EE 8510: Multi-user Information Theory

Distributed Source Coding for Sensor Networks: A Coding Perspective

Final Project Paper

By

Vikrham Gowreesunker

Acknowledgment: Dr. Nihar Jindal

Distributed Source Coding for Sensor Networks: A Coding Perspective

I. Introduction

Distributed Source Coding finds its roots in the work of Slepian-Wolf in 1973[13]. Around that time period, several other authors such as Cover, Wyner and Ziv followed up with some ground breaking extensions to other source models. After almost two decades, research in the area of Distributed Source Coding was reborn because of the growing attention to distributed sensor networks. Advances in Micro-Electro-Mechanical Systems (MEMS) and the demand for sensor network technology in defense programs and non-intrusive surveillance and monitoring programs has seen investment pouring in. Recognizing the trend, BusinessWeek named Sensor Networks one of the hottest technology for this new century and the promise is rapidly turning into a reality. From an Information Theoric perspective, there are relatively few known results in capacity achieving regions for the various sensor networks models. In coding community, there are several application specific efficient coding schemes, but few general methodologies. However, the sensor networks field is relatively new and research is guaranteed to continue for some time.

In this paper, we first review the general sensor network topology and then we focus on a simplified model. We will present recent progress in term of practical coding schemes for the Slepian-Wolf coding and the Wyner-Ziv coding problems. We find that the key to the current progress is a proper combination of theoretical work done decades ago and new algorithm breakthroughs.

II. Slepian-Wolf Model

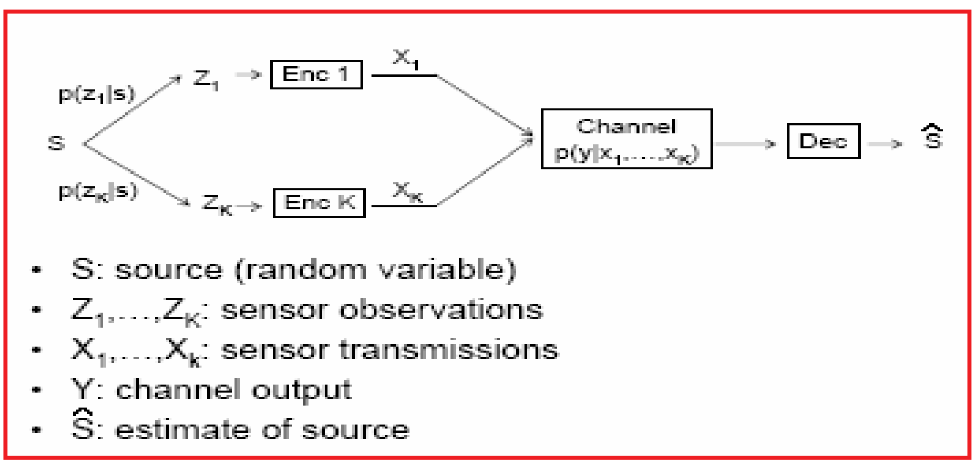

Figure 1.

Consider the deployment of some sensors over a constrained geographical area, with the sensors having to report their observations to a central base station that will then relay the data to a central data monitoring station. Generally, the data gathering sensors are low cost, energy-efficient nodes while the base stations doing the relays are fusion centers that would typically consume more energy. Several setups exist for how the sensor nodes communicate with the fusion center. We will look at a simple model. An event S happens and there are k sensors which records that event and need to communicate their finding to the fusion center. By the definition of this problem, there will be some correlation in the data, Z_k, since all the sensors are monitoring the same event. Figure 1 [8] illustrates the basic sensor model. For the rest of this paper we consider a simplified model where we ignore the channel.

Slepian-Wolf studied the case of lossless encoding of data from 2 sensors X and Y, (X_i,Y_i) drawn i.i.d with distribution p(x,y). Their achievable region is shown in figure 2 [16]. There are 3 parts to this figure. Part (a) shows joint encoding, where both encoders communicate and it is well known that the best rate is H(X,Y). Part (b) shows separate encoding with the best rate not known until Slepian-Wolf's work [13]. Part (c) shows the achievable region for the Slepian-Wolf problem. It reveals the surprising fact that the best rate for distributed, non-cooperative encoding is the same as that for joint encoding at the source, i.e. R1+R2 >= H(X,Y). This result can be proved by random binning, which is actually non-constructive. Pseudo-random binning [4] and algebraic binning [9] have been proposed for constructive codes. There are two types of codes for the SWC, *Asymmetric Coding* and *Symmetric Coding*. The former is when codes are designed to achieve point A and B in figure 2 and timesharing is used to achieve everything in between. The latter is when codes are designed to specifically achieve a rate on the line from A to B. In the results section we look at limit achieving asymmetric codes.

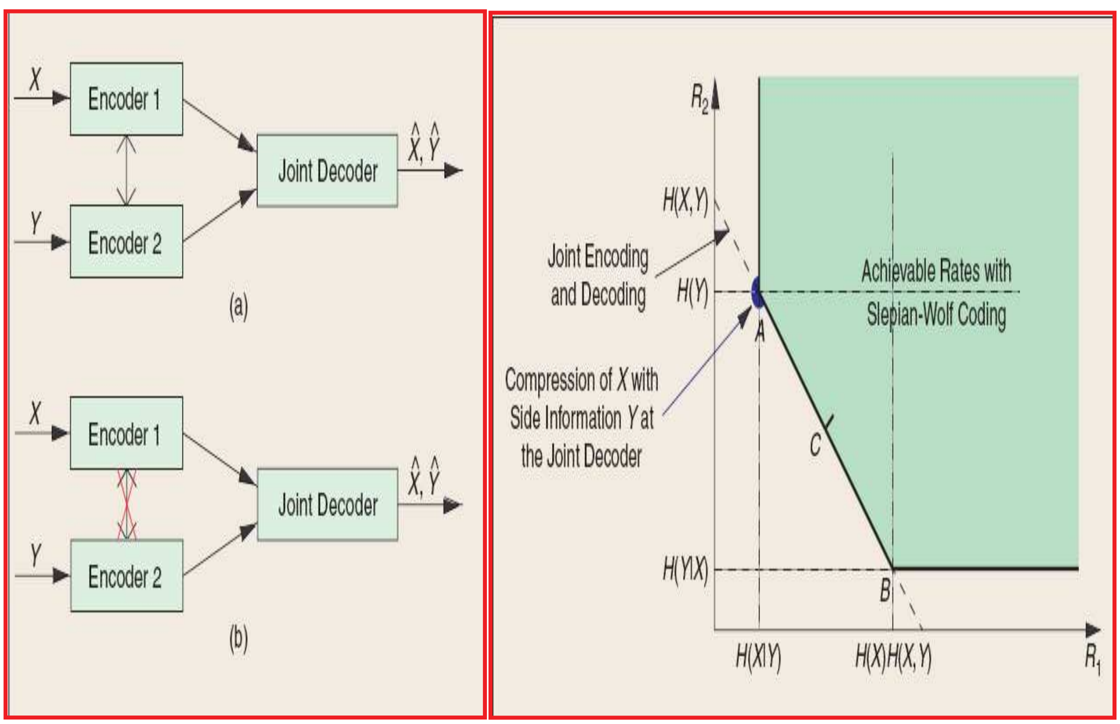

Figure 2.

Cover [2] extended Slepian-Wolf's results for 2 discrete correlated sources to the cases of ergodic processes, countably infinite alphabets and arbitrary number of correlated sources. Another way to look at Slepian-Wolf coding is to recast it into a problem of source coding (X) with side information (Y) at the receiver. We will see the benefit of this scheme in the next section

III. Slepian-Wolf Practical Coding Schemes

By restating the Slepian-Wolf problem as that of lossless encoding with side information available only at decoder, the apparent source coding problem can be viewed as a channel coding problem. In 1974, Wyner [15] drew the parallel between Slepian-Wolf coding and channel coding. This is quite interesting because it would allow us to use our extensive knowledge of channel coding in source coding. In what was known as the “Wyner’s Scheme”, he postulated that one could use linear block code and send syndrome bits to achieve SW limits. Although he first looked at the scheme of binary symmetric channels and hamming distortion measure, the concept can be extended to all binary linear codes and near capacity achieving schemes such as Turbo codes and Low Density Parity Check codes [16].

Linear codes depend on the correlation model between X and Y. In their tutorial paper, Xiong et al. [16] focused on the widely studied binary symmetric correlation model. This model draws an i.i.d sequence $\{X_i,Y_i\}$ which are correlated with a Bernoulli(0.5), leading to a virtual binary symmetric correlation channel [16]. Although not the most practical setup, it is the perfect channel to show the progress of coding scheme over years. The first codes that approached Slepian-Wolf limits were published in the early 2000’s. The schemes borrowed ideas from turbo codes [4] and parity bits [5] but were not quite the same a Wyner’s Scheme which recommends using syndrome bits. In 2003, Liveris et al incorporated the Wyner’s Scheme in their coding design [6] to successfully achieve near Slepian-Wolf limits and distinctly outperforming non-syndrome coding schemes. Figure 3 features results from Liveris et al, comparing their results to non-syndromes codes.

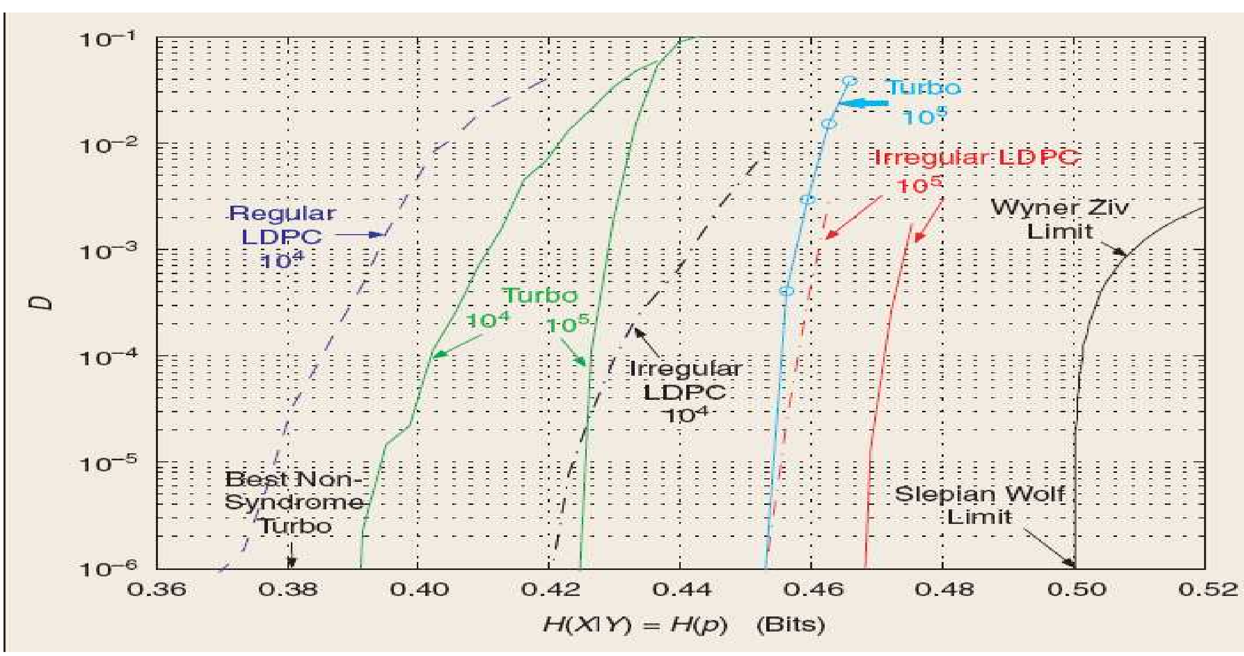

Figure 3

Figure 3 shows a plot of probability of error v/s correlation (higher correlation leads to lower H(X/Y)). We note that there is only a 0.03 bit difference between their best code and the Slepian-Wolf limit. In contrast, the best non-syndrome turbo code is only 0.12 bits from the same limit.

This work can be extended to any correlation model. Xiong et al, states that "if the correlation between the source output X and the side information Y can be modeled with a "virtual" correlation channel, then a good channel code over this channel can provide us with a good Slepian-Wolf code through the syndromes and the associated coset codes." This said, we have transformed the source coding problem of Slepian-Wolf into a channel coding problem that allows us to use high performing iterative codes such as turbo and LDPC codes. Lan et al [6] have recently extended the work to different correlation models and multiterminal cases. Still coding for source with memory and arbitrary number of sources with non binary alphabets evades us, mainly because the channel coding dual are not really well studied.

IV. Wyner-Ziv Model

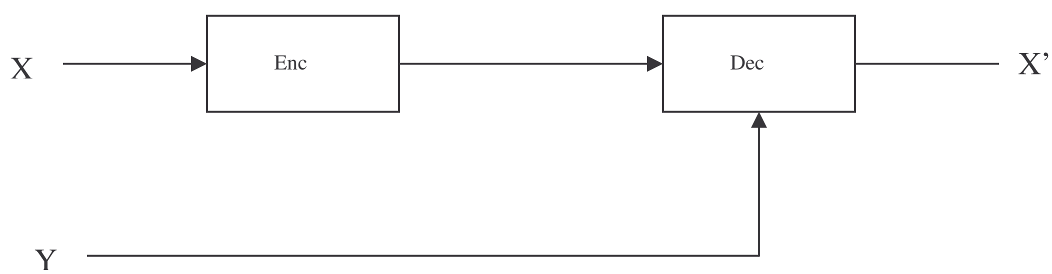

Figure 4

Figure 4 shows the basic setup for the Wyner-Ziv coding problem [14]. Instead of lossless compression as in the Slepian-Wolf case, we now encode X with respect to a distortion criterion and we seek the best rate at which we can send the encoded version X while respecting the distortion criterion.

This problem can be broken down in two parts; first quantization of the source X and then the encoding and transmission of X to the decoder. The first part is a source coding problem which uses results from Quantization Theory while the second part can viewed as a channel coding problem as we explained earlier in the Slepian-Wolf section.

V. Wyner-Ziv Results and Practical Coding Schemes

In general, there is a strict rate loss when we compare the Wyner-Ziv problem and the joint encoding case. However, there are 2 special cases where it has been shown that there is no rate loss when compared to the joint encoding scheme. The first case is that where X and Y are related by X = Y + Z, with both Y and Z are zero mean Gaussian random variables and the distortion metric is the MSE [14]. This result was further extended to show no rate loss when only Z is zero mean Gaussian, leaving arbitrary distribution for X and Y, while using the MSE metric [14].

The widely used coding schemes for the WZC use a source coding scheme like the Trellis Coded Quantization (TCQ) to achieve granular gain and channel coding scheme like Turbo or LDPC code. The work of Pradhan et al on Distributed Source Coding using syndromes (DISCUS) have set the tone for both SWC and WZC [11]. Since its appearance in 1999, several coding paradigm followed, including the Slephian-Wolf Coded Quantization (SWCQ) as described in [16]. This scheme makes use of efficient algebraic binning by grouping source codewords into coset channel codes and then sending one index per code. The results for quadratic Gaussian WZC is closer to theoretical limits than any previous codes. Other efficient practical codes, like PRISM [8] in video communication systems, are more application specific.

VII. Conclusion

While results in Distributed Source Coding has been known for 3 decades now, significant progress in developing practical codes have started only recently. A few factors can be attributed to this gap. First, technology and computational development have provided us with faster and more powerful hardware and simulation tools, which in turn have paved the way to developing powerful tools such as turbo codes. Second, is the renewed interest in the problems of Distributed Source Coding. Applications such as sensor networks have motivated more attention to this problem. And most importantly is the work of researchers who have successfully combined established results with new tools to give us better and more efficient codes.

The problems of distributed source coding extends to more complex cases such as multiterminal source coding, chief executive officer problem, cross-layer design, and problem where sources have memory. There is still much progress to be made in terms of achievable rates and even more work needs to be done for practical and efficient codes. However, as we continue to exploit the various source-channel coding dualities and progress in both communities continue to grow, we look forward to rapidly improving practical codes becoming available.

VI. References

1. R. Barron, B. Chen, and G. Wornell, “The duality between information embedding and source coding with side information and some applications,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1159-1180, May 2003
2. T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991
3. M. Gastpar and M. Vetterli. “Source-channel communication in sensor networks”. In Leonidas J. Guibas and Feng Zhao, editors, *2nd International Workshop on Information Processing in Sensor Networks (IPSN'03), Palo Alto, CA*. Lecture Notes in Computer Science vol. 2634, Springer: New York, April 2003, p. 162-177.

4. J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes", IEEE Communications Letters, vol. 5, pp. 417-419, October 2001.
5. V. Goyal, "Multiple description coding: compression meets the network," IEEE Signal Processing Mag., vol. 18, no. 5, pp. 74–93, 2001.
6. C. Lan, A. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-Wolf coding of multiple *M*-ary sources using LDPC codes," *Proc. DCC'04*, Snowbird, UT, 2004, p. 549.
7. Liveris, Z. Xiong, and C. Georghiades, "Compression with side information at the decoder using LDPC *Lett.*, vol. 6, no. 10, pp. 440–442, 2002.
8. Jindal, N. "EE8510 class notes"
9. P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo like codes," in *Proc. ISIT'02*, Lausanne, Switzerland, 2002, p. 91.
10. R. Puri and K. Ramchandran, "PRISM: A video coding architecture based on distributed compression principles," submitted to IEEE Trans. Image Processing, 2003.
11. 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
12. S.S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding with side information," IEEE Trans. Inform. Theory, vol. 49, pp.1181–1203, May 2003
13. D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471-480, 1973.
14. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1-10, 1976
15. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 2–10, 1974.
16. Zixiang Xiong; Liveris, A.D; Cheng, S. "Distributed source coding for sensor networks". *Signal Processing Magazine, IEEE.* Volume 21, Issue 5, Sept. 2004 Page(s):80 – 94R.
17. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," IEEE Trans. Inform. Theory, vol. 48, pp. 1250-1276, June 2002.