

Clustering Analysis of Wireless Sensor Network Based on Network Coding with Low-Density Parity Check

Maria Hammouti

Dept. Electronics, Informatics and Telecommunications
ENSAO, Mohammed I University
Oujda, Morocco

El Miloud Ar-reyouchi

Dept. Telecommunication and Computer Science
Abdelmalek Essaadi University
Tetouan, Morocco

Kamal Ghoumid

Dept. Electronics, Informatics and Telecommunications
ENSAO, Mohammed I University
Oujda, Morocco

Ahmed Lichioui

Broadcasting TV/FM
SNRT
Rabat, Morocco

Abstract—The number of nodes in wireless sensor networks (WSNs) is one of the fundamental parameters when it comes to developing an algorithm based on Network Coding (NC) with LDPC (Low Density Parity Check) code because it directly affects the size of the generator matrix of the LDPC code and to its dispersion. Optimizing Wireless Communication Systems by decreasing BER (Bit Error Rate) is one approach to analyze the network into clusters (at the level of their nodes). In this paper, the authors present a fully distributed clustering algorithm and they consider different node values by cluster, then they select the curves that have the best compromise. They examine the effects of SNR (Signal-to-noise ratio) quantization on system performance obtained for different scenarios (by varying the parameter corresponding to the number of the symbol during the forwarding phase). Finally, the results prove that the increased number nodes improve LDPC code properties.

Keywords—*Clustering Techniques; Network Coding; LDPC codes; distributed algorithms; wireless sensor network*

I. INTRODUCTION

A wireless sensor network (WSN) is a group of specialized transducers with a communications infrastructure for target tracking, environmental monitoring and recording conditions at diverse locations. WSN consists of multiple detection stations called sensor nodes, each of which is small, lightweight and portable.

Hierarchical clustering [1], [2] is particularly useful for applications that require scalability to hundreds or thousands of nodes. Moreover, clustering [3] can stabilize the network topology at the level of sensors and thus cuts on topology maintenance overhead. The need for clustering in WSNs has been first motivated, and a brief description of the implied hierarchical network pattern has been given. The sensor nodes periodically transmit their data to the corresponding particular node called cluster head (CH) nodes. The CH nodes aggregate the data and broadcast them to the base station (BS) either directly or through the intermediate communication with other CH nodes.

NC [4] is a recent field of the information theory that breaks with this assumption. Instead of only forwarding data, nodes may recombine several input packets into one or several output packets.

The concept of NC has been first introduced for satellite communications in [5] and then fully developed in [4] for general networks.

With NC, a network node allows combining several packets (messages) that it has generated or received into one or several outgoing packets. NC replaces the conventional routing (data packets have given by store-and-forward mechanisms in which the intermediate nodes only repeat data packets that they have received).

Recently, NC has been applied to wireless networks and received significant popularity as a means of improving network capacity and the End-to-End Delays of transmissions [6]. It can also increase the capacity and reduces the delay [7] in networks and at point-to-point communications [8], in broadcast networks [9]. NC increases the energy efficiency in broadcast networks [10]; it optimizes capacity and energy consumption in the network [11], [12]. Also, it can be combined with routing for improved performance [13], and adaptive algorithms of appropriate node mobility networks [14].

In fact, the unreliability and broadcast nature of wireless links make wireless networks a natural setting for network coding. Moreover, network protocols in wireless networks, e.g., wireless mesh networks and mobile ad hoc networks, are not fully developed yet and hence there is more freedom to apply network coding in such environments compared to wire line networks such as the Network Management System [6].

Low-Density Parity Check (LDPC) codes [15] deliver very good performance when decoded with SPA [16]. As LDPC codes are being used in a wide range of applications, the search for efficient implementations of decoding algorithms is being pursued intensively.

Error-correcting coding schemes using LDPC codes and belief propagation decoders based on the sum-product algorithm (SPA) have recently achieved some of the highest performance results in the literature. The SPA has been presented to decode LDPC codes.

The main contribution in this paper is reducing the BER value (through a distributed coding using LDPC codes) in ANCC (Adaptive Coded Cooperation Network) for WSN by using a very effective method adapted to this type of network. It studies two novel scopes in communication. The authors focus on the analysis of the BER. Regarding NC, they propose a coding algorithm adequate for dense and extensive. They create a set of scenarios in which several parameters have been modified to analyze how they affect the BER versus SNR curve. What is more, this paper aims to get the right values of these parameters to find a compromise, which seeks to decrease the BER.

The rest of the paper is organized as follows. In Section II, the model on which to base our approach is described. In section III, the authors discuss the results; they compare the different approaches, and they analyze their applicability. The conclusions as well as future work are discussed in Section IV.

II. NETWORK MODEL OVERVIEW

A. Description Working Scenario and ANCC Scheme

The network coding scheme to be used as a basis for development is the ANCC [18],[19] that adapt precisely to extensive and dense network like WSN, on which each fixed period, the N sensor nodes should send the information collected to the CH. To avoid interference, in the type of WSN in question, it is often established access scheme TDMA (Time Division Multiple Access) in which each node has a particular time for the transmission while all other nodes remain in silence.

The fundamental idea of the scheme ANCC network coding is that neighboring nodes can linearly combine the correctly received information and send it to the cluster head in a later phase of the communication. In particular, the ANCC scheme uses LDGM codes (Low Density Generator Matrix) [19]. This type of LDPC code fits to WSN scenario by the structure of its generator matrix. In LDGM codes, the generator matrix \mathbf{G} is being constructed as the concatenation of two matrices: the identity matrix I_N and one dispersed and random matrix \mathbf{P} . The matrices involved in encoding and decoding are as follows [19]:

- Generator matrix \mathbf{G} , of size $N \times 2N$ (since the encoding rate is $1/2$; the matrix \mathbf{P} is of size $N \times N$):

$$\mathbf{G} = [I_N, \mathbf{P}] \quad (1)$$

- Parity check matrix \mathbf{H} of size $N \times 2N$:

$$\mathbf{H} = [\mathbf{P}^T, I_N] \quad (2)$$

- Syndrome vector δ this vector is the result of checking a word with the parity check matrix \mathbf{H} . When the decoding performed by syndrome, we must have syndrome validates a coded word c must be the null vector as follows from:

$$\delta = c \cdot \mathbf{H}^T \quad (3)$$

The Network Coding scheme ANCC gets applied the coding LDGM of distributed form:

1) *Diffusion phase (or initial transmission):* In this phase, each of the N nodes in the WSN transmitted orderly (with TDMA) its corresponding information symbol (systematic symbol) to the head cluster, while the rest of the nodes remain in listening mode.

2) *Forwarding phase (or combined retransmission combined):* the N nodes transmit a linear combination (coded symbol). It composed of symbols s , randomly chosen from the symbols of a node, which had received correctly in the previous phase.

3) *Decoding phase:* once the Forwarding phase has been completed, the CH has already received all the symbols (systematic and codified) and it is capable constructing the parity check matrix \mathbf{H} , using, for this, the headers received in Forwarding phase. After building the matrix \mathbf{H} , the cluster head decodes the received symbols obtaining gain codification (decrease of BER) proper to the use of an LDGM code.

B. Topologies

Hierarchical network allows assigning different roles to nodes; exploiting that to the control node and link activity. Grouping sensor nodes into clusters has been widely pursued by the research community to achieve the network scalability objective. Every cluster would have a leader, often referred to the CH. The strategy is to divide the total set of nodes in several clusters or groupings involving a high number of nodes of each one. To ensure the functioning of the system, each cluster is chosen so that all its elements may have a direct link to the most important node of the cluster, which has already called the cluster head (CH) node (see Figure 1). This mode takes charge of being the destination node broadcasts performed by the other nodes of the cluster, being the element that receives and decodes all the information of this grouping (cluster) and, therefore, allows data communication in a clustered network.

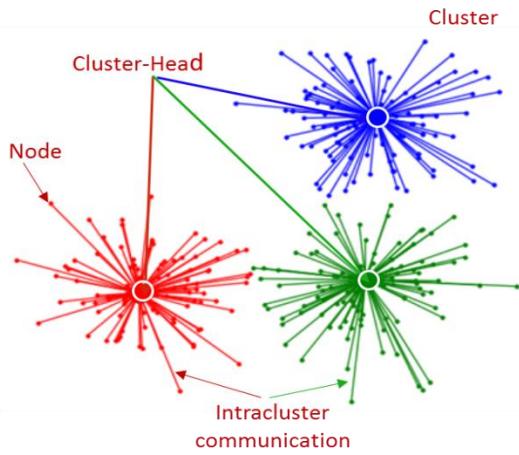


Fig. 1. Network model with clustering

C. Parameters affecting our system

Grouping sensor nodes into clusters has been widely pursued by the research community to achieve the network scalability objective. Every cluster would have a leader often referred to as the CH.

The main parameters to have analyzed for the scenario are:

1) *The number of iterations of the Sum-Product algorithm SPA*: It is analyzed for a concrete example the improvement of having a greater number of iterations in the SPA.

2) *The parameter s*: Several options were tested to see how it affects the number of nodes involved as a response of linear combination sent by each node during forwarding phase to the BER obtained in an ANCC scheme.

3) *The number of nodes per cluster*: In this case, it is mainly analyzed how the number of nodes that intervene in the distributed encoding LDPC affects the BER and how the results present themselves for various options.

We define a relationship between the symbols, which has combined in the forwarding phase and the number of nodes that can intervene in this combination. The Dispersion Grade (DG) refers to the dispersion having the generator matrix of the LDPC code, this DG is calculated as:

$$DG = \frac{s}{N - 1} \quad (4)$$

Where s is a parameter that indicates the values of the coded symbol (randomly chosen) which are correctly received in the previous phase, N is the total number of nodes. A parameter DG is the percentage of symbols involved in the response of forwarding phase (among the symbols that could intervene). The goal is to see if it can get a range of values of DG between which algorithm behave optimally, with in order to establish a priori the values of the parameter s and the number of nodes that give the best BER-SNR performance.

D. Methods of Performance Evaluation

A typical BER curve, for LDPC code, is shown in figure 2. Three regions [20] namely low SNR region, waterfall region, and error floor region, can be identified. In the low SNR region, BER decreases slowly as SNR increases. For the intermediate value of SNR, the BER decreases rapidly in the waterfall region with an increase in SNR. In this region, the coding gain approaches the theoretical limit. The error Floor is due to the weight distribution of LDPC codes. Normally, LDPC codes do not have large minimum distances. Hence, lowering the Error floor region result in better codes, which in some cases, may result in faster convergence in decoding.

The method have to be followed to determine which scenario is better than another is based on the inspection of their BER versus SNR curves, choosing to get a better compromise between the values of BER obtained at high and low values of SNR. The studies carried out show that there is a relationship between the position of the threshold th_1 , the slope in the Waterfall region and BER values obtained in the of Error Floor region.

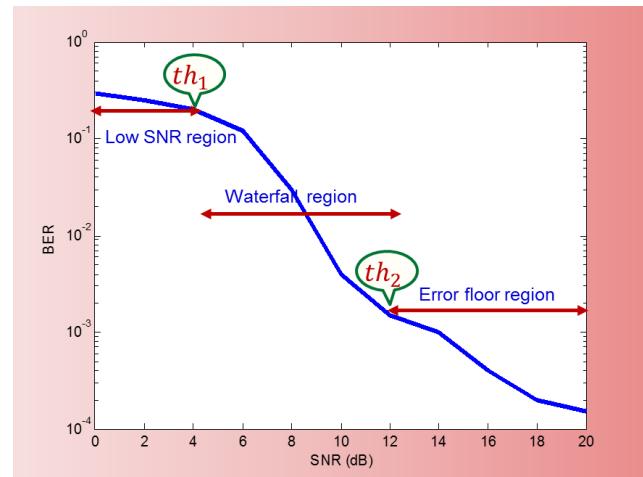


Fig. 2. Regions BER-SNR curve for a LDPC codes

Therefore, the goal is to find a compromise among which this threshold th_1 , it is located at the lowest possible, so that the SNR for BER values obtained in the region of Error Floor will be better and this region appears for reasonable SNR values in a wireless communication.

III. RESULTS AND DISCUSSION

In this section, we analyze the data results obtained from simulations for the clustering analysis of different scenarios, which corresponds to the sending and forwarding the data nodes of each aggrupation of their cluster head node. We start by cutting the network in groups of nodes called clusters, thereby giving, at the network, a hierarchical structure [21].

Within each phase, behavior is also studied in each cluster for different values of the parameter s .

For the simulations, we use a total of 1500 TDMA network nodes, depending on the number of nodes per cluster and so that it can give an exact value of cluster head nodes. We have 1500 nodes, which means $N=1500$ nodes.

The number of nodes defined as N , the size of the generator matrix will be $N \times 2N$. This leads to one of the objectives is to make a cluster of large size, encompassing the largest possible number of nodes. If the total number of nodes is defined as N_{Total} and divides into regular cluster of N nodes, therefore, number of cluster head node $N_{Cluster}$ is:

$$N_{Cluster} = \frac{N_{Total}}{N} \quad (5)$$

Table I gives the exact value of cluster head.

TABLE I. TOTAL NUMBER (1500 NODES) DIVIDED BY CLUSTER NODE FOR OBTAINED AN EXACT CH NUMBER

Number of nodes	Number of cluster head
150	10
250	6
375	4
500	3

A. Influence of the number of iterations

Before embarking on an approach, analyzing the results of different nodes per cluster there is a need to explore the influence of iteration number. By Sum-Product Algorithm, we measured, with MATLAB Profiler tool, a distributed LDPC coding for a WSN with 100 nodes.

SPA performs a series of operations on the received symbols so that they can comply with the condition of the syndrome shown in equation 3.

The aim is to analyze if the use of a large number of iterations improve, significantly, BER values obtained to compensate the time delay that occurs during decoding.

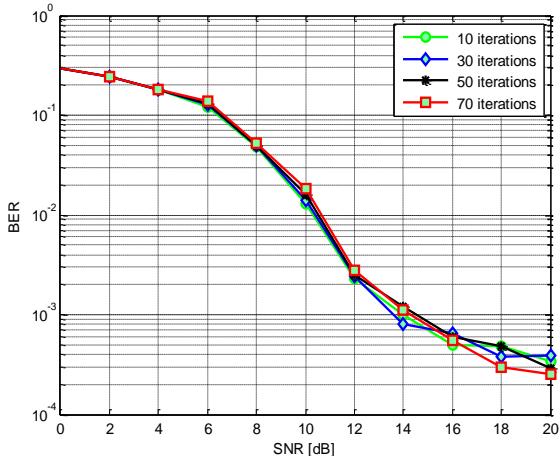


Fig. 3. BER-SNR curve obtained for a distributed LDPC coding in WSN 150 nodes. The combinations of the forwarding phase $s = 7$ nodes and the maximum number of iterations of the decoding algorithm is varied by taking the values 10, 30, 50 and 70

Figure 3 shows the different curves obtained in simulations for the case of a cluster of 150 nodes where the value of the parameter s is 7, using different maximum numbers of iterations in the decoding algorithm.

Then we concluded that increasing the number of iterations does not offer a significant advantage. Therefore, from now on, we use 10 (as maximum number for the sum-product algorithm) iterations in our system for all decoders.

B. Clustering analysis

1) Clusters of 60 nodes

We analyze one grouping of 60 nodes. In this case, our approach is to compare simulation results with experimental data to obtain valid and accurate results.

a) Simulation results

With 60 nodes, the generator matrix will have a size of 60×120 , which it could be considered a small size; due to high DG. The range in which we study the behavior of the parameter s is comprised between 4 and 8 in order to offers a better SNR-BER curve.

The figure 4 shows the different curves obtained for a grouping of 60 nodes by varying the value of the parameter s and using a maximum of 10 iterations of the decoding SPA.

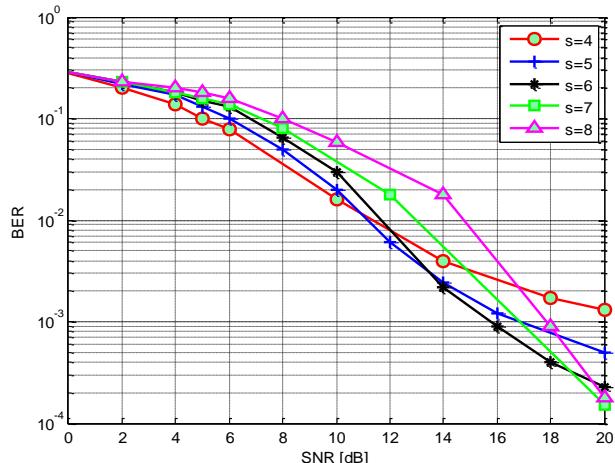


Fig. 4. BER-SNR curve obtained for a distributed LDPC coding in a 60 WSN nodes and the value of the parameter s varies between 4 and 8

The curve, which presents a better compromise between the position of the threshold values u_1 and BER obtained in the region of error floor, is the curve with $s = 6$. DG (with the principal parameters) presenting these curves are summarized in Table II.

TABLE II. DG FOR 60 NODES AND DIFFERENT VALUES OF THE PARAMETER S

Parameters s	DG
4	6.77 %
5	8.47 %
6	10.16 %
7	11.86 %
8	13.55 %

These values of the DG are too high and may be associated to sparse matrices that widely dispersed, which does not appear good option to use the groupings of this size.

b) Experimental Results

To verify our next simulation results, we analysis the experimental results, we choose 60 sensor nodes within a cluster view the Figure 5.

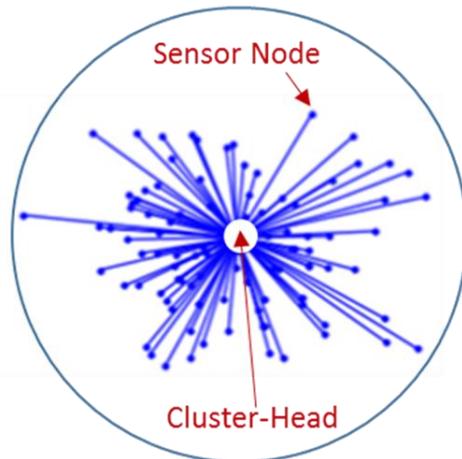


Fig. 5. Experimental topology of the wireless sensor network a (cluster of 60 nodes) the value of the parameter s is 6

The following figure (Fig 6) shows the experimental results.

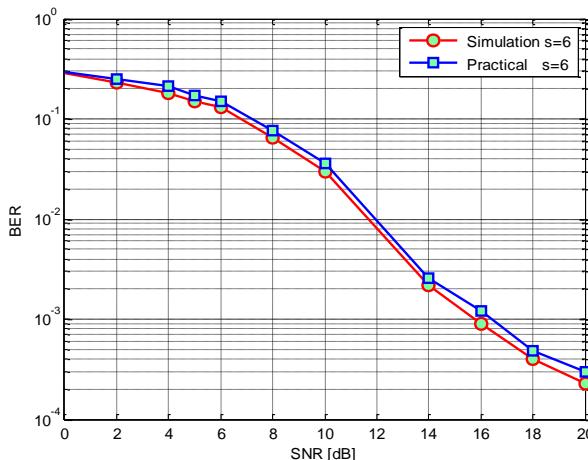


Fig. 6. Comparison between Simulations and Experimental Results obtained for a distributed LDPC coding in a 60 WSN nodes, and the value of the parameter s is 6

We note that the results of this simulation are very similar indeed; to those obtained by practice. Consequently our preliminary simulation results fully confirm the validity of our next simulations results.

2) Clusters of 150 nodes

With 150 nodes, the generator matrix of network coding with LDPC codes will have a size of 150x300. In addition, the parameter s is varied from 5 to 11 (to view which of them offers a BER versus SNR curve with a better compromise) while the SPA is kept constant at 10 iterations. Figure 7 shows the different curves obtained for a cluster containing 150 nodes by varying the value of the parameter s . It can be appreciated that for values of s , greater than 8, the Error Floor region appears for very high SNR values (over 18 dB), which provokes a good BER obtained at very high SNR values, often situated above the range of SNR in which is located a wireless system.

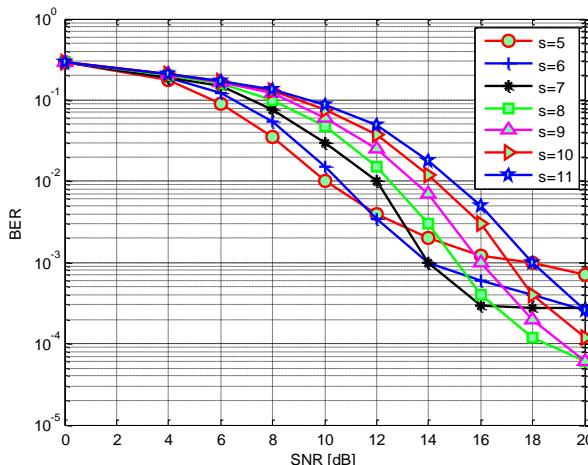


Fig. 7. BER-SNR curve obtained for a distributed LDPC coding in a 150 WSN nodes. The value of the parameter s varies between 5 and 11

The DG values are summarized in the following table.

TABLE III. DG FOR 150 NODES AND DIFFERENT VALUES OF THE PARAMETER S

Parameters s	DG
6	4.02 %
7	4.69 %
8	5.36 %

By Analyzing the graphs, we can see that the curves, with greater compromise, are given for the parameter values $s = 6, 7$ and 8 , whose matrices offer the DG that are displayed in Table III

3) Clusters of 250 nodes

In this case, all nodes are divided into several clusters of 250 nodes. With 250 nodes, the generator matrix of network coding with LDPC codes will have a size of 250x500. In addition, the parameter s is varied from 7 to 12, while the SPA always keeps the same value 10 iterations.

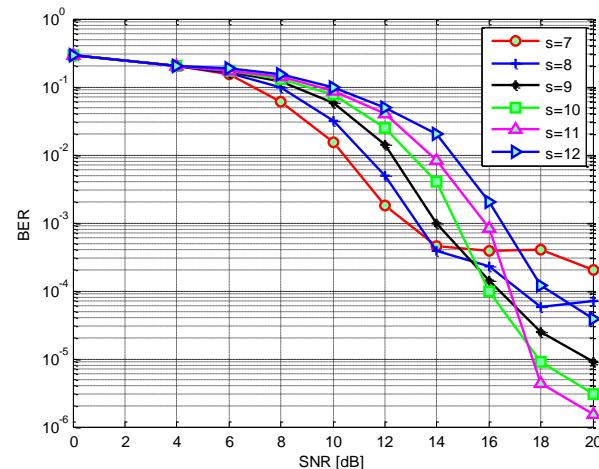


Fig. 8. BER-SNR curve obtained for a distributed LDPC coding in WSN 250 nodes. The parameter s varies between 7 and 12

The figure 8 represents the different curves obtained for a cluster of 250 nodes by varying the value of the parameter s and using a maximum of 10 iterations of the decoding SPA.

This figure shows that BER-SNR curves which obtain a greater compromise and whose DG are shown in the Table IV.

TABLE IV. DG FOR 250 NODES AND DIFFERENT VALUES OF THE PARAMETER S

Parameters s	DG
8	3.21 %
9	3.61 %
10	4.01 %

4) Clusters of 375 nodes

In this case, we divide the nodes into clusters of 375 nodes. With 375 nodes, the generator matrix of the NC with LDPC code has a size of 375x750. In addition, as in previous cases, the parameter s is varied from 7 to 11 and BER versus SNR.

The Figure 9 shows the different curves obtained for a cluster of 375 nodes, varying the value of the parameter s . In

all, the simulations were performed using a maximum of 10 iterations of the algorithm Sum-Product decoding

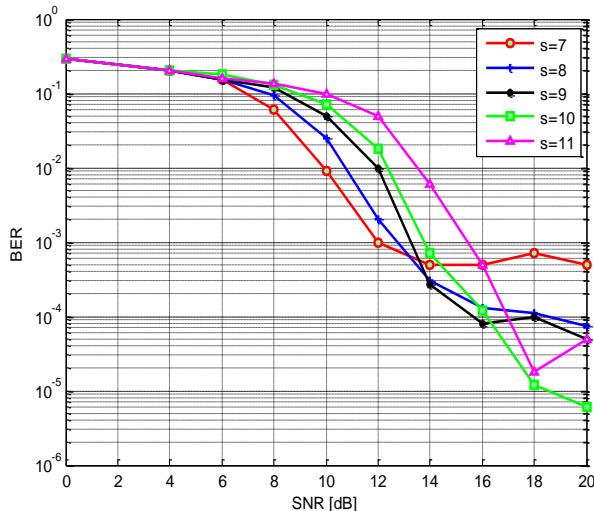


Fig. 9. BER-SNR curve obtained for a distributed LDPC coding in WSN 375 nodes. The parameter s varies between 7 and 11

By observing the resulting graphics to these groups, it can be seen that the best curves are obtained for values of the parameter s between 8 and 10. For these parameter values s its DG are marked in Table V.

TABLE V. DG FOR 250 NODES AND DIFFERENT VALUES OF THE PARAMETER S

Parameters s	DG
8	2.13 %
9	2.40 %
10	2.67 %

5) Clusters of 500 nodes

In the final case, the total nodes are divided into several clusters (grouping) so that each cluster includes 500 nodes.

Figure 10 shows the curves obtained in simulations clusters of 500 nodes, using a maximum of 10 iterations of the decoding algorithm Sum-Product

With 500 nodes, the generator matrix of the LDPC code has a size of 500×1000 . In addition, the parameter s is varied between 8 and 13 and the BER simulation results versus SNR are shown in the following figure.

By observing the resulting graphics to these groups, it can be seen that the curves, that offer a better compromise, are given for the parameter values s comprised between 9 and 12. The DGs obtained are detailed in Table VI.

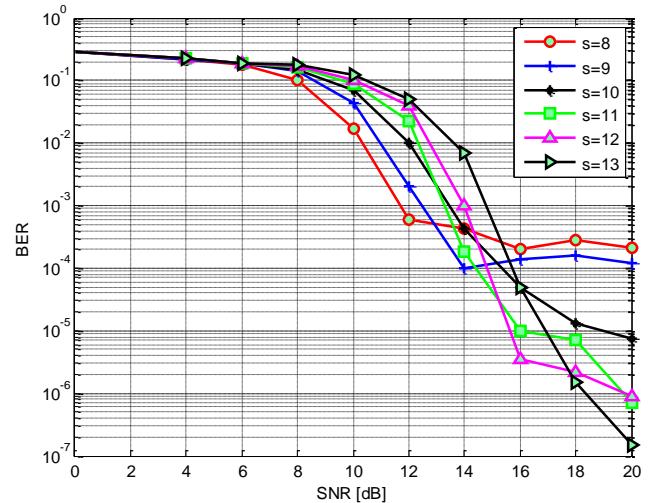


Fig. 10. BER-SNR curve obtained for a distributed LDPC coding in WSN 500 nodes. The parameter s varies between 8 and 13

TABLE VI. DG FOR 500 NODES AND DIFFERENT VALUES OF THE PARAMETER S

Parameters s	DG
9	1.8 %
10	2 %
11	2.2 %
12	2.4%

C. Better options of Clustering analysis

We examine the best BER versus SNR curves obtained for the scenarios described above. We choose for each scenario the best compromise solution.

We compare and analyze results of the proposed scenarios:

- Cluster of 150 nodes and parameter $s = 7$.
- Cluster of 250 nodes and parameter $s = 9$.
- Cluster of 375 nodes and parameter $s = 10$.
- Cluster of 500 nodes and parameter $s = 11$.

We have the following figure 11

Dispersion grades that correspond to those curves are shown in Table VII.

The curves represented in the figure 11 are the curves that offer better compromise for each cluster of those studied.

By analyzing the curves obtained in Figure 11, we see that the slope of the curve BER-SNR in the Waterfall region increases when using a greater number of nodes per cluster.

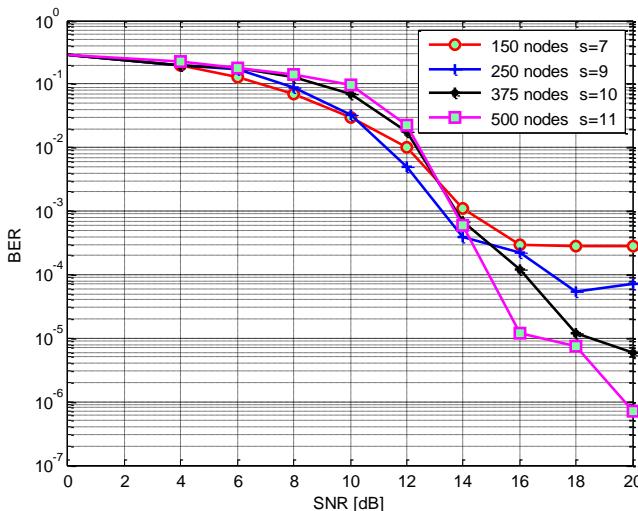


Fig. 11. Comparison of BER-SNR curves obtained for a distributed LDPC coding which offers a better compromise

TABLE VII. DG FOR BETTER OPTIONS OF DIFFERENT VALUES OF THE PARAMETER S

Number of nodes	Parameters s	DG
150	7	4.69 %
250	9	3.61 %
375	10	2.67 %
500	11	2.2 %

IV. CONCLUSIONS

We examine and we analyze the best BER curves versus SNR obtained for different scenarios. In light of these results, it can be concluded that the best BER vs. SNR curves obtained correspond to the clusters of 375 and 500 nodes. As a result, we prove that an increasing the number of the node improves LDPC code properties. The application of the LDPC code with NC can improve consistently, considerably and significantly the BER versus SNR when we increase the number of node per cluster in Hierarchical clustering for WSN. On the other side, it is also important to mention that the increase of iterations number has a negligible impact on the performance of our system.

Our future work (which is the next phase of the present paper) is to improve the BER in WSN at the level of clusters heads using NC with low-density parity-check (LDPC) codes.

ACKNOWLEDGMENT

This research was made possible in part by the « Société nationale de radiodiffusion et de télévision »SNRT.

REFERENCES

- [1] C.R. Lin and M. Gerla, Adaptive clustering for mobile wireless networks, IEEE Journal on Selected Areas Communications, 15(7), 1997,pp:1265–1275.
- [2] S. Banerjee and S. Khuller, A clustering scheme for hierarchical control inmulti-hop wireless networks, in Proceedings of 20th Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'01), Anchorage, AK, April 2001.
- [3] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, Next century challenges: scalable coordination in sensor networks, in Proceedings of the ACM/IEEE MOBICOM Intl. Conference, Boston, MA, August 2000, pp. 6–11.
- [4] R. Ahlsweide, N. Cai, S. R. Li, and R. W. Yeung, Network information flow, IEEE Transactions on Information Theory, July 2000.
- [5] R. W. Yeung and Z. Zhang, “Distributed source coding for satellite communications,” IEEE Trans. Inform. Theory, 45(4), May 1999, pp. 1111–1120.
- [6] El Miloud AR REYOUCHII, Kamal Ghoumid, Koutaiba Ameziane, and Otman El Mrabet, “The Improvement Of End To End Delays In Network Management System Using Network Coding”, International Journal of Computer Networks & Communications (IJCNC), 5(6), Novembre 2013, pp:65-84.
- [7] El Miloud Ar reyouchi, Kamal Ghoumid, Koutaiba Ameziane, and Otman El Mrabet, “The potentials of Network Coding for improvement of Round Trip Time in wireless Narrowband RF communications”, Publisher: IEEE , 2014 , pp: 765 – 770.
- [8] Katti, S., Rahul, H., H., W., Katabi, D., Medard, M., y Crowcroft, J. Xors in the air: Practical wireless network coding. IEEE/ACM Transactions on Networking, 16(3), 2008:pp:497-510.
- [9] Eryilmaz, A., Ozdaglar, A., Medard, M., y Ahmed, E. On the delay and throughput gains of coding in unreliable networks. IEEE Transactions on Information Theory, 54(12)2008,pp:5511-5524.
- [10] Fragouli, C., Widmer, J., y Le Boudec, J.-Y. Efficient broadcasting using network coding. IEEE/ACM Transactions on Networking, 16(2)2008:450-463.
- [11] Sagduyu, Y. E. y Ephremides, A. On joint mac and network coding in wireless ad hoc networks. IEEE Transactions on Information Theory, 53(10) ,2007:3697-3713.
- [12] Sagduyu, Y. E. y Ephremides, A. Cross-layer optimization of mac and network coding in wireless queueing tandem networks. IEEE Transactions on Information Theory, 54(2)2008:554-571.
- [13] Sengupta, S., Rayanchu, S., y Banerjee, S. An analysis of wireless network coding for unicast sessions: The case for coding-aware routing. In Proceedings of the IEEE International Conference on Computer Communications, 2007. INFOCOM'07,2007, pp. 1028-1036.
- [14] Yang, D.-N. y Chen, M.-S. Data broadcast with adaptive network coding in heterogeneous wireless networks. IEEE Transactions on Mobile Computing, 8(1)2009:109-125.
- [15] R. G. Gallager, Low Density Parity Check Codes. Cambridge, MA: MIT Press, 1963.
- [16] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” IEEE Trans. Inf. Theory, vol. 45, no. 3,1999: pp. 399–431.
- [17] Xingkai Bao and Jing Li. Matching code-on-graph with network-on-graph: Adaptive network coding for wireless relay networks. In Proc. Allerton Conf. on Commun., Control and Computing IL, 2005.
- [18] Xingkai Bao and Jing Li. Adaptive network coded cooperation (ancc) for wireless relay networks: matching code-on-graph with network-on-graph. Wireless Communications, IEEE Transactions on, 7(2) Febraruay 2008:574 –583.
- [19] Garcia-Frijs, J. y Zhong, W. Approaching shannon performance by iterative decoding of linear codes with low-density generator matrix. IEEE Communication Letters, 7(6) 2003 :266-268.
- [20] Tao Tian, C. Jones, J.D. Villasenor, and R.D. Wesel. Construction of irregular ldpc codes with low error floors. In Communications, 2003. ICC '03. IEEE International Conference on, volume 5, may 2003.
- [21] Colette Johnen and Le Nguyen. Self-stabilizing weight-based clustering algorithm for ad hoc sensor networks. In Algorithmic Aspects of Wireless Sensor Networks,2006,pp 83–94.