

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

# Design and Performance Analysis of Network Code Division Multiplexing for Wireless Sensor Networks

Jing Yue<sup>†‡</sup>, Zihuai Lin<sup>†</sup>, Guoqiang Mao<sup>\*‡</sup>, and Branka Vucetic<sup>†</sup>

<sup>†</sup>School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia

<sup>‡</sup>National ICT Australia (NICTA), Australia

<sup>\*</sup>School of Computing and Communications, University of Technology Sydney, Sydney, Australia

E-mail: {jing.yue,zihuai.lin,branka.vucetic}@sydney.edu.au, g.mao@ieee.org

**Abstract**—In this paper, we investigate the performance of a wireless sensor network, in which multiple groups of source nodes communicate with their respective destination nodes with the help of a common relay network. A network code division multiplexing (NCDM) scheme is proposed to remove the inter-session interference among multiple transmission sessions at each destination. We focus on analyzing the soft processing algorithm of the NCDM scheme. Based on the analysis results, a new code design criteria for the construction of the generator matrix is proposed. Simulation results show that by following the proposed code design criteria, the bit error ratio (BER) performance gap between the scheme we studied and the serial session scheme can be managed effectively. In serial session scheme, source nodes in a number of groups communicate with their respective destinations in a time division manner.

**Index Terms**—Network code division multiplexing, wireless sensor networks, soft processing algorithm, log-likelihood ratio, bit error ratio.

## I. INTRODUCTION

As a special channel coding strategy developed for cooperative communication networks [1], distributed coding has attracted significant attention recently. Distributed coding technique has been applied in the design of conventional channel codes, such as distributed turbo codes [2], distributed space-time codes [3] and distributed low-density parity-check (LDPC) codes [4]. In the above distributed coding schemes, transmission reliability over point-to-point wireless communication channels has been efficiently improved.

The distributed coding schemes discussed above are developed for small-scale unicast relay networks. However, many practical applications are in large-scale wireless sensor networks (WSNs). WSNs is the key elements in Internet of Things (IoT). IoT is a scenario in which objects are provided with unique identifiers and data transmission are over a network without requiring human-to-human or human-to-computer interaction [5]. In WSNs, a large number of sensor nodes gather information from the surrounding environment and then deliver the gathered information to a common destination via other sensor nodes, serving as relay nodes. The sensor nodes, which gather information, are referred to as

source nodes. As the source and the relay nodes are in different spatial locations, the destination node receives replicas of the transmitted signal via multiple nodes, its signal will have higher spacial diversity and hence better performance [6].

However, relay nodes may consume considerable amounts of radio spectrum and energy. If not used properly, they may cause further spectrum congestion and interference. Network coding (NC) [7] is a novel approach, in which the relay nodes are allowed to encode the packets received from multiple source nodes and send the encoded packets instead. Coding operations enable the relay nodes to compress the information, and whenever possible, to reduce the number of transmissions and bandwidth consumption and/or maximizing the transmission reliability. The idea of NC can be applied to WSNs. In [8], extrinsic information transfer (EXIT) charts are employed to design the irregular low-density generator matrix (LDGM) codes based on a single session model.

In this paper, we consider a network with multiple sessions. We further assume that at each destination node, only the information from its own session is required and the information from other sessions are not allowed to be decoded due to the security reasons. This assumption is realistic in some application scenarios in IoT. For example, in smart meter networks, where multiple sensors, measuring water, gas and energy consumptions, communicate with multiple utility control centres [9]. The information from various utilities, such as for water, gas and energy consumptions, cannot be shared. We propose to design novel network coding mechanisms, referred to as network code division multiplexing (NCDM). The NCDM scheme can provide security protection on physical layer. Furthermore, in order to optimize the overall system performance, including minimizing interference, and maximizing the spectral efficiency, diversity and coding gains, we analyze the NCDM scheme design and its performance. Based on the analysis results, we propose a new code design criteria for the NCDM scheme.

The remainder of the paper is organized as follows: the system model is introduced in Section II. Section III proposes the NCDM scheme for the parallel session model. Section IV analyzes the soft processing algorithm of the NCDM scheme and introduces a code design criteria. Section V presents the

simulation results. Finally, Section VI concludes the paper.

## II. SYSTEM MODEL

We consider a wireless network containing  $L$  transmission sessions. In each transmission session, a group of source nodes communicate with their common destination node through a group of relay nodes. The  $i$ th session is denoted by  $\varphi_i$ , where  $i \in \{1, 2, \dots, L\}$ . We assume that in the  $i$ th session there are  $S_i$  source nodes. The total number of source nodes in the  $L$  sessions is  $M = \sum_{i=1}^L S_i$ . Denote the destination node of session  $\varphi_i$  by  $D_i$ . Assume that the network consists of  $N$  relay nodes which are shared among all the  $L$  sessions. The  $j$ th relay node is denoted by  $R_j$ ,  $j \in \{1, 2, \dots, N\}$ . The system model is shown in Fig. 1.

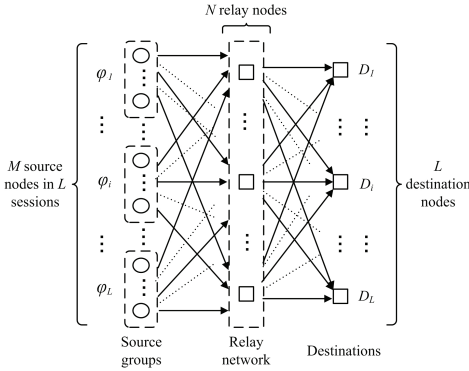


Fig. 1. A network graph used to describe the system model.

The data transmission from the source nodes to their destination nodes is carried out in three phases: a broadcast phase, a relay phase and a data recovery phase.

In the *broadcast phase*, all the source nodes broadcast their data packets to all the relay and the destination nodes. Each data packet is composed of three parts: an information data part, a Cyclic Redundancy Check (CRC) part and a header. The indices of the source nodes and the session to which the source node belongs are contained in the header. The information data parts of all data packets have the same length, as well as the CRC parts and the headers. A certain MAC layer protocol, e.g., Time-Division Multiple Access (TDMA) or Carrier Sense Multiple Access (CSMA), is assumed to be employed. This ensures that the source node transmissions do not cause interference to each other.

In the *relay phase*, each relay node listens to the transmission from the source nodes, decodes the received data packets, checks the correctness of the decoded packets by using CRC and puts the correctly decoded packets into its buffer. Since each relay node only performs hard decision decoding and CRC check, the decoding process is very fast and the delay due to the decoding at each relay node can be neglected. Then each relay node selects uniformly at random a number of  $d_r$  data packets from its buffer to perform network coding by using linear combinations of the packets in the field of  $\text{GF}(2)$  [10].  $d_r$  is selected with probability  $\Omega_{d_r}$  according to the degree

distribution  $\Omega(x) = \sum \Omega_{d_r} x^{d_r}$ . Since the number  $d_r$  is much smaller than the total number of source nodes, i.e.,  $d_r \ll M$ , the probability of the event that the number of data packets in a relay node's buffer is smaller than  $d_r$  is negligibly small and we can always assume that the network coding process is achievable. Then at each relay node, a network-coded data packet is formed. The connection information of the source and the relay nodes is contained in the header of each network-coded data packet. Finally, all the relay nodes broadcast their network-coded data packets to the destination nodes using a TDMA or CSMA MAC layer protocol.

In the *data recovery phase*, data packets received from the source and the relay nodes in the broadcast and relay phases, respectively, are organized into a data matrix. We consider the quasi-static Rayleigh fading channels. Without loss of generality, we focus our analysis on the  $i$ th row of the received data matrix. At the destination node  $D_i$ , the received signal can be expressed as  $\mathbf{r}_i = \alpha_i \otimes \sqrt{E_b}(\mathbf{J} - 2\beta) + \mathbf{n}_i$ , where  $\mathbf{r}_i$  is a row vector of size  $1 \times (M + N)$ .  $\alpha_i$  is the fading coefficient vector of size  $1 \times (M + N)$ . The fading coefficients remain constant over the length of one data packet but change independently between adjacent packets. Thus, we can assume perfect channel state information (CSI) at the destination nodes.  $\otimes$  represents the element-wise multiplication of the vector.  $E_b$  is the average transmitted energy per bit.  $\mathbf{J}$  is a  $1 \times (M + N)$  vector with all elements equal to 1. The binary sequence  $\beta$  is a random row of the data matrix  $[\mathbf{m}_1 \cdots \mathbf{m}_i \cdots \mathbf{m}_L] \odot \mathbf{G}_i$ , where  $\odot$  represents the multiplication operation over the field of  $\text{GF}(2)$ .  $[\mathbf{m}_1 \cdots \mathbf{m}_i \cdots \mathbf{m}_L]$  represents the information matrix of the  $L$  sessions.  $\mathbf{G}_i$  is the corresponding LDGM generator matrix seen from  $D_i$  of size  $M \times (M + N)$ . The parameter  $\mathbf{n}_i = [n_{i,1}, n_{i,2}, \dots, n_{i,M+N}]$  is the additive white Gaussian noise with a zero mean and a double sided power spectral density of  $N_0/2$ . Next, each destination node implements the NCDM scheme to remove the interference from the source nodes of other sessions.

Note that since all the connection information are contained in the header of each data packet, the destination nodes know how the checks at the relay nodes are formed and can correspondingly replicate the code graph and perform message-passing decoding.

## III. THE NCDM SCHEME FOR A PARALLEL SESSION MODEL WITH MORE THAN TWO SESSIONS

In this section, we will introduce the NCDM scheme for the parallel session model, which can be used to remove the inter-session interference at each destination. By a parallel session model, we refer to source nodes in multiple sessions communicating with their respective destinations through a common relay network simultaneously. The fundamental idea behind the NCDM process takes advantage of the property of  $\mathbf{G} \odot \mathbf{H}^T = \mathbf{0}$ . Here  $\mathbf{G}$  and  $\mathbf{H}$  are the generator matrix and the parity check matrix of the LDGM codes, respectively. Throughout the paper, we denote the LDGM generator matrix seen from the  $i$ th destination node  $D_i$  by  $\mathbf{G}_i$  and the corresponding parity check matrix by  $\mathbf{H}_i$ .  $\mathbf{G}_i$  is of size

$M \times (M + N)$ . Let  $\mathbf{G}_i^t$  represent the  $t$ th submatrix of  $\mathbf{G}_i$  and of size  $S_t \times (M + N)$ ,  $t \in \{1, 2, \dots, L\}$ .  $S_t$  is the number of source nodes in the  $t$ th session. Since  $\mathbf{G}_i$  is a systematic matrix, it can be written as  $\mathbf{G}_i = [\mathbf{I}_M \mathbf{P}_i]$ , where  $\mathbf{I}_M$  represents an  $M \times M$  identity matrix.  $\mathbf{P}_i$  is an  $M \times N$  parity part matrix. Let  $\mathbf{P}_i^t$  be an  $S_t \times N$  parity part matrix of  $\mathbf{G}_i^t$ .  $\mathbf{H}_i^t$  is the parity-check matrix corresponding to the generator matrix  $\mathbf{G}_i^t$ .

To obtain session  $\varphi_i$ 's information at the destination  $D_i$ ,  $i \in \{1, 2, \dots, L\}$ , the interference introduced from other sessions to  $\varphi_i$  needs to be eliminated. This can be done by multiplying the generator matrix  $\mathbf{G}_i$  by  $(\mathbf{H}_i^t)^T$ . Using the property that  $\mathbf{G}_i^t \odot (\mathbf{H}_i^t)^T = \mathbf{0}$ ,  $\forall t \in \{1, 2, \dots, L\} \setminus i$ , the interference from other sessions can be removed completely. However, the parity check matrix  $(\mathbf{H}_i^t)^T$  is also presented to other submatrices of  $\mathbf{G}_i^j$ ,  $j \in \{1, 2, \dots, L\} \setminus t$ , as an undesirable by-product. For example, when  $t = 1$ , after carrying out the first session  $\mathbf{G}_i^1$ 's interference cancellation process, the successive process of  $\mathbf{G}_i^2 \odot (\mathbf{H}_i^1)^T \odot (\mathbf{H}_i^2)^T$  is not equal to zero due to the contribution of  $(\mathbf{H}_i^1)^T$ . As a result, the interference from the second session's information cannot be eliminated and the interference cancellation process stops. To solve this problem,

in this paper, we deliberately introduce a term  $\left( (\tilde{\mathbf{H}}_i^1)^T \right)^{-1}$ ,

so that the product of  $(\mathbf{H}_i^1)^T$  and  $\left( (\tilde{\mathbf{H}}_i^1)^T \right)^{-1}$  approximately equals to an identity matrix  $\mathbf{I}$ . The interference cancellation process continues until  $\varphi_i$ 's information at  $D_i$  is obtained. Based on the above discussion, the equivalent generator matrix for the  $i$ th session seen from destination  $D_i$  is given by

$$\tilde{\mathbf{G}}_i = \begin{cases} \mathbf{G}_i \odot \left( \prod_{t,t \in \{1,2,\dots,L-1\} \setminus i} (\mathbf{H}_i^t)^T \odot \left( (\tilde{\mathbf{H}}_i^t)^T \right)^{-1} \right) \\ \odot (\mathbf{H}_i^L)^T, & \text{if } i \in \{1, \dots, L-1\} \\ \mathbf{G}_i \odot \left( \prod_{t,t \in \{1,2,\dots,L-2\}} (\mathbf{H}_i^t)^T \odot \left( (\tilde{\mathbf{H}}_i^t)^T \right)^{-1} \right) \\ \odot (\mathbf{H}_i^{L-1})^T, & \text{if } i = L \end{cases},$$

where  $\left( (\tilde{\mathbf{H}}_i^t)^T \right)^{-1}$  is an inverse matrix of  $(\tilde{\mathbf{H}}_i^t)^T$ .  $(\tilde{\mathbf{H}}_i^t)^T$  is a matrix obtained by setting the linear dependent rows in  $(\mathbf{H}_i^t)^T$  to zeros and keeping other rows unchanged.

Let  $\Phi_i$  represent the term  $\left( \prod_{t,t \in \{1,\dots,L-1\} \setminus i} (\mathbf{H}_i^t)^T \odot \left( (\tilde{\mathbf{H}}_i^t)^T \right)^{-1} \right) \odot (\mathbf{H}_i^L)^T$ . In a matrix form,  $\Phi_i$  can be expressed as

$$\Phi_i = \begin{bmatrix} \mathbf{0}_{S_1 \times A_i} & \mathbf{0}_{S_1 \times S_i} & \mathbf{0}_{S_1 \times (B_i - S_L)} & \mathbf{P}_i^1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{S_{i-1} \times A_i} & \mathbf{0}_{S_{i-1} \times S_i} & \mathbf{0}_{S_{i-1} \times (B_i - S_L)} & \mathbf{P}_i^{i-1} \\ \mathbf{0}_{S_i \times A_i} & \mathbf{I}_{S_i} & \mathbf{0}_{S_i \times (B_i - S_L)} & \mathbf{0}_{S_i \times N} \\ \mathbf{0}_{S_{i+1} \times A_i} & \mathbf{0}_{S_{i+1} \times S_i} & \mathbf{0}_{S_{i+1} \times (B_i - S_L)} & \mathbf{P}_i^{i+1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{S_L \times A_i} & \mathbf{0}_{S_L \times S_i} & \mathbf{0}_{S_L \times (B_i - S_L)} & \mathbf{P}_i^L \\ \mathbf{0}_{N \times A_i} & \mathbf{0}_{N \times S_i} & \mathbf{0}_{N \times (B_i - S_L)} & \mathbf{I}_N \end{bmatrix}, \quad (1)$$

$\underbrace{\hspace{15em}}_{(M-S_L) \text{ columns}} \quad \underbrace{\hspace{2em}}_{N \text{ columns}}$

where  $A_t = \sum_{i'=1}^{t-1} S_{i'}$ , when  $2 \leq t \leq L$  and  $A_t = 0$ , when  $t = 1$ .  $B_t = \sum_{i'=t+1}^L S_{i'}$ , when  $1 \leq t \leq L-1$  and  $B_t = 0$ ,

when  $t = L$ .  $(\mathbf{P}_i^t)^T$  is the transpose of the matrix  $\mathbf{P}_i^t$  [11]. For the first  $M - S_L$  columns, all the elements are equal to zero except an identity submatrix  $\mathbf{I}_{S_i}$  of size  $S_i \times S_i$ , see (1). For the last  $N$  columns of  $\Phi_i$ , the first  $L$  submatrices correspond to the parity part matrices of  $\mathbf{G}_i^t$ ,  $t \in \{1, \dots, L\} \setminus i$ . The last  $N$  rows form an identity matrix, as shown in (1).

Let  $[\underline{m}_1 \ \dots \ \underline{m}_i \ \dots \ \underline{m}_L]$  represent an arbitrary row of the data matrix  $[\underline{\mathbf{m}}_1 \ \dots \ \underline{\mathbf{m}}_i \ \dots \ \underline{\mathbf{m}}_L]$ . At destination node  $D_i$ , after the NCDM scheme, we can obtain the equivalent received signal as

$$\mathbf{r}_i \Phi_i = \left[ \alpha_i \otimes \sqrt{E_b} (\mathbf{J} - 2[\underline{m}_1 \ \dots \ \underline{m}_i \ \dots \ \underline{m}_L] \mathbf{G}_i) \right] \Phi_i + \mathbf{n}_i \Phi_i, \quad (2)$$

where  $\alpha_i$  is the fading coefficient of size  $1 \times (M + N)$ .

#### IV. PERFORMANCE ANALYSIS OF THE SOFT PROCESSING ALGORITHM AND CODE DESIGN CRITERIA

In [12], the performance of the NCDM scheme for the parallel session model with two sessions was analyzed from the perspective of matrix. In this section, we analyze the soft processing algorithm of the NCDM scheme. Then, based on the analysis results, a code design criteria is summarized to better control the BER performance.

##### A. Soft Processing Algorithm

For a binary random variable  $X \in \{\pm 1\}$ , its log-likelihood ratio (LLR) is defined as  $l(X) \doteq \log \left( \frac{P_r\{X=+1\}}{P_r\{X=-1\}} \right)$ . Subsequently, the LLR for a transmitted symbol  $x_{k'}$  at the destination node is given by  $l_{k'} = \frac{4\alpha_{i,k'} r_{k'} E_b}{N_0}$ , where  $\alpha_{i,k'}$  is the  $k'$ th element of the fading coefficient  $\alpha_i$ , which is assumed to be known at the receiver. For additive white Gaussian noise (AWGN) channel,  $\alpha_{i,k'} = 1$ .  $r_{k'}$  is the received symbol corresponding to the transmitted symbol  $x_{k'}$ .

Once the LLR values are obtained for all the transmitted symbols, we then need to find the corresponding input LLR values for each check node. This can be done by multiplying the LLR values with each column of the matrix  $\Phi_i$ . The obtained vectors of the LLR values after the multiplication are then the LLR values of the input symbols for the check nodes. Let  $\Psi_j$  be the set of position indices of the non-zero elements of the  $j$ th column of the matrix  $\Phi_i$ . Let  $z$  be the output symbol of the  $j$ th check node. Then  $z = \oplus \sum_{k' \in \Psi_j} x_{k'}$ , where  $\oplus$  represents the addition operation over the field of GF(2). The output LLR values of each check node can be computed by  $L_j = 2 \tanh^{-1} \left( \prod_{k' \in \Psi_j} \tanh \left( \frac{l_{k'}}{2} \right) \right)$ . Here  $l_{k'}$ ,  $k' \in \Psi_j$ , is the LLR value for the  $k'$ th input symbol of the check node  $j$ .

##### B. Analysis of the Soft Processing Algorithm

Let  $\Upsilon_j^t$  represent the set formed by the indices of positions of all the non-zero elements in the  $j$ th column of  $\mathbf{P}_i^t$ . Let  $z_j$  be the output symbol of the  $j$ th check node of the generator matrix, and let  $x_{t,k'}$  represent the  $k'$ th transmitted symbol of the  $t$ th session, where  $t \in \{1, 2, \dots, L\}$ . When  $M < j \leq M + N$ , the output symbol of the  $j$ th check node can be written as  $z_j = \oplus \sum_{t=1}^L \sum_{k' \in \Upsilon_j^t} x_{t,k'}$ . The encoded vector corresponding to the generator matrix

seen from the  $i$ th destination node can be written as  $\mathbf{x}_i = [x_{1,1}, \dots, x_{1,S_1}, \dots, x_{t,1}, \dots, x_{t,S_t}, \dots, x_{L,1}, \dots, x_{L,S_L}, z_{M+1}, \dots, z_j, \dots, z_{M+N}]$ .

Let  $l_{t,k'}$  be the LLR value of the  $k'$ th received symbol corresponding to the transmitted symbol from the  $t$ th session. It can be computed by

$$l_{t,k'} = \frac{4\alpha_{t,k'} r_{t,k'} E_b}{N_0}, \quad (3)$$

where  $\alpha_{t,k'}$  is the  $k'$ th element of the fading coefficient  $\alpha_t$ , which is assumed to be known at the receiver for decoding. When  $0 \leq j \leq M$ , the received symbol corresponding to the transmitted symbol  $x_{t,k'}$  is  $r_{t,k'} = \alpha_{t,k'}(1 - 2x_{t,k'}) + n_{t,k'}$ . When  $M < j \leq M + N$ , the received symbol is  $r_j = \alpha_j(1 - 2z_j) + n_j$ .

After performing the NCDM scheme, the output LLR values of each check node can be represented by

$$L_j = 2 \tanh^{-1} \left( \prod_{t \in \{1,2,\dots,L\}} \prod_{k' \in \Upsilon_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right) \right). \quad (4)$$

The values of the first  $M$  output LLR values can be split into three regions, i.e.,  $0 < j \leq \sum_{i'=1}^{i-1} S_{i'}$ ,  $\sum_{i'=1}^{i-1} S_{i'} < j \leq \sum_{i'=1}^i S_{i'}$ , and  $\sum_{i'=1}^i S_{i'} < j \leq M$ . For the regions where  $0 < j \leq \sum_{i'=1}^{i-1} S_{i'}$  and  $\sum_{i'=1}^i S_{i'} < j \leq M$ , the output LLR values are  $L_j = 0$ . For the region where  $\sum_{i'=1}^{i-1} S_{i'} < j \leq \sum_{i'=1}^i S_{i'}$ , the output LLR value can be written as  $L_j = l_{i,j}$ .

For the last  $N$  output LLR values, i.e.,  $M < j \leq M + N$ , we expect that they are in the form of  $L_j = l_{i,j}$ . Here  $l_{i,j}$  represents the LLR value for the  $j$ th received symbol at the destination node  $D_i$ . However, the  $j$ th output LLR value when  $M < j \leq M + N$  can be written as (6).

For the first part  $\prod_{t,t \in \{1,\dots,L\} \setminus i} \prod_{k' \in \Upsilon_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right)$  in (6), its expected value is 1. Let  $|x|$  represent the absolute value of  $x$ . Since  $0 \leq \left| \tanh \left( \frac{l_{t,k'}}{2} \right) \right| < 1$ , the more product terms there are, the smaller the value  $\left| \prod_{t,t \in \{1,\dots,L\} \setminus i} \prod_{k' \in \Upsilon_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right) \right|$  is. Thus, in order to obtain better performance, the column weight of the generator matrix of each user should be as small as possible. Also it can be seen that the larger the value of  $L$  is, the smaller the value  $\left| \prod_{t,t \in \{1,\dots,L\} \setminus i} \prod_{k' \in \Upsilon_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right) \right|$  is. It means that as the number of the parallel sessions increases, the value of this part becomes smaller. Accordingly, the performance of the system becomes worse.

For the second part  $\tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} \oplus \sum_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'} \right) \right) + n_{i,j} \right) \right]$  in (6), its expected form is  $\tanh \left( \frac{l_{i,j}}{2} \right)$ , i.e.,  $\tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} \right) \right) + n_{i,j} \right) \right]$ . Now, we focus on analyzing the second part of (6). There are four possible cases for the values of the combination of  $\bigoplus_{k' \in \Upsilon_j^i} x_{i,k'}$  and  $\bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'}$ . The four cases are shown below

$$1: \bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'} = 0, \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} = 0$$

$$2: \bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'} = 0, \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} = 1$$

$$3: \bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'} = 1, \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} = 0$$

$$4: \bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'} = 1, \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} = 1$$

For the first two cases, i.e., case 1 and case 2, the second part of (6) can be written as the expected form, i.e.,  $\tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} \right) \right) + n_{i,j} \right) \right]$ . For case 3, the second part of (6) can be written as  $\tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} \right) \right) + n_{i,j} \right) - \frac{4\alpha_{i,j}^2 E_b}{N_0} \right]$ . And for case 4, the second part can be written as  $\tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \bigoplus_{k' \in \Upsilon_j^i} x_{i,k'} \right) \right) + n_{i,j} \right) + \frac{4\alpha_{i,j}^2 E_b}{N_0} \right]$ . In order to achieve the expected form of the second part, we should try to avoid the case 3 and case 4.

Therefore, the value of  $\bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'}$  is expected to be zero. It can be seen that the value of this part has direct relationship with the number of the parallel sessions  $L$ .  $\bigoplus_{t,t \in \{1,\dots,L\} \setminus i} \sum_{k' \in \Upsilon_j^t} x_{t,k'}$  is the sum of the transmitted symbols from the  $L-1$  sessions except the  $i$ th session, where  $x_{t,k'}$  equals to 0 or 1 with equal probability, i.e.,  $P_r(x_{t,k'} = 0) = P_r(x_{t,k'} = 1) = \frac{1}{2}$ .

### C. Code Design Criteria

As mentioned in Section III,  $\mathbf{G}_i$  is the generator matrix seen from the  $i$ th destination node  $D_i$ .  $\mathbf{G}_i^t$  is the  $t$ th submatrix of  $\mathbf{G}_i$ . Based on the above analysis, we can summarize the code design criteria as

- When  $L$  is an even number, the column weight of each submatrix of the generator matrix  $\mathbf{G}_i$ , i.e.,  $\mathbf{G}_i^t$ , where  $t \in \{1, 2, \dots, L\}$ , must be an even number.
- When  $L$  is an odd number, the column weight of each submatrix of the generator matrix  $\mathbf{G}_i$  can be an odd or even number.

## V. SIMULATION RESULTS

In this section, we present simulation results for the parallel session model with the NCDM scheme. We consider a WSN with  $L = 3$  and 4 transmission sessions, respectively. Each transmission session has 500 source nodes, sending data packets to their corresponding destination nodes via a common multi-hop relay network. There are 1500 relay nodes when  $L = 3$ , and 2000 relay nodes when  $L = 4$ . We assume that all source nodes have data packets to transmit. All the source-relay channels in the network are spatially independent and have equal transmitting power. In the presented figures, the label "PSM" corresponds to the parallel session model, and "SSM" refers to the serial session model.

Fig. 2 (a) presents the performance comparison of various column weights of  $\mathbf{G}_i^t$ , (i.e.,  $\rho = 2$  and  $\rho = 3$ ), for an even number of sessions (i.e.,  $L = 4$ ). When the number of sessions  $L$  is an even number, if the column weight  $\rho$  of  $\mathbf{G}_i^t$  is an even number, the BER performance gap between the serial session model and parallel session model is smaller than the

$$\begin{aligned}
L_j &= 2 \tanh^{-1} \left\{ \prod_{t,t \in \{1, \dots, L\} \setminus i} \prod_{k' \in \mathcal{T}_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right) \tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \oplus_{t,t \in \{1, \dots, L\}} \sum_{k' \in \mathcal{T}_j^t} x_{t,k'} \right) \right) + n_{i,j} \right) \right] \right\} \\
&= 2 \tanh^{-1} \left\{ \prod_{t,t \in \{1, \dots, L\} \setminus i} \prod_{k' \in \mathcal{T}_j^t} \tanh \left( \frac{l_{t,k'}}{2} \right) \tanh \left[ \frac{2\alpha_{i,j} E_b}{N_0} \left( \alpha_{i,j} \left( 1 - 2 \left( \oplus_{k' \in \mathcal{T}_j^i} x_{i,k'} \oplus \sum_{t,t \in \{1, \dots, L\} \setminus i} \sum_{k' \in \mathcal{T}_j^t} x_{t,k'} \right) \right) + n_{i,j} \right) \right] \right\} \quad (6)
\end{aligned}$$

one for the case when  $\rho$  is an odd number. Note that when the column weight  $\rho$  is an odd number, the BER performance of the parallel session model gets worse. Fig. 2 (b) presents the performance comparison of various column weights of  $\mathbf{G}_i^t$ , (i.e.,  $\rho = 2$  and  $\rho = 3$ ), with an odd number of sessions (i.e.,  $L = 3$ ). The smaller the column weight  $\rho$  of  $\mathbf{G}_i^t$  is, the closer the BER performance of parallel session model with NCDM to that of the serial session model. From Fig. 2 (b) we can see that when  $\rho = 2$ , the BER performance of the parallel session model is very close to the one for the serial session model. The observation is consistent with the theoretical results.

larger than that of the AWGN channel. That's because the effect of the residual interference from other sessions to the performance is relatively large for fading channel.

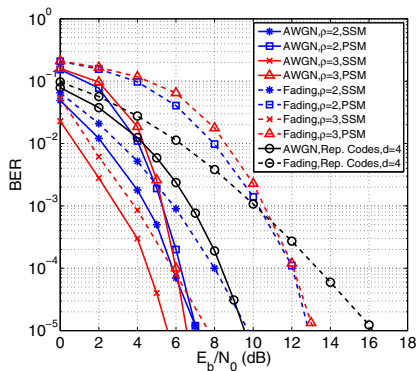
For comparison, the BER performance for repetition codes is also plotted. After using the NCDM scheme at each destination node, the code rate becomes  $\frac{1}{L+1}$ . Thus we compare the performance of our proposed scheme with the repetition code with parameter  $d = L$ , i.e., the messages are repeated  $L$  times. From Fig. 2, it can be seen that the proposed system achieves a significantly BER performance gain compared with the system with the repetition codes in a high signal to noise ratio (SNR) region.

## VI. CONCLUSION

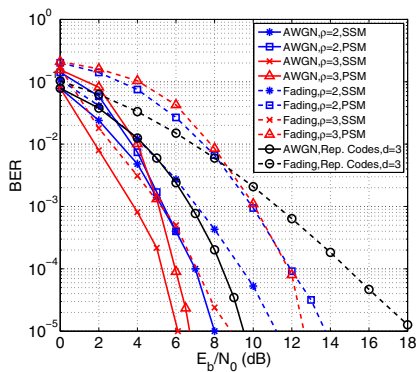
In this work, we proposed a novel network code division multiplexing scheme for the parallel session model. We analyzed the soft processing algorithm of the NCDM scheme. Based on the analytical results, a code design criteria for the generator matrix construction method was proposed. Simulation results showed that the problems caused by the NCDM scheme can be managed effectively and the BER performance can be improved significantly by following the design criteria.

## REFERENCES

- [1] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 283–289, Feb. 2006.
- [2] B. Zhao and M. Valenti, "Distributed turbo codes: Towards the capacity of the relay channel," *IEEE VTC'03-Fall*, vol. 1, pp. 322–326, Oct. 2003.
- [3] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1195–2006, Jul. 2006.
- [4] A. Chakrabarti, A. Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 280–291, Feb. 2007.
- [5] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.
- [6] M. Dohler, Y. Li, B. Vucetic, A. Aghvami, M. Arndt, and D. Barthel, "Performance analysis of distributed space-time block-encoded sensor networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 6, pp. 1776–1789, Nov. 2006.
- [7] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [8] K. Pang, Z. Lin, Y. Li, and B. Vucetic, "Design of distributed network-channel codes for wireless sensor networks," in *Proc. of IEEE ICC'11*, Kyoto, Japan, Jun. 2011, pp. 1–5.
- [9] (2011) Newington smart village. [Online]. Available: <http://www.ausgrid.com.au/Common/Network-projects/Network-projects/Smart-grid-projects/Smart-Grid-Smart-City/Newington-Smart-Village.aspx>
- [10] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*, 2nd ed. New Jersey: Prentice Hall, 2004.
- [11] J. Gentle, *Matrix Algebra: Theory, Computations, and Applications in Statistics*, 1st ed. New York: Springer, 2007.
- [12] J. Yue, K. Pang, Z. Lin, Y. Li, B. Bai, and B. Vucetic, "Distributed network channel coding for multiple access relay interference channels," in *Proc. IEEE Globecom'12*, Anaheim, CA, USA, Dec. 2012, pp. 1–6.



(a) The number of sessions is  $L = 4$ .



(b) The number of sessions is  $L = 3$ .

Fig. 2. The BER performance comparison between two different column weights of generator matrix, i.e.,  $\rho = 2$  and  $\rho = 3$ , over AWGN and Rayleigh fading channels, respectively. The number of sessions is  $L = 3$  and 4.

For the Rayleigh fading channel, the BER performance gap between the serial session and parallel session model is much