it is also possible to obtain the moments in the framework of free probability theory, by directly working with the free multiplicative convolution of $\mu_{\mathsf{MP}}(x; y_a)$ and $\mu_{\mathsf{MP}}(x; y_b)$ [15], [18]. However, the explicit computation of the moments using the free multiplicative convolution is more complicated, and we have thus chosen to use the direct combinatorial result of Bai *et al.* [4].

Proving the second statement is similar to the case of Wigner [1]. The number of distinct elements c_{ij}^a and c_{ij}^b appearing in cycles of length 8ℓ pertaining to the fourth moment is at most $4\min\left\{[(d_n'^a - 1)/4], [(d_n'^b - 1)/4]\right\} < d_n'^a, d_n'^b$. Hence, by Lemma 3.3, the only cycles contributing to the fourth moment are those with a double tree skeleton (where each element appears at least twice), whose contribution drops as $\mathcal{O}(1/n^2)$ [4]. This completes the proof of the Lemma. ∎

Finally, the following lemma gives an upper bound on the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$.

Lemma 3.5: There exist constants $C(y_a, y_b)$ and $\gamma(y_a, y_b)$ such that for all $\ell \geqslant 1$, the ℓth moment of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, denoted by $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$, is upper bounded as follows:

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant C(y_a, y_b)\gamma^\ell(y_a, y_b).$$

Proof: First, note that by forming the moment generating function of $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$ [4], the ℓth moment is the coefficient of z^ℓ in

$$\frac{1}{y_a(\ell+1)}\left[1 + y_a\left(\sum_{k=1}^{\infty} m^{(k)}_{\mathsf{MP}}(y_b) z^k\right)\right]^{\ell+1}. \tag{31}$$

Similarly, by an application of Theorem 1.1 of [4], the ℓth moment of the Marchenko–Pastur distribution $m^{(\ell)}_{\mathsf{MP}}(y)$ is the coefficient of z^ℓ in

$$\frac{1}{y(\ell+1)}\left[1 + y\left(\sum_{k=1}^{\infty} z^k\right)\right]^{\ell+1} = \frac{1}{y(\ell+1)}\left[1 + y\frac{z}{1-z}\right]^{\ell+1} \tag{32}$$

which can be explicitly computed as

$$m^{(\ell)}_{\mathsf{MP}}(y) = \sum_{i=0}^{\ell-1} \frac{y^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i}. \tag{33}$$

Suppose that there exist constants $\alpha(y)$ and $c(y)$, such that

$$m^{(\ell)}_{\mathsf{MP}}(y) \leqslant c(y)\alpha^\ell(y) \tag{34}$$

for all $\ell \geqslant 1$. Then, from (31), the ℓth moment $m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b)$ is upper bounded by the coefficient of z^ℓ in

$$\frac{1}{y_a(\ell+1)}\left[1+y_a\left(\sum_{k=1}^{\infty} c(y_b)\alpha(y_b)^k z^k\right)\right]^{\ell+1} = \frac{1}{y_a(\ell+1)}\left[1+y_a c(y_b)\frac{\alpha(y_b)z}{1-\alpha(y_b)z}\right]^{\ell+1}. \quad (35)$$

By comparing (35) and (32), and using (33), we get the following upper bound for the moments of $\mu_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$:

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant c(y_b)\alpha(y_b)^\ell \sum_{i=0}^{\ell-1} \frac{\big(y_a c(y_b)\big)^i}{i+1}\binom{\ell}{i}\binom{\ell-1}{i} \leqslant c(y_b)c\big(y_a c(y_b)\big)\Big(\alpha(y_b)\alpha\big(y_a c(y_b)\big)\Big)^\ell. \quad (36)$$

Thus, we only need to show that the constants $\alpha(y)$ and $c(y)$ exist. For simplicity, suppose that $y < 1$. Then, the ratio of the ith and $(i-1)$th terms of the summand of (33) is the following:

$$y\frac{(\ell-i+1)(\ell-i)}{i(i+1)}. \quad (37)$$

The largest integer i for which this ratio stays greater or equal to 1, namely i^*, corresponds to the maximum summand in the aforementioned summation. Solving for i^* yields

$$i^* = \left\lfloor \frac{\sqrt{(2\ell y+y+1)^2+4\ell(\ell+1)y(1-y)}-(2\ell y+y+1)}{2(1-y)} \right\rfloor. \quad (38)$$

Hence, the ℓth moment can be lower/upper bounded as

$$m^{(\ell)}_{\mathsf{MP}}(y) \geqslant \frac{y^{i^*}}{i^*+1}\binom{\ell}{i^*}\binom{\ell-1}{i^*} \quad (39)$$

and

$$m^{(\ell)}_{\mathsf{MP}}(y) \leqslant \frac{\ell y^{i^*}}{i^*+1}\binom{\ell}{i^*}\binom{\ell-1}{i^*}. \quad (40)$$

Recall that $m^{(1)}_{\mathsf{MP}}(y) = 1$, so in the following, we assume that $\ell \geqslant 2$. We can use the following version of the Stirling's bounds on $n!$:

$$\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\exp\left(\frac{1}{12n+1}\right) \leqslant n! \leqslant \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\exp\left(\frac{1}{12n}\right)$$

in order to simplify the upper bound on $m^{(\ell)}_{\mathsf{MP}}(y)$. Letting $t_\ell := i^*/\ell$, we get

$$\begin{aligned} m^{(\ell)}_{\mathsf{MP}}(y) &\leqslant \frac{\ell^2}{2\pi i^*(\ell-i^*)}\Big(y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \times \exp\left(\frac{1}{6\ell}-\frac{1}{6(1-t_\ell)\ell+1}-\frac{1}{6t_\ell\ell+1}\right) \\ &< \frac{\ell^2}{2\pi(\ell-1)}\Big(y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \quad (41) \\ &< \frac{1}{\pi}\Big(2y^{t_\ell}4^{H(t_\ell)}\Big)^\ell \quad (42) \end{aligned}$$

where $H(x) := -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function, and the last inequality follows from $\ell < 2^\ell$ for all $\ell > 1$. Also, for a given value of y, one can find the value of ℓ from (38) which maximizes the expression $y^{t_\ell}4^{H(t_\ell)}$. Note that for $y < 1$, we have $y^{t_\ell}4^{H(t_\ell)} < 4$. Hence, by maximizing over ℓ, one can find the constants $\pi \leqslant \alpha(y) < 4$ and $c(y) = 1/\pi$ that satisfy the upper bound. The proof for $y > 1$ is very similar and is thus omitted for brevity. In general, for $y \in (0, \infty)$, one can show that there exists a constant $\alpha(y)$ such that $\pi < \alpha(y) \leqslant 4\max\{y, 1\}$ and $c(y) = 1/\pi$, satisfying the upper bound on the moments of the Marchenko–Pastur measure $\mu_{\mathsf{MP}}(x; y)$. The upper bound given in (36) implies that there exist constants $\gamma(y_a, y_b)$, $C(y_a, y_b)$ such that

$$m^{(\ell)}_{\boxtimes^2\mathsf{MP}}(y_a, y_b) \leqslant C(y_a, y_b)\gamma^\ell(y_a, y_b).$$

A loose estimate of these parameters is $C(y_a, y_b) = 1/\pi^2$ and $\gamma(y_a, y_b) = 16\max\{1, y_a\}\max\{1, y_b\}$, which clearly satisfies the upper bound of the lemma. This proves the statement of the lemma. ■

Proof of Theorem 2.3: In what follows, we may drop the dependence on n, y_a, and y_b for notational convenience. Let d'_e be the greatest even integer less than or equal to $d' = \min\{[(d'^a - 1)/4], [(d'^b - 1)/4]\}$ and let r be an integer such that $r \leqslant d'_e$ for all $n > n_0$, for some n_0.

Let $m_{\mathscr{G}}(t)$ and $m_{\boxtimes^2\mathsf{MP}}(t)$ denote the characteristic functions of the distributions $M_{\mathscr{G}}(x)$ and $M_{\boxtimes^2\mathsf{MP}}(x)$, respectively. Using the inequality

$$\left|\exp(it) - 1 - \frac{it}{1!} - \cdots - \frac{(it)^{r-1}}{(r-1)!}\right| \leqslant \frac{|t|^r}{r!} \quad (43)$$

which can easily be verified by induction on r, one can obtain the following bound on the tail of the characteristic function $m_{\mathscr{G}}(t)$:

$$\left|m_{\mathscr{G}}(t) - \sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!}\right| \leqslant m^{(r)}_{\mathscr{G}}\frac{|t|^r}{r!}. \quad (44)$$

From Lemma 3.4, we know that $m^{(\ell)}_{\mathscr{G}} \to m^{(\ell)}_{\boxtimes^2\mathsf{MP}}$ almost surely as $n \to \infty$, for $\ell = 1, 2, \ldots, r$. Hence

$$\limsup_n \left|m_{\mathscr{G}}(t) - \sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!}\right| \leqslant m^{(r)}_{\boxtimes^2\mathsf{MP}}\frac{|t|^r}{r!}. \quad (45)$$

We have

$$\begin{aligned} &\limsup_n \frac{1}{\pi}\int_0^T \Big|m_{\mathscr{G}}(t) - m_{\boxtimes^2\mathsf{MP}}(t)\Big|\frac{dt}{t} \\ &\leqslant \limsup_n \frac{1}{\pi}\int_0^T \frac{1}{t}\left|\sum_{\ell=0}^{r-1} m^{(\ell)}_{\mathscr{G}}\frac{(it)^\ell}{\ell!} + \sum_{\ell=r}^{\infty} m^{(\ell)}_{\boxtimes^2\mathsf{MP}}\frac{(it)^\ell}{\ell!} - m_{\boxtimes^2\mathsf{MP}}(t)\right| dt \\ &\quad + \frac{2}{\pi}\int_0^T m^{(r)}_{\boxtimes^2\mathsf{MP}}\frac{|t|^{r-1}}{r!}dt. \quad (46) \end{aligned}$$

It is straightforward to show that

$$f(t) := \frac{1}{t}\left|\sum_{\ell=0}^{r-1} m_{\mathscr{G}}^{(\ell)}\frac{(it)^\ell}{\ell!}+\sum_{\ell=r}^{\infty} m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\frac{(it)^\ell}{\ell!}-m_{\boxtimes^2\mathsf{MP}}(t)\right| \leqslant \frac{1}{t}\sum_{\ell=1}^{r-1}\left|m_{\mathscr{G}}^{(\ell)}-m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\right|\frac{|t|^{\ell-1}}{\ell!} =: g(t). \quad (47)$$

Moreover, we have

$$\int_0^T g(t)dt < \left(\sum_{\ell=1}^{r-1}\left|m_{\mathscr{G}}^{(\ell)}-m_{\boxtimes^2\mathsf{MP}}^{(\ell)}\right|\right)e^T. \quad (48)$$

From lemma 3.4, and by an application of the Markov's inequality, we have

$$\mathbb{P}\left(\left|m_{\mathscr{G}}^{(\ell)}-\mathbb{E}_{P_{\mathsf{G}}}\{m_{\mathscr{G}}^{(\ell)}\}\right| > \mathcal{O}\left(\frac{1}{\sqrt[8]{n}}\right)\right) \leqslant \mathcal{O}\left(\frac{1}{n^{3/2}}\right) \quad (49)$$

for all $\ell = 1, 2, \ldots, d'_e$, where d'_e is the greatest even integer less than or equal to $d' = \min\{[(d'^a-1)/4], [(d'^b-1)/4]\}$. Hence, we have

$$\left|m_{\mathscr{G}}^{(\ell)}-\mathbb{E}_{P_{\mathsf{G}}}\{m_{\mathscr{G}}^{(\ell)}\}\right| \leqslant \mathcal{O}\left(\frac{1}{\sqrt[8]{n}}\right) \quad (50)$$

for some large n, almost surely. Therefore, $\int_0^T g(t)dt$ is bounded almost surely. Now, applying the Fatou's lemma [12, p. 23] to the nonnegative sequence $g(t) - f(t)$ yields

$$\limsup_n \int_0^T f(t)dt \leqslant \int_0^T \limsup_n f(t)dt = 0 \quad (51)$$

almost surely. Hence, from (46) and the Zolotarev's inequality given in Lemma 3.1, we will have the following asymptotic bound:

$$\limsup_n \mathscr{L}\big(M_{\mathscr{G}_n}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\big) \leqslant \frac{2}{\pi r} m_{\boxtimes^2\mathsf{MP}}^{(r)}\frac{T^r}{r!}+2e\frac{\ln T}{T}. \quad (52)$$

Utilizing the upper bound of Lemma 3.5 on $m_{\boxtimes^2\mathsf{MP}}^{(r)}$ and optimizing the bound of (52) with respect to T yield

$$\frac{C(y_a, y_b)\gamma^r(y_a, y_b)}{\pi^2 r!}T^{r+1} = e(\ln T - 1) \quad (53)$$

which is a nonalgebraic equation in T. An approximate solution can be obtained by taking

$$T = \mathscr{D}\left(\frac{r}{e\gamma(y_a, y_b)}\right) \quad (54)$$

where

$$\mathscr{D} := \left\{\frac{\gamma(y_a, y_b)}{C(y_a, y_b)}\pi^2 e^2\frac{\big(\ln r - 1\big)}{r}\right\}^{\frac{1}{r+1}}. \quad (55)$$

With this choice of T, the asymptotic bound becomes

$$\limsup_n \mathscr{L}\left(M_{\mathscr{G}}(x), M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right) \leqslant \mathscr{A}\frac{1}{r}+\mathscr{B}\frac{\ln r}{r}$$

where

$$\mathscr{A} = \frac{2C(y_a, y_b)}{\pi^2}\mathscr{D}^r+\frac{2e^2\gamma(y_a, y_b)\big(\ln\mathscr{D}-\ln\gamma(y_a, y_b)-1\big)}{\mathscr{D}}$$

and

$$\mathscr{B} := \frac{2e^2\gamma(y_a, y_b)}{\mathscr{D}}. \quad (56)$$

In particular, if we let $r = \liminf_n r_n$, the aforementioned bound holds as $n \to \infty$. This proves the statement of Theorem 2.3.

Proof of Theorem 2.4: The proof is identical to the proof of Theorem 2.3, up to the application of Zolotarev's inequality given in Lemma 3.1.

Recall from Lemma 2.2 that the density $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ has a point mass of $\big(1-\min\{1, 1/y_a, 1/y_b\}\big)$ at $x = 0$. Hence, denoting the nonatomic portion of this density by $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, the full density is given by

$$\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \Big(1-\min\big\{1, 1/y_a, 1/y_b\big\}\Big)\delta_0 + \underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b). \quad (57)$$

Hence, by taking only the nonatomic part $\underline{\mu}_{\boxtimes^2\mathsf{MP}}$ with appropriate renormalization, one gets the density

$$\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}}\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b).$$

Now, consider $\mathbf{A} \in \{-1, 1\}^{N_a\times n}$ and $\mathbf{B} \in \{-1, 1\}^{N_b\times n}$ based on the linear codes $\mathcal{C}^a$ and $\mathcal{C}^b$, respectively. Similarly, since the rank of the matrix $\big(\frac{1}{N_a}\mathbf{A}^*\mathbf{A}\big)\big(\frac{1}{N_b}\mathbf{B}^*\mathbf{B}\big)$ is at most $\min\{N_a, N_b, n\}$, the asymptotic density has a point mass of at least $\big(1-\min\{1, \frac{1}{y_a}, \frac{1}{y_b}\}\big)$ at $x = 0$. Hence, by removing $\big(1-\min\{1, \frac{1}{y_a}, \frac{1}{y_b}\}\big)$ from its point mass, and appropriate renormalization, one gets the density

$$\widehat{\mu}_{\mathscr{G}}(x) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}}\left(\mu_{\mathscr{G}}(x)-\Big(1-\min\big\{1, \tfrac{1}{y_a}, \tfrac{1}{y_b}\big\}\Big)\delta_0\right).$$

Let the distribution functions corresponding to $\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and $\widehat{\mu}_{\mathscr{G}}(x)$ be denoted by $\widehat{M}_{\mathscr{G}}(x)$ and $\widehat{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, respectively. It is easy to see that

$$\left|M_{\mathscr{G}}(x)-M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right| = \min\left\{1, \tfrac{1}{y_a}, \tfrac{1}{y_b}\right\}\left|\widehat{M}_{\mathscr{G}}(x)-\widehat{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)\right|. \quad (58)$$

Since $\mu_{\mathsf{MP}}(x; y)$ is supported on the interval $\big[(1-\sqrt{y})^2, (1+\sqrt{y})^2\big]$, by basic linear algebra, one can show that the support of $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is contained in the interval $\big[(1-\sqrt{y_a})^2(1-\sqrt{y_b})^2, (1+\sqrt{y_a})^2(1+\sqrt{y_b})^2\big]$. Also, by the cubic function root formula, since the coefficient of the leading term of (14) has its root at $z = 0$, the density has only a pole at $x = 0$ [11]. Hence, if $x = 0$ is not contained in the support of $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, it is clearly bounded. The condition of $y_a, y_b \neq 1$ implies that the lower bound on the support given by $(1-\sqrt{y_a})^2(1-\sqrt{y_b})^2$ is positive. Hence, $\underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ and consequently the density

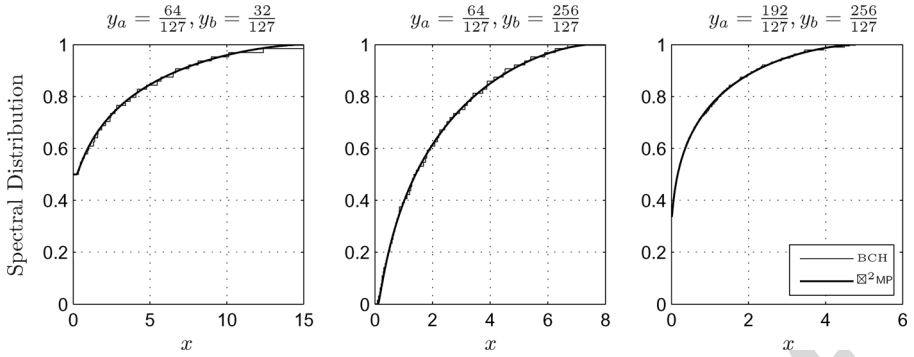

Fig. 1. Empirical spectral distribution of the product of two random matrices based on the $[127, 99, 9]$ and $[127, 92, 11]$ BCH codes versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$.

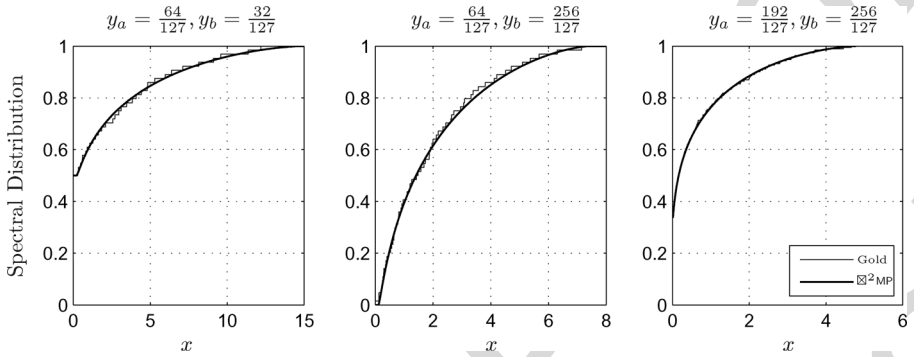

Fig. 2. Empirical spectral distribution of the product of two random matrices based on the $[127, 14, 56]$ Gold code versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$, for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$.

$\widehat{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ are bounded, and therefore, Lemma 3.2 can be used (instead of the Zolotarev's inequality) in order to bound the Kolmogorov distance of the two distributions $\widehat{M}_{\mathscr{G}}$ and $\widehat{M}_{\boxtimes^2\mathsf{MP}}$. The resulting upper bound is similar to that of the main theorem of [2], and is repeated here for completeness. Let

$$\tau(y_a, y_b) := \frac{1}{\min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\}} \sup_x \underline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) < \infty.$$

Then, we have the following almost sure bound for all x:

$$\limsup_n \left| M_{\mathscr{G}}(x) - M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b) \right| \leqslant \mathscr{C}\left(\frac{1}{r} + \frac{1}{r^2}\right)$$

where

$$\mathscr{C} := \frac{24\left\{e \min\left\{1, \frac{1}{y_a}, \frac{1}{y_b}\right\} \tau(y_a, y_b) \gamma(y_a, y_b)\right\}^{\frac{r}{r+1}}}{\pi\left\{\frac{6}{C(y_a, y_b) r}\right\}^{\frac{1}{r+1}}} \tag{59}$$

and $\gamma(y_a, y_b)$ and $C(y_a, y_b)$ are the same as in Lemma 3.5. Hence, the statement of Theorem 2.4 follows.

IV. Numerical Experiments

A. BCH Codes

As for the first set of experiments, we consider random matrices from binary BCH codes. It is known that certain BCH codes have large dual distances [7]. In particular, the Carlitz–Uchiyama bound [7] implies that a binary BCH code of length $n := 2^m - 1$ and designed distance $2t + 1$ with $2t - 1 < 2^{\lceil m/2 \rceil} + 1$ has a dual distance of at least $2^{m-1} - (t-1)2^{m/2}$. For instance, for the $[127, 99, 9]$, the Carlitz–Uchiyama bound implies $d' \geqslant 31$. BCH codes are known for their empirically random-like properties, and, for instance, have been successfully used in the design of near-capacity achieving turbo block codes [10]. Fig. 1 shows the empirical spectral distribution of the product of two random matrices based on the $[127, 99, 9]$ and $[127, 92, 11]$ BCH codes versus $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$. In order to obtain the analytical curves, $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is first obtained numerically from (14), by the cubic function root formula and using (57). Then, $\overline{M}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is obtained by numerical integration of $\overline{\mu}_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$. Finally, $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$ is obtained from the relation given in (18). As predicted by the theorems of this paper, the spectral distribution of the product of two random matrices from BCH codes resembles $M_{\boxtimes^2\mathsf{MP}}(x; y_a, y_b)$.

TABLE I
Distance Distribution of the Gold Code

i	B_i
$2^{m-1} - 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} + 2^{(m-3)/2})$
2^{m-1}	$(2^m - 1)(2^{m-1} + 1)$
$2^{m-1} + 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} - 2^{(m-3)/2})$

B. Gold Sequences

Gold sequences are a class of pseudorandom sequences which can be obtained by XOR-ing the shifted versions of two shift register sequences generated by two distinct primitive polynomials [6]. Let $h_1(x)$ and $h_2(x)$ be two primitive polynomials of degree m over $\mathrm{GF}(2)$, such that $h_1(\alpha) = 0$ and $h_2(\alpha^\delta) = 0$ for some integer δ. Suppose that $m \neq 0 \bmod 4$. If $\delta = 2^h + 1$ or $\delta = 2^{2h} - 2^h + 1$, and m/e is odd, with $e := \gcd(m, h)$, then the two polynomials $h_1(x)$ and $h_2(x)$ are denoted by the preferred pair of polynomials. Let $\mathbf{u}$ and $\mathbf{v}$ denote two shift register sequences of length $2^m - 1$, corresponding to the preferred pair of polynomials $h_1(x)$ and $h_2(x)$, respectively. Then, the set of Gold sequences $\mathcal{G}(\mathbf{u}, \mathbf{v})$ is defined as

$$\mathcal{G}(\mathbf{u}, \mathbf{v}) := \{\Re^a \mathbf{u}, \Re^b \mathbf{v}, \text{ or } \Re^a \mathbf{u} \oplus \Re^b \mathbf{v} | 0 \leqslant a, b \leqslant 2^m - 2\}$$

where $\oplus$ and $\Re$ denote the binary XOR and cyclic shift operators, respectively. The set $\mathcal{G}(\mathbf{u}, \mathbf{v})$ consists of $2^{2m} - 1$ binary sequences of length $2^m - 1$, with desirable cross-correlation properties [6].

The set $\mathcal{G}(\mathbf{u}, \mathbf{v}) \cup \{\mathbf{0}\}$ is a binary block code denoted by the Gold code. The dual of the Gold code, for m odd, is the $[2^m - 1, 2^m - 2m - 1, 5]$ double-error-correcting BCH code [7]. Hence, the dual distance of the gold code for odd m is 5, independent of its length. For m odd, the distance distribution of the Gold code is given in Table I.

Fig. 2 shows the empirical spectral distribution of the product of two random matrices based on the $[127, 14, 56]$ Gold code versus $M_{\boxtimes^2 \mathsf{MP}}(x; y_a, y_b)$ for $(y_a, y_b) = (\frac{64}{127}, \frac{32}{127})$, $(\frac{64}{127}, \frac{256}{127})$, and $(\frac{192}{127}, \frac{256}{127})$. Although the dual distance of the Gold code is $d' = 5$, the spectral distribution of the product of two random matrices from the Gold code curiously resembles $M_{\boxtimes^2 \mathsf{MP}}(x; y_a, y_b)$. This observation suggests that the condition of large dual distance is only a sufficient condition for the results of this paper to hold. The problems of strengthening this sufficient condition or finding necessary conditions remain open.

ACKNOWLEDGMENT

The authors would like to thank N. Raj Rao for insightful suggestions, Roland Speicher and Horng-Tzer Yau for instructive comments, and Olgica Milenkovic and Alexander Barg for useful discussions.

REFERENCES

[1] G. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge, U.K.: Cambridge Univ. Press, 2009, to be published.

[2] B. Babadi and V. Tarokh, "Spectral distribution of random matrices from binary linear block codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3955–3962, Jun. 2011.

[3] B. Babadi and V. Tarokh, "Spectral distribution of the product of two random matrices based on binary block codes," in *Proc. 49th Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 27–30, 2011, pp. 917–919.

[4] Z. D. Bai, B. Miao, and B. Jin, "On limit theorem for the eigenvalues of product of two random matrices," *J. Multivariate Anal.*, vol. 98, pp. 76–101, 2007.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1991, vol. 2.

[6] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 154–156, Jan. 1968.

[7] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1988.

[8] V. A. Marchenko and L. A. Pastur, "The distribution of eigenvalues in certain sets of random matrices," *Math. USSR-Sbornik*, vol. 72, pp. 507–536, 1967.

[9] S. Pafka, M. Potters, and I. Kondor, Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization 2004 [Online]. Available: arxiv: cond-mat/0402573

[10] R. M. Pyndiah, "Near-optimum decoding of product codes: Block turbo codes," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1003–1010, Aug. 1998.

[11] N. Raj Rao and A. Edelman, "The polynomial method for random matrices," *Found. Comput. Math.*, vol. 8, no. 6, pp. 649–702, 2008.

[12] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[13] V. M. Sidel'nikov, "Weight spectrum of binary Bose-Chaudhuri-Hoquinghem codes," *Problems Inf. Transmiss.*, vol. 7, pp. 11–17, 1971.

[14] J. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.

[15] R. Speicher, Combinatorial aspects of free probability theory. Goettingen, Germany, 2005 [Online]. Available: http://www.math.uni-sb.de/ag/speicher/speicher_publ_surveys.html

[16] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA: Now Publishers Inc., 2004, Foundations and Trends in Communications and Information Theory.

[17] B. Valkó, Lecture notes on random matrices 2009 [Online]. Available: http://www.math.wisc.edu/\;valko/courses/833/833.html

[18] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*. Providence, RI: Amer. Math. Soc., 2002, CRM Monographs.

[19] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Ann. Math.*, vol. 62, no. 3, pp. 548–564, Nov. 1955.

[20] V. M. Zolotarev, "Estimates of the difference between distributions in the Lévy metric," *Collect. Articles. Part I, Trudy Mat. Inst. Steklov.*, vol. 112, pp. 224–231, 1971.

Behtash Babadi (S'08) received the Ph.D. and M.Sc. degrees in engineering sciences from Harvard University, Cambridge, MA, USA in 2011 and 2008, respectively, and the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran in 2006. He is currently a post-doctoral fellow at the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology, Cambridge, MA, USA and the Department of Anesthesia, Critical Care and Pain Medicine at Massachusetts General Hospital, Boston, MA, USA and a research associate at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His research interests include biological signal processing, adaptive signal processing, information theory, and compressed sensing.

Vahid Tarokh (M'97–SM'02–F'09) received the M.Sc. degree from the University of Windsor, Windsor, ON, Canada, in 1992, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, in 1995. He was with AT&T Labs-Research until August 2000, where he was Head of the Department of Wireless Communications and Signal Processing. He then joined the Department of Electrical Engineering and Computer Sciences (EECS) at the Massachusetts Institute of Technology, Cambridge, as an Associate Professor. In 2002, he joined Harvard University as a Professor and Senior Fellow. His recent research interest includes non-linear information theory and communications, environmentally friendly networks, non-linear signal processing, Tomography, Capsule Endoscopy, Body Area Networks, EEG Signal Analysis and Brain Computer Interaction. His recent awards include a 2011 Guggenheim fellowship in Applied Mathematics, the 2012 IEEE COMSOC Cognitive Networks Technical Committee Publications Award, and the 2013 IEEE Eric E. Sumner Award. He holds two honorary degrees.