Polarization and Polar Codes

Polarization and Polar Codes

Eren Şaşoğlu

University of California, San Diego
La Jolla, CA 92093-0436
USA
esasoglu@ucsd.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is E. Şaşoğlu, Polarization and Polar Codes, Foundations and Trends® in Communications and Information Theory, vol 8, no 4, pp 259–381, 2011

ISBN: 978-1-60198-596-5

Foundations and Trends® in Communications and Information Theory

Volume 8 Issue 4, 2011

Editorial Board

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2011, Volume 8, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Foundations and Trends® in
Communications and Information Theory
Vol. 8, No. 4 (2011) 259–381

DOI: 10.1561/0100000041

Polarization and Polar Codes

Eren Şaşoğlu

University of California, San Diego, 9500 Gilman Drive #0436, La Jolla, CA 92093-0436, USA, esasoglu@ucsd.edu

Abstract

This tutorial treats the fundamentals of polarization theory and polar coding. Arıkan's original results on binary source and channel polarization methods are studied. Error probability and complexity analyses are offered. The original results are generalized in several directions. Early developments in the field are discussed, pointers to some of the important work omitted from this tutorial are given.

Contents

1

Introduction

Figure 1.1 depicts the setting for the fundamental problem in communication theory. A sender has K bits of information to send, which, after appropriate processing, are transmitted through a noisy channel that accepts input symbols one at a time and produces a sequence of output symbols. The task of the communication engineer is to design an encoding/decoding scheme that ensures that the K bits are (i) transmitted in as few uses of the channel as possible, and (ii) correctly reproduced at the receiver with as high a probability as desired. In [42], Shannon showed that these seemingly conflicting requirements can be met simultaneously so long as K and N (the number of channel uses) are large and K/N (called the rate of transmission) is below the *capacity* of the channel.

Shannon's proof of the channel coding theorem shows not only that reliable communication at rates below capacity is possible, but also that almost all encoding schemes, i.e., channel codes, with rates below

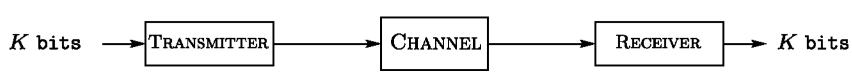

Fig. 1.1

channel capacity will perform well as long as optimal decoders are used at the receiver. Unfortunately, optimal decoding is in general prohibitively difficult — its complexity grows exponentially in the coding length — and how to construct practical coding schemes, and especially low-complexity decoders, is not immediately clear from Shannon's coding theorem alone.

Significant progress has been made in the past sixty years toward developing practical and capacity-achieving coding methods. The bulk of the research effort to this end can be broadly divided into two groups: algebraic coding and iterative coding. Research in algebraic coding was motivated primarly by the recognition that for channels of practical interest, the words of a code must be as different from each other as possible in order to ensure their distinguishability at the receiver. Iterative codes (e.g., Turbo codes and LDPC codes), on the other hand, are designed to work well with a low-complexity decoding algorithm. Despite remarkable advances in both fields, especially in iterative coding, finding codes that (i) operate at rates close to capacity, (ii) have low computational complexity, and (iii) have provable reliability guarantees was an elusive goal until recently.[1]

Polar codes, invented recently by Arıkan [4], have all of these desirable properties. In particular,

- they achieve the symmetric capacity of all binary-input memoryless channels. Consequently, they are capacity-achieving for symmetric channels, which include several channel classes of practical relevance such as the binary-input additive white Gaussian noise channel, the binary symmetric channel, and the binary erasure channel.
- they are low-complexity codes, and therefore are practical: the time and space complexities of the encoding/decoding algorithms Arıkan proposes in [4] are $O(N \log N)$, where N is the blocklength.
- the block error probability of polar codes is roughly $O(2^{-\sqrt{N}})$ [9]. This performance guarantee is analytical, and is not only based on empirical evidence.

[1] See [12] for a historical account of the development of coding theory in general.

- for symmetric channels, polar code construction is deterministic. That is, the above statements are true not only for ensembles of codes, but also for individual polar codes. Further, construction of polar codes can be accomplished with time complexity $O(N)$ and space complexity $O(\log N)$ [45].

The design philosophy of polar codes is fundamentally different from those of both algebraic codes and iterative codes (although the codes themselves are closely related to the algebraic Reed–Muller codes). It is interesting to note that the invention of these codes is the culmination of Arıkan's efforts to improve the rates achievable by convolutional codes and *sequential decoding* [6], a decoding method developed in the late 1950s.

The technique underlying polar codes is 'channel polarization': creating extremal channels — those that are either noiseless or useless — from mediocre ones. Soon after the publication of [4], Arıkan showed that a similar technique can be used to construct optimal source codes [5] — he calls this technique 'source polarization'. It is clear in his work that a single *polarization* principle underlies both techniques; channel polarization and source polarization are specific applications of this principle.

1.1 Extremal Distributions and Polarization

Suppose we are interested in guessing (i.e., decoding) the value of a binary N-vector U_1^N after observing a related random vector Y_1^N. Here, U_1^N may represent a codeword chosen randomly from a channel code, and Y_1^N the output of a channel when U_1^N is the input. Alternatively, U_1^N may be viewed as the output of a random source, and Y_1^N as side information about U_1^N. In order to minimize the probability of decoding error, one chooses the value of U_1^N that maximizes[2]

$$p(u_1^N \mid y_1^N) = \prod_{i=1}^{N} p(u_i \mid y_1^N, u_1^{i-1}).$$

[2] Throughout, we will denote probability distributions by p as long as their arguments are lower case versions of the random variables they represent. For example, we will write $p(x,y \mid z)$ for $p_{XY|Z}(x,y \mid z)$, denoting the joint distribution of X and Y conditioned on Z.

There are two extremal cases in terms of the probability of decoding error. First, if U_1^N is a function of Y_1^N — i.e., if the above probability is either 0 or 1 — then its value can always be guessed correctly. Second, if U_1^N is independent of Y_1^N and uniformly distributed, then all guesses are equally good and will be correct with probability $1/2^N$. The first of these cases is trivial provided that the function computations can be done easily, and the second is hopeless.

A more interesting extremal case is one in which the conditional distribution of U_1^N is neither $\{0,1\}$-valued nor uniform, but it is *polarized* in the sense that all distributions in the product formula above are either $\{0,1\}$-valued or uniform. One can view this as a case where all randomness in U_1^N is concentrated in a subset of its components. Clearly, one cannot in general correctly decode such a random vector with high probability. On the other hand, decoding U_1^N again becomes trivial if one has prior knowledge of its random component. The polarized structure in the probability distribution even suggests that U_1^N can be decoded *successively*: suppose, for the sake of argument, that the odd-numbered factors in the product formula above are $\{0,1\}$-valued distributions whereas the even-numbered factors are uniform. Then, if one has prior knowledge of the even indices of U_1^N, then the odd indices can be determined in increasing order as follows. The decoder first computes U_1 as a function of Y_1^N, then produces U_2 (which is already available to it) then uses its knowledge of U_1 and U_2 to compute U_3 as a function of (Y_1^N, U_1^2), etc.

A realistic model of the input/output process of a noisy channel or the output/side information process of a data source rarely fits this description. On the other hand, one may attempt to transform the process in question into one that does fit it. This is precisely the aim of Arıkan's polarization technique. In its original form, this technique consists in combining two identically distributed binary random variables so as to create two disparate random variables and repeating this operation several times to amplify the disparity, eventually approaching a polarized set of random variables. We will see this technique along with how to apply it to channel and source coding in Section 2. In Section 3 we will review the complexity of polar encoding, decoding, and code construction. As we have already mentioned, the practical appeal of polar codes is due to the low complexity requirements of these tasks along with provable reliability guarantees.

There has been considerable amount of research effort in polarization theory and polar coding since the publication of [4] in 2009. Arguably the main reason for this interest is the technique's ease of applicability to settings other than binary source and channel coding. In the rest of this monograph (Sections 4–6), we will review some of the main generalizations of the theory. We will begin in Section 4 by studying how discrete memoryless processes of arbitrary alphabet sizes, not just binary ones, can be polarized by recursive transforms. We will see that this can be accomplished through a linear transform similar to Arıkan's when the alphabet size is prime. Interestingly, linear transforms lose their ability to polarize *all* stationary memoryless processes when the underlying alphabet size is not a prime number. There are, however, non-linear transforms that do polarize all stationary memoryless processes for all finite alphabet sizes. In Section 4.2 we will study sufficient conditions for a recursive transform to polarize all such processes, and give an example of a family of transforms that satisfy these conditions for all finite alphabet sizes. The complexity and the error probability behavior of codes obtained by such transforms will be as in the binary case.

While the error probability guarantees of polar codes are unprecedented, it is of interest to know whether even stronger codes can be obtained by combining more than two random variables in each recursion of a polarizing construction. This study is undertaken in Section 5: we will first show that a large class of recursive linear transforms that combine several random variables at a time polarize memoryless processes with prime alphabet sizes. We will then characterize how a single recursion of a given polarizing transform affects error probability behavior, from which results on the large-blocklength behavior follow easily. The implications of this characterization are of a mixed nature: while in the binary case one cannot improve on the $O(2^{-\sqrt{N}})$ error probability decay by combining a small number of random variables at a time, strong improvements become possible as the alphabet size grows.

In Section 6, we will make use of the polarization theorems of earlier sections to study *joint* polarization of multiple processes. We will see that recursive transforms, applied separately to multiple processes, not only polarize the individual processes, but the correlations between the processes are also polarized. These results will immediately lead to polar coding theorems for two-user settings such as the separate encoding of correlated sources and the multiple-access channel.

References

[1] E. Abbe, "Randomness and dependencies extraction via polarization," in *Proceedings of the Information Theory and Applications Workshop*, Feb. 2011.

[2] E. Abbe and E. Telatar, "Polar codes for the m-user MAC," [Online] Available: arXiv:1002.0777, August 2010.

[3] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Communications Letters*, vol. 14, no. 8, pp. 752–754, August 2010.

[4] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[5] E. Arıkan, "Source polarization," in *Proceedings of the International Symposium on Information Theory*, pp. 899–903, 13–18 June 2010.

[6] E. Arıkan, "A survey of Reed–Muller codes from polar coding perspective," in *Proceedings of the Information Theory Workshop*, 1–5, Jan. 2010.

[7] E. Arıkan, "Systematic polar coding," *IEEE Communications Letters*, vol. 15, no. 8, pp. 860–862, August 2011.

[8] E. Arıkan, "Polar coding for the Slepian–Wolf problem based on monotone chain rules," in *Proceedings of the International Symposium on Information Theory*, pp. 566–570, July 2012.

[9] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proceedings of the International Symposium on Information Theory*, pp. 1493–1495.

[10] J. A. Bondy and U. S. R. Murty, *Graph Theory*. New York: Springer, 2008.

[11] A. Clark, *Elements of Abstract Algebra*. New York: Dover, 1971.

[12] D. J. Costello Jr. and G. D. Forney Jr., "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, June 2007.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[14] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[15] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for the deterministic broadcast channel," in *Proceedings of the International Zurich Seminars on Communications*, pp. 51–54, Feb–Mar 2012.

[16] A. Goli, H. Hassani, and R. Urbanke, "Universal bounds on the scaling behavior of polar codes," in *Proceedings of the International Symposium on Information Theory*, July 2012.

[17] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Rate-splitting multiple access for discrete memoryless channel," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 873–890, March 2001.

[18] H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proceedings of the Annual Allerton Conference on Communications, Control, and Computing*, pp. 16–21, Sept–Oct 2009.

[19] H. Hassani, R. Mori, T. Tanaka, and R. Urbanke, "Rate-dependent analysis of the asymptotic behavior of channel polarization," submitted to *IEEE Transactions on Information Theory*, [Online] Available: arXiv:1110.0194, Oct 2011.

[20] H. Hassani and R. Urbanke, "On the scaling of polar codes: I. The behavior of polarized channels," in *Proceedings of the International Symposium on Information Theory*, pp. 874–878, June 2010.

[21] E. Hof and S. Shamai, "Secrecy-achieving polar coding," in *Proceedings of the Information Theory Workshop*, pp. 1–5, August 2010.

[22] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proceedings of the International Symposium on Information Theory*, pp. 1488–1492, July 2009.

[23] M. Karzand, "Polar codes for degraded relay channels," in *Proceedings of the International Zurich Seminar on Communications*, pp. 59–62, Feb–Mar 2012.

[24] M. Karzand and E. Telatar, "Polar codes for q-ary source coding," in *Proceedings of the International Symposium on Information Theory*, pp. 909–912, June 2010.

[25] S. B. Korada, "Polar codes for channel and source coding," PhD Dissertation, EPFL 2009.

[26] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6253–6264, Dec 2010.

[27] S. B. Korada, A. Montanari, E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proceedings of the International Symposium on Information Theory*, pp. 884–888, June 2010.

[28] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.

[29] O. O. Koyluoglu and H. El Gamal, "Polar coding for secure transmission and key agreement," *International Symposium on Pers. Ind. Mob. Radio Comm.*, pp. 2698–2703, Sep 2010.

[30] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, "Hardware architectures for successive cancellation decoding of polar codes," *International Conference on Acou., Sp., and Sig. Proc.*, pp. 1665–1668, May 2011.

[31] F. J. MacWilliams and N. J. Sloane, *The Theory of Error-Correcting Codes*. North-Holland, 1977.

[32] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6428–6443, Oct 2011.

[33] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proceedings of the International Symposium on Information Theory*, pp. 1496–1500, July 2009.

[34] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," in *Proceedings of the International Symposium on Information Theory*, pp. 894–898, June 2010.

[35] W. Park and A. Barg, "Multilevel polarization for nonbinary codes and parallel channels," in *Proceedings of the Annual Allerton Conference on Communications, Control, and Computing*, pp. 228–234, Sep 2011.

[36] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," in *Proceedings of the International Symposium on Information Theory*, pp. 11–15, Aug 2011.

[37] E. Şaşoğlu, "An entropy inequality for q-ary random variables and its application to channel polarization," in *Proceedings of the International Symposium on Information Theory*, pp. 1360–1363, June 2010.

[38] E. Şaşoğlu, "Polarization in the presence of memory," in *Proceedings of the International Symposium on Information Theory*, pp. 189–193, June 2010.

[39] E. Şaşoğlu, "Polar coding theorems for discrete systems," PhD Dissertation, EPFL 2011.

[40] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proceedings of the Information Theory Workshop*, pp. 144–148, Oct 2009.

[41] E. Şaşoğlu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," [Online]. Available: arXiv:1006.4255, June 2010.

[42] C. E. Shannon, "A mathematical theory of communication," *Bell System Technology Journal*, vol. 27, pp. 379–423, 623–565, July and October 1948.

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

[44] I. Tal and A. Vardy, "List decoding of polar codes," in *International Symposium on Information Theory*, pp. 1–5, Aug 2011.

[45] I. Tal and A. Vardy, "How to construct polar codes," [Online]. Available: arXiv:1105.6164, May 2011.

[46] T. Tanaka and R. Mori, "Refined rate of channel polarization," in *International Symposium on Information Theory*, pp. 889–893, June 2010.

[47] E. Telatar, private communication.

[48] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769–772, Nov 1973.