

# Factorization of Joint Probability Mass Functions into Parity Check Interactions

Muhammet Fatih Bayramoğlu and Ali Özgür Yılmaz  
 Dept. of Electrical and Electronics Eng., Middle East Technical University  
 Email: {fatih,aoyilmaz}@eee.metu.edu.tr

**Abstract**—We show that any joint probability mass function (PMF) can be expressed as a product of parity check factors and factors of degree one with the help of some auxiliary variables, if the alphabet size is appropriate for defining a parity check equation. In other words, marginalization of a joint PMF is equivalent to a soft decoding task as long as a finite field can be constructed over the alphabet of the PMF. In factor graph terminology this claim means that a factor graph representing such a joint PMF always has an equivalent Tanner graph. We provide a systematic method based on the Hilbert space of PMFs and orthogonal projections for obtaining this factorization.

## I. INTRODUCTION

Most of the problems faced in communication systems are in the form of marginalization of joint PMFs. If the joint PMF is in the form of a product of some local functions (factors or interactions) then the marginalization task can be accomplished by the sum-product algorithm [1], [2]. However, the factorization structures of joint PMFs are not apparent always. Therefore, a systematic method showing the factorization structure of joint PMFs proves useful.

We propose a method for this purpose which is based on the Hilbert space of PMFs and orthogonal projections. The Hilbert space of PMFs is proposed in our recent work [3] and has potential applications one of which is proposed in this paper.

Our proposed method factorizes joint PMFs into soft parity check interactions (SPCI). We define an SPCI as a generalized form of parity check constraints. A parity check constraint guarantees that the weighted sum of the variables included in the parity check always equals to zero. However, in SPCIs we allow the weighted sum to admit all the values with certain probabilities. It is shown that SPCIs sharing the same set of parity check coefficients form a subspace. Then the factorization of joint PMFs is achieved by projecting them onto these subspaces.

Since our method employs parity checks, it is applicable to PMFs of certain alphabet sizes. The alphabet size of the random variables should be a prime number or its powers, for which a finite field exist. This may seem as a severe restriction. However, in the case of communication problems this restriction does not cause a big trouble since the alphabet sizes in the communication problems are either two or its powers usually.

It is known that the soft decoding operation is a special case of the marginalization of joint PMFs. In this work we show

that the reverse is also true for certain alphabet sizes. In other words, we show that marginalization sum can be handled by a soft decoder. This soft decoder belongs not to an arbitrary code but to the dual code of the Hamming code.

The paper is organized as follows. In the next section the Hilbert space of PMFs will be briefly introduced. Third section explains the factorization of joint PMFs in detail. In the fourth section, we show that the soft decoder of the dual Hamming code can be employed as a universal marginalization machine.

## II. THE HILBERT SPACE OF PMFS

The Hilbert space of PMFs is summarized in this section. Readers may refer to [3] for a more detailed explanation of the Hilbert space of the PMFs.

Consider an experiment with a set of outcomes (alphabet)  $\mathcal{A}$  which is discrete and has a finite number of elements. The probabilities assigned to these outcomes define a PMF such that  $p(x) = Pr\{x\}$  for every  $x$  in  $\mathcal{A}$ . Each different assignment of the probabilities to the outcomes defines a different PMF. We denote the set of all possible PMFs defined over the alphabet  $\mathcal{A}$  by  $\mathcal{V}_{\mathcal{A}}$  which is formally defined as

$$\mathcal{V}_{\mathcal{A}} \triangleq \{p(x) : \mathcal{A} \rightarrow [0, 1] : \sum_{\forall x \in \mathcal{A}} p(x) = 1\}. \quad (1)$$

The addition and the scalar multiplication operations are necessary to construct an algebraic structure over  $\mathcal{V}_{\mathcal{A}}$ . The addition of PMFs is denoted by  $\boxplus$  and defined as

$$p(x) \boxplus q(x) \triangleq \frac{1}{Z} p(x)q(x), \quad (2)$$

where  $p(x)$ ,  $q(x)$  are PMFs in  $\mathcal{V}_{\mathcal{A}}$  and  $Z$  is the normalization constant. The scalar multiplication is denoted by  $\boxtimes$  and is given as

$$\alpha \boxtimes p(x) \triangleq \frac{1}{Z} (p(x))^\alpha \quad (3)$$

where  $\alpha$  is in  $\mathbb{R}$  and  $Z$  is the normalization constant once again. This normalization constant is necessary to ensure the closure of the  $\mathcal{V}_{\mathcal{A}}$  under the addition and the scalar multiplication operations. Hence, its value is  $Z = \sum_{\forall x \in \mathcal{A}} p(x)q(x)$  for the case of addition and  $Z = \sum_{\forall x \in \mathcal{A}} (p(x))^\alpha$  for the case of scalar multiplication. Note that the PMFs are denoted not only by letter  $p$  but also by other lower case letters in the paper.

It can be shown that the set  $\mathcal{V}_{\mathcal{A}}$  together with the operations  $\boxplus$  and  $\boxtimes$  forms a vector space over  $\mathbb{R}$  [3].

The geometric structure over this vector space can be defined by means of an inner product. This vector space admits the following function as an inner product [3].

$$\langle p(x), q(x) \rangle: \mathcal{V}_{\mathcal{A}} \times \mathcal{V}_{\mathcal{A}} \rightarrow \mathbb{R} \triangleq \sum_{x \in \mathcal{A}} \left( \log \frac{(p(x))^{|\mathcal{A}|}}{\prod_{y \in \mathcal{A}} p(y)} \log \frac{(q(x))^{|\mathcal{A}|}}{\prod_{y \in \mathcal{A}} q(y)} \right) \quad (4)$$

where  $|\mathcal{A}|$  denotes the cardinality of the set  $\mathcal{A}$ . This definition can be simplified by introducing the following mapping.

$$\mathcal{L}\{p(x)\}: \mathcal{V}_{\mathcal{A}} \rightarrow \mathbb{R}^{|\mathcal{A}|} \triangleq \sum_{i=0}^{|\mathcal{A}|-1} \left( \log \frac{(p(x_i))^{|\mathcal{A}|}}{\prod_{y \in \mathcal{A}} p(y)} \right) \mathbf{e}_i \quad (5)$$

where  $x_i$  denotes the  $i^{\text{th}}$  element of the set  $\mathcal{A}$  and  $\mathbf{e}_i$  denotes the  $i^{\text{th}}$  canonical basis vector of  $\mathbb{R}^{|\mathcal{A}|}$ . Then the inner product of PMFs simply becomes

$$\langle p(x), q(x) \rangle = \sum_{i=0}^{|\mathcal{A}|-1} (\mathbf{p})_i (\mathbf{q})_i = \langle \mathbf{p}, \mathbf{q} \rangle \quad (6)$$

where  $\mathbf{p}, \mathbf{q}$  are vectors in  $\mathbb{R}^{|\mathcal{A}|}$  such that  $\mathbf{p} = \mathcal{L}\{p(x)\}$ ,  $\mathbf{q} = \mathcal{L}\{q(x)\}$ , and  $(\mathbf{p})_i$  ( $(\mathbf{q})_i$ ) denotes the  $i^{\text{th}}$  component of the vector  $\mathbf{p}$  ( $\mathbf{q}$ ). This identity shows that  $\mathcal{L}\{\cdot\}$  is an isometric transformation from  $\mathcal{V}_{\mathcal{A}}$  to  $\mathbb{R}^{|\mathcal{A}|}$ .

The mapping  $\mathcal{L}\{\cdot\}$  have further important properties. It is linear and one-to-one [3]. These properties allow us to find the dimension of the vector space  $\mathcal{V}_{\mathcal{A}}$ . The dimension of  $\mathcal{V}_{\mathcal{A}}$  is not very simple to calculate; whereas, the dimension of the range space of the  $\mathcal{L}\{\cdot\}$  is. For any  $p(x) \in \mathcal{V}_{\mathcal{A}}$ , let  $\mathbf{p} = \mathcal{L}\{p(x)\}$  then

$$\sum_{i=0}^{|\mathcal{A}|-1} (\mathbf{p})_i = \sum_{i=0}^{|\mathcal{A}|-1} \log \frac{(p(x_i))^{|\mathcal{A}|}}{\prod_{y \in \mathcal{A}} p(y)} = 0. \quad (7)$$

Therefore, the range space of  $\mathcal{L}\{\cdot\}$  becomes the set  $\{\mathbf{p} \in \mathbb{R}^{|\mathcal{A}|} : (1, 1, \dots, 1)\mathbf{p} = 0\}$ , which is clearly a  $|\mathcal{A}| - 1$  dimensional subspace of  $\mathbb{R}^{|\mathcal{A}|}$ . Hence,  $\mathcal{V}_{\mathcal{A}}$  is a  $|\mathcal{A}| - 1$  dimensional vector space. Moreover,  $\mathcal{V}_{\mathcal{A}}$  is a Hilbert space since it is a finite dimensional inner product space.

#### A. The Hilbert Space of Joint PMFs

The Hilbert space structure can be applied to joint PMFs of combined experiments as long as each individual experiments has a finite alphabet. Consider a combined experiment consisting of  $N$  individual discrete experiments with alphabets  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_N$ . Then the alphabet of the combined experiment, which is denoted by  $\mathcal{S}$ , is

$$\mathcal{S} = \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_N.$$

Hence, the alphabet size of the combined experiment is  $|\mathcal{S}| = \prod_{i=1}^N |\mathcal{A}_i|$ . Consequently, the dimension of this Hilbert space is

$$\dim \mathcal{V}_{\mathcal{S}} = \prod_{i=1}^N |\mathcal{A}_i| - 1. \quad (8)$$

If all of the individual experiments are defined over the same alphabet denoted by  $\mathcal{A}$  then  $\dim \mathcal{V}_{\mathcal{S}} = |\mathcal{A}|^N - 1$ .

### III. FACTORIZATION OF JOINT PMFS

In this section the factorization of joint PMFs is analyzed in a systematic way. Let the joint PMF under concern be  $p(x_1, x_2, \dots, x_N)$  which is an element of  $\mathcal{V}_{\mathcal{S}}$  as defined in the previous section. Suppose that this joint PMF can be expressed as

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^K \phi_i(\mathcal{X}_i) \quad (9)$$

where  $\mathcal{X}_i$ 's are the subsets of the set  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$  and the arguments of the functions  $\phi_i(\mathcal{X}_i)$  are the elements of  $\mathcal{X}_i$ . The functions  $\phi_i(\mathcal{X}_i)$ 's are called factor functions or interactions.

The factor functions are not necessarily PMFs in general. However, a proper PMF can be defined for each factor function by properly scaling them as follows.

$$q_i(x_1, x_2, \dots, x_N) = \frac{\phi_i(\mathcal{X}_i)}{\sum_{\mathcal{X}_i} \phi_i(\mathcal{X}_i)}.$$

Although  $q_i$  has all  $x_1, x_2, \dots, x_N$  as arguments in this notation, its value is independent of the arguments in  $\mathcal{X} \setminus \mathcal{X}_i$  and it is still a function of the members of  $\mathcal{X}_i$  only. After this scaling (9) can be rewritten as

$$p(x_1, x_2, \dots, x_N) = \frac{1}{Z} \prod_{i=1}^K q_i(x_1, x_2, \dots, x_N). \quad (10)$$

Note that  $p(x_1, x_2, \dots, x_N)$  and  $q_i(x_1, x_2, \dots, x_N)$ s are all members of the Hilbert space  $\mathcal{V}_{\mathcal{S}}$ , and the representation of (10) in this Hilbert space is

$$p(x_1, x_2, \dots, x_N) = \boxplus_{i=1}^K q_i(x_1, x_2, \dots, x_N). \quad (11)$$

#### A. Soft Parity Check Interactions

A random variable is defined as a mapping from the event space to the real line. This is also true for discrete experiments as well. However, if the number of outcomes of the discrete experiment is appropriate, defining a discrete random variable as a mapping from event space to a Galois field may inspire new ideas. This section is built on this idea. Therefore, in the rest of the paper it is assumed that it is possible to make a one-to-one matching between the event space and a Galois field. In other words, we assume that

$$\mathcal{A} = \text{GF}(|\mathcal{A}|), \quad (12)$$

where  $\text{GF}(|\mathcal{A}|)$  denotes the Galois field of order  $|\mathcal{A}|$ . Furthermore, it is assumed that combined experiments consist of individual experiments with identical event spaces. That is,

$$\mathcal{S} = \mathcal{A}^N = \text{GF}^N(|\mathcal{A}|).$$

Working on Galois fields allows us to define interactions (factor functions, joint PMFs) based on algebraic operations. An example for such an interaction is the soft parity check interaction (SPCI). We define SPCI as follows.

**Definition 1. Soft Parity Check Interaction:** A joint PMF  $p(x_1, x_2, \dots, x_N)$ , in  $\mathcal{V}_{\mathcal{S}}$ , where  $\mathcal{S} = \text{GF}^N(|\mathcal{A}|)$ , is called a

soft parity check interaction if there exists a  $q(x) \in \mathcal{V}_{\text{GF}(|\mathcal{A}|)}$  and a vector  $\mathbf{a} = (a_1, a_2, \dots, a_N) \in \text{GF}^N(|\mathcal{A}|)$  such that

$$p(\mathbf{x}) = \frac{1}{|\mathcal{A}|^{N-1}} q(\mathbf{a}\mathbf{x}^T),$$

where  $\mathbf{x}$  denotes  $(x_1, x_2, \dots, x_N)$  and  $^T$  denotes transposition. Moreover, the vector  $\mathbf{a}$  is called the **parity check coefficient vector** of the SPCI and the weight of this vector is called the **order** of the SPCI  $p(\mathbf{x})$ .

As its name implies, an SPCI, relates the random variables by a parity check equation. The term ‘‘soft’’ arises from the fact that the parity check equation is not guaranteed to be satisfied. That is, the weighted sum of the random variables has a probability distribution rather than being guaranteed to be zero.

*Example 1.* Let  $p_1(x_1, x_2)$  and  $p_2(x_1, x_2)$  be two PMFs which are given, with a slight abuse of notation, as

$$p_1(x_1, x_2) = \begin{bmatrix} 0.2 & 0.1 & 0.1/3 \\ 0.1/3 & 0.2 & 0.1 \\ 0.1 & 0.1/3 & 0.2 \end{bmatrix}$$

$$p_2(x_1, x_2) = \frac{1}{238} \begin{bmatrix} 144 & 18 & 6 \\ 3 & 18 & 36 \\ 3 & 4 & 6 \end{bmatrix}$$

where  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of the matrices represent the value of  $p_{1,2}(x_1 = i - 1, x_2 = j - 1)$ . In this example  $p_1(x_1, x_2) = (1/3)q(x_1 + 2x_2)$  where  $q(x) = [0.6 \ 0.1 \ 0.3]$  with a similar abuse of notation. Hence  $p_1(x_1, x_2)$  is an SPCI. On the other hand  $p_2(x_1, x_2)$  is not an SPCI since such an expression is not possible for it.

The SPCIs have some important properties. Firstly, the marginal functions associated with an SPCI will be investigated. If the order of the SPCI is one then the  $i^{\text{th}}$  marginal function becomes

$$\sum_{\forall(\mathcal{X} \setminus x_i)} \frac{1}{|\mathcal{A}|^{N-1}} q(\mathbf{a}\mathbf{x}^T) = \begin{cases} q(a_i x_i) & , \quad a_i \neq 0 \\ \frac{1}{|\mathcal{A}|} & , \quad \text{otherwise} \end{cases}.$$

In other words, SPCIs of order one provide local evidence about the variable whose associated coefficient is nonzero. If the order is greater than one then

$$\sum_{\forall(\mathcal{X} \setminus x_i)} \frac{1}{|\mathcal{A}|^{N-1}} q(\mathbf{a}\mathbf{x}^T) = \frac{1}{|\mathcal{A}|} \quad (13)$$

for all  $i \in \{1, 2, \dots, N\}$ , which means SPCIs of order greater than one do not provide any local evidence. However, these SPCIs provide information when used together with other SPCIs. Hence, we say that SPCIs of order greater than one provide purely extrinsic information.

Secondly, in a sum-product algorithm point of view, message computation for SPCIs is less complex. In general, for a factor function in  $\mathcal{V}_S$ , the message computation complexity is  $|\mathcal{A}|^N$  [1]. The reduced complexity message computation algorithm for low-density parity-check decoding presented in

[4] is directly applicable to SPCIs as well. Hence, message computation for an SPCI is  $N|\mathcal{A}| \log |\mathcal{A}|$ .

Finally and the most importantly, the set of SPCIs sharing the same parity check coefficients, as stated by Theorem 1, is a subspace of  $\mathcal{V}_S$ . The set of SPCIs with the parity check coefficient vector  $\mathbf{a}$  is denoted by  $\mathcal{V}_S^{\mathbf{a}}$  and defined as follows.

$$\mathcal{V}_S^{\mathbf{a}} = \left\{ p(\mathbf{x}) = \frac{1}{|\mathcal{A}|^{N-1}} q(\mathbf{a}\mathbf{x}^T) : q(x) \in \mathcal{V}_{\text{GF}(|\mathcal{A}|)} \right\}$$

**Theorem 1.** For any nonzero  $\mathbf{a}$  in  $\text{GF}^N(|\mathcal{A}|)$ ,  $\mathcal{V}_S^{\mathbf{a}}$  is a  $|\mathcal{A}| - 1$  dimensional subspace of  $\mathcal{V}_S$ .

*Proof:* For each  $\mathbf{a}$ , we can define the following mapping.

$$\mathcal{T}_{\mathbf{a}} \{q(x)\} : \mathcal{V}_{\text{GF}(|\mathcal{A}|)} \rightarrow \mathcal{V}_S \triangleq \frac{1}{|\mathcal{A}|^{N-1}} q(\mathbf{a}\mathbf{x}^T)$$

Clearly this mapping is one-to-one and it can be easily shown that it is also linear. It is well known from linear algebra that the range space of a linear mapping is a subspace of the codomain. Moreover, if the mapping is one-to-one the dimension of the range space is equal to the dimension of the domain of the mapping. Hence,

$$\dim \mathcal{V}_S^{\mathbf{a}} = \dim \mathcal{V}_{\text{GF}(|\mathcal{A}|)} = |\mathcal{A}| - 1 \quad (14)$$

■

Now the relations between two different subspaces defined by two different parity check coefficient vectors can be investigated. These relations are explained by the following theorems.

**Theorem 2.** For any two nonzero parity check coefficient vectors  $\mathbf{a}$  and  $\mathbf{b}$  in  $\text{GF}^N(|\mathcal{A}|)$ ,  $\mathcal{V}_S^{\mathbf{a}} = \mathcal{V}_S^{\mathbf{b}}$  if  $\mathbf{a} = \alpha\mathbf{b}$  for an  $\alpha$  in  $\text{GF}(|\mathcal{A}|)$ .

*Proof:* For any  $p(\mathbf{x})$  in  $\mathcal{V}_S^{\mathbf{a}}$  there exist a  $q_1(x)$  in  $\mathcal{V}_{\text{GF}(|\mathcal{A}|)}$  such that  $p(\mathbf{x}) = q_1(\mathbf{a}\mathbf{x}^T)$ . Let  $q_2(x) = q_1(\alpha x)$ . Clearly  $q_2(x)$  is in  $\mathcal{V}_{\text{GF}(|\mathcal{A}|)}$ . Then,

$$p(\mathbf{x}) = \frac{1}{|\mathcal{A}|^{N-1}} q_1(\alpha\mathbf{b}\mathbf{x}^T) = \frac{1}{|\mathcal{A}|^{N-1}} q_2(\mathbf{b}\mathbf{x}^T).$$

Therefore,  $p(\mathbf{x})$  is also an element of  $\mathcal{V}_S^{\mathbf{b}}$ . Hence,

$$\mathcal{V}_S^{\mathbf{a}} = \mathcal{V}_S^{\mathbf{b}},$$

if  $\mathbf{a} = \alpha\mathbf{b}$ . ■

**Theorem 3.** For any two nonzero parity check coefficient vectors  $\mathbf{a}$  and  $\mathbf{b}$  in  $\text{GF}^N(|\mathcal{A}|)$ , the subspace  $\mathcal{V}_S^{\mathbf{a}}$  is orthogonal to the subspace  $\mathcal{V}_S^{\mathbf{b}}$  if  $\mathbf{a} \neq \alpha\mathbf{b}$  for any  $\alpha$  in  $\text{GF}(|\mathcal{A}|)$ .

*Proof:* For any  $p_1(\mathbf{x}) \in \mathcal{V}_S^{\mathbf{a}}$  and  $p_2(\mathbf{x}) \in \mathcal{V}_S^{\mathbf{b}}$ , the inner product of these two SPCIs is

$$\langle p_1(\mathbf{x}), p_2(\mathbf{x}) \rangle = \sum_{\forall \mathbf{x}} \left( \log \frac{(p_1(\mathbf{x}))^{(|\mathcal{A}|^N)}}{\prod_{\forall \mathbf{y}} p_1(\mathbf{y})} \log \frac{(p_2(\mathbf{x}))^{(|\mathcal{A}|^N)}}{\prod_{\forall \mathbf{y}} p_2(\mathbf{y})} \right).$$

Let  $q_1(\mathbf{ax}^T) = |\mathcal{A}|^{N-1}p_1(\mathbf{x})$  and  $q_2(\mathbf{bx}^T) = |\mathcal{A}|^{N-1}p_2(\mathbf{x})$ . Then the inner product can be rewritten as

$$\begin{aligned} & \langle p_1(\mathbf{x}), p_2(\mathbf{x}) \rangle = \\ & \sum_{\forall \mathbf{x}} \left( \log \frac{(q_1(\mathbf{ax}^T))^{(|\mathcal{A}|^N)}}{\prod_{\forall \mathbf{y}} q_1(\mathbf{ay}^T)} \log \frac{(q_2(\mathbf{bx}^T))^{(|\mathcal{A}|^N)}}{\prod_{\forall \mathbf{y}} q_2(\mathbf{by}^T)} \right). \end{aligned}$$

In order to simplify the notation we can use operator  $\mathcal{L}\{\cdot\}$ . Let  $\mathbf{q}_1 = \mathcal{L}\{q_1(x)\}$  and  $\mathbf{q}_2 = \mathcal{L}\{q_2(x)\}$ . Then the inner product can be simplified as

$$\langle p_1(\mathbf{x}), p_2(\mathbf{x}) \rangle = |\mathcal{A}|^{2N-2} \sum_{\forall \mathbf{x}} (\mathbf{q}_1)_{\mathbf{ax}^T} (\mathbf{q}_2)_{\mathbf{bx}^T},$$

where the constant  $|\mathcal{A}|^{2N-2}$  arises from the differences between the alphabet sizes of  $\mathcal{S}$  and  $\text{GF}(|\mathcal{A}|)$ . Then, for some dummy variables  $c_1, c_2$  in  $\text{GF}(|\mathcal{A}|)$  the summation above can be grouped as follows.

$$\begin{aligned} \frac{\langle p_1(\mathbf{x}), p_2(\mathbf{x}) \rangle}{|\mathcal{A}|^{2N-2}} &= \sum_{\forall c_1} \sum_{\forall c_2} \sum_{\forall \mathbf{x} \in \mathcal{K}} (\mathbf{q}_1)_{c_1} (\mathbf{q}_2)_{c_2} \\ &= \sum_{\forall c_1} \left( (\mathbf{q}_1)_{c_1} \sum_{\forall c_2} \left( (\mathbf{q}_2)_{c_2} \sum_{\forall \mathbf{x} \in \mathcal{K}} 1 \right) \right) \\ &= \sum_{\forall c_1} \left( (\mathbf{q}_1)_{c_1} \sum_{\forall c_2} (\mathbf{q}_2)_{c_2} |\mathcal{K}| \right) \end{aligned}$$

where  $\mathcal{K} = \{\mathbf{x} \in \text{GF}^N(|\mathcal{A}|) : \mathbf{ax}^T = c_1 \wedge \mathbf{bx}^T = c_2\}$ . If  $\mathbf{a}$  was equal to  $\alpha\mathbf{b}$  then there were either  $|\mathcal{A}|^{N-1}$  or no  $\mathbf{x}$  vectors satisfying the conditions of set  $\mathcal{K}$  depending on the values of  $c_1$  and  $c_2$ . However, since  $\mathbf{a}$  is not a scaled version of  $\mathbf{b}$  there are always  $|\mathcal{A}|^{N-2}$  elements in  $\mathcal{K}$  regardless of the values of  $c_1$  and  $c_2$ . Hence, the inner product becomes

$$\begin{aligned} \langle p_1(\mathbf{x}), p_2(\mathbf{x}) \rangle &= |\mathcal{A}|^{3N-4} \left( \sum_{\forall c_1} (\mathbf{q}_1)_{c_1} \right) \left( \sum_{\forall c_2} (\mathbf{q}_2)_{c_2} \right) \\ &= 0, \end{aligned}$$

where the last line follows from (7). Finally, the subspace  $\mathcal{V}_S^{\mathbf{a}}$  is orthogonal to  $\mathcal{V}_S^{\mathbf{b}}$  since any  $p_1(\mathbf{x})$  in  $\mathcal{V}_S^{\mathbf{a}}$  is orthogonal to any  $p_2(\mathbf{x})$  in  $\mathcal{V}_S^{\mathbf{b}}$ . ■

The next question to be asked after Theorem 3 is what the number of different subspaces is. This question is equivalent to asking the number of distinct vectors in  $\text{GF}^N(|\mathcal{A}|)$  such that every pair of vectors are linearly independent. Note that the answer to this question is equal to the number of columns of a parity check matrix of a Hamming code defined over  $\text{GF}(|\mathcal{A}|)$  having  $N$  rows. As explained in [5], the number of distinct vectors in  $\text{GF}^N(|\mathcal{A}|)$  which are pairwise linearly independent is  $\frac{|\mathcal{A}|^{N-1}}{|\mathcal{A}|-1}$  and so is the number of distinct subspaces. Then we can state the following theorem.

**Theorem 4.** Let  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M$  be pairwise linearly independent vectors in  $\text{GF}^N(|\mathcal{A}|)$  where  $M = \frac{|\mathcal{A}|^{N-1}}{|\mathcal{A}|-1}$ . Then the orthogonal direct sum of the subspaces  $\mathcal{V}_S^{\mathbf{a}_1}, \mathcal{V}_S^{\mathbf{a}_2}, \dots, \mathcal{V}_S^{\mathbf{a}_M}$  is

equal to  $\mathcal{V}_S$ . In other words

$$\mathcal{V}_S = \bigoplus_{i=1}^M \mathcal{V}_S^{\mathbf{a}_i}. \quad (15)$$

*Proof:* The orthogonal direct sum of subspaces is a subspace. Hence, the right hand side of the equation above is a subspace of  $\mathcal{V}_S$  and its dimension is given as

$$\dim \bigoplus_{i=1}^M \mathcal{V}_S^{\mathbf{a}_i} = \sum_{i=1}^M \dim \mathcal{V}_S^{\mathbf{a}_i} = |\mathcal{A}|^N - 1 \quad (16)$$

due to Theorem 1. As explained in Section II-A the dimension of the  $\mathcal{V}_S$  is also  $|\mathcal{A}|^N - 1$ . Consequently,  $\mathcal{V}_S = \bigoplus_{i=1}^M \mathcal{V}_S^{\mathbf{a}_i}$ . ■

This theorem has important consequences. Any joint PMF  $p(\mathbf{x})$  can be projected onto the subspaces  $\mathcal{V}_S^{\mathbf{a}_i}$ s by using the inner product. Theorem 4 states that the *vector summation* of these projections is equal to the original joint PMF. In other words

$$\begin{aligned} p(\mathbf{x}) &= p_{\mathbf{a}_1}(\mathbf{x}) \boxplus p_{\mathbf{a}_2}(\mathbf{x}) \boxplus \dots \boxplus p_{\mathbf{a}_M}(\mathbf{x}) \\ &= \frac{1}{Z} \prod_{i=1}^M p_{\mathbf{a}_i}(\mathbf{x}) \end{aligned} \quad (17)$$

where the last line follows from the definition of the  $\boxplus$  operation and  $p_{\mathbf{a}_i}(\mathbf{x})$  denotes the projection of  $p(\mathbf{x})$  onto the subspace  $\mathcal{V}_S^{\mathbf{a}_i}$ . These projections can be calculated by

$$p_{\mathbf{a}_i}(\mathbf{x}) = \sum_{i=1}^{|\mathcal{A}|-1} \langle p(\mathbf{x}), \psi_{ij}(\mathbf{x}) \rangle \square \psi_{ij}(\mathbf{x}), \quad (18)$$

where  $\psi_{ij}(\mathbf{x})$  denotes the  $j^{\text{th}}$  orthonormal basis PMF of the  $i^{\text{th}}$  subspace. Moreover, since  $p_{\mathbf{a}_i}(\mathbf{x})$ s are SPCIs we can write  $p(x)$  as

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^M q_i(\mathbf{a}_i \mathbf{x}^T), \quad (19)$$

where all scaling coefficients are merged in  $Z$  and  $q_i(\mathbf{a}_i \mathbf{x}) = |\mathcal{A}|^{N-1} p_{\mathbf{a}_i}(\mathbf{x})$ .

*Example 2.* Consider the  $p_2(x_1, x_2)$  given in Example 1. It can be factorized as

$$p_2(x_1, x_2) = \frac{1}{Z} q_1(x_1) q_2(x_2) q_3(x_1 + x_2) q_4(x_1 + 2x_2)$$

where  $q_1(x) = \frac{1}{10}[6 \ 3 \ 1]$ ,  $q_2(x) = \frac{1}{3}[1 \ 1 \ 1]$ ,  $q_3(x) = \frac{1}{6}[4 \ 1 \ 1]$ , and  $q_4(x) = \frac{1}{10}[6 \ 1 \ 3]$ . Actually, we can omit writing  $q_2(x_2)$  since it is a constant.

## B. Parity Check Interactions

Any SPCI can be transformed into usual parity check factor function, which is nothing but an indicator function, by employing an auxiliary variable in  $\text{GF}(|\mathcal{A}|)$  as follows.

$$\frac{q(\mathbf{ax}^T)}{|\mathcal{A}|^{N-1}} = \frac{1}{|\mathcal{A}|^{N-1}} \sum_{\forall u \in \text{GF}(|\mathcal{A}|)} I(\mathbf{ax}^T - u) q(u), \quad (20)$$

where  $I(x)$  is the indicator function and its value is one if  $x = 0$  and zero otherwise. This transformation allows expressing

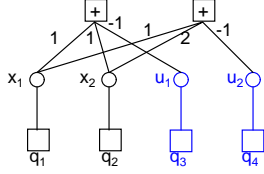


Fig. 1. Tanner graph of  $p_2(x_1, x_2)$  given in Examples 1,2, and 3.

any joint PMF as a product of parity check factors and factors of degree one.

$N$  of the parity check coefficient vectors of the SPCIs in (19) can be selected as the  $N$  canonical basis vectors of  $\text{GF}^N(|\mathcal{A}|)$ . Then the product in (19) can be grouped as

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^N q_i(x_i) \prod_{i=N+1}^M q_i(\mathbf{a}_i \mathbf{x}^T). \quad (21)$$

The second product above can be transformed into parity check constraints using (20) as follows.

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^N q_i(x_i) \sum_{\mathbf{u}} \prod_{i=N+1}^M I(\mathbf{a}_i \mathbf{x}^T - u_{i-N}) q_i(u_{i-N}),$$

where  $\mathbf{u}$  denotes  $(u_1, u_2, \dots, u_{M-N})$ . Let  $r(\mathbf{x}, \mathbf{u})$  be a PMF defined over  $\text{GF}^M(|\mathcal{A}|)$  as follows.

$$r(\mathbf{x}, \mathbf{u}) = \frac{1}{Z} \left( \prod_{i=1}^N q_i(x_i) \right) \left( \prod_{i=N+1}^M q_i(u_{i-N}) \right) \cdot \left( \prod_{i=N+1}^M I(\mathbf{a}_i \mathbf{x}^T - u_{i-N}) \right) \quad (22)$$

Clearly,  $p(\mathbf{x}) = \sum_{\mathbf{u}} r(\mathbf{x}, \mathbf{u})$ . Hence,  $r(\mathbf{x}, \mathbf{u})$  carries all the information that  $p(\mathbf{x})$  has for  $x_i$ 's. As (22) displays,  $r(\mathbf{x}, \mathbf{u})$  can be expressed as a product of parity check factors and factors of degree one which was our goal. Note that this factorization can be represented by a Tanner graph.

*Example 3.* The Tanner graph of  $p_2(x_1, x_2)$  in the previous examples is shown in Figure 1 which represents the following factorization.

$$r(x_1, x_2, u_1, u_2) = I(x_1 + x_2 - u_1) I(x_1 + 2x_2 - u_2) \cdot q_1(x_1) q_2(x_2) q_3(u_1) q_4(u_2).$$

#### IV. UNIVERSAL MARGINALIZATION MACHINE

The third product in (22) represents parity check constraints imposed by a linear code. The value of this product evaluates as

$$\prod_{i=N+1}^M I(\mathbf{a}_i \mathbf{x}^T - u_{i-N}) = \begin{cases} 1, & \mathbf{H}[\mathbf{x} \ \mathbf{u}]^T = 0 \\ 0, & \text{otherwise} \end{cases}$$

where the matrix  $\mathbf{H}$  is

$$\mathbf{H} = \begin{bmatrix} \mathbf{a}_{N+1} & -1 & 0 & \cdots & 0 \\ \mathbf{a}_{N+2} & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_M & 0 & 0 & \cdots & -1 \end{bmatrix} = [\mathbf{P} \quad -\mathbf{I}]. \quad (23)$$

The generator matrix  $\mathbf{G}$  of this code is  $[\mathbf{I} \ \mathbf{P}^T]$ . Remember that the vectors  $\mathbf{a}_{N+1}, \mathbf{a}_{N+2}, \dots, \mathbf{a}_M$  were all pairwise linearly independent. Moreover, these vectors are also linearly independent with the columns of the identity matrix, since the weights of these vectors are two or more. Hence, all columns of  $\mathbf{G}$  are pairwise linearly independent, which means that  $\mathbf{G}$  is the parity check matrix of a Hamming code. Therefore,  $\mathbf{H}$  is the parity check matrix of the  $(\frac{|\mathcal{A}|^N - 1}{|\mathcal{A}| - 1}, N)$  *dual Hamming code*.

If a soft decoder for this code existed, which gives the exact marginal a posteriori PMFs for each code symbol, then this soft decoder can be utilized to compute the marginal PMFs of  $N$  random variables having *any* joint PMF. Hence, we call such a soft detector as the universal marginalization machine (UMM). The UMM can be configured to marginalize a joint PMF by applying certain  $q_i(x_i)$ 's and  $q_i(u_{i-N})$ 's as inputs to the UMM.

This approach shows that the marginalization sum, which is the central part of the many communication problems, can be handled by a soft decoder. This is an important result in a practical point of view, since soft decoders can be approximated by analog VLSI structures [6]. For instance an analog equalizer can be implemented in this way.

#### V. CONCLUSION AND FUTURE DIRECTIONS

In this paper we have presented a method for factorizing joint PMFs into parity check factors. This factorization allows marginalizing a joint PMF by the soft decoder of the dual Hamming code if a Galois field exists in the order of the alphabet size of the PMF.

This work may be continued by extending the idea to the alphabet sizes for which a Galois field does not exist. Another interesting topic to work on might be employing the fast Fourier transform algorithm for obtaining the projections.

#### REFERENCES

- [1] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor Graphs and the Sum-Product Algorithm", IEEE Transactions on Information Theory, vol.47, No.2, pp.498-519 February 2001
- [2] H. A. Loeliger, "An Introduction to Factor Graphs", IEEE Signal Processing Magazine, Vol. 21, Issue 1, pp.28-41 Jan. 2004
- [3] M. F. Bayramoglu and A. Ö. Yilmaz, "A Hilbert Space of Probability Mass Functions and Applications on the Sum-Product Algorithm", Proc. 5<sup>th</sup> Int. Symp. On Turbo Codes, pp.338-343, Lausanne, Sept. 2008
- [4] L. Barnault and D. Declercq, "Fast Decoding Algorithms for LDPC over  $\text{GF}(2^q)$ ", Proc. ITW2003, pp.70-73, Paris, April 2003
- [5] Richard E. Blahut, "Algebraic Codes for Data Transmission", Cambridge Univ. Press 2003
- [6] H.-A. Loeliger, F. Lustenberger, M. Helfenstein, and F. Tarkoy, "Probability Propagation and Decoding in Analog VLSI", IEEE Tran. on Information Theory, pp.837-843, February 2001