Short-Packet Physical-Layer Network Coding

Shakeel Salamat Ullah, Soung Chang Liew, *Fellow, IEEE,* Gianluigi Liva, *Senior Member, IEEE,* and Taotao Wang, *Member, IEEE*

Abstract

This paper explores the application of physical-layer network coding (PNC) for short-packet transmissions. PNC can potentially reduce the communication delay in relay-assisted wireless networks and can thus be instrumental in realizing short-packet communication systems with stringent delay requirements. In this work, first, we first derive an achievability bound for channel-coded short-packet PNC systems. Based on the random-coding error-exponent, the bound serves as a benchmark for short-packet PNC operating with traditional preamble-aided channel estimation and XOR channel decoding. Second, we design a blind channel estimation algorithm and a code-aided channel estimation algorithm for short-packet PNC systems. Both outperform the traditional preamble-aided channel estimation for PNC systems operating with mismatched channel-state-information. As a case study, we compare three algorithms for packets of 128 symbols over a two-way relay channel. The results show that the blind algorithm outperforms the code-aided algorithm and preamble-aided algorithm by almost 0.2 and 1.5 dB respectively. Furthermore, the blind algorithm achieves the target packet error rate of 10^{-4} within 0.5 dB of the random coding bound of an imaginary system in which perfect channel-state-information is available at the relay at no cost (i.e., channel estimation is not required in the imaginary system). The bound and the algorithms give us a fundamental framework for applying PNC to short-packet transmissions.

Index Terms

Physical-layer network coding, short-packet transmissions, random coding bound, mismatched channel-state-information, expectation-maximization belief-propagation.

I. Introduction

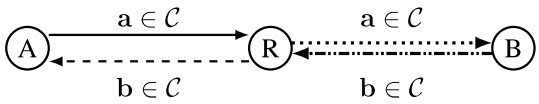
S. Salamat Ullah, S. C. Liew, and T. Wang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Email: {ssullah, soung, ttwang}@ie.cuhk.edu.hk. Gianluigi Liva is with the Institute of Communication and Networking, German Aerospace Center (DLR), Germany. Email: Gianluigi.Liva@dlr.de.

Parts of the paper related to the random coding bound for physical-layer network coding (covered mostly in Section IV) were presented at *Proc. IEEE ITW* 2017 [1] and *Proc. IEEE WCNC* 2018 [2].

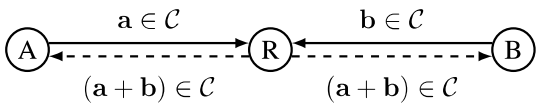
THIS paper investigates the application of physical-layer network coding (PNC) [3]–[5] for short-packet transmissions. Short-packet transmissions are envisioned to be a fundamental building block of the services that will be provided in future wireless communications [6]–[8]. Short-packet transmissions are required for applications that have to satisfy stringent delay constraints [9]–[12]. For such applications, there may be scenarios in which a node is not within range of the network [13] or in which a node experiences a low signal-to-noise ratio. To extend the network coverage or to improve the packet error rate performance at the boundary of the network, relays can be employed [14]–[16]. However, an additional hop increases the communication delay for two-way communication in relay-assisted systems (the assumption here is that packet lengths remain the same in both relay-assisted and non-relay-assisted systems). For such a scenario, PNC can be instrumental in reducing the communication delay.

To see that, consider the application of PNC in a two-way relay channel (TWRC) setting. In TWRC, two end users that are out of transmission range of each other exchange information via a relay. In a TWRC operated with PNC, the two end users transmit packets simultaneously to the relay in the uplink phase (see Fig. 1b). The relay performs PNC channel estimation and decoding on the superimposed received signal and broadcasts back a network-coded packet (i.e., an XOR codeword when the two end users employ binary codes) to the end users in the downlink phase. Upon receiving the network-coded packet, the end users subtract their self-information to obtain the intended information from the other user. PNC requires two time slots for the end users to deliver a packet to each other, whereas the traditional relaying requires four time slots (see Fig. 1a). Let us refer to the delay due to the exchange of two packets of the two users as the cycle time. Specifically, the cycle time is the interval between two transmission initiations by a source node. PNC has half the cycle time of conventional relaying in a TWRC.

The benefit of reduced cycle time is that the end users can can send information collected locally to each other more frequently and hence will have low "age of information" (a metric for data freshness at the receiver [17]–[19]) when PNC is employed. In the above, we assumed that there are already no packets in the queue of the end users at a given time slot. If packets arrive at each end user in a random and bursty manner, a buffer will be needed to queue up packets that cannot be immediately transmitted. In this case, the overall difference in delay between PNC and traditional relaying can be even more pronounced. In other words, besides the cycle time, packets may also incur queuing delay before they get to the head of the queue to be transmitted. PNC can potentially sustain a throughput twice that of traditional relaying due



(a) Conventional relaying. First time slot (⟶), second time slot (⋯►), third time slot (-⋯►), and fourth time slot (- -►).



(b) PNC. First time slot (⟶), and second time slot (- -►).

Fig. 1. Users A and B exchange their messages $\mathbf{a}$ and $\mathbf{b}$, respectively, belonging to a codebook $\mathcal{C}$ through a relay R in a TWRC.

to the reduction of four time slots to two time slots for one round of packet exchange. Each system will have a delay vs offered load (the traffic load entering the queue) curve with the delay going to infinity when the offered load approaches the sustainable throughput (see Fig. 2). In the figure, the normalized sustainable throughput of PNC and conventional relaying is 0.5 and 0.25 uplink packets per time slot respectively. Given the same offered load (even if this offered load is lower than the sustainable throughput of traditional relaying), packets in the PNC system will incur much shorter delay (a simple M/D/1 queuing model will readily indicate this result [20]).

The aforementioned delay advantage of PNC is attainable only if the transmissions are reliable and that packets seldom need to be retransmitted due to errors. To that end, the achievability bounds that provide benchmarks of packet error rate performance of PNC in the short packet-length regime need to be developed. In addition, the corresponding channel estimation and decoding algorithms that approach the benchmark performance, especially with mismatched channel-state-information (CSI) [21], [22], need to be designed. The performance with mismatched CSI needs to be considered since for short-packet transmissions, small preambles (or lack thereof) may compromise the accuracy of estimated CSI.

A. Contributions

This work addresses the issue of obtaining reliability benchmarks and the issue of designing

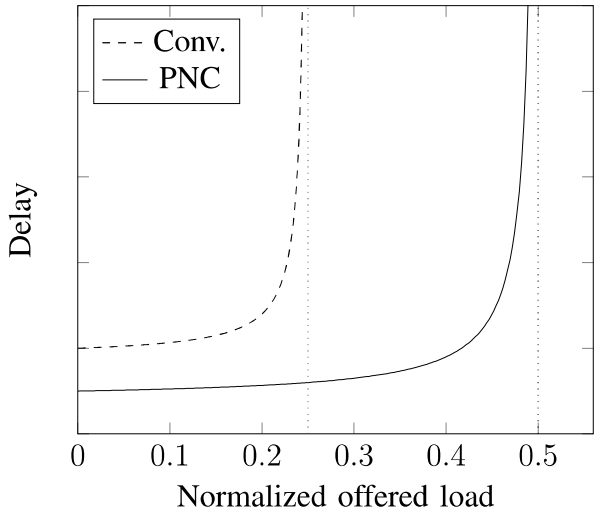


Fig. 2. Typical delay vs normalized offered load (i.e., number of uplink packets per time slot) of conventional relaying and PNC in a TWRC.

the channel estimation and decoding algorithms of PNC for short-packet communications. We focus on the uplink phase of PNC where nodes A and B transmit to the relay R simultaneously (see Fig. 1(b)). The signal in the downlink phase is not superimposed and is transmitted by relay R only (i.e., the conventional reliability benchmarks and estimation-decoding algorithms apply) [5]. A summary of our key contributions and results is as follows:

1) We obtain an achievability bound for decoding error probabilities based on random-coding error-exponent for the uplink phase of PNC in a two-way relay setup where the relay employs practical and low-complexity XOR channel decoding (XOR-CD). The bound is developed for the setting where the CSI is obtained from the preamble and fed to the channel decoder. Besides serving as a benchmark for packet-error rate performance, the random coding bound (RCB) also characterizes the trade-off between preamble-length and data-length for a given overall packet-length in the mentioned setting.
2) We design a blind channel estimator and a code-aided channel estimator for short-packet PNC systems operating with mismatched CSI and XOR-CD. For the blind channel estimation, we make use of the expectation-maximization (EM) algorithm [23] to obtain CSI estimate. No preamble symbols are required for the algorithm. For this method, channel estimation and decoding are performed in a disjoint manner, i.e., information is passed from the channel estimator to the XOR channel decoder only once, without any feedback

from the XOR channel decoder to the channel estimator for additional iterations. This is the first work in which blind channel estimation is applied to PNC. For the code-aided channel estimation, channel estimation and decoding are performed jointly, with iterations between the two components. We combine EM for channel estimation with belief-propagation (BP) [24] for XOR-CD. A small preamble is required for initialization of EM in this case. The two EM-based channel estimation and decoding algorithms have low complexity compared with those in the literature [25], [26], thanks to XOR-CD. Despite the lower complexity, the algorithms still achieve high reliability as they both make use of all the transmitted symbols for CSI estimation.

3) We numerically evaluate the bound and the algorithms for BPSK modulated PNC systems employing XOR-CD. For comparison, we also evaluate the performance of a PNC system employing traditional preamble-aided channel estimation and XOR-CD. We assume that packets of 128 symbols are transmitted over a TWRC. The results show that the blind algorithm outperforms the code-aided and preamble-aided counterparts by almost 0.2 and 1.5 dB respectively. Furthermore, the blind algorithm achieves the target packet error rate of 10^{-4} within 0.5 dB of the RCB of an imaginary system in which perfect CSI is available at the relay at no cost (i.e., channel estimation is not required in the imaginary system).

The rest of the paper is organized as follows. Section II presents the related work. System model is given in Section III. RCB for short-packet PNC systems employing XOR-CD is derived in Section IV. Section V presents a blind algorithm and a code-aided algorithm for channel estimation in PNC systems employing XOR-CD. Computational complexity of the algorithms is analyzed in Section VI. Numerical results are provided in Section VII. Finally, Section VIII concludes the paper.

II. Related Work

Prior works [27]–[30] regarding reliability benchmarks of PNC provide little insight as they did not consider short-packet transmissions. The only exception is the work in [31] in which an error-exponent is derived for PNC systems employing ML PNC channel decoder in the TWRC setting. The analysis in [31] could be used to obtain performance bounds of short-packet PNC systems. However, the analysis is related to PNC systems that employ ML PNC channel decoders only. The ML PNC channel decoder has prohibitive computational complexity, and it is not amenable to practical implementations. XOR-CD [1] is more relevant to short-packet

PNC systems since XOR-CD does not induce high decoding latency. In this work, we obtain an achievability bound for decoding error probabilities based on random-coding error-exponent for the uplink phase of PNC in the two-way relay setting, where practical and low-complexity XOR-CD is employed at the relay. This is the first work that provides an achievability bound for a PNC system operating with a practical and low-complexity PNC channel decoder.

Prior works [5], [32]–[34] on PNC channel estimators and decoders assume long-packet PNC systems. In long-packet PNC systems, ample pilot symbols are potentially available for accurate CSI estimation, enabling use of low-complexity PNC channel decoders. However, in short-packet PNC systems, the assumption of accurate CSI is untenable since short preambles (or lack thereof) in such systems may compromise the accuracy of estimated CSI. Moreover, tradeoffs between channel estimation resources and channel coding resources need to be considered since increasing the channel estimation resources will have significant impact on channel coding resources due to small packet size. In [25], [26], the authors considered the impact of mismatched CSI, and proposed an EM-BP framework [35] that solves the channel estimation and decoding problem jointly in an iterative manner. However, their PNC receivers employed multi-user-decoding, which is required for CSI estimation in their works, for PNC channel decoding. In this work, we devise a blind and a code-aided CSI estimation algorithms based on XOR-CD for short-packet PNC to compensate for the limitations imposed by having a small number of pilots. Unlike in [25], [26], the EM algorithms in our framework do not require the messages of both users to be decoded for CSI estimation. This is the first work that derives the EM algorithm for a PNC system operating with a practical and low-complexity PNC channel decoder.

We remark that in the design of the channel estimation and decoding algorithms, we do not focus on the code design for short block-lengths in the paper[1]. We instead focus on the interaction between the channel estimator and the PNC channel decoder. More specifically, we concentrate on the design of PNC channel estimation and decoding algorithms considering mismatched CSI for short-packet transmissions. For the algorithms, we adopt a low-density parity-check (LDPC) code that has a fair decoding performance in the short block-length regime for conventional single-user systems. Codes that achieve state-of-the-art decoding performance and that are designed specifically for short-packet communications can also be employed in the algorithms presented

[1]There has been an increasing interest in the design of short block-length codes that can achieve stringent reliability and delay constraints. Readers are referred to [36] and the references therein for more details on the design of short block-length codes.

here.

III. System Model

We focus our attention on the uplink phase of a PNC system in a TWRC. We assume the two users transmit simultaneously on the same frequency band using an (N, K) binary linear block code $\mathcal{C}$. The two users employ the binary phase shift keying (BPSK) modulation. The two transmissions are symbol-synchronous, i.e., at the receiver side, the symbols of the two users are aligned[2]. The two codewords transmitted by users A and B are denoted by $\mathbf{a}$ and $\mathbf{b}$ respectively. We denote the corresponding BPSK modulated codewords by the length-N row vectors $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$. Two preambles, one per transmitted codeword, are also transmitted, as depicted in Fig. 3. We denote the preambles of user A and B by the length-L row vectors $\mathbf{w}^{\mathrm{A}}$ and $\mathbf{w}^{\mathrm{B}}$ respectively. For symbol-synchronous systems, orthogonal time-overlapping sequences are possible, e.g., Walsh-Hadamard sequences. The preambles can be assumed to consist of such orthogonal sequences to ensure that they are linearly independent, i.e., $\langle \mathbf{w}^{\mathrm{A}}, \mathbf{w}^{\mathrm{B}} \rangle = 0$. In the preamble and data transmission phases, the relay observes $\mathbf{y}^{\mathrm{p}} \in \mathbb{C}^{1\times L}$ and $\mathbf{y} \in \mathbb{C}^{1\times N}$ respectively. We have

$$\begin{cases} \mathbf{y}^{\mathrm{p}} = h^{\mathrm{A}}\mathbf{w}^{\mathrm{A}} + h^{\mathrm{B}}\mathbf{w}^{\mathrm{B}} + \mathbf{n}^{\mathrm{p}} = \mathbf{h}\mathbf{w} + \mathbf{n}^{\mathrm{p}}, \\ \mathbf{y} = h^{\mathrm{A}}\mathbf{x}^{\mathrm{A}} + h^{\mathrm{B}}\mathbf{x}^{\mathrm{B}} + \mathbf{n} = \mathbf{h}\mathbf{x} + \mathbf{n}, \end{cases}$$

where h^{A}, h^{B} are the two (independent) complex channel coefficients, $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}]$, $\mathbf{w} = [(\mathbf{w}^{\mathrm{A}})^T, (\mathbf{w}^{\mathrm{B}})^T]^T$ ($(.)^T$ denotes the transpose of a vector/matrix), $\mathbf{x} = [(\mathbf{x}^{\mathrm{A}})^T, (\mathbf{x}^{\mathrm{B}})^T]^T$ and $\mathbf{n}^{\mathrm{p}} \in \mathbb{C}^{1\times L}$, $\mathbf{n} \in \mathbb{C}^{1\times N}$ are the additive white Gaussian noise (AWGN) contributions, with noise samples modeled as independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance $2\sigma^2$.

We refer to the overall length $N_{\text{eff}} = N + L$ as *effective block length*. The preamble length impacts the spectral efficiency of each transmission as follows. Denote by $R = K/N$ the rate of $\mathcal{C}$. We have that each user is employing $L + N$ channel uses to transmit K information bits, resulting in an *effective rate*

$$R_{\text{eff}} := \frac{K}{N + L} < R \quad \text{bpcu},$$

where bpcu stands for bits per channel use. Assuming BPSK modulation, we have that the signal-to-noise ratio *per user* given by

$$\frac{E_b}{N_0} = \frac{1}{2R_{\text{eff}}\sigma^2}$$

[2]Milder conditions on symbol synchronism have been analyzed, among others, in [33], [37]–[39].

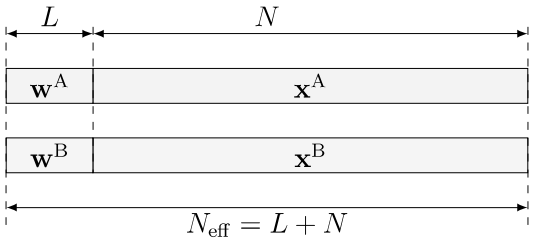


Fig. 3. Uplink frame definition.

where E_b is the energy per information bit *per user* and N_0 is the single-sided noise power spectral density. Here we assume that $|h^{\mathrm{A}}|^2 = |h^{\mathrm{B}}|^2 = 1$.

A. ML and XOR PNC Channel Decoding

Assuming h^{A} and h^{B} are perfectly known, the relay aims at decoding a binary linear combination (i.e., the XOR codeword) $\mathbf{c} = \mathbf{a} + \mathbf{b}$. When maximum-likelihood (ML) PNC channel decoding is employed, the relay computes

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \arg\max_{\mathbf{c}\in\mathcal{C}} \phi_{\mathrm{ML}}(\mathbf{y}, \mathbf{c}) \tag{1}$$

where we define the *ML decoding metric* as

$$\phi_{\mathrm{ML}}(\mathbf{y}, \mathbf{c}) := \sum_{\substack{\mathbf{a},\mathbf{b}:\ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a},\mathbf{b}\in\mathcal{C}}} p_{\mathbf{Y}|\mathbf{AB}}\left(\mathbf{y}|\mathbf{a}, \mathbf{b}\right). \tag{2}$$

In (2), the vector-wise channel transition probability density is

$$p_{\mathbf{Y}|\mathbf{AB}}\left(\mathbf{y}|\mathbf{a}, \mathbf{b}\right) = \prod_{i=1}^{N} p_{Y|AB}(y_i|a_i, b_i) \tag{3}$$

with

$$p_{Y|AB}(y|a, b) := \frac{1}{2\pi\sigma^2} \exp\left(\frac{-|y - h^{\mathrm{A}}\mu(a) - h^{\mathrm{B}}\mu(b)|^2}{2\sigma^2}\right). \tag{4}$$

Here $\mu(\cdot)$ is the modulation operation, with $\mu(0) = +1$ and $\mu(1) = -1$. From the equations above, we may restate (1) as

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \arg\max_{\mathbf{c}\in\mathcal{C}} \left[\sum_{\substack{\mathbf{a},\mathbf{b}:\ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a},\mathbf{b}\in\mathcal{C}}} \left(\prod_{i=1}^{N} p_{Y|AB}(y_i|a_i, b_i) \right) \right].$$

In general, sub-optimal PNC channel decoding such as XOR-CD is often used [40], [41]. When dealing with XOR-CD, the demodulation and channel decoding tasks are separated. The (soft) demodulator provides the decoder with the bit-wise soft estimate

$$\lambda(y_i, c_i) := \sum_{\substack{a_i, b_i: \\ a_i + b_i = c_i}} p_{Y|AB}(y_i|a_i, b_i). \tag{5}$$

Channel decoding then takes place in the same way as for a point-to-point channel, i.e., any off-the-shelf binary decoder can be employed. We introduce the reference decoder which computes

$$\hat{\mathbf{c}}_{\mathrm{XC}} = \arg\max_{\mathbf{c} \in \mathcal{C}} \phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c}), \tag{6}$$

where we define the decoding metric

$$\phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c}) := \prod_{i=1}^{N} \lambda(y_i, c_i) \tag{7}$$

$$= \prod_{i=1}^{N} \sum_{\substack{a_i, b_i: \\ a_i + b_i = c_i}} p_{Y|AB}(y_i|a_i, b_i) \tag{8}$$

$$= \sum_{\substack{\mathbf{a}, \mathbf{b}: \ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a}, \mathbf{b} \in \mathbb{F}_2^N}} p_{\mathbf{Y}|\mathbf{AB}}(\mathbf{y}|\mathbf{a}, \mathbf{b}). \tag{9}$$

We refer to the XOR-CD scheme using the metric in (7) as *maximum metric (MM) XOR-CD*[3].

B. Practical XOR-CD

Comparing (2) with (7), we observe that the ML decoding metric does not admit a trivial factorization, whereas the decoding metric under an XOR-CD is given by the product of N factors (one per observation). Observe also that the implementation of an XOR-CD scheme according to the rule (6) still entails in general a complexity growing exponentially with the block length. However, thanks to factorization of (7), one may use decoding algorithms which require input as bit-wise (or symbol-wise) metrics only; one can then also find fast algorithms to (6) that yield good solutions. This is not the case with ML PNC decoder, where the factorization of the ML decoding metric is not available. The calculation of the metric in (7) can be obtained, for instance, in the case of convolutional codes by applying the Viterbi decoding algorithm.

[3]Note that in (9) $\mathbf{a}$ and $\mathbf{b}$ are arbitrary vectors in $\mathbb{F}_2^N$ (i.e., they are not necessarily codewords), whereas in (2) the sum runs over all codeword pairs $\mathbf{a}, \mathbf{b}$ such that $\mathbf{a} + \mathbf{b} = \mathbf{c}$. The MM XOR-CD is hence obtained as a relaxation of the optimum ML channel decoding.

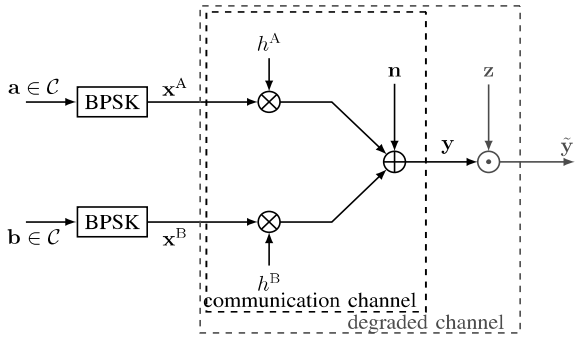


Fig. 4. Actual communication channel model and degraded channel model of uplink transmission in a TWRC setting.

For some codes, however, the metric in (7) cannot be obtained by any practical low-complexity channel decoding algorithm. For those codes, we have to resort to sub-optimum algorithm with respect to MM XOR-CD.

In our study, we adopt an LDPC code. The calculation of the metric in (7) for such a code is rather complicated. To obtain approximate solution to (6), we use BP decoding algorithm over the factor graph of the code where the likelihood of the XOR coded-bit c_i is replaced by soft estimate $\lambda(y_i, c_i)$, computed in (5), for channel decoding.

IV. Random Coding Bound for PNC Systems Employing XOR-CD

In this section, we derive RCB for short-packet PNC systems employing XOR-CD. We first construct the degraded channel model and provide the ML PNC channel decoding metric for the degraded channel. We then show that ML decision over the degraded channel is equal to the one provided by MM XOR-CD over the original channel. Using the result, we derive the random-coding error-exponent [42], [43] under XOR-CD to obtain RCB for PNC systems. The bound is derived under the assumption of perfect CSI available at the relay, and it is extended to a case where CSI is obtained using the pilot symbols in Appendix A.

A. Equivalent Degraded Channel

We refer to the proposed degraded channel as equivalent degraded channel (EDC). To construct EDC, we modify the original channel by appending to the channel output a block which

performs the multiplication of each channel output by a coefficient picked uniformly at random in $\{-1,+1\}$ (see Fig. 4). We denote by $\tilde{\mathbf{y}}$ the modified channel output. We have

$$\tilde{\mathbf{y}} = \mathbf{y} \odot \mathbf{z},$$

where $\odot$ denotes the Hadamard product and $\mathbf{z} = (z_1, \ldots, z_N)$ with elements modeled as i.i.d. Bernoulli random variables with $P_Z(-1) = P_Z(1) = 1/2$. Accordingly, we have

$$P_{\tilde{Y}|Y}(\tilde{y}|y) = \begin{cases} 1/2 & \text{if } \tilde{y} = +y \\ 1/2 & \text{if } \tilde{y} = -y \end{cases}$$

When ML PNC decoding is employed for EDC, the relay computes

$$\hat{\mathbf{c}}_{\text{MLD}} = \arg\max_{\mathbf{c}\in\mathcal{C}} \phi_{\text{ML}}^{\text{EDC}}(\tilde{\mathbf{y}}, \mathbf{c}),$$

where the ML PNC channel decoding metric for EDC is defined as

$$\phi_{\text{ML}}^{\text{EDC}}(\tilde{\mathbf{y}}, \mathbf{c}) := \sum_{\substack{\mathbf{a},\mathbf{b}:\ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a},\mathbf{b}\in\mathcal{C}}} p_{\tilde{\mathbf{Y}}|\mathbf{AB}}(\tilde{\mathbf{y}}|\mathbf{a},\mathbf{b}) \tag{10}$$

$$= \sum_{\substack{\mathbf{a},\mathbf{b}:\ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a},\mathbf{b}\in\mathcal{C}}} \prod_{i=1}^{N} p_{\tilde{Y}|AB}(\tilde{y}_i|a_i,b_i) \tag{11}$$

with symbol-wise channel transition probability density

$$\begin{aligned} p_{\tilde{Y}|AB}(\tilde{y}_i|a_i,b_i) =& \sum_{y_i=\pm\tilde{y}_i} p_{Y|AB}(y_i|a_i,b_i) p_{\tilde{Y}|Y}(\tilde{y}_i|y_i) \\ =& \frac{1}{2}\Big[p_{Y|AB}(\tilde{y}_i|a_i,b_i) + p_{Y|AB}(-\tilde{y}_i|a_i,b_i)\Big]. \end{aligned} \tag{12}$$

We now proceed to show that the ML PNC channel decoding on EDC is equivalent to XOR-CD on the original (non-degraded) channel. The following two lemmas will be useful for the purpose.

Lemma 1. *For the PNC system under consideration, we have*

$$p_{Y|AB}(y|a,b) = p_{Y|AB}(-y|\bar{a},\bar{b}),$$

where $\bar{a}$ ($\bar{b}$) is the binary complement of a (b).

Proof. Due to the adoption of BPSK modulation at the end users, we have $\mu(a) = (-1)^a$ and $\mu(b) = (-1)^b$. From (4), we have

$$
\begin{aligned}
&p_{Y|AB}(-y|\bar{a},\bar{b}) \\
&= \frac{1}{2\pi\sigma^2}\exp\left(\frac{-|-y-h^{\mathrm{A}}\mu(\bar{a})-h^{\mathrm{B}}\mu(\bar{b})|^2}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-|-y-h^{\mathrm{A}}(-1)^{a+1}-h^{\mathrm{B}}(-1)^{b+1}|^2}{2\sigma^2}\right) \\
&= \frac{1}{2\pi\sigma^2}\exp\left(\frac{-|-y+h^{\mathrm{A}}(-1)^{a}+h^{\mathrm{B}}(-1)^{b}|^2}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-|y-h^{\mathrm{A}}(-1)^{a}-h^{\mathrm{B}}(-1)^{b}|^2}{2\sigma^2}\right) \\
&= p_{Y|AB}(y|a,b).
\end{aligned}
$$

□

We next present Lemma 2 that states the function given therein is an even function of its associated observation.

Lemma 2. *For the PNC system under consideration, the summation of the channel transition probability of the original (non-degraded) channel over the XOR coded-bit is an even function of the observation variable Y, i.e.,*

$$\sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(y_i|a_i,b_i) \ = \sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(-y_i|a_i,b_i).$$

Proof. Take left hand side of the above equation, expand the summation and then apply Lemma 1 to get

$$
\begin{aligned}
\sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(y_i|a_i,b_i) &= p_{Y|AB}(y_i|a_i,a_i+c_i) + p_{Y|AB}(y_i|\bar{a}_i,\bar{a}_i+c_i) \\
&= p_{Y|AB}(-y_i|\bar{a}_i,\bar{a}_i+c_i) + p_{Y|AB}(-y_i|a_i,a_i+c_i) \\
&= \sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(-y_i|a_i,b_i).
\end{aligned}
$$

□

We next show that the ML PNC channel decoding over EDC outputs a decision equal to that of XOR-CD over the original (non-degraded) channel.

Theorem 1. *The ML decision over the EDC coincides with the MM XOR-CD decision over the (original) communication channel.*

Proof. To prove the theorem, we have to show that the decoding metric of ML PNC channel decoding over EDC is equal to (or some scalar multiple of) the decoding metric of XOR-CD over the (original) communication channel, i.e.,

$$\phi_{\mathrm{ML}}^{\mathrm{EDC}}(\tilde{\mathbf{y}}, \mathbf{c}) \propto \phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c}),$$

where $\phi_{\mathrm{ML}}^{\mathrm{EDC}}(\tilde{\mathbf{y}}, \mathbf{c})$ is defined in (10) and $\phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c})$ is defined in (7).

After applying Lemma 1, i.e., $p_{Y|AB}(-\tilde{y}_i|a_i, b_i) = p_{Y|AB}(\tilde{y}_i|\bar{a}_i, \bar{b}_i)$, to (12) and then putting it in (11), we get

$$\phi_{\mathrm{ML}}^{\mathrm{EDC}}(\tilde{\mathbf{y}}, \mathbf{c}) = \sum_{\substack{\mathbf{a},\mathbf{b}:\ \mathbf{a}+\mathbf{b}=\mathbf{c} \\ \mathbf{a},\mathbf{b}\in\mathcal{C}}} \prod_{i=1}^{N} \frac{1}{2}\left[p_{Y|AB}(\tilde{y}_i|a_i, b_i) + p_{Y|AB}(\tilde{y}_i|\bar{a}_i, \bar{b}_i)\right].$$

Observe that a_i and b_i in the equation above turn out to be dummy variables. Their individual values do not matter; only their sum c_i matters. So, for a given $\mathbf{c}$, we have

$$\phi_{\mathrm{ML}}^{\mathrm{EDC}}(\tilde{\mathbf{y}}, \mathbf{c}) \propto \prod_{i=1}^{N} \sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(\tilde{y}_i|a_i, b_i), = \prod_{i=1}^{N} \sum_{a_i,b_i:\ a_i+b_i=c_i} p_{Y|AB}(y_i|a_i, b_i) = \phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c}),$$

where the first equality follows from Lemma 2 since $\tilde{y}_i = \pm y_i$ and the summation term is an even function of y_i. The second equality follows from (8). □

Observation 1. *From the above theorem, we see that the performance of XOR-CD over the original (non-degraded) channel, expressed in* (3)*, can be fully characterized by analyzing the transmission with a linear block code $\mathcal{C}$ over the (virtual) memory-less point-to-point channel, i.e.,*

$$p_{\tilde{\mathbf{Y}}|\mathbf{C}}(\tilde{\mathbf{y}}|\mathbf{c}) \;\; = \;\; \prod_{i=1}^{N} p_{Y|C}(\tilde{y}_i|c_i)$$

where the transition probability density, obtained by applying Lemma 1 on (12)*,*

$$p_{Y|C}(\tilde{y}_i|c_i) = p_{Y|C}(y_i|c_i) = \frac{1}{2}\left[\sum_{\substack{a_i,b_i: \\ a_i+b_i=c_i}} p_{Y|AB}(y_i|a_i, b_i)\right]. \tag{15}$$

B. Random Coding Error Exponents

Given Observation 1, the derivation of the random coding error exponent under XOR-CD follows simply by deriving the error exponent for the virtual point-to-point channel with transition

probability density given by (15). Recall that Gallager's RCB [42] on the average block error probability $\bar{P}_{\mathrm{B}}$ of random (N, K) codes has the form

$$\bar{P}_{\mathrm{B}} \leq 2^{-NE_G(R)}, \tag{16}$$

where N is a block length, $R = K/N$ is the code rate and $E_G(R)$ is the random coding error exponent. The bound (16) holds also for the ensemble of linear random codes [42] and thus applies to the PNC systems employing XOR-CD.

We denote in the following the random variable associated with $\tilde{y}$ by $\tilde{Y}$ and the two independent and uniformly distributed binary random variables associated with XOR-coded bits by C, C'. Under perfect CSI, the random coding error exponent is

$$E_G(R) = \max_{0 \leq \rho \leq 1} \left[E_0(\rho) - \rho R \right]$$

where ρ is the auxiliary variables over which optimization is performed to get the maximum value of the right hand side, and

$$E_0(\rho) := -\log_2 \mathbb{E}\left[\left(\frac{\mathbb{E}\left[p_{Y|C}(Y|C')^{\frac{1}{1+\rho}} \middle| Y \right]}{p_{Y|C}(Y|C)^{\frac{1}{1+\rho}}} \right)^{\rho} \right],$$

where the inner expectation is with respect to the random variable C' and outer expectation is with respect to the random variables Y and C. Observe that, remarkably, the random coding error exponent under XOR-CD, given above, for linear random codes is exactly the same as the random coding error exponent under ML PNC channel decoding, given in [44], for linear codes.

The random coding error exponent can also be used to obtain upper bound on packet error rate for mismatched CSI PNC systems. The bound may act as a tool to characterize the trade-off between preamble-length and data-length for a given packet-length, E_b/N_0, and target packet error rate of short-packet PNC systems employing preamble-based CSI estimation (see Appendix A).

V. Channel Estimation Algorithms for PNC Systems Employing XOR-CD

In this section, we present a blind algorithm and a code-aided algorithm for channel estimation in PNC systems employing XOR-CD. In both channel estimation algorithms, we make use of the EM algorithm to obtain the CSI estimate. Once the final estimate of CSI is available, XOR-CD is performed to obtain XOR coded-bits. In the blind algorithm, channel estimation and decoding are performed in a disjoint manner, i.e., information is passed from the channel estimator to the

XOR channel decoder only once, without any feedback from the XOR channel decoder to the channel estimator for additional iterations. In the code-aided algorithm, an EM algorithm for channel estimation is combined with the BP algorithm for channel decoding, i.e., information is fed back from the XOR channel decoder to the channel estimator for additional iterations for joint channel estimation and channel decoding.

The EM algorithm iteratively finds the (local) optimal estimate of the parameter by alternating between the expectation (E) step and the maximization (M) step (the parameter and the E and M steps are defined in the subsections below) [23]. An initial estimate of the parameter is also required to bootstrap the EM algorithm. For the blind channel estimation, we obtain the initial estimate of parameter via the K-means algorithm [45]. In this scenario, we assume that the data is transmitted from the end users to the relay without any preamble, i.e., the preamble length $L = 0$ and $N_{\text{eff}} = N$ (see Fig. 3). For code-aided channel estimation, we combine EM algorithm for channel estimation with the BP decoding algorithm for XOR-CD. To bootstrap the algorithm, we obtain the initial estimate of the parameter via the preamble. In this case, we assume that a small preamble of length L is transmitted by each end user before the transmission of the codewords.

Before presenting the proposed algorithms for channel estimation in PNC systems employing XOR-CD, we first present a (conventional) preamble-aided channel estimation for PNC systems as a reference.

Notation: For the ease of presentation, we drop the subscripts of the probability terms from here onward. For example, we will write $p_{\mathbf{Y}|\mathbf{AB}}(\mathbf{y}|\mathbf{a},\mathbf{b})$ as $p(\mathbf{y}|\mathbf{a},\mathbf{b})$.

A. Preamble-Aided Channel Estimation

In the preamble-aided channel estimation, the relay estimates the channel coefficients via the two preambles. For the channel estimation step, we model h^{A} and h^{B} as two unknown parameters. Assuming orthogonal preambles, i.e., $\langle \mathbf{w}^{\mathrm{A}}, \mathbf{w}^{\mathrm{B}} \rangle = 0$, we can compute the ML CSI estimate as

$$\hat{h}^{\mathrm{A}} = \frac{1}{L}\langle \mathbf{w}^{\mathrm{A}}, \mathbf{y}^{\mathrm{p}} \rangle, \qquad \hat{h}^{\mathrm{B}} = \frac{1}{L}\langle \mathbf{w}^{\mathrm{B}}, \mathbf{y}^{\mathrm{p}} \rangle. \tag{17}$$

Given orthogonal preambles, it is easy to show that $\hat{h}^{\mathrm{A}}$ and $\hat{h}^{\mathrm{B}}$ can be modeled as two independent complex Gaussian random variables with means h^{A} and h^{B}, respectively, and variance

$$\sigma_{\mathrm{h}}^2 = \frac{1}{L} 2\sigma^2. \tag{18}$$

Once the ML CSI estimate is available, the metric $\lambda(y_i, c_i)$, computed in (5) using perfect CSI, is obtained assuming the estimated CSI as the perfect CSI to perform XOR-CD for PNC systems.

B. Blind Channel Estimation

Here we establish a framework for the blind channel estimation in PNC systems employing XOR-CD. For the blind channel estimation, we use the EM algorithm to estimate the "centroids" of the superposed constellation of the received symbols at the relay. Let $\boldsymbol{\theta} \stackrel{\triangle}{=} [\theta_1, \ldots, \theta_M]$ be a vector whose elements θ_l, $l = 1, \ldots, M$, are the centroids of the received superposed constellation. The EM algorithm provides the estimate of $\boldsymbol{\theta}$. Assuming $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}]$ and BPSK modulation, we have $\boldsymbol{\theta}$ as some permutation of the vector $[h^{\mathrm{A}}+h^{\mathrm{B}}, h^{\mathrm{A}}-h^{\mathrm{B}}, -h^{\mathrm{A}}+h^{\mathrm{B}}, -h^{\mathrm{A}}-h^{\mathrm{B}}]$ whose estimate is provided by the EM algorithm. Once the estimate of $\boldsymbol{\theta}$ is available, we use XOR-CD to solve the ambiguity within the elements of $\boldsymbol{\theta}$ and to eventually obtain the CSI estimate $\hat{\mathbf{h}}$.

To apply the EM algorithm, we consider $\mathbf{y}$ to be the observed (or incomplete) data, $\mathbf{x}$ to be the hidden data, $\{\mathbf{x}, \mathbf{y}\}$ to be the complete data, and $\boldsymbol{\theta}$ to be the unknown parameter. The E and M steps of the EM algorithm in the k^{th} iteration are as follows:

E-Step: Given the estimate $\hat{\boldsymbol{\theta}}^{(k-1)}$, compute the conditional expectation

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k-1)}) \log\left(p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})\right) \tag{19}$$

M-Step: Compute $\hat{\boldsymbol{\theta}}^{(k)}$ by

$$\hat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)})$$

We can greatly simplify the above steps for our specific problem. Note first that in the algorithm, we do not take codebook into account. We have

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(\mathbf{x}_i),$$

where $p(\mathbf{x}_i) = 1/M$, and M is the total number of clusters. Since we operate in a memory-less channel, we can factorize the complete likelihood $p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})$ as

$$p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i|\mathbf{x}_i, \boldsymbol{\theta}) p(\mathbf{x}_i)$$

where $\mathbf{x}_i = [\mathbf{x}_i^{\mathrm{A}}, \mathbf{x}_i^{\mathrm{B}}]^T$. Note that $\mathbf{x}_i^{\mathrm{A}}$ and $\mathbf{x}_i^{\mathrm{B}}$ are the i^{th} elements of row vectors $\mathbf{x}^{\mathrm{A}}$ and $\mathbf{x}^{\mathrm{B}}$ respectively, whereas $\mathbf{x}_i$ is the i^{th} column of the data matrix $\mathbf{x}$. Putting the above equation in (19) and ignoring $p(\mathbf{x}_i)$ (as it is just a constant and does not affect the overall algorithm), we get

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k-1)}) \sum_{i=1}^{N} \log\left(p(y_i|\mathbf{x}_i, \boldsymbol{\theta})\right).$$

The above equation can further be simplified as

$$\begin{aligned} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) &= \sum_{i=1}^{N} \sum_{\mathbf{x}_i} \log\left(p(y_i|\mathbf{x}_i, \boldsymbol{\theta})\right) \sum_{\mathbf{x}_j : j \in J_i} p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k-1)}) \\ &= \sum_{i=1}^{N} \sum_{\mathbf{x}_i} \log\left(p(y_i|\mathbf{x}_i, \boldsymbol{\theta})\right) p(\mathbf{x}_i|y_i, \hat{\boldsymbol{\theta}}^{(k-1)}), \end{aligned}$$

where $J_i = \{j \in \mathbb{N} | 1 \leq j \leq N, j \neq i\}$. Here,

$$p(\mathbf{x}_i|y_i, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{\mathbf{x}_j : j \in J_i} p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k-1)}) \tag{21}$$

is the *a-posteriori-probability* (APP) of $\mathbf{x}_i$. In the algorithm, we do not take codebook into account. The hidden data $\mathbf{x}_i$ thus only depends on y_i and that leads us to (21).

Note that $\mathbf{x}_i$ can only take M values in total. We have $\mathbf{x}_i \in \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_M\}$ where $\tilde{\mathbf{x}}_l$, $l = 1, \ldots, M$, denotes the given value of $\mathbf{x}_i$. As an example, for BPSK modulated PNC systems, we have $\mathbf{x}_i \in \{-1, +1\} \times \{-1, +1\}$. Since $\boldsymbol{\theta}$ is a function of $\mathbf{x}_i$ (i.e., $\theta_l = \mathbf{h}\tilde{\mathbf{x}}_l$), we get

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) = \sum_{i=1}^{N} \sum_{l=1}^{M} \log\left(p(y_i|\tilde{\mathbf{x}}_l, \theta_l)\right) p(\tilde{\mathbf{x}}_l|y_i, \hat{\boldsymbol{\theta}}^{(k-1)}). \tag{22}$$

Note further that for our specific problem, we have

$$p(y_i|\tilde{\mathbf{x}}_l, \theta_l) = p(y_i|\theta_l) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y_i - \theta_l)^H (y_i - \theta_l)}{2\sigma^2}\right),$$

where $(.)^H$ denotes the Hermitian transpose of a vector/matrix. Given the above conditional probability distribution, it is possible to analytically maximize the Q function as well. Taking the derivative of (22) with respect to θ_l and setting it equal to zero, we obtain

$$\hat{\theta}_l^{(k)} = \frac{y_i p(\tilde{\mathbf{x}}_l|y_i, \hat{\boldsymbol{\theta}}^{(k-1)})}{\sum_{i=1}^{N} p(\tilde{\mathbf{x}}_l|y_i, \hat{\boldsymbol{\theta}}^{(k-1)})}, \qquad l = 1, \cdots, M, \tag{23}$$

where the posterior probability can be obtained as

$$p(\tilde{\mathbf{x}}_l|y_i, \hat{\boldsymbol{\theta}}^{(k-1)}) = \frac{p(\tilde{\mathbf{x}}_l) p(y_i|\tilde{\mathbf{x}}_l, \hat{\boldsymbol{\theta}}^{(k-1)})}{\sum_{j=1}^{M} p(\tilde{\mathbf{x}}_j) p(y_i|\tilde{\mathbf{x}}_j, \hat{\boldsymbol{\theta}}^{(k-1)})} \tag{24}$$

Note that for the EM iterations, we only need to compute (23) and (24), which perform both the E and M steps simultaneously.

For blind channel estimation, we do not have any preamble that we can use to bootstrap the EM algorithm. To bootstrap the EM algorithm, $\hat{\boldsymbol{\theta}}^{(0)}$ is provided by the standard K-means algorithm [45]. Assuming BPSK modulation, the K-means algorithm provide estimated centroids (or estimate of elements of $\boldsymbol{\theta}$) of $\hat{\boldsymbol{\theta}}^{(0)}$. We skip the detailed presentation of the K-means algorithm in the paper since the K-means algorithm employed here is very similar to the one used in machine learning for finding clusters in the data [46].

The EM algorithm suffers from identifiability problem, e.g., we do not know whether a given element of $\hat{\boldsymbol{\theta}}^{(k)}$ represents $h^{\mathrm{A}} + h^{\mathrm{B}}$, $h^{\mathrm{A}} - h^{\mathrm{B}}$,$-h^{\mathrm{A}} + h^{\mathrm{B}}$ or $-h^{\mathrm{A}} - h^{\mathrm{B}}$. To obtain $\hat{\mathbf{h}}$ for BPSK modulated PNC systems, first, we multiply $\hat{\theta}_l^{(k)}$ by the sign of its real component to obtain the modified vector $\hat{\boldsymbol{\theta}}_m^{(k)}$. Second, we cluster the four elements of $\hat{\boldsymbol{\theta}}_m^{(k)}$ in two clusters via the K-means algorithm. The K-means algorithm provides us two centroids: the one corresponding to $h^{\mathrm{A}} + h^{\mathrm{B}}$ and $-h^{\mathrm{A}} - h^{\mathrm{B}}$; the one corresponding to $h^{\mathrm{A}} - h^{\mathrm{B}}$ and $-h^{\mathrm{A}} + h^{\mathrm{B}}$. Note that both the $h^{\mathrm{A}} + h^{\mathrm{B}}$ and $-h^{\mathrm{A}} - h^{\mathrm{B}}$ correspond to XOR coded-bit 0, and similarly both the $+h^{\mathrm{A}} - h^{\mathrm{B}}$ and $-h^{\mathrm{A}} + h^{\mathrm{B}}$ correspond to XOR coded-bit 1. Therefore, we can assume that one of the centroid represents $h^{\mathrm{A}} + h^{\mathrm{B}}$ and the other one represents $h^{\mathrm{A}} - h^{\mathrm{B}}$. We can then linearly solve the two equations that we get from the centroids to obtain $\hat{h}^{\mathrm{A}}, \hat{h}^{\mathrm{B}}$. Using these estimates, the metric $\lambda(y_i, c_i)$ is computed to perform XOR-CD for PNC systems. If syndrome check fails, we run the decoder again using the alternative hypothesis, i.e., the centroid that previously represented $h^{\mathrm{A}} + h^{\mathrm{B}}$ ($h^{\mathrm{A}} - h^{\mathrm{B}}$) now represents $h^{\mathrm{A}} - h^{\mathrm{B}}$ ($h^{\mathrm{A}} + h^{\mathrm{B}}$ respectively). If the decoding is successful using any one of the hypothesis, we then not only know the final estimated CSI but also successfully decode the XOR coded-bits.

We remark that we do not take codebook $\mathcal{C}$ into account in the calculation of the Q function in the E-step. Thus, in this algorithm, there is no feedback provided by the channel decoder to the EM estimator and hence PNC channel estimation and decoding are performed separately in a disjoint manner.

C. Code-Aided Channel Estimation

In this subsection, we establish a framework for code-aided channel estimation in PNC systems employing XOR-CD. In this framework, we combine the EM algorithm for channel estimation with the BP decoding algorithm for XOR-CD.

For the application of the EM algorithm, we consider $\mathbf{y}$ to be the observed (or incomplete) data, XOR codeword $\mathbf{c}$ to be the hidden data, $\{\mathbf{c}, \mathbf{y}\}$ to be the complete data, and $\mathbf{h}$ to be the unknown parameter. The E and M steps of the EM algorithm in the k^{th} EM iteration are as follows:

E-Step: Given $\hat{\mathbf{h}}^{(k-1)}$, compute the conditional expectation

$$Q(\mathbf{h}, \hat{\mathbf{h}}^{(k-1)}) = \sum_{\mathbf{c}:\mathbf{c}\in\mathcal{C}} p(\mathbf{c}|\mathbf{y}, \hat{\mathbf{h}}^{(k-1)}) \log\left(p(\mathbf{y}, \mathbf{c}|\mathbf{h})\right) \tag{25}$$

M-Step: Compute $\hat{\mathbf{h}}^{(k)}$ by

$$\hat{\mathbf{h}}^{(k)} = \arg\max_{\mathbf{h}} Q(\mathbf{h}, \hat{\mathbf{h}}^{(k-1)}) \tag{26}$$

Assuming that the complete likelihood can be factorized as

$$p(\mathbf{y}, \mathbf{c}|\mathbf{h}) = p(\mathbf{y}|\mathbf{c}, \mathbf{h})p(\mathbf{c}) = \prod_{i=1}^{N} p(y_i|c_i, \mathbf{h}) \mathbb{1}_{\mathcal{C}}(\mathbf{c})/|\mathcal{C}|, \tag{27}$$

where $\mathbb{1}_{\mathcal{C}}(\mathbf{c}) = 1$ when $\mathbf{c} \in \mathcal{C}$ and zero otherwise, and $|\mathcal{C}|$ is the number of codewords belonging to $\mathcal{C}$. Putting the above equation in (25) and ignoring $\mathbb{1}_{\mathcal{C}}(\mathbf{c})$ and scaling factor $|\mathcal{C}|$, we get

$$\begin{aligned}
Q(\mathbf{h}, \hat{\mathbf{h}}^{(k-1)}) &= \sum_{\mathbf{c}:\mathbf{c}\in\mathcal{C}} p(\mathbf{c}|\mathbf{y}, \hat{\mathbf{h}}^{(k-1)}) \sum_{i=1}^{N} \log\left(p(y_i|c_i, \mathbf{h})\right) \\
&= \sum_{i=1}^{N} \sum_{\mathbf{c}:\mathbf{c}\in\mathcal{C}} p(\mathbf{c}|\mathbf{y}, \hat{\mathbf{h}}^{(k-1)}) \log\left(p(y_i|c_i, \mathbf{h})\right) \\
&= \sum_{i=1}^{N} \sum_{c_i} \log\left(p(y_i|c_i, \mathbf{h})\right) \sum_{\substack{c_j: j\in J_i, \\ \mathbf{c}\in\mathcal{C}}} p(\mathbf{c}|\mathbf{y}, \hat{\mathbf{h}}^{(k-1)}) \\
&= \sum_{i=1}^{N} \sum_{c_i} \log\left(p(y_i|c_i, \mathbf{h})\right) p(c_i|\mathbf{y}, \hat{\mathbf{h}}^{(k-1)}),
\end{aligned} \tag{28}$$

where $J_i = \{j \in \mathbb{N} | 1 \leq j \leq N, j \neq i\}$.

Unfortunately, factorization of the term $p(\mathbf{y}|\mathbf{c}, \mathbf{h})$ into $\prod_{i=1}^{N} p(y_i|c_i, \mathbf{h})$ provided in (27) is not possible, as discussed in Section III-B. Nevertheless, since XOR-CD is equivalent to ML channel decoding on EDC introduced in Section IV-A, we make use of the factorization given in Observation 1. More specifically, we replace the probability term $p(\mathbf{y}|\mathbf{c}, \mathbf{h})$ by the decoding metric $\phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c})$, where we have, from (8) and (15),

$$\phi_{\mathrm{XC}}(\mathbf{y}, \mathbf{c}) = \prod_{i=1}^{N} \lambda(y_i, c_i) = \prod_{i=1}^{N} \sum_{\substack{a_i, b_i: \\ a_i + b_i = c_i}} p(y_i|a_i, b_i, \mathbf{h}) \propto \prod_{i=1}^{N} p(y_i|c_i, \mathbf{h}).$$

In XOR-CD, $\lambda(y_i, c_i)$ (or $p(y_i|c_i, \mathbf{h})$) is considered proper likelihood of c_i. XOR-CD, therefore, only provides an approximation to APP $p(c_i|\mathbf{y}, \mathbf{h})$, even if the underlying code has cycle free factor graph. Given these considerations, we can see the Q function obtained in (28) is only an approximation to the actual Q function given in (25). Notwithstanding that, since the Q function in (28) is tractable, we carry on with it for the design and evaluation of code-aided channel estimator.

To maximize the Q function in (28), we perform numerical optimization with respect to $\mathbf{h}$. Analytical optimization of the Q function is not possible. This is because we have "log of sum" term in (28), i.e.,

$$\log\left(p(y_i|c_i, \mathbf{h})\right) = \log\left(\frac{1}{2}\left[\sum_{\substack{a_i, b_i: \\ a_i + b_i = c_i}} p(y_i|a_i, b_i)\right]\right),$$

which does not permit a closed form solution of the optimal value of $\mathbf{h}$. The summation inside the log term, thanks to EM algorithm, is over the coded-bits a_i and b_i only rather than the whole codewords $\mathbf{a}$ and $\mathbf{b}$. We would have the latter case if we had tried to directly obtain optimal estimate of $\mathbf{h}$ without using the EM algorithm. In our case, the required numerical optimization is significantly simpler than that in the case where summation is required over the whole codewords inside the log term.

For code-aided channel estimation algorithm, we assume that a small preamble of length L is available. To bootstrap the EM algorithm, we obtain $\hat{\mathbf{h}}^{(0)}$ by performing preamble-aided channel estimation according to (17). The E and M steps can then be computed by obtaining Q function according to (28) and performing numerical optimization of the Q function respectively.

For short-packet PNC systems, the EM algorithm employed for code-aided channel estimation, like the EM algorithm employed for blind channel estimation, suffers from the identifiability problem. The problem exists even if we bootstrap the algorithm using the estimate $\hat{\mathbf{h}}^{(0)}$ obtained from the preamble. This is due to the fact that $\hat{\mathbf{h}}^{(0)}$ obtained from small preambles in short-packet transmissions suffer from high variance (given in (18)). The convergence of the EM algorithm depends critically on its initialization [47]. Given the small preamble for initial CSI estimation, the EM algorithm may often converge to undesirable or incorrect maxima, i.e., $\hat{\mathbf{h}}^{(k)}$ may converge to $[h^{\mathrm{A}}, -h^{\mathrm{B}}]$ instead of $[h^{\mathrm{A}}, h^{\mathrm{B}}]$. To solve the issue, we compute the metric $\lambda(y_i, c_i)$ according to the current CSI estimate (or using the null hypothesis) to perform XOR-CD. If syndrome check fails, we run the decoder again using the alternative hypothesis. If the decoding is successful using

any one of the hypothesis, we then not only know the final estimated CSI but also successfully decode the XOR coded-bits.

We remark that in the EM algorithm for code-aided channel estimation, the codebook $\mathcal{C}$ is taken into the account in the calculation of the Q function in the E step. Thus, in this algorithm, there is feedback provided by the channel decoder to the EM algorithm, and hence PNC channel estimation and decoding is performed jointly.

VI. Computational Complexity

This section discusses the computational complexity of channel estimation and decoding algorithms presented in the previous section. To begin with, we first discuss the computational complexity of XOR-CD. In our study, we adopt an LDPC codes and we employ the BP algorithm for XOR-CD. The complexity here is the same as for the binary BP used for channel decoding in single user (point-to-point) communication systems (see Section III-B). More specifically, the computational complexity of XOR-CD (for the case where the Tanner graph of the code does not contain state-variable nodes) is $\mathcal{O}\left(I_{XOR}(4|\mathcal{E}| - N + K)\right)$ [48], where I_{XOR} is the number of BP iterations performed for channel decoding, $|\mathcal{E}|$ is the total number of edges in the Tanner graph of the code, and N and K are the total number of XOR channel coded bits and information bits respectively.

For preamble-aided channel estimation, we need to estimate CSI of each user according to (17). The complexity of the CSI estimation then is $\mathcal{O}(L)$. The overall complexity of preamble-aided channel estimation followed by channel decoding then is $\mathcal{O}(L + I_{XOR}\left(4|\mathcal{E}| - N + K\right))$.

The blind channel estimation algorithm makes use of the K-means algorithm and the EM algorithm to obtain the CSI estimates. The complexity of the K-means algorithm is $\mathcal{O}(I_K MN)$ [45], [46], where I_K is the total iterations specified for the K-means algorithm, M is the the number of clusters, and N is the total number of received data symbols (or the total number of channel coded bits in BPSK modulated PNC system). The EM algorithm in blind channel estimation requires (23) and (24) to be computed for $l = 1, 2, \cdots, M$, in its each iteration, I_{BEM}. For computing (23) for each $l = 1, 2, \cdots, M$, we require $\mathcal{O}(M)$ operations. Whereas for computing (24) for each $l = 1, 2, \cdots, M$, we require $\mathcal{O}(N)$ operations. The complexity of calculating (23) and (24) in each iteration of the EM algorithm, I_{BEM}, is $\mathcal{O}(M^2 + MN)$. The total complexity of the blind channel estimation then is $\mathcal{O}\left(I_K MN + I_{BEM}(M^2 + MN)\right)$.

TABLE I

OVERALL COMPUTATIONAL COMPLEXITY OF THE CHANNEL ESTIMATION AND DECODING ALGORITHMS.

Channel Estimation and Decoding Algorithm	**Computational Complexity**
Preamble-aided channel estimation with XOR-CD	$\mathcal{O}(L + I_{XOR}\,(4\lvert\mathcal{E}\rvert - N + K))$
Blind channel estimation with XOR-CD	$\mathcal{O}\left((I_K + I_{BEM})N + I_{XOR}(4\lvert\mathcal{E}\rvert - N + K)\right)$
Code-aided channel estimation with XOR-CD	$\mathcal{O}(L + I_{CA}[N + I_{XOR}(4\lvert\mathcal{E}\rvert - N + K)] + I_{XOR_F}(4\lvert\mathcal{E}\rvert - N + K))$

Note that for BPSK modulated PNC, $M = 4$ (see Section V-B). Since the value of M does not change for BPSK modulated PNC, we can treat M as a constant. The complexity of blind channel estimation followed by XOR-CD then is $\mathcal{O}\left((I_K + I_{BEM})N + I_{XOR}(4|\mathcal{E}| - N + K)\right)$.

The complexity of the EM algorithm in the code-aided channel estimation can be computed as follows. The complexity of computing the Q function in (28) in the E-step is $\mathcal{O}(N)$ once APP is available through XOR-CD. The complexity of the M-step, on the other hand, in (26) depends on the underlying algorithm used for optimization. In our work, we use a Quasi-Newton algorithm. The complexity of the algorithm is $\mathcal{O}(n^2 + nC(f))$ [49], where n is the number of variables over which the optimization is performed and $C(f)$ is the cost of computing the objective function. In our setup, the optimization is performed over the real and imaginary parts of h^{A} and h^{B}, thus we have $n = 4$. The objective function in our case is the Q function given in (28), and its computational complexity is $\mathcal{O}(N)$. The complexity of performing one iteration of the code-aided algorithm (that includes the complexity of EM algorithm and XOR-CD) is $\mathcal{O}\left(4^2 + 4N + I_{XOR}(4|\mathcal{E}| - N + K)\right) = \mathcal{O}\left(N + I_{XOR}(4|\mathcal{E}| - N + K)\right)$. The overall complexity, including that of the preamble-aided channel estimation to initialize the code-aided algorithm and final XOR-CD performed at the end of EM iterations, is $\mathcal{O}(L + I_{CA}[N + I_{XOR}(4|\mathcal{E}| - N + K)] + I_{XOR_F}(4|\mathcal{E}| - N + K))$, where I_{CA} is the number of EM iterations specified for the code-aided algorithm and I_{XOR_F} is the number of BP iterations for final XOR-CD performed at the end of EM iterations.

The summary of the computational complexity of all the presented channel estimation and decoding algorithms is given in Table I.

VII. NUMERICAL RESULTS

This section presents the numerical results. As PNC channel estimation and decoding is concerned with obtaining the XOR coded-bits at the relay, the numerical results are presented

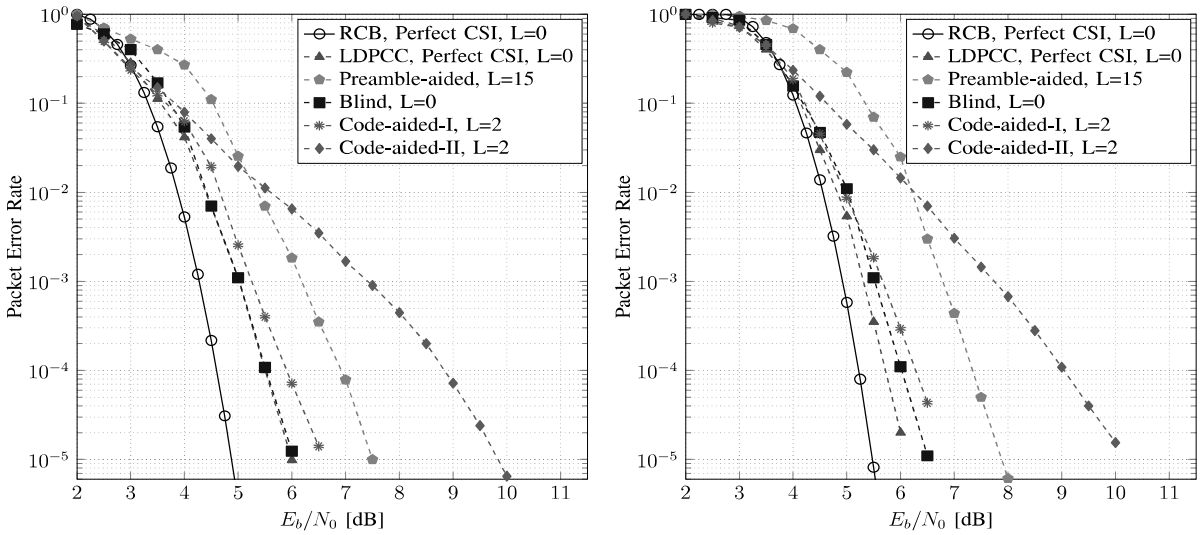
TABLE II

VALUES OF VARIOUS PARAMETERS USED IN OUR NUMERICAL EVALUATIONS.

Parameters	Values
Channel state information of user A (h^{A})	1
Channel state information of user B (h^{B})	1 and $1j$
Modulation	BPSK
Channel Code for XOR-CD	$(128, 64)$ IRA LDPC
K-means iterations I_K for blind channel estimation	20
EM iterations for blind channel estimation I_{BEM}	25
Maximum BP iterations for XOR-CD in blind and preamble-aided channel estimation I_{XOR}	500
Maximum EM iterations in code-aided channel estimation I_{CA}	16
Maximum BP iterations for XOR-CD I_{XOR} in code-aided channel estimation	25
EM iterations in final XOR-CD at the end of code-aided algorithm I_{XOR_F}	100
Preamble length L in preamble-aided channel estimation	15
Packet length N_{eff}	128
Information bits K	64

only for the uplink phase of packet transmission in TWRC. We consider a PNC system operating with packet length N_{eff} of 128 symbols over a TWRC. We assume both end users employ BPSK modulation. Note that the error performance of PNC is also affected by the relative phase offset between the signals of the two end nodes received at the relay [41]. Thus, we consider a phase-asynchronous BPSK modulated PNC system where relative phase offset is 0 and $\pi/2$ radian (i.e., both extreme cases [41]). In particular, we provide results for $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}] = [1, 1]$ and $\mathbf{h} = [1, i]$. The values of various simulation parameters are summarized in Table II.

We employ an $(128, 64)$ irregular repeat-accumulate (IRA) [50] LDPC code for XOR-CD. A regular variable node degree of 4 is used for the information part. For the blind channel estimation, we set the K-means iterations $I_K = 20$ and maximum EM iterations for channel estimation $I_{BEM} = 25$. For XOR-CD followed by blind channel estimation and by preamble-aided channel estimation, we set the maximum BP iterations to $I_{XOR} = 500$. For the code-aided channel estimation, we set the maximum EM iterations to $I_{CA} = 16$ and BP iterations to $I_{XOR} = 25$. For the final XOR-CD at the end of EM iterations in code-aided algorithm, we



(a) $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}] = [1, 1]$ (b) $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}] = [1, i]$

Fig. 5. Packet error rate vs E_b/N_0 of PNC system in a TWRC setting. Number of information bits is $K = 64$, and the packet length is $N_{\text{eff}} = 128$. LDPC code (LDPCC) is employed for channel decoding. The preamble length is L and the number of data symbols is $N_{\text{eff}} - L = 128 - L$.

set the maximum BP iterations to $I_{XOR_F} = 100$. In this way, the complexity of the channel decoding in all of the algorithms remains constant.

In the results, we show the upper bound on the average packet error rate vs E_b/N_0 provided by the RCB for PNC systems where perfect CSI is available at the relay. Packet error rates of PNC systems when the discussed channel estimation algorithms are employed with XOR-CD are also shown in the results. As a reference, the packet error rate performance of a PNC system without preamble and perfect CSI for XOR-CD is also provided. For the preamble-aided channel estimation, we consider the preamble length $L = 15$, which provides us the best performance for packet length $N_{\text{eff}} = 128$ (see Appendix A). For the code-aided channel estimation, we consider the following variants of the algorithm:

1) In the first variant of the algorithm, upon failing of syndrome check in the last phase of decoding, the alternative hypothesis (see Section V-C) is checked by performing XOR-CD again using the alternative hypothesis. The algorithm is referred to as code-aided-I in the figures.
2) In the second variant of the algorithm, the alternative hypothesis is not checked upon failing of the syndrome check in the last phase of decoding. The algorithm is referred to

as code-aided-II in the figures.

Fig. 5 shows the numerical results. **Response[5]:** In our evaluation, the preamble length is $L = 0$ for RCB, the (ideal) perfect-CSI PNC system, and the blind CSI estimation algorithm. For the code-aided CSI estimation algorithms, the preamble length is $L = 2$, and for the preamble-aided CSI estimation algorithm, the preamble length is $L = 15$. From the figures, we see that RCB provides us a good reference on the packet error rate performance of short-packet PNC systems. We also see that the blind channel estimator outperforms its code-aided-I, code-aided-II and preamble-aided counterparts by around $0.1, 3$ and 1.5 dB, respectively, for both $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}] = [1, 1]$ and $\mathbf{h} = [h^{\mathrm{A}}, h^{\mathrm{B}}] = [1, i]$, and it attains the target packet error rate of 10^{-4} within around 0.5 dB additional E_b/N_0 from the RCB.

The results show that the preamble-aided channel estimation is not a good candidate for short-packet PNC systems as it results in performance degradation. Even the code-aided-II algorithm, which uses data symbols as well as preamble for channel estimation, leads to poor packet error rate performance for short-packet PNC systems. This is due to the fact that the small preambles in the short-packet PNC systems do not provide a good CSI estimate for the initialization of the EM algorithm. EM algorithm depends critically on the initial CSI estimate $\hat{\mathbf{h}}^{(0)}$ for its convergence.

The code-aided-I and the blind channel estimation algorithms outperform the preamble-aided algorithm in channel estimation and decoding of PNC systems employing XOR-CD. The packet error rate performance of XOR-CD with these two algorithms almost overlaps the performance of XOR-CD in a PNC system where perfect CSI is assumed (with preamble length $L = 0$). The good performance is due to the fact that the alternative hypothesis regarding the CSI estimate is also checked in both code-aided-I and blind channel estimation algorithms, when decoding using the initial CSI estimate fails the syndrome check (see Section V-B and V-C). The blind channel estimation algorithm has slightly better performance than the code-aided-I algorithm. This is due to the overhead of the preamble in the code-aided-I algorithm. Thanks to the EM algorithm, the preamble overhead in code-aided-I is tiny compared to the one in preamble-aided channel estimation. The upside of using the preamble is that it allows simpler initialization of the EM algorithm as compared to the K-means algorithm.

Fig. 6 shows the mean square error (MSE) results of various channel estimation algorithms. We define the MSE of the CSI estimator as $\mathrm{MSE}(\hat{\mathbf{H}}) \triangleq \frac{1}{2}\left[\mathrm{MSE}(\hat{H}^{\mathrm{A}}) + \mathrm{MSE}(\hat{H}^{\mathrm{B}})\right]$, where $\mathrm{MSE}(\hat{H}^{\mathrm{A}}) = \mathbb{E}_{\hat{H}^{\mathrm{A}}}\left[|\hat{H}^{\mathrm{A}} - h^{\mathrm{A}}|^2\right]$, and $\mathrm{MSE}(\hat{H}^{\mathrm{B}}) = \mathbb{E}_{\hat{H}^{\mathrm{B}}}\left[|\hat{H}^{\mathrm{B}} - h^{\mathrm{B}}|^2\right]$. Here $\hat{H}^{\mathrm{A}}$ and $\hat{H}^{\mathrm{B}}$ are the random variables associated with $\hat{h}^{\mathrm{A}}$ and $\hat{h}^{\mathrm{B}}$ respectively, and $\hat{\mathbf{H}} = [\hat{H}^{\mathrm{A}}, \hat{H}^{\mathrm{B}}]$. From the figure,

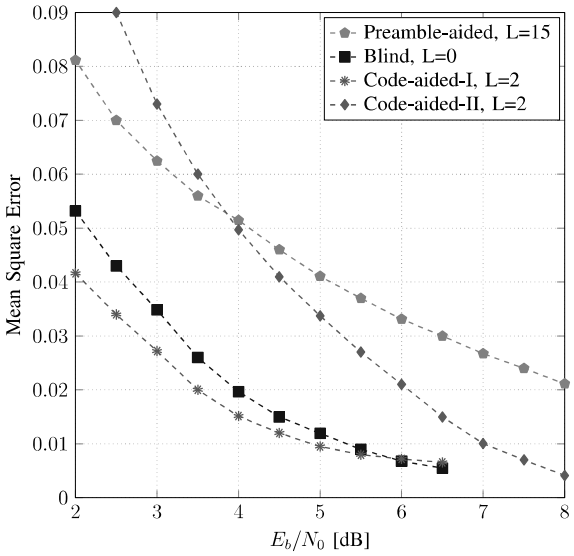


Fig. 6. Mean square error results of various algorithms estimating CSI of a PNC system in a TWRC setting. Number of information bits is $K = 64$, the packet length is $N_{\text{eff}} = 128$, and $\mathbf{h} = [h^{\text{A}}, h^{\text{B}}]$. LDPC code (LDPCC) is employed for channel decoding. The preamble length is L and the number of data symbols is $N_{\text{eff}} - L = 128 - L$.

we see that the code-aided-I algorithm has the minimum MSE, followed by the blind algorithm and the code-aided-II algorithm. The high MSE of the code-aided-II algorithm, compared with the code-aided-I algorithm, is due to the fact that the alternative hypothesis is not tested when decoding using the initial CSI estimate fails the syndrome check (see Section V-C). From the figure, we also see that the preamble-aided channel estimation algorithm has the highest MSE.

Preamble-aided channel estimation algorithm exhibits better packet error rate performance for the PNC system than the code-aided-II algorithm, even though we have relatively fewer data symbols and higher MSE for the preamble-aided algorithm. For the preamble-aided algorithm, data symbols $N = 113$ and preamble length $L = 15$, whereas for the code-aided-II algorithm, data symbols $N = 126$, preamble length $L = 2$. The packet length $N_{\text{eff}} = 128$ in both cases. We observed in our simulations that for the code-aided II algorithm, high MSE is caused by incorrect convergence of the algorithm to the undesirable CSI estimate (see Section V-C). Consider a scenario where $E_b/N_0 = 7$ dB and $\mathbf{h} = [h^{\text{A}}, h^{\text{B}}] = [1, i]$. For the code-aided-II algorithm, we then have the packet error rate of around 3×10^{-3} and MSE of 0.011. We can see that these results are consistent. More specifically, for the code-aided-II algorithm, the squared error of estimated CSI (when the algorithm converges to the undesirable CSI estimate) is $(|1-i|^2+|1+i|^2)/2 = 2$, and its contribution to the MSE term is $2 \times (3 \times 10^{-3}) = 0.006$ (which is lower than the actual

MSE of 0.011). On the other hand, for the preamble-aided channel estimator, MSE is around 0.027 at $E_b/N_0 = 7$ dB, which is also consistent with (18). The aforementioned scenario explains the higher packet error rate and lower MSE of the code-aided-II algorithm with respect to the preamble-aided algorithm.

Overall, from the results, we see that the code-aided-I and the blind algorithms have the best performance for short-packet PNC systems employing XOR-CD. The algorithms thus enable the application of PNC to short-packet communications.

VIII. Conclusion

The paper investigates the application of PNC for short-packet transmissions. Short-packet transmissions are envisioned as a key building block of the services to be provided in future wireless communications. For the application of PNC to short-packet transmission, first, we develop an achievability bound that provides a reference for the packet error rate performance in the short packet-length regime. The bound is based on random-coding error-exponent for the uplink phase of PNC in a two-way relay setting where practical and low-complexity XOR-CD is employed at the relay and where CSI is acquired from the preamble and fed to the channel decoder. Second, we design a blind algorithm and a code-aided algorithm for channel estimation in short-packet PNC systems operating with XOR-CD and mismatched CSI. The channel estimation and decoding algorithms put forth in this work have low complexity compared to those already discussed in literature—thanks to XOR-CD. The algorithms still achieve high reliability under mismatched CSI setting as they make use of all the transmitted symbols for CSI estimation. The bound and the algorithms present a fundamental framework for applying PNC to short-packet transmissions.

Appendix A
Random Coding Bound for Mismatched CSI PNC

Under mismatched CSI, the relay does not possess perfect knowledge of the channel coefficients but rather the estimate $\hat{\mathbf{h}}$. The decoder hence operates with the mismatched metric

$$q(y, c; \hat{\mathbf{h}}) = \sum_{a,b:\ a+b=c} \exp\left(\frac{-|y - \hat{h}^{\mathrm{A}}\mu(a) - \hat{h}^{\mathrm{B}}\mu(b)|^2}{2\sigma^2}\right),$$

where $\hat{\mathbf{h}} = [\hat{h}^{\mathrm{A}}, \hat{h}^{\mathrm{B}}]$. The random coding error exponent is in this case [51]–[53]

$$E_G(R; \hat{\mathbf{h}}) = \max_{0\le\rho\le 1} \sup_{s\ge 0} \Big[E_0(\rho, s; \hat{\mathbf{h}}) - \rho R\Big] \tag{29}$$

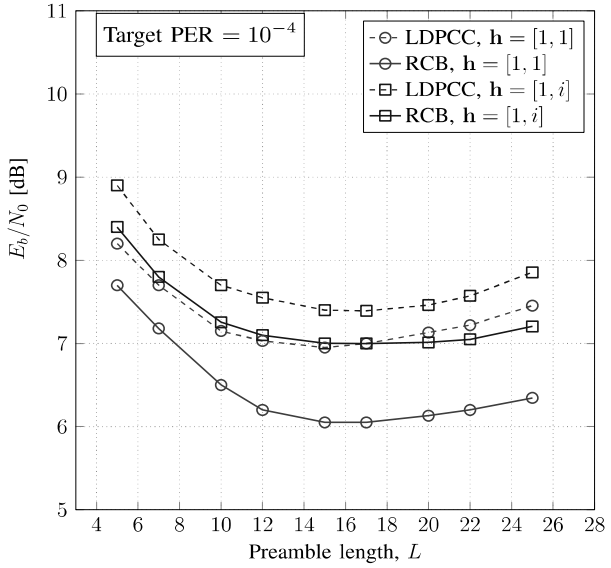


Fig. 7. E_b/N_0 required to achieve target packet error rate of 10^{-4} with preamble length L by a PNC system employing preamble-aided channel estimation and XOR-CD. Number of information bits is $K = 64$, the packet length is $N_{\text{eff}} = 128$, and $\mathbf{h} = [h^{\text{A}}, h^{\text{B}}]$. LDPC code (LDPCC) is employed for channel decoding. The preamble length is L and the number of data symbols is $N_{\text{eff}} - L = 128 - L$.

where ρ and s are the auxiliary variables over which optimization is performed to get the maximum value of the right hand side. Here

$$E_0(\rho, s; \hat{\mathbf{h}}) := -\log_2 \mathbb{E}\left[\left(\frac{\mathbb{E}\left[q(Y, C'; \hat{\mathbf{h}})^s \middle| Y\right]}{q(Y, C; \hat{\mathbf{h}})^s}\right)^{\rho}\right],$$

where the inner expectation is with respect to the random variable C' and outer expectation is with respect to the random variables Y and C.

Using RCB, we can obtain benchmarks on PER performance of short-packet PNC systems for packet and preamble of given sizes. More specifically, the bound acts as a benchmark for the case where we first obtain the CSI estimate from the preamble and we then feed it to the channel decoder assuming the estimated CSI as exact. To analyze the impact of mismatched CSI due to limited preamble length on packet error rate (PER), we do the following: first, estimate the CSI using (17) and compute RCB for that particular CSI estimate; second, compute the expected value of RCB where expectation is carried over the estimated CSI distributed according to (18). The upper bound on the average PER (i.e., RCB) for $\hat{\mathbf{h}}$ is $P_{\text{B}}(\hat{\mathbf{h}}) \leq 2^{-NE_G(R;\hat{\mathbf{h}})}$, where N is the number of channel uses for codeword transmission (i.e., it does not include preamble transmission. See Fig. 3). The preamble length L impacts the estimation of CSI $\hat{\mathbf{h}}$ according

to (17), and as a result, RCB varies in the equation above. Here $R = K/N$ is the rate, and $E_G(R; \hat{\mathbf{h}})$ is the mismatched random coding error exponent calculated in (29).

Denote by $\hat{H}^{\mathrm{A}}$ and $\hat{H}^{\mathrm{B}}$ the random variables associated with $\hat{h}^{\mathrm{A}}$ and $\hat{h}^{\mathrm{B}}$, we obtain an upper bound on average PER under mismatched CSI $\bar{P}_{\mathrm{B}}^{\mathrm{MM}} \leq \mathbb{E}\left\{2^{-NE_G(R;\hat{\mathbf{H}})}\right\}$, where $\hat{\mathbf{H}} = [\hat{H}^{\mathrm{A}}, \hat{H}^{\mathrm{B}}]$ and the expectation is with respect to the random variables $\hat{H}^{\mathrm{A}}$ and $\hat{H}^{\mathrm{B}}$.

Fig. 7 shows the minimum E_b/N_0 required by a PNC system employing preamble-aided channel estimation and XOR-CD to obtain packet error rate of 10^{-4} with a preamble length L. The results are given for RCB and the LDPC code (LDPCC). We assume information bits $K = 64$ and packet length $N_{\mathrm{eff}} = 128$. Both users employ BPSK modulation to transmit their packets (i.e., 128 symbols). Since preambles of length L are inserted in the beginning of the packets, the data part is shortened by puncturing the channel-coded bits to keep the packet length N_{eff} fixed to 128 symbols. From the figure, we see that for the current setup, preamble length $L = 15$ exhibits the best packet error rate performance. Moreover, we see that RCB can serve as a good tool for identifying good preamble-length regimes for PNC systems employing preamble-aided channel estimation and XOR-CD.

Response[5]: The impact of the preamble length on the packet error rate is summarized below. For the preamble-aided algorithm, increasing the pilot symbols too much increases the overhead, resulting in sub-par performance (see Fig. 7). Decreasing the number of pilots too much, on the other hand, results in mismatched CSI, which again degrades the performance. This behaviour is typical [54], and the code-aided II algorithm also follows this behaviour of preamble-aided channel estimation. The code-aided I algorithm with $L = 2$ can approach the packet error rate performance of an ideal PNC system operating with perfect CSI without any pilot symbols. We cannot reduce the pilot symbols further for our setting (i.e., we need at least two pilots to initialize the code-aided algorithm, and setting $L = 0$ will require us to use the blind algorithm). By increasing the number of pilot symbols in the code-aided algorithm, we only increase the overhead and degrade the overall performance. The blind algorithm does not make use of any pilot symbol for CSI estimation, and its performance hence depends on the received data symbols only.

REFERENCES

[1] S. S. Ullah, G. Liva, and S. C. Liew, "Physical-layer network coding: A random coding error exponent perspective," in *2017 IEEE Information Theory Workshop (ITW)*, Nov 2017, pp. 559–563.

[2] S. S. Ullah, G. Liva, and S. C. Liew, "Short packet physical-layer network coding with mismatched channel state information," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2018, pp. 1–5.

[3] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: physical-layer network coding," in *Proc. ACM MobiCom*. Los Angeles, CA, Sep. 2006, pp. 358–365.

[4] P. Popovski and H. Yomo, "The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *2006 IEEE International Conference on Communications*, vol. 9, June 2006, pp. 3885–3890.

[5] S. C. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond," *Physical Communication*, vol. 6, pp. 4–42, Mar. 2013.

[6] F. Schaich, T. Wild, and Y. Chen, "Waveform contenders for 5g - suitability for short packet and low latency transmissions," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, May 2014, pp. 1–5.

[7] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, March 2018.

[8] ITU-R, "Rep. ITU-R M.2412-0: Guidelines for evaluation of radio interface technologies for IMT-2020," International Telecommunication Union, Tech. Rep., Oct 2017.

[9] 5G-PPP, "5G and the factories of the future." The 5G Infrastructure Public Private Partnership, Oct 2015, White Paper.

[10] ——, "5G automotive vision." The 5G Infrastructure Public Private Partnership, Oct 2015, White Paper.

[11] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *International Conference on 5G for Ubiquitous Connectivity (5GU)*, Nov 2014, pp. 146–151.

[12] G. Durisi, T. Koch, J. stman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna rayleigh-fading channels," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 618–629, Feb 2016.

[13] G. Durisi, T. Koch, and P. Popovski, "Towards massive, ultra-reliable, and low-latency wireless: The art of sending short packets," *Proc. IEEE*, vol. 104, no. 9, Sep. 2016.

[14] Y. Hu, J. Gross, and A. Schmeink, "On the capacity of relaying with finite blocklength," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1790–1794, March 2016.

[15] Y. Hu, A. Schmeink, and J. Gross, "Blocklength-limited performance of relaying under quasi-static rayleigh channels," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4548–4558, July 2016.

[16] Y. Hu, M. C. Gursoy, and A. Schmeink, "Relaying-enabled ultra-reliable low-latency communications in 5g," *IEEE Network*, vol. 32, no. 2, pp. 62–68, March 2018.

[17] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, Nov 2017.

[18] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, March 2012, pp. 2731–2735.

[19] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, pp. 162–259, 2017. [Online]. Available: http://dx.doi.org/10.1561/1300000060

[20] L. Kleinrock, *Theory, Volume 1, Queueing Systems*. New York, NY, USA: Wiley-Interscience, 1975.

[21] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[22] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[23] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society*, vol. 29, pp. 1–38, 1976.

[24] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[25] T. Wang and S. C. Liew, "Joint channel estimation and channel decoding in physical-layer network coding systems: An EM-BP factor graph framework," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 4, pp. 2229–2245, April 2014.

[26] Z. Situ, I. W. Ho, T. Wang, S. C. Liew, and S. C. Chau, "Ofdm modulated pnc in v2x communications: An ici-aware approach against cfos and time-frequency-selective channels," *IEEE Access*, vol. 7, pp. 4880–4897, 2019.

[27] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5641–5654, Nov 2010.

[28] Y. Hao, D. Goeckel, Z. Ding, D. Towsley, and K. K. Leung, "Achievable rates for network coding on the exchange channel," in *MILCOM 2007 - IEEE Military Communications Conference*, Oct 2007, pp. 1–7.

[29] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 438–460, March 2011.

[30] K. Lu, S. Fu, Y. Qian, and H. Chen, "On capacity of random wireless networks with physical-layer network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 763–772, June 2009.

[31] J. Kim, "Performance analysis of physical layer network coding," Ph.D. dissertation, The University of Michigan, 2009.

[32] S. Zhang and S. C. Liew, "Channel coding and decoding in a relay system operated with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 788–796, Jul. 2009.

[33] L. Lu and S. C. Liew, "Asynchronous physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 819–831, Feb. 2012.

[34] D. Wubben and Y. Lang, "Generalized sum-product algorithm for joint channel decoding and physical-layer network coding in two-way relay systems," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, Dec 2010, pp. 1–5.

[35] J. Dauwels, S. Korl, and H. A. Loeliger, "Expectation maximization as message passing," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, Sept 2005, pp. 583–586.

[36] G. Liva, L. Gaudio, T. Ninacs, and T. Jerkovits, "Code design for short blocks: A survey," *CoRR*, vol. abs/1610.00873, 2016. [Online]. Available: http://arxiv.org/abs/1610.00873

[37] F. Rossetto and M. Zorzi, "On the design of practical asynchronous physical layer network coding," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Perugia, Italy, Jun. 2009, pp. 469–473.

[38] Q. Yang and S. C. Liew, "Asynchronous convolutional-coded physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1380–1395, Mar. 2014.

[39] P.-C. Wang, Y.-C. Huang, and K. R. Narayanan, "Asynchronous physical-layer network coding with quasi-cyclic codes," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 2, pp. 309–322, Feb. 2015.

[40] S. S. Ullah, S. C. Liew, L. Lu, and L. You, "Physical-layer network coding: A high performance PHY-layer decoder," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[41] S. S. Ullah, S. C. Liew, and L. Lu, "Phase asynchronous physical-layer network coding: Decoder design and experimental study," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2708–2720, April 2017.

[42] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[43] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[44] J. Kim and W. E. Stark, "Error exponent of exclusive-or multiple-access channels," in *Proc. IEEE International Symposium on Information Theory*, Seoul, South Korea, June 2009, pp. 1709–1713.

[45] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[46] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[47] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, 2nd ed., ser. Wiley series in probability and statistics. Hoboken, NJ: Wiley, 2008.

[48] J. Chen, A. Dholakia, E. Eleftheriou, M. P. C. Fossorier, and X.-Y. Hu, "Reduced-complexity decoding of ldpc codes," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1288–1299, Aug 2005.

[49] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[50] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd Int. Symp. Turbo codes and related topics*. Citeseer, 2000.

[51] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[52] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov 2000.

[53] J. Scarlett, A. Martinez, and A. G. i. Fabregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[54] G. Liva, G. Durisi, M. Chiani, S. S. Ullah, and S. C. Liew, "Short codes with mismatched channel state information: A case study," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.