Hierarchical Pairwise Error Probability for Hierarchical Decode and Forward Strategy in PLNC

Jan Sykora

***Abstract*—Pairwise error probability (PEP) is an essential tool allowing to design practical optimized coding schemes. It reveals the connection between the performance and the decoding metric that is directly related to the codeword and/or constellation properties. The application of this principle to the *hierarchical* PEP used to describe the decoding performance of *many-to-one* message functions (hierarchical network code maps) in physical layer network coding (PLNC) is addressed in this letter. Unlike for the single-user case, the *hierarchical PEP* reveals a complicated dependence on the structure of the hierarchical codeword/constellation. The structure is defined in terms of *hierarchical distance* and *hierarchical self-distance spectra*. We show that the network coded modulation (NCM) minimizing the hierarchical decoding error probability should have zero self-distance spectrum leading to *self-folded NCM design criterion*.**

***Index Terms*—Hierarchical decode and forward, pairwise error probability, physical layer network coding.**

I. Introduction

PLNC *Background:* Physical Layer Network Coding (PLNC) communication networks deliver the information from sources to destinations through the complex network of relays. The information flow does not route individual source messages but their many-to-one functions, similarly to Network Coding. But in addition to this, the signals of multiple transmitting nodes are allowed to interact at the receiving relay and the relay performs all processing directly in the constellation space. The extraction of the many-to-one message function out of the interacting signals can be done in multiple *hierarchical* encapsulation levels, and we name it *hierarchical* information. Network Coded Modulation (NCM) denotes the set of component constellation space codebooks aware of the network structure and designed for a particular relay strategy. Relay strategies can have many forms: Compute and Forward (CF), Denoising, Compress and Forward, Hierarchical Decode and Forward (HDF), etc. Final destination node determines the desired message by solving the decoding task on multiple collected received signals carrying hierarchical messages with independent many-to-one message functions. The overview of selected strategies can be found in [1]–[7], and selected results related to the error rate performance are in [8], [9].

Motivation: The error probability is an important performance indicator and also an obvious code and receiver design optimization goal. The exact error probability evaluation based on the transition probabilities is too complex (apart of trivial uncoded cases) to be practically useful. It also gives only a limited insight for a *synthesis* of the code. A pairwise error probability can be used to upper-bound the true error rate. As a side-effect it also *connects* the performance target with the metric used by the demodulator and decoder. This can be used for the code synthesis.

This letter focuses on the HDF relay strategy [5] in a single-stage H-MAC channel.[1] The relay wants to decode the H-message. Multiple combinations of component codewords correspond to one H-message. As a consequence, the hierarchical symbol/codeword, is generally *a set of multiple* constellation points or codewords $\mathcal{U}(b)$. It is called called H-constellation.[2] The fact that H-constellation (or H-codeword) has *multiple* representations of the H-message makes the pair-wise error analysis *substantially more difficult.* Also the *interpretation* of the results reveals some surprising facts which do *not* have their equivalents in the classical single user system.

Contribution and Main Results: We will define hierarchical pairwise error probability and we will show how this can be used in the isomorphic layered NCM design. The evaluation of the hierarchical error probability on the relay implies *H-message relay decoding strategy*, i.e. HDF.

(1) We derive the expression for *Hierarchical Pair-wise Error Probability (H-PEP)* revealing its dependence on the structure of the hierarchical constellation. The structure is defined in terms of *hierarchical distance* and *hierarchical self-distance spectra*.
(2) We show that the NCM minimizing the hierarchical decoding error probability should have *zero* self-distance spectrum and we call this case *self-folded NCM*. The self-folding property provides a neat NCM design criterion.

II. System Model and Definitions

H-MAC Stage: In order to streamline the development of this letter, we focus on a generic *single* H-MAC stage with a *single* receiving relay (Fig. 1). It is a smallest building block of a more complex PLNC system. The component messages $b_k \in [1:M]$ of all transmitting nodes are encoded by NCM into $\mathbf{c}_k = \mathcal{C}_k(b_k)$, $k \in [1:K]$, where K is the number of the transmitters in H-MAC. The relay wants to decode H-message which is Hierarchical Network Code (HNC) map (many-to-one function) $b = \chi(\tilde{b})$ of the component messages, where the concatenation of all component messages is $\tilde{b} = [b_1, \ldots, b_K]$. The "tilde" denotes "concatenation" and will also be used in the notations of codewords $\tilde{\mathbf{c}} = [\mathbf{c}_1^T, \ldots, \mathbf{c}_K^T]^T$, symbols $\tilde{c} =$

Manuscript received July 1, 2015; revised September 29, 2015; accepted October 30, 2015. Date of publication November 5, 2015; date of current version January 7, 2016. This work was supported in part by the FP7-ICT DIWINE project, and in part by the MSMT LD12062. The associate editor coordinating the review of this paper and approving it for publication was X. Zhou.

The author is with Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic (e-mail: jan.sykora@fel.cvut.cz).

Digital Object Identifier 10.1109/LCOMM.2015.2498164

[1] A prefix "H-" is used to denote "hierarchical" entity, i.e. the one that is generally many-to-one function of the components, or a processing (e.g. codebook, codeword, MAC channel, etc.) related to the hierarchical entities.

[2] A trivial uncoded example has two component BPSK sources with symbols $\{\pm 1\}$. The channel combined constellation points are $\{-2, 0, 2\}$. For XOR type of many-to-one hierarchical function, there are two H-symbols, one represented by $\{0\}$ and second one being a pair of possible points $\{\pm 2\}$.

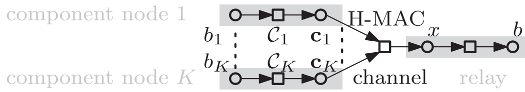

Fig. 1. System model for H-MAC Stage.

$[c_1^T, \ldots, c_K^T]^T$, codebooks $\tilde{\mathcal{C}} = [\mathcal{C}_1, \ldots, \mathcal{C}_K]$. The observation is described by $p(\mathbf{x}|\tilde{b}) = p(\mathbf{x}|\tilde{\mathbf{c}}(\tilde{b}))$, where $\mathbf{x} = \mathbf{u}(\tilde{\mathbf{c}}) + \mathbf{w}$, $\mathbf{u}(\tilde{\mathbf{c}})$ is the channel-combined constellation symbol depending on all component codewords $\tilde{\mathbf{c}}$, and $\mathbf{w}$ is Gaussian noise.

Isomorphic Layered NCM: We call the set of component codebooks $\{\mathcal{C}_k\}_{k=1}^K$ *isomorphic layered NCM* if there exists one-to-one mapping H-code $\mathcal{C}$ and H-codeword HNC map $\mathbf{c} = \chi_c(\tilde{\mathbf{c}})$ such that $\mathbf{c} = \mathcal{C}(b)$. The isomorphism implies that the component-wise encodings can be equivalently expressed for the relation between the desired H-message and the corresponding H-codeword. Since the H-decoding can be based on the decoding metric related to the H-codeword which is first obtained from the component-wise channel observation model $p(\mathbf{x}|\tilde{\mathbf{c}})$, we call that approach also layered. The examples of isomorphic layered NCM are CF [1] or layered block linear NCM [5].

The analysis in this letter can be generally applied on the level of complete messages b, codewords $\mathbf{c}$ and vector observation $\mathbf{x}$ but it could also be applied in the symbol-wise manner (c_n, x_n), e.g. for the uncoded case. We will use a generic notation using b, c, x to cover both.

III. Hierarchical Pairwise Error Probability

A. H-PEP for Isomorphic NCM

H-PEP Definition: Assume the H-message decoding metric is $\mu_b(x)$ and that the decoder *decision* processing *maximizes* its value, $\hat{b} = \arg\max_b \mu_b(x)$. H-PEP is $P_{2H}(b'|b) = \Pr\{\mu_b(x) < \mu_{b'}(x)|b = \chi(\tilde{b}_x)\}$ where $\tilde{b}_x$ are *actually* transmitted component data and $b \neq b'$ are some given H-messages. We will also use a simplified notation $P_{2H} = \Pr\{\mu_b < \mu_{b'}|b\}$.

The H-PEP is thus the probability that the H-metric for correct H-message b is smaller than the one for some other message b' provided that the received signal is consistent with b, i.e. all component transmitted data are such that $b = \chi(\tilde{b}_x)$. The form of the metric used in H-PEP is arbitrary. It does not even need to be the metric leading to the optimal performance. In such a case it would simply analyze the performance under that suboptimal metric and potentially suggest how to optimize the code for that given (suboptimal) metric. The most common example is the MAP metric.

H-PEP for Isomorphic NCM: The evaluation of H-PEP related directly to the *message* level HNC map metric is a difficult task. It becomes much easier when the metric is related to the *code* level HNC map. It directly employs the codewords into the calculation which will provide better insight on what the code should optimally look like. If the NCM is *isomorphic* then $P_{2H}(b'|b) = P_{2H}(c'|c)$ where $c = \mathcal{C}(b)$ and $c' = \mathcal{C}(b')$. From now on, we will assume isomorphic NCM and thus

$$P_{2H}(c'|c) = \Pr\{\mu_c(x) < \mu_{c'}(x)|c\}. \tag{1}$$

We will also assume the use of *MAP decoding metric*. The H-metric is a marginalization

$$\mu_c(x) = p(x|c) = \frac{1}{p(c)} \sum_{\tilde{c}:c} p(x|\tilde{c})p(\tilde{c}) \tag{2}$$

where we used notation $(\tilde{c} : c) = \{\tilde{c} : c = \chi_c(\tilde{c})\}$. Notice a very important fact that the step from message H-PEP evaluation to codeword based one (which is trivial in a single user case) requires a *specific assumption and further treatment*.

In (1), we first focus on the conditioning by the received signal consistent with c. There are multiple of component codewords $\tilde{c}_x$ in the received signal consistent with c $P_{2H} = \Pr\{\mu_c(x) < \mu_{c'}(x)| \bigcup_{\tilde{c}_x:c} \tilde{c}_x\}$. Events $\tilde{c}_x$ are *disjoint*. If we assume that they are also *equally probable* $\Pr\{\tilde{c}_x : c\} =$ const then[3]

$$P_{2H} = \frac{1}{M_{\tilde{c}:c}} \sum_{\tilde{c}_x:c} \Pr\{\mu_c(x) < \mu_{c'}(x)|\tilde{c}_x\} \tag{3}$$

where $M_{\tilde{c}:c}$ is the size of sub-codebook $\tilde{C}(c)$. $\tilde{C}(c)$ is a subset of $\tilde{C}$ where we take only the entries consistent with c. The H-PEP is the average over all consistent source node component codes. This is additional level of averaging over those being present in the traditional pairwise error probability calculation.

The Most Probable Event: We can upper-bound the H-PEP by *the most probable pairwise H-constellation event*

$$P_{2H} \leq P_{2H_m} = \max_{\tilde{c}_x:c} \Pr\{\mu_c(x) < \mu_{c'}(x)|\tilde{c}_x\}. \tag{4}$$

Notice that the hierarchical codewords c, c' are still fixed. The overall error rate behavior can be then (similarly as in classical single user code case) upper-bounded by the *overall* most probable pairwise event

$$P_{2H_{\max}} = \max_{c \neq c', \tilde{c}_x:c} \Pr\{\mu_c(x) < \mu_{c'}(x)|\tilde{c}_x\}. \tag{5}$$

B. H-PEP for Gaussian Memoryless Channel

Gaussian Channel: In the next step, we will constrain the treatment to a special case of *Gaussian memoryless channel*. The observation likelihood for *all component* codes is $p(x|\tilde{c}) = \frac{1}{\pi^m \sigma_w^{2m}} \exp\left(-\|x - u(\tilde{c})\|^2/\sigma_w^2\right)$ where m is a complete dimensionality of the signals (both per-symbol and length of the message), σ_w^2 is the variance per dimension, and $u(\tilde{c})$ is noiseless *constellation space point* observed at the receiver. Notice that channel model inherently contained in $u(\tilde{c})$ can be *arbitrary*.[4] *H-constellation* is the set of $u(\tilde{c})$ consistent with c, i.e. $\mathcal{U}(c) = \{u(\tilde{c}) : c = \chi_c(\tilde{c})\}$ where we properly count for multiplicities.

On top of assuming uniformly distributed $\tilde{c}$, $\Pr\{\tilde{c}\} = 1/M_{\tilde{c}}$, $M_{\tilde{c}} = |\tilde{\mathcal{C}}|$, we also assume *uniform* c, i.e. $\Pr\{c\} = \text{const} = 1/M_c$ where $M_c = |\mathcal{C}|$ is the size of H-codebook. Then the metric is

$$\mu_c = \frac{M_c}{M_{\tilde{c}}} \sum_{\tilde{c}:c} p(x|\tilde{c}) = \frac{M_c}{M_{\tilde{c}}} \frac{1}{\pi^m \sigma_w^{2m}} \sum_{\tilde{c}:c} e^{-\frac{1}{\sigma_w^2}\|x - u(\tilde{c})\|^2} \tag{6}$$

and its normalized form $\dot{\mu}_c = \mu_c \pi^m \sigma_w^{2m} M_{\tilde{c}}/M_c$ is

$$\dot{\mu}_c = \sum_{\tilde{c}:c} \exp\left(-\|x - u(\tilde{c})\|^2/\sigma_w^2\right). \tag{7}$$

[3]For multiple *disjoint* and *equally probable* events B_i, it holds $\Pr\left\{A| \bigcup_{i=1}^M B_i\right\} = \frac{1}{M}\sum_{i=1}^M \Pr\{A|B_i\}$.

[4]In a special case of a linear flat fading channel, used now *only* as an example and not needed for the rest of the derivation, it would be $u = \sum_k h_k c_k$.

H-distance: The H-distance for the complete message is $u^{\mathrm{Hmin}}(c) = \arg\min_{u(\tilde{c}):\chi_c(\tilde{c})=c} \|x - u(\tilde{c})\|^2$. We can factorize (7)

$$\dot{\mu}_c = e^{-\frac{1}{\sigma_w^2}\|x-u^{\mathrm{Hmin}}(c)\|^2} \sum_{\tilde{c}:c} e^{-\frac{1}{\sigma_w^2}\left(\|x-u(\tilde{c})\|^2-\|x-u^{\mathrm{Hmin}}(c)\|^2\right)} \quad (8)$$

where all differences in the summation are *non-negative* $\|x - u(\tilde{c})\|^2 - \|x - u^{\mathrm{Hmin}}(c)\|^2 \geq 0$. Finally, taking the negative scaled logarithm $\rho_c = -\sigma_w^2 \ln \dot{\mu}_c$, we get the decoder metric

$$\rho_c = \|x - u^{\mathrm{Hmin}}(c)\|^2 - \sigma_w^2 \eta_c \quad (9)$$

where the correction term is

$$\eta_c = \ln \sum_{\tilde{c}:c} e^{-\frac{1}{\sigma_w^2}\left(\|x-u(\tilde{c})\|^2-\|x-u^{\mathrm{Hmin}}(c)\|^2\right)}. \quad (10)$$

Clearly, it holds $0 \leq \eta_c \leq \ln(M_{\tilde{c}}/M_c)$. The correction term is zero $\eta_c = 0$ if exactly one H-constellation point is the minimal H-distance point $u^{(1)}(\tilde{c}) = u^{\mathrm{Hmin}}(c)$ and all others (if any) are at much larger distance, $\forall i \neq 1$, $\|x - u^{(i)}(\tilde{c})\|^2 \gg \|x - u^{\mathrm{Hmin}}(c)\|^2$. Notice that many practical component constellations (e.g. 2 component BPSK and XOR HNC with H-constellation $\{0, \{-2, 2\}\}$), do not fit under these conditions. On the other side, if for all $\tilde{c} : c$ the H-constellation point is the minimal H-distance point, then the correction is non-zero but *constant* and *independent* of x, $\eta_c = \ln(M_{\tilde{c}}/M_c)$. The non-zero value of η_c on its own does not present a problem from the H-PEP evaluation point of view. However its dependence on x is a problem. The presence and behavior of the correction term also nicely demonstrates that the pure H-distance is *not* generally optimal decoding metric.

H-PEP: The H-PEP for a given set of messages and for c-consistent received signal ($\chi_c(\tilde{c}_x) = c$) is then

$$\begin{aligned} P_{2H_e} &= \Pr\{\rho_c(x) > \rho_{c'}(x)|\tilde{c}_x\} \\ &= \Pr\{\|x - u^{\mathrm{Hmin}}(c)\|^2 - \|x - u^{\mathrm{Hmin}}(c')\|^2 \\ &\quad - \sigma_w^2 \eta_{c,c'} > 0|\tilde{c}_x\} \end{aligned} \quad (11)$$

where $\eta_{c,c'} = \eta_c - \eta_{c'}$. The inequality is reversed since we used negative scaled logarithm metric. Let us now denote the c-consistent noiseless part of the received signal for given $\tilde{c}_x$ as $u(\tilde{c}_x)$. It must be a member of the H-constellation set for the c H-symbol, i.e. $u(\tilde{c}_x) \in \mathcal{U}(c)$. The condition of c-consistent received signal is thus reflected in having $x = u(\tilde{c}_x) + w$ and consequently

$$\begin{aligned} P_{2H_e} = \Pr\Big\{&\|u(\tilde{c}_x) + w - u^{\mathrm{Hmin}}(c)\|^2 \\ &- \|u(\tilde{c}_x) + w - u^{\mathrm{Hmin}}(c')\|^2 - \sigma_w^2 \eta_{x,c,c'} > 0\Big\} \end{aligned} \quad (12)$$

where the correction terms under this condition are $\eta_{x,c,c'} = \eta_{x,c} - \eta_{x,c'}$

$$\eta_{x,c} = \ln \sum_{\tilde{c}:c} e^{-\frac{1}{\sigma_w^2}\left(\|u(\tilde{c}_x)+w-u(\tilde{c})\|^2-\|u(\tilde{c}_x)+w-u^{\mathrm{Hmin}}(c)\|^2\right)}, \quad (13)$$

$$\eta_{x,c'} = \ln \sum_{\tilde{c}':c'} e^{-\frac{1}{\sigma_w^2}\left(\|u(\tilde{c}_x)+w-u(\tilde{c}')\|^2-\|u(\tilde{c}_x)+w-u^{\mathrm{Hmin}}(c')\|^2\right)}. \quad (14)$$

The expression of the distances difference that appears in P_{2H_e} (and with minor modification in $\eta_{x,c}$, $\eta_{x,c'}$) can be further manipulated ($\langle \cdot; \cdot \rangle$ denotes inner product)

$$\begin{aligned} &\|u(\tilde{c}_x) + w - u^{\mathrm{Hmin}}(c)\|^2 - \|u(\tilde{c}_x) + w - u^{\mathrm{Hmin}}(c')\|^2 \\ &= \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c)\|^2 - \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c')\|^2 \\ &+ 2\Re\left[\left\langle u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c); w\right\rangle\right] - 2\Re\left[\left\langle u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c'); w\right\rangle\right] \\ &= \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c)\|^2 - \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c')\|^2 \\ &- 2\Re\left[\left\langle u^{\mathrm{Hmin}}(c) - u^{\mathrm{Hmin}}(c'); w\right\rangle\right]. \end{aligned} \quad (15)$$

$\xi = -2\Re\left[\left\langle u^{\mathrm{Hmin}}(c) - u^{\mathrm{Hmin}}(c'); w\right\rangle\right]$ is Gaussian real-valued scalar zero-mean random variable with the variance $\sigma_\xi^2 = 2\sigma_w^2 \|u^{\mathrm{Hmin}}(c) - u^{\mathrm{Hmin}}(c')\|^2$. Then

$$P_{2H_e} = \Pr\left\{\xi > \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c')\|^2 - \|u(\tilde{c}_x) - u^{\mathrm{Hmin}}(c)\|^2 + \sigma_w^2 \eta_{c,c'}\right\}. \quad (16)$$

A similar manipulation can be done for the correction terms

$$\eta_{x,c} = \ln \sum_{\tilde{c}:c} e^{-\frac{1}{\sigma_w^2}\left(\|u(\tilde{c}_x)-u(\tilde{c})\|^2-\|u(\tilde{c}_x)-u^{\mathrm{Hmin}}(c)\|^2-2\Re\left[\left\langle u(\tilde{c})-u^{\mathrm{Hmin}}(c);w\right\rangle\right]\right)}, \quad (17)$$

$$\eta_{x,c'} = \ln \sum_{\tilde{c}':c'} e^{-\frac{1}{\sigma_w^2}\left(\|u(\tilde{c}_x)-u(\tilde{c}')\|^2-\|u(\tilde{c}_x)-u^{\mathrm{Hmin}}(c')\|^2-2\Re\left[\left\langle u(\tilde{c}')-u^{\mathrm{Hmin}}(c');w\right\rangle\right]\right)}. \quad (18)$$

C. Hierarchical Distance and Self-Distance Spectrum

The properties of the quantities determining the H-PEP clearly depend on two types of the H-constellation/codeword distances. The first one is the distance between the points belonging to *different* H-symbols and the second one is the distance between the points belonging to the *same* H-symbol. For this purpose, we define hierarchical distance and self-distance spectrum.

Hierarchical distance (H-distance) spectrum is a set

$$\mathcal{S}_H(c, c') = \left\{\|u(\tilde{c}) - u(\tilde{c}')\|^2 : c = \chi_c(\tilde{c}) \neq c' = \chi_c(\tilde{c}')\right\}. \quad (19)$$

We also define $\mathcal{S}_H = \bigcup_{c,c'} \mathcal{S}_H(c, c')$.

Hierarchical self-distance (H-self-distance) spectrum is a set

$$\mathcal{S}_{\bar{H}}(c) = \left\{\|u(\tilde{c}^{(a)}) - u(\tilde{c}^{(b)})\|^2 : c = \chi_c(\tilde{c}^{(a)}) = \chi_c(\tilde{c}^{(b)}) \wedge \tilde{c}^{(a)} \neq \tilde{c}^{(b)}\right\}. \quad (20)$$

We also define $\mathcal{S}_{\bar{H}} = \bigcup_c \mathcal{S}_{\bar{H}}(c)$.

D. NCM Design Rules Based on H-PEP

We use (16, 17, 18) to establish qualitative design rules for NCM that minimize H-PEP. The situation is however less straightforward than in the classical single user code. There are several observations we need to keep in our mind before we start. There are multiple mutually correlated random variables in the expression and these cannot be easily factorized into a single one as in single-user code case. All ξ, $\eta_{x,c}$, $\eta_{x,c'}$ directly depend on Gaussian noise w and are continuos valued correlated variables. But also the hierarchical minimum distance points $u^{\mathrm{Hmin}}(c)$, $u^{\mathrm{Hmin}}(c')$ depend on the received signal and therefore also on w. These variables are however discrete one. There are random and dependent on w but *constrained* to be inside the H-constellation and their influence on H-PEP can be thus controlled through the H-distance and H-self-distance spectrum. In order to *minimize* H-PEP, we should consider the following.

(1) The distance $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c')\|^2$ in (16) should be as large as possible. Notice that $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c')\|^2 \in \mathcal{S}_H(c, c')$ for arbitrary noise w realization.

(2) The self-distance $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c)\|^2$ in (16) should be as small as possible. Notice that $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c)\|^2 \in \mathcal{S}_{\bar{H}}(c)$ for arbitrary noise w realization.

(3) The variance of the ξ variable is proportional to $\|u^{\text{Hmin}}(c) - u^{\text{Hmin}}(c')\|^2 \in \mathcal{S}_H(c, c')$ which is constrained by the H-distance spectrum.

(4) The correction term $\eta_{x,c,c'}$ should be as large as possible which in turn means maximizing $\eta_{x,c}$ and minimizing $\eta_{x,c'}$. Behavior of $\eta_{x,c}$ is dictated by H-self-distance spectrum while the behavior of $\eta_{x,c'}$ is jointly dictated by both H-distance and H-self-distance spectrum.

(5) The maximum value of $\eta_{x,c}$ is $\ln(M_{\tilde{c}}/M_c)$ and it is reached when arguments of the exponentials are zero, i.e. when all self-distances are zero $\mathcal{S}_{\bar{H}}(c) = \{0\}$. All H-constellation/codeword points for given c are *identical*. We will call this a *self-folded* H-constellation/codebook or *self-folded NCM*. If the H-constellation/codebook is self-folded then the arguments of the exponentials in $\eta_{x,c'}$ are also all zeros and thus $\eta_{x,c'} = \ln(M_{\tilde{c}}/M_c)$ and the overall correction term is zero regardless of the noise $\eta_{x,c,c'} = 0$. Self-folded NCM also causes the self-distance in (16) to be zero and thus

$$P_{2H_e}^{\text{SF}} = \Pr\left\{\xi > \|u(\tilde{c}_x) - u^{\text{Hmin}}(c')\|^2\right\}. \tag{21}$$

(6) Now let us have a look at the situation when the NCM is *not* self-folded. Let us assume that the spread in self-distances is symmetric for all c. If it was not a symmetric one then the case that would make an advantage for $P_{2H_e}(c'|c)$ would become a disadvantage for $P_{2H_e}(c|c')$ in terms of the possible compensation as discussed in point (6b) below.

(6a) Let us also assume that some point pair $\tilde{c}_x, \tilde{c}$ in the H-constellation maximizes the self-distance $\|u(\tilde{c}_x) - u(\tilde{c})\|^2$ to some particular value $d_{\bar{H}}^2$. The expression $\eta_{x,c}$ will not be the maximal one (as for self-folded case) but it will be somewhat smaller. For the given pair of points, the argument $\|u(\tilde{c}_x) - u(\tilde{c})\|^2$ of the exponential in (17) increases to the value $d_{\bar{H}}^2$. The second term $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c)\|^2$ will highly likely (at least for high SNR) be zero since the minimum H-distance point is the closest to the received signal. The degradation of the first noiseless term in (17) is thus $d_{\bar{H}}^2$ at least for that given point pair.

(6b) This degradation can be possibly compensated by the improvement in the term (18). In the most favorable case for the improvement, the points $u(\tilde{c}_x)$, $u(\tilde{c}')$, $u^{\text{Hmin}}(c')$ lie in the line and the maximal value of $\|u(\tilde{c}_x) - u(\tilde{c}')\|^2 - \|u(\tilde{c}_x) - u^{\text{Hmin}}(c')\|^2$ is $d_{\bar{H}}^2$. Where, by the assumption of the symmetry, points $u(\tilde{c}')$, $u^{\text{Hmin}}(c')$ are constrained to the distance $\|u(\tilde{c}') - u^{\text{Hmin}}(c')\|^2 = d_{\bar{H}}^2$. So the noiseless terms in the exponentials of (18) can, at the best, just compensate the degradation of the argument of (17) but practically it will be even worse.

(6c) The noise terms in both (17) and (18), i.e. $2\Re\left[\langle u(\tilde{c}) - u^{\text{Hmin}}(c); w\rangle\right]$ and $2\Re\left[\langle u(\tilde{c}') - u^{\text{Hmin}}(c'); w\rangle\right]$ are given by the self-distances only. The left-hand sides in the inner products are different but under the assumption of symmetric self-distances $\mathcal{S}_{\bar{H}}(c) \approx \mathcal{S}_{\bar{H}}(c')$ they will make the noise term highly correlated and thus both will be affecting the arguments of the exponentials in (17) and (18) the same way.

(6d) The main expression (16) also contains the self-distance. A positive value of $\|u(\tilde{c}_x) - u^{\text{Hmin}}(c)\|^2$ decreases the right-hand side of the inequality and increases the H-PEP.

(6e) As we see, the nonzero spread of the self-distances *cannot* improve the H-PEP and will make highly likely the things only worse.

IV. Conclusions and Final Results

The analysis of the H-PEP behavior lead us to the final main result identifying the self-folded NCM design criterion.

Conjecture: Self-folded NCM (H-constellation/codebook) Minimizes H-PEP: Assume isomorphic NCM in Gaussian memoryless channel, decoding MAP H-metric, and uniform component messages and HNC map such that $\Pr\{\tilde{c}\} = 1/M_{\tilde{c}}$, $\Pr\{c\} = 1/M_c$, $\Pr\{\tilde{c} : c\} = M_c/M_{\tilde{c}}$. Self-folded NCM, i.e. the one with zero H-self-distance spectrum $\mathcal{S}_{\bar{H}} = \{0\}$, minimizes H-PEP which is (Q is complementary Gaussian CDF)

$$P_{2H_e}^{\text{SF}} = Q\left(\sqrt{\|u(\tilde{c}_x) - u^{\text{Hmin}}(c')\|^2/(2\sigma_w^2)}\right). \tag{22}$$

It is important to note that the self-folding property is expected to be *natural*, i.e. naturally performed by the channel combining the component signals into the H-constellation also fully respecting the channel parameterization. Notice that the modulo-lattice preprocessing (used in CF) achieves the hierarchical self-folding but it achieves that by the *force*. The price paid for this enforcement is the distortion of the noise which becomes modulo-equivalent Gaussian.

References

[1] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[2] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 773–787, Jun. 2009.

[3] S. Fu, K. Lu, T. Zhang, Y. Qian, and H.-H. Chen, "Cooperative wireless networks based on physical layer network coding," *IEEE Wireless Commun.*, vol. 17, no. 6, pp. 86–95, Dec. 2010.

[4] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[5] J. Sykora and A. Burr, "Layered design of hierarchical exclusive codebook and its capacity regions for HDF strategy in parametric wireless 2-WRC," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3241–3252, Sep. 2011.

[6] J. Sykora and A. Burr, "Advances in wireless network coding—The future of cloud communications," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 1–190.

[7] S. C. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond," *Phys. Commun.*, vol. 6, pp. 4–42, Mar. 2013, doi: 10.1016/j.phycom.2012.05.002.

[8] Z. Chen, B. Xia, Z. Hu, and H. Liu, "Design and analysis of multi-level physical-layer network coding for Gaussian two-way relay channels," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1803–1817, Jun. 2014.

[9] A. Y.-C. Peng, S. Yousefi, and I.-M. Kim, "On error analysis and distributed phase steering for wireless network coding over fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 11, pp. 5639–5649, Nov. 2009.