Chapter One

Introduction

Andreas Greven, Gerhard Keller
Universität Erlangen

Gerald Warnecke
Otto-von-Guericke University Magdeburg

The concept of entropy arose in the physical sciences during the 19th century, in particular in thermodynamics and in the development of statistical physics. From the very beginning the objective was to describe the equilibria and the evolution of thermodynamic systems. The development of the theory followed two conceptually rather different lines of thought. Nevertheless, they are symbiotically related, in particular through the work of Boltzmann.

The historically older line adopted a macroscopic point of view on the systems of interest. They are described in terms of a relatively small number of real variables, namely, temperature, pressure, specific volume, mass density, etc., whose values determine the macroscopic properties of a system in thermodynamic equilibrium. Clausius, building on the previous intuition of Carnot, introduced for the first time in 1867 a mathematical quantity, S, which he called *entropy*. It describes the heat exchanges that occur in thermal processes via the relation

$$\mathrm{d}S = \frac{\mathrm{d}Q}{T}.$$

Here Q denotes the amount of heat and T is the absolute temperature at which the exchange takes place. Among the scientists who developed this idea further are some of the most influential physicists of that time, most notably Gibbs and Planck, as well as the mathematician Carathéodory.

The other conceptual line started from a microscopic point of view on nature, where macroscopic phenomena are derived from microscopic dynamics: any macrostate is represented by many different microstates, i.e. different configurations of molecular motion. Also, different macrostates can be realized by largely differing numbers of corresponding microstates. Equilibria are those macrostates which are most likely to appear, i.e. they have the largest number of corresponding microstates (if one talks about the simplest case of unconstrained equilibria). In particular, two names are associated with this idea: Boltzmann and Maxwell. Boltzmann argued that the Clausius entropy S associated with a system in equilibrium is proportional to the logarithm of the number W of microstates which form the macrostate of this equilibrium,

$$S = k \ln W.$$

The formula is engraved on his tombstone in Vienna. Here the idea in the background, even though not explicitly formulated, is a relation between the encoded information specifying a microstate and the complexity of the system.

Since then both of these approaches to entropy have led to deep insights into the nature of thermodynamic and other microscopically unpredictable processes. In the 20th century these lines of thought had a considerable impact on a number of fields of mathematics and this development still continues. Among the topics where these ideas were most fruitful are stochastic processes and random fields, information and coding, data analysis and statistical inference, dynamical systems and ergodic theory, as well as partial differential equations and rational mechanics. However, the mathematical tools employed were very diverse. They also developed somewhat independently from their initial physical background. After decades of specialization and diversification, researchers involved in one of the areas of research mentioned above are often unprepared to understand the work of the others. Nowadays, it takes a considerable effort to learn and appreciate the techniques of a complementary approach, and the structure of this book reflects this and the state of the whole field. On the other hand, it was our intention with the symposium and this book to help bridge the gaps between the various specialized fields of research dealing with the concept of entropy. It was our goal to identify the unifying threads by inviting outstanding representatives of these research areas to give surveys for a general audience. The reader will find examples of most of the topics mentioned above in this book. To a large extent, the idea of entropy has developed in these areas in the context of quite different mathematical techniques, as the reader will see, but on the other hand there are common themes which we hope to exhibit in this book and which surfaced in the discussions taking place during the meeting.

We found two major methods in which entropy plays a crucial role. These are *variational principles* and *Lyapunov functionals*. They provide common tools to the different research areas.

Variational principles come in many guises. But they always take the following form, which we discuss for the case of the Gibbs variational principle. There, the sum of an energy functional and the entropy functional are minimized or maximized, depending on a choice of sign (see, for example, Section 2.3.1 on p. 29 and Section 3.6 on p. 50). Along the macroscopic line of thought such a variational problem appears as a first principle. It is clear that such a principle is crucial in thermodynamics and statistical physics. However, the reader will learn that it also plays a key role in the description of dynamical systems generated by differentiable maps, for populations evolving in a random medium, for the transport of particles and even in data analysis. As we proceed in the description of the various contributions we will show in more detail how this variational principle plays its role as an important tool.

For the microscopic line of thought, take the situation of a system where a macrostate corresponds to many microstates. In the Gibbs variational principle, one term reflects the number of possibilities to realize a state or path of a dynamical system or of a stochastic process under a macroscopic constraint such as, for example, a fixed energy. Then for a gain, e.g. in energy, one may have to 'pay' by a reduction in the number of ways in which microstates can achieve this gain. This results in a competition between energy and entropy that is reflected in the variational principle. Bearing this in mind it is not surprising that in many contributions to this book, a major role is played by information or information gain. There are two sides to the coin: one side is the description of the *complexity* of the problem to specify a microstate satisfying a macroscopic constraint; the other is the problem of *coding* this information. It is therefore also plausible that there are deeper connections between problems which at the surface look quite different, e.g. the description of orbits of a map on the one hand and data analysis on the other. We hope that these seemingly mysterious relations become clearer by reading through this book.

Another such common theme in many areas touched on in this book is the close connection of entropy to stability properties of dynamical processes. The entropy may, for instance, be interpreted as a Lyapunov function. Originally introduced for systems of ordinary differential equations, Lyapunov functions are an important tool in proving asymptotic stability and other stability properties for dynamical systems. These are functions that are increasing or decreasing with the dynamics of a system, i.e. the trajectories of solutions to the system cut the level sets of the function. This is exactly the type of behavior that the entropy exhibits due to the second law of thermodynamics. It is also exhibited in the axiomatic framework of Lieb–Yngvason when considering sequences of adiabatically accessible states, especially if the sequence of transitions is irreversible. For time-dependent partial differential equations in the framework of evolution equations, i.e. considering them as ordinary differential equations on infinite-dimensional spaces, such as Sobolev spaces, one has the analogous tool of Lyapunov functionals. Basically, in all these cases asymptotic stability tells us in which state (or states), in the long run, a dynamical system ends up.

There are other theoretical implications concerning the stability of a system as well, which are coming up in the chapters by Dafermos and Young. For instance, on the one hand, Sections 6.3 and 6.4 on pp. 110 and 113 of the chapter by Dafermos are devoted to illuminating the implications of entropy inequalities for stability properties of weak, i.e. discontinuous, solutions to hyperbolic conservation laws. Similarly, entropy inequalities are also used as a tool for systems of parabolic differential equations. On the other hand this might, at first sight, contrast with the role played by entropy in the chapter by Young, where it is connected to the instability of dynamical systems with chaotic behavior. However, instability of individual trajectories is a feature of the microscopic level which is responsible for the macroscopic stability of the actually observed invariant measures. Keeping in mind that the ergodic theory of dynamical systems is an equilibrium theory that cannot describe the dynamics of transitions between invariant measures, the role of entropy in variational principles, which single out the actually observed invariant measure among all other ones, comes as close as possible to the role of a Lyapunov functional.

1.1 Outline of the Book

Now we outline and comment on the contents of this book. In the first part, basic concepts, terminology and examples from both the macroscopic and the microscopic line of thought are introduced in a way that should provide just the right amount of detail to prepare the interested reader for further reading. The second part comprises five contributions tracing various lines of evolution that have emerged from the macroscopic point of view. The third part collects five contributions from the 'probabilistic branch,' and in a final part we offer the reader four contributions that illustrate how entropic ideas have penetrated coding theory and the theory of dynamical systems.

Due to limited time at the symposium, a number of further important topics related to entropy could not be dealt with appropriately, and as a consequence there are a number of deplorable omissions in this book, in particular regarding the role of entropy in statistics, in information theory, in the ergodic theory of amenable group actions, and in numerical methods for fluid dynamics.

Part 1 The concept of entropy emerging from the macroscopic point of view is emphasized by Ingo Müller's first contribution (p. 19). Starting from Clausius' point of view, Boltzmann's statistical ideas are added into the mix and the role of entropy as a governing quantity behind a number of simple real world phenomena is discussed: the elastic properties of rubber bands as an illustration for configurational entropy, the different atmospheres of the planets in our Solar System as a result of the 'competition between entropy and energy,' and the role of entropy in the synthesis of ammonia are just a few examples.

The approach to entropy in statistical physics developed its full strength in the framework of probability theory. It had a large impact on information theory, stochastic processes, statistical physics and dynamical systems (both classical and quantum). Concepts, terminology and fundamental results which are basic for all these branches are provided by Hans-Otto Georgii (p. 37), the second introductory chapter of this book.

Part 2 In the second part we have collected expositions on the macroscopic approach. There are two chapters on thermodynamics in the context of continuum mechanics. The first one, by Hutter and Wang, reviews alternative approaches to nonequilibrium thermodynamics, while in the second Müller surveys one of these approaches more extensively—the theory of extended thermodynamics—which he developed with various collaborators. Next, in the chapter by Dafermos on conservation laws, we pursue the bridge from continuum mechanics to differential equations further. Finally, we have two chapters that reflect the views of physicists on the fundamental issues concerning the understanding of entropy and irreversibility. Surprisingly, it turns out that there has not been until now one unified, universally accepted theory on the *thermodynamics of irreversible processes*. There are competing theories and concepts with a certain overlap and some fundamental differences. This diversity is a bit confusing, but bringing more clarity and unity to this field is also a challenging research goal. The contributions we have collected here are attempting to expose this diversity to a certain extent and clearly point out some of the conceptual differences that need to be resolved.

Interestingly, the uncertainty about some of the basic physical concepts involved is reflected by the fact that the state of the mathematical theory of the equations that follow from these theories is also very inadequate. Numerical methods play an increasingly important role in the effort to sort this out. This is because numerical computations allow us to explore more deeply consequences of mathematically complicated theories, i.e. the large systems of nonlinear differential equations they lead to. They may allow us to make quantitative predictions and comparisons with physical reality. Entropy also plays a key role in numerical analysis and numerical computations, an important topic that is missing here. However, the analytical tools are actually discussed in the chapter by Dafermos. The link between analytical theory and numerical approximations is particularly close in this field since key existence results were actually obtained via convergence proofs for numerical schemes. In numerical computations it is important to have verified that the method being used is consistent with an entropy condition, otherwise unphysical shock discontinuities may be approximated. Also, it is useful to monitor the entropy production in computations, because false entropy production is an indicator of some error in the scheme or the implementation.

Let us now turn to the contributions. The first chapter in this part is a survey by Kolumban Hutter and Yongqi Wang (p. 57) on some theories of irreversible thermodynamics. It highlights two different forms of the second law of thermodynamics, namely, the Coleman–Noll approach to the Clausius–Duhem inequality and the Müller–Liu entropy principle in a detailed comparison. The structure of these descriptions of thermodynamics is exposed in very concise form. The interesting lesson to the nonexpert is that there are different theories available that have much common ground but marked differences as well, a theme we will encounter again in Uffink's chapter. The second approach that Hutter and Wang discuss is then exposed much more extensively in the next chapter, by Ingo Müller (p. 79). The reader is given an introduction to the theory of *extended thermodynamics*, which provides a macroscopic theory for irreversible thermodynamic processes. It builds a vital link to the microscopic approach of kinetic gas theory formulated in terms of the Boltzmann equation and the density function it has as solution. From these one can rigorously derive an infinite sequence of balance equations involving higher and higher moments of the density function. In order to have a finite number of equations, these are truncated and certain closure laws have to be derived. This approach leads to symmetric hyperbolic systems of partial differential equations that are automatically endowed with an entropy function. The entropy production plays an important role in deriving closure laws for these systems. The reader who wishes to learn more should consult the book on extended thermodynamics that Müller together with Ruggeri recently published in updated form.

Shock discontinuities, i.e. an instantaneous transition from one physical state to a quite different state, are a good approximation to physical reality when we are modelling at the scale of continuum physics. This means that the free mean path between collisions of neighboring particles in a fluid is orders of magnitude smaller than the scale in which we are making observations. In a close-up view of shock waves we see that they do have a finite width. Since at least the work of Becker in the 1930s, it has been known that this width is only of the order of a few free mean path lengths, i.e. in the realm of microscopic kinetic gas theory. It is one of the important achievements of extended thermodynamics that its moment theories provide a precise description of the transition between the physical states across these shocks.

The chapter by Constantine Dafermos (p. 107) is intended to be a teaser to the book he recently wrote, where the ideas and concepts as well as the methods touched on in the chapter are expanded in much more detail. In the theory of partial differential equations a very significant open problem is the lack of an existence and uniqueness theory for nonlinear hyperbolic systems in more than one space dimension. Though such a theory may not be available in full generality such a theory should be found at least in the class of symmetric systems of conservation laws. Such systems appear naturally in various physical contexts, such as the moment systems derived from kinetic theory and discussed by Müller in his chapter. Dafermos gives a further example from isothermal thermoelasticity.

The peculiarity of these nonlinear differential equations is that they generically tend not to have smooth differentiable solutions. One has to use the concept of distributions which provides a framework in which functions may be differentiated in a generalized sense infinitely often. In this context one has a well-defined concept of solutions to differential equations that have jump discontinuities. This is not just a mathematical trick, namely, when you are looking for a solution, just make the solution space in which you are searching bigger and you may find something, possibly useless. However, making the solution space larger has its price. One tends to obtain too many solutions, i.e. one actually does get an abundance of unphysical solutions along with the reasonable one. If we allow jumps in a solution, then the solution may jump up and down almost arbitrarily in an oscillatory manner. This behaviour and certain types of mathematically feasible jump discontinuities are not observed in nature and must be excluded. Next to providing initial conditions in time and appropriate boundary conditions in the case of bounded solution domains, an additional mathematical condition is needed to specify the physically relevant solutions. Solutions of this nature are a very good description of shock waves on a macroscopic scale. These waves are well known and audible in connection with detonations or supersonic flight.

It has become a generally accepted jargon in the field to call such mathematical selection principles *entropy conditions*, or *E-conditions*, as in Chapter 6, even if they do not involve the concept of entropy explicitly. This is due to the fact that in all physically relevant hyperbolic systems of differential equations this selection is due to the second law of thermodynamics in one form or another. A microscopic approach to this question is presented for a special case in the contribution by Varadhan (Chapter 9).

The kinetic theory of gases has as its foundation the classical mechanics of gas molecules. Its mathematical formulation contains no irreversibility of any kind. The sheer enormity of the number of particles involved makes classical mechanics useless for quantitative description and prediction. Going to a macroscopic formulation by averaging procedures, the reversibility of classical mechanics is to a certain extent lost. This is a quite remarkable fact and the cause of some uneasiness with the notion of irreversibility, as discussed by Uffink and by Lieb and Yngvason in their respective chapters. Mathematically, irreversibility clearly shows up in the structure of shock wave solutions of conservation laws. In one-dimensional systems the characteristic curves of one family merge in the shock and most of the information they carry is lost. Already, for the simple Burgers equation, equation (6.9) in the chapter by Dafermos, one can consider for time $t \geqslant T > 0$ an admissible solution with only one shock. It may result from infinitely many different monotonely nonincreasing initial data at time $t = 0$. They have all merged into the same solution by the time $t = T$. This loss of information on the initial data is exactly linked to the property that the mathematical entropy has a jump discontinuity at the shock. But, as long as the entropy and other physical states remain smooth we may go back in time along characteristics to reconstruct the initial data.

A problem of the macroscopic theory of thermodynamics is that there are two levels of theory involved. One level, closer to continuum mechanics, is usually formulated in precise mathematical language of differential forms and differential equations. It involves the energy balance in the first law of thermodynamics, the Gibbs relation stating that the reciprocal of temperature is an integrating factor for the change in heat allowing the entropy to be defined as a total differential, and the second law, e.g. in the form of Clausius as a differential inequality. Such equations and the differential equations incorporating irreversible thermodynamics into continuum mechanics are discussed in the chapters by Müller, Hutter and Wang, and Dafermos.

A second, more fundamental level of macroscopic theory lies at the basis of these analytic formulations, and this is addressed in the final two chapters of this part. As pointed out by Hutter and Wang, there is some controversy about the foundations of irreversible thermodynamics leading to a certain extent to differing formulations of the second law of thermodynamics. But this is not surprising since some of the basic concepts, even the definition of fundamental quantities such as temperature and entropy, have not yet found a universally accepted form. The chapter by Jos Uffink (p. 121) summarizes a more detailed paper he recently published on the foundations of thermodynamics. He discusses historically as well as conceptually the basic notions of time-reversal invariance, irreversibility, irrecoverability, quasistatic processes, adiabatic processes, the *perpetuum mobile* of the second kind, and entropy. Clearly, there is much confusion on the precise meaning of some of them. Mathematical formalism and physical interpretation are not as clear as in other areas of physics, making thermodynamics a subject hard to grasp for nonspecialists and, as all authors of this part seem to agree, for the specialists as well. The distinctions, comparisons of theories and criticisms pointed out by Uffink are a helpful guide to understanding the development of the field and discussions of theoretical implications. He covers some of the same historical ground as Müller's survey in Part 1, but with a different emphasis. In passing he sheds a critical light on the controversial and more philosophical issues of the arrow of time and the heat death of the Universe, the statements in connection with entropy that have stirred the most interest among the general public.

The final contribution of Part 2 is a survey by Elliott Lieb and Jakob Yngvason (p. 147) of their very recent work attempting to clarify the issues of introducing entropy, adiabatic accessibility and irreversibility by using an axiomatic framework and a more precise mathematical formulation. This is a clear indication that the foundations of thermodynamics remain an active area of research. Such a more rigorous attempt was first made almost a century ago by Carathéodory. Uffink discusses these two more rigorous approaches of Carathéodory and Lieb and Yngvason in his chapter. Here we have a more detailed introduction to the latter approach by its proponents. To a mathematician this is clearly the type of formalization to be hoped for, because it allows precise formulations and the proof of theorems. They provide a clear and precise set of axioms. From these they prove theorems such as the existence and, up to an affine transformation, uniqueness of the nonincreasing entropy function on states as well as statements on its scaling properties. Another theorem states that temperature is a function derived from entropy. The fact that energy flows from the system with higher temperature to the system with lower temperature is a consequence of the nondecrease of total entropy. An interesting fact, also pointed out by Uffink, is that the Lieb–Yngvason theory derives this entropy from axioms that are time-reversal invariant, another variant of the issue already mentioned above. The goal of this theory is to have very precise mathematical formulations that find universal acceptance in the scientific communities involved. As Lieb and Yngvason point out at the end of Section 8.2, discovering the rules of the game is still part of the problem. This intellectually challenging field is therefore open to further exploration.

Part 3 The third part of this book deals with the role of entropy as a concept in probability theory, namely, in the analysis of the large-time behavior of stochastic processes and in the study of qualitative properties of models of statistical physics. Naturally, this requires us to think along the lines of the microscopic approach to entropy. We will see, however, that there are strong connections to Part 2 via hyperbolic systems and via Lyapunov functions (compare the sections on the Burgers equation in the contributions of Dafermos and Varadhan), and to Part 4, in particular to dynamical systems, due to similar phenomena observed in the large-time behavior of dynamical systems and stochastic systems. The first two contributions of Part 3 deal with the role of entropy for stochastic processes, the last three deal with entropy in models of statistical physics. All these contributions show that variational principles play a crucial role. They always contain a term reflecting constraints of some sort imposed on the system and an entropy term representing the number of possibilities to realize the state of the system under these constraints.

The fundamental role that entropy plays in the theory of stochastic processes shows up most clearly in the theory of *large deviations*, which tries to measure probabilities of rare (untypical) events on an exponential scale. Historically, this started by asking the question how the probability that a sum of n independent random variables exceeds a level $a \cdot n$, when a is bigger than the mean of a single variable, decays as n tends to infinity. More precisely, can we find the *exponential rate* of decay of this event as n tends to infinity? The role of entropy in the theory of large deviations is the topic treated in the contribution by Shrinivasa R. S. Varadhan (p. 199). He starts by explaining how entropy enters in calculating the probability of seeing untypical frequencies of heads in n-fold and independent coin tossing. If one observes the whole empirical distribution of this experiment, one is led to Sanov's Theorem now involving relative entropies. Along this line of thought one passes next from independent trials to Markov chains and spatial random fields with a dependence structure which is given by Gibbs measures, the key object of statistical physics. In all these situations one is led to *variational problems* involving entropy on the one hand, since one has to count possible realizations, and a term playing the role of an energy arising from a potential defined as the logarithm of relative probabilities. This competition between entropy and energy is explained in various contexts ranging from coin tossing to occupation numbers of Markov chains. At this point it is useful to remark that in Part 4 very similar structures play a role in the contribution by Young. However, these questions cannot only be asked in spatial situations given, for example, by the realization of a random field at a given time, but one can take the whole space-time picture of the process into view. This allows us, for example, to study small random perturbations of dynamical systems and to determine asymptotically as the perturbation gets small the probabilities of paths of the stochastic process leaving small tubes around the path of the dynamical system in an exponential scale. The contribution concludes with a discussion of the hydrodynamic rescaling of stochastic particle systems and their relation to Burgers equations. Of particular interest in this model is the problem of selecting the proper solution in the presence of shocks and here again entropy is the key. This connects directly with the contribution by Dafermos in Part 2.

An important application of large-deviation theory is the study of stochastic processes evolving in a random medium. This means that the parameters of a stochastic evolution are themselves produced by a random mechanism. The contribution of Frank den Hollander (p. 215) discusses an interesting and typical example for this situation in which the relative entropy is the key quantity. It is an example from the area of spatial population models. Here particles live on the integer lattice, they migrate to the left and to the right with certain probabilities and split into several new particles (or become extinct) with a probability distribution of the offspring size which depends on the location. The large-time behavior exhibits phase transitions from extinction to explosion of the population. Here the *variational problem* energy versus entropy of statistical physics appears in a different guise. The energy is replaced by the local growth parameter given by the random medium. The entropy enters since it regulates asymptotically the numbers of paths which can be used by individuals migrating from a point A to a point B, when A and B become very distant. However, since the mathematical structure is exactly the one occurring in statistical physics, the system exhibits various phase transitions in the parameter regulating the drift of motion. Depending on that drift we see populations growing or becoming extinct and here one can distinguish both local and global extinction. Even though the model is rather special, the phenomena and structures exhibited are quite representative for systems in random media.

The two contributions by Enzo Olivieri and Christian Maes use stochastic evolutions to discuss the phenomenon of *metastability* and *entropy production*, two key topics for the broader theme of entropy in statistical physics. Metastable behavior of a physical system is a phenomenon which should be understood by means of probabilistic microscopic models of the system. To make a step in this direction is the goal of the contribution by Olivieri (p. 233). He discusses two examples of metastable behavior, supersaturated vapor as an example of a conservative system (the number of particles is conserved), and ferromagnetic systems below the Curie temperature for a nonconservative context. He discusses these two systems in the framework, respectively, of van der Waals and Curie–Weiss theory in great detail. He continues the discussion with a critical review of the shortcomings of this approach and passes on to the description of the 'restricted ensemble approach' and the 'pathwise approach' to this problem of describing metastable behavior. This pathwise approach is then formulated precisely in the context of the stochastic Ising model for ferromagnets and finally a description of the general probabilistic mechanism behind metastable behavior via first-exit problems of Markov chains is provided. The concept of the first descent tube of trajectories is used to relate metastable behavior to appropriate concepts of time entropy. Again, in deciding which route a system takes to pass from the metastable to the stable state, a key role is played by the competition between the number of paths producing particular features of the route and the probability of such paths. Hence again a *variational problem* describing this competition is in the background, and again it has the form of the sum of an entropy and an energy term.

The concept of entropy in nonequilibrium is treated in the contribution by Maes (p. 251). The key quantity in this approach is the entropy production rate. One proves that it is always nonnegative and shows that the system is time reversible if and only if there is no entropy production. The second point is to study the local fluctuations of the entropy production rate. Here these ideas are developed in the framework of spatially extended systems with many interacting components and a simple stochastic evolution mechanism, the so-called asymmetric exclusion process. In this system particles move according to independent random walks with a drift and the additional rule is imposed that every site can only be occupied by at most one particle. Due to this term the system exhibits interaction between the components. (Nevertheless, the equilibria of the system are product measures.) Here the entropy production is defined by considering the system and the time-reversed system and forming the logarithmic ratio of the number of those microstates with which the original macrostate can be realized.

The third part of the book concludes with a contribution by Joel Lebowitz and Christian Maes (p. 269) of a quite different form, namely, a dialogue on entropy between a scientist and an angel. Their conversation touches on such topics as the Boltzmann equation, the second law of thermodynamics and the von Neumann entropy in quantum systems. The reader will see many of the topics in this part in a different light and feel stimulated to have another look at the previous contributions.

Part 4 The final part of this book is devoted to several offspring of the probabilistic concept of entropy discussed already in Part 3, namely, to its use in information compression and dynamical systems. Although these two directions deal with nearly complementary aspects of entropy, they both use Shannon's concept of entropy as a measure of information (see also Section 3.2 on p. 39).

For a thorough understanding of why information compression and uncertainty produced by dynamical systems are two sides of the same coin, *coding* is the basic ingredient. Information compression is the art of describing data given as a (long) string of symbols from some alphabet in a 'more efficient' way by a (hopefully shorter) string from a possibly different alphabet, here from the set $\{0, 1\}$. The classical result in this situation is Shannon's source-coding theorem for memoryless sources saying roughly that for randomly generated data there exists a simple coding procedure such that the average code length for data of length n is nearly n times the entropy of the source. In dynamical systems one wants to measure the 'diversity' of the set of all orbits of length n when such orbits are observed with 'finite precision,' i.e. one splits the state space into a finite number of sets and observes only which sets are visited in the successive time steps. This led Kolmogorov and Sinai to code individual orbits by recording the outcomes of such coarse-grained observations at times 1 to n and to apply ideas from information theory to these code words and their probability distributions.

This circle of ideas is the starting point of Fabio Benatti's contribution (p. 279). Beginning with a condensed description of the basic notions and results leading to the concept of *Kolmogorov–Sinai entropy* for the dynamics of a classical measure-theoretic dynamical system, Benatti gradually develops the appropriate concepts for quantum dynamical systems which generalize the classical ones. In this way discrete probability distributions are replaced by density matrices, products of probability distributions by tensor products of density matrices, the Shannon entropy of a discrete probability distribution by the von Neumann entropy of a density matrix and the coding theorem for classical memoryless sources by an analogous result for quantum memoryless sources. Finally, the generalization from discrete probability distributions to general probability spaces translates into a passage from density matrices to positive operators on Hilbert spaces. Whereas these transfers from the classical to the quantum world are relatively straightforward, the situation changes drastically when one tries to pass from the classical Kolmogorov–Sinai entropy to a quantum dynamical entropy. The idea to model a coarse-grained observation device for a classical system as a finite measurable partition of the classical phase space has no obvious quantum counterpart. Two approaches (which both coincide with Kolmogorov–Sinai entropy on classical dynamical systems) are discussed: the *Connes–Narnhofer–Thirring entropy*, where partitions of the state are used, and the *Alicki–Fannes entropy*, which is based on partitions of unity.

It is worthwhile noting that the above-mentioned problem to find a shortest description of a given finite string of data is not inherently probabilistic, but that Kolmogorov and, independently, Chaitin offer a purely deterministic approach using the concept of complexity of such a string. It is reassuring, though, that both theories meet—at least in the classical context as pointed out in Section 14.3 of Benatti's contribution. A much more prominent role is played by complexity in Jorma Rissanen's contribution (p. 299). Again the coding of strings of data is the starting point, but now the optimal code word is the shortest 'program' for a universal Turing machine from which the machine can reproduce the encoded data. Following some unpublished work from Kolmogorov's legacy, we look at (essentially) the shortest descriptions which consist of two parts: one that describes 'summarizing properties' of the data and another one that just specifies the given data string as a more or less random string within the class of all strings sharing the same summarizing properties. Then the summarizing properties represent only the 'useful' information contained in the data. Unfortunately, it is impossible to implement a data analysis procedure based on Kolmogorov's complexity because the complexity of a string is uncomputable. A more modest goal can be achieved, however. Given a suitably parametrized class of models for the observed data, one can define the complexity of a string of data relative to the model class and decompose the shortest description of the data relative to the model class into some 'useful' information and a noninformative part. Now the close relation between effective codes and probability distributions expressed by Shannon's source-coding theorem allows us to formalize the notion of a 'parametrized class of models' by a parametrized class of probability distributions so that it is not too surprising that actual calculations of complexity and information of data involve not only Shannon entropy but also classical statistical notions like the Fisher information matrix.

There is one more notion from dynamical systems which Benatti connects to entropy: *chaos* in classical and also in quantum systems. The basic idea, briefly indicated here, is the following. Unpredictability of a deterministic dynamical system, which is caused by an extreme sensitivity to initial conditions, is noticed by an external observer because typical coded trajectories of the system are highly complex and resemble those produced by a random source with positive entropy. This is the theme of Lai-Sang Young's contribution (p. 313), where entropy means dynamical entropy, i.e. the logarithmic growth rate of the number of essentially different orbits as mentioned before. If not really all orbits are taken into account but only most of them (in the sense of an underlying invariant probability measure), this growth rate coincides with the Kolmogorov–Sinai entropy introduced above. If all orbits are counted, the corresponding growth rate is called topological entropy. Both notions are connected through a variational principle saying that the topological entropy is the supremum of all Kolmogorov–Sinai entropies with respect to invariant measures. For differentiable dynamical systems with a given invariant measure the Kolmogorov–Sinai entropy, which is a global dynamical characteristic, can be linked closely to more geometric local quantities, namely, to the Lyapunov exponents which describe how strongly the map expands or contracts in the directions of certain subspaces, and to certain partial dimensions of the invariant measure in these directions. In the simplest case, when the invariant measure is equivalent to the

Riemannian volume on the manifold and has at least one positive Lyapunov exponent, this very general relationship specializes to *Pesin's Formula*. In fact, the validity of Pesin's Formula characterizes exactly those invariant measures which are physically observable and known as the *Sinai–Ruelle–Bowen* measures. This characterization can be cast in the form of a *variational principle* involving the Kolmogorov–Sinai entropy and, as 'energy' terms, the average Lyapunov exponents.

There are more interpretations of entropy in dynamical systems. Young describes how the counting of periodic orbits and how the volume growth of cells of various dimension is linked to topological entropy, and how large-deviation techniques, omnipresent in Part 3 of this book, can be successfully applied to dynamical systems with positive entropy. Michael Keane (p. 329) describes the breakthrough that Kolmogorov–Sinai entropy, called mean entropy in this contribution, brought for the classification of measure-theoretic dynamical systems. It culminated in *Ornstein's* proof that two Bernoulli systems are isomorphic if and only if they have the same Kolmogorov–Sinai entropy. As, in turn, many systems occurring 'in nature' are isomorphic to Bernoulli systems, this result clarifies the measure-theoretic structure of a great number of dynamical systems.

All the results above, however, are rather meaningless for systems of zero entropy. The structure theory of such systems is therefore a challenging field of research, and Keane presents an example—the so-called binomial transformation—that conveys the flavour of these kinds of system, which in many respects are rather different from positive entropy systems.

1.2 Notations

In view of the broad spectrum of contributions from rather different fields—all with their own history, mathematical techniques and their time-honored notations—we feel that a few additional remarks might be useful to help readers to interpret the various notations correctly.

For the benefit of readers not familiar with continuum mechanics we would like to point out a subtle point concerning time derivatives that may be somewhat confusing at first. The fundamental laws of mechanics, such as conservation of mass, i.e. no mass is created or lost, or Newton's law, namely, that the time change in momentum is equal to the acting forces, are most simply and familiarly formulated in the so-called *Lagrangian framework*, where one is following trajectories of massive particles, single-point particles or (in continuum mechanics) an arbitrary volume of them. The rate of change in time, i.e. the time derivative, in this framework is written as the total derivative $\mathrm{d}/\mathrm{d}t$ or frequently by a dot over the physical state variable that is changing in time, e.g. $m\dot{\boldsymbol{u}} = \boldsymbol{f}$ for Newton's law with mass m, velocity $\boldsymbol{u}$ and force field $\boldsymbol{f}$. This derivative is called the *material derivative* and must be distinguished from the partial derivative in time arising in the context of the *Eulerian framework*. In the latter all physical state variables are simply described as a field, not in the sense of algebra, but rather meaning that they are given as functions of time and space, maybe vector-valued or higher-order tensor-valued. Particle trajectories are not given explicitly but may be obtained by integrating the velocity field. The rate of change in time in this framework is given by the partial derivatives $\partial/\partial t$, the rate of change in space by the spatial gradient ∇ consisting of the partial spatial derivatives of the physical variables. The connection between the two types of time derivatives $\mathrm{d}/\mathrm{d}t$ and $\partial/\partial t$ is given by the chain rule. Following a particle trajectory of an Eulerian variable such as the mass density ρ gives the derivatives in the following form

$$\frac{\mathrm{d}\rho}{\mathrm{d}t}(t, x(t)) = \frac{\partial\rho}{\partial t} + \nabla\rho \cdot \dot{x} = \frac{\partial\rho}{\partial t} + \boldsymbol{u} \cdot \nabla\rho,$$

where $\boldsymbol{u}$ denotes the velocity field in the Eulerian framework. This point is discussed in more detail in any good introductory textbook on continuum mechanics.

The standard set-up used in probability theory is a basic probability space $(\Omega, \mathcal{A}, P)$ with a set Ω describing possible outcomes of random experiments, a σ-algebra $\mathcal{A}$ describing possible events, and a basic measure P representing the probabilities of these events. On that space random variables are defined, often called X, which are maps from Ω into some set E, the possible values which the random variable can attain. This might be real numbers, vectors, or even measures. In the latter case such random variables are called random measures. One very important example is the empirical distribution arising from repeating a random experiment n times. The law specifying the probability that this observed distribution is close to a given one is then in fact a probability law on the set of probability laws. Random variables are in principle maps from one measure space into another, i.e. $X : \Omega \to E$, and one would expect to write $\{\omega \in \Omega \mid X(\omega) \in A\}$ for the event that X takes values in a set A, but in fact one typically writes $\{X \in A\}$. Hence you might encounter symbols like L_n, Y_t, etc., which are random objects but no ω floats around to remind you, and the state spaces of the random variables might vary from $\mathbb{R}$ over spaces of continuous functions to spaces of measures. Furthermore, for the expectation $\int_\Omega X(\omega) P(\mathrm{d}\omega)$ of the random variable X one writes $\int X \,\mathrm{d}P$ and even shorter $E[X]$, and here neither ω nor Ω or P appears explicitly. There is a good reason for this since the interest is usually not so much in Ω and P but in the values a certain random variable takes and in its distribution. This distribution is a probability measure on the space E where the random variable takes its values. Analysing properties of this random variable the basic space is then very often of little relevance.

Another point to be careful about is to distinguish in convergence relations whether random variables, i.e. measurable functions, actually converge in some sense (almost surely, in mean or in probability), or whether only their laws converge, that is, expectations of continuous bounded functions of those variables converge to a limit thus determining a limiting probability distribution.

In ergodic theory and the theory of dynamical systems (with discrete time), the basic object is a measurable transformation, call it f, acting on a more or less abstract state space, often called X. Typical examples of state spaces are manifolds or sequence spaces like $A^{\mathbb{Z}}$, where A is a finite symbol set. Properties of trajectories $x, f(x), f^2(x), \ldots$ obtained from repeated application of the transformation are studied relative to an invariant measure, say μ, where invariance means that $\mu(f^{-1}A) = \mu(A)$ for each measurable subset A of X. In this book only the most common setting is treated where μ is a probability measure. Typically, a given map f has many invariant measures, and one problem is to single out (often using a variational principle) those which are dynamically most relevant. In particular, for each such measure (X, μ) is a probability space and the scene is set for carrying over concepts from the world of probability to dynamical systems. Note, however, that in contrast to the probabilistic approach, where properties of the underlying probability space play a minor role, the specific structure of X and μ can be of central interest in the context of dynamical systems.