

# An Upgrading Algorithm with Optimal Power Law

Or Ordentlich, Ido Tal

**Abstract**—Consider a channel  $W$  along with a given input distribution  $P_X$ . In certain settings, such as in the construction of polar codes, the output alphabet of  $W$  is often ‘too large’, and hence we replace  $W$  by a channel  $Q$  having a smaller output alphabet. We say that  $Q$  is upgraded with respect to  $W$  if  $W$  is obtained from  $Q$  by processing its output. In this case, the mutual information  $I(P_X, W)$  between the input and output of  $W$  is upper-bounded by the mutual information  $I(P_X, Q)$  between the input and output of  $Q$ . In this paper, we present an algorithm that produces an upgraded channel  $Q$  from  $W$ , as a function of  $P_X$  and the required output alphabet size of  $Q$ , denoted  $L$ . We show that the difference in mutual informations is ‘small’. Namely, it is  $O(L^{-2/(|\mathcal{X}|-1)})$ , where  $|\mathcal{X}|$  is the size of the input alphabet. This power law of  $L$  is optimal.

## I. INTRODUCTION

In his seminal paper on polar codes, Arikan introduced synthetic channels [1, equation (5)], also called bit-channels. These synthetic channels have a binary input alphabet and an intractably large output alphabet. Namely, the output alphabet size of such a channel is at least  $2^N$ , where  $N$  is the length of the polar code. When *constructing* a polar code, the vast size of the output alphabet is very much an issue. We note that in many settings more general than the seminal one, we search for channels that are ‘very noisy’. A crucial observation is that instead of considering the original synthetic channel, one may approximate it by another channel having a much smaller output alphabet size [2]. Specifically, if the approximating channel is upgraded with respect to the original channel and shown to be ‘very noisy’, then this must also be the case for the original channel.

## II. SETTING

We are given a channel  $W : \mathcal{X} \rightarrow \mathcal{Y}$  along with an input distribution  $P_X$ . We denote the mutual information between the input and output of  $W$  as  $I(P_X, W) \triangleq I(X; Y)$ , where  $X$  and  $Y$  are random variables with joint distribution

$$P_{X,Y}(x, y) = P_X(x)W(y|x). \quad (1)$$

Let  $Q : \mathcal{X} \rightarrow \mathcal{Z}$  be a channel with the same input alphabet as  $W : \mathcal{X} \rightarrow \mathcal{Y}$ . We say that  $Q$  is *upgraded* with respect to  $W$  if we can obtain  $W$  by processing the output of  $Q$ . That is, if there exists a third channel  $\Phi : \mathcal{Z} \rightarrow \mathcal{Y}$  such that, for every  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ ,

$$W(y|x) = \sum_{z \in \mathcal{Z}} Q(z|x)\Phi(y|z).$$

Put another way, we want  $X$ ,  $Z$ , and  $Y$  to form a Markov chain, in that order.

O. Ordentlich is with the school of computer science and engineering, Hebrew University of Jerusalem, Israel (email: or.ordentlich@mail.huji.ac.il).

I. Tal is with the Department of Electrical Engineering, Technion, Haifa 32000, Israel (email: idotal@ee.technion.ac.il).

Our goal in this paper, given a fixed input alphabet size  $|\mathcal{X}|$ , an input distribution  $P_X$ , a channel  $W : \mathcal{X} \rightarrow \mathcal{Y}$ , and a parameter  $L$ , is to construct a channel  $Q : \mathcal{X} \rightarrow \mathcal{Z}$  that is upgraded with respect to  $W$  and whose output alphabet size satisfies  $|\mathcal{Z}| \leq L$ . Our method produces such a  $Q$  for which

$$I(P_X, Q) - I(P_X, W) = O(L^{-2/(|\mathcal{X}|-1)}). \quad (2)$$

By [3, Section IV], the above power law of  $L$  is optimal.

## III. THE ALGORITHM

Similarly to the method in [4], we use the ‘one-hot’ representation of  $x \in \mathcal{X}$  to affect a reduction from the non-binary alphabet  $\mathcal{X}$  to the binary alphabet  $\mathcal{X}'$ . Namely, w.l.o.g. let us assume that  $\mathcal{X} = \{1, 2, \dots, q\}$ . We will replace  $x \in \mathcal{X}$  by a length  $q-1$  vector  $f(x) = (x_1, x_2, \dots, x_{q-1})$ , such that

$$x_i = \begin{cases} 1 & \text{if } x = i \\ 0 & \text{otherwise} \end{cases}$$

For each  $1 \leq i \leq q-1$ , we apply the binary-input upgrading algorithm in [3, Section VI] to  $P_{X_i Y | X_1^{i-1} = 0_1^{i-1}}$ , and require that the output alphabet size of the upgrading channel satisfy  $|\mathcal{Z}^{(i)}| \leq \Lambda$ , where  $\Lambda = \lfloor L^{1/(q-1)} \rfloor$ . Denote the resulting joint distribution  $\beta_{X_i, Z_i, Y}^{(i)}(x', z', y)$ . From these  $q-1$  distributions, we define our final distribution on  $(X, Z, Y)$ . The output alphabet is

$$\mathcal{Z} = \mathcal{Z}^{(1)} \times \mathcal{Z}^{(2)} \times \dots \times \mathcal{Z}^{(q-1)},$$

and the joint distribution is

$$P_{X,Z,Y}^*(x, z, y) = P_Y(y) \cdot \left( \prod_{i=1}^{q-1} \beta_{Z_i|Y}^{(i)}(z_i|y) \right) \cdot \left( \prod_{i=1}^{q-1} \gamma_{X_i|Z_i, X_1^{i-1}}^{(i)}(x_i|z_i, x_1^{i-1}) \right), \quad (3)$$

where, for  $1 \leq i \leq q$ ,

$$\gamma_{X_i|Z_i, X_1^{i-1}}^{(i)}(x_i|z_i, x_1^{i-1}) = \begin{cases} \beta_{X_i|Z_i}^{(i)}(x_i|z_i) & \text{if } x_1^{i-1} = 0_1^{i-1}, \\ 1 & \text{if } x_1^{i-1} \neq 0_1^{i-1} \text{ and } x_i = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

## REFERENCES

- [1] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [2] R. Mori and T. Tanaka, “Performance and construction of polar codes on symmetric binary-input memoryless channels,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2009)*, Seoul, South Korea, 2009, pp. 1496–1500.
- [3] A. Kartowsky and I. Tal, “Greedy-merge degrading has optimal power-law,” *IEEE Trans. Inform. Theory*, vol. 65, no. 2, pp. 917–934, February 2019.
- [4] A. Bhatt, B. Nazer, O. Ordentlich, and Y. Polyanskiy, “Information-distilling quantizers,” *arXiv preprint arXiv:1812.03031*, 2018.