Binary-Tree Encoding for Uniform Binary Sources in Index Modulation Systems

Coon, J., Badiu, M-A., Liu, Y., Yarkin, F. & Dang, S.

Published PDF deposited in Coventry University's Repository

Original citation:
Coon, J, Badiu, M-A, Liu, Y, Yarkin, F & Dang, S 2019, 'Binary-Tree Encoding for Uniform Binary Sources in Index Modulation Systems' IEEE Journal of Selected Topics in Signal Processing, vol. (In-press), pp. (In-press).
https://dx.doi.org/10.1109/JSTSP.2019.2914531

DOI 10.1109/JSTSP.2019.2914531
ISSN 1932-4553
ESSN 1941-0484

Publisher: IEEE

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

Binary-Tree Encoding for Uniform Binary Sources in Index Modulation Systems

Justin P. Coon, *Senior Member, IEEE*, Mihai-Alin Badiu, Ye Liu, Ferhat Yarkin, *Student Member, IEEE*, Shuping Dang, *Member, IEEE*

***Abstract*—The problem of designing bit-to-pattern mappings and power allocation schemes for orthogonal frequency-division multiplexing (OFDM) systems that employ subcarrier index modulation (IM) is considered. We assume the binary source conveys a stream of independent, uniformly distributed bits to the pattern mapper, which introduces a constraint on the pattern transmission probability distribution that can be quantified using a binary tree formalism. Under this constraint, we undertake the task of maximizing the achievable rate subject to the availability of channel knowledge at the transmitter. The optimization variables are the pattern probability distribution (i.e., the bit-to-pattern mapping) and the transmit powers allocated to active subcarriers. To solve the problem, we first consider the relaxed problem where pattern probabilities are allowed to take any values in the interval $[0, 1]$ subject to a sum probability constraint. We develop (approximately) optimal solutions to the relaxed problem by using new bounds and asymptotic results, and then use a novel heuristic algorithm to project the relaxed solution onto a point in the feasible set of the constrained problem. Numerical analysis shows that this approach is capable of achieving the maximum mutual information for the relaxed problem in low and high-SNR regimes and offers noticeable benefits in terms of achievable rate relative to a conventional OFDM-IM benchmark.**

***Index Terms*—OFDM, index modulation, binary tree, mutual information, achievable rate.**

I. Introduction

As a subclass of permutation modulation [1], index modulation (IM) has recently attracted significant interest [2], [3] due to its feature of "achieving more by doing less". The central idea of IM lies in the observation that, in addition to encoding information in a signal, one can encode information in the order in which a signal is conveyed in a given domain. The idea of encoding information using permutations or combinations has been applied in several contexts. For example, by using different transmit antennas and channel uniqueness, permutation modulation has been employed in the spatial domain in the form of so-called spatial modulation [4], [5]. Similar ideas have been applied to the medium/channel domain by manipulating the radiation patterns of antennas [6], [7]. Permutation modulation has also been used in the subcarrier index domain in multicarrier systems, such as orthogonal frequency-division multiplexing (OFDM). This approach is commonly referred to as subcarrier-IM or simply IM [8], [9]. Finally, the use of permutation methods in conjunction with different modes in orbital angular momentum transmissions has been studied [10], [11].

To facilitate the use of combinatorial patterns for encoding, a codebook for the mapping between patterns and the source messages (bit sequences) must be specified. Many existing works that study permutation modulation in digital communication systems assume that the number of possible patterns is a power of two [4], [12], [13]. However, such an assumption limits the applicability of permutation modulation, e.g., conventional spatial modulation (with a single active antenna in each transmission period) is only applicable when the number of antennas at the transmitter is a power of two.

Another typical approach that has been studied is to assume that only a subset of all possible patterns contains valid patterns, and the size of the subset is a power of two [9], [14]–[16]. However, this approach is not able to utilize the full potential of permutation modulation in terms of data rate, because a certain number of possible permutations that could have been used to carry information are neglected [17]. The study detailed in [18] considers the possibility of using all permutation patterns with uniform probability, but no treatment of how to realize the uniform probability distribution in digital communication systems is given in that work.

To address these issues related to the mapping of source bit sequences to permutation patterns, a few recent contributions have focused on the adaptation of binary Huffman coding [19] for permutation/index codebook design [11], [17], [20]–[23]. Here, a bijective mapping between information bit sequences and the permutation/index patterns is constructed with the aid of a full binary tree; patterns are associated with leaves in the tree, and corresponding bit sequences are defined according to a labeling rule (used in the Huffman algorithm) pertaining to the respective paths from each leaf to the root. Importantly, in contrast to conventional application scenarios for source compression where the source symbol distribution is known a priori, the probability distribution of the patterns observed during transmission in permutation modulation systems *is dependent upon the binary source* [17]. In this sense, the Huffman mapping is applied in permutation modulation schemes in a reversed manner. We adopt the term *binary-tree encoding* rather than *Huffman coding* for the bit-to-pattern mapping operation for the remainder of this paper in order to highlight this subtle, but important difference.

J. P. Coon, M.-A. Badiu, Y. Liu, and F. Yarkin are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK, (Email: justin.coon@eng.ox.ac.uk).

M.-A. Badiu is also with the Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, Aalborg, 9220, Denmark.

S. Dang is with Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia.

This work was supported by EPSRC grant numbers EP/R511742/1 and EP/N002350/1. The authors also wish to acknowledge the support of the Telecommunications Research Laboratory of Toshiba Research Europe Ltd.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

Binary-tree encoding for permutation modulation schemes enables one to choose the probability distribution of the permutation patterns to achieve certain design criteria, e.g., achievable rate maximization [11], [17], [21], [23] or symbol-error rate (SER) minimization [21], [23], [24]. However, existing works along this direction fall short in a number of ways. For example, the support of the (random) patterns, when constrained by full binary tree structures, is discrete. As a result, optimization problems for maximizing achievable rates or minimizing SERs are of mixed-integer forms, and an exhaustive search over all admissible probability distributions may be required to find the global optimum. However, the number of admissible distributions has not been characterized in the literature, and thus the complexity of exhaustive searching is not well understood. A common way to reduce optimization complexity that has been treated in the literature is to relax the full-binary-tree constraint on the pattern probability distribution [11], [21]. However, the problem of how to project the relaxed probability distribution to a feasible distribution that satisfies the full-binary-tree constraint remains open. An alternative strategy that has received attention recently has been to focus on high and low signal-to-noise ratio (SNR) regimes. For the limited case of single-active-antenna spatial modulation, analytic forms of the asymptotically optimal probability distributions for the permutation patterns were reported in [21]. A generalization that activates multiple resources per channel use, which is the scenario of interest for multicarrier communication systems such as OFDM-IM [9], is desirable.

In this work, we study the subclass of permutation modulation where K out of N resources are active during each channel use. We concentrate our investigation on OFDM-IM systems [9], because OFDM-IM is a primary user of the permutation modulation subclass that we study, and any results obtained for full binary trees would be directly applicable to other permutation modulation schemes. Our main goal is to optimize the bit-to-pattern mapping operation and transmit power allocation strategy for achievable rate maximization when channel state information is available at the transmitter. We make the following contributions.

1) We give a complete and rigorous formulation of the bit-to-pattern mapping problem using the formalism of full binary trees, which covers all admissible pattern probability distributions given a uniform binary source. To this end, we report a new method to generate a reduced set of these trees and establish bounds on the number of trees in this set, which have not been reported in the mathematics or engineering literature to the best of our knowledge.
2) We formulate a relaxation of the achievable rate optimization problem with pattern probabilities and transmit powers as the optimization variables and give a number of analytic bounds and high/low-SNR asymptotic results that can be used to (approximately) solve the problem. While a similar study is carried out in [21], our results are distinct, because (i) they correspond to a different system (i.e., OFDM vs. MIMO, which assumes a different model of the channel input-output relationship) and (ii) our scheme allows for $K > 1$ active resources, while in [21] a single antenna is selected for each channel use.
3) We propose an efficient, heuristic algorithm that projects a relaxed pattern probability distribution onto the feasible set of distributions that obey the full binary tree constraints, and demonstrate that this method yields an achievable rate that is superior to a conventional OFDM-IM benchmark.

The rest of the paper is organized as follows. In Section II, the basic OFDM-IM model is described, with emphasis being placed on the binary-tree encoding operation. Section III explores the fundamental properties of binary trees; in this section details of the new tree construction algorithm are provided along with proof of completeness and bounds on the number of trees of a given size are reported. In Section IV, a relaxation of the achievable rate optimization problem is explained, and several analytic bounds and asymptotic results related to this problem are given. The fully constrained optimization problem is treated in Section V, where the aforementioned heuristic projection algorithm is outlined. A numerical analysis of all results reported in the paper are included in Section VI, and conclusions are drawn in Section VII.

II. System Model

A. Binary-Tree Encoding

Consider a binary sequence $\{B_n\}_{n\in\mathbb{N}}$, which is conveyed from a maximum entropy source to an OFDM-IM encoder. The maximum entropy property of the source implies the sequence elements $B_n \in \{0,1\}$ are independent, uniformly distributed random variables. The encoder partitions[1] the sequence $\{B_n\}$ into two subsequences $\{B_{n_k}\}$ and $\{B_{n_\ell}\}$, where $k,\ell \in \mathbb{N}$ and $k \neq \ell$. One subsequence (say, $\{B_{n_k}\}$) is mapped to a sequence of M-ary complex-valued constellation symbols. For example, if 16-QAM is employed, $M = 16$, and each group of $m = \log_2 M = 4$ bits in $\{B_{n_k}\}$ is mapped to a QAM symbol. The other subsequence is used to assign the M-ary symbols to subcarriers in preparation for transmission. In the IM system considered in this paper, we assume each OFDM symbol vector is comprised of G groups of N subcarriers, and $K \leq N$ subcarriers in each group are active, while the remaining $N - K$ subcarriers are nulled[2]. In keeping with convention, we use the term *subcarrier activation pattern (SAP)* to denote a pattern of K active subcarriers (out of N).

We are interested in system designs that maximize the achievable rate of OFDM-IM; hence, we consider bit-to-SAP mapping strategies that cover the full set of available SAPs. Since there are $\binom{N}{K}$ SAPs, it is generally not possible to construct a fixed-length bit-to-SAP mapping scheme that satisfies this condition[3]. For example, with $N = 4$ and $K = 2$,

[1]The exact detail of how this partitioning is accomplished is beyond the scope of this paper.

[2]We will focus on the case where the inequality is strict, since $K = N$ corresponds to a conventional OFDM system.

[3]In the context of OFDM-IM, a fixed-length coding scheme implies that the lengths of bit streams mapped to each respective SAP are the same. In contrast to the concept of the fixed-length coding scheme, a variable-length coding scheme encodes bit streams with different lengths by mapping these to different SAPs according to certain criteria. More rigorous and general definitions of both terms can be found in [25].

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

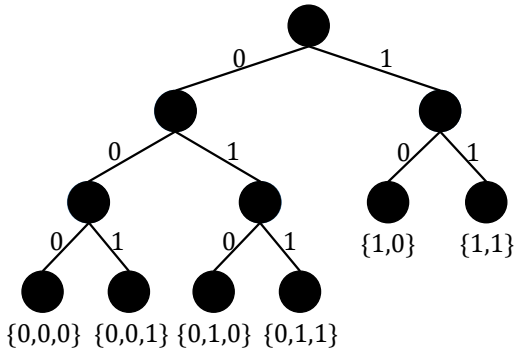


Fig. 1: A possible bit-to-SAP mapping when $N = 4$ and $K = 2$. Each edge is labeled with zero or one. Each SAP is associated with a unique leaf. For each leaf, the set below is the bit sequence mapped to that leaf.

six SAPs are available. By using a fixed-length mapping scheme, it would be possible to map two bits to one of four SAPs, leaving two SAPs unused.

To overcome this issue, we employ a variable-length scheme based on *full binary trees*. A tree is a full binary tree if every node other than the leaf nodes has exactly two children. Every full binary tree comprised of v internal nodes has $v+1$ leaves. Considering the total set of v-node full binary trees[4], the maximum depth of a tree in the set ranges from $\lfloor \log_2 v \rfloor + 1$ to v.

It is well known that one can map symbols from a source alphabet to uniquely and instantaneously decodable bit sequences using full binary trees. Indeed, this method is employed in the celebrated Huffman source coding algorithm. For the IM system considered herein, we apply a *reverse mapping* approach, which entails the use of a chosen binary tree to map source bit sequences to SAPs. Each edge of the tree is labeled with a zero or a one, and the tree is constructed such that it has $\binom{N}{K}$ leaves. Each SAP in the set of $\binom{N}{K}$ admissible patterns is associated with a leaf. The bit-to-SAP mappings are determined by tracing the unique path from the root node to each leaf, recording the bit labels for each edge in order along the way. Fig. 1 provides an illustration of this procedure for the example of $N = 4$ and $K = 2$.

Similar to Huffman source coding, the use of full binary trees to develop a bit-to-SAP mapping rule ensures each mapping is unique and instantaneously *encodable*. Uniqueness results from the binary tree structure. Instantaneous encodability simply means that the encoder can map bit sequences to SAPs using the minimum amount of information. To illustrate this point, we can again turn to Fig. 1. Suppose the SAP bit sequence is $\{0, 0, 1, 1, 0, 1, 1\}$. Reading left to right and referring to Fig. 1, we see that the encoder only needs to read the first three bits to map them to the SAP associated with the second leaf. The encoder would then read $\{1, 0\}$, which also yields a valid SAP (the fifth leaf), and so on. In this example, it is clear that the encoder does not need to interpret long sequences of bits in order to decide upon the correct mapping relating to the first few bits.

It is also important to note that *every possible* two and three-bit sequence is accounted for in the mapping shown in Fig. 1. This property extends to mappings based on other values of N and K. For a maximum entropy bit source, each subsequence consisting of q bits will appear with probability $1/2^q$. Thus, we immediately deduce that an SAP associated with a leaf node at level q below the root will be transmitted with probability $1/2^q$. *This feature of binary-tree encoding imposes a constraint on the system, which much of the literature published on this topic to date has largely ignored.* In this work, we will exploit the structure imposed by this constraint to develop efficient optimization procedures for OFDM-IM systems.

B. OFDM Model

Once bit-to-symbol mappings (both constellation and SAP) have been completed, each length-GN OFDM symbol vector is processed with a GN-point inverse discrete Fourier transform (DFT) and a cyclic prefix of adequate length (to mitigate the effects of channel dispersion) is appended to each time-domain symbol array prior to filtering, up-conversion, and transmission. At the receiver, the received signal is down-converted, filtered, and sampled. The cyclic prefix is then removed from each received baseband symbol vector before processing with a DFT. It is well-known that this sequence of procedures converts the dispersive channel into a parallel channel, and the signal on each subcarrier is (ideally) free of interference from other subcarriers.

We now formalize the OFDM model. Define $C := \binom{N}{K}$. We can uniquely associate each SAP with an index in the set $\mathcal{U} := \{1, \ldots, C\}$. For each $i \in \mathcal{U}$, denote by $\mathcal{S}_i \subseteq \{1, \ldots, N\}$ the set of indices of the K subcarriers that are active under pattern i, where equality holds when $K = N$ (which corresponds to a conventional OFDM system). The index symbol U is randomly distributed over $\mathcal{U}$ with probabilities $p_i := \mathbb{P}(U = i)$, $i \in \mathcal{U}$. The channel input-output relationship for subcarrier $l \in \{1, \ldots, N\}$ conditioned on the SAP can be written as

$$Y_l \mid U = i \text{ equals } \begin{cases} \sqrt{g_l} e^{\mathrm{j}\theta_l} X_l + Z_l, & \text{if } l \in \mathcal{S}_i, \\ Z_l, & \text{if } l \notin \mathcal{S}_i, \end{cases} \tag{1}$$

where $\sqrt{g_l} e^{\mathrm{j}\theta_l}$ is the complex channel coefficient for subcarrier l (with $\mathrm{j} = \sqrt{-1}$); the input symbols $\{X_l\}$ are zero mean and independent over the subcarriers with ρ_{li} being the transmit power on subcarrier l for index i; the noise is independent over the subcarriers with $Z_l \sim \mathcal{CN}(0, \sigma^2)$. Throughout this paper, we will assume the channel gains $\{g_l\}$ are known at the transmitter. We will return to this model in the context of mutual information optimization in Sections IV and V.

III. Full Binary Trees

One of the goals of this work is to develop a method of computing the bit-to-SAP mapping that maximizes the achievable rate of an OFDM-IM system. This is equivalent to determining the full binary tree that defines the optimal mapping. To achieve this aim, we will need a method of

[4]For the rest of the paper, unless explicitly stated otherwise, a v-node full binary tree is one with v internal nodes.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

considering all full binary trees of a given size as well as all SAP-to-leaf assignments. At first glance, this is a complicated problem. The number of v-node trees is given by the Catalan number

$$c_v = \frac{1}{v+1}\binom{2v}{v} \sim \frac{4^v}{v^{3/2}\sqrt{\pi}}$$

and the number of SAP-to-leaf assignments is $(v+1)!$. However, it is possible to significantly simplify the problem by making use of symmetry. The important aspect of the mapping is not in the exact tree that is chosen, but rather in the level of the leaf node that a given SAP is assigned to. As noted in Section II, an SAP assigned to a leaf at level q has probability $1/2^q$ of being transmitted. We can transpose leaf nodes at a given level in any way we wish and still achieve the same SAP probability distribution. This reasoning leads us to consider a smaller set of trees, which we call the *reduced set of v-node full binary trees* $\mathcal{T}_v$. Each tree in this set actually corresponds to an automorphism group of the complete set. Moreover, consider a given tree $t \in \mathcal{T}_v$ and denote the number of leaves at level q by n_q. Due to the symmetry stated above, the number of ways of assigning $v+1$ objects (i.e., SAPs, where $v+1 = C$) to the leaf nodes such that we attain a unique probability distribution is

$$\binom{v+1}{n_1, n_2, \ldots, n_v} = \frac{(v+1)!}{n_1! n_2! \cdots n_v!} \tag{2}$$

which can be considerably smaller than the total $(v+1)!$ permutations. We now give preliminary results on the construction and enumeration of the set $\mathcal{T}_v$, which will be useful in determining systematic optimization procedures and analyzing computational complexity.

A. Construction

In order to choose the best tree for encoding, we require a method of constructing all trees in $\mathcal{T}_v$. The approach we propose is outlined in Algorithm 1, which is valid for $v \geq 2$. The initial set $\mathcal{T}_1 = \{\tau\}$ consists of the single full binary tree τ with one root and two leaves (at level one). This *protograph* is recursively appended to trees to obtain the set $\mathcal{T}_v$. The algorithm is presented in a somewhat informal way here for clarity; we formalize it slightly in Appendix A in order to prove the following proposition.

Proposition 1. *Algorithm 1 returns all elements in $\mathcal{T}_v$.*

Proof: See Appendix A. □

As an example of the output of Algorithm 1, Fig. 2 shows the sets generated for $v = 1, 2, 3$. Note that the tree shown for the set $\mathcal{T}_1$ is the protograph τ. The number of protographs contained in a graph of $\mathcal{T}_v$ is v.

B. Enumeration

As noted above, the number of ordered full binary trees with v internal nodes is given by the Catalan number c_v. The reduced set of v-node full binary trees contains significantly fewer elements. For example, Fig. 2 shows that two trees are contained in $\mathcal{T}_3$; yet, by considering all orderings of these two trees, we can enumerate five ordered trees ($c_3 = 5$).

Algorithm 1: Recursive Construction of $\mathcal{T}_v$

initialize $\mathcal{T}_1 = \{\tau\}$; $\mathcal{T}_{k'} = \{\}\ \forall k' = 2, \ldots, v$; $k = 2$;
while $k \leq v$ **do**
  for $t \in \mathcal{T}_{k-1}$ **do**
    Append τ to the left-most leaf on the lowest level of t, and add the new tree to $\mathcal{T}_k$;
    Append τ to the left-most available leaf on the next-to-lowest level of t where possible, and add the new tree to $\mathcal{T}_k$;
  end
  $k \leftarrow k + 1$;
end

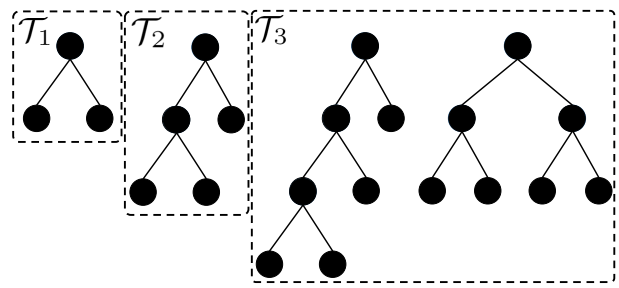


Fig. 2: Illustration of the sets $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$ generated by Algorithm 1.

Let T_v denote the number of trees in the set $\mathcal{T}_v$. From Algorithm 1, we can infer the relations

$$T_v \leq 2T_{v-1} \leq \cdots \leq 2^{v-1}T_1 = 2^{v-1} \tag{3}$$

since each step in the for loop *at most* doubles the number of elements in $\mathcal{T}_k$. This bound captures the slower exponential growth in the number of trees in the reduced set compared to the set of ordered trees. Numerical results have shown that the bound overestimates the rate of increase in v. Published results on full binary trees have attempted to obtain generating functions for the number of trees in unordered, unlabelled sets (see, e.g., [26] and references therein). However, it appears that results on the reduced sets that we are interested in remain undiscovered.

It is possible to obtain a tighter bound on T_v by analyzing Algorithm 1. The bound is given as a recurrence relation in the following proposition.

Proposition 2. *The number of trees in $\mathcal{T}_v$ is upper bounded by*

$$T_v \leq 2T_{v-1} - \delta_v - \sum_{q=2}^{\lfloor \log_2(v-1) \rfloor} T_{v-2^q} \tag{4}$$

where $\delta_v = 1$ if v is a power of two and $\delta_v = 0$ otherwise, and the summation is empty when $v < 5$.

Proof: See Appendix B. □

The accuracy of each of the two bounds given above is illustrated for sets of up to twenty internal nodes in Fig. 3. From the figure, we see that the loose bound slightly overestimates the growth rate of T_v. The recursion is exact up to $v = 9$, but slowly diverges for larger v, although it clearly remains

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

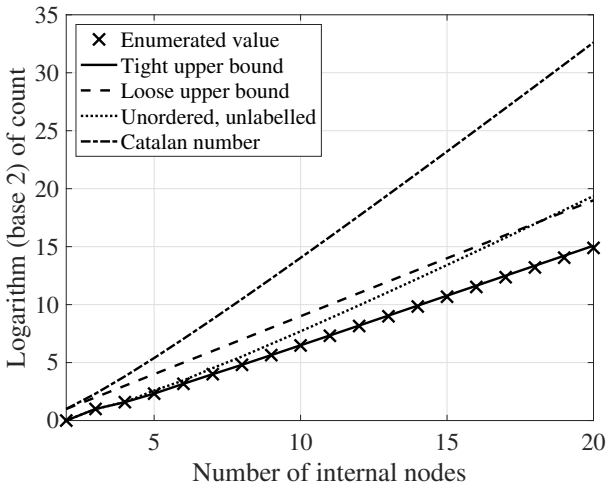


Fig. 3: Enumeration of and bounds on the number of full binary trees as a function of the number of internal nodes. The enumerated result corresponds to Algorithm 1. The loose bound is 2^{v-1} (cf. (3)). The tight bound corresponds to the result given in Proposition 2, where the recursion is performed over the bounds on T_v rather than the exact enumerated values. The "Unordered, unlabelled" plot corresponds to the enumeration given in [26]. The number of ordered full binary trees (the Catalan number) is plotted as a reference.

fairly tight up to $v = 20$. Practically, we will be interested in reasonably small v; hence, the recursion is a useful tool for analyzing the IM systems studied in this paper.

IV. Mutual Information Optimization: Relaxation

We now provide details of new results and methods related to the optimization of the mutual information in OFDM-IM systems. As noted in Section II, the SAP probabilities are constrained by the binary tree chosen for encoding. Before we treat these constraints, we will consider the relaxed problem, for which it is assumed that SAPs can be transmitted with any probability. This will give an upper bound on the achievable rate for the constrained system, and we will use the approaches developed herein to treat that case in Section V.

Consider a single set of N subcarriers that adhere to the model described in Section II. We collect the N received symbols in the vector $\boldsymbol{Y} := (Y_1, \ldots, Y_N)$. Furthermore, we collect the K transmitted symbols in the vector $\boldsymbol{X} := (X_1, \ldots, X_N)$, noting that X_l is nonzero only when subcarrier l is active, as given by the encoded SAP. Define the SAP probability vector $\mathbf{p} = (p_i)$ and the power vector $\boldsymbol{\rho} = (\rho_{li})$. We are interested in the probabilities in $\mathbf{p}$ and transmit powers in $\boldsymbol{\rho}$ that maximize the mutual information

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= h(\boldsymbol{Y}) - h(\boldsymbol{Y} \mid \boldsymbol{X}) \\ &= h(\boldsymbol{Y}) - h(\boldsymbol{Z}). \end{aligned} \tag{5}$$

Conditioned on $U = i$, we assume $X_l \sim \mathcal{CN}(0, \rho_{li})$ when $l \in \mathcal{S}_i$. Choosing X_l to be Gaussian is not proven to achieve capacity, but the assumption provides a tractable expression. In this case, the complex random vector $\boldsymbol{Y}$ has probability density function (pdf)

$$f_{\boldsymbol{Y}}(\mathbf{y}) = \sum_{i \in \mathcal{U}} p_i f_{\boldsymbol{Y}|U}(\mathbf{y} \mid U = i) \tag{6}$$

where

$$f_{\boldsymbol{Y}|U}(\mathbf{y} \mid U = i) = \prod_{l \in \mathcal{S}_i} f_{\text{CN}}(y_l; g_l \rho_{li} + \sigma^2) \prod_{m \notin \mathcal{S}_i} f_{\text{CN}}(y_m; \sigma^2) \tag{7}$$

with $f_{\text{CN}} : \mathbb{C} \to [0, \infty)$ being the complex Gaussian pdf $f_{\text{CN}}(t; \nu) = e^{-|t|^2/\nu}/(\pi\nu)$ with mean zero and variance ν.

Writing $I(\boldsymbol{X};\boldsymbol{Y}) = I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$, the optimization problem is formulated as

$$\begin{aligned} \underset{\mathbf{p},\boldsymbol{\rho}}{\text{maximize}} \quad & I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) \\ \text{subject to} \quad & \sum_{i \in \mathcal{U}} p_i = 1 \\ & \sum_{l \in \mathcal{S}_i} \rho_{li} \le P, \quad \forall i \in \mathcal{U} \\ & p_i \ge 0, \quad \forall i \in \mathcal{U} \\ & \rho_{li} \ge 0, \quad \forall l \in \mathcal{S}_i,\, i \in \mathcal{U}. \end{aligned} \tag{8}$$

Note that the relaxation alluded to earlier manifests in the simple constraint $\sum_{i \in \mathcal{U}} p_i = 1$. If we were to consider only probability vectors $\mathbf{p}$ that adhere to the binary-tree encoding methodology, this constraint would be defined differently (see Section V). We now detail several strategies for solving, either approximately or exactly, the optimization problem stated in (8).

A. Concavity and Numerical Optimization

The following result that can be used to solve (8) numerically.

Lemma 1. *For fixed $\boldsymbol{\rho}$, the problem*

$$\begin{aligned} \underset{\mathbf{p}}{\textit{maximize}} \quad & I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) \\ \textit{subject to} \quad & \sum_{i \in \mathcal{U}} p_i = 1 \\ & p_i \ge 0, \quad \forall i \in \mathcal{U}. \end{aligned} \tag{9}$$

is concave.

Proof: See Appendix C. □

For the special case where g_l is constant for all l and a balanced power distribution is chosen (i.e., $\rho_{li} = \rho$ for all l and i), Lemma 1 leads to the following.

Proposition 3. *When the channel gains and transmit powers are constant across frequency, the optimal SAP probability distribution is uniform.*

Proof: See Appendix D. □

More generally, Lemma 1 suggests that it may be economic to solve (8) by employing a block coordinate descent (BCD) approach [27], in which one would alternately maximize the mutual information in either $\mathbf{p}$ or $\boldsymbol{\rho}$ while keeping the other vector fixed at the previously obtained optimum value. The method requires the constraints of the problem to be convex, which is clearly satisfied. Furthermore, the maximization over each of the vectors $\mathbf{p}$ and $\boldsymbol{\rho}$, keeping the other constant, must be unique. Lemma 1 implies this condition is met in part, but it is not clear whether the condition may be violated for the maximization of $I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$ over $\boldsymbol{\rho}$ for a fixed $\mathbf{p}$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

in some parameterizations of $\{g_l\}$ and σ^2. Nevertheless, the smoothness of the objective function provides some assurance that a BCD approach will converge to a local extremum.

One may encounter numerical problems when using the BCD technique to solve (8) since, in general, the evaluation of $I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$ requires high-dimensional numerical integration or time-consuming Monte Carlo methods. In practice, we have found that the BCD method can only be employed to optimize systems with three or four subcarriers per group; larger systems require different approaches. In this regard, one may think of applying random search algorithms, such as genetic and simulated annealing algorithms, which are computationally simple, quick, and robust for complex systems and are capable of providing efficient solutions for many ill-structured global optimization problems [28].

B. A Lower Bound

It is possible to obtain an approximate solution to (8) by considering a lower bound on the mutual information rather than the mutual information, itself. The following proposition provides one such bound.

Proposition 4. *For transmit powers $\boldsymbol{\rho}$ and SAP probabilities $\mathbf{p}$, $I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$ satisfies the lower bound*

$$I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) \geq -\ln\left(\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}} \frac{p_i p_j}{\det(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_j)}\right) - N\ln\left(e\sigma^2\right) \tag{10}$$

where $\boldsymbol{\Xi}_i$ is a diagonal matrix with the lth element of the diagonal ξ_{li} satisfying

$$\xi_{li} = \begin{cases} g_l\rho_{li} + \sigma^2, & \text{if } l \in \mathcal{S}_i, \\ \sigma^2, & \text{otherwise.} \end{cases} \tag{11}$$

Proof: See Appendix E. □

The bound given above is a result of Jensen's inequality and is, thus, not particularly tight. In fact, a slightly different application of the inequality yields a marginally tighter bound [29, Th. 2]. However, the utility in Proposition 4 is not in the accuracy of the bound, but rather in the ease with which this bound can be optimized over the SAP probabilities. These optimal probabilities are captured in the following proposition.

Proposition 5. *Let $\boldsymbol{A} = (a_{ij})$ with $a_{ij} = 1/\det(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_j)$. Suppose $\boldsymbol{A}$ is nonsingular, and let $\boldsymbol{B} = \boldsymbol{A}^{-1}$, with b_{ij} denoting the element in the ith row and jth column of $\boldsymbol{B}$. The SAP probabilities that maximize the lower bound given in* (10) *are given by*

$$p_i = \frac{\left(\sum_{j\in\mathcal{U}} b_{ij}\right)^+}{\sum_{i\in\mathcal{U}}\left(\sum_{j\in\mathcal{U}} b_{ij}\right)^+}, \quad \forall i \in \mathcal{U} \tag{12}$$

where $(x)^+ = \max\{x, 0\}$.

Proof: See Appendix F. □

Note that the probabilities given in (12) are dependent upon the subcarrier powers. The BCD approach can be employed in a fairly straightforward manner to compute the power values by alternately computing (12) for fixed powers, then fixing these probabilities in (10) and computing the maximizing power values. Alternatively, one can, in theory, substitute (12) into (10) and compute the optimal powers directly. However, the nonlinear form of (12) can cause problems using this approach.

A condition that must be satisfied in order to invoke Proposition 5 is that $\boldsymbol{A}$ must be nonsingular. It is possible that this condition is not met, for example when only a single subcarrier in the set of K active subcarriers is allocated power. Such cases can typically be dealt with by using other results reported in this section (e.g., the asymptotic results detailed below). In general, we have found that Proposition 5 is applicable to a wide range of system configurations.

C. Closed-Form Asymptotics

It is naturally preferable to solve (8) analytically. To make progress in this direction, we apply the following strategy: first, we find the probabilities $\mathbf{p}^\star(\boldsymbol{\rho})$ that maximize the mutual information for any given values of the transmit powers, i.e., the optimal probabilities are functions of the powers; then, the mutual information $I(\mathbf{p}^\star(\boldsymbol{\rho}), \boldsymbol{\rho}, \sigma^2)$ that corresponds to the optimal probabilities found previously is maximized over the powers in $\boldsymbol{\rho}$.

1) Probability Optimization: To obtain a closed-form expression for the optimal SAP probability distribution as a function of the powers, we first resort to a high-SNR analysis, which gives rise to the following result.

Proposition 6. *For fixed powers $\boldsymbol{\rho}$, let $I^\star(\boldsymbol{\rho}, \sigma^2) = \max_{\mathbf{p}} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$ and define the probability values*

$$q_i = q_i(\sigma^2) = \frac{\prod_{l\in\mathcal{S}_i}(g_l\rho_{li} + \sigma^2)}{\sum_{j\in\mathcal{U}}\prod_{m\in\mathcal{S}_j}(g_m\rho_{mj} + \sigma^2)}. \tag{13}$$

Then, $I(\mathbf{q}, \boldsymbol{\rho}, \sigma^2)$ is a lower bound of $I^\star(\boldsymbol{\rho}, \sigma^2)$ that is tight in the high SNR regime, i.e., as $\sigma^2 \to 0$.

Proof: See Appendix G. □

In addition to a simple, closed-form expression for the optimal SAP probability distribution at high SNR, this result also provides an upper bound on the achievable rate, as stated in the following corollary.

Corollary 1. *The function $\mu(\sigma^2) := \ln\left[\sum_{i\in\mathcal{U}}\prod_{l\in\mathcal{S}_i}\left(\frac{g_l\rho_{li}}{\sigma^2} + 1\right)\right]$ is an upper bound on the maximal mutual information $I^\star(\boldsymbol{\rho}, \sigma^2)$, which is tight as $\sigma^2 \to 0$.*

Proof: See Appendix H. □

We now turn our attention to the low-SNR case, for which we obtain the following beautifully intuitive result, which is a somewhat discrete version of the well known waterfilling principle at low SNR.

Proposition 7. *For fixed powers $\boldsymbol{\rho}$, let $i^\star = \arg\max_i \sum_{l\in\mathcal{S}_i}\ln\left(\frac{g_l\rho_{li}}{\sigma^2} + 1\right)$, i.e., $i^\star$ corresponds to the group of K strongest subcarriers. Then, the index probabilities*

$$r_i = \begin{cases} 1, & \text{if } i = i^\star, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

maximize the mutual information at low SNR, which satisfies the asymptotic equivalence $I^\star(\boldsymbol{\rho},\sigma^2) \sim \sum_{l\in\mathcal{S}_{i^\star}} \ln\left(\frac{a_{li^\star}}{\sigma^2}+1\right)$, *as* $\sigma^2 \to \infty$.

Proof: See Appendix I. □

2) Power Optimization: Propositions 6 and 7 and Corollary 1 yield closed-form expressions for the mutual information, which depend upon the powers $\boldsymbol{\rho}$. As a result, these expressions can be used to develop optimal power allocation rules in the high and low-SNR regimes. It turns out that the optimal rules follow our conventional understanding of power allocation in OFDM systems, as formalized in the following proposition.

Proposition 8. *For high SNR (as* $\sigma^2 \to 0$*), allocating powers for the subcarriers of each SAP according to the waterfilling strategy is optimal under power constraints for each pattern. For low SNR (as* $\sigma^2 \to \infty$*), allocating powers according to the waterfilling strategy is optimal.*

Proof: See Appendix J. □

The waterfilling result for the low-SNR case is somewhat unsurprising given that Proposition 7 indicates the mutual information expression is the same as that for OFDM with only K active subcarriers. On the other hand, the optimality of waterfilling at high SNR is not immediately obvious from Corollary 1. The waterfilling strategy gives equal powers when $\sigma^2 \to 0$. However, this allocation scheme is near-optimal at high, but finite SNR (in which case the resulting powers are not equal), because the waterfilling strategy results from optimizing the upper bound given by Corollary 1, which is very close to the maximal mutual information at high SNR (and asymptotically tight). These results lead us to a simple mutual information optimization strategy for $\mathbf{p}$ and $\boldsymbol{\rho}$ at high and low SNR: one should perform waterfilling power allocation for each subcarrier pattern and then compute the corresponding probabilities according to Proposition 6 or 7 and select the result that maximizes the objective.

V. Achievable Rate Optimization: Constrained

We now consider a more practical rate optimization problem that is effectively the same as (8) but with a nonlinear constraint on the probabilities $\{p_i\}$. As discussed in Sections II and III, SAP probabilities depend on two things: (1) the full binary tree that corresponds to the bit-to-SAP mapping operation, and (2) the ordering of the SAP-to-leaf assignment.

Let $\mathcal{P}_v$ denote the set of feasible probability vectors of length C that can be constructed by considering all non-redundant SAP-to-leaf assignments for all binary trees in $\mathcal{T}_v$ *including null assignments*. For example, for $C = 4$, $\mathcal{P}_2$ is constructed by considering all mappings of three (out of four) SAPs to two leaves on the second level and one leaf on the first level of the single tree in $\mathcal{T}_2$ (cf. Fig. 2). There are $3!/2! = 3$ mappings of three SAPs to the leaves, and $\binom{4}{3} = 4$ ways of choosing the *active* SAPs. The inclusion of null assignments in this way ensures we consider the case of not using some SAPs that may correspond to poor channel conditions. Under this definition of $\mathcal{P}_v$, the number of elements (probability vectors) in $\mathcal{P}_v$ is

$$P_v = \binom{C}{v+1} \sum_{t\in\mathcal{T}_v} \frac{(v+1)!}{n_{t1}!\cdots n_{tv}!} \tag{15}$$

where n_{tq} denotes the number of leaves at level q in tree t. We further define the union $\mathcal{P} = \cup_{v=0}^{C-1}\mathcal{P}_v$, where $\mathcal{P}_0$ consists of the C vectors with one element equal to one and the rest equal to zero.

The constrained optimization problem can now be formulated as

$$\begin{aligned} \underset{\mathbf{p},\boldsymbol{\rho}}{\text{maximize}} \quad & I(\mathbf{p},\boldsymbol{\rho},\sigma^2) \\ \text{subject to} \quad & \sum_{i\in\mathcal{U}} p_i = 1 \\ & \mathbf{p} \in \mathcal{P} \\ & \sum_{l\in\mathcal{S}_i} \rho_{li} \le P, \quad \forall i \in \mathcal{U} \\ & \rho_{li} \ge 0, \quad \forall l \in \mathcal{S}_i,\, i \in \mathcal{U}. \end{aligned} \tag{16}$$

We propose two methods of solving this problem here: an enumerative approach, and a projection from the relaxation.

A. Enumerative Approach

This approach is based on the enumeration of all possible probability distributions of the SAPs, i.e., all $\mathbf{p} \in \mathcal{P}$. The allocated powers $\boldsymbol{\rho}$ are optimized for each probability distribution $\mathbf{p}$. The pair $(\mathbf{p},\boldsymbol{\rho})$ that yields the highest mutual information is the solution to the problem stated in (16).

For a given probability distribution $\mathbf{p}$, the power allocation problem may not be solved analytically. In this case, one can invoke the asymptotic results stated in Proposition 8 to obtain the power values. First, the waterfilling power allocation solution would be calculated for each SAP. Then, the distribution $\mathbf{p} \in \mathcal{P}$ that maximizes the mutual information would be chosen.

B. Projection from the Relaxation

A much more computationally efficient method of treating (16) can be developed by first considering the relaxation studied in the previous section. First, we relax the constraint $\mathbf{p} \in \mathcal{P}$ to find a solution to (8). Any of the approaches used in Section IV can be applied. We let $\mathbf{p}'$ denote the probability distribution computed in this step. We then project $\mathbf{p}'$ onto the feasible vector $\mathbf{p}^\star \in \mathcal{P}$ and take this to be the partial solution to (16). The power allocation vector $\boldsymbol{\rho}^\star$ is then computed to maximize the mutual information.

The projection of the relaxed solution $\mathbf{p}'$ onto a point in the set $\mathcal{P}$ can be accomplished efficiently by using the Huffman coding algorithm [19]. To this end, we interpret the elements of $\mathbf{p}'$ as source symbol probabilities, then generate a full binary tree according to the Huffman algorithm. As discussed in Section II, SAPs associated with a leaf node in the tree at level q will be transmitted with probability $1/2^q$. Hence, the probabilities in $\mathbf{p}'$ are replaced with the corresponding probabilities derived from the tree structure to yield a candidate for $\mathbf{p}^\star$.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

Algorithm 2: Projection from the Relaxation

initialize Compute $\mathbf{p}'$ by (approximately) solving (8), set $\bar{\mathbf{p}} = \mathrm{sort}(\mathbf{p}')$, set $k = 1$;
while $k \leq C$ **do**
  $s_k = \sum_{l=1}^{C-k+1} \bar{p}_l$;
  $\{q_1, \ldots, q_{C-k+1}\} =$
   $\mathrm{Huffman}(\bar{p}_1/s_k, \ldots, \bar{p}_{C-k+1}/s_k)$;
  $\mathbf{p}_k^\star = \mathrm{unsort}(1/2^{q_1}, \ldots, 1/2^{q_{C-k+1}}, 0, \ldots, 0)$;
  $k \leftarrow k + 1$;
end
return $\mathbf{p}^\star = \arg\min_{\mathbf{t} \in \{\mathbf{p}_1^\star, \ldots, \mathbf{p}_C^\star\}} \mathrm{dist}(\mathbf{t}, \mathbf{p}')$;

It is important to note that this basic approach will only yield trees (and associated probability distributions) with C leaves, i.e., the algorithm maps $\mathbf{p}'$ to $\mathcal{P}_C$ only. To ensure we consider mappings to all points in $\mathcal{P}$, we require a slightly modified approach. The full details of the complete projection algorithm are given in Algorithm 2, and an example depicting how the algorithm works is shown in Fig. 4. The function $\mathrm{sort}(\cdot)$ in Algorithm 2 arranges the set of C arguments in decreasing order; $\mathrm{unsort}(\cdot)$ performs the inverse mapping (again, acting on C elements). The function $\mathrm{Huffman}(\cdot)$ takes a set of "source probabilities" and returns the corresponding set of depths, or path lengths from the root to a given leaf. Finally, the function $\mathrm{dist}(\mathbf{p}_1, \mathbf{p}_2)$ computes the distance between the discrete probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$. In the next section, we consider three distance measures: Euclidean distance, for which

$$\mathrm{dist}(\mathbf{p}_1, \mathbf{p}_2) \coloneqq \|\mathbf{p}_1 - \mathbf{p}_2\|^2 \tag{17}$$

Kullback–Leibler (KL) divergence, for which

$$\mathrm{dist}(\mathbf{p}_1, \mathbf{p}_2) \coloneqq \sum_{i \in \mathcal{U}} p_{1i} \ln \frac{p_{1i}}{p_{2i}} \tag{18}$$

and total variation distance, for which

$$\mathrm{dist}(\mathbf{p}_1, \mathbf{p}_2) \coloneqq \max_{i \in \mathcal{U}} \{|p_{1i} - p_{2i}|\}. \tag{19}$$

Algorithm 2 is heuristic. It is not guaranteed to produce the solution to (16). However, results have shown it performs very well in practical scenarios (see Section VI).

VI. Numerical Results

In this section, we present a numerical analysis of the methods described above. We begin with a discussion of the mutual information. We then give a brief analysis of the error rate of the systems described herein. In what follows, we define $SNR \coloneqq P/(N\sigma^2)$, which can be interpreted as the average *transmit* SNR per subcarrier. All mutual information results are given in units of nats, and all curves were obtained via Monte Carlo sampling when closed-form expressions were not available. For all systems, we set $N = 4$ and $K = 2$. It should be noted, however, that we also performed extensive simulations for $(N, K) = (6, 4)$ and $(N, K) = (8, 6)$, and observed very similar trends to the case where $N = 4$ and $K = 2$.

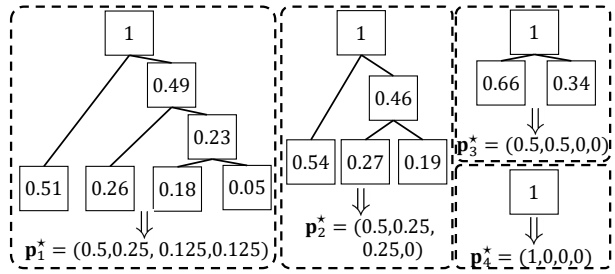


Fig. 4: Illustration of Algorithm 2 for a system with $N = 4$ and $K = 1$, where $\mathbf{p}' = (0.51, 0.26, 0.18, 0.05)$. The algorithm constructs four full binary trees according to the input distributions $(0.51, 0.26, 0.18, 0.05)$, $(0.54, 0.27, 0.19)$, $(0.66, 0.34)$, and (1), respectively. The output probability distributions of the algorithm are as follows: $(0.5, 0.25, 0.25, 0)$ for the Euclidean distance metric, $(0.5, 0.25, 0.125, 0.125)$ for the KL divergence metric, and $(0.5, 0.25, 0.25, 0)$ for the total variation distance metric.

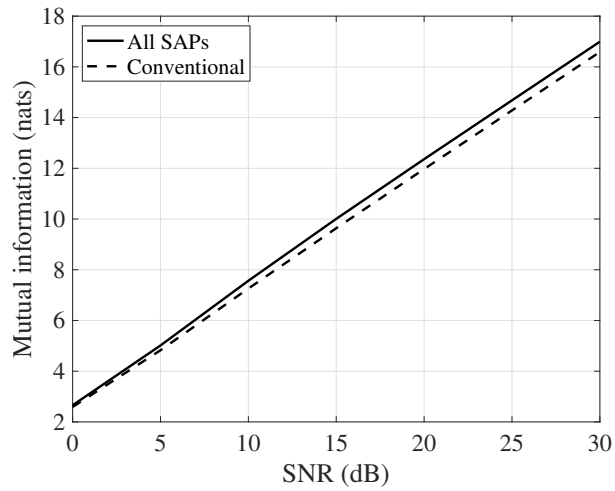


Fig. 5: Mutual information vs. SNR for equal channel gains and uniform power allocation.

A. Mutual Information

We begin with a simple case. Consider a system operating in AWGN with equal channel gains (i.e., $g_l = 1$ for all l). Proposition 3 states that the optimal SAP probability distribution in this system is uniform. Adopting this result, Fig. 5 shows the mutual information for an OFDM-IM system that uses all six SAPs (each with probability $1/6$) compared to one that limits the number of utilized SAPs to four where uniform power allocation is applied. The small improvement offered by the former approach simply arises as a result of the additional SAPs that are used[5]. However, we note that a uniform SAP distribution is infeasible given a uniform binary source[6].

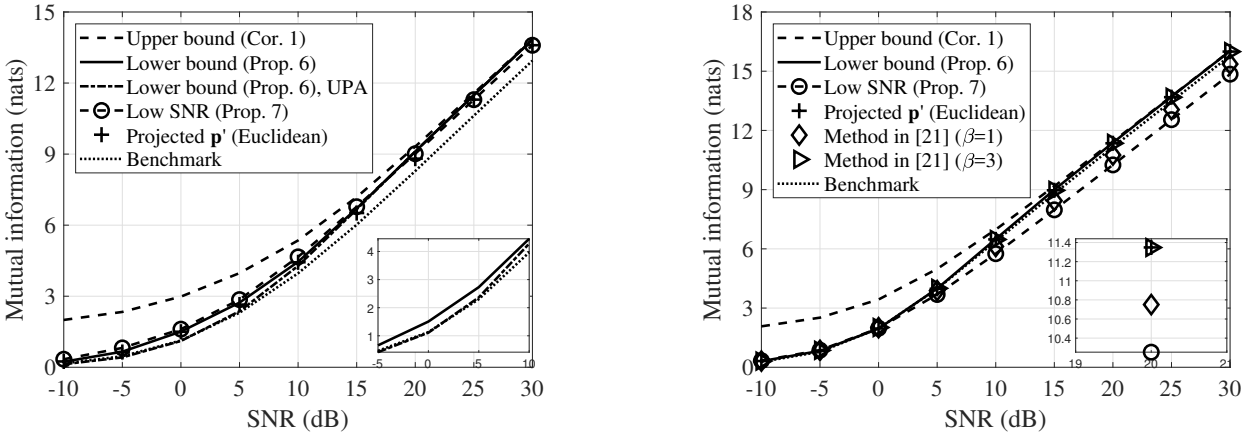
To better understand the advantages that can be brought by utilizing all SAPs along with the binary-tree encoding strategy, we now analyze the case where the channel gains are defined by $g_l = \eta^{l-1}$, $\forall\ l \in \{1, \ldots, N\}$ for some $\eta \in (0, 1)$. We assume full channel knowledge is available at the transmitter,

[5]One would expect an improvement when finite constellations are used for signalling, rather than Gaussian signals. However, this study is beyond the scope of this paper.

[6]This fact has typically been ignored in the literature to date.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing



(a) $\eta = 0.2$ (b) $\eta = 0.7$

Fig. 6: Mutual information vs. SNR for $\eta = 0.2$ and $\eta = 0.7$, where $g_l = \eta^{l-1}$ for $l = 1, \ldots, N$. Waterfilling power allocation is used for all results except for the benchmark, which uses uniform power allocation. The curve labelled "Projected $\mathbf{p}'$ (Euclidean)" gives an achievable rate when all SAPs are allowed to be utilized and power allocation is employed.

so that SAP probabilities and power allocation can be optimized. Fig. 6a shows the mutual information for $\eta = 0.2$, and Fig. 6b gives results for $\eta = 0.7$. In both figures, the different curves represent different SAP probability assignment strategies and bounds. The first four curves in Fig. 6a and the first three curves in Fig. 6b exhibit the mutual information computed by using the respective analytic results. The fifth and the forth curves ("Projected **p'** (Euclidean)") in Fig. 6a and Fig. 6b, respectively, illustrate the mutual information attained by employing Algorithm 2.[7] In this case, $\mathbf{p}'$ is computed by using the analytic form given in Proposition 6. Note that this curve represents an achievable rate for OFDM-IM systems that utilize all SAPs. The curves "Method in [21] ($\beta = 1$)" and "Method in [21] ($\beta = 3$)" in Fig. 6b represent the mutual information computed by using the method employed in [21]. In such a method, SAP probabilities p_i are chosen from a feasible domain up to a binary-tree depth β, i.e., $p_i \in \{0, 1, 2^{-1}, \ldots, 2^{-\beta}\}$, and the probabilities drawn from such a domain are substituted into a Monte Carlo estimator of the mutual information one-by-one; the probability set that provides the maximum mutual information is chosen for the transmission. For all curves other than the "Benchmark", "Lower bound (Prop. 6), UPA", "Method in [21] ($\beta = 1$)", and "Method in [21] ($\beta = 3$)", waterfilling power allocation is employed, since this approach is optimal at high and low SNR in the relaxed setting (cf. Proposition 8). Here, the curves "Lower bound (Prop. 6), UPA", "Method in [21] ($\beta = 1$)", and "Method in [21] ($\beta = 3$)" employ uniform power allocation. The benchmark curve relates to a standard OFDM-IM system where the four SAPs, chosen according to the lexicographic principle discussed in [16], are transmitted with equal probability and uniform power allocation[8].

In Fig. 6a, we see that the (relaxed) lower bound of Proposition 6 and the low-SNR result of Proposition 7 are similar, and that convergence to the upper bound of Corollary 1 occurs at high SNR. Moreover, the fully constrained result (where $\mathbf{p} \in \mathcal{P}$) denoted by the "$\times$" markers is very close to the analytic curves corresponding to the relaxed optimization. The benchmark curve is noticeably lower than all results that offer SAP probability optimization and power allocation. This was also seen in simulations for $(N, K) = (6, 4)$ and $(N, K) = (8, 6)$ systems (not shown). The lower bound of Proposition 6 without waterwilling power allocation ("Lower bound (Prop. 6), UPA") is also notably lower than all the results at especially low-to-medium SNR values; however, it is still more beneficial than the benchmark at medium-to-high SNR values. Turning our attention to Fig. 6b, we see that the advantages offered by optimization diminish for less variable channel conditions. The optimized scenario ("Projected $\mathbf{p}'$ (Euclidean)") still offers an advantage that saturates the upper bound from mid-to-high SNR values, but it is marginal. It is important to note that this approach outperforms the method used in [21] when $\beta = 1$ and provides an almost identical mutual information to this benchmark when $\beta = 3$. The approach also exhibits a much lower complexity, since the method used in [21] exhaustively searches all binary tree probabilities up to a binary tree depth of β; clearly, the complexity of this method increases markedly with β.

For this simple system ($N = 4$ and $K = 2$), the results shown in Fig. 6 point to a need to understand how frequency selectivity affects performance. To this end, Fig. 7 illustrates the mutual information as a function of η. The first, fourth, and fifth curves relate to those with the same labels shown

[7]Curves corresponding to other distance functions are not shown because they yield results that are nearly identical to the Euclidean case.

[8]Note that the curves that employ waterfilling power allocation and SAP probability optimization include the case where only four SAPs may be chosen, which would correspond to the benchmark curve but where waterfilling is employed. Such a selection, if deemed to be optimal, would naturally arise through the SAP probability optimization procedure. As a result, the true benchmark only utilizes uniform power allocation here.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

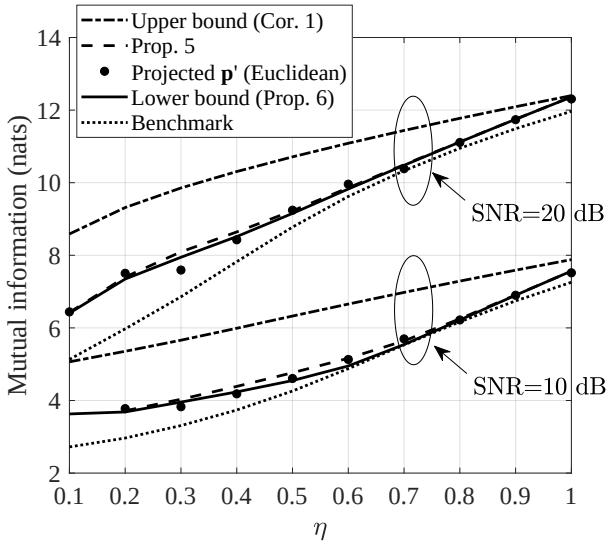
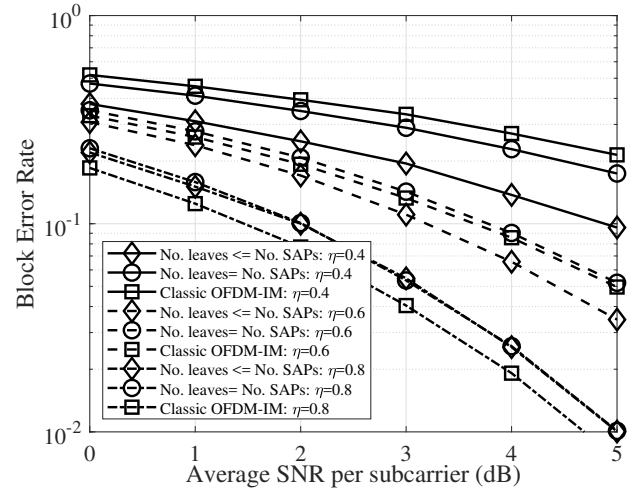


Fig. 7: Mutual information vs. η, where $g_l = \eta^{l-1}$ for $l = 1, \ldots, N$. The curve labelled "Projected $\mathbf{p}'$ (Euclidean)" gives an achievable rate when all SAPs are allowed to be utilized and power allocation is employed.

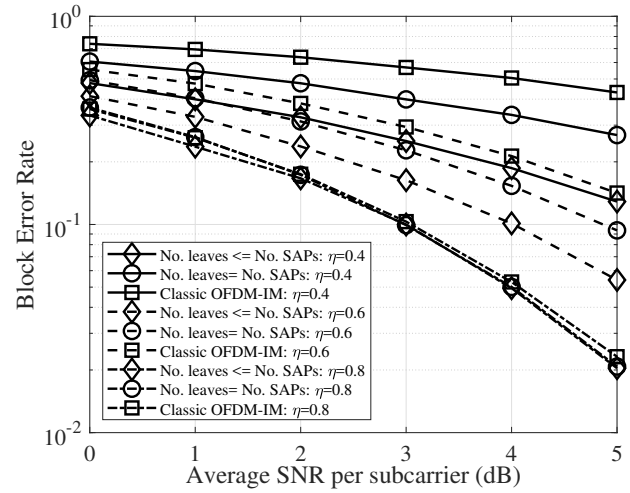
in Fig. 6. The second curve shows the mutual information attained by using the probabilities given in Proposition 5. The third curve ("Projected $\mathbf{p}'$ (Euclidean)") illustrates the mutual information attained by employing Algorithm 2, where $\mathbf{p}'$ is computed using Proposition 5. Again, waterfilling is used for all systems except the benchmark, where uniform power allocation is used. The advantages offered by optimization in highly frequency selective channels are apparent in this example[9]. It is observed that one data point is missing for the $SNR = 10$ dB curves related to Proposition 5. This omission results from the fact that the matrix $\boldsymbol{A}$ in Proposition 5 is singular for the corresponding paramterization. Hence, for this point, one would choose a different method of obtaining $\mathbf{p}'$ in the initialization step of Algorithm 2. To conclude this discussion, it is important to note that Algorithm 2 roughly achieves the same mutual information promised by the analytic lower bound. The upper bound is only tight at high SNR; hence, it is not particularly tight for most η values in this figure.

B. Block Error Rate

Apart from achievable rate, error performance is another key performance metric of a communication system. It should be apparent that a scheme that is designed to maximize the achievable rate does not necessarily optimize the error performance. Nevertheless, it is important to consider the effects that the designs detailed in Sections IV and V have on the error rate. Note that, because the lengths of the bit sequences encoded as SAPs and modulated signals are variable, the measurement of bit errors would be difficult to assess in a standardized manner. Therefore, to maintain brevity and clarity, we choose to evaluate the block-error rate (BLER)



(a) BPSK



(b) QPSK

Fig. 8: BLER vs. average SNR per subcarrier.

instead of bit-error rate for the [30]–[32]. Here, a block is a group of N subcarriers.

For simplicity and optimality, we adopt the maximum likelihood (ML) detection scheme at the receiver to estimate the received signal vector $\boldsymbol{Y}$ consisting of received modulated symbols and nulls on N independent subcarriers[10]. We assume that channel knowledge is available at both transmitter and the receiver. The estimated signal vector $\hat{\boldsymbol{X}}$ satisfies

$$\hat{\boldsymbol{X}} = \arg\min_{\dot{\boldsymbol{X}}} \left\{ \boldsymbol{Y} - \boldsymbol{G}\dot{\boldsymbol{X}} \right\} \tag{20}$$

where $\dot{\boldsymbol{X}}$ is a candidate transmit vector (obtained by following the model described in Section II) and $\boldsymbol{G} = \mathrm{diag}\{\sqrt{g_1}e^{j\theta_1}, \ldots, \sqrt{g_N}e^{j\theta_N}\}$ is the diagonal channel coefficient matrix. We write the BLER conditioned on the transmitted signal vector $\boldsymbol{X}$ as $\mathsf{P}_{\mathrm{BLER}}(\boldsymbol{X}|\boldsymbol{G}) = \mathbb{P}(\boldsymbol{X} \neq \hat{\boldsymbol{X}}|\boldsymbol{G})$. Averaging over $\boldsymbol{X}$ gives our measure of interest: $\bar{\mathsf{P}}_{\mathrm{BLER}}(\boldsymbol{G}) = \mathbb{E}[\mathsf{P}_{\mathrm{BLER}}(\boldsymbol{X}|\boldsymbol{G})]$, which is now only dependent

[9]Again, similar results were observed for $(N, K) = (6, 4)$ and $(N, K) = (8, 6)$ systems.

[10]Independence can be assumed by considering the case where a subcarrier-level block interleaver is employed [33].

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

on the channel state captured in $\boldsymbol{G}$. This measure is useful for evaluating performance in a slow-fading environment. It also allows us to observe how channel variations affect error performance.

Following the simulation setting for the mutual information analysis, we configured the channel gains to be $g_l = \eta^{l-1}$ and let $\{\theta_l\}$ be uniformly distributed over $[0, 2\pi)$, $\forall\ l \in \{1, \ldots, N\}$. We also normalized $\sigma^2 = 1$ for the noise power and let $N = 4$ and $K = 2$ as an example. We do not apply rate adaptation in this study; consequently, we employ a uniform power allocation scheme in all simulations related to error performance. This allows us to focus on the effect that binary-tree optimization (i.e., bit-to-SAP optimization) has on performance. We numerically examined the BLER for OFDM-IM with the SAP probability distribution optimized under two conditions. The first condition only requires the number of leaves in the binary tree that defines bit-to-SAP mapping to be equal to or smaller than the number of SAPs. The second condition restricts the encoder to only consider full binary trees with C leaves, which reduces the achievable rate at low SNR. Also, the classic OFDM-IM scheme studied in [9] was adopted as a benchmark. We adopt the lexicographic codebook design for the classic OFDM-IM system to select four out of six SAPs for comparison purposes [16]. The numerical results are presented in Fig. 8, which were obtained by collecting 10^3 block-error events for each SNR point (subject to random additive white Gaussian noise).

In Fig. 8, it is apparent that the rate-optimized OFDM-IM systems designed according to the first condition outperform those designed under the second condition. This behavior correlates with the fact that the first condition is less restrictive than the second. Perhaps more interestingly, we note that the rate-optimized system does not always outperform the classical OFDM-IM scheme. When signals are subject to deep fading, the rate-optimized system designed according to the first condition may only utilize a single SAP consisting of the best K subcarriers, and the system is reduced to OFDM with K active subcarriers. This system is capable of performing better than those that encode information in the index domain as well as signal space, since block errors arising from incorrect SAP decoding do not occur.

VII. Conclusions

In this paper, we provided a thorough treatment of the rate-optimization problem for OFDM-IM systems with channel knowledge at the transmitter. We cast the problem as one of mapping bit sequences to activation patterns, which enabled us to utilize a binary tree formalism for algorithm development and analysis. To this end, we presented new results on full binary trees, both in terms of algorithmic construction and enumeration. We also reported a number of new analytic bounds and asymptotic results related to the relaxed mutual information optimization problem where SAP probabilities can take any values in the interval $[0, 1]$ subject to a sum probability constraint. We then used the results pertaining to the relaxed problem to develop a heuristic algorithm for obtaining a feasible solution to the constrained problem. Numerical results indicate that this solution is nearly optimum (relative to the relaxed upper bound), particularly in the low and high SNR regimes, and the optimized approach is capable of offering a rate advantage over the conventional OFDM-IM benchmark of [9].

A number of open problems remain. First, it is not clear whether an analytic form for the optimal power values exists for all SNR values; only low and high SNR results were reported here. In fact, it is not readily apparent that the mutual information is concave in the power vector $\boldsymbol{\rho}$; hence, a general analytic form may not be forthcoming. As an alternative, it would be preferable to develop an efficient numerical approach to solving the relaxed optimization problem. The use of BCD was briefly discussed here, but further work is needed to determine whether this method would be a viable solution in practice. It is also not known whether the heuristic projection algorithm (Algorithm 2) is, in fact, optimal in some sense. Furthermore, the present work could be placed in a practical context by evaluating the coded BLER of OFDM-IM. In this case, one would need to choose appropriate coding schemes for the *synchronization channel* [34] that arises through the use of variable-length coding [17]. This is a rich area of research, which requires significant attention to make progress. Finally, and more generally, only rate-optimization for uniform sources was considered; it would be fruitful to study the BLER-minimization problem as well as nonuniform sources and non-Gaussian signalling (i.e., finite signal constellations).

Appendix A
Proof of Proposition 1

Define the mapping $w_\ell : \mathcal{T}_{v-1} \to \mathcal{T}_v$ for any integers $v > 1$ and $\ell \geq 0$, such that w_ℓ appends the protograph τ to the left-most available node at height ℓ, relative to the deepest leaf, of a tree in $\mathcal{T}_{v-1}$. Hence, for a tree $t' \in \mathcal{T}_{v-1}$ with maximum depth d, w_ℓ connects two edges to the left-most leaf node in level $d - \ell$ of t'. These edges are, in turn, each connected to a leaf node at depth $d - \ell + 1$. Note that w_0 always maps a tree to a new tree with one additional internal node and one additional leaf, whereas w_ℓ, for $\ell > 0$, will return the empty set if no height-ℓ leaves exist in the tree on which w_ℓ acts. Algorithm 1 applies w_0 and w_1 to each tree in $\mathcal{T}_k$ with every step of the for loop. The following lemma guarantees that w_0 and w_1 generate unique trees.

Lemma 2. *When applied to elements of $\mathcal{T}_{v-1}$, the mappings w_0 and w_1 yield nonisomophic trees.*

Proof: The mapping w_0 is one-to-one. Hence, the image $w_0(\mathcal{T}_{v-1})$ consists of T_{v-1} trees. None of these trees are isomorphic, since no trees in $\mathcal{T}_{v-1}$ are isomorphic. Similarly, where admissible, the mapping w_1 is one-to-one. In the inadmissible case where w_1 returns the empty set, the mapping is many-to-one; but this can be ignored, since no tree is generated. Thus, the image $w_1(\mathcal{T}_{v-1})$ consists of *at most* T_{v-1} trees. Again, none of these trees are isomorphic, since no trees in $\mathcal{T}_{v-1}$ are isomorphic. Furthermore, we deduce that $w_0(\mathcal{T}_{v-1}) \cap w_1(\mathcal{T}_{v-1}) = \{\}$, since every $t \in w_0(\mathcal{T}_{v-1})$ has two leaves at the deepest level and every $t' \in w_1(\mathcal{T}_{v-1})$ has more than two leaves at the deepest level. $\square$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

We now conclude the proof with the following lemma.

Lemma 3. *For $v \geq 2$,*

$$\mathcal{T}_v = w_0(\mathcal{T}_{v-1}) \cup w_1(\mathcal{T}_{v-1}) \tag{21}$$

is a complete reduced set of v-node full binary trees.

Proof: Let $\mathcal{T}_1 = \{\tau\}$. It is easy to verify (cf. Fig. 2) that $\mathcal{T}_2 = w_0(\mathcal{T}_1) \cup w_1(\mathcal{T}_1)$ and $\mathcal{T}_3 = w_0(\mathcal{T}_2) \cup w_1(\mathcal{T}_2)$ are complete sets. Moreover, Lemma 2 ensures $\mathcal{T}_v$ contains no isomorphic trees for $v > 1$. Hence, to prove that $\mathcal{T}_v$ is a complete reduced set of full binary trees for $v \geq 4$, we must show that

$$w_\ell(\mathcal{T}_{v-1}) \subseteq w_0(\mathcal{T}_{v-1}) \cup w_1(\mathcal{T}_{v-1}) \tag{22}$$

for $2 \leq \ell \leq v-1$ and $v \geq 4$.

Assume the lemma is true for all $v = 2, \ldots, k-1$. Choose $t' \in \mathcal{T}_{k-1}$. Consider the mapping $w_\ell(t')$. We treat several possibilities. If the operation maps to the empty set, the set relation is satisfied since by definition (no tree is generated). On the other hand, if $w_\ell(t')$ is nonempty and the deepest level of $w_\ell(t')$ contains exactly two leaves (and hence the same can be said for t' since $\ell \geq 2$), then we must show that there exists a tree $t \in \mathcal{T}_{k-1}$ such that $w_\ell(t') = w_0(t)$. Note that, in this case, w_0 has an inverse, and the composition $w_0^{-1} \circ w_\ell$ commutes. Thus, we write

$$t = w_0^{-1} \circ w_\ell(t') = w_\ell \circ w_0^{-1}(t') = w_\ell(t'') \tag{23}$$

where, in the second and third equalities, it is understood that w_ℓ operates on the level at height ℓ relative to the deepest leaf in t'. But, by the inductive hypothesis and Lemma 2, we have that $t'' = w_0^{-1}(t') \in \mathcal{T}_{k-2}$. It follows that

$$t = w_\ell(t'') \in w_0(\mathcal{T}_{k-2}) \cup w_1(\mathcal{T}_{k-2}) = \mathcal{T}_{k-1} \tag{24}$$

as required.

Now suppose $w_\ell(t')$ is nonempty and the deepest level of $w_\ell(t')$ contains more than two leaves. In this case, we must show that there exists a tree $t \in \mathcal{T}_{k-1}$ such that $w_\ell(t') = w_1(t)$. We take a similar approach, recognizing that w_1 has an inverse, and the composition $w_1^{-1} \circ w_\ell$ commutes. It follows that

$$t = w_1^{-1} \circ w_\ell(t') = w_\ell \circ w_1^{-1}(t') = w_\ell(t'') \tag{25}$$

and, by induction, $t'' = w_1^{-1}(t') \in \mathcal{T}_{k-2}$. Finally, we have that

$$t = w_\ell(t'') \in w_0(\mathcal{T}_{k-2}) \cup w_1(\mathcal{T}_{k-2}) = \mathcal{T}_{k-1} \tag{26}$$

as required. □

Appendix B
Proof of Proposition 2

The proposition can be seen to hold (with equality) for $v = 2, 3, 4$ by explicit construction of $\mathcal{T}_v$. For $v > 4$, consider the mappings $\{w_\ell\}$ given in Appendix A. As noted in the proof of Lemma 2 in that appendix, w_0 is one-to-one. Thus, $|w_0(\mathcal{T}_{v-1})| = T_{v-1}$. Moreover, from Lemma 2 and the definition of $\mathcal{T}_v$ given in Lemma 3, we know that

$$T_v = |w_0(\mathcal{T}_{v-1}) \cup w_1(\mathcal{T}_{v-1})| = T_{v-1} + |w_1(\mathcal{T}_{v-1})|. \tag{27}$$

The set $\mathcal{T}_{v-1}$ can be partitioned into two subsets: one set that contains trees that are mapped to v-node trees in $\mathcal{T}_v$ under w_1 and one set does not admit a mapping under w_1. We call trees in the first subset $\mathcal{T}_{v-1}^o$ *open* trees and trees in the second subset $\mathcal{T}_{v-1}^c$ *closed* trees. Since $\mathcal{T}_{v-1}^o = \mathcal{T}_{v-1} - \mathcal{T}_{v-1}^c$, we have

$$T_v = T_{v-1} + |\mathcal{T}_{v-1}^o| = 2T_{v-1} - |\mathcal{T}_{v-1}^c|. \tag{28}$$

To lower bound $|\mathcal{T}_{v-1}^c|$, we apply the following reasoning. A $(v-1)$-node closed tree is formed by appending a closed subtree of size, say, r internal nodes to a subtree of size $v-1-r$. Each set $\mathcal{T}_r$ with $r = 2^q - 1$ for some positive integer q has exactly one *dense* closed tree, i.e., a tree where every level is fully connected to the pervious and next levels, and the deepest level has 2^q leaves. For every $q \in 2, \ldots, \lfloor \log_2(v-1) \rfloor$, we can enumerate $v - 1 - r = v - 2^q$ closed trees. This is a lower bound, since other combinations of r-node closed subtrees and $(v-1-r)$-node trees exist. Finally, we note that if $v-1$ is one less than a power of two, $\mathcal{T}_{v-1}$ contains a dense subtree, which gives rise to the δ_v parameter stated in the proposition.

Appendix C
Proof of Lemma 1

Referring to (9), the equality constraint is affine and the inequality constraints are convex. From (5), we can write

$$I(\boldsymbol{X}; \boldsymbol{Y}) = -\int f_{\boldsymbol{Y}}(\mathbf{y}) \ln f_{\boldsymbol{Y}}(\mathbf{y}) \prod_{l=1}^{N} \mathrm{d}y_l \, \mathrm{d}y_l^* - n \ln(\pi e \sigma^2). \tag{29}$$

Hence, we must prove that $h(\boldsymbol{Y})$ is concave in $\mathbf{p}$. Let us interpret $f_{\boldsymbol{Y}}$ (cf. (6)) as a function of $\mathbf{p}$ for a fixed $\mathbf{y}$. Then the mapping $f_{\boldsymbol{Y}} : [0,1]^N \to [0, \infty)$ is linear in this context. Furthermore, $h(\boldsymbol{Y}) = \int u(f_{\boldsymbol{Y}}(\mathbf{y})) \, \mathrm{d}\mathbf{y}$, where $u(x) = -x \ln x$ is concave. It is well known that a composition of a concave function and a linear function is concave, and that concavity is preserved under nonnegative weighted integration [35, sec. 3.2]. The weight function here is simply one, so it follows that $h(\boldsymbol{Y})$ is concave in $\mathbf{p}$.

Appendix D
Proof of Proposition 3

Starting from Lemma 1, we form the KKT conditions

$$\begin{aligned} -\mathbf{p} \leq \mathbf{0}, \qquad \mathbf{1}^T \mathbf{p} - 1 = 0, \qquad \boldsymbol{\lambda_0} \geq \mathbf{0}, \\ \operatorname{diag}(\boldsymbol{\lambda}_0)\mathbf{p} = \mathbf{0}, \qquad \nabla L(\mathbf{p}, \boldsymbol{\lambda}_0, \nu) = 0 \end{aligned} \tag{30}$$

where $\boldsymbol{\lambda}_0 \in \mathbb{R}^C$ and $\nu \in \mathbb{R}$ are the Lagrange multipliers. The gradient of the Lagrangian is

$$\nabla L = \nabla h(\boldsymbol{Y}) + \boldsymbol{\lambda}_0 + \nu \mathbf{1} \tag{31}$$

and the gradient of $h(\boldsymbol{Y})$ with respect to $\mathbf{p}$ can be written as

$$\nabla h(\boldsymbol{Y}) = -\int \ln(\boldsymbol{a}(\mathbf{y})^T \mathbf{p}) \boldsymbol{a}(\mathbf{y}) + \boldsymbol{a}(\mathbf{y}) \, \mathrm{d}\mathbf{y} \tag{32}$$

where $\boldsymbol{a}(\mathbf{y}) := (f_{\boldsymbol{Y}|U}(\mathbf{y} \mid U = i))_{i=1,\ldots,C}$. Hence, the optimal vector $\mathbf{p}$ must satisfy

$$\int \ln(\boldsymbol{a}(\mathbf{y})^T \mathbf{p}) \boldsymbol{a}(\mathbf{y}) \, \mathrm{d}\mathbf{y} = \boldsymbol{\lambda}_0 + (\nu - 1)\mathbf{1}. \tag{33}$$

where we have used the fact that $\int \boldsymbol{a}(\mathbf{y}) \, \mathrm{d}\mathbf{y} = \mathbf{1}$. Furthermore, since the problem is concave, any $\mathbf{p}$, $\boldsymbol{\lambda}_0$, and ν that satisfy (33)

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

and the rest of the conditions in (30) are primal and dual optimal, and thus yield the maximum mutual information. Choose $p_i = 1/C$ for all i. In this case, the inequality constraints are inactive, which implies $\boldsymbol{\lambda}_0 = \mathbf{0}$. As a result, (33) is satisfied (along with the rest of the KKT conditions), and thus uniform SAP probabilities is optimal.

Appendix E
Proof of Proposition 4

Using Jensen's inequality and considering the concavity of the logarithm, $h(\boldsymbol{Y})$ can be lower bounded as

$$h(\boldsymbol{Y}) = -\mathbb{E}\left[\ln f_{\boldsymbol{Y}}(\mathbf{y})\right] \geq -\ln \mathbb{E}\left[f_{\boldsymbol{Y}}(\mathbf{y})\right]. \tag{34}$$

The expectation can be evaluated to give

$$\mathbb{E}\left[f_{\boldsymbol{Y}}(\mathbf{y})\right] = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}} \frac{p_i p_j}{\pi^N \det(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_j)} \tag{35}$$

where $\boldsymbol{\Xi}_i$ is a diagonal matrix with the lth element of the diagonal ξ_{li} satisfying (11). Substituting (35) into (34), factoring out $1/\pi^N$, and using (5), we arrive at the bound stated in the proposition.

Appendix F
Proof of Proposition 5

Let $a_{ij} = 1/\det(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_j)$. Proposition 4 can be written as

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &\geq -\ln\left(\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}} a_{ij} p_i p_j\right) - N\ln\left(e\sigma^2\right) \\ &= -\ln(\mathbf{p}^T \boldsymbol{A}\mathbf{p}) - N\ln\left(e\sigma^2\right) \end{aligned} \tag{36}$$

where $\boldsymbol{A} = (a_{ij})$. Thus, maximizing the bound with respect to the activation probabilities (subject to constraints) can be achieved by solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{p}}{\text{minimize}} \quad & \mathbf{p}^T \boldsymbol{A}\mathbf{p} \\ \text{subject to} \quad & \sum_{i\in\mathcal{U}} p_i = 1 \text{ and } p_i \geq 0,\ \forall i \in \mathcal{U}. \end{aligned} \tag{37}$$

When $\boldsymbol{A}$ is nonsingular, this classical quadratic program yields the solution

$$\mathbf{p}^\star = (\nu \boldsymbol{A}^{-1}\mathbf{1})^+ \tag{38}$$

where ν is chosen to satisfy the equality constraint $\sum_{i\in\mathcal{U}} p_i = 1$. Letting b_{ij} denote the ijth element of $\boldsymbol{A}^{-1}$, we have that $p_i^\star = \nu(\sum_{j\in\mathcal{U}} b_{ij})^+$ and $\nu = 1/\sum_{i\in\mathcal{U}}(\sum_{j\in\mathcal{U}} b_{ij})^+$. The result stated in the proposition follows readily.

Appendix G
Proof of Proposition 6

In the following we denote $a_{li} = g_l \rho_{li}$, for all l. From (29), together with (6) and (7), we have

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &= -N\ln(\pi e \sigma^2) \\ &- \sum_{i\in\mathcal{U}} p_i \int \prod_{l=1}^{N} \mathrm{d}y_l\, \mathrm{d}y_l^* \prod_{l\in\mathcal{S}_i} f_{\text{CN}}(y_l; a_{li} + \sigma^2) \prod_{l\notin\mathcal{S}_i} f_{\text{CN}}(y_l; \sigma^2) \\ &\times \ln\left[\sum_{j\in\mathcal{U}} p_j \prod_{m\in\mathcal{S}_j} f_{\text{CN}}(y_m; a_{mj} + \sigma^2) \prod_{m\notin\mathcal{S}_j} f_{\text{CN}}(y_m; \sigma^2)\right] \end{aligned} \tag{39}$$

In the integral corresponding to the ith term above, we make the following change of variables: $y_l \to \sqrt{a_{li} + \sigma^2} y_l$ and $y_l^* \to \sqrt{a_{li} + \sigma^2} y_l^*$, $l \in \mathcal{S}_i$, and $y_l \to \sigma y_l$ and $y_l^* \to \sigma y_l^*$, $l \notin \mathcal{S}_i$, and obtain

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &= -N\ln(\pi e \sigma^2) - \sum_{i\in\mathcal{U}} p_i \int \prod_{l=1}^{N} \mathrm{d}y_l\, \mathrm{d}y_l^* \frac{e^{-|y_l|^2}}{\pi} \\ &\times \ln\left[\sum_{j\in\mathcal{U}} \frac{p_j}{(\pi\sigma^2)^{n-k} \prod_{m\in\mathcal{S}_j} \pi(a_{mj} + \sigma^2)} \prod_{m\in\mathcal{S}_j\setminus\mathcal{S}_i} e^{-\frac{\sigma^2 |y_m|^2}{a_{mj}+\sigma^2}}\right. \\ &\left.\times \prod_{m\in\mathcal{S}_i\setminus\mathcal{S}_j} e^{-\frac{(a_{mi}+\sigma^2)}{\sigma^2}|y_m|^2} \prod_{m\in(\mathcal{S}_j\cap\mathcal{S}_i)\cup(\mathcal{S}_j^c\cap\mathcal{S}_i^c)} e^{-|y_m|^2}\right] \end{aligned} \tag{40}$$

In the argument of the log in (40), we factor out $\frac{1}{(\pi\sigma^2)^N}\prod_{m=1}^{N}\exp\left(-|y_m|^2\right)$, such that, upon taking the log and integrating, eq. (40) becomes

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &= -\sum_{i\in\mathcal{U}} p_i \int \prod_{l=1}^{N} \mathrm{d}y_l\, \mathrm{d}y_l^* \frac{e^{-|y_l|^2}}{\pi} \\ &\times \ln\left[\frac{p_i}{\prod_{m\in\mathcal{S}_i}(1 + a_{mi}/\sigma^2)} + \sum_{j\in\mathcal{U}\setminus\{i\}} \frac{p_j}{\prod_{m\in\mathcal{S}_j}(1 + a_{mj}/\sigma^2)}\right. \\ &\left.\times \prod_{m\in\mathcal{S}_j\setminus\mathcal{S}_i} e^{\frac{a_{mj}|y_m|^2}{a_{mj}+\sigma^2}} \prod_{m\in\mathcal{S}_i\setminus\mathcal{S}_j} e^{-\frac{a_{mi}}{\sigma^2}|y_m|^2}\right] \end{aligned} \tag{41}$$

Now, we define the probability values

$$q_i = \frac{\prod_{m\in\mathcal{S}_i}(a_{mj} + \sigma^2)}{\sum_{j\in\mathcal{U}}\prod_{l\in\mathcal{S}_j}(a_{lj} + \sigma^2)} \tag{42}$$

and obtain

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &= \ln\left[\sum_{i\in\mathcal{U}}\prod_{l\in\mathcal{S}_i}\left(\frac{a_{li}}{\sigma^2} + 1\right)\right] - \int \prod_{l=1}^{N} \mathrm{d}y_l\, \mathrm{d}y_l^* \frac{e^{-|y_l|^2}}{\pi} \\ &\times \sum_{i\in\mathcal{U}} p_i \ln\left[\frac{p_i}{q_i} + \sum_{j\in\mathcal{U}\setminus\{i\}} \frac{p_j}{q_j} \prod_{m\in\mathcal{S}_j\setminus\mathcal{S}_i} e^{\frac{a_{mj}|y_m|^2}{a_{mj}+\sigma^2}}\right. \\ &\left.\times \prod_{m\in\mathcal{S}_i\setminus\mathcal{S}_j} e^{-\frac{a_{mi}}{\sigma^2}|y_m|^2}\right] \end{aligned} \tag{43}$$

We observe that the limit of the ith term of the sum in (43) as $\sigma^2 \to 0$ is $p_i \ln \frac{p_i}{q_i}$ (with the convention $0 \cdot \ln 0 = 0$).[11] Thus, the mutual information obeys the asymptotic equivalence

$$I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) \sim \ln\left[\sum_{i\in\mathcal{U}}\prod_{l\in\mathcal{S}_i}\left(\frac{a_{li}}{\sigma^2} + 1\right)\right] - \sum_{i\in\mathcal{U}} p_i \ln\frac{p_i}{q_i}. \tag{44}$$

Note that $\sum_{i\in\mathcal{U}} p_i \ln \frac{p_i}{q_i}$ is the Kullback-Leibler divergence between p and q, which is always positive and equals zero if and only if the two distributions are identical. Thus, at high SNR ($\sigma^2 \to 0$), the probabilities that maximize the mutual information are $p_i = q_i$, $i \in \mathcal{U}$.

[11] By the dominated convergence theorem, the order of the limit and the integral can be exchanged.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

Appendix H
Proof of Corollary 1

For an arbitrary probability distribution p, define $\mathcal{U}_+ = \{i \in \mathcal{U} \mid p_i > 0\}$. Based on (43), we write

$$\mu(\sigma^2) - I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) = -\sum_{i\in\mathcal{U}_+} p_i \ln \frac{q_i}{p_i} + \int \prod_{l=1}^{N} \mathrm{d}y_l \, \mathrm{d}y_l^* \frac{e^{-|y_l|^2}}{\pi} \times \sum_{i\in\mathcal{U}_+} p_i \ln \left[1 + \sum_{j\in\mathcal{U}\setminus\{i\}} \frac{p_j q_i}{q_j p_i} \prod_{m\in\mathcal{S}_j\setminus\mathcal{S}_i} e^{\frac{a_{mj}|y_m|^2}{a_{mj}+\sigma^2}} \times \prod_{m\in\mathcal{S}_i\setminus\mathcal{S}_j} e^{-\frac{a_{mi}}{\sigma^2}|y_m|^2} \right] \quad (45)$$

Firstly, using Jensen's inequality, we have $-\sum_{i\in\mathcal{U}_+} p_i \ln \frac{q_i}{p_i} \geq -\ln \sum_{i\in\mathcal{U}_+} p_i \frac{q_i}{p_i} = -\ln \sum_{i\in\mathcal{U}_+} q_i \geq 0$. Secondly, the log in the integrand in (45) is positive, as $\ln(1+t) > 0$ for $t > 0$); therefore, the integral is also positive. Thus, $\mu(\sigma^2) > I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2)$, for any $\mathbf{p}$ and σ^2. Consequently, $\mu(\sigma^2) > \max_{\mathbf{p}} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) = I^\star(\boldsymbol{\rho}, \sigma^2)$. Finally, $\lim_{\sigma^2\to 0}[\mu(\sigma^2) - I^\star(\boldsymbol{\rho}, \sigma^2)] = \lim_{\sigma^2\to 0}[\mu(\sigma^2) - I(\mathbf{q}, \boldsymbol{\rho}, \sigma^2)] = 0$, as argued in Proposition 6.

Appendix I
Proof of Proposition 7

Starting from (41), we observe that, as $\sigma^2 \to \infty$, the mutual information obeys the asymptotic equivalence

$$\begin{aligned} I(\mathbf{p}, \boldsymbol{\rho}, \sigma^2) &\sim -\sum_{i\in\mathcal{U}} p_i \ln \left[\sum_{j\in\mathcal{U}} \frac{p_j}{\prod_{m\in\mathcal{S}_j} \left(\frac{a_{mj}}{\sigma^2} + 1\right)} \right] \\ &= -\ln \left[\sum_{i\in\mathcal{U}} \frac{p_i}{\prod_{l\in\mathcal{S}_i} \left(\frac{a_{li}}{\sigma^2} + 1\right)} \right] \\ &\leq -\ln \left[\frac{\sum_{i\in\mathcal{U}} p_i}{\prod_{l\in\mathcal{S}_{i^\star}} \left(\frac{a_{li^\star}}{\sigma^2} + 1\right)} \right] \\ &= \ln \left[\prod_{l\in\mathcal{S}_{i^\star}} \left(\frac{a_{li^\star}}{\sigma^2} + 1 \right) \right] = \sum_{l\in\mathcal{S}_{i^\star}} \ln \left(\frac{a_{li^\star}}{\sigma^2} + 1 \right) \end{aligned}$$

The inequality is achieved when $p_i = r_i$, $i \in \mathcal{U}$.

Appendix J
Proof of Proposition 8

Consider the high-SNR regime. According to Proposition 6 and Corollary 1, the optimal mutual information for given powers satisfies the asymptotic equivalence

$$I^\star(\boldsymbol{\rho}, \sigma^2) \sim \ln \left[\sum_{i\in\mathcal{U}} \prod_{l\in\mathcal{S}_i} \left(\frac{g_l \rho_{li}}{\sigma^2} + 1 \right) \right], \quad \text{as } \sigma^2 \to 0. \quad (46)$$

We now maximize the mutual information over the powers at high SNR by formulating the optimization problem

$$\begin{aligned} \underset{\boldsymbol{\rho}}{\text{maximize}} \quad & \ln \left[\sum_{i\in\mathcal{U}} \exp \sum_{l\in\mathcal{S}_i} \ln \left(\frac{g_l \rho_{li}}{\sigma^2} + 1 \right) \right] \\ & \sum_{l\in\mathcal{S}_i} \rho_{li} \leq P \text{ and } \rho_{li} \geq 0, \quad \forall i \in \mathcal{U}, l \in \mathcal{S}_i. \end{aligned} \quad (47)$$

Given that the objective function is increasing in every variable, the constraints are satisfied with equality. Moreover, the problem actually decouples in $\binom{N}{K}$ separate problems. We cast the ith problem as

$$\begin{aligned} \underset{\boldsymbol{\rho}}{\text{maximize}} \quad & \sum_{l\in\mathcal{S}_i} \ln \left(\frac{g_l \rho_{li}}{\sigma^2} + 1 \right) \\ & \sum_{l\in\mathcal{S}_i} \rho_{li} \leq P \text{ and } \rho_{li} \geq 0. \end{aligned} \quad (48)$$

For each $i \in \mathcal{U}$, the optimal powers are found via the waterfilling strategy.

For the low-SNR regime, from Proposition 7, we optimize $I^\star(\boldsymbol{\rho}, \sigma^2) \sim \sum_{l\in\mathcal{S}_{i^\star}} \ln \left(\frac{a_{li^\star}}{\sigma^2} + 1\right)$ under the constraint $\sum_{l\in\mathcal{S}_i^\star} \rho_{li^\star} = P$. Waterfilling power allocation is optimum.

Acknowledgment

The authors wish to thank Dr. Hachem Yassine and Mr. Jinchuan Tang for their helpful suggestions regarding the proof of Proposition 3 and the structuring of numerical simulations.

References

[1] D. Slepian, "Permutation modulation," *Proceedings of the IEEE*, vol. 53, no. 3, pp. 228–236, Mar 1965.

[2] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16 693–16 746, 2017.

[3] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: From classic RF to visible light communications and data storage," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1905–1938, thirdquarter 2018.

[4] R. Y. Mesleh, H. Haas, S. Sinanović, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, July 2008.

[5] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, Jan 2014.

[6] A. K. Khandani, "Media-based modulation: A new approach to wireless transmission," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 3050–3054.

[7] E. Basar and I. Altunbas, "Space-time channel modulation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7609–7614, Aug 2017.

[8] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (SIM) OFDM," in *IEEE GLOBECOM Workshops (GC Wkshps)*, Dec 2011, pp. 728–732.

[9] E. Başar, U. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, Nov 2013.

[10] A. E. Willner, "Communication with a twist," *IEEE Spectrum*, vol. 53, no. 8, pp. 34–39, August 2016.

[11] Y. Yang, W. Cheng, W. Zhang, and H. Zhang, "Mode modulation for wireless communications with a twist," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 704–10 714, Nov 2018.

[12] Y. Yang and B. Jiao, "Information-guided channel-hopping for high data rate wireless communication," *IEEE Communications Letters*, vol. 12, no. 4, pp. 225–227, April 2008.

[13] M. Maleki, H. R. Bahrami, S. Beygi, M. Kafashan, and N. H. Tran, "Space modulation with CSI: Constellation design and performance evaluation," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1623–1634, May 2013.

[14] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalized spatial modulation," in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, Nov 2010, pp. 1498–1502.

[15] J. Wang, S. Jia, and J. Song, "Generalized spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1605–1615, April 2012.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2019.2914531, IEEE Journal of Selected Topics in Signal Processing

[16] S. Dang, G. Chen, and J. P. Coon, "Lexicographic codebook design for OFDM with index modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8373–8387, Dec 2018.

[17] Y. Liu and J. P. Coon, "Mitigating bit-synchronization errors in Huffman-coding-aided index modulation," accepted by IEEE Communications Letters, Dec 2018.

[18] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the achievable rate of OFDM with index modulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1919–1932, April 2016.

[19] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, Sept 1952.

[20] A. I. Siddiq, "Low complexity OFDM-IM detector by encoding all possible subcarrier activation patterns," *IEEE Communications Letters*, vol. 20, no. 3, pp. 446–449, March 2016.

[21] W. Wang and W. Zhang, "Huffman coding-based adaptive spatial modulation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5090–5101, Aug 2017.

[22] Z. Hu, J. Liu, and F. Chen, "On the mutual information and Huffman coding for OFDM-IM," in *2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, May 2018, pp. 298–302.

[23] M. I. Kadir, H. Zhang, S. Chen, and L. Hanzo, "Entropy coding aided adaptive subcarrier-index modulated OFDM," *IEEE Access*, vol. 6, pp. 7739–7752, 2018.

[24] W. Wang and W. Zhang, "Transmit signal designs for spatial modulation with analog phase shifters," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3059–3070, May 2018.

[25] S. Roman, *Introduction to Coding and Information Theory*, ser. Undergraduate Texts in Mathematics. Springer New York, 1996.

[26] F. Murtagh, "Counting dendrograms: a survey," *Discrete Applied Mathematics*, vol. 7, no. 2, pp. 191–199, 1984.

[27] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific Belmont, 1999.

[28] Z. B. Zabinsky, *Random Search Algorithms*. American Cancer Society, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470400531.eorms0704

[29] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*. IEEE, 2008, pp. 181–188.

[30] S. Dang, J. P. Coon, and G. Chen, "Adaptive OFDM with index modulation for two-hop relay-assisted networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1923–1936, Mar. 2018.

[31] Z. Wang, S. Dang, and D. T. Kennedy, "Multi-hop index modulation-aided OFDM with decode-and-forward relaying," *IEEE Access*, vol. 6, pp. 26 457–26 468, 2018.

[32] J. Zhao, S. Dang, and Z. Wang, "Full-duplex relay-assisted orthogonal frequency-division multiplexing with index modulation," *IEEE Systems Journal*, pp. 1–12, 2018.

[33] Y. Xiao, S. Wang, L. Dan, X. Lei, P. Yang, and W. Xiang, "OFDM with interleaved subcarrier-index modulation," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1447–1450, 2014.

[34] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, vol. 6, pp. 1–33, 2009.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

Justin P. Coon (S'02 – M'05 – SM'10) received the B.Sc. degree (Hons.) in electrical engineering from the Calhoun Honours College, Clemson University, USA, and the Ph.D. degree in communications from the University of Bristol, U.K., in 2000 and 2005, respectively. He was with Toshiba Research Europe Ltd. (TREL) from 2004 until 2013. Prof. Coon was also a Visiting Fellow with the School of Mathematics at the University of Bristol from 2010 until 2012, and held a concurrent appointment as Reader in the Department of Electrical and Electronic Engineering at the same university from 2012 to 2013. Since 2013, he has been an Associate Professor with the Department of Engineering Science at the University of Oxford and the Emmott Fellow in Engineering at Oriel College, where he teaches and conducts research on communication theory, network theory, and signal processing for communications. Prof. Coon has published over 170 papers on these topics. He is an active member of the IEEE, having served as an editor for several journals. Prof. Coon is also a Fellow of the Institute of Mathematics and its Applications.

Mihai-Alin Badiu received the Dipl.-Ing., M.S. and Ph.D. degrees in electrical engineering from the Technical University of Cluj-Napoca, Romania, in 2008, 2010 and 2012, respectively. He is currently a Postdoctoral Research Assistant in the Department of Engineering Science at the University of Oxford, UK. From 2012 to 2019, he held positions as Postdoctoral Researcher and Assistant Professor in the Department of Electronic Systems at Aalborg University, Denmark. In 2018, he was a part-time PDRA at University of Oxford. From 2016 to 2018, he held an Individual Postdoctoral Fellowship from the Danish Council for Independent Research. His research interests are in the areas of wireless network theory, signal processing for communications, and probabilistic modelling and inference.

Ye Liu received the B. Eng. degree (Hons.) in electronic and communication engineering and the Ph.D. degree in electronic engineering from the City University of Hong Kong in 2009 and 2014, respectively. He was a Visiting Student at the Laboratory of Information, Networking and Communication Sciences, Paris, France. He was also an Intern at Nokia Bell Labs, France, and was a Post-Doctoral Research Associate with the Signal Processing and Networks Research Group at Loughborough University. He is currently a Post-Doctoral Research Assistant with the Department of Engineering Science at the University of Oxford. His research interests include index modulation, non-orthogonal multiple access, resource allocation, and network coding.

Ferhat Yarkin (S'16) received the B.Sc. and M.Sc. degrees (Hons.), in electronics and communication engineering, from the Istanbul Technical University, Istanbul, Turkey, in 2015 and 2017, respectively. From 2016 to 2018, he was a Research and Teaching Assistant and a member of the Wireless Communication Research Group, Istanbul Technical University. He is currently a member of the Oxford Communications Research Group in the Department of Engineering Science at the University of Oxford, where he is working toward a D.Phil degree. His research interests include multiple-input, multiple-output systems; cooperative communication; spatial modulation; and index modulation. Mr. Yarkin has been a reviewer for several IEEE and Elsevier journals.

Shuping Dang (S'13 – M'18) received B.Eng (Hons) in Electrical and Electronic Engineering from the University of Manchester (with first class honors) and B.Eng in Electrical Engineering and Automation from Beijing Jiaotong University in 2014 via a joint '2+2' dual-degree program. He also received D.Phil in Engineering Science from University of Oxford in 2018. Dr. Dang joined in the R&D Center, Huanan Communication Co., Ltd. after graduating from University of Oxford and is currently working as a Postdoctoral Fellow with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST). He serves as a reviewer for a number of key journals in communications and information science, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. His current research interests include artificial intelligence assisted communications, novel modulation schemes and cooperative communications.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.