Signal Shaping for BICM at Low SNR

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

IEEE Transactions on Information Theory (ISSN: 0018-9448)

Citation for the published paper:

http://dx.doi.org/10.1109/TIT.2012.2231900

Downloaded from: http://publications.lib.chalmers.se/publication/176217

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.
Signal Shaping for BICM at Low SNR

Erik Agrell and Alex Alvarado

Abstract—The generalized mutual information (GMI) of bit-interleaved coded modulation (BICM) systems, sometimes called the BICM capacity, is investigated at low signal-to-noise ratio (SNR). The combinations of input alphabet, input distribution, and binary labeling that achieve the Shannon limit ~1.59 dB are completely characterized. The main conclusion is that a BICM system with probabilistic shaping achieves the Shannon limit at low SNR if and only if it can be represented as a zero-mean linear projection of a hypercube. Hence, probabilistic shaping offers no extra degrees of freedom to optimize the low-SNR BICM-GMI, in addition to what is provided by geometrical shaping. The analytical conclusions are confirmed by numerical results, which also show that for a fixed input alphabet, probabilistic shaping can improve the BICM-GMI in the low and medium SNR range.

Index Terms—Binary labeling, bit-interleaved coded modulation, generalized mutual information, Hadamard transform, probabilistic shaping, Shannon limit, wideband regime.

I. INTRODUCTION

The most important breakthrough for coded modulation (CM) in fading channels came in 1992, when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [1], usually referred to as a pragmatic approach for CM [2], [3]. Despite not being fully understood theoretically, BICM has been rapidly adopted in commercial systems such as wireless and wired broadband access networks, 3G/4G telephony, and digital video broadcasting, making it the de facto standard for current telecommunications systems [3] Ch. 1.

Signal shaping refers to the use of non-equally spaced and/or non-equally likely symbols, i.e., geometrical shaping and probabilistic shaping, resp. Signal shaping has been studied during many years, cf. [4], [5] and references therein. In the context of BICM, geometrical shaping was studied in [6]–[8], and probabilistic shaping, i.e., varying the probabilities of the bit streams, was first proposed in [9], [10] and developed further in [11]–[14]. Probabilistic shaping offers another degree of freedom in the BICM design, which can be used to make the discrete input distribution more similar to the optimal distribution (which is in general unknown). This is particularly advantageous at low and medium SNR.

For the additive white Gaussian noise (AWGN) channel, the so-called Shannon Limit (SL) ~1.59 dB represents the average bit energy-to-noise ratio needed to transmit information reliably when the signal-to-noise ratio (SNR) tends to zero [15], [16], i.e., in the wideband regime. When discrete input alphabets are considered at the transmitter and a BICM decoder is used at the receiver, the SL is not always achieved as first noticed in [17]. This was later shown to be caused by the selection of the binary labeling [18]. The behavior of BICM in the wideband regime was studied in [17]–[21] as a function of the alphabet (\(\mathbb{X}\)) and the binary labeling (\(\mathbb{L}\)), assuming a uniform input distribution. First-order optimal (FOO) constellations were defined in [21] as the triplet \([\mathbb{X}, \mathbb{P}, \mathbb{L}]\) that make a BICM system achieve the SL, where \(\mathbb{P}\) represents the input distribution.

In this paper, the results of [21] are generalized to nonuniform input distributions and give a complete characterization of FOO constellations for BICM in terms of \([\mathbb{X}, \mathbb{P}, \mathbb{L}]\). More particularly, the geometrical and/or probabilistic shaping rules that should be applied to a constellation to make it FOO are found. The main conclusion is that probabilistic shaping offers no extra degrees of freedom in addition to what is provided by geometrical shaping for BICM in the wideband regime.

II. PRELIMINARIES

A. Notation

Bold italic letters \(\mathbf{x}\) denote row vectors. Block letters \(\mathbf{X}\) denote matrices or sometimes column vectors. The identity matrix is \(I\). The inner product between two row vectors \(\mathbf{a}\) and \(\mathbf{b}\) is denoted by \(\langle \mathbf{a}, \mathbf{b} \rangle\) and their element-wise product by \(\mathbf{a} \odot \mathbf{b}\). The Euclidean norm of the vector \(\mathbf{a}\) is denoted by \(\|\mathbf{a}\|\). Random variables are denoted by capital letters \(X\) and random vectors by boldface capital vectors \(\mathbf{X}\). The probability density function (pdf) of the random vector \(Y\) is denoted by \(p_Y(y)\) and the conditional pdf by \(p_{Y|X}(y|x)\). A similar notation applies to probability mass functions of a random variable, which are denoted by \(P_Y(y)\) and \(P_{Y|X}(y|x)\). Expectations are denoted by \(\mathbb{E}\).

The empty set is denoted by \(\emptyset\) and the binary set by \(\mathcal{B} = \{0, 1\}\). The negation of a bit \(b\) is denoted by \(\overline{b} = 1 - b\). Binary addition (exclusive-OR) of two bits \(a\) and \(b\) is denoted by \(a \oplus b\). The same notation \(a \oplus b\) denotes the integer that results from taking the bitwise exclusive-or of two integers \(a\) and \(b\).

B. System Model

We consider transmissions over a discrete-time memoryless vectorial fast fading channel. The received vector at any discrete time instant is

\[
\mathbf{Y} = \mathbf{H} \odot \mathbf{X} + \mathbf{Z}
\]  

(1)

where \(\mathbf{X}\) is the channel input and \(\mathbf{Z}\) is Gaussian noise with zero mean and variance \(N_0/2\) in each dimension [11], [13].

Research supported by The British Academy and The Royal Society (via the Newton International Fellowship scheme), U.K., and by the European Community’s Seventh’s Framework Programme (FP7/2007-2013) under grant agreement No. 271986. This work was presented in part at the Information Theory and Applications (ITA) Workshop, San Diego, CA, February 2012, and at the IEEE International Symposium on Information Theory, Cambridge, MA, July 2012.

E. Agrell is with the Dept. of Signals and Systems, Chalmers Univ. of Technology, SE-41296 Göteborg, Sweden (email: agrell@chalmers.se). A. Alvarado is with the Dept. of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom (email: alex.alvarado@ieee.org).
The channel is represented by the $N$-dimensional vector $H$. It contains the real fading coefficients $H_i$, which are random, possibly dependent, with the same pdf $p_H(h)$. We assume that $H$ and $N_0$ are perfectly known at the receiver or can be perfectly estimated, and that the technical requirements on $X$ and $H$ in [21] Sec. I-D] are satisfied.

The conditional transition pdf of the channel in (1) is

$$p_{Y|X,H}(y|x, h) = \frac{1}{(N_0\pi)^{N/2}} \exp \left( -\frac{\|y - h \otimes x\|^2}{N_0} \right).$$

The SNR is defined as

$$\rho \triangleq \frac{E[H^2]}{E_0} = \frac{R_c E_t^2}{N_0}$$

where $E_a \triangleq E[\|X\|^2]$ is the average transmitted symbol energy, $R_c$ is the transmission rate in information bits per symbol, and $E_t^2 \triangleq E[H^2]E_a/R_c$ is the average received energy per information bit.

The generic BICM scheme in Fig. 1 is considered. The transceiver is, in the simplest case, a single binary encoder concatenated with an interleaver and a memoryless mapper $\Phi$. Multiple encoders and/or interleavers may be needed to achieve probabilistic shaping [10–13]. At the receiver, using the channel output $Y$, the demapper $\Phi^{-1}$ computes metrics $L_k$ for the individual coded bits $C_k$ with $k = 0, \ldots, m-1$, usually in the form of logarithmic likelihood ratios. These metrics are then passed to the deinterleaver(s) and decoder(s) to obtain an estimate of the information bits.

The mapper $\Phi$ is defined via the input alphabet $X = \{x_0, \ldots, x_{M-1}\}$, with $m$ bits used to index the $M = 2^m$ symbols vectors $x_i \in \mathbb{R}^N$ for $i = 0, \ldots, M-1$. We associate with each symbol $x_i$ the codeword (binary labeling) $c_i \triangleq [c_{i,0}, \ldots, c_{i,m-1}] \in \mathbb{B}^m$ and the probability $0 \leq P_i \leq 1$, where $P_i \triangleq p_X(x_i)$. The binary labeling is denoted by $L = [c_0^T, \ldots, c_{M-1}^T] \in \mathbb{B}^{M \times m}$ and the input distribution by $\mathbb{P} = [P_0, \ldots, P_{M-1}] \in [0,1]^M$.

In the following, the labeling used throughout this paper is defined. This can be done without loss of generality, as will be explained in Sec. II.C

**Definition 1 (Natural binary code):** The natural binary code (NBC) is the binary labeling $\mathbb{N}_m \triangleq [n(0)^T, \ldots, n(M-1)^T]^T$, where $n(i) = [n_{i,0}, \ldots, n_{i,m-1}] \in \mathbb{B}^m$ denotes the base-2 representation of the integer $0 \leq i \leq M-1$, with $n_{i,m-1}$ being the most significant bit.

This definition of the NBC is different from the one in [21]. The difference lies only in the bit ordering, i.e., in this paper we consider the last column of $\mathbb{N}_m$ to contain the most significant bits of the base-2 representation of the integers $i = 0, 1, \ldots, M-1$. It follows from Definition 1 that

$$n_{2^l,k} = \begin{cases} 1, & k = l, \\ 0, & k \neq l \end{cases}$$

for $k = 0, \ldots, m-1$ and $l = 0, \ldots, m-1$, and

$$n_{i,j,k} = n_{i,k} \oplus n_{j,k}$$

for $i = 0, \ldots, M-1, j = 0, \ldots, M-1$ and $k = 0, \ldots, m-1$.

**C. Probabilistic Shaping in BICM**

Assuming independent, but possibly nonuniformly distributed, bits $C_0, \ldots, C_{m-1}$ at the input of the modulator (cf. Fig. 1), the symbol probabilities are given by [21] eq. (8) [13] eq. (8) [22] eq. (9)

$$P_i = \prod_{k=0}^{m-1} P_{C_k}(c_{i,k})$$

for $i = 0, \ldots, M-1, m - 1$, where $P_{C_k}(u)$ for $u \in \mathbb{B}$ is the probability of $C_k = u$. Since $P_{C_k}(1) = 1 - P_{C_k}(0)$, the distribution $\mathbb{P}$ is fully specified by the vector of bit probabilities $b \triangleq [P_{C_0}(0), \ldots, P_{C_{m-1}(0)}]$.

Throughout this paper, we assume that $0 < P_{C_k}(0) < 1$ for all $k = 0, \ldots, m-1$; i.e., all constellation points are used with a nonzero probability. This can be done without loss of generality, because if $P_{C_k}(0) = 0$ or $P_{C_k}(0) = 1$ for some $k$, then half of the constellation points will never be transmitted. If this is the case, the corresponding branches in Fig. 1 are removed, $m$ is reduced by one, and the mapper $\Phi$ is redefined accordingly. The result is another BICM scheme with identical performance, which satisfies $0 < P_{C_k}(0) < 1$ for all $k$.

For any constellation $\{X, \mathbb{P}, L\}$, a set of equivalent constellations can be constructed by permuting the rows of $X$, $L$, and $\mathbb{P}$, provided that the same permutation is applied to all three matrices. Specifically, denote the permutation that maps the NBC into the desired labeling $L$ by $\Pi$, i.e., $\Pi(\mathbb{N}_m) = L$. The BICM system defined by the alphabet $\Pi(X)$, the distribution $\Pi(\mathbb{P})$, and the labeling $\Pi(\mathbb{N}_m) = L$ is entirely equivalent to the system with alphabet $X$, distribution $\mathbb{P}$, and labeling $\mathbb{N}_m$. Without loss of generality, the analysis in this paper is therefore restricted to the latter case.

Based on the previous discussion, from now on we use the name constellation to denote the pair $[X, \mathbb{P}]$, where the NBC labeling is implicit. Thus, $L = \mathbb{N}_m$ and $c_{i,k} = n_{i,k}$ for all $i$.
and $k$, which simplifies the analysis. Note that $\mathbb{P}$ cannot be chosen arbitrarily in BICM; only distributions that satisfy

$$P_i = \prod_{k=0}^{m-1} P_{C_k}(n_i,k)$$

(6)

for some vector of bit probabilities $b$ will be considered in the paper. An important special case is the uniform distribution, for which $b = [1/2, \ldots, 1/2]$ and $\mathbb{P} = \mathbb{U}_m \triangleq [1/M, \ldots, 1/M]^T$.

### D. The Hadamard Transform

The Hadamard transform (HT), or Walsh–Hadamard transform, is a discrete, linear, orthogonal transform, whose coefficients take values in $\pm 1$. It is popular in image processing and can be used to analyze various aspects of binary labelings in digital communications and source coding.

**Definition 2:** The HT $\hat{X} = [\hat{x}_0^T, \ldots, \hat{x}_{M-1}^T]^T$ of a matrix (or vector) $X = [x_0^T, \ldots, x_{M-1}^T]^T$ with $M = 2^m$ rows is

$$\hat{x}_i \triangleq \frac{1}{M} \sum_{j=0}^{M-1} x_j h_{i,j}, \quad i = 0, \ldots, M - 1$$

(7)

where for all $i = 0, \ldots, M - 1$ and $j = 0, \ldots, M - 1$

$$h_{i,j} = \prod_{k=0}^{m-1} (-1)^{n_i,k n_j,k}.$$  

(8)

Because $n_{0,k} = 0$ for $k = 0, \ldots, m - 1$, setting $i = 0$ in (7–8) shows that the first HT vector

$$\hat{x}_0 = \frac{1}{M} \sum_{j=0}^{M-1} x_j$$

(9)

can be interpreted as the uniformly weighted mean of the alphabet. This is a property that the HT shares with, e.g., the discrete Fourier transform.

It can be shown from (8) that

$$\sum_{i=0}^{M-1} h_{i,l} h_{i,j} = \begin{cases} M, & j = l, \\ 0, & j \neq l \end{cases}$$

(10)

for all $j = 0, \ldots, M - 1$ and $l = 0, \ldots, M - 1$. Therefore, the inverse transform is identical to the forward transform, apart from a scale factor:

$$x_j = \sum_{i=0}^{M-1} \hat{x}_i h_{i,j}, \quad j = 0, \ldots, M - 1.$$  

(11)

### E. A New Transform

In this section, we define a linear transform between vectors or matrices, which depends on the input distribution $\mathbb{P}$ via the bit probabilities $b$. Its usage will become clear in Section III-C.

**Definition 3:** Given the bit probabilities $b = [P_{C_0}(0), \ldots, P_{C_{m-1}}(0)]$, the transform $\tilde{X} = [\tilde{x}_0^T, \ldots, \tilde{x}_{M-1}^T]^T$ of a matrix (or vector) $X = [x_0^T, \ldots, x_{M-1}^T]^T$ is

$$\tilde{x}_i \triangleq \sum_{j=0}^{M-1} x_j g_{i,j} \sqrt{P_j}, \quad i = 0, \ldots, M - 1$$

(12)

where $P_j$ is given by (5). The coefficients $g_{i,j}$ are defined as

$$g_{i,j} \triangleq \prod_{k=0}^{m-1} (1)^{n_i,k n_j,k} \sqrt{P_{C_k}(0)} + (1)^{n_i,k n_j,k} \sqrt{P_{C_k}(1)}$$

(13)

for all $i = 0, \ldots, M - 1$ and $j = 0, \ldots, M - 1$, where the bars represent negation ($\bar{b} = 1 - b$, see Sec. II-A).

**Remark 1:** For equally likely symbols, i.e., $\mathbb{P} = \mathbb{U}_m$, the transform becomes the identity operation $\tilde{X} = X$, because then $g_{i,i} = \sqrt{M}$ for $i = 1, \ldots, M$ and $g_{i,j} = 0$ for $i \neq j$.

The transform coefficients $g_{i,j}$ are nonsymmetric in the sense that in general $g_{i,j} \neq g_{j,i}$. They have some appealing properties given by the following lemma, which will be used in the proofs of Theorems 3, 4, and 8.

**Lemma 1:** For any $j = 0, \ldots, M - 1$ and $l = 0, \ldots, M - 1$,

$$\sum_{i=0}^{M-1} h_{i,l} g_{i,j} = \begin{cases} M, & j = l, \\ 0, & j \neq l \end{cases}$$

(14)

$$\sum_{i=0}^{M-1} h_{i,l} g_{i,j} = M h_{l,t} \sqrt{P_j \oplus t}$$

(15)

where $P_j$ is given by (5) and $h_{l,t}$ is defined in (8).

**Proof:** See the Appendix. \hfill $\Box$

We pay particular attention to two important special cases of (15). First, if $l = 0$, then $h_{0,l} = h_{0,t} = 1$ and $P_{j \oplus t} = P_j$ for $j = 0, \ldots, M - 1$. Second, if $l = 2^k$ for any integer $k = 0, \ldots, m - 1$, then by (8), $h_{l,i} = h_{i,l} = (-1)^{n_{i,k}}$ for any $i = 0, \ldots, M - 1$ and by (5)

$$P_j \oplus t = \prod_{k=0}^{m-1} P_{C_k}(n_{j \oplus 2^k,k}).$$

Using first (5) and then (4), we obtain

$$P_j \oplus t = \left( \prod_{k=0}^{m-1} P_{C_k}(n_{j,k}) \right) \frac{P_{C_k}(n_{j,k} \oplus 1)}{P_{C_k}(n_{j,k})} = P_j \frac{P_{C_k}(n_{j,k})}{P_{C_k}(n_{j,k})}. $$

Substituting these two cases ($l = 0$ and $l = 2^k$) into (15) proves the following corollary.

**Corollary 2:** For any $j = 0, \ldots, M - 1$,

$$\sum_{i=0}^{M-1} g_{i,j} = M \sqrt{P_j}$$

(16)

$$\sum_{i=0}^{M-1} (-1)^{n_{i,k}} g_{i,j} = M (-1)^{n_{j,k}} \sqrt{P_j \frac{P_{C_k}(n_{j,k})}{P_{C_k}(n_{j,k})}}.$$  

(17)

The fact that the sums $\sum_{i=0}^{M-1} g_{i,j} = 0$ whenever $j \neq l$, independently of the input distribution, implies that the coefficients $g_{i,j}$ form an orthogonal basis. As a consequence, the transform is invertible, as shown in the next theorem.

**Theorem 3:** The inverse transform $\tilde{X} = [\tilde{x}_0^T, \ldots, \tilde{x}_{M-1}^T]^T$ of a matrix (or vector) $X = [x_0^T, \ldots, x_{M-1}^T]^T$ is, given the bit
It is readily verified that $G$ is upper triangular in matrix notation. The mean values in each column of (19) are nonzero, completes the proof.

**Proof:** For $j = 0, \ldots, M - 1$,

$$
\sum_{i=0}^{M-1} \bar{x}_i g_{i,j} = \sum_{i=0}^{M-1} g_{i,j} \sum_{i=0}^{M-1} \bar{x}_i \sqrt{P_i} = \sum_{i=0}^{M-1} \bar{x}_i \sqrt{P_i} \sum_{i=0}^{M-1} g_{i,j}.
$$

Applying (14) and dividing both sides by $M \sqrt{P_j}$, which by Sec. II-C is nonzero, completes the proof. \hfill \square

**Example 1:** If the bit probabilities are $b = [0.35, 0.50]$, then the symbol probabilities $p$ are $P = [0.175, 0.325, 0.175, 0.325]^T$. The transform coefficients $g_{i,j}$ in (13) are the elements at row $i$, column $j$ of $G$.

It is readily verified that $G^T G = M I$, which is (14) in matrix notation. The mean values in each column of (19) are $[0.418, 0.570, 0.418, 0.570]^T$, which in agreement with (16) are the square roots of the elements in $P$. Similarly, it can be shown that $G$ in (19) satisfies (15) and (17).

If the Gray-labeled 4-ary pulse amplitude modulation (PAM) constellation $[X, P]$ is considered, $X = [-3, -1, 3, 1]^T$. Rewriting (12) in matrix notation, the transform can be calculated as $X = G D^{1/2} \Xi = [-2.654, -0.746, 2.654, 0.746]^T$, where $D \triangleq \text{diag}(P)$. This nonequally spaced 4-PAM alphabet will be illustrated and analyzed in Example 2. The inverse transform (13) can be written as $\Xi = (1/M) D^{-1/2} G^T X$. For a uniform distribution, $G = D^{-1/2} = \sqrt{M}I$, which agrees with Remark 1.

In Sec. IV-B we will need to apply the HT and the new transform after each other to the same alphabet. However, the two transforms do not commute, and the result will therefore depend on which order the transforms are applied. Of particular interest for our analysis is the setup in Fig. 2 where $X$ and $S$ are related via the transform defined above. Their HTs $\tilde{X}$ and $\tilde{S}$ are however not related via the same transform. Instead, a relation between $X$ and $S$ can be established via the following theorem.

**Theorem 4:** If $S = \tilde{X}$, then their HTs $\tilde{S}$ and $\tilde{X}$ satisfy

$$
\tilde{s}_i = \psi_i \sum_{j=0}^{M-1} \tilde{x}_j \prod_{k=0}^{m-1} (P_{C_k}(0) - P_{C_k}(n_{j,k})),
$$

$$
i = 0, \ldots, M - 1,
$$

and a product over $\varnothing$ is defined as 1.

**Proof:** See the Appendix. \hfill \square

**Remark 2:** The summation in (20) can be confined to $\sum_{j=i}^{M-1}$, because whenever $j < i$, there exists at least one bit position $k$ for which $n_{j,k} \neq n_{i,k} = 0$. Analogously, the summation in (21) can be confined to $\sum_{j=i}^{M-1}$.

**Example 2:** Expression (20) can be written as $\tilde{S} = T \tilde{X}$, or $\tilde{X} = T^{-1} \tilde{S}$. The element at row $i$, column $j$ of $T$ and $T^{-1}$ are given by (20) and (21) as, resp., $\psi_i \prod_{k=0}^{m-1} (P_{C_k}(0) - P_{C_k}(n_{i,k}))$ and $(1/\psi_i) \prod_{k=0}^{m-1} (P_{C_k}(n_{i,k}) - P_{C_k}(0))$. With $b$, $X$ and $\tilde{X}$ from Example 1 we obtain $[\psi_0, \ldots, \psi_3] = [1, 0.954, 1, 0.954]$ and

$$
\begin{bmatrix}
1 & -0.300 & 0 & 0 \\
0 & 0.954 & 0 & 0 \\
0 & 0 & 1 & -0.300 \\
0 & 0 & 0 & 0.954
\end{bmatrix},
T^{-1} =
\begin{bmatrix}
1 & 0.315 & 0 & 0 \\
0 & 1.048 & 0 & 0 \\
0 & 0 & 1 & 0.315 \\
0 & 0 & 0 & 1.048
\end{bmatrix}
$$

which, as predicted by Remark 2, are upper triangular.

Another relation between $\tilde{X}$ and $\tilde{S}$ can be deduced from Fig. 2. Defining the Hadamard matrix $H$ as the matrix with elements $h_{i,j}$ for $i, j = 0, \ldots, M - 1$, the HT relations (7) and (11) yield $S = (1/M) H S$ and $\tilde{X} = H \tilde{X}$. From Example 1 $S = \tilde{X} = G D^{1/2} X$, we conclude that $\tilde{S} = (1/M) H G D^{1/2} X$, which implies that $T = (1/M) H G D^{1/2} X$. Because $H^{-1} = (1/M) H$ (see (10)) and $G^{-1} = (1/M) G^T$, the inverse relation is $T^{-1} = (1/M^2) H G^{-1/2} D^{-1/2} X$. It is straightforward to verify that $T$ and $T^{-1}$ calculated in this manner, using the numerical values of $G$ and $D$ in Example 1, indeed yield (23).

### III. BICM at Low SNR

#### A. Mutual Information

The mutual information (MI) in bits per channel use between the random vectors $X$ and $Y$ for an arbitrary channel parameter $H$ perfectly known at the receiver is defined as

$$
I(X; Y | H) \triangleq \mathbb{E} \left[ \log_2 \frac{p_{Y|X,H}(Y|X,H)}{p_{Y|H}(Y|H)} \right]
$$

where the expectation is taken over the joint pdf $p_{X,Y,H}$, and $p_{Y|X,H}$ is given by (2).
The MI between $X$ and $Y$ conditioned on the value of the $k$th bit at the input of the modulator is defined as 
\[
I(X;Y|H,C_k) \triangleq E \left[ \log_2 \frac{p_{X,Y|H,C_k}(Y|X,H,C_k)}{p_{Y|H,C_k}(Y|H,C_k)} \right]
\]
where the expectation is taken over the joint pdf $p_{X,Y,H,C_k}$.

**Definition 4 (BICM Generalized Mutual Information):**
The BICM generalized mutual information (BICM-GMI) is defined as \cite{2,17,18,27}
\[
I(\rho) \triangleq \sum_{k=0}^{m-1} I(C_k;Y|H)
= mI(X;Y|H) - \sum_{k=0}^{m-1} I(X;Y|H,C_k)
\tag{24}
\]
where the second line follows by the chain rule. We will analyze the right-hand side of \eqref{24} as a function of $\rho$, for a given pdf $p_H$. According to \eqref{2}, $\rho$ can be varied in two ways, either by varying $N_0$ for a fixed constellation $[X,P]$, or, equivalently, by rescaling the alphabet $X$ linearly for fixed $N_0$ and input distribution $P$.

Martinez et al. \cite{27} recognized the BICM decoder in Fig. 1 as a mismatched decoder and showed that the BICM-GMI in \eqref{24} corresponds to an achievable rate of such a decoder. This means that reliable transmission using a BICM system at rate $R_c$ is possible if $R_c \leq I(\rho)$. Since from \cite{3} $E_b/N_0 = \rho/R_c$, the inequality $R_c \leq I(\rho)$ gives \cite{3}
\[
\frac{E_b}{N_0} \geq \frac{\rho}{I(\rho)}\tag{25}
\]
for any $\rho$. Focusing on the wideband regime, i.e., asymptotically low SNR, we make the following definition.

**Definition 5 (Low-GMI Parameters):** The low-GMI parameters of a constellation $[X,P]$ are defined as $[\mu,E_s,\alpha]$, where
\[
\mu \triangleq E[X],
E_s \triangleq E[\|X\|^2],
\alpha \triangleq \frac{dI(\rho)}{d\rho}\bigg|_{\rho=0}.
\]

In the wideband regime, the average bit energy-to-noise ratio needed for reliable transmission is, using \eqref{25} and the definition of $\alpha$, lower-bounded by
\[
\frac{E_b}{N_0} \geq \lim_{\rho \to 0} \frac{\rho}{I(\rho)} = \frac{1}{\alpha}.
\tag{26}
\]
Furthermore, since in the wideband regime $E_b/N_0 \geq \log_2 2 = -1.59$ dB\cite{15}, $\alpha^{-1} \geq -1.59$ dB.

The first-order behavior of the BICM-GMI in \eqref{24} is fully determined by $\alpha$, which, as we shall see later (e.g., in \eqref{26}), in turn depends on $\mu$ and $E_s$. This is why we designate this triplet as low-GMI parameters. The same definitions can be applied to other MI functions $f(\rho)$ such as the coded modulation MI (CM-MI) \cite{21}. In this paper, however, we are only interested in the BICM-GMI.

The main contributions of this paper are to characterize the low-GMI parameters for arbitrary constellations, including those with nonuniform distributions (Sec. III-C), and to identify the set of constellations for BICM that maximize $\alpha$, i.e., minimize $E_b/N_0$ in the wideband regime (Sec. IV-B).

**B. Low-GMI Parameters for Uniform Distributions**
The low-GMI parameters $[\mu,E_s,\alpha]$ have been analyzed in detail for arbitrary input alphabets $X$ under the assumption of uniform probabilities \cite{21}. Under this assumption, they can be expressed as given by the following theorem.

**Theorem 5:** For a constellation $[X,U_m]$, the low-GMI parameters are
\[
\mu = \frac{1}{M} \sum_{i=0}^{M-1} x_i,\tag{27}
E_s = \frac{1}{M} \sum_{i=0}^{M-1} \|x_i\|^2,\tag{28}
\alpha = \frac{\log_2 e}{M^2 E_s} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} (-1)^{n_i,k} x_{i|k}\|x_{i|k}\|^2.\tag{29}
\]

**Proof:** Expressions \eqref{27} and \eqref{28} follow directly from Definition 5, while \eqref{29} was proved in \cite[eq. (50)]{21}.

The low-GMI parameters can be conveniently expressed as functions of the HT $\mu$ of the alphabet $X$, as shown in the following theorem.

**Theorem 6:** The low-GMI parameters can be expressed as
\[
\mu = \tilde{x}_0,\tag{30}
E_s = \sum_{i=0}^{M-1} \|\tilde{x}_i\|^2,\tag{31}
\alpha = \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \|\tilde{x}_{2k}\|^2.\tag{32}
\]

**Proof:** The expression \eqref{30} is obtained from \cite[eq. (16)]{9}, \cite[eq. (51)]{31} from \cite[eq. (16)]{21}, and \eqref{32} from \cite[Th. 11]{21}.

**C. Low-GMI Parameters for Nonuniform Distributions**
The next theorem is analogous to Theorem 5 but applies to an arbitrary input distribution.

**Theorem 7:** For a constellation $[X,P]$, the low-GMI parameters are
\[
\mu = \sum_{i=0}^{M-1} P_i x_i,\tag{33}
E_s = \sum_{i=0}^{M-1} P_i \|x_i\|^2,\tag{34}
\alpha = \frac{\log_2 e}{E_s} \sum_{i=0}^{M-1} P_i \sum_{j=0}^{M-1} \sum_{k=0}^{m-1} (-1)^{n_i,k+n_j,k} \frac{P_{C_x}(n_{i,k})}{P_{C_x}(n_{j,k})}.\tag{35}
\]
TABLE 1
LOW-GMI PARAMETERS AND FOO CONDITIONS FOR BICM USING UNIFORM AND NONUNIFORM INPUT DISTRIBUTIONS. THE RESULTS FOR \( \{X, U_m\} \)
ARE FROM [21] (cf. THEOREMS 5 AND 2) AND THE ONES FOR \( \{X, P\} \) OR \( \{X, U_m\} \) ARE FROM THEOREMS 7, 8 AND 11.

<table>
<thead>
<tr>
<th>( {X, U_m} )</th>
<th>( {X, P} ) or ( {X, U_m} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( \bar{x}_0 )</td>
</tr>
<tr>
<td>( E_s )</td>
<td>( \sum_{i=0}^{M-1} P_i |x_i|^2 )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>( \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \left( \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \frac{P_i x_i}{\sqrt{P_{C_k(n_{i,k})}}} \right)^2 )</td>
</tr>
<tr>
<td></td>
<td>( \frac{\log_2 e}{E_s} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_i P_j \langle x_i, x_j \rangle )</td>
</tr>
</tbody>
</table>

FOO Condition \( \bar{x}_j = 0, \forall j \notin \{1, 2, 4, \ldots, M/2\} \) \( \mu = 0 \) and \( \bar{x}_j = 0, \forall j \notin \{0 \} \cup \{1, 2, 4, \ldots, M/2\} \).

\textbf{Proof:} Again, (33) and (34) follow from Definition 5 while (35) requires some analysis. It was shown in [21] Th. 10 that

\[
\alpha = \frac{\log_2 e}{2E_s} \sum_{k=0}^{m-1} \left[ \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \frac{P_i x_i}{\sqrt{P_{C_k(n_{i,k})}}} \right]^2 + \left[ \sum_{i=0}^{M-1} \frac{P_i x_i}{\sqrt{P_{C_k(n_{i,k})}}} \right]^2 - 2\|\mu\|^2. \tag{36}
\]

Substituting (33) and writing the squared norms as the inner products of two identical vectors yields

\[
\alpha = \frac{\log_2 e}{2E_s} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_i P_j \langle x_i, x_j \rangle \left[ (-1)^{n_{i,k}} (-1)^{n_{j,k}} \frac{P_i x_i}{\sqrt{P_{C_k(n_{i,k})}}} \right] \left[ \frac{1}{\sqrt{P_{C_k(n_{i,k})}P_{C_k(n_{j,k})}}} - 2 \right].
\]

The expression in brackets can be simplified as

\[
\frac{(-1)^{n_{i,k}+n_{j,k}} + 1}{\sqrt{P_{C_k(n_{i,k})}P_{C_k(n_{j,k})}}} - 2 \begin{cases} 2P_{C_k(n_{i,k})}, & n_{j,k} = n_{i,k} \\ \frac{-1}{2}, & n_{j,k} \neq n_{i,k} \end{cases}
\]

which completes the proof of (35).

Theorem 7 shows that the low-GMI parameters depend on the input alphabet \( X \), the binary labeling (via \( n_{i,k} \) in the expression for \( \alpha \)), and the input distribution (via \( P_{C_k} \)). While the low-GMI parameters of an alphabet \( X \) with uniform probabilities are conveniently expressed in terms of its HT \( \bar{X} \) (cf. Theorem 6), no similar expressions are known for the low-GMI parameters of a general constellation in (33)–(35). This has so far prevented the analytic optimization of such constellations. The new transform introduced in Section 11.2, however, solves this problem by establishing an equivalence between an arbitrary constellation, possibly with nonuniform probabilities, and another constellation with uniform probabilities.

\textbf{Theorem 8:} The low-GMI parameters \( \{\mu, E_s, \alpha\} \) of any constellation \( \{X, P\} \) are equal to the low-GMI parameters of \( \{\bar{X}, U_m\} \).

\textbf{Proof:} Let the low-GMI parameters of \( \{\bar{X}, U_m\} \) be denoted by \( \{\mu', E_s', \alpha'\} \). First, (27) and (12) yield

\[
\mu' = \frac{1}{M} \sum_{i=0}^{M-1} x_i g_{i,j} \sqrt{P_j} = \frac{1}{M} \sum_{j=0}^{M-1} x_j \sqrt{P_j} \sum_{i=0}^{M-1} g_{i,j}. \tag{37}
\]

Applying (16) to the inner sum in (37) reveals that

\[
\mu' = \sum_{j=0}^{M-1} P_j x_j = \mu.
\]

Second, (28) and (12) yield

\[
E_s' = \frac{1}{M} \sum_{i=0}^{M-1} \|x_i\|^2 = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} x_j g_{i,j} \sqrt{P_j} = \frac{1}{M} \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} \langle x_j, x_i \rangle \sqrt{P_j} \sum_{i=0}^{M-1} g_{i,j}. \tag{38}
\]

Evaluating the inner sum using (14) gives

\[
E_s' = \frac{1}{M} \sum_{j=0}^{M-1} P_j \|x_j\|^2 = E_s.
\]

For the third and last part of the theorem, (29) yields

\[
\alpha' = \frac{\log_2 e}{M^2 E_s} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \bar{x}_i \|\bar{x}_i\|^2.
\]

where \( \bar{x}_i \) is given by (12). The inner sum can be expanded as

\[
\sum_{i=0}^{M-1} (-1)^{n_{i,k}} \bar{x}_i = \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \sum_{j=0}^{M-1} x_j g_{i,j} \sqrt{P_j}
\]

\[
= \sum_{j=0}^{M-1} x_j \sqrt{P_j} \sum_{i=0}^{M-1} (-1)^{n_{i,k}} g_{i,j}.
\]

Applying (17) to the inner sum, we obtain

\[
\sum_{i=0}^{M-1} (-1)^{n_{i,k}} \bar{x}_i = \sum_{j=0}^{M-1} x_j P_j (-1)^{n_{j,k}} \sqrt{P_{C_k(n_{j,k})}}.
\]
We take the inner product of this vector with itself and substitute the obtained expression for the squared norm in (38). This yields, after rearranging terms,

\[ \alpha' = \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \langle x_j, x_j \rangle \cdot P_i P_j (\bar{n}_{i,k} + n_{j,k}) \]

\[ = \alpha. \]  

The square root in (39) is \( P_{C_k}(\bar{n}_{i,k})/P_{C_k}(n_{j,k}) \) if \( n_{j,k} = n_{i,k} \) or 1 if \( n_{j,k} = \bar{n}_{i,k} \). In both cases, it can be expressed as \( P_{C_k}(\bar{n}_{i,k})/P_{C_k}(n_{j,k}) \) (or, equivalently, \( P_{C_k}(n_{j,k})/P_{C_k}(\bar{n}_{i,k}) \)). Comparing this result with (35) gives (40).

Theorem 8 shows that the constellation \( [X, F] \) can be mapped to another constellation \( [X, U_m] \) with the same low-GMI parameters, where \( X \) is related to \( X \) via (12) and (18). This relation between \( [X, F] \) and \( [X, U_m] \) will be applied in Sec. IV-B to prove Theorem 10 which is the main result of the paper. To summarize, Table II lists the low-GMI parameters for BICM given by Theorems 6 and 7. The equivalence of the parameters for \( [X, F] \) or \( [X, U_m] \) comes from Theorem 8.

D. Numerical Examples

In this Section, we show examples of how the transform defined in Sec. II-B works and we also present equivalent constellations \( [X, F] \) and \( [X, U_m] \). All results are for the AWGN channel. The (G)MIs are numerically evaluated using Gauss–Hermite quadratures following (28) Sec. III.

Example 3: Consider the equally spaced 16-ary square quadrature amplitude modulation (16-QAM) alphabet \( X = X_{QAM} \) labeled by the binary reflected Gray code (BRGC) \( [X, F] \) with bit probabilities \( b_1 = [0.35, 0.5, 0.35, 0.5] \), shown with black circles in Fig. 3. The input distribution \( F = F_1 \) given by (9) is \( P_0 = P_2 = P_8 = P_{10} = 0.031, P_1 = P_3 = P_4 = P_6 = P_9 = P_{11} = P_{12} = P_{14} = 0.057, \) and \( P_5 = P_7 = P_{13} = P_{15} = 0.106. \) These symbol probabilities are indicated in Fig. 3 where the area of the circle representing \( x_j \) is proportional to the corresponding probability \( P_j \).

Another alphabet \( X_{QAM} \) is obtained by applying the transform in (12) to the constellation \( [X_{QAM}, F_1] \). The white circles in Fig. 3 represent the symbols in \( X_{QAM} \) using a uniform distribution \( U_4 \). The alphabet \( X_{QAM} \) is still a rectangular 16-QAM constellation, but a nonuniformly spaced one. Every row in Fig. 3 can be regarded as a probabilistically shaped (black) or a geometrically shaped (white) 4-PAM constellation; in fact, the same 4-PAM constellations as in Example 1.

The low-GMI parameters of the two constellations \( [X_{QAM}, F_1] \) and \( [X_{QAM}, U_4] \), given by Theorems 7 and 8, are identical, as predicted by Theorem 8. These are

\[ \mu = 0, \quad E_s = 7.60, \quad \alpha = 1.10. \]  

The BICM-GMI for the constellations \( [X_{QAM}, F_1] \) and \( [X_{QAM}, U_4] \) are shown in Fig. 4. In this figure, we also show the capacity of the AWGN channel \( C_AW \) (21) Eq. (22)) and the CM-MI and BICM-GMI for 16-QAM using a uniform input distribution, i.e., \( [X_{QAM}, U_4] \). The results show that the BICM-GMIs of the original constellation \( [X_{QAM}, F_1] \) and the transformed constellation \( [X_{QAM}, U_4] \) are in general different; however, they converge in the low-SNR regime. The endpoints of the BICM-GMI curves for \( [X_{QAM}, F_1] \) and \( [X_{QAM}, U_4] \) are shown with a white circle, whose value follows from (41) and (36). The endpoint for the BICM-GMI curve for the constellation \( [X_{QAM}, U_4] \) is shown with a white square (17) Eq. (18).

Example 4: Consider the NBC-labeled 4-ary phase-shift keying (PSK) alphabet \( X = X_{PSK} \), where \( x_j = [\cos(2\pi j/M + \pi/M), \sin(2\pi j/M + \pi/M)] \) with \( j = 0, \ldots, M - 1 \). The constellation for \( M = 8 \) for two input distributions \( F \) are
have the same low-GMI parameters. The circle areas is proportional to the symbol probabilities $P_j$. Dashed lines join symbols whose labels differ in one bit only.

Fig. 5. 8-PSK constellation $[X_{PSK}, P_2]$ with bit probabilities $b_2$ (black circles) and the transformed constellation $[X_{PSK}, U_3]$ (white circles), which have the same low-GMI parameters. The circle areas is proportional to the symbol probabilities $P_j$. Dashed lines join symbols whose labels differ in one bit only.

Fig. 6. 8-PSK constellation $[X_{PSK}, P_3]$ with bit probabilities $b_3$ (black circles) and its transform $[X_{PSK}, U_3]$ (white circles).

shown in Figs. 5 and 6 where again the circle areas are proportional to the symbol probabilities $P_j$. We denote these input distributions by $P_2$ and $P_3$, which are generated by $b_2 = [0.5, 0.7, 0.9]$ and $b_3 = [0.9, 0.7, 0.3]$, resp. The transforms $X_{PSK}$ are irregular, not resembling a PSK alphabet. Nevertheless, the low-GMI parameters for pairs of constellations $[X_{PSK}, P_2]$ and $[X_{PSK}, U_3]$ are again equal. Particularly, $\alpha = 0.67$ for $[X_{PSK}, P_2]$ and $\alpha = 0.76$ for $[X_{PSK}, P_3]$.

The BICM-GMI for the two 8PSK constellations in Figs. 5 and 6 is shown in Fig. 7. Also shown are the BICM-GMI and CM-MI for 8-PSK with uniform input distributions, for which $\alpha = 0.62$ [21 eq. (56)] and $\alpha = \log_2 e$, resp., and the capacity of the AWGN channel. Again, the results show that the BICM-GMIs of the original constellation $[X_{PSK}, P_2]$ and the transformed constellation $[X_{PSK}, U_3]$ are in general different but converge in the low-SNR regime. The endpoints of the BICM-GMI curves are obtained from (26).

Example 5: Consider the eight-level star-shaped QAM alphabet shown in Fig. 8 (black circles), which we denote by $X_{8QAM}$. This alphabet is used with bit probabilities $b_4 = [0.5, 0.5, 0.85]$, giving an input symbol probability $P_4$. In this figure we also show the transformed constellation $[X_{8QAM}, U_3]$, which according to Theorem 8 has the same low-GMI parameters as $[X_{8QAM}, P_4]$. This can be appreciated in Fig. 9 where the corresponding BICM-GMIs are shown. Fig. 9 also shows how probabilistic shaping improves the BICM-GMI considerably over a wide range of SNRs.

The results in Figs. 7, 8, and 9 also show other interesting properties of probabilistic shaping for BICM. In the high-SNR regime, the use of a nonuniform distribution results in a loss in GMI, i.e., the curves flatten out at a value below $m$ [bit/symbol], but for a wide range of moderately high SNR,
the BICM-GMI is higher with probabilistic shaping. For the 16-QAM alphabet in Example 3, in the medium SNR regime, the AWGN capacity is also shown. The white circle and square show the endpoints \( \alpha \) for \( P = U_m \). For the "if" part, suppose that (43) and (44) hold for a given \( \mathbf{X} \). Second, we prove the "only if" part by showing that if \( [\mathbf{X}, \mathbb{P}] \) is FOO, then (43) and (44) hold.

For the "if" part, suppose that (43) and (44) imply that \( [\mathbf{X}, \mathbb{P}] \) is FOO. Second, we prove the "only if" part by showing that if \( [\mathbf{X}, \mathbb{P}] \) is FOO, then (43) and (44) hold.

The constellation \( [\mathbf{X}, \mathbb{P}] \) is FOO if and only if the HT \( \hat{\mathbf{X}} \) of \( \mathbf{X} \) satisfies both the following conditions:

\[
\hat{x}_0 = \sum_{k=0}^{m-1} x_{2^k} (P_{C_0}(1) - P_{C_0}(0)) \quad (43)
\]

\[
\hat{x}_j = 0, \quad \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}. \quad (44)
\]

**Proof:** We will prove the theorem in two steps. First, we prove the "if" part by showing that (43) and (44) imply that \( [\mathbf{X}, \mathbb{P}] \) is FOO. Second, we prove the "only if" part by showing that if \( [\mathbf{X}, \mathbb{P}] \) is FOO, then (43) and (44) hold.

For the "if" part, suppose that (43) and (44) hold for a given constellation \( [\mathbf{X}, \mathbb{P}] \). Applying (44) in (20) yields for the HT \( \hat{\mathbf{X}} \) of \( \mathbf{X} \):

\[
\hat{s}_j = \psi_j \left[ \hat{x}_0 + \sum_{l=0}^{m-1} \prod_{k=0}^{m-1} \left( P_{C_0}(0) - P_{C_0}(n_{0,k}) \right) + \sum_{l=0}^{m-1} \prod_{k=0}^{m-1} \left( P_{C_0}(0) - P_{C_0}(n_{2^l,k}) \right) \right],
\]

\[
\hat{s}_j = 0, \quad \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}. \quad (45)
\]

Since the bits \( n_{0,k} \) for \( k = 0, \ldots, m-1 \), the first product in (45) is always zero, except when \( j = 0 \). (Recall that a product over \( \emptyset \) in (20) is defined as 1.) Furthermore, because of (4), the second product in (45) is zero whenever \( j \notin \{0, 2^l\} \) for some integer \( l \). We can therefore identify three cases for (45). First,

\[
\hat{s}_j = 0, \quad j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\}. \quad (46)
\]

Second, \( j = 0 \) yields

\[
\hat{s}_0 = \psi_0 \left[ \hat{x}_0 + \sum_{l=0}^{m-1} \prod_{k=0}^{m-1} \left( P_{C_0}(0) - P_{C_0}(n_{2^l,k}) \right) \right] = 0 \quad (47)
\]
because of (43). And third, letting $j = 2^i$ for an integer $i,$
\[\tilde{s}_{2^i} = \psi_{2^i} \sum_{l=0}^{m-1} \tilde{x}_{2^i} \prod_{k=0}^{m-1} (P_{C_k}(0) - P_{C_{n_{2^i,k}}}(0)),\]
i = 0, \ldots, m - 1. \quad (48)

When $l \neq i,$ the product in (48) includes two factors, $k = l$ and $k = i.$ For $k = i,$ $n_{2^i,k} = 0$ and the whole product is 0. Therefore, only $l \neq i$ contributes to the sum in (48). When $l = i,$ the product in (48) is again over $\emptyset$ and (48) becomes
\[\tilde{s}_{2^i} = \psi_{2^i} \tilde{x}_{2^i}.\]
\[\tilde{x}_{2^i} = \begin{cases} \psi_{2^i} \tilde{x}_{2^i}, & j \in \{1, 2, \ldots, 2^{m-1}\}, \\ 0, & \text{otherwise}. \end{cases}\]
\[(49)\]

Combining the three cases (46), (47), and (49) yields
\[\tilde{s}_j = \begin{cases} \psi_j \tilde{x}_j, & j \in \{1, 2, \ldots, 2^{m-1}\}, \\ 0, & \text{otherwise}. \end{cases}\]
\[(50)\]

Using Theorem 9, we conclude that the constellation $[X, \cup_{m}]$ is FOO. Finally, Theorem 8 implies that $[X, P]$ is also FOO, which completes the proof of the “if” part.

For the “only if” part, assume that $[X, P]$ is FOO. By Theorem 8 $[X, \cup_{m}]$ is also FOO, and by Theorem 9 $\tilde{s}_j = 0$ for any $i$ that is not a power of two, where $\tilde{S}$ is the HT of $S = \overline{X}.$ We will now use Theorem 4 to translate the condition on $\tilde{S}$ into conditions on $\overline{X}.$

If $\tilde{s}_i = 0$ for $i \notin \{1, 2, \ldots, 2^{m-1}\},$ then the summation over $i = 0, \ldots, M - 1$ in (24) can be reduced to a summation over $i = 2^l$ for $l = 0, \ldots, m - 1,$
\[\tilde{x}_j = \sum_{l=0}^{m-1} \frac{\tilde{s}_{2^l}}{\psi_{2^l}} \prod_{k=0}^{m-1} (P_{C_k}(n_{2^l,k}) - P_{C_k}(0)),\]
j = 0, \ldots, M - 1. \quad (51)

Due to (43), the product in (50) is nonzero only if the product is over $k \in \emptyset$ or over the single-element set $k \in \{l\},$ i.e., if $j = 0$ or $j = 2^l$ for some integer $l,$ resp. Again, three cases can be identified. First, if $j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\},$ then the product in (50) includes at least one $k \neq l$ for every $l$ and
\[\tilde{x}_j = 0, \quad j \notin \{0\} \cup \{1, 2, \ldots, 2^{m-1}\}. \quad (52)\]

Second, for $j = 0,$ the product comprises only one factor, $k = l,$ and
\[\tilde{x}_0 = \sum_{l=0}^{m-1} \frac{\tilde{s}_{2^l}}{\psi_{2^l}} (P_{C_l}(1) - P_{C_l}(0)). \quad (53)\]

And third, setting $j = 2^l,$
\[\tilde{x}_{2^l} = \sum_{l=0}^{m-1} \frac{\tilde{s}_{2^l}}{\psi_{2^l}} \prod_{k=0}^{m-1} (P_{C_k}(n_{2^l,k}) - P_{C_k}(0)),\]
i = 0, \ldots, m - 1. \quad (54)

As explained after (43), the product in (53) is 1 if $l = i$ (product over $\emptyset$) and 0 otherwise. Thus, the summation in (53) can be reduced to just one term, $l = i,$ and
\[\tilde{x}_{2^i} = \frac{\tilde{s}_{2^i}}{\psi_{2^i}}, \quad i = 0, \ldots, m - 1. \quad (55)\]

Combining the three cases (51), (52), and (54) yields
\[\tilde{x}_j = \begin{cases} \tilde{s}_{2^l} \frac{P_{C_l}(1) - P_{C_l}(0)}{\psi_{2^l}}, & j = 0, \\ \tilde{s}_{2^i}, & j \in \{1, 2, \ldots, 2^{m-1}\}, \\ 0, & \text{otherwise}. \end{cases}\]
\[(56)\]

which satisfies (43) and (44). This completes the proof of the “only if” part.

\[\text{Remark 3: Only (43) depends on the input distribution, not (44). In view of Theorem 9 the only difference between FOO constellations with uniform and nonuniform distributions lies in } \tilde{x}_0. \text{ The final theorem gives this statement a more intuitive interpretation.}\]

Theorem 11: The constellation $[X, \mathbb{P}]$ is FOO if and only if both the following conditions hold:
\[\mu = 0, \quad (57)\]
\[\tilde{x}_j = 0, \quad \forall j \notin \{0\} \cup \{1, 2, 4, \ldots, M/2\}. \quad (58)\]

Proof: We wish to prove that if (43) (or equivalently (56)) holds, then (53) and (54) are equivalent.

For any constellation $[X, \mathbb{P},]$ the mean $\mu = \tilde{s}_0,$ where $\tilde{S}$ is the HT of $S = \overline{X}.$ This follows from Theorem 8 and (30). The mean $\mu$ can be calculated by letting $i = 0$ in (20) and (22) as
\[\mu = \psi_0 \sum_{i=0}^{M-1} \tilde{x}_i \prod_{k=0}^{m-1} (P_{C_k}(0) - P_{C_k}(1)) \quad (59)\]
where $\psi_0 = 1.$

In this theorem, we are only interested in constellations that satisfy (43) or equivalently (56). For such constellations, the sum in (57) includes at most $m + 1$ nonzero terms, namely,
\[\mu = \tilde{x}_0 + \sum_{l=0}^{M-1} \tilde{x}_{2^l} (P_{C_l}(0) - P_{C_l}(1)) \quad (60)\]

This expression makes (43) and (56) equivalent. \hfill \Box

Based on Theorem 9 the result in Theorem 11 can be understood as follows. If a constellation with a uniform input distribution is FOO, it will still be FOO for any other input distribution $b$ provided that the input alphabet is translated to be zero mean. In view of the geometrical interpretation of Theorem 9 given in [21] Th. 12], the result in Theorem 11 also states that a constellation is FOO if and only if its input alphabet is a linear projection of a hypercube and it has zero mean.

We also note that the zero-mean condition in Theorem 11 is the same that guarantees FOO for the CM-MI [21] Footnote 13. This implies that the only difference between FOO constellations for the CM-MI and the BICM-GMI lies on the extra constraint on the input alphabet to be a linear projection of a hypercube.

C. Numerical Examples

In this subsection we give numerical examples to illustrate the analytical results presented in this paper.

3The parameter $\alpha$ for the CM-MI is [21] Th. 7 $\alpha = \log_2 e (1 - \|\mu\|^2/E_b).$
Example 6: Consider the so-called 8-AMPM alphabet
\[ X_{8AMPM} = \begin{bmatrix} -1 & 1 & -3 & -1 & 1 & 3 & -1 & 1 \\ 0 & -2 & 2 & 0 & 2 & 0 & 2 & 0 \end{bmatrix}^T \] (58)
which corresponds to a projected hypercube. The constellation \( X'_{8AMPM} \) was shown to be FOO for \( P = U_3 \) in [21, Example 4]. In view of Theorem 11, it is FOO for any \( P \) if it has zero mean. Using (58) and (6) in (33), we find (after some algebra) that
\[ \mu = \left[ 1 + 2 \left( P_{C_1}(0) - P_{C_4}(0) - P_{C_2}(0) \right) \right] \] (59)
For example, the bit probabilities \( b_5 = [0.40, 0.55, 0.60] \) give an input distribution \( P_5 \) for which the mean (59) is \( \mu = [0.10, -0.40] \). We define another alphabet \( X''_{8AMPM} \) by subtracting \( \mu \) from each element in \( X_{8AMPM} \). Fig. 10 shows the translated constellation \( X'_{8AMPM}, P_5 \) along with \( X''_{8AMPM}, U_3 \), where \( X''_{8AMPM} \) is the transform of \( X'_{8AMPM} \) for the distribution \( P_5 \). They are both zero-mean projected hypercubes and thus FOO according to Theorem 11. This can be observed from the BICM-GMI curve for \( X'_{8AMPM}, P_5 \) in Fig. 12.

The exemplified method holds in full generality: Any alphabet that is FOO with a uniform input distribution is FOO also with an arbitrary nonuniform distribution, if it is translated to zero mean. Furthermore, all nonuniform FOO constellations can be constructed in this manner.

For certain distributions, the mean (59) is zero without translation. Specifically, \( \mu = 0 \) if and only if
\[ P_{C_6}(0) = P_{C_5}(0) = P_{C_1}(0)/2 + 1/4. \] (60)
Clearly, the uniform case \( P_{C_6}(0) = P_{C_5}(0) = P_{C_1}(0) = 1/2 \) analyzed in [21] fulfills (60). More interestingly, when any other vector of bit probabilities fulfilling (60) is used, the resulting constellation will be FOO. This is the case for instance with \( b_5 = [0.70, 0.90, 0.70] \). The obtained constellation, denoted by \( X'_{8AMPM}, P_6 \) is illustrated in Fig. 11 along with its transform \( X'_{8AMPM}, U_3 \). Graphically, both alphabets look like cubes, although viewed from different angles, which is precisely what Theorems 9 and 11 predict.

In Fig. 12, we show the BICM-GMI for the zero-mean constellations \( X''_{8AMPM}, P_5 \) and \( X''_{8AMPM}, P_6 \) as well as for the constellations \( X'_{8AMPM}, U_3 \) and \( X'_{8AMPM}, U_3 \). As expected, all the GMIs converge at the SL for low SNR. For these two cases, the transformed alphabets with uniform input distributions give larger GMI for all SNRs compared to the corresponding nonuniform ones. This is however not always the case, cf. Fig. 7 with \( P_2 \).

Example 7: M-PAM alphabets have been shown to be FOO if the NBC is used with a uniform input distribution, i.e, the constellation \( X'_{PAM}, U_m \) with \( X'_{PAM} = [-(M-1), -(M-3), \ldots, M-1] \) is FOO [13] [21, Th. 14]. In this example, we study the first-order behavior of the 8-PAM alphabet \( X'_{PAM} = [-7, 7, -1, 1, -5, 5, -3, 3] \), where the order of the points represents the BRGC. The constellation \( X'_{PAM}, P \) is known not to be FOO for \( U_3 [17, Th. 3] \).

The BICM-GMI of \( X'_{PAM} \) is shown in Fig. 13 for the set of bit probabilities \( b_7 = [0.5, p, p] \) for different values of \( p \). For \( p = 0.5 \), the uniform distribution is obtained. As \( p \) decreases, the Gray-labeled constellation approaches a zero-mean binary alphabet, which is FOO [13, Fig. 2] [21].

Fig. 10. The FOO constellations \( X'_{8AMPM}, P_5 \) (black circles) and \( X'_{8AMPM}, U_3 \) (white circles).

Fig. 11. The constellations \( X_{8AMPM}, P_6 \) (black circles) and \( X_{8AMPM}, U_3 \) (white circles). They both have zero mean and look like cubes, which indicates that they are FOO.

Fig. 12. BICM-GMI for the four FOO constellations in Figs. 10 and 11. The SL is shown with a white circle.
BRGC to the NBC, but this also comes with a heavy penalty. Alternatively, the SL can be attained by switching from the high-SNR regimes: The SL can be approached by decreasing the input distribution, such that the two systems have the same theoretic shaped constellation and a constellation with uniform distribution, is considerably weaker than for a wide range of SNRs.

Fig. 3 (b)]. Fig. 13 illustrates the tradeoff between the low- and high-SNR regimes: The SL can be approached by decreasing $p$, but this causes a decrease in GMI in the high-SNR regime. Alternatively, the SL can be attained by switching from the BRGC to the NBC, but this also comes with a heavy penalty at higher SNRs.

V. CONCLUSIONS

There exists a closed-form mapping between any probabilistically shaped constellation and a constellation with uniform input distribution, such that the two systems have the same low-SNR first-order behavior (Definition 3 and Theorem 8). Thus, the combination of probabilistic and geometric shaping is equivalent to pure geometric shaping at low SNR.

We are particularly interested in BICM systems that attain the SL $-1.59$ dB at asymptotically low SNR, i.e., in FOO constellations. Somewhat disappointingly, the set of probabilistically shaped FOO constellation is no larger than the set of FOO constellations with uniform distributions, disregarding translations of the whole input alphabet. Both sets can be fully characterized as the set of linear projections of a hypercube, translated to have zero mean for the considered input distribution (Theorems 9 and 10; cf. Figs. 10 and 11). Although non-FOO constellations for BICM can be improved by probabilistic shaping (Fig. 13), it is impossible to make them FOO except in degenerate cases (by setting some probabilities to zero).

APPENDIX

PROPERTIES OF THE TRANSFORM

In this Appendix, some theoretical properties of the new transform defined in Section II-E are proved. In the first section, the “sum-product lemma” is established, which will be used extensively throughout the appendix. In the following two sections, Lemma 1 and Theorem 3 are proved.

A. The Sum-Product Lemma

Many properties of NBC-labeled constellations and their transforms can be expressed as the sum of products of certain functions, where each function depends on one bit position. Such expressions, which occur frequently in the two subsequent sections, can be resolved using this general lemma.

**Lemma 12**: Let $f_{k,i}$ for $k = 0, \ldots, m - 1$ and $u \in B$ be any real numbers. Then

$$
\sum_{i=0}^{M-1} \prod_{k=0}^{m-1} f_{k,n_{i,k}} = \prod_{k=0}^{m-1} (f_{k,0} + f_{k,1}).
$$

**Proof**: A summation over $i = 0, \ldots, 2^m - 1$ is equivalent to $m$ sums over $i_k \in B$, where $k = 0, \ldots, m - 1$ and $i = i_0 + 2i_1 + \cdots + 2^{m-1}i_{m-1}$. With this notation, $n_{i,k} = i_k$ and

$$
\sum_{i=0}^{m-1} \prod_{k=0}^{m-1} f_{k,n_{i,k}} = \sum_{i_0 = B} \sum_{i_1 = B} \cdots \sum_{i_{m-1} = B} \prod_{k=0}^{m-1} f_{k,i_k}
$$

Following (13) yields

$$
\sum_{i=0}^{M-1} \prod_{k=0}^{m-1} f_{k,n_{i,k}} = \sum_{i_0 \in B} \sum_{i_1 \in B} \cdots \sum_{i_{m-1} \in B} f_{0,i_0} f_{1,i_1} \cdots f_{m-1,i_{m-1}}
$$

$$
\sum_{i_0 \in B} \sum_{i_1 \in B} \cdots \sum_{i_{m-1} \in B} f_{0,i_0} f_{1,i_1} \cdots f_{m-1,i_{m-1}} = \sum_{k=0}^{m-1} f_{k,0} \sum_{i_k \in B} f_{1,i_k} \cdots \sum_{i_{m-1} \in B} f_{m-1,i_{m-1}}
$$

$$
= \prod_{k=0}^{m-1} \sum_{i_k \in B} f_{k,i_k}.
$$

B. The Transform Coefficients

Lemma which lists two fundamental properties of the transform coefficients $g_{i,j}$, was given in Section II-E.

**Proof of Lemma 7**: The two parts of Lemma 11 will now be proved separately. First, the definition of $g_{i,j}$ in (13) yields

$$
\sum_{i=0}^{M-1} g_{i,j} g_{i,j} = \prod_{i=0}^{M-1} \sum_{k=0}^{m-1} (-1)^{n_{i,k} n_{i,k}} \sqrt{P_{C_k}(0)}
$$

$$
+ (-1)^{n_{i,k} n_{i,k}} \sqrt{P_{C_k}(1)}
$$

$$
\cdot \left[ (-1)^{n_{i,k} n_{i,k}} \sqrt{P_{C_k}(0)}
+ (-1)^{n_{i,k} n_{i,k}} \sqrt{P_{C_k}(1)} \right]
$$

$$
= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} \left[ (-1)^{n_{i,k} n_{i,k} + n_{i,k} n_{i,k}} P_{C_k}(0)
+ (-1)^{n_{i,k} n_{i,k} + n_{i,k} n_{i,k}} P_{C_k}(1) \right]
$$

$$
+ (-1)^{n_{i,k} n_{i,k} + n_{i,k} n_{i,k}} \sqrt{P_{C_k}(0)} P_{C_k}(1)
+ (-1)^{n_{i,k} n_{i,k} + n_{i,k} n_{i,k}} P_{C_k}(1)\right]
$$

$$
= \sum_{i=0}^{M-1} \prod_{k=0}^{m-1} (-1)^{n_{i,k} n_{i,k} + n_{i,k} n_{i,k}}
$$

$$
\cdot \left[ (-1)^{n_{i,k} n_{i,k}} P_{C_k}(0) + P_{C_k}(1)
+ (-1)^{n_{i,k}} \left[ (-1)^{n_{i,k}} + (-1)^{n_{i,k}} \right] \sqrt{P_{C_k}(0)} P_{C_k}(1) \right]
$$
where the last equality follows by repeatedly using the identities \( \bar{u} = 1 - u \) and \((-1)^u = (-1)^{-u} \) for \( u \in \mathcal{B} \). Lemma 12 now yields

\[
\sum_{i=0}^{M-1} g_{i,l} g_{i,j} = \prod_{k=0}^{m-1} \left[ (-1)^{n_{i,j}+n_{l,k}} P_{C_k}(0) + P_{C_k}(1) \right.
\]

\[
+ \left. \left[ (-1)^{n_{i,j}+n_{l,k}} \right] \sqrt{P_{C_k}(0)P_{C_k}(1)} \right]
\]

\[
= \prod_{k=0}^{m-1} \left[ \left( (-1)^{n_{i,j}+n_{l,k}} + 1 \right) P_{C_k}(0) \right.
\]

\[
+ \left. \left( 1 + (-1)^{n_{i,j}+n_{l,k}} \right) P_{C_k}(1) \right]
\]

\[
= \prod_{k=0}^{m-1} \left( 1 + (-1)^{n_{i,j}+n_{l,k}} \right)
\]

where the last step follows because \( P_{C_k}(0) + P_{C_k}(1) = 1 \). The factors in \( 61 \) are either 2 or 0, depending on whether \( n_{i,j} = n_{l,k} \) or \( n_{i,j} \neq n_{l,k} \) for the particular bit position \( k \). Thus, \( \sum_{k=0}^{m-1} g_{i,l} g_{i,j} \) is either \( 2^m = M \) or 0, depending on whether \( l \) and \( j \) have all bits equal or not. This completes the proof of (44).

To prove the second part of Lemma 11 which is (15), we observe from (8) and (13) that

\[
\sum_{i=0}^{M-1} h_{i,l} g_{i,j} = \prod_{k=0}^{m-1} \left[ (-1)^{n_{i,j}+n_{l,k}} \sqrt{P_{C_k}(0)} \right]
\]

\[
= \sum_{k=0}^{M-1} \prod_{i=0}^{m-1} \phi_{k,n_{i,j}}
\]

where

\[
\phi_{k,u} = (-1)^{u_{n_{i,j}+n_{l,k}}} \sqrt{P_{C_k}(0)} + (-1)^{u_{n_{i,j}+n_{l,k}}} \sqrt{P_{C_k}(1)}
\]

Intending to apply Lemma 12 to (62), we first calculate the quantity

\[
\phi_{k,0} + \phi_{k,1} = (-1)^{n_{l,k}} \sqrt{P_{C_k}(0)} + \sqrt{P_{C_k}(1)}
\]

\[
+ (-1)^{n_{l,k}} \sqrt{P_{C_k}(0)} + (-1)^{n_{l,k}+n_{l,k}} \sqrt{P_{C_k}(1)}
\]

for \( k = 0, \ldots, m - 1 \). For reasons that will soon become clear, we extract a common factor \((-1)^{n_{l,k}}\) from all terms, obtaining

\[
\phi_{k,0} + \phi_{k,1} = (-1)^{n_{l,k}} \left[ (((-1)^{n_{l,k}+n_{l,k}}) \sqrt{P_{C_k}(0)} \right]
\]

\[
+ (((-1)^{n_{l,k}+n_{l,k}}) \sqrt{P_{C_k}(1)} \right]
\]

\[
(63)
\]

The coefficient in front of \( \sqrt{P_{C_k}(0)} \) is 2 if \( n_{j,k} = n_{l,k} \) and 0 otherwise. Similarly, the coefficient in front of \( \sqrt{P_{C_k}(1)} \) is 0 if \( n_{j,k} = n_{l,k} \) and 2 otherwise. Thus, for every \( k \), the expression depends on either \( P_{C_k}(0) \) or \( P_{C_k}(1) \) but not both, namely,

\[
\phi_{k,0} + \phi_{k,1} = \begin{cases} 
(-1)^{n_{l,k}+n_{l,k}} \cdot 2 \sqrt{P_{C_k}(0)}, & n_{j,k} = n_{l,k} \\
(-1)^{n_{l,k}+n_{l,k}} \cdot 2 \sqrt{P_{C_k}(1)}, & n_{j,k} \neq n_{l,k}
\end{cases}
\]

\[
= 2(-1)^{n_{l,k}} \sqrt{P_{C_k}(n_{j,k} \oplus n_{l,k})}
\]

\[
(63)
\]

according to (5).

Using (63), we are now ready to apply Lemma 12 to (62), which yields

\[
\sum_{i=0}^{M-1} h_{i,l} g_{i,j} = \prod_{k=0}^{m-1} \left[ (-1)^{n_{l,k}+n_{l,k}} \sqrt{P_{C_k}(n_{j,k} \oplus n_{l,k})} \right]
\]

\[
\sum_{k=0}^{m-1} \phi_{k,n_{i,j}}
\]

\[
= \sum_{k=0}^{M-1} \prod_{i=0}^{m-1} \phi_{k,n_{i,j}}
\]

\[
(62)
\]

\[
\sum_{k=0}^{M-1} \prod_{i=0}^{m-1} \phi_{k,n_{i,j}}
\]

\[
(64)
\]

C. Two Consecutive Transforms

Theorem 4 which characterizes the vectors obtained by applying the new transform and the HT sequentially to a given alphabet, was given in Section 11.

Proof of Theorem 4 To prove (20), we first write \( \tilde{S} \) as a function of \( \tilde{X} \) using (7) and (12), as illustrated in Fig. 2. For \( i = 0, \ldots, M - 1 \),

\[
\tilde{s}_i = \frac{1}{M} \sum_{j=0}^{M-1} s_j h_{i,j}
\]

\[
= \frac{1}{M} \sum_{j=0}^{M-1} h_{i,j} \sum_{l=0}^{M-1} x_l g_{j,l} \sqrt{P_l}
\]

\[
= \frac{1}{M} \sum_{j=0}^{M-1} \sum_{l=0}^{M-1} x_l \sqrt{P_l} h_{i,j} g_{j,l}
\]

\[
= \sum_{l=0}^{M-1} \sum_{j=0}^{M-1} x_l h_{i,l} \sqrt{P_l} \tilde{P}_{i \oplus l}
\]

\[
(53)
\]

\[
(53)
\]

using (15) for the last equality. To obtain \( \tilde{S} \) as a function of \( \tilde{X} \), we express \( x_l \) in terms of its inverse HT in (11), to obtain

\[
\tilde{s}_i = \sum_{l=0}^{M-1} \tilde{x}_l h_{i,l} \sqrt{P_l} \tilde{P}_{i \oplus l}
\]

\[
= \sum_{j=0}^{M-1} \tilde{x}_j h_{i,j}
\]

\[
(64)
\]

where

\[
\eta_{i,j} \triangleq \sum_{l=0}^{M-1} h_{i,l} \tilde{h}_{j,l} \sqrt{P_l} \tilde{P}_{i \oplus l}
\]
The coefficient $\eta_{i,j}$ can be expressed using (68) and (5) as

$$
\eta_{i,j} = \sum_{l=0}^{M-1} \prod_{k=0}^{m-1} (-1)^{n_{i,k} + n_{j,k}} \cdot \sqrt{P_{C_k}(n_{i,k})P_{C_k}(n_{i,k} \oplus n_{l,k})}
$$

$$
\prod_{k=0}^{m-1} \left[ \sqrt{P_{C_k}(0)P_{C_k}(n_{i,k})} + (-1)^{n_{i,k} + n_{j,k}} \sqrt{P_{C_k}(1)P_{C_k}(\tilde{n}_{i,k})} \right]
$$

(65)

where to pass from (65) to (66) we used Lemma 12 with $f_{k,u} = (-1)^{u(n_{i,k} + n_{i,k})} \sqrt{P_{C_k}(u)P_{C_k}(n_{i,k} \oplus u)}$. Depending on the value of $n_{i,k}$, this product can be partitioned into two subproducts

$$
\eta_{i,j} = \prod_{k=0}^{m-1} \left[ P_{C_k}(0) + (-1)^{n_{i,k}} P_{C_k}(1) \right] \cdot \prod_{k=0}^{m-1} \left[ \sqrt{P_{C_k}(0)P_{C_k}(1) + (-1)^{n_{i,k}} P_{C_k}(0)} \right]
$$

(66)

making use of $\psi_i$ defined in (22). Since the two factors in (67) depend on $n_{i,k}$ as

$$
P_{C_k}(0) + (-1)^{n_{i,k}} P_{C_k}(1) = \begin{cases} 1, & n_{i,k} = 0, \\ P_{C_k}(0) - P_{C_k}(1), & n_{i,k} = 1, \\ 0, & n_{j,k} = 0, \\ 1, & n_{j,k} = 1, \end{cases}
$$

$$
1 + (-1)^{n_{i,k}} \frac{1}{2}
$$

they can be expanded into four factors as

$$
\eta_{i,j} = \prod_{k=0}^{m-1} \left[ P_{C_k}(0) - P_{C_k}(1) \right] \cdot \prod_{k=0}^{m-1} \left[ (P_{C_k}(0) - P_{C_k}(1)) \right]
$$

(68)

The two products of ones can be omitted. The product of zeros may look strange but is perfectly legitimate; its value is by definition 1 if the set $\{k : [n_{i,k}, n_{i,k}] = [1, 0]\}$ is empty and 0 otherwise. Furthermore, since $P_{C_k}(0) - P_{C_k}(n_{j,k}) = 0$ for all $k$ in this set, the second and third factors of (68) can be merged into

$$
\eta_{i,j} = \prod_{k=0}^{m-1} \left[ P_{C_k}(0) - P_{C_k}(n_{j,k}) \right]
$$

(69)

which together with (64) completes the proof of (20).

To prove the reverse relationship (21), we first multiply both sides of (64) with $h_{i,M-1}\eta_{i,l}/\psi_i\psi_l$ and sum over $i$:

$$
\prod_{i=0}^{M-1} \frac{h_{i,M-1}\eta_{i,l}}{\psi_i\psi_l} = \prod_{j=0}^{M-1} \frac{\eta_{i,l}}{\psi_i\psi_l}
$$

(70)

The Hadamard coefficient $h_{i,M-1}$ can be factorized using (8) with $n_{M-1,k} = 1, \forall k$ as

$$
h_{i,M-1} = \prod_{k=0}^{m-1} (-1)^{n_{j,k}}.
$$

(71)

Using (71) in (70) and replacing the ratios $\eta_{i,j}/\psi_i$ and $\eta_{i,l}/\psi_l$ with their factorizations according to (69), we obtain for the inner sum on the right-hand side of (70)

$$
\prod_{i=0}^{M-1} \frac{h_{i,M-1}\eta_{i,j}}{\psi_i\psi_l} = \prod_{i=0}^{M-1} \frac{\eta_{i,j}}{\psi_i}\prod_{k=0}^{m-1} \phi_{k,n_{i,k}}
$$

(72)

where

$$
\phi_{k,u} = \begin{cases} (1)^u, & u = n_{j,k} = n_{i,k}, \\ (1)^u(P_{C_k}(0) - P_{C_k}(u)), & u = n_{j,k} \neq n_{i,k}, \\ (1)^u(P_{C_k}(0) - P_{C_k}(n_{j,k}))(P_{C_k}(0) - P_{C_k}(u)), & u \neq n_{j,k} = n_{i,k}. \end{cases}
$$

(73)

The fourth case is always 0, because either $P_{C_k}(0) - P_{C_k}(n_{j,k})$ or $P_{C_k}(0) - P_{C_k}(u)$ is 0. Furthermore, the second and third cases can be combined into

$$
\phi_{k,u} = \begin{cases} (1)^u, & u = n_{j,k} = n_{i,k}, \\ (1)^u(P_{C_k}(0) - P_{C_k}(n_{j,k})), & n_{j,k} \neq n_{i,k}, \\ 0, & u \neq n_{j,k} = n_{i,k}. \end{cases}
$$

(74)

We wish to apply Lemma 12 to the right-hand side of (72). In order to do so, we first calculate from (73)

$$
\phi_{k,0} + \phi_{k,1} = \begin{cases} 1 + 0, & n_{j,k} = n_{i,k} = 0, \\ (P_{C_k}(0) - P_{C_k}(n_{j,k}))(P_{C_k}(0) - P_{C_k}(n_{j,k})), & n_{j,k} \neq n_{i,k}, \\ 0 - 1, & n_{j,k} = n_{i,k} = 1. \end{cases}
$$

(75)

Now by Lemma 12 (72) can be expressed as

$$
\prod_{i=0}^{M-1} \frac{h_{i,M-1}\eta_{i,j}}{\psi_i\psi_l} = \prod_{k=0}^{m-1} (\phi_{k,0} + \phi_{k,1}).
$$

By (74), this product will be nonzero only if $j$ and $l$ match in all bit positions $k = 0, \ldots, m - 1$, i.e., if $j = l$. Thus, again utilizing (71),

$$
\prod_{i=0}^{M-1} \frac{h_{i,M-1}\eta_{i,j}}{\psi_i\psi_l} = \begin{cases} \prod_{k=0}^{m-1} (-1)^{n_{i,k}}, & j = l, \\ 0, & j \neq l. \end{cases}
$$

(75)
Finally, we combine (70) and (75) into
\[
\sum_{i=0}^{M-1} \frac{h_{i,M-1} \psi_{l,i}}{\psi_l} = h_{l,M-1} \tilde{x}_l. \tag{76}
\]
Dividing both sides by \(h_{l,M-1}\) yields on the left-hand side the coefficient \(h_{l,M-1} \psi_{l,i}/(h_{l,M-1} \psi_l)\), which can be expressed using (8) and (69) as
\[
\frac{h_{l,M-1} \psi_{l,i}}{h_{l,M-1} \psi_l} = \prod_{k=0}^{m-1} (1) \prod_{k=0}^{m-1} (P_{C_k}(0) - P_{C_k}(n_{i,k}))
\]
This expression, substituted into (70), completes the proof of (21).

REFERENCES