A Note on Slepian-Wolf Bounds for Several Node Grouping Configurations

Benjamin Rosen, Adnan M. Abu-Mahfouz, Senior Member, IEEE, and Ling Cheng, Senior Member, IEEE

Abstract—The Slepian-Wolf bound on the admissible coding rate forms the most fundamental aspect of distributed source coding. As such, it is necessary to provide a framework with which to model more practical scenarios with respect to the arrangement of nodes in order to make Slepian-Wolf coding more suitable for multi-node Wireless Sensor Networks. This paper provides two practical scenarios in order to achieve this aim. The first is by grouping the nodes based on correlation while the second involves simplifying the structure using Markov correlation. It is found that although the bounds of these scenarios are more restrictive than the original Slepian-Wolf bound, the overall model and bound are simplified.

Index Terms—distributed source coding, slepian wolf, achievable bounds, wireless sensor networks

I. INTRODUCTION

With the massive improvements in processing power and reduction in cost, modern computers are now more suited than ever to be used in applications requiring a lot of data, such as Wireless Sensor Networks (WSNs), in which many nodes observe an event and communicate with a central node. This setup necessitates the implementation of advanced coding techniques to maximise the data rate while minimising the interference amongst nodes. Slepian-Wolf (SW) coding [1] has proven itself as a viable means to performing distributed coding that is suitable for a system of correlated nodes, since WSNs employ many nodes that are in close proximity, meaning that the nodes are highly correlated with one another [2]. Fundamentally, the SW coding rate is bounded, which limits the maximum compression that can be achieved. Although much research has been done on the techniques to achieve this bound [3]–[9], the literature is lacking in the adjustment of this bound depending on the nature of the correlation between nodes.

This paper attempts to bridge this knowledge gap by providing the adjusted SW bound for a variety of correlation structures. Specifically, we present a simpler bound than the general SW bound by introducing the concepts of “strong” and “weak” correlation, and using an arbitrary adjacency matrix to represent the correlation structure, as well as considering nodes in Markov chain relations. We develop this concept to find the bound in a specific case, where the nodes are organised into a disjoint grouping structure. These bounds simplify the modelling for the multi-node bound of the coding rate, while sacrificing the maximum achievability. However, this also has the added effect of reducing the complexity of the coding and decoding schemes. These results are also important with regards to security, since the shared information (given by the bound) results in more information leakage, giving more power to an eavesdropper tapping into the channel [10], [11].

This paper is organised as follows: Section II gives a brief overview of SW coding, while Section III presents the system model. The total rate bound for different node grouping cases when considering different correlation groupings and Markov chain relations between nodes are discussed in Section IV and V before concluding in Section VI.

\[
R_x \geq H(X|Y) \\
R_y \geq H(Y|X) \\
R_x + R_y \geq H(X,Y)
\]

Fig. I shows that, although the admissible rate region is greater in the former case, it only differs by the maximum achievable rates. This paper has inspired an entire field dedicated to lossless Distributed Source Coding (DSC).

Cover [12] extended this contribution by first proving the Slepian-Wolf rate region in a simpler manner. He then extends their work to an arbitrary N number of sources. Of importance to this research is the fact that N sources corresponds to \(2^N - 1\) rate region equations, meaning the maximum admissible rate region can be accurately obtained for any number of sources. Cover gives the following equations for \(N = 3\), for example:
The system to be modelled is a part of a WSN, where the edge nodes are correlated with one another. Although they communicate through the WSN, there is only one receiver at the end of the network, denoted by the Sink. Fig. 2 shows the modelling for this system, where $X_1, \ldots, X_N$ are the correlated edge nodes. This paper is particularly focused on the minimum encoding rate over all the nodes, represented by the supernode $S$.

The grouping of nodes, and thus the bound on the total compression rate, are directly affected by the correlation between nodes. As mentioned in [2], the correlation represents a “virtual” channel that allows the nodes to communicate, although explicit communication is not present. Therefore, this paper focuses on the different arrangement of nodes, based on different correlation scenarios, and how it affects the total rate. We also consider the effect on the total rate when taking into account a Markov chain correlation between nodes.

### III. System Model

The system to be modelled is a part of a WSN, where the edge nodes are correlated with one another. Although they communicate through the WSN, there is only one receiver at the end of the network, denoted by the Sink. Fig. 2 shows the modelling for this system, where $X_1, \ldots, X_N$ are the correlated edge nodes. This paper is particularly focused on the minimum encoding rate over all the nodes, represented by the supernode $S$.

The grouping of nodes, and thus the bound on the total compression rate, are directly affected by the correlation between nodes. As mentioned in [2], the correlation represents a “virtual” channel that allows the nodes to communicate, although explicit communication is not present. Therefore, this paper focuses on the different arrangement of nodes, based on different correlation scenarios, and how it affects the total rate. We also consider the effect on the total rate when taking into account a Markov chain correlation between nodes.

### IV. Total Rate for Single Grouping Cases

In the Section 2, the total rate was given for the two and three node case, where all nodes are correlated with one another. We now give a simplified form of Cover’s [12] total rate when considering $N$ nodes.
Theorem 1. For a system of $N$ correlated sources $X_1, X_2, \ldots, X_N$, a valid total rate $R_T = \sum_{i=1}^{N} R_i$ (in bits/symbol) is given by

$$R_T \geq H(X_1) + \sum_{i=2}^{N} H(X_i|X_{i-1}, \ldots, X_1)$$  \hspace{1cm} (5)

Proof: We begin with the final inequality given by Cover in [12] for $N$ nodes:

$$\sum_{i=1}^{N} R_i \geq H(X_1, X_2, \ldots, X_N)$$  \hspace{1cm} (6)

Using the chain rule for entropy and substituting $R_T$ yields

$$R_T \geq H(X_1) + H(X_2|X_1) + \cdots + H(X_N|X_{N-1}, \ldots, X_1)$$  \hspace{1cm} (7)

which can be simplified using a summation.

Conceptually, the total rate given in Theorem 1 represents a single valid point on the $N-1$ hypersurface defining the upper bound. Any point on this surface will have the same total rate, with a different distribution of individual rates. The result shown here is a parallel to the “corner points” given in literature for the two node case [13].

![Fig. 3. Modelling a DSC system with Markov correlation](image)

Next we consider the lemma of this bound calculation if the correlation is modelled as a Markov chain. This implies that each node is correlated with its direct neighbours, but is independent of all other nodes (as shown in Fig. 3).

Lemma 1. For a system of $N$ nodes, where the correlation between them forms a Markov chain $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_N$, a valid bound on the total rate $R_T$ is given by

$$R_T \geq H(X_1) + \sum_{i=2}^{N} H(X_i|X_{i-1})$$  \hspace{1cm} (8)

Proof: From Theorem 1 we have

$$R_T \geq H(X_1) + H(X_2|X_1) + \cdots + H(X_N|X_{N-1}, \ldots, X_1)$$  \hspace{1cm} (9)

By the definition of Markov correlation, $H(X_3|X_2, X_1) = H(X_3|X_2)$ and $H(X_4|X_3, X_2, X_1) = H(X_4|X_3)$. Thus, every term in (9) to the right of the inequality (besides for the first) will reduce to $H(X_i|X_{i-1})$.

We can also consider a system in which there is a combination of Markov and regular correlation structures.

Theorem 2. For a system of $N$ sources, correlated by a combined regular and Markov correlation $X_1, \ldots, X_{k-1} \rightarrow X_k \rightarrow \cdots \rightarrow X_{k+r}, X_{k+r+1}, \ldots, X_N$, $k + r + 1 < N$, the bound on the total rate is

$$R_T \geq H(X_1) + \sum_{i=2}^{k} H(X_i|X_{i-1}, \ldots, X_1) + \sum_{j=k+1}^{N} H(X_j|X_{j-1}) + \sum_{i=k+r+2}^{N} H(X_i|X_{i-1}, \ldots, X_{k+r})$$  \hspace{1cm} (10)

Proof: The bound follows Theorem 1 for $X_i$, $i = 1, \ldots, k$ and Theorem 2 for $i = k+1, \ldots, k+r+1$, while for $k+r+1 < i < N$, Theorem 1 holds again, with $X_{k+r}$ replacing $X_1$ since $H(X_i|X_{i-1}, \ldots, X_1) = H(X_i|X_{i-1}, \ldots, X_{k+r})$ and the Markov independence breaks the entropy chain.

V. TOTAL RATE FOR CATEGORISED CORRELATION CASES

In order to simplify the SW bound, it is necessary to organise the nodes into different arrangements based on more practical correlation aspects. Although there are many methods to quantifying the correlation between two entities, we use a correlation function $f_C(X_1, X_2)$, which evaluates the correlation between two nodes $X_1$ and $X_2$ by using a value (distance) metric, in which $\tau$ is the threshold, such that $f_C(X_1, X_2) \geq \tau$ means the correlation between the nodes is strong, with weak correlation otherwise. We use a graph to represent the inter-correlation between $N$ nodes, which can be represented using an adjacency matrix $C$, where $c_{i,j} = f_C(X_i, X_j)$ and $c_{i,j} = 0$ if the correlation is weak, or if $i = j$. We can construct a set for each row $i$ of $C$, defined as $S_i = \{j|j = 1, 2, \ldots, i-1; c_{i,j} \neq 0\}$, which contains all the indices of nodes correlated with $X_i$, where $j < i$. Furthermore, we use the notation $X_{S_i}$ to represent all nodes referenced by the set $S_i$, namely $X_{j_1}, X_{j_2}, \ldots, X_{j_{|S_i|}}$. We now give the SW bound when using this correlation structure.

Theorem 3. Given an adjacency matrix $C$ containing the correlation information for $X_1, X_2, \ldots, X_N$ the bound on the admissible total rate is given by

$$R_T \geq H(X_1) + \sum_{i=2}^{N} H(X_i|X_{S_i})$$  \hspace{1cm} (11)

Proof: The definition of $S_i$ is designed to capture the lower triangular part of $C$. This definition allows the same structure as that in Theorem 1 to be followed, which is the expansion of $H(X_1, X_2, \ldots, X_N)$. The difference lies in the fact that $X_{S_i}$ excludes any terms that are not correlated with $X_i$, giving the desired expansion.

In the next scenario, we fix $\tau$ such that $C$ becomes sparse, implying that only specific subsets of nodes are strongly correlated with one another, but not with any other group of nodes. The modelling for this disjoint grouping case is shown in Fig. 4. Clearly, the set of all nodes is divided into $k$ independent sets. Let $T_i$ be a set containing the indices of the nodes in the $i$th independent set, with $i = 1, 2, \ldots, k$. 
Necessarily, $T_p \cap T_q = \emptyset \ \forall p,q \in \{1,2,\ldots,k\}, p \neq q$. Furthermore, let $X_{i,j}$ be the node referenced by the $j$th element of $T_i$, with $j = 1,2,\ldots,|T_i|$. 

**Theorem 4.** For a system of $N$ correlated sources $X_1,X_2,\ldots,X_N$, and disjoint set distribution $T_i$, $i = 1,2,\ldots,k$, a valid total rate is given by

$$R_T \geq \sum_{i=1}^{k} H(X_{i,1}) \sum_{j=2}^{|T_i|} H(X_{i,j}|X_{i,j-1},\ldots,X_{i,1}) \quad (12)$$

**Proof:** For each $T_i$, the total rate is given by Theorem 1 replacing $X_1,X_2,\ldots,X_N$ with $X_{i,1},X_{i,2},\ldots,X_{i,|T_i|}$ accordingly:

$$R_{i,1} + R_{i,2} + \cdots + R_{i,|T_i|} \geq H(X_{i,1})$$

$$+ \sum_{j=2}^{|T_i|} H(X_{i,j}|X_{i,j-1},\ldots,X_{i,1}) \quad (13)$$

Since each subset is independent, the bound for each total rate associated with $T_i$ can be added without any further manipulations.

As in the Section IV we can also determine the adjusted bound when the nodes within each disjoint set are correlated with a Markov relation.

**Lemma 2.** For a system of $k$ disjointly correlated nodes represented by $T_i$ and modelled with a Markov correlation between them ($X_{i,1} \rightarrow X_{i,2} \rightarrow \cdots \rightarrow X_{i,|T_i|}, i = 1,2,\ldots,k$), a valid total rate is given by

$$R_T \geq \sum_{i=1}^{k} H(X_{i,1}) \sum_{j=2}^{|T_i|} H(X_{i,j}|X_{i,j-1}) \quad (14)$$

**Proof:** The proof follows along the same line of reasoning as Theorem 4 using Lemma 1 instead of Theorem 1.

**VI. CONCLUSION**

The Slepian-Wolf bound has been shown in both its original form, as well as the simplification when modelling the correlation between nodes as a Markov chain. These bounds have then been extended to include an arbitrary correlation structure, by introducing and defining the concepts of “strong” and “weak” correlation and using an adjacency matrix to model the correlation. We also consider the effect on the bound when modelling the correlation with a Markov chain structure. A final simplification is demonstrated by making the adjacency matrix sparse, causing the nodes to be arranged into disjoint groups. Although the bounds presented in this paper are more restrictive than the Slepian-Wolf bound, the modelling complexity is reduced owing to the more practical correlation considerations. By outlining a simple yet efficient modelling approach, we show the potential of improving the coding/decoding complexity of practical SW algorithms and schemes.

**ACKNOWLEDGEMENT**

This research was partially supported by the Council for Scientific and Industrial Research, Pretoria, South Africa, through the Smart Networks collaboration initiative and IoT-Factoy Program (Funded by the Department of Science and Innovation (DSI), South Africa), and partially by South Africa’s National Research Foundation (129311).

**REFERENCES**

[1] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Transactions on Information Theory, vol. 19, no. 4, pp. 471–480, July 1973.

[2] Z. Xiong, A. D. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” IEEE Signal Processing Magazine, vol. 21, no. 5, pp. 80–94, Sept 2004.

[3] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (discus): design and construction,” in Data Compression Conference, 1999, Proceedings. DCC ’99, Mar 1999, pp. 158–167.

[4] D. Schonberg, K. Ramchandran, and S. S. Pradhan, “Distributed code constructions for the entire slepian-wolf rate region for arbitrarily correlated sources,” in Data Compression Conference, 2004. Proceedings. DCC 2004, March 2004, pp. 292–301.

[5] V. Stankovic, A. D. Liveris, Z. Xiong, and C. N. Georghiades, “Design of slepian-wolf codes by channel code partitioning,” in Data Compression Conference, 2004. Proceedings. DCC 2004, March 2004, pp. 302–311.

[6] P. Tan, K. Xie, and J. Li, “Slepian-wolf coding using parity approach and syndrome approach,” in 2007 41st Annual Conference on Information Sciences and Systems, March 2007, pp. 708–713.

[7] X. Zhu, L. Zhang, and Y. Liu, “A distributed joint source-channel coding scheme for multiple correlated sources,” in 2009 Fourth International Conference on Communications and Networking in China, Aug 2009, pp. 1–6.

[8] B. Lu and J. Garcia-Frias, “Analog joint source-channel coding for transmission of correlated senders over separated noisy channels,” in 2015 49th Annual Conference on Information Sciences and Systems (CISS), March 2015, pp. 1–5.

[9] S. Elenjia, U. Abdu-Aguay, M. Ambrose, M. Tomlinson, and M. Zak, “Design of binary LDPC codes for slepian-wolf coding of correlated information sources,” in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov 2017, pp. 1120–1124.

[10] R. Balmahoon, A. J. H. Vinck, and L. Cheng, “Information leakage for correlated sources with compromised source symbols over wiretap channel ii,” in 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2014, pp. 1302–1308.

[11] R. Balmahoon and L. Cheng, “Information leakage of heterogeneous encoded correlated sequences over an eavesdropped channel,” in 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 2949–2953.

[12] T. Cover, “A proof of the data compression theorem of slepian and wolf for ergodic sources (corresp.),” IEEE Transactions on Information Theory, vol. 21, no. 2, pp. 226–228, March 1975.

[13] V. Toto-Zarasoa, A. Roumy, and C. Guillemot, “Rate-adaptive codes for the entire slepian-wolf region and arbitrarily correlated sources,” in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, March 2008, pp. 2965–2968.