A Recursive Construction of the Set of Binary Entropy Vectors

John MacLaren Walsh & Steven Weber
Department of Electrical and Computer Engineering
Drexel University, Philadelphia, PA 19104
Email: jwalsh@ece.drexel.edu and sweber@ece.drexel.edu

***Abstract*—The primary contribution is a finite terminating algorithm that determines membership of a candidate entropy vector in the set of binary entropy vectors Φ_N. We outline the relationship between Φ_N and its unbounded cardinality discrete random variable counterpart $\bar{\Gamma}^*_N$ (or its normalization $\bar{\Omega}^*_N$). We discuss connections between Φ_N and $\bar{\Omega}^*_N$. For example, for any outer bound, say the Shannon outer bound $\mathcal{P}_N$, to $\bar{\Omega}^*_N$, we provide a finite terminating algorithm to find a polytopic inner bound on $\bar{\Omega}^*_N$ that agrees on tight faces of the outer bound.**

I. Introduction

Recent work in multiterminal information theory has shown that many important open problems (such as the rate regions for multiterminal source coding and the network coding capacity region) can be solved if an efficient description of the set of entropy vectors under various distribution constraints can be obtained [1], [2]. Two related approaches have emerged to do this. One [1] directly studies $\bar{\Gamma}^*_N$, the closure of the set of entropy vectors for discrete unbounded cardinality random variables. Another [2], [3] studies a normalized counterpart $\bar{\Omega}^*_N$. Both of these regions are difficult to characterize for arbitrary N, as Matùš recently definitively proved [4], and there does not presently exist a finite terminating algorithm to verify membership in them for $N \geq 4$. This paper points out that by restricting the discrete random variables to be binary, one can obtain efficient descriptions of the entropy region, called the set of binary entropy vectors Φ_N, with a finite terminating algorithm that operates recursively in N. Possible applications of this new characterization of Φ_N involving inner bounds to $\bar{\Gamma}^*_N$ and $\bar{\Omega}^*_N$ which agree on tight faces with outer bounds to these sets are also discussed.

II. The Set of Binary Entropy Vectors Φ_N

We review the definition of the set of entropy vectors $\bar{\Gamma}^*_N$. Consider all subsets $\mathbf{X}_\mathcal{A} := [X_i | i \in \mathcal{A}]$ of N discrete random variables $\mathbf{X} := [X_1, \ldots, X_N]$, and stack the $2^N - 1$ entropies of each nonempty subset into a vector $\mathbf{h} := [H(\mathbf{X}_\mathcal{A}) \, | \mathcal{A} \subseteq \{1, \ldots, N\}]$, called an entropy vector. The entropy vector $\mathbf{h} = \mathbf{h}(p_\mathbf{X})$ is clearly a function of the joint distribution $p_\mathbf{X}$ on the N discrete random variables $\mathbf{X}$[1]. Define the set of possible entropy vectors $\bar{\Gamma}^*_N := \text{cl } \mathbf{h}(\mathcal{D})$ as the closure of the image under the function $\mathbf{h}$ of the set $\mathcal{D}$ of viable joint probability mass functions $p_\mathbf{X}$ (i.e., those which are non-negative and sum to one). The set Φ_N is then easily described as the subset of $\bar{\Gamma}^*_N$ obtained by requiring that the random variables X_i be binary. Specifically, defining $\mathcal{D}_{\text{bin}} \subset \mathcal{D}$ to be the set of joint distributions among N bits $X_1, \ldots, X_N$, then $\Phi_N := \mathbf{h}(\mathcal{D}_{\text{bin}})$. The version of this set for just two bits Φ_2 is plotted in Fig. 1.

A related concept of limiting the cardinality was introduced in [2], [3] by introducing the normalized entropy function $\mathbf{h}^o_M$ defined on the domain $\mathcal{D}_M \subset \mathcal{D}$ of probability mass functions on N random variables each with cardinality M, where the normalized entropy of the subset $\mathcal{A}$ random variables is given by $\frac{1}{\log_2(M)} H(X_\mathcal{A})$. For a valid joint PMF $p_\mathbf{X}$ in $\mathcal{D}_M$, the normalized entropy vector $\mathbf{h}^o_M(p_\mathbf{X})$ is obtained by stacking these normalized entropies for each subset $\mathcal{A}$. The set of normalized entropy vectors Ω^*_N is then obtained as union over all cardinalities M of the image of $\mathcal{D}_M$ under $\mathbf{h}^o_M$

$$\Omega^*_N := \bigcup_{M=2}^{\infty} \mathbf{h}^o_M(\mathcal{D}_M)$$

[1]Entropies are denoted both as functions of rvs ($H(X)$) and functions of distributions ($H(p_X)$).

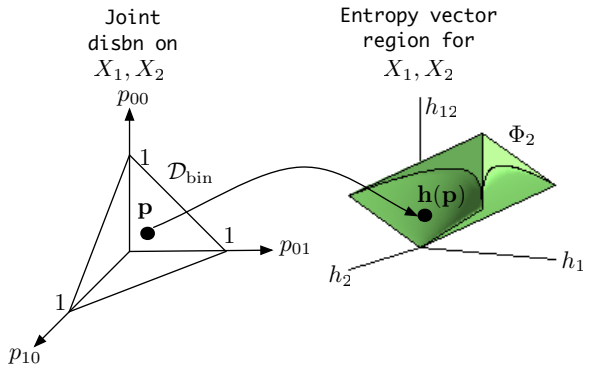

Fig. 1. The entropy vector region Φ_N for $n = 2$ bits is the set of vectors $\mathbf{h} = (h_1 = H(X_1), h_2 = H(X_2), h_{12} = H(X_1, X_2))$ for each joint distribution $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11}) \in \mathcal{D}_{\text{bin}}$ on (X_1, X_2).

Here we can recognize the set of binary entropy vectors Φ_N as the first set in this union (i.e., for $M = 2$). The sets $\Phi_N, \bar{\Gamma}^*_N, \bar{\Omega}^*_N$ are all related. In the next section we justify our working with Φ_N instead of the more traditional $\bar{\Gamma}^*_N$ or its normalized counterpart $\bar{\Omega}^*_N$. We describe properties of Φ_2 in §IV, and then give a simple description of Φ_N in §V.

III. Why Restrict To Binary: Why Φ_N and not $\bar{\Gamma}^*_N$ or $\bar{\Omega}^*_N$?

While it may appear from Fig. 1 that the set Φ_N is far more complex than $\bar{\Gamma}^*_N$ (which is a convex cone), there are in fact many reasons for restricting our attention to Φ_N. To begin with, many problems, including network coding rate regions and many source coding problems, involve finite cardinality intermediate (dummy) random variables, for which Φ_N is naturally the proper choice, and the unlimited cardinality case does not fit. What's more, Φ_N is closed and bounded and thus needs no topological closure to be taken, while $\bar{\Gamma}^*_N$ is not bounded and does require a closure to be taken and Ω^*_N requires a topological closure. Also, as we will shown in §V, we can obtain a recursive in N exact functional description for the boundaries of Φ_N for any N. We can not presently do this with $\bar{\Gamma}^*_N$ or $\bar{\Omega}^*_N$, or it has not yet been done. Even more exciting, as we shall discuss in detail in §VI, is the fact that for $N \geq 4$ it may turn out that convΦ_N is a polytope, while $\bar{\Gamma}^*_N$ has recently been proven not be a polyhedron [4]. Thus, expressions of rate regions in terms of convΦ_N could be much more easy to work with and determine. Even if Φ_N turns out not to be polytope for $N \geq 4$, we will see in §VI that at least for $N = 2, 3$ (and perhaps for other N) that $i)$ conv$\Phi_N = \bar{\Omega}^*_N$ (for conv the convex hull) and $ii)$ cone$\Phi_N = \bar{\Gamma}^*_N$ (for cone the conic hull). For $N \geq 4$ convΦ_N is an inner bound to $\bar{\Omega}^*_N$ (and may be tight) and coneΦ_N is an inner bound to $\bar{\Gamma}^*_N$ (and may be tight), and easily expressible inner bounds for these sets are hard to come by [3]. As we briefly discuss in §VI-A, because we can determine membership in Φ_N with a finite terminating algorithm, we can use it to generate inner bounds to $\bar{\Gamma}^*_N$ and $\bar{\Omega}^*_N$ which agree on tight faces of any specified polyhedral outer bound.

IV. Properties of Φ_2

We first demonstrate how to obtain Φ_2. Suppose we are given $H(X_1) = h_1$ and $H(X_2) = h_2$. We know from a basic Shannon inequality that the maximum value for $H(X_1, X_2)$ is obtained when the two bits are independent of one another. This gives the top surface of Φ_2 as depicted in Fig. 2. Because the entropy of a bit as a function of p (the probability the bit is one) is symmetric about $p = \frac{1}{2}$, we know the marginal probabilities $p_i := \mathbb{P}[X_i = 1]$ up to a two fold ambiguity. Also, given a particular pair of such marginal probabilities for the two bits, the set of joint pmfs $p(X_1, X_2)$ having these two marginal distributions is a line segment as depicted in Fig. 3. This is because the two marginal distribution constraints together with the fact that the distribution must sum to one, form three linearly independent linear constraints on the distribution space, which is 4 dimensional, leaving one dimension left for the space of solutions. The set of joint distributions having $H(X_1) = h_1$ and $H(X_2) = h_2$ may then be written as the union of four line segments, one for each pair of bitwise marginal ambiguities. Along each one of these ambiguities, the joint entropy $H(X_1, X_2)$ as a function of λ (which parameterizes the line segment in joint distribution space having these marginals) appears as in the left of Fig. 4 because it is concave . Thus, given h_1 and h_2, the joint entropy is minimized at one of the two ends of the line segment in joint distribution space. This series of steps is detailed in Figs 2, 3, and 4 below.

V. Explicit Recursive Construction of Φ_N

The figures above show that given the joint entropy and marginal entropies for two bits, we can determine all possible joint two bit distributions having these entropies. By extension, this means that given the marginal and pairwise entropies among three bits, we can can determine the set of possibilities for the pairwise distributions. Each of these possible collection of three pairwise distributions then yields a line segment of possible solutions for the joint distribution on the three bits, whose range of joint entropies can be computed.

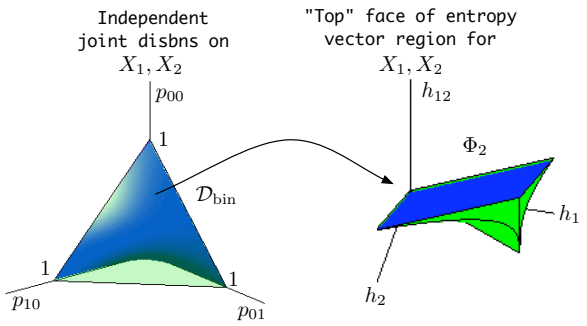

Fig. 2. Entropy is maximized for X_1, X_2 independent. The independent distributions form a curved, non-convex subspace of the set of joint distributions $\mathcal{D}_{\text{bin}}$, shown in blue on the left. This subspace maps to the top face of the binary entropy vector region Φ_2, shown in blue on the right.

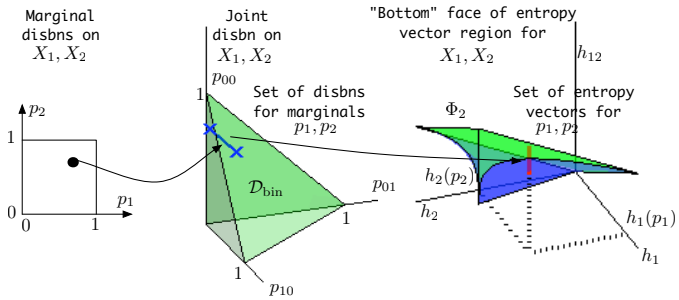

Fig. 3. Each pair of marginals maps to a line segment in the joint distribution space $\mathcal{D}_{\text{bin}}$, shown in blue in the center. This line punctures the surface of $\mathcal{D}_{\text{bin}}$ at two points, and maps to a vertical line segment in Φ_2, shown in red on the right, at $h_1(p_1), h_2(p_2)$. The minimum joint entropy h_{12} for fixed marginal entropies (the bottom of this red segment) corresponds to one of the two puncture points on $\mathcal{D}_{\text{bin}}$. This gives the the bottom face of Φ_2, shown in purple on right.

Thus, the technique shown here to characterize Φ_2 can be used in a recursive manner in N to obtain Φ_N, as we presently shall describe in more detail. We begin by establishing two equivalent representations for the joint distribution of a collection of binary random variables.

Lemma 1: Let $\mathbf{X} = (X_1, \ldots, X_m)$ be a vector of m binary random variables. Denote the joint distribution as $\mathbf{p}_{[m]} = (p_{[m]}(\mathbf{x}),\ \mathbf{x} \in \{0,1\}^m)$. Let $\mathbf{q}_{[m]} = (p_{\mathcal{A}}(\mathbf{0}),\ \mathcal{A} \subseteq [m])$ be the set of probabilities that each subset of random variables $\mathbf{X}_{\mathcal{A}} = (X_i,\ i \in \mathcal{A})$ each take value zero, i.e., $p_{\mathcal{A}}(\mathbf{0}) = \mathbb{P}(X_i = 0,\ i \in \mathcal{A})$. By convention, we take $p_{\emptyset}(\mathbf{0}) = 1$. Then $\mathbf{p}_{[m]}, \mathbf{q}_{[m]}$ are equivalent in that there is a one to one mapping between them.

The proof of Lemma 1 in Appendix A formalizes the construction of this matrix and its inverse for all N.

Equipped with these two parameterizations of the joint distribution on N binary random variables, we can introduce a recursion in N which can characterize with a finite terminating algorithm membership in Φ_N, as formalized in the following theorem.

Theorem 1: Fix $N \in \mathbb{N}$ and let $\mathbf{h} = (h_{\mathcal{A}}, \mathcal{A} \subseteq [N], \mathcal{A} \neq \emptyset) \in \mathbb{R}^{2^N-1}$ be a vector. Algorithm 1 determines whether or not $\mathbf{h} \in \Phi_N$ in finite time, and if so, it returns those joint distributions $\mathcal{D}_N^{\text{bin}}(\mathbf{h})$ such that $\mathcal{H}(\mathbf{p}) = \mathbf{h}$ for each $\mathbf{p} \in \mathcal{D}_N^{\text{bin}}(\mathbf{h})$, or will return that no such distribution exists.

Before proceeding with the proof of Theorem 1, we find it instructive to build intuition about the presented algorithm by demonstrating its application to the set Φ_2 discussed in the previous section.

Example: Algorithm 1 for $N = 2$. Let $N = 2$ and $\mathbf{h} = (h_1, h_2, h_{12}) \in \mathbb{R}^3$ be an arbitrary vector. We apply Algorithm 1 to determine whether or not $\mathbf{h} \in \Phi_2$. We consider the nonempty subsets of $\{1,2\}$ of size $k = 1$, i.e., $\mathcal{A} = \{1\}$ and $\mathcal{A} = \{2\}$. The sets $\tilde{\mathcal{Q}}_1$ and $\tilde{\mathcal{Q}}_2$ are both empty. We define variables $p_1(0)$ and $p_2(0)$ and matrices $\mathbf{M}_1^{-1} = \mathbf{M}_2^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ which yield the marginal

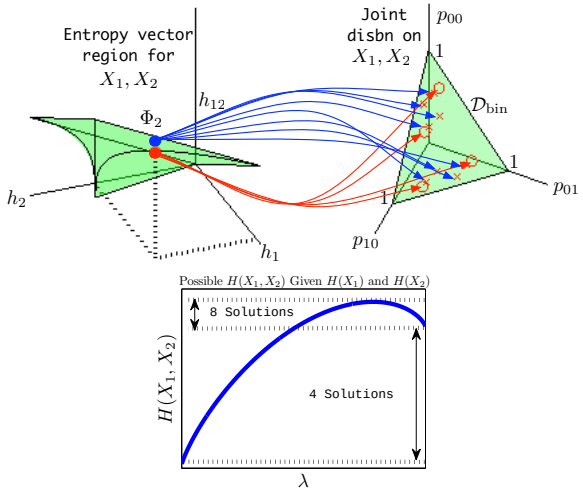

Fig. 4. The inverse map from a particular entropy vector $\mathbf{h} = (h_1, h_2, h_{12})$ to the joint distributions $\mathbf{p}$ with those entropies. The number of joint distributions for a given entropy vector is either 4 (red, bottom) or 8 (blue, top) depending upon the joint entropy, due to the shape of the joint entropy for particular marginals p_1, p_2 graphed to the right (λ parameterizes line segment in $\mathcal{D}_{\text{bin}}$ having these marginals).

Algorithm 1 Determine whether $\mathbf{h} \in \Phi_N$.

Require: $N > 1$ and $\mathbf{h} = (h_{\mathcal{A}},\ \mathcal{A} \subset [N],\ \mathcal{A} \neq \emptyset) \in \mathbb{R}^{2^N-1}$

1: Initialize $\mathcal{Q}_\emptyset = \emptyset$
2: **for** $k = 1 \ldots N$ **do**
3: **for all** $\mathcal{A} \subseteq [N]$, $|\mathcal{A}| = k$ **do**
4: Initialize $\tilde{\mathcal{Q}}_\mathcal{A} = \emptyset$
5: **for all** $(\mathbf{q}_\mathcal{B},\ \mathcal{B} \subseteq \mathcal{A},\ |\mathcal{B}| = k-1) \in \prod_{\mathcal{B}\subset\mathcal{A},|\mathcal{B}|=k-1} \mathcal{Q}_\mathcal{B}$ **do**
6: **if** $q_{\mathcal{B},i}(\mathbf{0}) = q_{\mathcal{B}',i}(\mathbf{0})$ for each $i \in \mathcal{B} \cap \mathcal{B}'$, for each $\mathcal{B}, \mathcal{B}' \subset \mathcal{A}$ with $|\mathcal{B}| = |\mathcal{B}'| = k-1$ **then**
7: Add $\tilde{\mathbf{q}} = \bigcup_{\mathcal{B}\subset\mathcal{A},|\mathcal{B}|=k-1} \mathbf{q}_\mathcal{B}$ to $\tilde{\mathcal{Q}}_\mathcal{A}$
8: **end if**
9: **end for**
10: Initialize $\mathcal{Q}_\mathcal{A} = \emptyset$
11: **for all** $\tilde{\mathbf{q}}_\mathcal{A} \in \tilde{\mathcal{Q}}_\mathcal{A}$ **do**
12: Set $\mathbf{q}_\mathcal{A}(p_\mathcal{A}(\mathbf{0})) = (\tilde{\mathbf{q}}_\mathcal{A}, p_\mathcal{A}(\mathbf{0}), 1)$
13: Set $\mathbf{p}_\mathcal{A}(p_\mathcal{A}(\mathbf{0})) = \mathbf{M}_\mathcal{A}^{-1}\mathbf{q}_\mathcal{A}(p_\mathcal{A}(\mathbf{0}))$
14: Set $\mathcal{P}_\mathcal{A} = \{p_\mathcal{A}(\mathbf{0}) \in [0,1] : \mathcal{H}(\mathbf{p}_\mathcal{A}(p_\mathcal{A}(\mathbf{0}))) = h_\mathcal{A}\}$
15: Add $\mathbf{q}_\mathcal{A} = (\tilde{\mathbf{q}}_\mathcal{A}, p_\mathcal{A}(\mathbf{0}))$ to $\mathcal{Q}_\mathcal{A}$ for each $p_\mathcal{A}(\mathbf{0}) \in \mathcal{P}_\mathcal{A}$, if any.
16: **end for**
17: **end for**
18: **end for**
19: **if** $\mathcal{Q}_{[N]} = \emptyset$ **then**
20: Return $\mathbf{h} \notin \Phi_N$
21: **else**
22: Return $\mathbf{h} \in \Phi_N$ and $\mathcal{D}_N^{\rm bin} = (\mathbf{p}_{[N]} = \mathbf{M}_{[N]}^{-1}\mathbf{q}_{[N]},\ \mathbf{q}_{[N]} \in \mathcal{Q}_{[N]})$
23: **end if**

distributions

$$\begin{aligned}\mathbf{p}_1(p_1(0)) &= \begin{bmatrix} p_1(0) \\ 1-p_1(0)\end{bmatrix} = \mathbf{M}_1^{-1}\mathbf{q}_1(p_1(0)) \\ \mathbf{p}_2(p_2(0)) &= \begin{bmatrix} p_2(0) \\ 1-p_2(0)\end{bmatrix} = \mathbf{M}_2^{-1}\mathbf{q}_2(p_2(0)) \quad (1)\end{aligned}$$

Next, we compute the solution sets

$$\begin{aligned}\mathcal{P}_1 &= \{p_1(0) \in [0,1] : \mathcal{H}(\mathbf{p}_1(p_1(0))) = h_1\} \\ \mathcal{P}_2 &= \{p_2(0) \in [0,1] : \mathcal{H}(\mathbf{p}_2(p_2(0))) = h_2\}. \quad (2)\end{aligned}$$

It is clear that if $h_i \in [0,1]$ then $\mathcal{P}_i$ will contain two points, else $\mathcal{P}_i = \emptyset$, for $i = 1, 2$. Suppose $h_i \in [0,1]$ for $i = 1, 2$ and define $\mathcal{P}_1 = \mathcal{Q}_1 = \{p_1^{(1)}(0), p_1^{(2)}(0)\}$ and $\mathcal{P}_2 = \mathcal{Q}_2 = \{p_2^{(1)}(0), p_2^{(2)}(0)\}$. Next let $k = N = 2$, and note $\mathcal{A} = \{1, 2\}$ is the sole size-2 subset of $\{1, 2\}$. Form

$$\mathcal{Q}_1 \times \mathcal{Q}_2 = \left\{ \begin{array}{c} (p_1^{(1)}(0), p_2^{(1)}(0)), (p_1^{(1)}(0), p_2^{(2)}(0)), \\ (p_1^{(2)}(0), p_2^{(1)}(0)), (p_1^{(2)}(0), p_2^{(2)}(0)) \end{array} \right\}. \quad (3)$$

Each of these four pairs trivially satisfies the condition in line 6 since the sets $\mathcal{B} = \{1\}, \mathcal{B} = \{2\}$ are disjoint. Thus $\tilde{\mathcal{Q}}_{1,2}$ includes all four pairs above. Form the four joint distributions expressed as a function of the variable $p_{12}(00)$ is given by $\mathbf{p}_{12}^{(i,j)}(p_{12}(00)) =$

$$\begin{bmatrix} p_{12}^{(i,j)}(00) \\ p_{12}^{(i,j)}(01) \\ p_{12}^{(i,j)}(10) \\ p_{12}^{(i,j)}(11) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_1^{(i)}(0) \\ p_2^{(j)}(0) \\ p_{12}(00) \\ 1 \end{bmatrix}, \quad (4)$$

with $i, j = 1, 2$. Next, compute the solution sets for each of the four pairs above:

$$\mathcal{P}_{1,2}^{(i,j)} = \{p_{12}(00) \in [0,1] : \mathcal{H}(\mathbf{p}_{12}^{(i,j)}(p_{12}(00))) = h_{12}\}, \quad (5)$$

for $i, j = 1, 2$. Each of the four pairs will have a solution set of size zero, one, or two, depending upon h_{12}, but they will each have the same cardinality, so there are a total of zero, four, or eight joint distributions. For simplicity, suppose there are two solutions for each pair, yielding $\mathcal{Q}_{12}$ as

$$\left\{ \begin{array}{rcl} q_{12}^{(1,1,1)} &=& (p_1^{(1)}(0),\ p_2^{(1)}(0),\ p_{12}^{(1,1,1)}(00)) \\ q_{12}^{(1,1,2)} &=& (p_1^{(1)}(0),\ p_2^{(1)}(0),\ p_{12}^{(1,1,2)}(00)) \\ q_{12}^{(1,2,1)} &=& (p_1^{(1)}(0),\ p_2^{(2)}(0),\ p_{12}^{(1,2,1)}(00)) \\ q_{12}^{(1,2,2)} &=& (p_1^{(1)}(0),\ p_2^{(2)}(0),\ p_{12}^{(1,2,2)}(00)) \\ q_{12}^{(2,1,1)} &=& (p_1^{(2)}(0),\ p_2^{(1)}(0),\ p_{12}^{(2,1,1)}(00)) \\ q_{12}^{(2,1,2)} &=& (p_1^{(2)}(0),\ p_2^{(1)}(0),\ p_{12}^{(2,1,2)}(00)) \\ q_{12}^{(2,2,1)} &=& (p_1^{(2)}(0),\ p_2^{(2)}(0),\ p_{12}^{(2,2,1)}(00)) \\ q_{12}^{(2,2,2)} &=& (p_1^{(2)}(0),\ p_2^{(2)}(0),\ p_{12}^{(2,2,2)}(00)) \end{array} \right\}. \quad (6)$$

We return $\mathcal{D}_2^{\rm bin}(\mathbf{h}) = \left\{\mathbf{p}_{12} = \mathbf{M}_{12}^{-1}\mathbf{q}_{12},\ \mathbf{q} \in \mathcal{Q}_{12}\right\}$, using $\mathbf{M}_{12}^{-1}$ in (4). ♢

Having illustrated how the algorithm may be applied to determine membership in the set Φ_2, we presently prove that the result provided by it for Φ_N is correct for any number of bits N.

Proof of Theorem 1.: Algorithm 1 determines whether or not $\mathbf{h} \in \Phi_N$ in finite time, and if so, it returns those joint distributions $\mathcal{D}_N^{\rm bin}(\mathbf{h})$ such that $\mathcal{H}(\mathbf{p}) = \mathbf{h}$ for each $\mathbf{p} \in \mathcal{D}_N^{\rm bin}(\mathbf{h})$. The general idea is to build up a set of candidate marginal distributions, one for each subset $\mathcal{A}$ of size k, and use these to construct a set of

candidate marginals for each subset $\mathcal{A}$ of size $k+1$, until $k = N$. The algorithm is "complete" in that i) if $\mathbf{p}$ is such that $\mathcal{H}(\mathbf{p}) = \mathbf{h}$ then the algorithm will include $\mathbf{p} \in \mathcal{D}_N^{\text{bin}}$, and "correct" in that ii) if $\mathbf{p} \in \mathcal{D}_N^{\text{bin}}$ then $\mathcal{H}(\mathbf{p}) = \mathbf{h}$. To see part i) (completeness), note that the algorithm will begin with a candidate set of marginal 1-way distributions $\mathbf{q}_i$ such that $\mathcal{H}(\mathbf{p}_i) = h_i$ for each $i = 1, \ldots, N$, and will use these to build up the 2-way marginals such that $\mathcal{H}(\mathbf{p}_\mathcal{A}) = h_\mathcal{A}$ for each $\mathcal{A} \subseteq [N]$ with $|\mathcal{A}| = 2$, and so on. At each step, the algorithm is guaranteed to find and retain each feasible marginal, which means each possible joint distribution $\mathbf{p}$ will be found. To see part ii) (correctness), note that at each step the algorithm only maintains a candidate marginal distribution if it has the correct entropy for each of its respective marginals. It follows that the set of final joint distributions each have the correct required entropy vector.

Before proceeding we define compatibility for a set of marginal distributions. Consider a set of m rvs $(X_1, \ldots, X_m)$ and a set of k subsets of $[m]$, say $(\mathcal{A}_1, \ldots, A_k)$ with each $A_k \subset [m]$. A set of marginal distributions $(\mathbf{p}_{\mathcal{A}_1}, \ldots, \mathbf{p}_{\mathcal{A}_k})$ is compatible if there exists a joint distribution $\mathbf{p}_{[m]}$ with those marginals:

$$\sum_{\mathbf{x} \in \{0,1\}^m : \mathbf{x}_{\mathcal{A}_i} = \mathbf{y}} p_{[m]}(\mathbf{x}) = p_{\mathcal{A}_i}(\mathbf{y}), \ \mathbf{y} \in \{0,1\}^{|\mathcal{A}_i|}, \quad (7)$$

for $i = 1, \ldots, k$. We focus on the case when each of the subsets $\mathcal{A}_i$ has the same cardinality, i.e., $|A_i| = d$ for $i = 1, \ldots, k$ for some integer $d < m$. Using the equivalent representation $(\mathbf{q}_{\mathcal{A}_1}, \ldots, \mathbf{q}_{\mathcal{A}_k})$, a necessary *but not sufficient* condition for the existence of a joint $\mathbf{p}_{[m]}$ satisfying (7) is that the marginals must agree on all common sub-words:

$$q_{\mathcal{A}_i, l}(\mathbf{0}) = q_{\mathcal{A}_j, l}(\mathbf{0}), \ l \in \mathcal{A}_i \cap \mathcal{A}_j, \ i, j \in \{1, \ldots, k\}. \quad (8)$$

As an example to highlight this necessary condition for compatibility, consider lines 5-9 where we form the candidate set of marginal distributions for some $\mathcal{A} \subseteq [N]$ of size $|\mathcal{A}| = k$. The largest possible set is formed by taking the cartesian product of the candidate marginal distributions for each $\mathcal{B} \subset \mathcal{A}$ of size $|\mathcal{B}| = k-1$, i.e., $\prod_{\mathcal{B} \subset \mathcal{A}, |\mathcal{B}| = k-1} \mathcal{Q}_\mathcal{B}$. This set is too large in that it contains combinations of $k-1$ way marginals that are incompatible. An ordered set $(\mathbf{q}_\mathcal{B}, \mathcal{B} \subseteq \mathcal{A}, |\mathcal{B}| = k-1)$ can only be compatible if each pair of elements, $\mathbf{q}_\mathcal{B}, \mathbf{q}_{\mathcal{B}'}$ agrees on its common sub-elements. That is, we must have $q_{\mathcal{B}, i}(\mathbf{0}) = q_{\mathcal{B}', i}(\mathbf{0})$ for each $i \in \mathcal{B} \cap \mathcal{B}'$. As an example, consider $N = 3$ and $\mathcal{A} = [N]$, and the collection

$$\left(\begin{array}{l} \mathbf{q}_{12} = (p_1^{(1)}(0), p_2^{(1)}(0), p_{12}^{(1)}(00)), \\ \mathbf{q}_{13} = (p_1^{(1)}(0), p_3^{(2)}(0), p_{13}^{(2)}(00)), \\ \mathbf{q}_{23} = (p_2^{(1)}(0), p_3^{(1)}(0), p_{23}^{(1)}(00)) \end{array} \right). \quad (9)$$

Then the pair $\mathbf{q}_{12}, \mathbf{q}_{13}$ could be compatible because they agree on their common element, $p_1^{(1)}(0)$, as can be $\mathbf{q}_{12}, \mathbf{q}_{23}$ as they agree on their common element, $p_2^{(1)}(0)$. The pair $\mathbf{q}_{13}, \mathbf{q}_{23}$ are definitely incompatible, however, as the former uses $p_3^{(2)}(0)$ while the latter uses $p_3^{(1)}(0)$. Thus the collection overall is incompatible. Thus the set $\tilde{\mathcal{Q}}_\mathcal{A}$ is the set of combinations of marginals of $\mathcal{A}$ of order $k-1$ which agree on their common elements.

Of course, the condition that all $k-1$-bit distributions give marginal distributions which agree on overlapping subsets is necessary, but insufficient in general, to guarantee existence of a joint k-bit distribution yielding the $k-1$ bit distribution as its marginals, that is, compatibility. This fact has proven important in [5] and other literature about region based approximations analysis of belief propagation. Algorithm 1 handles marginal distributions with matching common elements, and hence in $\tilde{\mathcal{Q}}_\mathcal{A}$, but that are not compatible in line 14, for in this case the obtained $\mathcal{A}$ joint distribution $\mathbf{p}_\mathcal{A}$ will have some negative elements for any $p_\mathcal{A}(\mathbf{0}) \in [0,1]$, and thus will not give a joint entropy of $h_\mathcal{A}$ for any $p_\mathcal{A}(\mathbf{0})$.

We now show that $\mathcal{H}(\mathbf{p}_\mathcal{A}(p_\mathcal{A}(\mathbf{0})))$ is a concave function of $p_\mathcal{A}(\mathbf{0})$ and thus the solution set $\mathcal{P}_\mathcal{A}$ at line 14 has either 0, 1, or 2 solutions. By the definition of entropy:

$$\mathcal{H}(\mathbf{p}_\mathcal{A}) = -\sum_{\mathbf{x} \in \{0,1\}^{|\mathcal{A}|}} p_\mathcal{A}(\mathbf{x}) \log p_\mathcal{A}(\mathbf{x}). \quad (10)$$

The transformation $\mathbf{p}_\mathcal{A} = \mathbf{M}_\mathcal{A}^{-1} \mathbf{q}_\mathcal{A}$ where $\mathbf{q}_\mathcal{A} = (\tilde{\mathbf{q}}_\mathcal{A}, p_\mathcal{A}(\mathbf{0}), 1)$ and $\tilde{\mathbf{q}}_\mathcal{A} = (p_\mathcal{B}(\mathbf{0}), \ \mathcal{B} \subset \mathcal{A})$ means that each element $p_\mathcal{A}(\mathbf{x})$ may be written as an affine function of $p_\mathcal{A}(\mathbf{0})$:

$$p_\mathcal{A}(\mathbf{x}) = M_\mathcal{A}^{-1}(\mathbf{x}, \mathcal{A}) p_\mathcal{A}(\mathbf{0}) + \sum_{\mathcal{B} \subset \mathcal{A} : \mathcal{B} \neq \emptyset} M_\mathcal{A}^{-1}(\mathbf{x}, \mathcal{B}) p_\mathcal{B}(\mathbf{0}) + M_\mathcal{A}^{-1}(\mathbf{x}, 2^{|\mathcal{A}|}) \cdot 1 = m_{\mathbf{x}, \mathcal{A}} p_\mathcal{A}(\mathbf{0}) + b_{\mathbf{x}, \mathcal{A}}.$$

Here we index each element of $\mathbf{M}_\mathcal{A}^{-1}$ by its row using the binary word $\mathbf{x}$ and its column using the corresponding subset $\mathcal{B} \subseteq \mathcal{A}$. It is known that the Shannon entropy is a concave function of its distribution. Here the Shannon entropy of a distribution that is an affine function of $p_\mathcal{A}(\mathbf{0})$ is evaluated. Since concavity is always preserved under composition with affine functions, the Shannon entropy will be a concave function of $p_\mathcal{A}(\mathbf{0})$. This establishes the entropy is concave in $p_\mathcal{A}(\mathbf{0})$, and from

here it follows easily that a concave function defined over the interval $p_{\mathcal{A}}(\mathbf{0}) \in [0,1]$ may have 0, 1, or 2 points that intersect the horizontal line of height $h_{\mathcal{A}}$. It further follows that the solution set $\mathcal{P}_{\mathcal{A}}$ is easily obtained by line search. ■

VI. Information Inequalities and convΦ_N

Frequently the coding problems we are working with whose rate regions may be expressed in terms of Φ_N allow the use of "time-sharing" between two codes, and thus two points in the rate region. This corresponds to a convex combination in rate region space, and thus in Φ_N, because the rate regions are linear maps on Φ_N and the convex combination can commute with this linear map. Thus, in these instances, it is actually of interest to characterize the convex hull convΦ_N rather than just Φ_N. Although Figure 1 demonstrated that the binary entropy vector set is not a polytope even for just two bits, one can use information inequalities such as the Shannon Type inequalities to obtain a polytope $\mathcal{P}_N$ outer bound to it, i.e., a polytope which contains Φ_N. Clearly, the smallest convex set containing Φ_N, convΦ_N, when polytopic, is the tightest polytope containing Φ_N, so it is interesting to see if it is equal to the $\mathcal{P}_N$. Specifically, define $\mathcal{P}_N$ to be the set of all vectors $\mathbf{h} \in \mathbb{R}_+^{2^N-1}$ (not necessarily entropy vectors) obeying the Shannon-type inequalities:

$$H(X_i, \mathbf{X}_{\mathcal{A}}) - H(\mathbf{X}_{\mathcal{A}}) - H(X_i, X_j, \mathbf{X}_{\mathcal{A}}) + H(X_j, \mathbf{X}_{\mathcal{A}})$$

is ≥ 0 for each subset $\mathcal{A} \in 2^{[N]}$ and each $i, j \in [N]$, with the convention that the entropy of the empty set is zero. All Shannon-type inequalities follow from these inequalities [6]. Additionally, binary random variables obey the inequalities $H(X_i) \leq 1, \quad i \in [N]$. Because any entropy vector must obey these inequalities, $\mathcal{P}_N$ is a polytopic outer bound to Φ_N. The outer bound formed by the set of Shannon information inequalities equals the convex hull of the binary entropy vector set in the case $N = 2$, i.e., conv$(\Phi_2) = \mathcal{P}_2$.

Proposition 1: $\mathcal{P}_2 = \text{conv}(\Phi_2)$.

Proof: The set $\mathcal{P}_2$ is generated by the vertices $(0,0,0), (0,1,1), (1,0,1), (1,1,1), (1,1,2)$. Since we already know that $\mathcal{P}_2$ is a convex outer bound for Φ_2, it must be an outer bound for conv(Φ_2). That the generating vertices of $\mathcal{P}_2$ are in Φ_2 can be seen with the following constructions: $(0,1,1)$ $(1,0,1)$ can be achieved by making X_1, X_2 independent with one bit uniform and the other deterministic, $(1,1,1)$ can be achieved by making both bits uniform and equal to each other with probability one, $(1,1,2)$ can be achieved by making both bits uniform and independent, while $(0,0,0)$ can be achieved by making the pair of bits equal to $(0,0)$ with probability one (w.p. 1). Since each of the generating vertices of $\mathcal{P}_2$ are in Φ_2, we must have $\mathcal{P}_2 \subseteq \text{conv}(\Phi_2)$, which then implies (together with the already proven containment in the other direction) equality. □

Proposition 2: $\text{conv}(\Phi_3) = \mathcal{P}_3$.

Proof: The set $\mathcal{P}_3$ is the polytope described by the 15 inequalities

$$\begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & \mathbf{I}_3 & -\mathbf{1}_{3\times 1} \\ \mathbf{I}_3 & \mathbf{A} & \mathbf{1}_{3\times 1} \\ \mathbf{A} & \mathbf{I}_3 & \mathbf{0}_{3\times 1} \end{bmatrix} \begin{bmatrix} H(X_1) \\ H(X_2) \\ H(X_3) \\ H(X_1,X_2) \\ H(X_1,X_3) \\ H(X_2,X_3) \\ H(X_1,X_2,X_3) \end{bmatrix} \leq \begin{bmatrix} \mathbf{1}_{3\times 1} \\ \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 1} \end{bmatrix}$$

$$\mathbf{A} := \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}$$

Using the vertex enumeration routine lrs [7], we obtain a list of 16 generating vertices for $\mathcal{P}_3$ from its Shannon linear inequality representation. By constructing simple distributions as in the previous proof, one may show that each of these vertices also lie in Φ_3, thereby establishing that $\mathcal{P}_3 \subseteq \text{conv}(\Phi_3)$ and thus the proposition. We now list each of these generating vertices and describe a distribution on three bits which achieves them.

- $(0,0,0,0,0,0,0)$: $X_1, X_2, X_3 = (0,0,0)$ (w.p. 1).
- $(1,0,0,1,1,0,1)$: X_1, X_2, X_3 independent, X_1 uniform, X_2, X_3 deterministic.
- $(0,0,1,0,1,1,1)$: X_1, X_2, X_3 independent, X_3 uniform, X_1, X_3 deterministic.
- $(1,0,1,1,2,1,2)$: X_1, X_2, X_3 independent, X_1, X_3 uniform X_2 deterministic.
- $(0,1,1,1,1,2,2)$: X_1, X_2, X_3 independent, X_2, X_3 uniform X_1 deterministic.
- $(1,1,1,2,2,2,3)$: All bits independent and uniform.
- $(0,1,0,1,0,1,1)$: X_1, X_2, X_3 independent, X_2 uniform, X_2, X_3 deterministic.
- $(1,1,0,2,1,1,2)$: X_1, X_2, X_3 independent, X_1, X_2 uniform X_3 deterministic.
- $(1,1,1,2,2,2,2)$: $X_3 = X_1 \oplus X_2$ with (w.p. 1) ($\oplus$ denotes binary addition), and X_1, X_2 independent and uniform.

- $(1,1,1,1,2,2,2)$: X_3 independent of X_1, X_2 and uniform. $X_1 = X_2$ (w.p. 1), and X_1 uniform.
- $(1,1,0,1,1,1,1)$: X_3 independent of X_1, X_2 and deterministic. $X_1 = X_2$ (w.p. 1), and X_1 uniform.
- $(1,1,1,1,1,1,1)$: $X_1 = X_2 = X_3$ (w.p. 1), and X_1 uniform.
- $(1,0,1,1,1,1,1)$: X_2 independent of X_1, X_3 and deterministic. $X_1 = X_3$ (w.p. 1), and X_1 uniform.
- $(1,1,1,2,1,2,2)$: X_2 independent of X_1, X_3 and uniform. $X_1 = X_3$ (w.p. 1), and X_1 uniform.
- $(0,1,1,1,1,1,1)$: X_1 independent of X_2, X_3 and deterministic. $X_2 = X_3$ (w.p. 1), and X_2 uniform.
- $(1,1,1,2,2,1,2)$: X_1 independent of X_2, X_3 and uniform. $X_2 = X_3$ (w.p. 1), and X_2 uniform. □

Thus, the polytopes $\mathcal{P}_2$ and $\mathcal{P}_3$ are simple characterizations of conv(Φ_2) and conv(Φ_3). A similar relationship has long been known to hold between $\mathcal{P}_2$ and $\bar{\Gamma}_2^*$ and $\mathcal{P}_3$ and $\bar{\Gamma}_3^*$, and is implied by this result. Namely, $\bar{\Omega}_2^* = \mathcal{P}_2$ and $\bar{\Omega}_3^* = \mathcal{P}_3$ and $\bar{\Gamma}_2^* = \text{ray}(\bar{\Omega}_2^*)$ and $\bar{\Gamma}_3^* = \text{ray}(\bar{\Omega}_3^*)$ [2], [8]. The proof of these facts submitted here is considerably shorter than that usually used, and it has been recently noted that all of the extreme vertices for $N = 2, 3$ correspond to binary quasi-uniform distributions [8]. Thus these results also show that conv$(\Phi_N) = \bar{\Omega}_N^*$ and cone$(\Phi_N) = \bar{\Gamma}_N^*$ for $N = 2, 3$. Work by Zhang and Yeung [9] and later work by Dougherty, Freiling, and Zeger [10] then showed that Shannon's Inequalities were no longer sufficient for $N \geq 4$, so that $\Gamma_N \neq \bar{\Gamma}_N^*$ for $N \geq 4$ by providing more information inequalities that held among 4 or more variables that were not of Shannon Type. However, recent brilliant work by Matúš [4] has shown that the situation for $\bar{\Gamma}_N^*$ becomes far more complex at $N = 4$ variables, namely it is not polyhedral for $N \geq 4$, so an infinite number of such linear inequalities are necessary to describe it. The relevant question for Φ_N is then: *is convΦ_N a polytope for $N \geq 4$*? If the answer is affirmative then there is a significant advantage to working with Φ_N instead of Γ_N, since conv(Φ_N) would be characterized by only a finite number of inequalities, but it could still serve the same purposes as $\bar{\Gamma}_N^*$.

A. Inner Bounds for conv(Φ_N) and for $\bar{\Omega}_N^*, \bar{\Gamma}_N^*$

Because we know that there is a gap between the outer bound $\mathcal{P}_N$ and conv(Φ_N) for $N \geq 4$, it is of interest to know where the outer bound is tight by obtaining an inner bound which agrees with it on some faces. Of course, because conv(Φ_N) is a subset of $\bar{\Omega}_N^*$, such an inner bound is also an inner bound to $\bar{\Omega}_N^*$, and thus on faces which it agrees with the outer bound $\mathcal{P}_N$, these bounds also agree with $\bar{\Omega}_N^*$.

Such an inner bound for conv(Φ_N), call it $\mathcal{I}_N$, can be easily obtained computationally as follows:

1) Enumerate the generating vertices of the polytope $\mathcal{P}_N$, by using a double description algorithm to convert the linear inequality representation into the generating vertices representation (e.g. with the program lrs).
2) For each of these vertices, determine if they lie in Φ_N, using the recursive characterization of Φ_N described in V. Keep only those vertices lying in Φ_N.
3) Take the convex hull of these vertices to get the polytope $\mathcal{I}_N$. $\mathcal{I}_N$ can be expressed in normal linear inequality form by using the double description method again (e.g. with the program lrs).

We could have picked as our inner bound any random collection of points in Φ_N. We selected the generating vertices of $\mathcal{P}_N$ because, by design, the inner bound $\mathcal{I}_N$ generated by this method will agree with the polytope $\mathcal{P}_N$ on exactly those faces where it is tight with conv(Φ_N). Thus, inner and outer bounds in information theory problems which are linear maps on these regions will agree, yielding an exact description of the rate region, when they depend on those regions of $\mathcal{P}_N$ which are tight on conv(Φ_N). No special property of the outer bound $\mathcal{P}_N$ other than the fact that it was polyhedral here, and indeed, tight faces of any polyhedral outer bound to $\bar{\Omega}_N^*$ can be determined in the same manner. These in turn can form bounds for $\bar{\Gamma}_N^*$ with the same property by taking conic hulls instead of convex hulls.

VII. Conclusions

We have provided a finite terminating algorithm which can determine whether or not a candidate entropy vector lies in the set Φ_N of entropy vectors for N binary random variables. We discussed the relationship between Φ_N and its unbounded cardinality counterparts $\bar{\Gamma}_N^*$ and $\bar{\Omega}_N^*$, and outlined how the new algorithm can be used to gain insights about these more general sets. Specifically, we outlined how tight faces of any polyhedral outer bound to $\bar{\Gamma}_N^*$ and $\bar{\Omega}_N^*$ could be found using the new algorithm.

Acknowledgments

The authors thank Raymong Yeung and Sormeh Shadbakht for helpful brief discussions and the Air Force Office of Scientific Research for their support under AFOSR FA9550-09-C-0014. John Walsh thanks Drexel

University and the National Science Foundation for their support in part under CCF-0728496.

REFERENCES

[1] Raymond W. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[2] B. Hassibi and S. Shadbakht, "Normalized Entropy Vectors, Network Information Theory and Convex Optimization," in *IEEE Information Theory Workshop*, July 2007, pp. 1 – 5.

[3] ——, "On a Construction of Entropic Vectors Using Lattice-Generated Distributions," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 501 – 505.

[4] František Matúš, "Infinitely Many Information Inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 41–44.

[5] J. Yedidia and W. Freeman and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Info. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[6] Zhen Zhang and Raymond W. Yeung, "On Characterization of Entropy Function via Information Inequalities," *IEEE Trans. Info. Theory*, vol. 44, no. 4, July 1998.

[7] D. Avis, "lrslib ver 4.2." [Online]. Available: http://cgm.cs.mcgill.ca/∼avis/C/lrs.html

[8] D. Fong, S. Shadbakht, B. Hassibi, "On the Entropy Region and the Ingleton Inequality," in *Mathematical Theory of Networks and Systems (MTNS)*, 2008.

[9] Zhen Zhang and Raymond W. Yeung, "A Non-Shannon-Type Conditional Inequality of Information Quantities," *IEEE Trans. Info. Theory*, vol. 43, no. 6, Nov. 1997.

[10] R. Dougherty, C. Freiling, and K. Zeger, "Networks, Matroids, and Non-Shannon Information Inequalities," *IEEE Trans. Info. Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

APPENDIX A
PROOF OF LEMMA 1

Let the joint distribution be $\mathbf{p}_{[m]} = (p_{[m]}(\mathbf{x}),\ \mathbf{x} \in \{0,1\}^m)$. Each element of the marginal distribution for the rvs $\mathbf{X}_\mathcal{A}$, say $p_\mathcal{A}(\mathbf{y})$, is found by summing the joint distribution $\mathbf{p}_{[m]}$ over the set of m-bit binary strings containing $\mathbf{y}$:

$$p_\mathcal{A}(\mathbf{y}) = \sum_{\mathbf{x}\in\{0,1\}^m:\ \mathbf{x}_\mathcal{A}=\mathbf{y}} p_{[m]}(\mathbf{x}),\ \mathbf{y} \in \{0,1\}^{|\mathcal{A}|}, \quad (11)$$

where $\mathbf{x}_\mathcal{A}$ returns the elements from $\mathbf{x}$ at positions indexed in $\mathcal{A}$. The marginal distribution $\mathbf{p}_\mathcal{A}$ may be obtained from the joint distribution $\mathbf{p}_{[m]}$ by a matrix $\mathbf{L}_{\mathcal{A},m}$ via $\mathbf{p}_\mathcal{A} = \mathbf{L}_{\mathcal{A},m}\mathbf{p}_{[m]}$, where $\mathbf{L}_{\mathcal{A},m}$ is an $|\mathcal{A}| \times m$ matrix with entries

$$L_{\mathcal{A},m}(i,j) = \begin{cases} 1, & \mathbf{x}^{(i)}_{|\mathcal{A}|} = \mathbf{y}^{(j)}_\mathcal{A} \\ 0, & \text{else} \end{cases}, \quad (12)$$

where $\mathbf{x}^{(i)}_{|\mathcal{A}|}$ is the i^{th} $|\mathcal{A}|$-bit word, ordered lexicographically, and $\mathbf{y}^{(j)}_\mathcal{A}$ is the subset of elements at positions $\mathcal{A}$ from the j^{th} m-bit word, ordered lexicographically. Let $\mathcal{I}(m) = (\mathcal{A} \subseteq [m] : \mathcal{A} \neq \emptyset)$ be the ordered set of all nonempty subsets of $[m]$, ordered in terms of cardinality, then among those with the same cardinality, ordered lexicographically. Let $\mathcal{M}(m) = (\mathbf{p}_\mathcal{A},\ \mathcal{A} \in \mathcal{I}(m))$ be an ordered set of marginal distributions, one for each nonempty subset of $[m]$. It is convenient to list $\mathbf{q}_{[m]}$ as defined in the theorem with the empty set in the last position. We obtain $\mathbf{q}_{[m]}$ from $\mathbf{p}_{[m]}$ via $\mathbf{q}_{[m]} = \mathbf{M}_{[m]}\mathbf{p}_{[m]}$, where $\mathbf{M}_{[m]}$ is the $2^m \times 2^m$ matrix with entries

$$M_{[m]}(i,j) = \begin{cases} 1, & \mathbf{x}^{(j)}_{\mathcal{A}^{(i)}} = \mathbf{0}, \\ 0, & \text{else} \end{cases}, \quad (13)$$

for $i = 1,\ldots,2^m-1,\quad j = 1,\ldots,2^m$, and a bottom row of ones, i.e., $M_{[m]}(2^m, j) = 1$. Here $\mathcal{A}^{(i)}$ is the i^{th} nonempty subset of $[m]$, ordered first by cardinality then lexicographically, and $\mathbf{x}^{(j)}$ is the j^{th} m-bit word. Thus $\mathbf{x}^{(j)}_{\mathcal{A}^{(i)}}$ consists of those bits at positions $\mathcal{A}^{(i)}$ from word $\mathbf{x}^{(j)}$. This completes the construction for obtaining $\mathbf{q}_{[m]}$ from $\mathbf{p}_{[m]}$. We now give a unique construction for obtaining $\mathbf{p}_{[m]}$ from $\mathbf{q}_{[m]}$ for each $\mathbf{x} \in \{0,1\}^m$. Define $\mathcal{A}(\mathbf{x})$ as the indices of the zeros in $\mathbf{x}$. Thus, for $\mathbf{x} = (0,1,0)$ we have $\mathcal{A}(010) = (1,3)$. The proof is by induction in $k = |\mathcal{A}(\mathbf{x})|$. The base case ($k = m$) is trivial. The only word with $k = m$ is $\mathbf{x} = \mathbf{0}$, and this probability, $p_{[m]}(\mathbf{0})$ is an element of both $\mathbf{p}_{[m]}$ and $\mathbf{q}_{[m]}$. Suppose we have a unique construction for each m-bit word $\mathbf{x}$ with $|\mathcal{A}(\mathbf{x})| \in \{k+1,\ldots,m\}$. That is, for each such $\mathbf{x}$ we can express $p_{[m]}(\mathbf{x})$ in terms of $\mathbf{q}_{[m]}$. Consider an arbitrary m-bit word $\mathbf{x}$ with $|\mathcal{A}(\mathbf{x})| = k$ zeroes. Then, using $\mathbf{y} = \mathbf{0}$ in (11) and splitting off the $\mathbf{x}$ term from the sum yields:

$$p_{[m]}(\mathbf{x}) = p_{\mathcal{A}(\mathbf{x})}(\mathbf{0}) - \sum_{\mathbf{y}\neq\mathbf{x}:\mathbf{y}_{\mathcal{A}(\mathbf{x})}=\mathbf{0}} p_{[m]}(\mathbf{y}). \quad (14)$$

But each $\mathbf{y}$ in the sum must satisfy $\mathbf{y}_{\mathcal{A}(\mathbf{x})} = \mathbf{0}$ and $\mathbf{y} \neq \mathbf{x}$, which means each such term has strictly more zeroes than $\mathbf{x}$, i.e.,

$$\mathbf{y}_{\mathcal{A}(\mathbf{x})} = \mathbf{0} \text{ and } \mathbf{y} \neq \mathbf{x} \ \Rightarrow\ \mathcal{A}(\mathbf{y}) > \mathcal{A}(\mathbf{x}). \quad (15)$$

Thus $\mathcal{A}(\mathbf{y}) > k$, which means we have a unique construction for each such $\mathbf{y}$. The final case is for $k = 0$, i.e., the word $\mathbf{x} = \mathbf{1}$. But we can write $p_{[m]}(\mathbf{1}) = 1 - \sum_{\mathbf{y}\neq\mathbf{1}} p_{[m]}(\mathbf{y})$ and note that each such $\mathbf{y}$ has $\mathcal{A}(\mathbf{y}) > 0$. Applying this procedure to each word will yield the $2^m \times 2^m$ inverse matrix $\mathbf{M}^{-1}_{[m]}$ so that $\mathbf{p}_{[m]} = \mathbf{M}^{-1}_{[m]}\mathbf{q}_{[m]}$. This establishes the proof.