Information Theory and Statistical Physics

Lecturer: Neri Merhav

Details About the Course

- **Prerequisite:** Information Theory (048733).

  Comment: Prior background in statistical physics is advantageous but not compulsory. The course will be self–contained as far as the required (elementary) physics background goes.

- **Target Audience and Objectives:** The course is aimed at EE graduate students in the area of Communications and Information Theory, as well as to graduate students in Physics who have basic background in Information Theory (in particular, students that graduated the EE–Physics track form the ideal target audience for this course). The main objective of the course is to expose the student to various aspects of the relationships between Information Theory and Statistical Physics. Accordingly, strong emphasis will be given to the analogy and parallelism between these two theories, 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 student to the meeting points between these two disciplines, we will enhance his/her background and perspective to carry out research in the field.

- **Short Syllabus:** Introduction; Elementary Statistical Physics and its Relation to Information Theory; Analysis Tools in Statistical Physics; Systems of Interacting Particles and Phase Transitions; The Random Energy Model (REM) and Random Channel Coding; Additional Topics (optional).

- **Requirements:** Homework assignments (30%) plus a critical summary (70%) on a paper from a given list (plus an oral discussion).

- **Administration:** The course will make use of a moodle website, which will be open to guest access too. The moodle site will be used to post lecture notes, homework assignments, the list of papers for critical summary, and messages about updates. The lecture notes can also be downloaded from my personal website (http://webee.technion.ac.il/people/merhav/). The course number is 048704 ("Selected topics in ...").

- **Bibliography:**

1. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, second edition, John Wiley & Sons, 2006.
2. R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.
3. M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford University Press, 2009. Available on–line at: [http://www.stanford.edu/∼montanar/BOOK/book.html].
4. H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: an Introduction*, (International Series of Monographs on Physics, no. 111), Oxford University Press, 2001.
5. J. P. Sethna, *Statistical mechanics: entropy, order parameters, and complexity*, Oxford University Press, 2007.
6. F. Mandl, *Statistical Physics*, John Wiley & Sons, 1971.
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics – Volume 5: Statistical Physics, Part 1*, 3rd edition, Elsevier, 1980.
8. J. Honerkamp, *Statistical physics – an advanced approach with applications*, 2nd edition, Springer–Verlag, 2002.
9. C. Kittel, *Elementary Statistical Physics*, John Wiley & Sons, 1958.
10. F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw–Hill, 1965.
11. A. H. W. Beck, *Statistical Mechanics, Fluctuations and Noise*, Edward Arnold Publishers, 1976.

12. M. Kardar, *Statistical Physics of Particles*, Cambridge University Press, 2007.
13. N. G. de Bruijn, *Asymptotic methods in analysis*, Dover Publications, 1981.
14. Papers from the current literature.

0. Introduction

This course, which is rather interdisciplinary, is intended to EE graduate students in the field of Communications and Information Theory, and also to graduates of the Physics Department (in particular, graduates of the EE–Physics program) who have basic background in Information Theory, which is a prerequisite to this course. As its name suggests, this course focuses on relationships and interplay between Information Theory and *Statistical Physics* – a branch of physics that deals with many–particle systems using probabilitistic/statistical methods in the microscopic level.

The relationships between Information Theory and Statistical Physics (+ thermodynamics) are by no means new, and many researchers have been exploiting them for many years. Perhaps the first relation, or analogy, that crosses our minds is that in both fields, there is a fundamental notion of *entropy*. Actually, in Information Theory, the term entropy was coined after the thermodynamic entropy. The thermodynamic entropy was first introduced by Clausius (around 1850), whereas its probabilistic–statistical interpretation is due to Boltzmann (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, and we will not cover all of them in this course, but just to give the idea 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 theortists, but also that of researchers in related fields like signal processing, image processing, and the like. It is about a philosopy, 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 underlying 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 moment, then the ME distribution is Gaussian with matching first and second order moments. Indeed, the Gaussian model is perhaps the most widespread model for physical processes in Information Theory as well as in signal– and image processing. But why maximum entropy? The answer to this philosophical 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 equilibrium, its entropy reaches its maximum. Of course, when it comes to problems in Information Theory and other related fields, this principle becomes quite heuristic, and so, one may question its relevance, but nevertheless, this approach has had an enormous impact on research trends throughout the last fifty years, after being proposed by Jaynes in the late fifties of the previous century, and further advocated by Shore and Johnson afterwards. In the book by Cover and Thomas, there is a very nice chapter on this, but we will not delve into this any further in this course.

- *Landauer's Erasure Principle.* Another aspect of these relations has to do with a piece of theory whose underlying guiding principle is that *information is a physical entity.* In every information bit in the universe there is a certain amount of energy. Specifically, Landauer's erasure principle (from the early sixties of the previous century), which is based on a 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. It is my personal opinion that these kind of theories should be taken with a grain of salt, but this is only my opinion. At any rate, this is not going to be included in the course either.

- *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 the most fundamental mathematical tool is the *Chernoff bound.* This is a topic that will be covered in the course and quite soon.

- *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 hot area in probability theory with many applications, both in Statistical Physics and in Information Theory, especially in modern theories of wireless communication (e.g., MIMO systems). This is again outside the scope of this course, but whoever is interested to 'taste' it, is invited to read the 2004 paper by Tulino and Verdú in *Foundations and Trends in Communications and Information Theory*, a relatively new journal for tutorial papers.

- *Spin Glasses and Coding Theory.* 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 cusomarily 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 course and we will study it extensively, at least from some aspects.

We can go on and on with this list and add more items in the context of these very fascinating meeting points between Information Theory and Statistical Physics, but for now, we stop here. We just mention that the last item will form the main core of the course. We will 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 analsysis, or improved bounds, or both.

At this point, let us have a brief review of the syllabus of this course, where as can be seen, the physics and the Information Theory 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 seen more readily. The detailed structure of the remaining part of this course is as follows:

1. *Elementary Statistical Physics and its Relation to Information Theory:* What is statistical physics? Basic postulates and the micro–canonical ensemble; the canonical ensemble: the Boltzmann–Gibbs law, the partition function, thermodynamical potentials and their relations to information measures; the equipartition theorem; generalized ensembles (optional);

Chernoff bounds and the Boltzmann–Gibbs law: rate functions in Information Theory and thermal equilibrium; physics of the Shannon limits.

2. *Analysis Tools in Statistical Physics:* The Laplace method of integration; the saddle–point method; transform methods for counting and for representing non–analytic functions; examples; the replica method – overview.

3. *Systems of Interacting Particles and Phase Transitions:* Models of many–particle systems with interactions (general) and examples; a qualitative explanation for the existence of phase transitions in physics and in information theory; ferromagnets and Ising models: the 1D Ising model, the Curie-Weiss model; randomized spin–glass models: annealed vs. quenched randomness, and their relevance to coded communication systems.

4. *The Random Energy Model (REM) and Random Channel Coding:* Basic derivation and phase transitions – the glassy phase and the paramagnetic phase; random channel codes and the REM: the posterior distribution as an instance of the Boltzmann distribution, analysis and phase diagrams, implications on code ensemble performance analysis.

5. *Additional Topics (optional)*: The REM in a magnetic field and joint source–channel coding; the generalized REM (GREM) and hierarchical ensembles of codes; phase transitions in the rate–distortion function; Shannon capacity of infinite–range spin–glasses; relation between temperature, de Bruijn's identity, and Fisher information; the Gibbs inequality in Statistical Physics and its relation to the log–sum inequality of Information Theory.

As already said, there are also plenty of additional subjects that fall under the umbrella of relations between Information Theory and Statistical Physics, which will not be covered in this course. One very hot topic is that of codes on graphs, iterative decoding, belief propagation, and density evolution. The main reason for not including these topics is that they are already covered in the course of Dr. Igal Sason: "Codes on graphs."

As a final note, I emphasize that prior basic background in Information Theory will be assumed, therefore, Information Theory is a prerequisite for this course. As for the physics part, prior background in statistical mechanics could be helpful, but it is not compulsory. The course is intended to be self–contained as far as the physics background goes. The bibliographical list includes, in addition to a few well known books in Information Theory, also several very good books in elementary Statistical Physics, as well as two books on the relations between these two fields.