**Wireless Multimedia Communications Laboratory**
**School of Electrical & Computer Engineering**
**College of Engineering**
**University of Tehran**

# Resource Optimized Distributed Source Coding for Complexity Constrained Data Gathering Wireless Sensor Networks

Hamidreza Arjmandi, Farshad Lahouti

To Appear in IEEE Sensors Journal, 2011.

DOI: 10.1109/JSEN.2011.2109947

# Resource Optimized Distributed Source Coding for Complexity Constrained Data Gathering Wireless Sensor Networks

Hamidreza Arjmandi, Farshad Lahouti

***Abstract*— This paper addresses the problem of efficient data gathering based on distributed source coding (DSC) in wireless sensor networks (WSN) with a complexity constrained data gathering node (DGN). A particular scenario of interest is a cluster of low complexity sensor nodes among which, one node is selected as the cluster head (CH) or the DGN. Utilizing DSC allows for reducing the required rate of communications by exploiting the dependency between the nodes observations in a distributed manner. We consider a DSC-based rate allocation structure, which takes into account the CH (DGN) memory and computational constraints. Specifically, this is accomplished respectively by limiting the number of nodes whose data may be stored at the CH and exploited during decoding, and the number of nodes that can be jointly (de)compressed using DSC. Based on this structure, we investigate two WSN resource optimization problems aiming at (i) minimizing the total network cost and (ii) maximizing the network lifetime. To these ends, optimal dynamic programming solutions based on a trellis structure are proposed that incur substantially smaller computational complexity in comparison to an exhaustive search. Also, a suboptimal yet high performance solution is presented whose complexity grows in polynomial order with the number of network nodes. Numerical results demonstrate that the proposed rate allocation structure and solutions, even with limited complexity, allow for exploiting most of the available dependency and hence the achievable compression gain.**

***Index Terms*—Wireless Sensor Networks, Data Gathering, Resource Optimization, Distributed Source Coding, Slepian-Wolf Coding.**

Manuscript received July 8, 2010; revised October 7, 2010 and December 3, 2010; accepted January 13, 2011.

H. Arjmandi and F. Lahouti are with the Wireless Multimedia Communications Laboratory, Center of Excellence on Applied Electromagnetic Systems, School of Electrical & Computer Engineering, University of Tehran, 14395-515, Tehran, Iran. Phone: (9821) 82084314, Fax: (9821) 88778690; http://wmc.ut.ac.ir; e-mails: {h.arjmandi, lahouti}@ut.ac.ir.

A preliminary report on this work has been presented at the IEEE International Conference on Telecommunications, Doha, Qatar, 2010.

H. Arjmandi, F. Lahouti, "Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks," *To Appear in IEEE Sensors Journal*, 2011.

## I. INTRODUCTION

WIRELESS sensor networks (WSN) have been widely researched and used in different applications recently. A data gathering sensor network consists of many small sensor nodes, which send their data to a data gathering node (DGN). The sensor nodes are restricted in energy, memory and computational capability [1]. When the data collected by the nodes are correlated, distributed source coding (DSC) can be used to reduce the required rate of communications. The Slepian-Wolf (SW) theorem states that separate encoding of two distributed sources, at the cost of increased complexity at the joint decoder, is as efficient as their joint encoding for lossless compression [2], [3]. The SW theorem also holds for the case with multiple sources [4]. Practical distributed source codes based on turbo or low-density parity check codes are presented in, e.g., [5]-[11]. These codes with large block lengths can perform near the bound predicted by the SW theorem.

Data gathering and aggregation in large WSNs are often facilitated via grouping the network nodes to a set of clusters. The nodes within a cluster communicate to a cluster head (CH) node through single-hop wireless links. The CH is often an ordinary wireless sensor node with similar complexity limitations, while it acts as a DGN within the cluster and communicates the collected and possibly aggregated nodes data to a central DGN [12]. This paper considers the problem of resource optimized distributed source coding in such complexity constrained wireless sensor clusters. This is distinct with the classic DSC, in which no complexity constraints are assumed at the DGN.

Data gathering in WSNs based on DSC is considered for different objectives in the literature [13]-[19]. Specifically, minimizing sum cost of the nodes and maximizing network lifetime are studied in [13] and [14] respectively, where an optimal DSC (with no complexity constraints) is assumed for all the nodes. In [15] and [16], cost optimal schemes for data gathering WSNs are presented based on pairwise DSC in which the dependency between pairs of nodes are exploited. The pairwise DSC setting is motivated based on WSN complexity constraints and the observation that most existing DSC schemes concern two correlated sources. Obviously, a CH may facilitate high performance distributed source decoding for a large number of nodes only at the cost of increased computational and memory requirements [8]-[11].

In this paper, we consider efficient data gathering in a WSN cluster, whose CH node is of limited complexity (memory and computational complexity) and employ asymmetric SW codes [1]. The data of the nodes are transmitted to the CH in a specific order. A SW code is used to optimally compress the data of each node before

H. Arjmandi, F. Lahouti, “Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks,” *To Appear in IEEE Sensors Journal*, 2011.

transmission, by exploiting its dependency with those of the prior nodes stored in the CH buffer as Side Information (SI). Due to the CH memory limitation, it is assumed that it may only store the data of up to $\beta$ nodes. In light of the CH computational complexity constraint, it is further assumed that it employs a DSC of limited order, $\alpha$. In fact, a simple sensor node used as the CH, may still manage the required computations within its capability by proper selection of the DSC-order ($\alpha$). Hence, the SI of each node in the transmission order is a subset of up to $\alpha - 1$ nodes within the set of $\beta$ nodes, whose data are stored at the CH. Given the limited computational and memory resources of typical WSN nodes, adjusting $\alpha$ and $\beta$ enables the designer to judiciously use these resources for an optimized performance in DSC applications. For this network, we consider two resource (rate) allocation problems for the nodes aiming at (i) minimizing the sum cost of the nodes and (ii) maximizing the network lifetime. To solve these problems an optimal dynamic programming solution based on a trellis structure is proposed. The presented algorithm optimally determines (i) an order for transmission of the nodes, (ii) identifies the set of nodes, whose data are stored at the CH in each stage and (iii) the corresponding subset selected as SI for each transmitting node. The presented algorithm obtains the optimal solution with a substantially smaller computational complexity in comparison to an exhaustive search, at the cost of a modest increase in memory requirement. To reduce the complexity to a polynomial order, we also present a suboptimal yet high performance algorithm, which is set up based on a modified trellis with tied states. Numerical results show that the proposed resource optimized DSC schemes significantly enhance the network performance in comparison with benchmark algorithms. Furthermore, the results demonstrate that the presented solutions may enable a CH with limited complexity to still capture most of the available dependency and achievable compression gain.

The rest of this paper is organized as follows. The system model is described in Section II. In Section III, the cost optimized rate allocation problem and the proposed optimal and suboptimal solutions are presented. In Section IV, the problem of network lifetime maximization is investigated. Sections V and VI, respectively present the performance evaluation and conclusions.

## II. System Model

We consider a WSN cluster with $N$ nodes that communicate with a CH. Let $\boldsymbol{\mathcal{S}}$ be the set of nodes indices, $\boldsymbol{\mathcal{S}} = \{1, 2, \dots, N\}$. For each node $i \in \boldsymbol{\mathcal{S}}$, the data that is to be sent to the CH is represented by $X_i$. The observation of a node is a random variable, that has a countable alphabet, e.g., following quantization of a continuous random variable, and is assumed to be independent and identically distributed over time. The nodes observations are spatially correlated and their joint probability mass function is known at the CH [13]. When $\boldsymbol{\mathcal{A}}$ is a subset of the nodes indexes then $X(\boldsymbol{\mathcal{A}}) \triangleq \{X_i : i \in \boldsymbol{\mathcal{A}}\}$.

A specific transmission order is identified by a permutation of the nodes indices $\boldsymbol{S}$ and is denoted by $\boldsymbol{\pi} = [\pi(1), \dots, \pi(N)]$. The transmission order also determines the decoding order at the CH. For example, for a three node network, $\boldsymbol{\pi} = [3, 1, 2]$ indicates that node 3 transmits and is decoded first, and node 2 the last. The data of a node that is received and decoded in the CH can be stored and used as SI to decode the data of the subsequent nodes in the order. As elaborated below, different transmission orders for the nodes, in a DSC-based data gathering setting, correspond to different rates for the nodes and hence different transmission cost and lifetime for the network. Therefore, the transmission order in such scenarios must be selected in an optimized manner. In a general setting without complexity constraints, considering the permutation $\boldsymbol{\pi}$, the SI nodes for node $\pi(i)$ can be considered as any subset of the nodes $\{\pi(1), \dots, \pi(i-1)\}$ whose data, $\{X_{\pi(1)}, \dots, X_{\pi(i-1)}\}$, are decoded and available at the CH.

### *A. Rate Allocation Structure with Complexity Constraints*

The data of the nodes is compressed frame by frame without loss with a DSC of order $\alpha, 1 \leq \alpha \leq N$. It is also assumed that the CH has a limited-size buffer, in which it can only store the data collected from $\beta, 1 \leq \beta \leq N$ nodes. Therefore, in this setting at stage $i$, the node $\pi(i)$ encodes its data considering a set of maximum $\alpha - 1$ nodes as side information ($SI(\pi(i))$) out of a maximum of $\beta$ nodes, whose data are available at the CH buffer or memory ($M(i)$). These sets are constrained as follows:

$$\begin{cases} M(i) = \{\pi(1), \dots, \pi(i-1)\}, & 1 \leq i \leq \beta + 1 \\ M(i) \subset M(i-1) \cup \{\pi(i-1)\}: & \beta + 1 < i \leq N \\ |M(i)| = \beta, \pi(i-1) \in M(i) & \end{cases} \tag{1}$$

$$\begin{cases} SI(\pi(i)) = M(i), & 1 \leq i \leq \alpha \\ SI(\pi(i)) \subset M(i+1) - \{\pi(i)\}: & \alpha < i \leq N. \\ |SI(\pi(i))| = \alpha - 1 & \end{cases} \tag{2}$$

Consider the triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$, where $\boldsymbol{M} = \{M(1), \ldots, M(N)\}$ and $\boldsymbol{SI} = \{SI(\pi(1)), \ldots, SI(\pi(N))\}$, the rate of the node $\pi(i)$, using an (ideal) asymmetric SW code for compression, is then given by [4]:

$$R_{\pi(i)} = H\left(X_{\pi(i)} \middle| X\left(SI(\pi(i))\right)\right), \qquad 1 \le i \le N, \tag{3}$$

where $H(.)$ denotes the discrete entropy function. We refer to this Rate Allocation structure with Constrained DSC order of $\alpha$, $1 \le \alpha \le N$ and limited buffer size of $\beta, 1 \le \beta \le N$ as $(\alpha, \beta)$-RAC.

**Remark 1.** If $\alpha > \beta$, the maximum order of DSC, which can be employed is $\beta$, as the buffer size is limited. Therefore, $(\alpha, \beta)$-RAC in the case of $\alpha > \beta$ coincides with $(\beta, \beta)$-RAC. As a result, we assume $\alpha \le \beta$ in the sequel.

**Property 1**. The rate of the nodes allocated by an arbitrary triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ satisfying $(\alpha, \beta)$-RAC is within the SW region for $N$ nodes.

*Proof*: If $\alpha = \beta = N$, the set of the nodes rates correspond to a corner point of the SW region for $N$ nodes [4], i.e., the rate of the node $\pi(i)$ is $H(X_{\pi(i)} | X_{\pi(1)}, \ldots, X_{\pi(i-1)})$. If $1 \le \alpha \le \beta < N$, we have $SI(\pi(i)) \subseteq \{\pi(1), \ldots, \pi(i-1)\}$ and as conditioning reduces the entropy [4], we have:

$$H\left(X_{\pi(i)} \middle| X\left(SI(\pi(i))\right)\right) \ge H(X_{\pi(i)} | X_{\pi(1)}, \ldots, X_{\pi(i-1)}), \ \forall\, 1 \le i \le N. \tag{4}$$

Therefore, it is straight forward to see that $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ results rates within the SW region for $N$ nodes.

**Property 2**. The size of the set $\mathcal{T}$ comprised of the triples $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ that satisfy $(\alpha, \beta)$-RAC for $1 \le \alpha \le \beta \le N$ is:

$$|\mathcal{T}| = N!\, \beta^{N-\beta} \binom{\beta-1}{\alpha-1}^{N-\beta} \prod_{i=\alpha+1}^{\beta} \binom{i-1}{\alpha-1}. \tag{5}$$

*Proof*: The total number of permutations of the nodes indexes is $N!$. For each permutation $\boldsymbol{\pi}$, considering the node $\pi(i)$, $1 \le i \le \alpha$, and noting (1) and (2), $M(i)$ and $SI(\pi(i))$ are fixed. For the node $\pi(i)$, $\alpha < i \le \beta$, $M(i)$ is fixed and $SI(\pi(i))$ have $\binom{i-1}{\alpha-1}$ choices. For $\beta < i \leq N$, the node $\pi(i)$ replaces one node in $M(i)$ ($\beta$ selections) to form $M(i+1)$ and for each selection $SI(\pi(i))$ have $\binom{\beta-1}{\alpha-1}$ choices according to (1). Hence for each $\boldsymbol{\pi}$, there are $\beta^{N-\beta}\binom{\beta-1}{\alpha-1}^{N-\beta}\prod_{i=\alpha+1}^{\beta}\binom{i-1}{\alpha-1}$ possible $(\boldsymbol{M},\boldsymbol{SI})$'s. Therefore, the total number of possible triples $(\boldsymbol{\pi},\boldsymbol{M},\boldsymbol{SI})$ satisfying $(\alpha,\beta)$-RAC is given by (5).

In the next subsection, we describe a cost function to quantify the cost for communicating the data of a node to the CH. Later we use this cost function to set up the resource optimization problems for data gathering in WSNs.

### *B. Cost Function*

The cost function of each node $i$, $F_i(.)$ is a function of rate of the node and related fixed parameters of that node. Although, $F_i$ may be an arbitrary such function in the subsequent analysis, however, we consider the following scenarios as concrete application cases of interest for numerical results in Section V.

*Case 1- Rate:* A desired objective in data gathering sensor networks is to minimize the sum of the nodes rates. In other words, for node $i$ the cost function is defined as follows,

$$F_i(R_i) = R_i. \tag{6}$$

*Case 2- Energy:* An alternative objective of interest in data gathering sensor networks is to minimize the total energy consumption. Consider a single-hop WSN cluster in which the channels between the nodes and the CH are independent and subject to Additive White Gaussian Noise (AWGN). The capacity of the channel from node $i$ to the CH with bandwidth $W_i$ Hz is given by [4],

$$C_i(P_i) = \frac{1}{2}\log\left(1+\frac{g_iP_i}{\sigma_i^2W_i}\right)\ \frac{\text{bits}}{\text{sample}},\ 1 \leq i \leq N \tag{7}$$

in which, $P_i$ is the transmission power in Watts, constant $g_i$ denotes the channel gain, and $\sigma_i^2$ is the noise variance. The node rate $R_i$ is to satisfy $R_i \leq C_i(P_i)$. Therefore, for a given $R_i$, the minimum required power is computed as,

$$P_i = (2^{2R_i}-1)\frac{\sigma_i^2W_i}{g_i}. \tag{8}$$

Considering that the SW code frame length is $T_i$ seconds, in which the channel gain and noise variance are constant, the cost function (energy consumption) of the node $i$ is given by,

$$F_i(R_i) = P_i T_i, \ \frac{\text{jouls}}{\text{frame}}, \tag{9}$$

where $P_i$ is given in (8).

## III. Total Cost Minimization in Data Gathering WSNs

In this section, we consider the problem of total cost minimization in DSC-based data gathering WSNs with the described complexity constraints. Specifically, the problem is to find the optimal triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$, such that the corresponding rate allocation with $(\alpha, \beta)$-RAC structure leads to minimization of the total (sum) cost of the network.

**Problem 1.** Consider $\mathcal{T}$ as the set of all triples of $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ corresponding to $(\alpha, \beta)$-RAC. The optimal triple $(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o})$ minimizing the total cost of the network is the solution of the following optimization problem:

$$(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o}) = \underset{(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI}) \in \mathcal{T}}{\operatorname{argmin}} \sum_{i=1}^{N} F_{\pi(i)}\big(R_{\pi(i)}\big), \tag{10}$$

where $F_{\pi(i)}(.)$ is the cost function of node $\pi(i)$ and $R_{\pi(i)}$ is its rate corresponding to the triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$.

In order to solve this optimization problem efficiently, we propose a dynamic programming solution based on a trellis structure. Different paths in the trellis simply correspond to different triples $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ within the $(\alpha, \beta)$-RAC structure. A metric is assigned to each branch of the trellis and subsequently a cost is associated with each trellis path. The objective is to find the path with minimum cost, which identifies a specific optimal triple $(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o})$. This consequently provides the rate allocated to each node as in (3).

### A. *Trellis Structure*

The trellis components are described below. Fig. 1 partially depicts an instance of such a trellis.

*Stages*: The total number of stages in the trellis is equal to $N+1$, corresponding to the total number of nodes, $N$, plus a terminating last stage. The stage $i$ is associated with the node, whose data is collected in the $i$'th order, $1 \leq i \leq N$.

*States*: The state $s$ in stage $i$ is denoted by the pair $(i,s)$ and indicates the nodes whose data have been collected so far. We split the set of these nodes to two, namely $B(i,s)$ or the ones whose data are stored in the buffer, and $A(i,s)$ as the rest of the nodes whose data are not stored in the buffer. We have:

- For $1 \leq i \leq \beta+1$, $A(i,s)=\{\}$. There is one state corresponding to each subset of nodes with $i-1$ elements that is determined by $B(i,s)$. Therefore, in stage $i$, there are a total of $\binom{N}{i-1}$ states.
- For $\beta+1 < i \leq N+1$, $A(i,s)$ contains $i-\beta-1$ nodes, and $B(i,s)$ contains $\beta$ nodes. The total number of states in stage $i$ is then given by $\binom{N}{\beta}\binom{N-\beta}{i-\beta-1}$.

*Branches*: The branch $b$ emanating from the state $(i,s)$ is identified by $(i,s,b)$ and corresponds to the node $n(i,s,b)$ that is to transmit in this stage and the set of its SI nodes denoted by $SI(n(i,s,b))$. Of course, the node $n(i,s,b)$ is not included in the set of nodes that identify the state $(i,s)$ whose data have been collected so far.

The SI of the node, $SI\big(n(i,s,b)\big)$, is determined as follows.

- For $1 \leq i \leq \alpha$, $SI\big(n(i,s,b)\big)=B(i,s)$. Corresponding to each specific node, there is only one branch emanating from the state $(i,s)$.
- For $\alpha < i \leq \beta$, each subset of nodes in $B(i,s)$ with $\alpha-1$ elements may be chosen as $SI\big(n(i,s,b)\big)$. Therefore, corresponding to each specific node, the number of branches emanating from the state $(i,s), \alpha < i \leq \beta$, is equal to $\binom{i-1}{\alpha-1}$.
- For $\beta < i \leq N$, the node $n(i,s,b)$ is to replace one of the nodes currently in $B(i,s)$. Each subset of the remaining $\beta-1$ nodes in $B(i,s)$ with $\alpha-1$ elements may serve as $SI\big(n(i,s,b)\big)$. Hence, corresponding to each specific node, the number of branches emanating from the state $(i,s), \beta < i \leq N$, is equal to $\beta\binom{\beta-1}{\alpha-1}$.

Consider the branch $(i,s,b)$ that enters the state $(i+1,s')$. For $1 \leq i \leq \beta$, the node $n(i,s,b)$ is added to the set $B(i,s)$ to form $B(i+1,s')$. Similar to $A(i,s)$, the set $A(i+1,s')$ is empty. For $\beta < i \leq N$, the node $n(i,s,b)$ replaces one element of the set $B(i,s)$ to form $B(i+1,s')$. The set $A(i+1,s')$ is then the union of the replaced element of $B(i,s)$ and the set $A(i,s)$.

Based on the above analysis, it may be deducted that the number of branches entering a state in stage $i$, is given by $i-1$, $(i-1)\binom{i-2}{\alpha-1}$ and $\beta(i-\beta-1)\binom{\beta-1}{\alpha-1}$ for $1 \le i \le \alpha+1$, $\alpha+1 < i \le \beta+1$ and $\beta+1 < i \le N+1$, respectively.

## *B. Trellis Search Algorithm*

Based on the described trellis structure, the search algorithm is set up by defining the branch metric and the state cost as follows.

*Metric*: A specific SI and therefore a specific rate is assigned to the node $n(i,s,b)$ corresponding to the branch $(i,s,b)$. According to the SW theorem and $(\alpha,\beta)$-RAC structure, this rate is denoted by $R_{n(i,s,b)}$ and is given by,

$$R_{n(i,s,b)} = H\left(X_{n(i,s,b)} \middle| X\left(SI\big(n(i,s,b)\big)\right)\right), \tag{11}$$

The metric of the branch $(i,s,b)$ is defined as the cost of the node $n(i,s,b)$, i.e., $F_{n(i,s,b)}\big(R_{n(i,s,b)}\big)$.

**Remark 2.** In certain cases, there may be an additional constraint that depends on the node rate. This situation arises for example in case 2 of Section II-B, where a peak power constraint for the nodes may also be assumed. The constraint may be incorporated in the trellis search algorithm, by modifying the metric or alternatively removing the branch, whose corresponding rate is in violation of the constraint.

*State Cost*: The cost of the state $(i,s)$, denoted by $SC(i,s)$, is defined as the minimum cost up to this state that is provided by the paths reaching this state. Considering the set of the branches entering the state $(i+1,s)$ as $L(i+1,s)$, we have

$$SC(i+1,s) = \min_{(i,s',b)\in L(i+1,s)} \left(SC(i,s') + F_{n(i,s',b)}\left(R_{n(i,s',b)}\right)\right). \tag{12}$$

The branch that results in the minimum cost at a given state, or alternatively its parent state is stored. Once the cost is computed for all the states up to the last stage, the optimal path resulting the smallest cost is traced back. This path identifies the optimal triple $(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o})$. Given that the optimal path in stage $i+1$, coincides with state $(i+1, s)$ and the branch resulting in the minimum cost at this state is $(i, s', b)$; we have $\pi^o(i) = n(i, s', b)$, $M^o(i) = B(i, s')$, $SI^o\big(\pi^o(i)\big) = SI\big(n(i, s', b)\big)$, $1 \leq i \leq N$.

**Proposition 1.** The proposed trellis structure and search algorithm provides the optimal solution for the optimization problem 1.

*Proof*: The principle of optimality for a dynamic programming solution is satisfied if the decisions in each stage of the trellis, given the state, are independent of the previous stages [21]. Considering the definition of the trellis components, given the current state and its cost, the chosen branches reaching the states in subsequent stages do not depend on the previously reached states or previously chosen branches. More precisely, the states constitute a first order Markov chain and the stated principle of optimality is satisfied. ■

We refer to the presented solution for problem 1 as the Total Cost Optimal Trellis-based (TCOT) solution in the sequel. The proposed TCOT solution finds the optimal triple $(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o})$ in each transmission round for a given network configuration. As soon as the network configuration changes the corresponding optimal triple needs to be recomputed. Therefore, as the objective is to minimize the total network cost [in this part of the manuscript], this still leads to an overall optimal solution for the problem.

*Example 1.* For data gathering in a WSN with 4 nodes and $\alpha = 2$, $\beta = 3$, Fig. 1 depicts the corresponding trellis diagram. To avoid confusion, the diagram is only partially presented, i.e., some states in certain stages are not shown

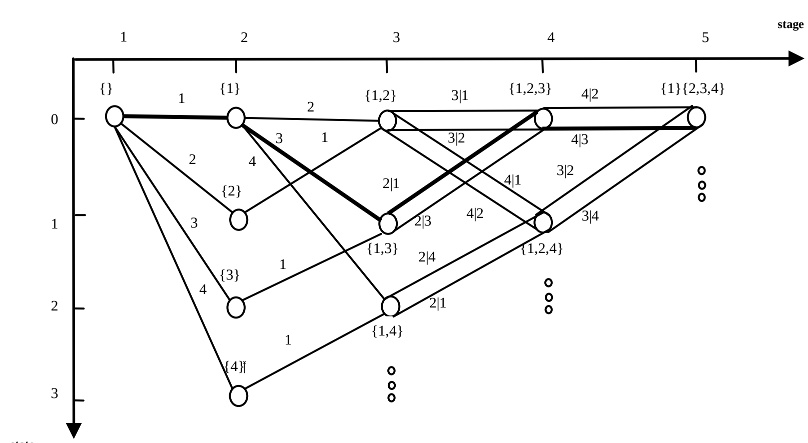


Fig. 1. An example of a trellis structure for a network with 4 nodes, $\alpha = 2, \beta = 3$.

and therefore some of the branches are removed. The label of each branch illustrates the node index and its SI nodes. Each state $(i,s)$ is marked with its corresponding sets $A(i,s)$ and $B(i,s)$, where in cases that $A(i,s)=\{\}$ the label shows only $B(i,s)$. For instance, the state 0 in stage 5 (state (5,0)) is marked by {1}{2,3,4} that indicates the sets $A(5,0)=\{1\}$ and $B(5,0)=\{2,3,4\}$ for this state. Therefore, in this state the data of the nodes {1,2,3,4} have been collected and processed so far, but only those of the nodes {2,3,4} are stored in the buffer to be potentially used as SI for subsequent nodes. Consider the state (4,1), where $B(4,1)=\{1,2,4\}$. Emanating from this state, the branch marked by 3|2 shows that the node 3 compresses and sends its data with rate $H(X_3|X_2)$ in this stage. As the buffer is already full ($\beta=3$), therefore the data of the node 3 that is decoded at the CH in this stage can replace that of either node 1 or node 4 from the nodes currently in the buffer. Note that $X_2$ may not be replaced since it is being used as SI for decoding $X_3$. If $X_1$ is replaced, the branch enters the state identified by $B=\{2,3,4\}$ and $A=\{1\}$, i.e., state (5,0). If $X_4$ is replaced, the branch enters the state identified by $B=\{1,2,3\}$ and $A=\{4\}$ (not shown in the figure).Suppose that the highlighted path in the trellis is identified as the optimal path by the proposed trellis search algorithm. This path corresponds to the triple $(\boldsymbol{\pi^o},\boldsymbol{M^o},\boldsymbol{SI^o})$ where $\boldsymbol{\pi^o}=[1,3,2,4]$, $\boldsymbol{M^o}=\{\{\,\},\{1\},\{1,3\},\{1,2,3\}\}$ and $\boldsymbol{SI^o}=\{\{\,\},\{1\},\{1\},\{3\}\}$. Therefore, the resulting optimal rate allocation is as follows: $R_1=H(X_1), R_3=H(X_3|X_1), R_2=H(X_2|X_1), R_4=H(X_4|X_3)$.

### C. *Solution Complexity*

We compare the complexity of the proposed TCOT algorithm with that of running full search algorithm in terms of the number of operations and memory requirements. These quantities are measured in number of floating point operations or floating point values to be stored, simply referred to as floats [22]. The computational complexity for full search strategy can be obtained by considering the number of total triples of $(\boldsymbol{\pi},\boldsymbol{M},\boldsymbol{SI})$ corresponding to $(\alpha,\beta)$-RAC provided in (5), the number of operations required to compute the total cost resulted by each triple and those to find the optimal triple. Hence, the number of operations (computational complexity) to find the solution of optimization problem 1 based on a full search strategy is:

$$CC_1=(Q+1)NN!\,\beta^{N-\beta}\binom{\beta-1}{\alpha-1}^{N-\beta}\prod_{i=\alpha+1}^{\beta}\binom{i-1}{\alpha-1} \tag{13}$$

where in (13), $Q$ is a constant that denotes the number of operations required to compute the cost function, e.g., based

H. Arjmandi, F. Lahouti, "Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks," *To Appear in IEEE Sensors Journal*, 2011.

on equations (6) or (9), per node.

The computational complexity of the proposed TCOT algorithm is as follows and is due to the operations required to compute the branch metrics and those to find the optimal input branch to each state.

$$CC_2 = (Q+2)[\,\sum_{i=1}^{\alpha+1}\binom{N}{i-1}(i-1) + \sum_{i=\alpha+2}^{\beta+1}\binom{N}{i-1}(i-1)\binom{i-2}{\alpha-1} + \binom{N}{\beta}\beta\binom{\beta-1}{\alpha-1}\sum_{i=\beta+2}^{N+1}\binom{N-\beta}{i-\beta-1}(i-\beta-1)\,]. \quad (14)$$

This is obtained by quantifying the number of states in each stage and the number of input branches to each state as presented in Section III-A. The three terms in (14) correspond to the number of operations required in stages $1 \le i \le \alpha+1$, $\alpha+1 < i \le \beta+1$, and $\beta+1 < i \le N+1$, respectively.

The memory required to run a full search is to store the triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ being examined and is equal to $MR_1 = \frac{1}{2}(N+(N-\alpha)\alpha+(N-\beta)\beta)$ floats. In order to run the proposed TCOT algorithm, it is required to store the optimal input branch (the indices of the node and its SI nodes) to each state and the set $B$ corresponding to all states of the trellis. The memory requirement is then obtained as follows. Note that the memory required to store an integer is considered as one half of a float.

$$MR_2 = \frac{1}{2}\sum_{i=1}^{\alpha+1}\binom{N}{i-1}(i) + \frac{1}{2}\sum_{i=\alpha+2}^{\beta+1}(i+\alpha-1)\binom{N}{i-1} + \frac{(\beta+\alpha)}{2}\binom{N}{\beta}\times\sum_{i=\beta+2}^{N+1}\binom{N-\beta}{i-\beta-1}. \quad (15)$$

Fig. 2 depicts the values of $CC_1, CC_2, MR_1, MR_2$ as a function of the number of network nodes $N$ for $\beta = 5$ and $\alpha = 3$. It is clear that the proposed TCOT solution substantially reduces the computational complexity compared to the full search at the cost of a much smaller increase in the memory requirement. The presented optimal solution incurs a manageable complexity for data gathering within a cluster of nodes or in WSNs with a small number of nodes. However, we still present an alternative and more efficient solution.

### *D. An Efficient Suboptimal Solution*

In this subsection, a suboptimum, yet high performance algorithm to solve the optimization problem 1 is proposed that is of polynomial complexity. The solution is also based on a trellis structure, however with a different definition for the states of stages $\beta+1 < i \le N+1$. We refer to this solution for problem 1 as the Total Cost Suboptimal

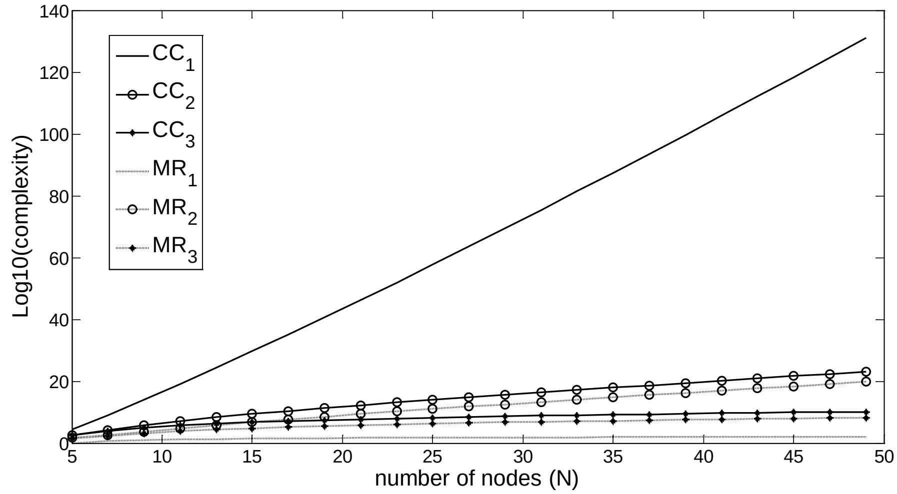


Fig. 2. Comparison of complexity for running full search, and the proposed TCOT and TCST solutions; $\beta = 5$, $\alpha = 3$ and $Q = 1$. Note that the scale is logarithmic.

Trellis-based (TCST) solution. In this scheme, a state is uniquely identified by the set $B(i,s)$ (described in Section III-A). Hence, the states in the original trellis with a given $B(i,s)$ and possibly different $A(i,s)$ are replaced with a single merged state in the new trellis. In fact, the set $A(i,s)$ for each state is now only identified by the path with minimum cost reaching that state. Hence, the principle of dynamic programming optimality as described in proposition 1 is now unsatisfied, and therefore the solution is suboptimal. The number of states in stage $i$, $\beta + 1 < i \leq N + 1$ is reduced to $\binom{N}{\beta}$ in the suboptimal trellis. As we shall demonstrate, this only negligibly affects the performance, but interestingly results in a polynomial complexity.

For the presented TCST solution, the number of trellis branches entering each state at stage $i$, is $i - 1$, $(i-1)\binom{i-2}{\alpha-1}$ and is smaller than $\binom{\beta-1}{\alpha-1}(\beta)$ for $1 \leq i \leq \alpha + 1$, $\alpha + 1 < i \leq \beta + 1$ and $\beta + 1 < i \leq N + 1$, respectively. Therefore, the worst case computational complexity of the TCST solution is determined as follows:

$$CC_3 = (Q+2)\left[\sum_{i=1}^{\alpha+1}\binom{N}{i-1}(i-1) + \sum_{i=\alpha+2}^{\beta+1}\binom{N}{i-1}(i-1)\binom{i-2}{\alpha-1} + (N-\beta)\binom{N}{\beta}\binom{\beta-1}{\alpha-1}\beta\right] \quad (16)$$

The memory required to run the TCST solution is then given by:

$$MR_3 = \frac{1}{2}\sum_{i=1}^{\alpha+1}(i-1)\binom{N}{i-1} + \frac{1}{2}\sum_{i=\alpha+2}^{\beta+1}(i+\alpha-1)\binom{N}{i-1} + \frac{1}{2}(N-\beta)\binom{N}{\beta}(\alpha+\beta) \quad (17)$$

For large $N$ and a limited $\beta$ ($\beta < \frac{N}{2}$), analyzing $CC_3$ or $MR_3$ in (16) and (17) indicate a polynomial complexity in terms of $N$ as $O(N^{\beta+1})$. Fig. 2 also depicts $CC_3$ and $MR_3$ as a function of the number of network nodes for comparison.

**Remark 3.** The suboptimal solution asymptotically tends to the optimal solution with increasing $\alpha$ and $\beta$. This is due to the fact that, for larger $\alpha$ and $\beta$ merging of states in the suboptimum algorithm occurs in a smaller number of stages and therefore, less states and branches are removed from the original trellis. Specifically, the suboptimal solution coincides with the optimal one when there is no limitation on memory and computational complexity ($\alpha = \beta = N$).

## IV. Lifetime Maximization in Data Gathering WSNs

In this section, we consider the problem of lifetime maximization for DSC-based data gathering WSNs given the system model and complexity constraints described in Section II. Specifically, the problem is to find the optimal triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$, such that the corresponding rate and power allocation with $(\alpha, \beta)$-RAC structure leads to the maximization of the minimum remaining energy in the set of network nodes subsequent to one data gathering round. A data gathering round involves transmission of a data frame by each and every network node.

**Problem 2**. Consider $\mathcal{T}$ as the set of all triples of $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$ corresponding to $(\alpha, \beta)$-RAC. Suppose that the initial energy of node $i$, $1 \le i \le N$ is $E_i$. The optimal triple $(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o})$ maximizing the network lifetime is the solution of the following optimization problem:

$$(\boldsymbol{\pi^o}, \boldsymbol{M^o}, \boldsymbol{SI^o}) = \arg \max_{(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI}) \in \mathcal{T}} \min_{1 \le i \le N} \left( E_{\pi(i)} - F_{\pi(i)}\left(R_{\pi(i)}\right)\right), \tag{18}$$

where $F_{\pi(i)}(.)$ is the energy consumption function of node $\pi(i)$ (case 2 Section II-B) and $R_{\pi(i)}$ is its rate corresponding to the triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI})$.

The dynamic programming solution proposed for problem 1 can also be employed to solve problem 2. The same trellis structure and search algorithm is used, while the branch metric and state cost (here utility[1]

[1] Since problems 2 is a maximization problem, we use the term state utility as opposed to state cost in this case.

follows.

*Metric*: Consider the branch $(i, s, b)$. A specific SI and therefore a specific rate is assigned to the node $n(i, s, b)$ corresponding to this branch. The metric of this branch is defined as the remaining energy of the related node, i.e., $E_{n(i,s,b)} - F_{n(i,s,b)}(R_{n(i,s,b)})$.

*State utility*: The utility of the state $(i, s)$, denoted by $SU(i, s)$, is defined as the maximum utility provided by the paths reaching this state. We have

$$SU(i+1, s) = \max_{(i,s',b)\in L(i+1,s)} \min\left\{E_{n(i,s',b)} - F_{n(i,s',b)}\left(R_{n(i,s',b)}\right), SU(i, s')\right\}. \tag{19}$$

The minimization in (19) quantifies the utility of a path reaching state $(i, s)$, i.e., the minimum remaining energy among the nodes corresponding to this path. The maximization over $L(i + 1, s)$, or the set of branches (paths) entering the state $(i + 1, s)$, identifies the path with maximum utility and hence determines the state utility. At each state, the branch that results in the maximum utility, or alternatively its parent state is stored.

Following proposition 1, it is straight forward to show that the presented solution for problem 2 is optimal. Naturally, in this case a suboptimal solution with polynomial complexity may also be developed. We refer to these proposed solutions for problem 2 respectively, as the Lifetime Maximization Optimal and Suboptimal Trellis-based (LMOT and LMST) algorithms.

## V. Performance Evaluation

We consider a cluster of $N$ wireless sensor nodes uniformly distributed in a unit square and a CH located at the center of the square. The data of the nodes are generated based on an $N$-dimensional jointly Gaussian distribution with covariance matrix $\boldsymbol{K}$ and mean vector $\boldsymbol{\mu}$. The covariance between the nodes $i$ and $j$ with a distance $d_{ij}$ is $k_{ij} = \sigma_0^2 e^{-\gamma d_{ij}^2}$, where $\gamma \geq 0$ is a parameter controlling the level of spatial dependency. A larger $\gamma$ implies a smaller correlation. The variance of the sensed data at each node is $k_{ii} = \sigma_0^2, 1 \leq i \leq N$ [13]. The energy consumption function of the nodes is computed as described in Section II.B case 2 (equation (9)). For the experiments, we
) is defined as
consider $g_i = 0.02/{d_i}^2$, where $d_i$ is the distance of node $i$ to the CH, $W_i = 1$, $\sigma_i^2 = 10^{-4}$, $T_i = 1$ and $E_i = 50$,

H. Arjmandi, F. Lahouti, "Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks," *To Appear in IEEE Sensors Journal*, 2011.

$1 \leq i \leq N$. The results are averaged over 50 random networks with $N = 10$ nodes.

*A. Benchmark Algorithms for Problem 1*

The simplest algorithm to address problem 1 is a random rate allocation. In this algorithm, given $\alpha$ and $\beta$, a triple $(\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{SI}) \in \mathcal{T}$ is selected randomly and the rate of the nodes are allocated accordingly. We refer to this scheme as B1 for the subsequent comparisons. Another benchmark algorithm can be set up based on the solution proposed in [13] for minimizing the total cost of WSN nodes using unconstrained DSC. In this case, the transmission order of the nodes is set based on their distance to the CH, i.e., the nearest node sends first. In this algorithm, which is referred to as B2 in the sequel, the SI is set as the data of $\alpha - 1$ nodes immediately prior to the current node in the transmission order.

*B. Benchmark Algorithms for Problem 2*

The proposed algorithm for maximizing the network lifetime takes the remaining energy of the nodes as initial values. Therefore, a simple algorithm which arranges the order of transmission of the nodes according to their level of remaining energy may be considered as a benchmark algorithm. An extended form of this metric is suggested in [20] in the context of routing for maximizing the lifetime of WSNs. Specifically, for the node $i$ a cost is defined as $c_i = {d_i}^2/E_i$ and the nodes transmit in order of their cost (the node with smaller cost sends first). For solving problem 2, we consider these algorithms in the followings as benchmark schemes B3 and B4, respectively. In both cases, the SI is set as the data of $\alpha - 1$ nodes immediately prior to the current node in the transmission order.

*C. Numerical Results*

Fig. 3 depicts the average total rate of the WSN nodes resulted by the proposed TCOT and TCST algorithms and the benchmark scheme B1 for the problem 1 as a function of parameter $\beta$ and for different values of $\alpha$. It is observed that proposed TCOT and TCST algorithms significantly reduce the total rate of the WSN in comparison to the benchmark algorithm B1. It is evident that for the proposed algorithms and a given DSC order, $\alpha$, increasing the CH storage size, $\beta$ beyond $\alpha$ leads to only 5% performance improvement. Hence, the results verify that for a given $\alpha$ a limited value of $\beta = \alpha$ can capture most of the available dependency and achievable compression gain. Fig. 3 also

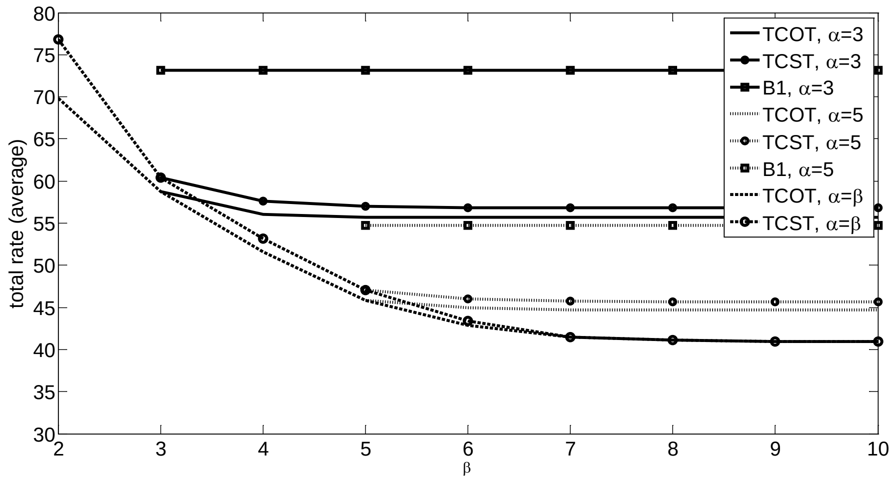


Fig. 3. Average total rate vs. $\beta$ for different values of $\alpha$. Results of the proposed TCOT and TCST solutions and the benchmark algorithm B1 averaged over 50 random networks with $N = 10$ nodes.

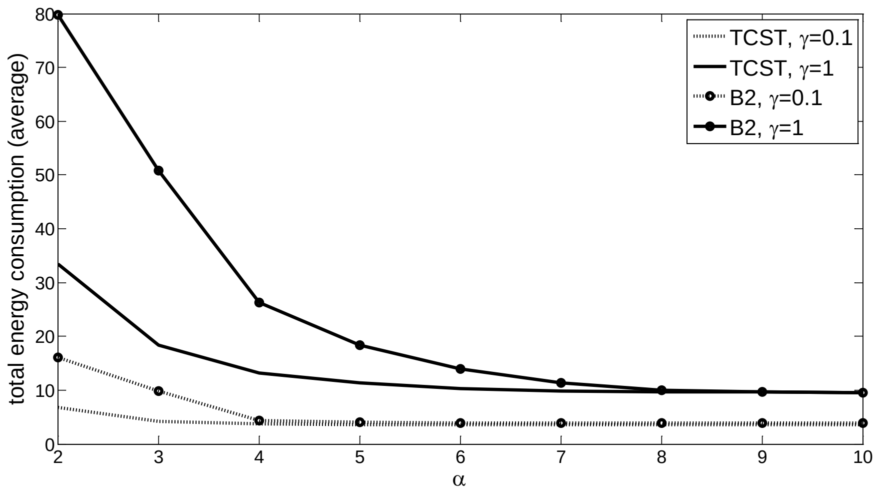


Fig. 4. Average total energy consumption vs. $\alpha$, for different value of γ ($\beta = \alpha$). Results of the proposed TCST solution and the benchmark algorithm B2 averaged over 50 random networks with $N = 10$ nodes.

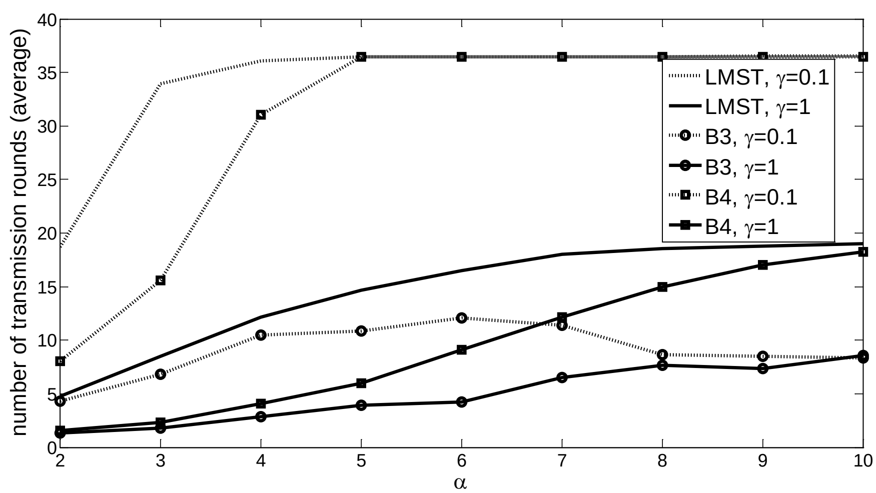


Fig. 5. Average number of transmission rounds vs. $\alpha$, for different value of γ ($\beta = \alpha$). Results of the proposed LMST solution and the benchmark algorithms B3 and B4 for 50 random networks with $N = 10$ nodes.

demonstrates that the performance of the proposed TCST algorithm with polynomial complexity still closely follows that of the proposed optimum solution (TCOT). The gap in their performance is in fact smaller than 2% and decreases and tends to zero with increasing $\beta$ and $\alpha$. Therefore, for the rest of this section, we simply focus on the performance of the proposed suboptimum schemes, TCST and LMST.

Fig. 4 depicts the average total energy consumption of the WSN nodes resulted by the proposed TCST algorithm and the benchmark scheme B2 for the problem 1 as a function of parameter $\alpha$ and for different values of $\gamma$ ($\beta = \alpha$). As expected, if the nodes data are more strongly correlated (smaller $\gamma$), the resulting energy consumption is smaller, as a higher level of compression is achieved. It is also observed that the energy consumption provided by the proposed algorithm in comparison with the algorithm B2 is noticeably smaller. Specially, the gain is more substantial, when the nodes data are less dependent (smaller $\gamma$). In fact, the effectiveness of the resource allocation algorithm plays a much more important role in such cases.

Fig. 5 depicts the average number of transmission rounds until the network lifetime ends, resulted by the proposed LMST algorithm and the benchmark schemes B3 and B4 as a function of $\alpha$, for different values of $\gamma$ ($\beta = \alpha$). The remaining energy of a node following a transmission serves as its initial energy for its subsequent transmission. It is observed that the proposed algorithm substantially improves the number of transmission rounds in comparison with the benchmark schemes. Specifically, in comparison with B4, this improvement is more considerable for small values of $\alpha$. As expected, if the nodes data are more strongly correlated (smaller $\gamma$), the resulting number of possible network transmission rounds are greater.

The results in Figs. 4 and 5 also verify that for a given $\gamma$ a limited value of $\alpha$ greater than a threshold, if effectively utilized, still allows capturing most of the available dependency and achievable compression gain. The values of the threshold for the proposed TCST and LMST solutions depend on the dependency between the nodes data. For example, for the TCST scheme in Fig. 4 we can consider the thresholds as $\alpha = 4$ for $\gamma = 0.1$ and $\alpha = 7$ for $\gamma = 1$.

## VI. Conclusions

In this paper, a DSC-based rate allocation structure in data gathering wireless sensor networks was proposed which takes into account the CH (DGN) memory limitation and complexity constraint. Based on this structure, we considered the problems of resource (rate) allocation for the nodes to minimize the network total cost or maximize the network lifetime, where the cost of a node is a general function of its rate and its related parameters. To solve these problems optimal dynamic programming solutions and alternative suboptimal solutions with polynomial complexity order in terms of number of nodes were presented. Numerical results indicate that even when the CH resources (buffer size or DSC complexity) are limited, if effectively utilized based on the proposed solutions, it still allows for capturing most of the available dependency and compression gain.

## ACKNOWLEDGMENT

This research has been supported in part by the Iran Telecommunications Research Center and Iran National Science Foundation.

## REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirici, “A survey on sensor networks,” *IEEE Commun. Mag.*, Vol. 40, No. 8, pp. 102-114, Aug. 2002.

[2] Z. Xiong, A.D. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Process. Mag.*, Vol. 21, No. 5, pp. 80-94, Sep. 2004.

[3] D. Slepian and J.K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theor*y, Vol. 19, pp. 471-480, July 1973.

[4] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Ed, Wiley, New-York, 2006.

[5] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” *IEEE Trans. Inform. Theory*, Vol. 49, pp. 626–643, Mar. 2003.

[6] J. Garcia-Frias, F. Cabarcas, “Approaching the Slepian–Wolf boundary using practical channel codes,” *Signal Process.* Vol. 86, No. 11, pp. 3096–3101, Nov. 2006.

[7] A. Aaron, B. Girod, “Compression with side information using turbo codes,” in *Proc. of the IEEE Data Compression Conf. (DCC)*, Snowbird, UT, USA, Apr. 2002, pp. 252-261.

[8] A. Liveris, C. Lan, K. Narayanan, Z. Xiong, and C. Georghiades, “Slepian-Wolf coding of three binary sources using LDPC codes,” in *Proc. of Int. Symp. on Turbo Codes and Related Topics*, Brest, France, Sep. 2003, pp. 63-66.

[9] K. Lajnef, C. Guillemot, P. Siohan, “Distributed coding of three binary and Gaussian correlated sources using punctured turbo codes,” *Signal Process.* Vol. 86, No. 11, pp. 3131–3149, Nov. 2006.

[10] C. Lan, A.D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, “Slepian-Wolf coding of multiple M-ary sources using LDPC codes,” *Proc. of the IEEE Data Compression Conf. (DCC)*, Mar. 2004, p. 549.

[11] M. Zamani and F. Lahouti, "Distributed source coding using symbol-based and non-binary turbo codes - applications to wireless sensor networks," *IET Commun.* Vol. 2, No. 8, pp. 1089-1097, Sep. 2008.

[12] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks – A survey," *IEEE Commun. Surveys & Tutorials*, Vol. 8, No. 4, pp. 48-63, 2006.

[13] R. Cristescu, B. Beferull-Lozano, and M.Vetterli, "Networked Slepian-Wolf: theory, algorithms, and scaling laws," *IEEE Trans. Inform. Theory*, Vol. 51, No. 12, pp. 4057-4073, Dec. 2005.

[14] S. Agnihotri, P. Nuggehalli, and H. S. Jamadagni, "On maximizing lifetime of a sensor cluster," *IEEE Int. Symp. World of Wireless, Mobile and Multimedia Networks,* Taormina, Italy, Jun. 2005, pp. 312-317.

[15] A. Roumy and D. Gesbert, "Optimal matching in wireless sensor networks," *IEEE Journal on Selected Topics in Signal Process*., Vol. 1, No. 4, pp. 725-735, Dec. 2007.

[16] S. Li and A. Ramamoorthy, "Rate and power allocation under the pairwise distributed source coding constraint," *IEEE Trans. Commun.*, Vol. 57, No. 12, pp. 3771-3781, Dec. 2009.

[17] W. Wang, D. Peng, H. Wang, H. Sharif and H. Chen, "Cross-layer multirate interaction with distributed source coding in wireless sensor networks," *IEEE Trans. Wireless Commun.* vol. 8, no. 2, pp. 787-795, Feb. 2009.

[18] Z. Tanga, I.A. Gloverb, A.N. Evansa, J. Hec, "An energy-efficient adaptive DSC scheme for wireless sensor networks," *Signal Process.* Vol. 87, No. 12, pp. 2896–2910, Dec. 2007.

[19] Y.W.P. Hong, Y.R. Tsai, Y.Y. Liao, C.H. Lin and K.J. Yang, "On the throughput, delay, and energy efficiency of distributed source coding in random access sensor networks," *IEEE Trans. Wireless Commun.*, Vol. 9, No. 6, pp. 1965-1975, Jun. 2010.

[20] H. O. Tan and I. Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," *ACM SIGMOD Record*, Vol. 32, No. 4, pp. 66–71, Dec. 2003.

[21] E. Denardo, *Dynamic Programming: Models and Applications*, Prentice Hall, Englewood Cliffs, 1982.

[22] F. Lahouti and A. K. Khandani, "Quantization of LSF parameters using a trellis modeling," *IEEE Trans. Speech and Audio Process,* Vol. 11, No. 5, pp. 400-412, Sep. 2003.