Statistical Physics and Information Theory
Statistical Physics and Information Theory

Neri Merhav

Department of Electrical Engineering
Technion — Israel Institute of Technology
Haifa 32000
Israel
merhav@ee.technion.ac.il

now
the essence of knowledge

Boston – Delft
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, 2009, Volume 6, 6 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.
Statistical Physics and Information Theory

Neri Merhav

Department of Electrical Engineering, Technion — Israel Institute of Technology, Haifa 32000, Israel, merhav@ee.technion.ac.il

Abstract

This monograph is based on lecture notes of a graduate course, which focuses on the relations between information theory and statistical physics. The course was delivered at the Technion during the Spring of 2010 for the first time, and its target audience consists of EE graduate students in the area of communications and information theory, as well as graduate students in Physics who have basic background in information theory. Strong emphasis is given to the analogy and parallelism between information theory and statistical physics, as well as to the insights, the analysis tools and techniques that can be borrowed from statistical physics and ‘imported’ to certain problem areas in information theory. This is a research trend that has been very active in the last few decades, and the hope is that by exposing the students to the meeting points between these two disciplines, their background and perspective may be expanded and enhanced. This monograph is substantially revised and expanded relative to an earlier version posted in arXiv (1006.1565v1 [cs.IT]).
Contents

1 Introduction 1

2 Basic Background in Statistical Physics 7
2.1 What is Statistical Physics? 7
2.2 Basic Postulates and the Microcanonical Ensemble 8
2.3 The Canonical Ensemble 17
2.4 Properties of the Partition Function and the Free Energy 21
2.5 The Energy Equipartition Theorem 32
2.6 The Grand-Canonical Ensemble 34

3 Physical Interpretations of Information Measures 39
3.1 Statistical Physics of Optimum Message Distributions 40
3.2 Large Deviations and Physics of Coding Theorems 42
3.3 Gibbs' Inequality and the Second Law 57
3.4 Boltzmann's H-Theorem and the DPT 73
3.5 Generalized Temperature and Fisher Information 88

4 Analysis Tools and Asymptotic Methods 95
4.1 Introduction 95
4.2 The Laplace Method 97
4.3 The Saddle-Point Method 101
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4 Extended Example: Capacity of a Disordered System</td>
<td>111</td>
</tr>
<tr>
<td>4.5 The Replica Method</td>
<td>116</td>
</tr>
<tr>
<td>5 Interacting Particles and Phase Transitions</td>
<td>123</td>
</tr>
<tr>
<td>5.1 Introduction — Sources of Interaction</td>
<td>124</td>
</tr>
<tr>
<td>5.2 Models of Interacting Particles</td>
<td>125</td>
</tr>
<tr>
<td>5.3 A Qualitative Discussion on Phase Transitions</td>
<td>131</td>
</tr>
<tr>
<td>5.4 Phase Transitions of the Rate–Distortion Function</td>
<td>135</td>
</tr>
<tr>
<td>5.5 The One-Dimensional Ising Model</td>
<td>139</td>
</tr>
<tr>
<td>5.6 The Curie–Weiss Model</td>
<td>142</td>
</tr>
<tr>
<td>5.7 Spin Glasses and Random Code Ensembles</td>
<td>147</td>
</tr>
<tr>
<td>6 The Random Energy Model and Random Coding</td>
<td>153</td>
</tr>
<tr>
<td>6.1 REM without a Magnetic Field</td>
<td>153</td>
</tr>
<tr>
<td>6.2 Random Code Ensembles and the REM</td>
<td>159</td>
</tr>
<tr>
<td>6.3 Random Coding Exponents</td>
<td>166</td>
</tr>
<tr>
<td>7 Extensions of the REM</td>
<td>177</td>
</tr>
<tr>
<td>7.1 REM Under Magnetic Field and Source–Channel Coding</td>
<td>178</td>
</tr>
<tr>
<td>7.2 Generalized REM (GREM) and Hierarchical Coding</td>
<td>187</td>
</tr>
<tr>
<td>7.3 Directed Polymers in a Random Medium and Tree Codes</td>
<td>198</td>
</tr>
<tr>
<td>8 Summary and Outlook</td>
<td>203</td>
</tr>
</tbody>
</table>

Acknowledgments                                                          205

References                                                               207
Introduction

This work focuses on some of the relationships and the interplay between information theory and statistical physics — a branch of physics that deals with many-particle systems using probabilistic and statistical methods in the microscopic level.

The relationships between information theory and statistical thermodynamics are by no means new, and many researchers have been exploiting them for many years. Perhaps the first relation, or analogy, that crosses one’s mind is that in both fields there is a fundamental notion of entropy. Actually, in information theory, the term entropy was coined in the footsteps of the thermodynamic entropy. The thermodynamic entropy was first introduced by Clausius in 1850, and its probabilistic-statistical interpretation was established by Boltzmann in 1872. It is virtually impossible to miss the functional resemblance between the two notions of entropy, and indeed it was recognized by Shannon and von Neumann. The well-known anecdote on this tells that von Neumann advised Shannon to adopt this term because it would provide him with “…a great edge in debates because nobody really knows what entropy is anyway.”

But the relationships between the two fields go far beyond the fact that both share the notion of entropy. In fact, these relationships have
many aspects. We will not cover all of them in this work, but just to
taste the flavor of their scope, we will mention just a few.

The maximum entropy (ME) principle. This is perhaps the oldest
concept that ties the two fields and it has attracted a great deal of
attention, not only of information theorists, but also that of researchers
in related fields like signal processing and image processing. The ME
principle evolves around a philosophy, or a belief, which, in a nutshell,
is the following: if in a certain problem, the observed data comes from
an unknown probability distribution, but we do have some knowledge
(that stems, e.g., from measurements) of certain moments of the under-
lying quantity/signal/random-variable, then assume that the unknown
underlying probability distribution is the one with maximum entropy
subject to (s.t.) moment constraints corresponding to this knowledge.
For example, if we know the first and the second moments, then the ME
distribution is Gaussian with matching first and second order moments.
Indeed, the Gaussian model is perhaps the most common model for
physical processes in information theory as well as in signal- and image
processing. But why maximum entropy? The answer to this philoso-
phical question is rooted in the second law of thermodynamics, which
asserts that in an isolated system, the entropy cannot decrease, and
hence, when the system reaches thermal equilibrium, its entropy reaches
its maximum. Of course, when it comes to problems in information the-
ory and other related fields, this principle becomes quite heuristic, and
so, one may question its justification, but nevertheless, this approach
has had an enormous impact on research trends throughout the last
50 years, after being proposed by Jaynes in the late fifties of the pre-
vious century [45, 46], and further advocated by Shore and Johnson
afterward [106]. In the book by Cover and Thomas [13, Section 12],
there is a good exposition on this topic. We will not put much emphasis
on the ME principle in this work.

Landauer’s erasure principle. Another aspect of these relations
has to do with a theory whose underlying guiding principle is that
information is a physical entity. Specifically, Landauer’s erasure
principle [63] (see also [9]), which is based on this physical theory
of information, asserts that every bit that one erases, increases the
entropy of the universe by \( k \ln 2 \), where \( k \) is Boltzmann’s constant.
The more comprehensive picture behind Landauer’s principle is that “any logically irreversible manipulation of information, such as the erasure of a bit or the merging of two computation paths, must be accompanied by a corresponding entropy increase in non-information bearing degrees of freedom of the information processing apparatus or its environment.” (see [6]). This means that each lost information bit leads to the release of an amount $kT \ln 2$ of heat. By contrast, if no information is erased, computation may, in principle, be achieved in a way which is thermodynamically a reversible process, and hence requires no release of heat. This has had a considerable impact on the study of reversible computing. Landauer’s principle is commonly accepted as a law of physics. However, there has also been some considerable dispute among physicists on this. This topic is not going to be included either in this work.

**Large deviations theory as a bridge between information theory and statistical physics.** Both information theory and statistical physics have an intimate relation to large deviations theory, a branch of probability theory which focuses on the assessment of the exponential rates of decay of probabilities of rare events, where one of the most elementary mathematical tools is the Legendre transform, which stands at the basis of the Chernoff bound. This topic will be covered quite thoroughly, mostly in Section 3.2.

**Random matrix theory.** How do the eigenvalues (or, more generally, the singular values) of random matrices behave when these matrices have very large dimensions or if they result from products of many randomly selected matrices? This is a very active area in probability theory with many applications, both in statistical physics and information theory, especially in modern theories of wireless communication (e.g., MIMO systems). This is again outside the scope of this work, but the interested reader is referred to [115] for a comprehensive introduction on the subject.

**Spin glasses and coding theory.** As was first observed by Sourlas [109] (see also [110]) and further advocated by many others, it turns out that many problems in channel coding theory (and also to some extent, source coding theory) can be mapped almost verbatim to parallel problems in the field of physics of spin glasses — amorphic
magnetic materials with a high degree of disorder and very complicated physical behavior, which is customarily treated using statistical-mechanical approaches. It has been many years that researchers have made attempts to “import” analysis techniques rooted in statistical physics of spin glasses and to apply them to analogous coding problems, with various degrees of success. This is one of main subjects of this work and we will study it extensively, at least from some aspects.

The above list of examples is by no means exhaustive. We could have gone much further and add many more examples of these very fascinating meeting points between information theory and statistical physics, but most of them will not be touched upon in this work. Many modern analyses concerning multiuser situations, such as MIMO channels, CDMA, etc., and more recently, also in compressed sensing, are based on statistical-mechanical techniques. But even if we limit ourselves to single-user communication systems, yet another very active problem area under this category is that of codes on graphs, iterative decoding, belief propagation, and density evolution. The main reason for not including it in this work is that it is already very well covered in recent textbooks, such as the one Mézard and Montanari [80] as well as the one by Richardson and Urbanke [98]. Another comprehensive exposition of graphical models, with a fairly strong statistical-mechanical flavor, was written by Wainwright and Jordan [118].

As will be seen, the physics and the information-theoretic subjects are interlaced with each other, rather than being given in two continuous, separate parts. This way, it is hoped that the relations between information theory and statistical physics will be made more apparent. We shall see that, not only these relations between information theory and statistical physics are interesting academically on their own right, but, moreover, they also prove useful and beneficial in that they provide us with new insights and mathematical tools to deal with information-theoretic problems. These mathematical tools sometimes prove a lot more efficient than traditional tools used in information theory, and they may give either simpler expressions for performance analysis, or improved bounds, or both.

Having said that, a certain digression is in order. The reader should not expect to see too many real breakthroughs, which are allowed
exclusively by statistical-mechanical methods, but could not have been achieved otherwise. Perhaps one exception to this rule is the replica method of statistical mechanics, which will be reviewed in this work, but not in great depth, because of two reasons: first, it is not rigorous (and so, any comparison to rigorous information-theoretic methods would not be fair), and secondly, because it is already very well covered in existing textbooks, such as [80] and [87]. If one cares about rigor, however, then there are no miracles. Everything, at the end of the day, boils down to mathematics. The point then is which culture, or scientific community, has developed the suitable mathematical techniques and what are the new insights that they provide; in many cases, it is the community of statistical physicists.

There are several examples of such techniques and insights, which are emphasized rather strongly in this work. One example is the use of integrals in the complex plane and the saddle-point method. Among other things, this should be considered as a good substitute to the method of types, with the bonus of lending itself to extensions that include the countable and the continuous alphabet case (rather than just the finite alphabet case). Another example is the analysis technique of error exponents, which stems from the random energy model (see Section 6 and onward), along with its insights about phase transitions. Again, in retrospect, these analyses are just mathematics and therefore could have been carried out without relying on any knowledge in physics. But it is nevertheless the physical point of view that provides the trigger for its use. Moreover, there are situations (see, e.g., Section 7.3), where results from statistical mechanics can be used almost verbatim in order to obtain stronger coding theorems. The point is then that it is not the physics itself that may be useful, it is the way in which physicists use mathematical tools.

One of the main take-home messages, that will hopefully remain with the reader after reading this work, is that whatever the field of statistical mechanics has to offer to us, as information theorists, goes much beyond the replica method. It is believed that this message is timely, because the vast majority of papers at the interface between the two disciplines are about applying the replica method to some information-theoretic problem.
The outline of the remaining part of this work is as follows: in Section 2, we give some elementary background in statistical physics and we relate fundamental thermodynamic potentials, like thermodynamical entropy and free energy with fundamental information measures, like the Shannon entropy and the Kullback–Leibler divergence. In Section 3, we explore a few aspects of physical interpretations of some fundamental results in information theory, like non-negativity of the Kullback–Leibler divergence, the data processing inequality, and the elementary coding theorems of information theory. In Section 4, we review some analysis tools commonly used in statistical physics, like the Laplace integration method, the saddle-point method, and the replica method, all accompanied by examples. Section 5 is devoted to a (mostly descriptive) exposition of systems with interacting particles and phase transitions, both in physics and information theory. Section 6 focuses on one particular model of a disordered physical system with interacting particles — the random energy model, which is highly relevant to the analysis of random code ensembles. Section 7 extends the random energy model in several directions, all relevant to problems in information theory. Finally, Section 8 contains a summary and an outlook on the interplay between information theory and statistical mechanics.

As with every paper published in *Foundations and Trends in Communications and Information Theory*, the reader is, of course, assumed to have some solid background in information theory. Concerning the physics part, prior background in statistical mechanics does not harm, but is not necessary. This work is intended to be self-contained as far as the physics background goes.

In a closing note, it is emphasized again that the coverage of topics, in this work, is by no means intended to be fully comprehensive, nor is it aimed at providing the complete plethora of problem areas, methods and results. The choice of topics, the approach, the flavor, and the style are nothing but the mirror image of the author’s personal bias, perspective, and research interests in the field. Therefore, this work should actually be viewed mostly as a monograph, and not quite as a review or a tutorial paper. This is also the reason that a considerable part of the topics, covered in this work, is taken from articles in which the author has been involved.
References


207
References


References


References


Full text available at: http://dx.doi.org/10.1561/0100000052
References


References


