

Distributed source coding

From Wikipedia, the free encyclopedia

Distributed source coding (DSC) is an important problem in information theory and communication. DSC problems regard the compression of multiple correlated information sources that do not communicate with each other.^[1] By modeling the correlation between multiple sources at the decoder side together with channel codes, DSC is able to shift the computational complexity from encoder side to decoder side, therefore provide appropriate frameworks for applications with complexity-constrained sender, such as sensor networks and video/multimedia compression (see distributed video coding^[2]). One of the main properties of distributed source coding is that the computational burden in encoders is shifted to the joint decoder.

Contents

- 1 History
- 2 Theoretical bounds
 - 2.1 Slepian–Wolf bound
 - 2.2 Wyner–Ziv bound
- 3 Virtual channel
- 4 Asymmetric DSC vs. symmetric DSC
- 5 Practical distributed source coding
 - 5.1 Slepian–Wolf coding – lossless distributed coding
 - 5.1.1 General theorem of Slepian–Wolf coding with syndromes for two sources
 - 5.1.2 Slepian–Wolf coding example
 - 5.1.2.1 Asymmetric case (,)
 - 5.1.2.2 Symmetric case
 - 5.2 Wyner–Ziv coding – lossy distributed coding
 - 5.3 Large scale distributed quantization
- 6 Non-asymmetric DSC
- 7 Non-asymmetric DSC for more than two sources
- 8 See also
- 9 References

History

In 1973, David Slepian and Jack Keil Wolf proposed the information theoretical lossless compression bound on distributed compression of two correlated i.i.d. sources X and Y .^[3] After that, this bound was extended to cases with more than two sources by Thomas M. Cover in 1975,^[4] while the theoretical results in the lossy compression case are presented by Aaron D. Wyner and Jacob Ziv in 1976.^[5]

Although the theorems on DSC were proposed on 1970s, it was after about 30 years that attempts were started for practical techniques, based on the idea that DSC is closely related to channel coding proposed in 1974 by Aaron D. Wyner.^[6] The asymmetric DSC problem was addressed by S. S. Pradhan and K. Ramchandran in

1999, which focused on statistically dependent binary and Gaussian sources and used scalar and trellis coset constructions to solve the problem.^[7] They further extended the work into the symmetric DSC case.^[8]

Syndrome decoding technology was first used in distributed source coding by the DISCUS system of SS Pradhan and K Ramachandran (Distributed Source Coding Using Syndromes).^[7] They compress binary block data from one source into syndromes and transmit data from the other source uncompressed as side information. This kind of DSC scheme achieves asymmetric compression rates per source and results in *asymmetric* DSC. This asymmetric DSC scheme can be easily extended to the case of more than two correlated information sources. There are also some DSC schemes that use parity bits rather than syndrome bits.

The correlation between two sources in DSC has been modeled as a virtual channel which is usually referred as a binary symmetric channel.^{[9][10]}

Starting from DISCUS, DSC has attracted significant research activity and more sophisticated channel coding techniques have been adopted into DSC frameworks, such as Turbo Code, LDPC Code, and so on.

Similar to the previous lossless coding framework based on Slepian–Wolf theorem, efforts have been taken on lossy cases based on the Wyner–Ziv theorem. Theoretical results on quantizer designs was provided by R. Zamir and S. Shamai,^[11] while different frameworks have been proposed based on this result, including a nested lattice quantizer and a trellis-coded quantizer.

Moreover, DSC has been used in video compression for applications which require low complexity video encoding, such as sensor networks, multiview video camcorders, and so on.^[12]

With deterministic and probabilistic discussions of correlation model of two correlated information sources, DSC schemes with more general compressed rates have been developed.^{[13][14][15]} In these *non-asymmetric* schemes, both of two correlated sources are compressed.

Under a certain deterministic assumption of correlation between information sources, a DSC framework in which any number of information sources can be compressed in a distributed way has been demonstrated by X. Cao and M. Kuijper.^[16] This method performs non-asymmetric compression with flexible rates for each source, achieving the same overall compression rate as repeatedly applying asymmetric DSC for more than two sources. Then, by investigating the unique connection between syndromes and complementary codewords of linear codes, they have translated the major steps of DSC joint decoding into syndrome decoding followed by channel encoding via a linear block code and also via its complement code,^[17] which theoretically illustrated a method of assembling a DSC joint decoder from linear code encoders and decoders.

Theoretical bounds

The information theoretical lossless compression bound on DSC (the Slepian–Wolf bound) was first purposed by David Slepian and Jack Keil Wolf in terms of entropies of correlated information sources in 1973.^[3] They also showed that two isolated sources can compress data as efficiently as if they were communicating with each other. This bound has been extended to the case of more than two correlated sources by Thomas M. Cover in 1975.^[4]

Similar results were obtained in 1976 by Aaron D. Wyner and Jacob Ziv with regard to lossy coding of joint Gaussian sources.^[5]

Slepian–Wolf bound

Distributed Coding is the coding of two or more dependent sources with separate encoders and joint decoder. Given two statistically dependent i.i.d. finite-alphabet random sequences X and Y , Slepian–Wolf theorem includes theoretical bound for the lossless coding rate for distributed coding of the two sources as below:^[3]

$$R_X \geq H(X|Y),$$

$$R_Y \geq H(Y|X),$$

$$R_X + R_Y \geq H(X, Y).$$

If both the encoder and decoder of the two sources are independent, the lowest rate we can achieve for lossless compression is $H(X)$ and $H(Y)$ for X and Y respectively, where $H(X)$ and $H(Y)$ are the entropies of X and Y . However, with joint decoding, if vanishing error probability for long sequences is accepted, the Slepian–Wolf theorem shows that much better compression rate can be achieved. As long as the total rate of X and Y is larger than their joint entropy $H(X, Y)$ and none of the sources is encoded with a rate larger than its entropy, distributed coding can achieve arbitrarily small error probability for long sequences.

A special case of distributed coding is compression with decoder side information, where source Y is available at the decoder side but not accessible at the encoder side. This can be treated as the condition that $R_Y = H(Y)$ has already been used to encode Y , while we intend to use $H(X|Y)$ to encode X . The whole system is operating in an asymmetric way (compression rate for the two sources are asymmetric).

Wyner–Ziv bound

Shortly after Slepian–Wolf theorem on lossless distributed compression was published, the extension to lossy compression with decoder side information was proposed as Wyner–Ziv theorem.^[5] Similarly to lossless case, two statistically dependent i.i.d. sources X and Y are given, where Y is available at the decoder side but not accessible at the encoder side. Instead of lossless compression in Slepian–Wolf theorem, Wyner–Ziv theorem looked into the lossy compression case.

Wyner–Ziv theorem presents the achievable lower bound for the bit rate of X at given distortion D . It was found that for Gaussian memoryless sources and mean-squared error distortion, the lower bound for the bit rate of X remain the same no matter whether side information is available at the encoder or not.

Virtual channel

Deterministic model

Probabilistic model

Asymmetric DSC vs. symmetric DSC

Asymmetric DSC means that, different bitrates are used in coding the input sources, while same bitrate is used in symmetric DSC. Taking a DSC design with two sources for example, in this example X and Y are two discrete, memoryless, uniformly distributed sources which generate set of variables \mathbf{x} and \mathbf{y} of length 7 bits and the Hamming distance between \mathbf{x} and \mathbf{y} is at most one. The Slepian–Wolf bound for them is:

$$\begin{aligned} R_X + R_Y &\geq 10 \\ R_X &\geq 5 \\ R_Y &\geq 5 \end{aligned}$$

This means, the theoretical bound is $R_X + R_Y = 10$ and symmetric DSC means 5 bits for each source. Other pairs with $R_X + R_Y = 10$ are asymmetric cases with different bit rate distributions between X and Y , where $R_X = 3, R_Y = 7$ and $R_X = 7, R_Y = 3$ represent two extreme cases called decoding with side information.

Practical distributed source coding

Slepian–Wolf coding – lossless distributed coding

It was understood that Slepian–Wolf coding is closely related to channel coding in 1974,^[6] and after about 30 years, practical DSC started to be implemented by different channel codes. The motivation behind the use of channel codes is from two sources case, the correlation between input sources can be modeled as a virtual channel which has input as source X and output as source Y . The DISCUS system proposed by S. S. Pradhan and K. Ramchandran in 1999 implemented DSC with syndrome decoding, which worked for asymmetric case and was further extended to symmetric case.^{[7][8]}

The basic framework of syndrome based DSC is that, for each source, its input space is partitioned into several cosets according to the particular channel coding method used. Every input of each source gets an output indicating which coset the input belongs to, and the joint decoder can decode all inputs by received coset indices and dependence between sources. The design of channel codes should consider the correlation between input sources.

A group of codes can be used to generate coset partitions,^[18] such as trellis codes and lattice codes. Pradhan and Ramchandran designed rules for construction of sub-codes for each source, and presented result of trellis-based coset constructions in DSC, which is based on convolution code and set-partitioning rules as in Trellis modulation, as well as lattice code based DSC.^{[7][8]} After this, embedded trellis code was proposed for asymmetric coding as an improvement over their results.^[19]

After DISCUS system was proposed, more sophisticated channel codes have been adapted to the DSC system, such as Turbo Code, LDPC Code and Iterative Channel Code. The encoders of these codes are usually simple and easy to implement, while the decoders have much higher computational complexity and are able to get good performance by utilizing source statistics. With sophisticated channel codes which have performance approaching the capacity of the correlation channel, corresponding DSC system can approach the Slepian–Wolf bound.

Although most research focused on DSC with two dependent sources, Slepian–Wolf coding has been extended to more than two input sources case, and sub-codes generation methods from one channel code was proposed by V. Stankovic, A. D. Liveris, etc. given particular correlation models.^[20]

General theorem of Slepian–Wolf coding with syndromes for two sources

Theorem: Any pair of correlated uniformly distributed sources, $X, Y \in \{0, 1\}^n$, with $d_H(X, Y) \leq t$, can be compressed separately at a rate pair (R_1, R_2) such that $R_1, R_2 \geq n - k, R_1 + R_2 \geq 2n - k$, where R_1 and R_2 are integers, and $k \leq n - \log\left(\sum_{i=0}^t \binom{n}{i}\right)$. This can be achieved using an $(n, k, 2t + 1)$ binary linear code.

Proof: The Hamming bound for an $(n, k, 2t + 1)$ binary linear code is $k \leq n - \log\left(\sum_{i=0}^t \binom{n}{i}\right)$, and we have Hamming code achieving this bound, therefore we have such a binary linear code \mathbf{C} with $k \times n$ generator matrix \mathbf{G} . Next we will show how to construct syndrome encoding based on this linear code.

Let $R_1 + R_2 = 2n - k$ and \mathbf{G}_1 be formed by taking first $(n - R_1)$ rows from \mathbf{G} , while \mathbf{G}_2 is formed using the remaining $(n - R_2)$ rows of \mathbf{G} . \mathbf{C}_1 and \mathbf{C}_2 are the subcodes of the Hamming code generated by \mathbf{G}_1 and \mathbf{G}_2 respectively, with \mathbf{H}_1 and \mathbf{H}_2 as their parity check matrices.

For a pair of input (\mathbf{x}, \mathbf{y}) , the encoder is given by $\mathbf{s}_1 = \mathbf{H}_1 \mathbf{x}$ and $\mathbf{s}_2 = \mathbf{H}_2 \mathbf{y}$. That means, we can represent \mathbf{x} and \mathbf{y} as $\mathbf{x} = \mathbf{u}_1 \mathbf{G}_1 + \mathbf{c}_{s1}$, $\mathbf{y} = \mathbf{u}_2 \mathbf{G}_2 + \mathbf{c}_{s2}$, where $\mathbf{c}_{s1}, \mathbf{c}_{s2}$ are the representatives of the cosets of $\mathbf{s}_1, \mathbf{s}_2$ with regard to $\mathbf{C}_1, \mathbf{C}_2$ respectively. Since we have $\mathbf{y} = \mathbf{x} + \mathbf{e}$ with $w(\mathbf{e}) \leq t$. We can get $\mathbf{x} + \mathbf{y} = \mathbf{uG} + \mathbf{c}_s = \mathbf{e}$, where $\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2]$, $\mathbf{c}_s = \mathbf{c}_{s1} + \mathbf{c}_{s2}$.

Suppose there are two different input pairs with the same syndromes, that means there are two different strings $\mathbf{u}^1, \mathbf{u}^2 \in \{0, 1\}^k$, such that $\mathbf{u}^1 \mathbf{G} + \mathbf{c}_s = \mathbf{e}$ and $\mathbf{u}^2 \mathbf{G} + \mathbf{c}_s = \mathbf{e}$. Thus we will have $(\mathbf{u}^1 - \mathbf{u}^2) \mathbf{G} = \mathbf{0}$. Because minimum Hamming weight of the code \mathbf{C} is $2t + 1$, the distance between $\mathbf{u}_1 \mathbf{G}$ and $\mathbf{u}_2 \mathbf{G}$ is $\geq 2t + 1$. On the other hand, according to $w(\mathbf{e}) \leq t$ together with $\mathbf{u}^1 \mathbf{G} + \mathbf{c}_s = \mathbf{e}$ and $\mathbf{u}^2 \mathbf{G} + \mathbf{c}_s = \mathbf{e}$, we will have $d_H(\mathbf{u}^1 \mathbf{G}, \mathbf{c}_s) \leq t$ and $d_H(\mathbf{u}^2 \mathbf{G}, \mathbf{c}_s) \leq t$, which contradict with $d_H(\mathbf{u}^1 \mathbf{G}, \mathbf{u}^2 \mathbf{G}) \geq 2t + 1$. Therefore, we cannot have more than one input pairs with the same syndromes.

Therefore, we can successfully compress the two dependent sources with constructed subcodes from an $(n, k, 2t + 1)$ binary linear code, with rate pair (R_1, R_2) such that

$$R_1, R_2 \geq n - k, R_1 + R_2 \geq 2n - k, \text{ where } R_1 \text{ and } R_2 \text{ are integers, and } k \leq n - \log\left(\sum_{i=0}^t \binom{n}{i}\right)$$

. \log indicates \log_2 .

Slepian–Wolf coding example

Take the same example as in the previous **Asymmetric DSC vs. Symmetric DSC** part, this part presents the corresponding DSC schemes with coset codes and syndromes including asymmetric case and symmetric case. The Slepian–Wolf bound for DSC design is shown in the previous part.

Asymmetric case ($R_X = 3, R_Y = 7$)

In this case, the length of an input variable \mathbf{Y} from source Y is 7 bits, therefore it can be sent lossless with 7 bits independent of any other bits. Based on the knowledge that \mathbf{x} and \mathbf{y} have Hamming distance at most one, for input \mathbf{x} from source X , since the receiver already has \mathbf{y} , the only possible \mathbf{x} are those with at most 1 distance from \mathbf{y} . If we model the correlation between two sources as a virtual channel, which has input \mathbf{x} and output \mathbf{y} , as long as we get \mathbf{y} , all we need to successfully "decode" \mathbf{x} is "parity bits" with particular error correction ability, taking the difference between \mathbf{x} and \mathbf{y} as channel error. We can also model the problem with cosets partition. That is, we want to find a channel code, which is able to partition the space of input X into several cosets, where each coset has a unique syndrome associated with it. With a given coset and \mathbf{y} , there is only one \mathbf{x} that is possible to be the input given the correlation between two sources.

In this example, we can use the $(7, 4, 3)$ binary Hamming Code \mathbf{C} , with parity check matrix \mathbf{H} . For an input \mathbf{x} from source X , only the syndrome given by $\mathbf{s} = \mathbf{H}\mathbf{x}$ is transmitted, which is 3 bits. With received \mathbf{y} and \mathbf{s} , suppose there are two inputs \mathbf{x}_1 and \mathbf{x}_2 with same syndrome \mathbf{s} . That means $\mathbf{H}\mathbf{x}_1 = \mathbf{H}\mathbf{x}_2$, which is $\mathbf{H}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. Since the minimum Hamming weight of $(7, 4, 3)$ Hamming Code is 3, $d_H(\mathbf{x}_1, \mathbf{x}_2) \geq 3$. Therefore the input \mathbf{x} can be recovered since $d_H(\mathbf{x}, \mathbf{y}) \leq 1$.

Similarly, the bits distribution with $R_X = 7, R_Y = 3$ can be achieved by reversing the roles of X and Y .

Symmetric case

In symmetric case, what we want is equal bitrate for the two sources: 5 bits each with separate encoder and joint decoder. We still use linear codes for this system, as we used for asymmetric case. The basic idea is similar, but in this case, we need to do coset partition for both sources, while for a pair of received syndromes (corresponds to one coset), only one pair of input variables are possible given the correlation between two sources.

Suppose we have a pair of linear code \mathbf{C}_1 and \mathbf{C}_2 and an encoder-decoder pair based on linear codes which can achieve symmetric coding. The encoder output is given by: $\mathbf{s}_1 = \mathbf{H}_1\mathbf{x}$ and $\mathbf{s}_2 = \mathbf{H}_2\mathbf{y}$. If there exists two pair of valid inputs $\mathbf{x}_1, \mathbf{y}_1$ and $\mathbf{x}_2, \mathbf{y}_2$ generating the same syndromes, i.e. $\mathbf{H}_1\mathbf{x}_1 = \mathbf{H}_1\mathbf{x}_2$ and $\mathbf{H}_1\mathbf{y}_1 = \mathbf{H}_1\mathbf{y}_2$, we can get following ($w(\cdot)$ represents Hamming weight):

$$\mathbf{y}_1 = \mathbf{x}_1 + \mathbf{e}_1, \text{ where } w(\mathbf{e}_1) \leq 1$$

$$\mathbf{y}_2 = \mathbf{x}_2 + \mathbf{e}_2, \text{ where } w(\mathbf{e}_2) \leq 1$$

$$\text{Thus: } \mathbf{x}_1 + \mathbf{x}_2 \in \mathbf{C}_1$$

$$\mathbf{y}_1 + \mathbf{y}_2 = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{e}_3 \in \mathbf{C}_2$$

where $\mathbf{e}_3 = \mathbf{e}_2 + \mathbf{e}_1$ and $w(\mathbf{e}_3) \leq 2$. That means, as long as we have the minimum distance between the two codes larger than 3, we can achieve error-free decoding.

The two codes \mathbf{C}_1 and \mathbf{C}_2 can be constructed as subcodes of the $(7, 4, 3)$ Hamming code and thus has minimum distance of 3. Given the generator matrix \mathbf{G} of the original Hamming code, the generator matrix \mathbf{G}_1 for \mathbf{C}_1 is constructed by taking any two rows from \mathbf{G} , and \mathbf{G}_2 is constructed by the remaining two rows of \mathbf{G} . The corresponding (5×7) parity-check matrix for each sub-code can be generated according to the generator matrix and used to generate syndrome bits.

Wyner–Ziv coding – lossy distributed coding

In general, a Wyner–Ziv coding scheme is obtained by adding a quantizer and a de-quantizer to the Slepian–Wolf coding scheme. Therefore, a Wyner–Ziv coder design could focus on the quantizer and corresponding reconstruction method design. Several quantizer designs have been proposed, such as a nested lattice quantizer,^[21] trellis code quantizer^[22] and Lloyd quantization method.^[23]

Large scale distributed quantization

Unfortunately, the above approaches do not scale (in design or operational complexity requirements) to sensor networks of large sizes, a scenario where distributed compression is most helpful. If there are N sources transmitting at R bits each (with some distributed coding scheme), the number of possible reconstructions scales 2^{NR} . Even for moderate values of N and R (say $N=10$, $R=2$), prior design schemes become impractical. Recently, an approach,^[24] using ideas borrowed from Fusion Coding of Correlated Sources, has been proposed where design and operational complexity are traded against decoder performance. This has allowed distributed quantizer design for network sizes reaching 60 sources, with substantial gains over traditional approaches.

The central idea is the presence of a bit-subset selector which maintains a certain subset of the received (NR bits, in the above example) bits for each source. Let \mathcal{B} be the set of all subsets of the NR bits i.e.

$$\mathcal{B} = 2^{\{1, \dots, NR\}}$$

Then, we define the bit-subset selector mapping to be

$$\mathcal{S} : \{1, \dots, N\} \rightarrow \mathcal{B}$$

Note that each choice of the bit-subset selector imposes a storage requirement (C) that is exponential in the cardinality of the set of chosen bits.

$$C = \sum_{n=1}^N 2^{|\mathcal{S}(n)|}$$

This allows a judicious choice of bits that minimize the distortion, given the constraints on decoder storage. Additional limitations on the set of allowable subsets are still needed. The effective cost function that needs to be minimized is a weighted sum of distortion and decoder storage

$$J = D + \lambda C$$

The system design is performed by iteratively (and incrementally) optimizing the encoders, decoder and bit-subset selector till convergence.

Non-asymmetric DSC

Non-asymmetric DSC for more than two sources

The syndrome approach can still be used for more than two sources. Let us consider a binary sources of length- n $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_a \in \{0, 1\}^n$. Let $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_s$ be the corresponding coding matrices of sizes $m_1 \times n, m_2 \times n, \dots, m_a \times n$. Then the input binary sources are compressed into $\mathbf{s}_1 = \mathbf{H}_1 \mathbf{x}_1, \mathbf{s}_2 = \mathbf{H}_2 \mathbf{x}_2, \dots, \mathbf{s}_a = \mathbf{H}_a \mathbf{x}_a$ of total $m = m_1 + m_2 + \dots + m_a$ bits. Apparently, two source tuples cannot be recovered at the same time if they share the same syndrome. In other words, if all source tuples of interest have different syndromes, then one can recover them losslessly.

General theoretical result does not seem to exist. However, for a restricted kind of source so-called Hamming source^[25] that only has at most one source different from the rest and at most one bit location not all identical, practical lossless DSC is shown to exist in some cases. For the case when there are more than two sources, the number of source tuple in a Hamming source is $2^n(an + 1)$. Therefore, a packing bound that $2^m \geq 2^n(an + 1)$ obviously has to satisfy. When the packing bound is satisfied with equality, we may call such code to be perfect (an analogous of perfect code in error correcting code).^[25]

A simplest set of a, n, m to satisfy the packing bound with equality is $a = 3, n = 5, m = 9$. However, it turns out that such syndrome code does not exist.^[26] The simplest (perfect) syndrome code with more than two sources have $n = 21$ and $m = 27$. Let

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix},$$

$$\mathbf{Q}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{pmatrix},$$

$$\mathbf{Q}_3 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \mathbf{G} = [\mathbf{0} | \mathbf{I}_9], \text{ and } \mathbf{G} = \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \mathbf{G}_3 \end{pmatrix} \text{ such that}$$

$\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ are any partition of \mathbf{G} .

$\mathbf{H}_1 = \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{Q}_1 \end{pmatrix}, \mathbf{H}_2 = \begin{pmatrix} \mathbf{G}_2 \\ \mathbf{Q}_2 \end{pmatrix}, \mathbf{H}_3 = \begin{pmatrix} \mathbf{G}_3 \\ \mathbf{Q}_3 \end{pmatrix}$ can compress a Hamming source (i.e., sources that have no more than one bit different will all have different syndromes).^[25] For example, for the symmetric case, a possible

set of coding matrices are $\mathbf{H}_1 =$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix},$$

$$\mathbf{H}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{pmatrix},$$

$$\mathbf{H}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

See also

- Linear code
- Syndrome decoding
- Low-density parity-check code
- Turbo Code

References

1. "Distributed source coding for sensor networks" by Z. Xiong, A.D. Liveris, and S. Cheng (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1328091)
2. "Distributed video coding in wireless sensor networks" by Puri, R. Majumdar, A. Ishwar, P. Ramchandran, K. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1657820&isnumber=34703)
3. "Noiseless coding of correlated information sources" by D. Slepian and J. Wolf (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1055037)
4. "A proof of the data compression theorem of Slepian and Wolf for ergodic sources" by T. Cover (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1055356)
5. "The rate-distortion function for source coding with side information at the decoder" by A. Wyner and J. Ziv (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1055508)
6. "Recent results in Shannon theory" by A. D. Wyner (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=1055171)
7. "Distributed source coding using syndromes (DISCUS): design and construction" by S. S. Pradhan and K. Ramchandran (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1055508)
8. "Distributed source coding: symmetric rates and applications to sensor networks" by S. S. Pradhan and K. Ramchandran (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=838176)
9. "Distributed code constructions for the entire Slepian–Wolf rate region for arbitrarily correlated sources" by Schonberg, D. Ramchandran, K. Pradhan, S.S. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1281474)
10. "Generalized coset codes for distributed binning" by Pradhan, S.S. Ramchandran, K. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1512420)
11. "Nested linear/lattice codes for Wyner–Ziv encoding" by R. Zamir and S. Shamai (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=706450)
12. "Distributed Video Coding" by B. Girod, etc. (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=1369699)
13. "On code design for the Slepian–Wolf problem and lossless multiterminal networks" by Stankovic, V. Liveris, A.D. Zixiang Xiong Georgiades, C.N. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1614079)
14. "A general and optimal framework to achieve the entire rate region for Slepian–Wolf coding" by P. Tan and J. Li (<http://portal.acm.org/citation.cfm?id=1226544>)
15. "Distributed source coding using short to moderate length rate-compatible LDPC codes: the entire Slepian–Wolf rate region" by Sartipi, M. Fekri, F. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4471935)
16. "A distributed source coding framework for multiple sources" by Xiaomin Cao and Kuijper, M. (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=4895364&arnumber=4895396&count=299&index=31)
17. [1] (http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6291702&refinements%3D4281469143%26sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A6362326%29) "Distributed Source Coding via Linear Block Codes: A General Framework for Multiple Sources" by Xiaomin Cao and Kuijper, M.
18. "Coset codes. I. Introduction and geometrical classification" by G. D. Forney (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=21245)
19. "Design of trellis codes for source coding with side information at the decoder" by X. Wang and M. Orchard (http://ieeexplore.ieee.org/xpls/freeabs_all.jsp?arnumber=917167)
20. "Design of Slepian–Wolf codes by channel code partitioning" by V. Stankovic, A. D. Liveris, Z. Xiong and C. N. Georgiades (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1281475)
21. "Nested quantization and Slepian–Wolf coding: a Wyner–Ziv coding paradigm for i.i.d. sources" by Z. Xiong, A. D. Liveris, S. Cheng and Z. Liu (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1289429)
22. "Wyner–Ziv coding based on TCQ and LDPC codes" by Y. Yang, S. Cheng, Z. Xiong and W. Zhao (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4784347)
23. "Design of optimal quantizers for distributed source coding" by D. Rebollo-Monedero, R. Zhang and B. Girod (http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1193992)
24. "Towards large scale distributed source coding" by S. Ramaswamy, K. Viswanatha, A. Saxena and K. Rose (http://www.sc.lece.ucsb.edu/pubs/pubs_D/d10_4.pdf)
25. "Hamming Codes for Multiple Sources" by R. Ma and S. Cheng (<http://arxiv.org/pdf/1001.4072>)
26. "The Non-existence of Length-5 Slepian–Wolf Codes of Three Sources" by S. Cheng and R. Ma (http://tulsagrad.ou.edu/samuel_cheng/papers/dcc10.pdf)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Distributed_source_coding&oldid=633690221"

Categories: [Information theory](#) | [Coding theory](#) | [Wireless sensor network](#) | [Data transmission](#)

- This page was last modified on 13 November 2014, at 17:22.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.