International Journal of Theoretical Physics, Vol. 20, No. 7, 1981

Mathematical Foundations of Quantum Field Theory: Fermions, Gauge Fields, and Supersymmetry Part I: Lattice Field Theories

David A. Edwards

Department of Mathematics, University of Georgia, Athens, Georgia

Received November 5, 1980

In this paper we explore the mathematical foundations of quantum field theory. From the mathematical point of view, quantum field theory involves several revolutions in structure just as severe as, if not more than, the revolutionary change involved in the move from classical to quantum mechanics. Ordinary quantum mechanics is based upon real-valued observables which are not all compatible. We will see that the proper mathematical understanding of Fermi fields involves a new concept of probability theory, the graded probability space. This new concept also yields new points of view concerning ergodic theorems in statistical mechanics.

1. INTRODUCTION

In Edwards (1981a) we presented a systematic approach to the mathematical foundations of quantum mechanics along the lines originally pioneered by von Neumann. This approach requires a substantial mathematical background that most physicists find unpleasant (e.g., measure theory). They prefer to sacrifice mathematical rigor—and even consistency—in exchange for a flexible formalism that is more readily learned. Thus physicists follow Dirac and not von Neumann.

In this paper we explore the mathematical foundations of quantum field theory. From the mathematical point of view, quantum field theory involves several revolutions in structure just as severe as the revolutionary change involved in the move from classical to quantum mechanics, if not more so. Ordinary quantum mechanics is based upon real-valued observables which are not all compatible. We will see that the proper mathematical understanding of Fermi fields involves a different conception of probability theory. This new conception will also yield new points of view concerning the need for ergodic theorems in statistical mechanics.

Fermi fields were first introduced by Jordan and Wigner in 1928 by the simple formal move of replacing commutators by anticommutators. They were viewed as annihilation and creation operators for Fermi particles in a "second-quantized" formalism. From our point of view, Fermi fields represent a totally different structure from that of second quantization. The physicist's approach allows him to move fluidly between what are for the mathematician radically different models. The structures and viewpoints we will explicate are at least tacitly present in much of the physics literature (e.g., the work of Schwinger). It was our struggle to comprehend from a mathematical viewpoint Kadanoff (1977) which led most directly to our present position.

In this first paper we will restrict our attention to lattice field theories, leaving the much more sophisticated continuum models to Paper II. In Section 2.2 we introduce gauge fields and in Section 2.3 Fermi fields. In Section 2.4 we discuss phase transitions and critical phenomena. In Section 2.5 we discuss thermodynamic limits. We conclude this paper with some comments on the Wilson model of quarks and strings on a lattice.

2. LATTICE FIELD THEORIES

2.1. The Ising Model. Since its introduction in the 1920s by Lenz and Ising, the Ising model has become one of the most important models in modern physics because it is the simplest nontrivial model of cooperative phenomena. The model consists of a finite collection $\sigma=\{\sigma_i|1\leqslant i\leqslant N\}$ of compatible two-valued observables σ_i, where the index i is viewed to vary over a square lattice $\mathcal{L}$ of side L ($N=L^d$, where d is the dimension of the lattice). If the values of σ_i are denoted by -1 and $+1$, then a configuration of σ determines a point in $\mathcal{C}=\{-1,+1\}^N$, and a joint probability measure μ for the σ_i's determines a probability measure on $\mathcal{C}$. From the state μ one can compute various moments such as $\langle\sigma_i\rangle$, $\langle\sigma_i\sigma_j\rangle$, etc. These moments can be empirically determined (at least approximately) by choosing a large ensemble of identically prepared systems (i.e., each is in a state determined by a prior probability distribution μ) and computing the average over the ensemble of the values empirically obtained for σ_i, $\sigma_i\sigma_j$, etc. (Of course, the actual moments $\langle\sigma_i\rangle$, $\langle\sigma_i\sigma_j\rangle$ would only be obtained in this way with probability one in the limit of an infinitely large ensemble.) In this way we obtain new observables $\langle\sigma_i\rangle$, $\langle\sigma_i\sigma_j\rangle$, etc., which we shall call second-order observables. In practice physicists usually do not measure the σ_i directly but instead do scattering experiments (e.g., neutron scattering in ferromagnets or light scattering in gases) which yield a scattering Green's function $G(k)$ which they interpret as the Fourier transform of $\langle \sigma_i \sigma_{i+r} \rangle$. This leads to two possible different views of the moments $\langle \sigma_i \sigma_{i+r} \rangle$. In the usual view, $\langle \sigma_i \sigma_{i+r} \rangle$ is not itself obtainable in a single experiment, and what one is actually doing in scattering experiments is making a sample average of $\sigma_i \sigma_{i+r}$ and invoking an (unproved) ergodic theorem to relate sample and ensemble averages. An alternative view is that the scatter directly couples to the *disposition* μ and thence directly yields the moments $\langle \sigma_i \sigma_{i+r} \rangle$. This view will be necessary when we study Fermi fields. In any case, the Green's functions are the objects of central interest to the physicist.

If we have a family of states μ_α depending upon a parameter α varying, say, in R^n, then we would naturally be interested in the analytic properties of $\langle \sigma_i \rangle_\alpha$, $\langle \sigma_i \sigma_j \rangle_\alpha$, etc., as a function of the parameter α. In this way we could introduce new (third-order) observables

$$\frac{d}{d\alpha} \langle \sigma_i \rangle_\alpha \Big|_{\alpha=\alpha^0}, \qquad \frac{d}{d\alpha} \langle \sigma_i \sigma_j \rangle_\alpha \Big|_{\alpha=\alpha^0},$$

$$\frac{\partial^2}{\partial \alpha_1 \partial \alpha_2} \langle \sigma_1 \rangle_{(\alpha_1, \alpha_2)} \Big|_{(\alpha_1, \alpha_2)=(\alpha_1^0, \alpha_2^0)}, \qquad \text{etc.}$$

In statistical mechanics particular attention is focused upon the probability measures first introduced by Gibbs. In the context of the Ising model, the Gibbs distributions are defined as follows: (a) The microcanonical distribution

$$\mu^M(\sigma) = \frac{1}{\#\{\sigma \mid H(\sigma) = E\}}$$

to each σ such that $H(\sigma) = E$, where H is the Hamiltonian of the system—(e.g., in the simplest case taken to be

$$H(\sigma) = -J \sum_{\substack{\text{nearest} \\ \text{neighbor}}} \sigma_i \sigma_j - h \sum_i \sigma_i$$

(b) The canonical distribution corresponding to a fixed mean energy $\langle H \rangle = E$ is given by

$$\mu^c(\sigma) \equiv \frac{e^{-\beta H(\sigma)}}{\sum_\sigma e^{-\beta H(\sigma)}}$$

where β is chosen so that

$$\langle H\rangle=\frac{\Sigma_\sigma H(\sigma)e^{-\beta H(\sigma)}}{\Sigma_\sigma e^{-\beta H(\sigma)}}=E$$

(Note that since we want μ_β^c to be a probability measure, it is necessary that β be an extended real number, but *nothing* in the model requires that $\beta>0$. This suggests that one should look for phenomena for which it would be appropriate to ascribe a "negative" temperature and develop the appropriate thermodynamics.) (c) The grand canonical distribution corresponding to a fixed mean energy $\langle H\rangle=E$ and a fixed mean number of particles $\langle N(\sigma)\rangle=\langle\Sigma\sigma_i\rangle=N$ (here we have shifted to a lattice gas language with $\sigma_i\in\{0,1\}$) is given by

$$\mu^G(\sigma)=\frac{\exp\{-[H(\sigma)-N(\sigma)]\}}{\Sigma_\sigma\exp\{-\beta[H(\sigma)-\bar{\mu}N(\sigma)]\}}$$

where β and $\bar{\mu}$ are chosen so that $\langle H\rangle=E$ and $\langle N(\sigma)\rangle=N$.

There are various ways of attempting to motivate these distributions. These distributions can be shown to minimize "information" (equivalently, maximize entropy) subject to their respective constraints. Furthermore, if one assumes that the microcanonical distribution applies to a large system, limit arguments lead to the canonical or grand-canonical distributions for the description of certain small subsystems. But the way we favor most of motivating these distributions is to simply recognize the use of these Gibbs measures as a basic postulate concerning the manner of preparation of the system. Alternatively, from the point of view of pure probability theory, these Gibbs measures define simple models of families of random variables which are strongly dependent upon their neighbors, and satisfy a generalized Markov property.

To the Gibbs canonical and grand canonical distributions one associates the partition functions

$$Z^c\equiv\sum_\sigma\exp[-\beta H(\sigma)]=\exp(-\beta Vf)\quad\text{and}$$

$$Z^G\equiv\sum_\sigma\exp\{-\beta[H(\sigma)-\bar{\mu}N(\sigma)]\}=\exp(\beta Vp)$$

where f is defined to be the free energy per unit volume, and p is defined to be the pressure. Furthermore, one associates β with inverse temperature (i.e., $\beta=1/kT$, with k equal to Boltzmann's constant) and $\bar{\mu}$ with the chemical potential. In this way one can associate traditional thermodynamical variables such as temperature and pressure with the *totally static* Ising model. Note that while pressure and temperature are observables in thermodynamics, their occurrence in the Ising model is more complex—at best, they can be related to higher-order observables. The introduction of the partition function Z^c and Z^G can be motivated without appealing to thermodynamics by observing that $\ln Z^c$ and $\ln Z^G$ provide generating functions for various correlation functions occurring in the model, e.g.,

$$\langle H\rangle = -\frac{\partial}{\partial\beta}\ln Z^c, \qquad \langle N\rangle = \frac{1}{\beta}\frac{\partial}{\partial\mu}\ln Z^G,$$

$$\langle \Sigma\sigma_i\rangle = \frac{1}{\beta}\frac{\partial}{\partial h}\ln Z^c, \qquad \text{etc.}$$

Obvious extensions of the Ising model are obtained by considering more general lattices, Hamiltonians, and spin variables (i.e., allowing σ_i to take on values in a range X more general than Z_2, e.g. $X=R$ or S^2). One can also consider quantum mechanical analogs where not all the σ_i are compatible; in this case one uses Gibbs–von Neumann density matrices such as $\rho^c = e^{-\beta H(\sigma)}/\mathrm{Tr}(e^{-\beta H(\sigma)})$ with its associated partition function $Z^c \equiv \mathrm{Tr}(e^{-\beta H(\sigma)})$.

2.2. Gauge Fields. We shall now consider some less obvious extensions of the Ising model. Suppose each σ_i takes values in a set X_i such that each X_i has two elements. In order to define an interesting Hamiltonian which couples nearest-neighbor spins σ_i we may assume given for each i a map $\langle -,-\rangle_i\colon X_i\times X_i\to R$ and for each nearest-neighbor pair i, j an isomorphism $U_{ij}\colon X_i\to X_j$. Given such data, a simple Hamiltonian is $H(\sigma)=\Sigma_{\text{n.n.}}\langle U_{ij}\sigma_i,\sigma_j\rangle_j$. From H we obtain a Gibbs measure and partition function in the standard way. One should note carefully that the $\langle -,-\rangle_i$ and U_{ij} are part of the underlying geometry of the problem (determining an inner product and connection on the bundle $X\to\mathfrak{L}$) and *not* new observables. We have no new Green's functions in this model. On the other hand, the inner product $\langle -,-\rangle$ and connection U will reveal themselves indirectly through changes in the various moments $\langle\sigma_i\sigma_j\rangle$ [more precisely, one must choose functions $f_i\colon X_i\to R$, and then one can form moments $\langle f_i(\sigma_i)\cdot f_j(\sigma_j)\rangle$]. Hence, they are higher-order observables similar in nature to parameters occurring in the Hamiltonian. (One might have *other* direct experimental means of determining the geometry.) U is usually called a gauge field or string variable and often mistakenly considered to be a first-order observable. In most models the choice of connection U is not itself fixed, but instead considered to be determined by a probability measure μ on the space of connections $\mathcal{C}=\text{conn}(X\to\mathcal{L})$, and the Green's functions are further averaged over μ, i.e., $G(i, j)=\langle\langle\sigma_i\sigma_j\rangle_U\rangle\mu$. In this way one has a chance of obtaining translation-invariant Green's functions. A typical choice for μ in this model where the space of connections is finite is determined by Gibbs weight function $e^{A(U)}$, where the $A(U)$ depends only on the "curvature" of U. For example, the Yang–Mills action $A^{\text{YM}}(U)$ is defined as follows. A closed loop $P=(i, j, k, l\cdot i)$ of nearest-neighbor bonds is called a plaquette. Let $\text{Iso}(X_i)$ be the two-element group of isomorphisms of X_i. For each i let χ_i: $\text{Iso}(X_i)\to R$ take the identity to 1 and the switch map to -1. For a plaquette P defined $A_P=\chi_i(U_{li}U_{kl}U_{jk}U_{ij})$, $A^{\text{YM}}(U)=(1/2g_0^2)\Sigma A_p$, where g_0^2 is the Yang–Mills coupling constant and the sum is over all plaquettes in $\mathcal{L}$. In more general models one must use Haar measure on the gauge group G in order to define a background measure on the space of connections $\mathcal{C}$ and also choose characters χ_i on $\text{Iso}(X_i)\simeq G$ [one then takes $A_P = \text{Re}\,\chi_i(U_{li}U_{kl}U_{jk}U_{ij})$]. Since μ depends upon the curvature of U, one can to some extent determine the local curvature of U by measuring the Green's functions. Thus, the curvature of U can also be considered a higher-order observable. From this point of view, one sees that the electromagnetic field, which is mathematically identified with the curvature of a connection, is *not* a first-order observable and hence one should *not* try to "quantize" it when one does quantum electrodynamics.

To stress this point we shall now discuss a gauged version of the Heisenberg model. The Heisenberg model is defined by associating to each lattice point i a partial algebra A_i of real-valued observables. In the simplest case A_i is assumed to be isomorphic to the partial algebra of Hermitian elements in $L(\mathbf{C}^2)$, where a family of elements in $L(\mathbf{C}^2)$ are called compatible (or simultaneously observable) if they lie in a commutative subalgebra of $L(\mathbf{C}^2)$. Each of the A_i has a natural embedding in the tensor product algebra $A=\otimes_{i\in\mathcal{L}}L(\mathbf{C}^2)_i$. Taking the tensor product of the natural traces on each $L(\mathbf{C}^2)_i$ yields a trace $\text{Tr}=\otimes_{i\in\mathcal{L}}\text{Tr}_i$ on $\otimes_{i\in\mathcal{L}}L(\mathbf{C}^2)_i$. Of special interest are the Pauli matrices

$$\sigma_x=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\qquad \sigma_y=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\qquad \sigma_z=\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix}$$

The Hamiltonian for the Heisenberg model is defined by

$$H=-\sum_{\substack{\text{nearest}\\ \text{neighbors}}}\sigma_x^i\sigma_x^j+\sigma_y^i\sigma_y^j+\sigma_z^i\sigma_z^j$$

The main Green's functions of interest are of the following type:

$$G(i,x;\,j,y)=\langle\sigma_x^i\sigma_y^j\rangle=\frac{\text{Tr}\left(\sigma_x^i\sigma_y^j e^{-H}\right)}{\text{Tr}(e^{-H})}$$

(Note that one may obtain an alternate description of the Ising model by choosing A_i isomorphic to the algebra with generators 1 and σ with $\sigma^2=1$ and defining $\mathrm{Tr}_i\,\sigma=0$ and $\mathrm{Tr}_i\,1=1$.) In order to gauge this model we must assume that the observables at i take values not in R but instead in a one-dimensional real vector space X_i. By *choosing* for each i an isomorphism $f_i\colon X_i\to R$, we can pull back the Pauli matrices $\sigma_x^i, \sigma_x^i, \sigma_z^i$ to obtain special X_i-valued observables $\bar\sigma_x^i, \bar\sigma_y^i, \bar\sigma_z^i$. Suppose, furthermore, that each X_i has an inner product $\langle -,-\rangle$ and that the bundle $X\to\mathfrak{L}$ has a connection U. Then, if i and j are nearest neighbors, $\bar\sigma_x^i$ and $\bar\sigma_y^j$ determine a *real*-valued observable by taking the inner product of the result of measuring $\bar\sigma_y^j$ (which is in X_j) with the image under $U(i,j)$ of the result of measuring $\bar\sigma_x^i$. Denote this new real-valued observable by $\langle U(i,j)\bar\sigma_x^i, \bar\sigma_y^j\rangle$. Since these new observables define elements $A=\otimes L(\mathbb{C}^2)_i$, one can form the Hamiltonian

$$H=-\sum_{\text{n.n.}}\langle U(i,j)\bar\sigma_x^i,\bar\sigma_x^j\rangle+\langle U(i,j)\bar\sigma_y^i,\bar\sigma_y^j\rangle+\langle U(i,j)\bar\sigma_z^i,\bar\sigma_z^j\rangle$$

This Hamiltonian defines new moments. But we can go further and also average over the choice of connections using the Yang–Mills action. We thus obtain the Green's functions of the gauged Heisenberg model. Note that although the Heisenberg model is a quantum mechanical model, there is nothing quantum mechanical about the gauge field.

2.3. Fermi Fields. In order to obtain a clear conception of the meaning of Fermi fields we will first introduce the notion of graded probability theory à la Kostant (1977). Graded probability theory is an extension of classical probability theory in a *different* direction from that of quantum probability theory. The main idea is to extend the commutative algebra of complex-valued observables $M(X)$ to a Z_2-graded-commutative algebra $A(X)$ whose quotient by its ideal of nilpotent elements $A'(X)$ yields back $M(X)$, i.e., we have a short exact sequence

$$0\to A'(X)\to A(X)\to M(X)\to 0$$

of graded commutative algebras. Furthermore, one assumes that $A(X)$ is (noncanonically) isomorphic to the tensor product of $M(X)$ with an exterior algebra $\wedge^*(\mathbb{C}^n)$. The elements of $A(X)$ can*not* be viewed as observables yielding values in $\wedge^*(\mathbb{C}^n)$ (or in anywhere else, for that matter). The only tie this model has to experiment is that we will be able to associate Green's functions to it. This yields quite an exotic conception of a field theory. For physicists, quantum field theory consists of a family of Green's functions, and for each such function an algorithm for generating certain graphs called Feynman diagrams, together with a procedure which associates to each diagram a formal analytic expression, together with (renormalization) techniques for converting the formal analytic expressions into finite numbers. All of the above combine to yield asymptotic expansions for the Green's functions. Graded probability theory yields a systematic procedure for generating the physicists' formulation. The relationship is somewhat analogous to the relationship between Newtonian and Ptolemaic astronomy (especially if you believe that gravity is an "occult" quantity).

The simplest examples of graded probability spaces occur as follows. If $V \to X$ is a complex vector bundle over X, then we can take $A(X)$ to be the space of sections of $\wedge^*(V)$. If $V = W \oplus \overline{W}$, where $\overline{W}$ is the conjugate bundle of the Hermitian bundle W^n, then V has a canonical orientation $\nu \in \wedge^{2n}(V)$. We thus obtain a canonical (linear) isomorphism $\mathbb{F}$: $\wedge^{2n}(V) \to M(X)$. $\mathbb{F}$ extends to a degree 0 linear function $\mathbb{F}$: $A(X) \to M(X)$ which is called the Fermionic integral. A measure on X determines a linear functional $\int$: $M(X) \to \mathbb{C}$. The quadruple $(X, A(X), \mathbb{F}, \int)$ will be called a Fermionic measure space. $A(X)$ comes with a natural conjugation extending the natural conjugation on $M(X)$ (e.g., if $w_1 \in W$ and $\bar{w}_2 \in \overline{W}$, then $\overline{w_1 w_2} = w_2 \wedge \bar{w}_1$). If $H = \overline{H} \in A(X)$, then we can form Green's functions $G(f) = (1/Z)\int \mathbb{F}(fe^{-H})$, where Z is the partition function $\int \mathbb{F}(e^{-H})$ and $f \in A(X)$. The case (pt., $\wedge^* V, \mathbb{F}, \delta_{\text{pt.}}$) is of particular importance. The elements of the graded Lie algebra of graded derivations of $\wedge^*(V)$ are called infinitesimal supersymmetry transformations. Those derivations of degree one are said (in the physics literature) to relate bosons to fermions, i.e., to map $\wedge^{\text{ev}}(V)$ to $\wedge^{\text{odd}}(V)$

"Quantized" versions of Fermionic measure spaces are simply "diagrams" (see Edwards, 1981a) of the form $\{(X_\alpha, A(X_\alpha), \mathbb{F} \int)\}$. One simple way to obtain such a diagram is to start with a noncommutative C^* algebra B and a state ρ on B, tensor B by $\wedge^*(V)$ to obtain the graded algebra $B \otimes \wedge^*(V) \xrightarrow{\pi} B$. Then

$$\{(X_\alpha, A(X_\alpha), \mathbb{F}, \rho|_\alpha) A(X_\alpha) = \pi^{-1}(\alpha),$$

where α is a commutative subalgebra of normal elements in B and X_α is the maximal ideal space of $\alpha\}$

is a diagram of Fermionic measure spaces. Note that in this case the elements f of $B \otimes \wedge^*(V)$ do *not* have any simple interpretation as observables, but can only be related to experiment via the Green's functions

$$G(f) = \frac{1}{Z} \rho \cdot \mathbb{F}(fe^{-H})$$

(There may be more general possible choices for G than those given by this formula; this leads to a graded Gleason problem.) A family $\{f_i\}$ of elements of $B\otimes\wedge^*(V)$ will be said to be *compatible* if they lie in a graded commutative subalgebra of $B\otimes\wedge^*(V)$. Thus in the usual relativistic quantum field theory of fermions the vanishing of the anticommutators $\{\psi_i(f),\psi_j(g)\}$ when f and g has spacelike separated supports means that $\psi_i(f)$ and $\psi_j(g)$ are *compatible* (i.e., Einstein causuality holds), and hence one can define Green's functions $\langle\psi_i(f)\bar{\psi}_j(g)\rangle$.

Lattice Fermionic field theories are now defined by associating to each lattice point $i\in\mathfrak{L}$ a Fermionic measure space $(X_i, A(X_i), \mathcal{F}, \int)$ [or, in quantum theories, a diagram of such spaces, or, more generally, a diagram $\{(X_i, A(X_i), \int\mathcal{F})\}$, where $\int\mathcal{F}\colon A(X_i)\to C$]. Consider the simplest case where each X_i is a point and $A(X_i)=\wedge^*(V_i)=\wedge^*(\overline{W_i}\oplus\overline{W}_i)$. Let U be a connection on the bundle $W\to\mathfrak{L}$. Then a simple choice for the Hamiltonian is

$$H=\sum_{\text{n.n.}}\pm\left[U(i,j)\psi_i^{\alpha}\right]\bar{\psi}_j^{\alpha}\in\left(\text{pt.},\ \bigotimes_{\mathfrak{L}}\wedge^*(V_i),\mathcal{F},\delta\right)$$

$$=\prod_{\mathfrak{L}}(X_i, A(X_i),\mathcal{F},\delta)$$

where ψ_i^{α} is a basis for W_i, and one chooses the sign $\pm$ depending upon whether j is to the left or right of i. In forming the Green's functions one may also wish to average over the choice of connection using the Yang–Mills action.

2.4. Phase Transitions and Critical Phenomena. One of the most striking common occurrences is the change of water into steam when its temperature is raised. Water and steam have vastly different properties; yet the mind tends to assume a single underlying substance whose properties have undergone a very rapid change. Atomic theory identifies the underlying substance as H_2O. If one studies the density of H_2O as a function of pressure and temperature, then one finds that for most choices of P and T one obtains a unique value of ρ, but for certain special values of P and T there are two or even three possible values of ρ corresponding to the different phases of H_2O called ice, water, and steam. This situation is summarized in Figure 1. Experimentally, it seems that ρ is a nice function of P and T except across the critical curve where it has a jump discontinuity. At the critical point, c, ρ is continuous. By going around c, one can change water into steam gradually, i.e., without there ever occurring the striking phenomena associated with a phase transition.

The magnetic properties of iron are a second prime example of critical phenomena. Here, one studies how the magnetic vector m depends upon the

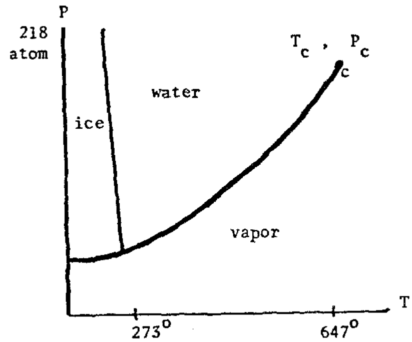

Fig. 1.

external magnetic fields h and the temperature T. For $h=0$ and $T<T_c$ there are two possible choices of m. One obtains the diagram shown in Figure 2. In Figure 2 the critical curve is the interval $[0, T_c]$ on the T axis. There are many other similar phenomena occurring throughout contemporary science.

Let us look more closely at some of the phenomena occurring near the critical point. The spontaneous magnetization is defined by $m_0(T)= \lim_{h\to 0^+} m(h,T)$, and one discovers that for $T<T_c, m_0(T)>0$. More precisely, one obtains a curve having the form shown in Figure 3. This figure suggests that $m_0(T)$ might behave like $(T_c-T)^\beta$ for T near to but less than T_c and β some critical exponent with $0<\beta<1$. More precisely

$$\beta = \lim_{T\to T_c^-} \frac{\ln m_0(T)}{\ln(T_c-T)}$$

If the limit exists, we write $m_0(T)\sim(T_c-T)^\beta$. Empirically, β is often found to be approximately 0.33.

At $T=T_c$ one finds the dependence of m on h shown in Figure 4. This figure suggests that $m\sim h^{1/\delta}$ when h is very small. Here δ is some number greater than unity and often found to be approximately 4.2.

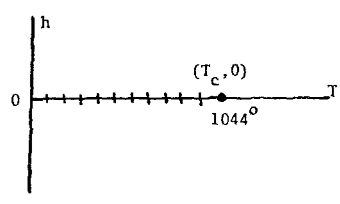

Fig. 2.

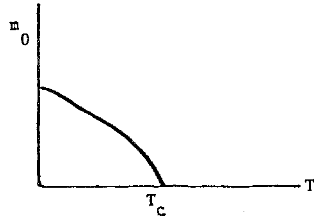

Fig. 3.

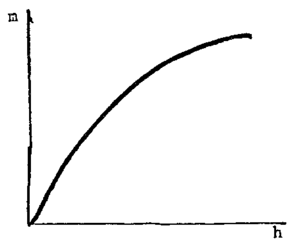

Fig. 4.

The magnetic susceptibility χ is defined by

$$\chi(T,h)\equiv\left(\frac{\partial m}{\partial h}\right)_T$$

For $h=0$ one finds the dependence of $\ln\chi$ on $\ln(T-T_c)$ shown in Figure 5. This figure suggests that $\chi\sim(t-t_c)^{-\gamma}$ for T near to but greater than T_c and $\gamma>0$. One similarly finds that the data suggest that $\chi\sim(T_c-T)^{-\gamma}$ when T is near to but less than T_c. Furthermore, the data also often suggest that $\gamma\sim\gamma'\sim1.3$.

The specific heat C is the rate of change of energy with respect to temperature. At $h=0$ one observes the dependence of C on T shown in Figure 6. This figure suggests that

$$C\sim\begin{cases}(T-T_c)^{-\alpha}, & T<T_c\\(T_c-T)^{-\alpha}, & T>T_c\end{cases}$$

The data also suggest that $\alpha\sim\alpha'\sim0.1$.

Similar critical exponents are defined for other systems such as fluids. Besides the relations $\alpha=\alpha'$ and $\gamma=\gamma'$, one also seems to have other relations occurring among the critical exponents, such as $\alpha'+2\beta+\gamma'=2$.

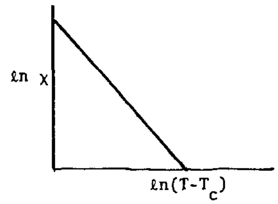

Fig. 5.

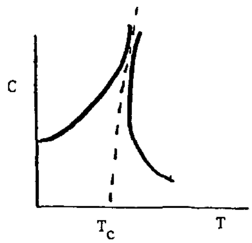

Fig. 6.

2.5. The Thermodynamic Limit. In the previous section we have described some typical critical phenomena. One goal of statistical mechanics is to derive these phenomena from an underlying microscopic model. If we have some microscopic model (e.g., the Ising model described in Section 2.1) with Gibbs measure $\mu = e^{-\beta H}/Z$, then one can associate to this model a specific heat $C \equiv \partial\langle H\rangle/\partial T$, where T is the temperature which is *defined* to be proportional to $1/\beta$. The free energy f is defined to be $-1/\beta \ln Z$. One thus has $H = f - T\,\partial f/\partial T$, and hence $C = -T\,\partial^2 f/\partial T^2$. Thus, in order that C have the singularity structure at the critical temperature T_c which was described in Section 2.4, it is necessary that f be a nonanalytic function of T at T_c. In the case of the Ising model for N spins, one has $Z = \Sigma_\sigma e^{-\beta H(\sigma)}$, where the sum is over the 2^N possible configurations of σ. Since the sum is finite, Z is a strictly positive analytic function of β. Hence, $f = -(1/\beta)\ln Z$ is an analytic function of β for $\beta > 0$. Thus, f has no singularities for $0 < T < \infty$, i.e., there is *no* critical temperature T_c. Thus, the critical phenomena of Section 2.4 *cannot* be derived from the Ising model of Section 2.1. If one feels that the model is a basically correct microscopic model, then one is led to reinterpret the experimental "facts" of Section 2.4. For example for large N the Ising model's specific heat might look as in Figure 7. Thus C_N would be an analytic function for all T having a maximum when N is large near T_c. Furthermore, the limit as $N \to \infty$ of C_N may very well be described

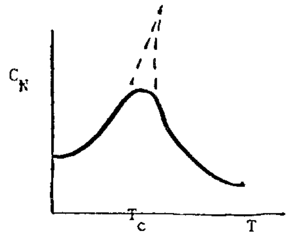

Fig. 7.

near T_c by a power law $C=(T-T_c)^{-\alpha}$. These heuristics suggest two programs of research. One, experimental, is aimed at exploring the critical region to see if one can find a maximum for C. If one meets with success here, one will have obtained a case study in the importance of theory for guiding experiments. But success in this project may be foredoomed if the size of the critical region is too small (e.g., 10^{-100}). For example, the specific heat of a superconductor looks experimentally as shown in Figure 8, i.e., it seems to have a discontinuity at T_c and not a power law behavior at all. But theory suggests that C really looks as shown in Figure 9 with the critical region simply too small for present techniques to probe. The second program of research is to rigorously derive the asymptotic properties of

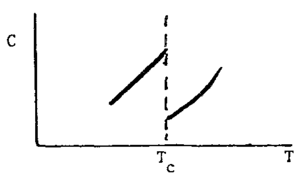

Fig. 8.

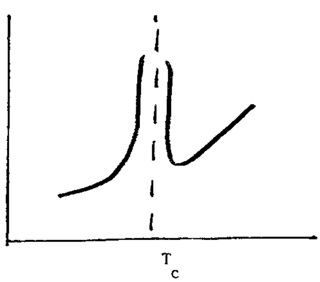

Fig. 9.

quantities such as C_N as $N \to \infty$. One might also try to formulate directly models for which $N = \infty$, and to derive their thermodynamic properties. For the Ising model and its relatives described in Section 2.1 much is already known concerning the models described in Sections 2.2 and 2.3 (see Kadanoff, 1977; and Ostwalder and Seiler, 1981; and the next section).

If the thermodynamic limit exists, then an important question which arises is whether or not the system has long-range order. This is a question concerning the decay rate of the Green's functions $G(fg^a)$, where g^a is g translated by the lattice vector a (and we are assuming $\langle f \rangle = \langle g^a \rangle = 0$). If $|G(fg^a)| \leq C^{|a|/\xi}$, then one says that the system has no long-range order. ξ is called a correlation length and $1/\xi = m$ is called a mass. Poles occurring in the Fourier transform $\tilde{G}(f, g^a)$ of $G(fg^a)$ are said to be related to the particle structure of the theory, the real part of the pole corresponding to the mass and the imaginary part to the (inverse) lifetime of the particle. This use of particle language is certainly a long way from our intuitive use of particle language for describing billiard balls! If one has a field taking values in $\mathbb{C}^N$, then, as the parameters in the model are varied so that a critical point is approached, long-range order sets in (i.e., one has spontaneous symmetry breakdown as in ferromagnetism), the mass m goes to zero, and $\tilde{G}$ acquires a pole at the origin. Physicists describe this pole by saying that the model has attained a massless Goldstone boson. If one now gauges this model, the long-range order disappears and m stays finite. This is because the local geometry itself is varying in a random fashion preventing the symmetry breakdown. In fact, if f and g are allowed to depend upon the $U(i, j)$ for finitely many pairs (i, j), then $\langle fg^a \rangle \leq Ce^{-m|a|}$ (i.e., the correlation in the geometry also exponentially decays). One says that the gauge field U has acquired a mass m. This process is known as the Higgs mechanism.

2.6. The Wilson Model of Quarks and Strings on a Lattice. Suppose one discovered a new liquid B. On the basis of past experience one would expect that by increasing its temperature one could make it undergo a phase transition into a gaseous phase Q. Suppose that such attempts to induce a phase transition fail. Then, either we do not have the technical ability to raise the temperature above the boiling point T_c of B or else $T_c = \infty$ and our failure to produce a gaseous state is not our fault. This is the present situation with regard to baryons. Baryons are believed to be composed of two or three quarks. Physicists expected that at very high energies the quarks would dissociate and one would see free quarks. This has not yet happened. The Wilson model of quarks and strings on a lattice is an attempt to understand these phenomena. (Of course, one is really interested in the relativistic model obtained by analytically continuing the continuum limit of the Wilson model.) In this model one associates the Fermionic measure space (pt., $\wedge^*(W_i \otimes \overline{W}_i), \mathcal{F}\delta$) to the lattice point i, and in forming the Green's functions also averages over the space of connections using the Yang–Mills action (see the end of Section 2.3). This model has *no* first-order observables. Wilson *conjectures* that in four dimensions this model has no phase transition (i.e., $T_c = \infty$), and hence that quarks are simultaneously asymptotically free and also confined.

REFERENCES

Domb, C., and Green, M. (1972). *Phase Transitions and Critical Phenomena*, 6 Vols. Academic Press, New York.

Edwards, D. (1979). "The Mathematical Foundations of Quantum Mechanics," *Synthese* **2**, 1–70.

Edwards, D. (1981). "Phase Transitions and Critical Phenomena," (Submitted).

Kadanoff, L. (1977). "The Application of Renormalization Group Techniques to Quarks and Strings," *Reviews of Modern Physics*, **49**, No. 2.

Kostant, B. (1977). "Graded Manifolds, Graded Lie Theory, and Prequantization," in *Differential Geometric Methods in Mathematical Physics*, Vol. 570. Springer-Verlag, Heidelberg.

Ma, S. (1976). *Modern Theory of Critical Phenomena*. W. A. Benjamin, Inc., New York.

Osterwalder, K., and Seiler, E. (1981). "Gauge Field Theories on the Lattice," (preprint).