

# Expurgated Random-Coding Ensembles: Exponents, Refinements and Connections

Jonathan Scarlett, Li Peng, Neri Merhav, Alfonso Martinez and Albert Guillén i Fàbregas

## Abstract

This paper studies expurgated random-coding bounds and exponents with a given (possibly suboptimal) decoding rule. Variations of Gallager's analysis are presented, yielding new asymptotic and non-asymptotic bounds on the error probability for an arbitrary codeword distribution. A simple non-asymptotic bound is shown to attain an exponent which coincides with that of Csiszár and Körner for discrete alphabets, while also remaining valid for continuous alphabets. The method of type class enumeration is studied for both discrete and continuous alphabets, and it is shown that this approach yields improved exponents for some codeword distributions. A refined analysis of expurgated i.i.d. random coding is given which yields an exponent with a  $O(\frac{1}{\sqrt{n}})$  prefactor, thus improving on Gallager's  $O(1)$  prefactor.

## I. INTRODUCTION

Achievable performance bounds for channel coding are typically obtained by analyzing the average error probability of an ensemble of codebooks with independently generated codewords. For memoryless channels, random codes with independent and identically distributed (i.i.d.) symbols achieve the channel capacity [1], characterize the error exponent of the best code at sufficiently high rates [2, Ch. 5], and provide tight bounds on the finite-length performance [3].

At low rates, the error probability of the best code in the random-coding ensemble can be significantly smaller than the average. In such cases, better performance bounds are obtained by considering an ensemble in which a subset of the randomly generated codewords are expurgated from the codebook. In particular, the error exponents resulting from such techniques are generally higher than the random-coding error exponent at low rates. Existing works exploring such techniques include those of Gallager [2, Sec. 5.7], Csiszár-Körner-Martón [4], [5, Ex. 10.18] and Csiszár-Körner [6]. The advantages of Gallager's approach include its simplicity and the fact that the analysis is not restricted to discrete

J. Scarlett and L. Peng are with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: jmscarlett@gmail.com, lp327@cam.ac.uk). N. Merhav is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel. (e-mail: merhav@ee.technion.ac.il). A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: alfonso.martinez@ieee.org). A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03. The work of N. Merhav was partially supported by the Israeli Science Foundation (ISF), grant no. 412/12.