Research Repository UCD

Title	Optimum Reversible Data Hiding and Permutation Coding
Authors(s)	Balado, Félix
Publication date	2015-11-19
Publication information	Balado, Félix. "Optimum Reversible Data Hiding and Permutation Coding." IEEE, 2015.
Conference details	7th IEEE International Workshop on Information Forensics and Security (WIFS), Rome, Italy, 16 - 19 November, 2015
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/7259
Publisher's statement	© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Optimum Reversible Data Hiding and Permutation Coding

Félix Balado

School of Computer Science
University College Dublin
Belfield Campus, Dublin 4, Ireland
Email: felix@ucd.ie

Abstract—This paper is mainly devoted to investigating the connection between binary reversible data hiding and permutation coding. We start by undertaking an approximate combinatorial analysis of the embedding capacity of reversible watermarking in the binary Hamming case, which asymptotically shows that optimum reversible watermarking must involve not only “writing on dirty paper”, as in any blind data hiding scenario, but also writing on the dirtiest parts of the paper. The asymptotic analysis leads to the information-theoretical result given by Kalker and Willems more than a decade ago. Furthermore, the novel viewpoint of the problem suggests a near-optimum reversible watermarking algorithm for the low embedding distortion regime based on permutation coding. A practical implementation of permutation coding, previously proposed in the context of maximum-rate perfect steganography of memoryless hosts, can be used to implement the algorithm. The paper concludes with a discussion on the evaluation of the general rate-distortion bound for reversible data hiding.

I. INTRODUCTION

Reversible data hiding, sometimes also called invertible, revertible, or lossless data hiding, harks back to the beginnings of research in digital watermarking. The origin of the subject lies in the development of blind watermarking techniques which can perfectly recover the original host from the information-carrying watermarked signal. This special property is desirable in the application of data hiding to medical imaging, legal data, cartographic information, and other sensitive areas in which data integrity is paramount.

The earliest work on reversible watermarking is apparently found in a 1997 patent [1], and the “reversible watermarking” name was itself coined later in the same year [2]. The first relevant reversible data hiding algorithm was devised some years later by Fridrich et al. [3], who proposed a method based on optimum compression of the least-significant bit plane of a host replacing its least-significant bit plane. However it was the work of Kalker and Willems which definitively set reversible watermarking on a sound footing, by revealing the information-theoretical limits of the problem (see [4], later generalised in [5] and [6]). This work showed that Fridrich et al.’s approach was only optimum for a particular distortion constraint. Ever since, a steady supply of papers proposing reversible watermarking algorithms have kept appearing in different venues, perhaps swimming against the tide: in recent years interest in data hiding research has arguably waned — especially after many (but by no means all) of the original

research challenges were eventually worked out. In this context, the intermittent resurfacing of reversible watermarking has probably been a consequence of the fact that the results by Kalker and Willems have mostly remained sidelined. This has forestalled movement towards some kind of closure of the subject. Although a few authors have cited [4] in passing, precious little has happened in the intervening years in the way of producing algorithms that can actually approach the rate-distortion limit: most authors never even compare their algorithms against this fundamental limit, but only against other methods. The remarkable exception is the work by Zhang et al. [7], [8].

In this article we reexamine the problem of optimum reversible data hiding in order to give it some further attention from a combinatorial point of view. In recent years permutation coding has been shown to be of relevance in two canonical information security problems: optimum perfect steganography of memoryless hosts [9] and optimum perfect histogram-based counterforensics [10]. Here we will show that permutation coding and combinatorics can also make theoretical and practical inroads into yet another standard information security problem.

II. NOTATION AND FRAMEWORK

Boldface lowercase Roman letters are column vectors. We will refer to a vector of n elements as an n -vector. The i -th element of vector $\mathbf{a}$ is denoted by a_i . The special symbol $\mathbf{0}$ is the null vector. $(\cdot)^t$ is the transpose operator. The indicator function is defined as $\mathbb{1}_{\{\theta\}} = 1$ if logical expression θ is true, and zero otherwise. The 2-norm of a vector $\mathbf{r}$ is $\|\mathbf{r}\| = \sqrt{\mathbf{r}^t \mathbf{r}}$. The Hamming distance between two n -vectors $\mathbf{r}$ and $\mathbf{s}$ is $\delta(\mathbf{r}, \mathbf{s}) \triangleq \sum_{i=1}^n \mathbb{1}_{\{r_i \neq s_i\}}$. The Hamming weight of $\mathbf{r}$ is $\omega(\mathbf{r}) \triangleq \delta(\mathbf{r}, \mathbf{0})$. Calligraphic letters are sets; $|\mathcal{V}|$ is the cardinality of set $\mathcal{V}$. A host sequence is denoted by the n -vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^t$. An information-carrying sequence is denoted by $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$. As different permutations of the elements of $\mathbf{x}$ can in general lead to the same vector, we will follow the convention that a *rearrangement* of $\mathbf{x}$ is a unique ordering of its elements. The number of rearrangements of $\mathbf{x}$ is the binomial coefficient $\binom{n}{\omega(\mathbf{x})}$. Italicised Roman capital letters are used to represent random variables. The expectation, variance, and entropy of X are denoted by $E\{X\}$, $\text{Var}\{X\}$,

and $H(X)$, respectively. Logarithms are base 2 throughout the paper.

We recall next the basic setting in noiseless reversible data hiding. The encoder wishes to modify the host $\mathbf{x}$ in order to produce a new vector $\mathbf{y}$ that encodes the message $m \in \{1, 2, \dots, 2^{\lfloor \rho n \rfloor}\}$, where ρ is the embedding rate, or embedded bits per element of $\mathbf{x}$. As in any other watermarking setting, the encoding function can be put as $\mathbf{y} = e(\mathbf{x}, m)$, where $\mathbf{y}$ is constrained to be close to $\mathbf{x}$ according to a distortion function $s(\cdot, \cdot)$. This is usually stated on average, namely $\mathbb{E}\{s(X, Y)\} \leq \Delta$. In the binary case that will mainly occupy us here distortion is usually measured through the Hamming distance, in which case $s(x, y) = \mathbb{1}_{\{x \neq y\}}$. In reversible watermarking the channel is typically considered to be noiseless, and thus the decoder receives $\mathbf{y}$ undistorted. The decoder applies a decoding function to $\mathbf{y}$ to retrieve the encoded message, that is to say, $d(\mathbf{y}) = m$. Furthermore, the encoder must be able to exactly retrieve the original host through a reconstruction function $r(\mathbf{y}) = \mathbf{x}$. The functions $e(\cdot, \cdot)$, $d(\cdot)$ and $r(\cdot)$ may all be secured through a symmetric key shared by encoder and decoder.

III. CAPACITY OF BINARY REVERSIBLE WATERMARKING

In this section we provide an informal rederivation of the capacity result by Kalker and Willems for the achievable rate of noiseless reversible watermarking in the binary Hamming case, by means of combinatorial arguments. We will draw on the combinatorial line of reasoning originally used by Feynman in [11] to informally rederive the capacity of the binary symmetric channel (BSC), although the problem addressed here will require a different approach. In any case, like Feynman's, our procedure does not explicitly rely on Shannon theory; in particular, Gel'fand and Pinsker's capacity formula for channels with side information at the encoder is not used.

We will assume that elements of the host $\mathbf{x} \in \{0, 1\}^n$ are drawn from an independent and identically distributed (i.i.d.) Bernoulli distribution with $\Pr\{X = 1\} = p$, where $p \geq 1/2$. Since $\omega(\mathbf{X})$ follows a Binomial distribution with parameters $t \triangleq \mathbb{E}\{\omega(\mathbf{X})\} = np$ (which we assume integer) and $\text{Var}\{\omega(\mathbf{X})\} = np(1 - p)$ one can make the rough approximation for large n of assuming that the Hamming weight of the host is fixed and equal to the mean, i.e., $\omega(\mathbf{x}) = t$. We will also assume that the distortion constraint is deterministic, i.e., $(1/n)\delta(\mathbf{x}, \mathbf{y}) \leq \Delta$ where $\Delta \leq 1/2$, rather than it being stated on average. If we define $d \triangleq n\Delta$ (which we assume integer), this means that the encoder can avail of a budget of at most d changes to the host $\mathbf{x}$ in order to generate $\mathbf{y} \in \{0, 1\}^n$ while complying with the distortion constraint.

Firstly observe that the number of codewords with Hamming weight $\omega(\mathbf{y}) = v$ is $\binom{n}{v}$. Therefore, the number of codewords that comply with the distortion constraint can be represented by means of $b_y = \log \sum_{v=t-d}^{\min\{t+d, n\}} \binom{n}{v}$ bits. Taking into account the range of v and $t \geq n/2$, we have that $\binom{n}{v}$ peaks when $v = \max\{t - d, n/2\}$ (assuming that n

is even). Hence, a trivial upper bound on b_y is

$$b_y < \log \left((2d + 1) \binom{n}{\max\{t - d, n/2\}} \right). \quad (1)$$

We have not yet incorporated the perfect reversibility constraint to our discussion. Since the Hamming weight of the host is $\omega(\mathbf{x}) = t$ there are $\binom{n}{t}$ possible hosts. This implies that the encoder must reserve at least $b_x = \log \binom{n}{t}$ bits out of b_y in order to be able to exactly describe the host to the decoder. Hence, at most $b_y - b_x$ bits can be used to convey the embedded message. Using now the upper bound (1), we find that the embedding rate $\rho \triangleq (1/n)(b_y - b_x)$ (message bits/host element) for reversible binary watermarking is bounded from above as $\rho < (1/n) \left\{ \log \binom{n}{\max\{t - d, n/2\}} - \log \binom{n}{t} \right\} + \epsilon_n$, where $\epsilon_n \triangleq (1/n) \log(2n\Delta + 1)$. The rate on the right hand side of this inequality is not achievable in a finite context. However as $n \rightarrow \infty$ we have that $\epsilon_n \rightarrow 0$, whereas the asymptotic upper bound

$$\rho^* \triangleq \frac{1}{n} \left\{ \log \binom{n}{\max\{t - d, n/2\}} - \log \binom{n}{t} \right\} \quad (2)$$

is in fact the asymptotically optimum embedding rate, because it can be achieved simply by using an encoding scheme comprising all codewords with *fixed* weight $\omega(\mathbf{y}) = \max\{t - d, n/2\}$. Observe that in this scheme the encoder only flips nonzero elements of $\mathbf{x}$: therefore for all $i \in \{1, 2, \dots, n\}$ such that $x_i = 0$ it must hold that $y_i = x_i$. In other words, and paraphrasing the well-known Costa result [12]: in asymptotically optimum binary reversible watermarking the encoder must write not only on dirty paper, but exclusively on the dirtiest parts of the paper.

We consider next the relationship of this approximate result with the information theoretical analysis by Kalker and Willems [4]. Applying Stirling's approximation for the factorial $z! \sim \sqrt{2\pi} z^{z+\frac{1}{2}} e^{-z}$ [13] to the binomial coefficients in (2), it is possible to see that, for large n , the asymptotically optimum rate can be approximated as

$$\rho^* \approx H_B(\max\{p - \Delta, 1/2\}) - H_B(p), \quad (3)$$

where $H_B(p)$ is the entropy of a Bernoulli random variable with parameter p . Therefore ρ^* also approximates the exact rate-distortion result originally proved in [4] for binary reversible watermarking (cf. right-hand side of expression (3) with Theorem 2 in [4]). Although the road that we have travelled to get to (3) is narrower (and bumpier) than the one followed in [4], we will see next that it can afford some new vistas to the practically-minded traveller.

IV. IMPLEMENTATION OF BINARY REVERSIBLE WATERMARKING USING PERMUTATION CODING

The combinatorial approximation of the asymptotically optimum rate-distortion function (2) only involves permutations of two multisets: the one formed by the elements of the binary n -vectors with $\omega(\mathbf{y}) = v^* \triangleq \max\{t - d, n/2\}$ and the one formed by the elements of the binary n -vectors with $\omega(\mathbf{x}) = t$. This immediately begs the question of whether permutation

coding, which can encode all unique permutations (rearrangements) of a multiset into unique indices, can be an avenue for implementing near-optimum reversible watermarking in the binary Hamming case. The reader is referred to [9] for the details of a practical implementation of permutation coding.

Let us first consider what kind of scheme would be required. In order to implement reversible watermarking based on an algorithmic reading of expression (2) one needs to find a surjective reconstruction function $r(\cdot) : \mathcal{Y} \rightarrow \mathcal{X}$ whose domain is $\mathcal{Y} = \{\mathbf{y} \in \{0, 1\}^n \mid \omega(\mathbf{y}) = v^*\}$ and codomain $\mathcal{X} = \{\mathbf{x} \in \{0, 1\}^n \mid \omega(\mathbf{x}) = t\}$, i.e., all rearrangements of an arbitrary binary n -vector $\mathbf{y}_0$ with weight v^* and all rearrangements of an arbitrary binary n -vector $\mathbf{x}_0$ with weight t , respectively. If we define the sets $\mathcal{W}_{\mathbf{x}} \triangleq \{\mathbf{y} \in \mathcal{Y} \mid r(\mathbf{y}) = \mathbf{x}\}$, then $r(\cdot)$ must be such that

- 1) $|\mathcal{W}_{\mathbf{x}}| = 2^{\lfloor \rho^* n \rfloor} \approx \binom{n}{v^*} / \binom{n}{t}$ for all $\mathbf{x} \in \mathcal{X}$, so that the optimum rate is approached. Labels from $\{1, 2, \dots, 2^{\lfloor \rho^* n \rfloor}\}$ must be assigned to the elements of $\mathcal{W}_{\mathbf{x}}$ in order to encode and decode the messages.
- 2) For every $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{W}_{\mathbf{x}}$, $\delta(\mathbf{y}, \mathbf{x}) \leq d$.

Therefore if the encoder produces $\mathbf{y} = e(\mathbf{x}, m) \in \mathcal{W}_{\mathbf{x}}$ then the decoder first recovers $\mathbf{x} = r(\mathbf{y})$ (perfect reversibility), which determines $\mathcal{W}_{\mathbf{x}}$ and thus allows decoding of the message m represented by $\mathbf{y}$.

The broad lines of the coding scheme are thus clear: reversibility at the decoder is guaranteed through a quantization-like function whose domain is always larger than its codomain, which in turn coincides with the original domain of the host. Messages are signalled through points in the domain within the Voronoi cell around a quantization centroid. However the devil is in the detail. To start with, the obvious way to implement the aforementioned scheme using permutation coding would require to evenly divide $\mathcal{Y}$ into $2^{\lfloor \rho^* n \rfloor}$ classes of vectors. This can be easily accomplished through permutation decoding, for example by reserving the least significant bits of the index representing each rearrangement of $\mathbf{y}_0$ for encoding the message carried by that rearrangement. Then we would need to find a rearrangement of $\mathbf{x}_0$ closest to each rearrangement of $\mathbf{y}_0$: this can be optimally determined through permutation decoding (see decoding algorithm in [14], taking into account that in the binary Hamming scenario $\delta(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$). However two main hurdles stand in the way of this approach: 1) there are $\binom{n-v^*}{t-v^*}$ equally valid centroids per rearrangement of $\mathbf{y}_0$; as permutation decoding relies on sorting [14], the large number of ties makes it unclear how to systematically determine the $\mathcal{W}_{\mathbf{x}}$ sets; and 2) most importantly, this approach leaves us in the dark about how to implement the encoder.

A. Near-Optimum Reversible Algorithm

In order to propose a workable algorithm we need to decrease the complexity of the problem. To this end, consider the cancellation identity [15] (also known as subset-of-a-subset identity) for the product of two binomial identities, which is

$$\binom{n}{t} \binom{t}{v^*} = \binom{n}{v^*} \binom{n-v^*}{t-v^*}, \quad (4)$$

for $0 \leq v^* \leq t \leq n$. The choice of variables that we have made in (4) is not arbitrary, since by taking logarithms in this identity and noting that $\binom{t}{v^*} = \binom{t}{t-v^*}$ it is straightforward to see that the rate in (2) can alternatively be written as

$$\rho^* = \frac{1}{n} \left\{ \log \binom{t}{t-v^*} - \log \binom{n-v^*}{t-v^*} \right\}. \quad (5)$$

This expression gives us a dual view of the combinatorial reasoning in Section III: it tells us that the number of embeddable messages can also be obtained by dividing the number ways in which the encoder can change $t - v^*$ elements from the t nonzero elements of the host by the number of ways in which the decoder can change $t - v^*$ elements from the $n - v^*$ zero elements of the received codeword (in order to recover the original host). This dual view is interesting because of the following reason: since $t - v^* \leq d$, in the small distortion regime the binomial coefficients in (5) can be much smaller than the ones in (2).

We can exploit the previous observations as follows. Firstly, observe that the decoder knows that if $y_i = 1$ then $x_i = 1$. So it only needs $\log \binom{n-v^*}{t-v^*}$ bits to find the centroid $\mathbf{x}$ uniquely associated to $\mathbf{y}$. According to (5), each of the $\binom{n-v^*}{t-v^*}$ different $(n-v^*)$ -subvectors may be represented by $t-v^*$ changes to the all-ones t -subvector in the original host in as many different ways as messages that can be transmitted, and so the encoder can exploit this fact. The main issue is that the decoder cannot know the t -subvector with Hamming weight $t - v^*$, and so it must determine the missing information solely from $\mathbf{y}$.

This suggests the following encoding algorithm:

- The encoder sequentially finds the $\binom{t}{t-v^*}$ possible modifications to the t -subvector of ones in $\mathbf{x}$ using permutation decoding.
- It then performs permutation coding of each resulting $\mathbf{y}$. If the permutation decoding of an $(n-t)$ -vector with Hamming weight $t - v^*$ using the least significant $\log \binom{n-t}{t-v^*}$ bits leads to the same $(n-t)$ -subvector in $\mathbf{x}$, then it marks $\mathbf{y}$ as an element of $\mathcal{W}_{\mathbf{x}}$.
- The encoder finally chooses the $\mathbf{y} \in \mathcal{W}_{\mathbf{x}}$ that represents the message that it wishes to encode.

The decoder performs permutation coding of $\mathbf{y}$ and obtains the $(n-t)$ -vector that allows perfect reconstruction of $\mathbf{x}$ through permutation decoding of the $\log \binom{n-t}{t-v^*}$ least significant bits of the encoding. Then, its knowledge of $\mathbf{x}$ allows it to retrace the steps of the encoder in order to obtain $\mathcal{W}_{\mathbf{x}}$ and determine what message is encoded by $\mathbf{y}$.

Remarks. Although not a problem for encoder-decoder agreement, the cardinality of the $\mathcal{W}_{\mathbf{x}}$ sets resulting from the algorithm is only approximately equal to $\binom{n}{v^*} / \binom{n}{t}$, and thus not all hosts can embed the exact same amount of information. Most importantly, the algorithm is only feasible in the small distortion regime where all $\binom{t}{t-v^*}$ rearrangements can be generated by encoder and decoder; however, this distortion regime is usually the one of interest in data hiding. Finally the decoder must know t and d .

In Figure 1 we compare the algorithm described in this section against the theoretical optimum on the right-hand side

Fig. 1. Empirical rate-distortion performance of near-optimum reversible embedding based on permutation coding (red squares) versus theoretical limits (blue solid lines)

of (3) for two hosts with length $n = 100$ and Hamming weights $t = 90$ and $t = 70$, respectively.

V. A BOUND ON THE GENERAL RATE-DISTORTION FUNCTION OF THE PROBLEM

In this final section we leave aside the special binary reversible watermarking problem and the combinatorial considerations that have kept us busy until now, and we briefly focus our attention on the general case. Our goal here will be the practical evaluation of the general capacity formula for reversible watermarking in a noiseless channel, which is [4]

$$\rho^* = \max_{E\{s(X,Y)\} \leq \Delta} H(Y) - H(X). \quad (6)$$

In spite of this rate-distortion expression having been available for over ten years, to the best of our knowledge no nonbinary reversible watermarking algorithm has ever been compared against it. This is most likely due to the difficulty in evaluating or estimating the first entropy in (6), which requires a distribution that is unavailable in practice. On the other hand the second entropy in (6) can always be approximated through the best compression rate attainable for the host, even if this is only empirically obtained.

An alternative is to exploit the chain rule of the entropy $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ to rewrite (6) in a more convenient way. As with perfect reversibility one must have that $H(X|Y) = 0$, if we define the watermark random variable $W \triangleq Y - X$ and use the quadratic distortion $s(X,Y) = (Y - X)^2$ we can put (6) as

$$\rho^* = \max_{E\{s(X,Y)\} \leq \Delta} H(Y|X) = \max_{E\{W^2\} \leq \Delta} H(W). \quad (7)$$

Using now the differential entropy bound on discrete entropy—independently found by Djakov [16], Massey [17] and Willems (unpublished, see [18, Problem 8.7])—to bound $H(W)$ from above, we can write the following upper bound on the rate-distortion function of reversible embedding:

$$\rho^* < \rho_u \triangleq \frac{1}{2} \log \left(2\pi e \left(\Delta + \frac{1}{12} \right) \right). \quad (8)$$

This bound is loose for $\Delta \leq 1/12$, which is not important in practical terms since Δ is much higher for typical peak-signal-to-noise ratios (PSNR).

VI. CONCLUSION

In this article we have mainly investigated the connection of combinatorics and permutation coding with binary reversible data hiding, which open novel theoretical and practical avenues for the problem. The initial results are promising, as they allow for the implementation of a near-optimum reversible algorithm in the low distortion regime—usually the most relevant one in data hiding. Further work is needed in order to find a low-complexity algorithm able to work with arbitrary distortion.

ACKNOWLEDGEMENT

This work has been supported by UCD Seed Funding under award SF1130.

REFERENCES

- [1] M. Barton, "Method and apparatus for embedding authentication information within digital data," U.S. Patent 5,646,997, July 1997.
- [2] F. Mintzer, J. Lotspiech, and N. Morimoto, "Safeguarding digital library contents and users: Digital watermarking," D-Lib Magazine, December 1997.
- [3] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding for all image formats," in *Proc. SPIE: Security and Watermarking of Multimedia Contents*, San Jose, USA, 2002, vol. 4675, pp. 572–583.
- [4] T. Kalker and F. M. J. Willems, "Capacity bounds and constructions for reversible data-hiding," in *14th Int. Conf. on Digital Signal Processing (DSP)*, Santorini, Greece, July 2002, vol. 1, pp. 71–76.
- [5] F. M. J. Willems and T. Kalker, "Reversible embedding methods," in *Proc. of 40th Annual Allerton Conf. on Communication, Control and Computing*, Monticello, USA, October 2002.
- [6] F. M. J. Willems and T. Kalker, "Coding theorems for reversible embedding," in *DIMACS Workshop Network Information Theory*, Piscataway, USA, March 2003, vol. 66 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pp. 61–78.
- [7] W. Zhang, B. Chen, and N. Yu, "Capacity-approaching codes for reversible data hiding," in *Proc. of the 13th Information Hiding Workshop*, Springer-Verlag, 2011, vol. 6958 of *LNCS*, pp. 255–269.
- [8] W. Zhang, B. Chen, and N. Yu, "Improving various reversible data hiding schemes via optimal codes for binary covers," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 2991–3003, 2012.
- [9] F. Balado and D. Haughton, "Optimum perfect universal steganography of finite memoryless sources," 2014, arXiv:1410.2659.
- [10] F. Balado, "The role of permutation coding in minimum-distortion perfect counterforensics," in *39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 6240–6244.
- [11] R. P. Feynman, *Feynman Lectures on Computation*, Addison-Wesley Longman Inc., Boston, MA, USA, 1998.
- [12] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [13] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, John Wiley & Sons, 3rd edition, 1968.
- [14] D. Slepian, "Permutation modulation," *Proc. of the IEEE*, vol. 53, no. 3, pp. 228–236, 1965.
- [15] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff, *Combinatorics and Graph Theory*, Springer, 2nd edition, 2008.
- [16] A. G. Djakov, "On a search model of false coins," in *Topics in Information Theory*, I. Csiszár and P. Elias, Eds., vol. 16 of *Colloquia Mathematica Societatis János Bolyai*, pp. 163–170. North Holland Publishing Co., Budapest, Hungary, 1977.
- [17] J. L. Massey, "On the entropy of integer-valued random variables," in *Proc. of Int. Workshop on Information Theory*, Beijing, China, July 1988, pp. C1.2–C1.4.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2nd edition, 2006.