Spin glasses: A mystery about to be solved

Frank den Hollander and Fabio Toninelli
EURANDOM, P.O. Box 513,
5600 MB Eindhoven, The Netherlands

March 1, 2004

Abstract

The study of spin glasses started some 30 years ago, as a branch of the physics of disordered magnetic systems. In the late 1970's and early 1980's it went through a period of intense activity, when experimental and theoretical physicists discovered that spin glasses exhibit new and remarkable phenomena. However, a real understanding of the behaviour of these systems was not achieved and little progress was made in the next 20 years, especially in mathematical terms. In the 1990's various related systems were studied with mounting success, most notably, neural networks and random energy models. Since a couple of years the field has again entered a phase of exciting development. Some of the main mathematical questions surrounding spin glasses are currently being solved and a full understanding is at hand. In this paper we sketch the main steps in this development, which is interesting not only for the physical and the mathematical relevance of this research field, but also because it is an example where scientific progress follows a tortuous path.

Acknowledgment. The authors are grateful to Aernout van Enter (Groningen) for detailed comments on a draft of this paper.

1. Ferromagnets. Let us begin with a brief history of magnetic materials. All matter is composed of a large number of atoms. Atoms carry a *spin*, i.e., a microscopic "magnetic moment" generated by the motion of the electrons around the nucleus. This spin, which in turn generates a microscopic magnetic field around the atom, can be viewed as a vector in three-dimensional space. To simplify matters, assume that for this vector only two opposite directions are allowed, *up* and *down*. In *ferromagnets*, materials capable of attracting pieces of iron placed in their vicinity, each spin has a tendency to align with the spins in its neighbourhood. At high temperature, the motion of the spins is so erratic that at any time about half of them are pointing up and half are pointing down. Consequently, the net macroscopic magnetisation is zero, i.e., the individual microscopic magnetic fields generated by the spins cancel each other out. As the temperature is lowered, the erratic motion of the spins reduces and the spins become more and more sensitive to their mutual interaction. The characteristic feature of ferromagnets is that there is a *critical temperature*, T_c, below which the spins exhibit *collective behaviour* in that a majority of them point in the same direction (either a majority up or a majority down). This phenomenon is called *spontaneous magnetisation* (see Fig 1).

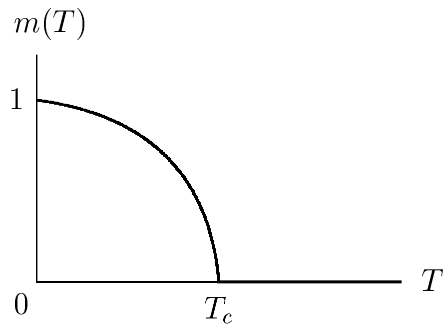

Figure 1: The magnetisation $m(T)$ as a function of the temperature T for a typical configuration of the spins; $m(T)$ is the difference between the number of up-spins and the number of down-spins divided by the total number of spins. By symmetry, configurations with the opposite magnetisation $-m(T)$ are equally likely.

Below T_c the individual microscopic magnetic fields sum up coherently to create a macroscopic magnetic field, which is what is ultimately responsible for the ferromagnet's capability to attract iron. It is important to emphasize that this seemingly natural picture took a long time to emerge – from 1895 (Curie) until 1944 (Onsager) – and that the genius of many illustrious theoretical physicists and mathematicians was necessary in order to fully establish that this is what actually happens.

The microscopic theory that explains the collective behaviour of atoms is called *statistical physics*. According to this theory, a system in equilibrium is described with the help of an energy functional, called *Hamiltonian*, which associates with each microscopic configuration of the system a macroscopic energy. In our case a configuration means a complete list of the orientations of all the spins. If the spins are located at the sites x in a macroscopic box Λ, and if $s_x \in \{+1, -1\}$ denotes the value of the spin at site x ($+1$ for up and -1 for down), then the configuration is

$$s = \{s_x \colon x \in \Lambda\}$$

and the Hamiltonian of the ferromagnet is

$$H(s) = -\sum_{\substack{x,y \in \Lambda \\ x \sim y}} s_x s_y,$$

where $x \sim y$ means that x and y are neighbouring sites. Thus, each pair of neighbouring aligned spins gets energy -1, each pair of neighbouring anti-aligned spins gets energy $+1$. At a given temperature T, the state of the system is described by the *Gibbs distribution* associated with H,

$$\mu_T(s) = \frac{1}{Z_T} e^{-H(s)/kT}, \qquad s \in \{+1, -1\}^{\Lambda},$$

where k is Boltzmann's constant and Z_T normalizes μ_T to a probability distribution: *$\mu_T(s)$ is the probability that the system assumes configuration s.* When T is lowered, μ_T tends to concentrate more and more around the configurations having minimal energy, the so-called *ground states* of the system. For the ferromagnet these ground states are those configurations where all the spins have the same value. Indeed, it is only when $s_x = +1$ for all x or $s_x = -1$ for all x that all terms in $H(s)$ give a negative contribution, leading to the maximal value for $\mu_T(s)$. This maximum is a pronounced peak when T is small, explaining why for low temperature in a typical configuration the majority of the spins is aligned.

2. Spin glasses. Now that we have briefly introduced some important concepts from the theory of magnetism, we are in a position to explain what spin glasses are. Consider a system of spins, as before, but assume that some pairs of neighbouring spins prefer to be aligned, while the others prefer to be anti-aligned. The former are said to have a *ferromagnetic* interaction, the latter an *anti-ferromagnetic* interaction. Say that for any given pair of spins the type of interaction is chosen *randomly* with equal probability. It is because of this randomness in the interactions that such systems are called *disordered*.

In terms of the Hamiltonian, the above model can be defined as

$$H(s) = -\sum_{\substack{x,y \in \Lambda \\ x \sim y}} J_{xy} s_x s_y,$$

where, for each $x \sim y$, J_{xy} can be either $+1$ (indicating a ferromagnetic interaction) or -1 (indicating an anti-ferromagnetic interaction), with probability $\frac{1}{2}$ each. This Hamiltonian was introduced in 1975 by Edwards and Anderson [8], in an attempt to describe a class of disordered magnetic systems found a few years earlier by experimental physicists and termed "spin glasses". Examples in this class are disordered magnetic alloys, i.e., metals containing random magnetic impurities, such as AuFe or CuMn.

What is the analogue in this case of the behaviour depicted in Figure 1? Even at low temperature there is no reason why the majority of the spins should be aligned. Indeed, due to the equal competition between ferromagnetic and anti-ferromagnetic interactions the corresponding magnetisation $m(T)$ will be zero for all T. One might thus conclude that the model simply has no critical temperature and therefore exhibits no interesting phenomena. However, in the early 1970's it was found *experimentally*, by Cannella and Mydosh [6] and by Tholence and Tournier [18], that there still is a critical temperature below which the system undergoes an *ordering transition*, in the sense that the spins act coherently in some sort of way (see Fig. 2). This fact came as a surprise to the physicists.

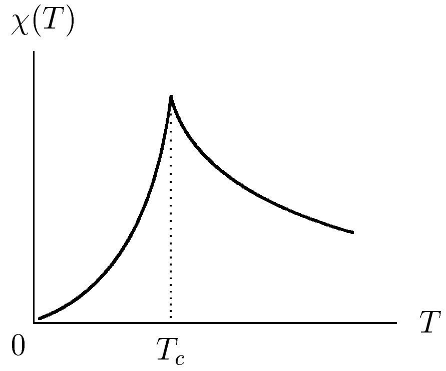

Figure 2: The magnetic susceptibility $\chi(T)$ as a function of the temperature T. $\chi(T)$ measures the sensitivity of the system to the application of a magnetic field and shows a cusp at the critical temperature T_c. This cusp signals a freezing of the spins in random directions.

In simplified terms, what happens is the following. Above T_c, the spins behave essentially independently from one another, i.e., their orientation is hardly influenced by the spins in their neighbourhood. As a result, the typical configurations of the system are those that are completely disordered. This is true both for the ferromagnet and for the spin glass. Below T_c, however, the spins show cooperative behaviour and can be found in *more than one* class of typical configurations. In the case of the ferromagnet described above, there are *two* classes of typical configurations, namely, those having magnetisation $+m(T)$ and $-m(T)$, respectively. These classes of configurations are called *pure states.* In the case of the spin glass, instead, there are *many* pure states, which are not characterised by a non-zero magnetisation, but rather by the occurrence of many "mesoscopic domains" (microscopically large but macroscopically small) in which the spins show some form of "local magnetic order". In fact, a whole "hierarchy" of such domains occurs. At present it is not yet clear what the features of these domains precisely are. The important point, however, is that the existence of a transition at T_c is experimentally observable.

The Edwards-Anderson model is far too difficult to be analysed theoretically in detail, even today. In fact, condensed matter physicists have been disputing heatedly in the past three decades about what precisely happens at low temperature. In 1975 Sherrington and Kirkpatrick [15] introduced a simplified version of this model. The difference with the Edwards-Anderson model is that each spin is influenced not only by its neighbouring spins, but by *all* the spins in the system. The corresponding Hamiltonian reads

$$H(s) = -\frac{1}{|\Lambda|^{1/2}} \sum_{\substack{x,y\in\Lambda \\ x\neq y}} J_{xy} s_x s_y,$$

where J_{xy} is $+1$ or -1, with probability $\frac{1}{2}$ each, for all $x \neq y$ (rather than for $x \sim y$ only), and a factor $1/|\Lambda|^{1/2}$ is added to normalise the interaction. In statistical physical jargon, the Sherrington-Kirkpatrick model is a *mean-field approximation* of the Edwards-Anderson model. Strange as it may seem, this type of approximation actually makes the model easier.

For a history of spin glasses up to 1986, we refer to Binder and Young [2].

3. Replica symmetry breaking. The article by Sherrington and Kirkpatrick carried the rather innocent title "A solvable model of a spin glass". The authors never imagined that they were giving birth to one of the most exciting enigmas of modern statistical physics. The solution they proposed, assuming so-called "replica symmetry", turned out to be incorrect, and even self-contradictory as they themselves realised very well. It was only a few years later, in 1980, that the Italian theoretical physicist Giorgio Parisi [14] proposed a different solution, known as the *continuous replica symmetry breaking scheme*, which could account for many of the experimental observations (both laboratory experiments and computer simulations).

Replica symmetry breaking theory predicts the existence of a collective behaviour with many *exotic features*, never before observed in any physical system. In simple words, Parisi's theory predicts that the Hamiltonian of the Sherrington-Kirkpatrick model has *many* ground states (growing in number as the volume of the system increases), which are *highly disordered* and which do not seem to be related to one another via simple transformations. In contrast, recall that the ferromagnetic Hamiltonian has only *two* ground states, one with all spins up and one with all spins down, which are *fully ordered* and which are related to one another via a global inversion of all the spins. Moreover, it turns out that for the Sherrington-Kirkpatrick model, by choosing a *different realisation of the disorder* (i.e., a different choice for the random interactions $J_{xy} = \pm 1$, again with probability $\frac{1}{2}$ each), the new ground states in general have nothing to do with the old ones. Even more surprisingly, if the disorder realisation is kept fixed but the volume of the system is increased, then the new ground states are not related to the old ones either ("chaotic size dependence"). In spite of this extremely irregular situation, according to Parisi's theory the collection of all the ground states has some regular, highly non-trivial, geometrical structure, called *ultrametricity*, which is *not* modified when the disorder realisation is changed.

So, what distinguishes the region above the critical temperature T_c from the one below, for the Sherrington-Kirkpatrick model? Suppose that we take two copies – two *replicas* – of the system, with the *same* realisation of the disorder, and compute the *overlap* between them, i.e.,

$$q(s^{(1)}, s^{(2)}) = \frac{1}{|\Lambda|} \sum_{x \in \Lambda} s_x^{(1)} s_x^{(2)},$$

where $s^{(1)}$ and $s^{(2)}$ are the configurations of the first and the second replica, respectively. Then, above T_c the overlap is zero for typical configurations (typical with respect to the Gibbs distribution and the disorder realisation), while below T_c it can assume a range of non-zero *random* values. This can be explained as follows. Recall that, at low temperature, the Gibbs distribution is peaked around the ground states of the system. Consequently, the configurations in the two replicas will each be very close to one of the ground states (not necessarily the same one), which causes a non-zero overlap. Due to the erratic nature of the ground states, the overlap does not have a fixed value: it varies randomly with the ground states.

Replica symmetry breaking theory came as a shock to the physics community, not only for the novelty of the phenomena predicted, but also for the way in which it was presented. It happens frequently that theories formulated by physicists are not mathematically rigorous, and contain a number of assumptions and simplifications that need to be justified. Often full mathematical proofs come only much later. Here the situation was more delicate: the works of Parisi and co-workers were not only non-rigorous, they were based on such strange and such daring techniques that it was hard to see how the relevant statements could be formulated in a proper mathematical language. This is why part of the mathematics community has regarded Parisi's theory as somewhat magic. Still, the phenomena predicted by the theory were so appealing, and its range of applications so wide, that it soon became a standard tool for theoretical physicists, who were much more excited by its power than worried by its lack of mathematical sense and precision. One could say that Parisi had discovered a new world.

A review of the results of replica symmetry breaking theory up to 1987 can be found in Mézard, Parisi and Virasoro [12].

4. Towards a solution. The reader might wonder at this point whether all the excitement about the Sherrington-Kirkpatrick model is really justified. After all, it is only an approximate version of the more difficult – but more realistic – Edwards-Anderson model, which remains unsolved. In fact, it is not yet clear how much we really learn about the Edwards-Anderson model from a detailed analysis of the Sherrington-Kirkpatrick model. According to a scenario put forward by Newman and Stein (see Newman [13]), the behaviour of the two models may well turn out to be qualitatively different: the main phenomena related to replica symmetry breaking may not occur in "short range" models like the Edwards-Anderson model. Still, the excitement is understandable. First, the study of the Sherrington-Kirkpatrick model has taught us a lot and continues to do so. In the attempts to understand this model, new ideas and techniques have been invented and further developed that are extremely interesting and that have turned out to be fruitful for other statistical physical models as well. Second – and more importantly – it has gradually become clear that the knowledge gained through the analysis of the Sherrington-Kirkpatrick model can be applied to a variety of – apparently unrelated – problems in mathematics, physics and engineering. These problems have therefore come to be considered as belonging to the realm of spin glasses. Examples are neural networks (models for memory and learning), error correcting codes (used in communications engineering to recover the information transmitted through a noisy channel) and random combinatorial optimisation (problems of decision in the presence of many mutually competing requirements).

From the moment the replica symmetry breaking theory came into being, trying to prove – or to disprove – the predictions of Parisi and co-workers became an exciting challenge for many among the best mathematical physicists. The task proved to be quite hard and quite frustrating, and for almost 20 years progress was painfully slow. Much effort was devoted to the search for and the study of mathematical models that would be easier than the Sherrington-Kirkpatrick model, but that would still exhibit replica symmetry breaking effects. In particular, the Generalized Random Energy Model, introduced by Derrida [7] in 1985, shows striking similarities with the Sherrington-Kirkpatrick model, yet is exactly solvable. The structure of the Gibbs distribution in this model has been analysed in full mathematical detail by Bovier and Kurkova [4]. Similarly, extensive rigorous results have been obtained by Bovier, Gayrard and Picco for the Hopfield model of neural networks (see [3] and references therein). The latter is a paradigm for auto-associative memory, i.e., systems that try to recognize words – or *patterns* – that were previously memorized. In this case, the spins should be interpreted as the states of the neurons located at the various sites: $s_x = +1$ if the neuron at site x is sending electric pulses, $s_x = -1$ if it is not. When varying the number of memorized patterns, the behavior can range from a ferromagnetic type to a spin glass type. For an overview of the expanding panorama of spin glasses up to 1998, see Bovier and Picco [5].

It gradually became clear – more through failures than through positive results – that completely new ideas were needed to make significant progress in the comprehension of replica symmetry breaking. It is only in the last few years that we are witnessing a rapid and unexpected boost in the mathematical understanding of the key questions. Surprisingly, the missing new ideas turned out to be relatively simple, although they were quite hard to find. The first steps in this breakthrough were taken in 2000-2002 by the Italian mathematical physicist Francesco Guerra [10], together with Fabio Toninelli [11], building on earlier work by Ghirlanda and Guerra [9]. As a result, some of the mathematical questions that had been tackled in vain in the preceding 20 years could finally be solved. One important result is the existence of the "thermodynamic limit" for the Sherrington-Kirkpatrick model. This means that physical quantities, like the energy of the ground states divided by the volume of the system, converge to a well defined limit when the volume of the system tends to infinity. The proof of this fact is quite standard in statistical physics for models with "short range" interactions, but it is not for mean-field models, especially not for disordered ones. Another important result is that with the help of certain rigorous comparison identities – so-called *sum rules* – the thermodynamic properties of the Sherrington-Kirkpatrick model can be compared with the corresponding expressions given by Parisi's theory. These sum rules concern the *free energy* $f(T)$ as a function of the temperature T, a quantity of central importance in statistical physics, from which all thermodynamic properties of the system can be deduced. This free energy is related to the Gibbs distribution μ_T via the relation $f(T) = -kT \log Z_T$. The result is that $f(T)$ can be related to the free energy predicted by Parisi's theory via an identity of the type

$$f(T) = f^{Parisi}(T) + R(T, |\Lambda|),$$

where $R(T, |\Lambda|)$ is an "error term" that not only depends on the temperature T but also on the volume $|\Lambda|$ of the system. Proving the validity of Parisi's theory is equivalent to showing that $R(T, |\Lambda|)$ tends to zero in the thermodynamic limit $|\Lambda| \to \infty$. A particularly important fact is that $R(T, |\Lambda|)$ turns out to be non-negative, so that Parisi's free energy at least is a lower bound for $f(T)$, a fact that itself is rich in physical implications (see Toninelli [19]). Subsequently, Aizenman, Sims and Starr [1] obtained Guerra's sum rules through a general variational principle and showed that Parisi's free energy arises from a restriction of this variational principle to "ultrametric structures". This restriction is optimal precisely when replica symmetry breaking theory correctly describes the Sherrington-Kirkpatrick model.

These new ideas provoked great excitement in the scientific community, and new feverish work began. The last part of this story is still in progress and is keeping the excitement high. In July 2003 the French mathematician Michel Talagrand, who has been working on the problem intensively and who has introduced many new ideas in this field since the mid 1990's (see [16]), has announced (see [17]) that he was able to complete the mathematical proof of Parisi's solution, extending the method of sum rules invented by Guerra. The details of the proof have not yet been made public, but it is not hard to imagine the impression this announcement has produced on the experts. It seems that the full mathematical justification of Parisi's theory, explaining the mysterious features of the Sherrington-Kirkpatrick model, is finally at hand.

Currently, research in this rapidly evolving field is being carried out by a number of groups, including the Interacting Stochastic Systems group at EURANDOM, the European institute for research on stochastic phenomena located at the Technical University of Eindhoven, The Netherlands. Fabio Toninelli works as a postdoc in the ISS-group. Frank den Hollander is supervisor of the ISS-group and scientific director of EURANDOM.

References

[1] M. Aizenman, R. Sims and S. Starr, *An extended variational principle for the SK spin-glass model*, Phys. Rev. B68 (2003) 214403.

[2] K. Binder and A.P. Young, Spin glasses: Experimental facts, theoretical concepts, and open questions, Rev. Mod. Phys. 58 (1986) 801–976.

[3] A. Bovier and V. Gayrard, Hopfield models as generalized random mean field models, in: [5], pp. 3–89.

[4] A. Bovier and I. Kurkova, Rigorous results on some simple spin glass models, Markov Proc. Related Fields 9 (2003) 209–242.

[5] A. Bovier and P. Picco (eds.), *Mathematical Aspects of Spin Glasses and Neuronal Networks*, Progress in Probability 41, Birkhäuser, Basel, 1998.

[6] V. Cannella and J.A. Mydosh, Magnetic ordering in gold-iron alloys, Phys. Rev. B 6 (1972) 4220–4237.

[7] B. Derrida, A generalization of the random energy model which includes correlations between energies, J. Physique Lett. 46 (1985) L401–L407.

[8] S.F. Edwards and P.W. Anderson, Theory of spin glasses, J. Phys. F: Metal Phys. 5 (1975) 965–974.

[9] S. Ghirlanda and F. Guerra, General properties of overlap distributions in disordered spin systems. Towards Parisi ultrametricity, J. Phys. A31 (1998) 9149–9155.

[10] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, Commun. Math. Phys. 233 (2003) 1–12.

[11] F. Guerra and F.L. Toninelli, The thermodynamic limit in mean field spin glass models, Commun. Math. Phys. 230 (2002) 71–79.

[12] M. Mézard, G. Parisi and M.A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987.

[13] C.M. Newman, *Topics in Disordered Systems*, Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel, 1997.

[14] G. Parisi, The order parameter for spin glasses: a function on the interval $[0, 1]$, J. Phys. A 13 (1980) 1101–1112.

[15] D. Sherrington and S. Kirkpatrick, A solvable model of a spin-glass, Phys. Rev. Lett. 35 (1975) 1792–1796.

[16] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians*, Ergebnisse der Mathematik und ihrer Grenzgebiete 46, Springer, Berlin, 2003.

[17] M. Talagrand, The generalized Parisi formula, C. R. Acad. Sci. Paris, Ser. I 337 (2003) 111–114.

[18] J.L. Tholence and R. Tournier, Susceptibility and remnent magnetization of a spin glass, J. Phys. (Paris) Colloq. 35 (1974) C4–229.

[19] F.L. Toninelli, About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean field spin glass model, Europhys. Lett. 60 (2002) 764–767.