REVIEW ARTICLE

Quantum spin chains and the Haldane gap

J. Phys.: Condens. Matter. **1** (1989) 3047–3072. Printed in the UK

REVIEW ARTICLE

Quantum spin chains and the Haldane gap

Ian Affleck
Canadian Institute for Advanced Research and Physics Department,
University of British Columbia, Vancouver, BC, V6T 2A6, Canada

Received 24 October 1988

Abstract. One-dimensional antiferromagnets have exotic disordered ground states. As was first argued by Haldane, there is an excitation gap for integer, but not half-integer, spin. We review the arguments for this behaviour based on field-theory mappings, the Lieb–Shultz–Mattis theorem, exactly solvable models, finite-chain diagonalisation and real experiments.

1. Introduction

The study of quantum spin chains has a long history, going back at least to the remarkable exact solution found by Bethe [1] in 1931 for the spin-$\frac{1}{2}$ case. Spin-wave theory, developed by Anderson [2], Kubo [3] and others in the 1950s, gave a very clear physical picture of the behaviour of higher-dimensional antiferromagnets, predicting long-range order and gapless Goldstone bosons. However, the one-dimensional case remained rather mysterious. It was known that long-range order could not occur in $d = 1$, and yet the Bethe *ansatz* predicts massless excitations. A vague notion was common that in $d = 1$ 'quasi-long-range order' occurs, corresponding to power-law correlations and gapless excitations which are not true Goldstone bosons. Such a Kosterlitz–Thouless phase [4] occurs in a variety of other one-dimensional quantum and two-dimensional finite-T classical systems, with SO(2) symmetry. The bosonisation approach to the $s = \frac{1}{2}$ problem, by Luther and Peschel [5] in 1975, when combined with Bethe *ansatz* results, makes this picture concrete. At this point any special features associated with the larger, SO(3) symmetry of the Heisenberg model were not very clear.

Thus, it came as quite a shock to many practitioners of this field when Haldane first suggested [6] in 1983 that integer-spin antiferromagnetic Heisenberg spin chains have a finite gap, and only the half-odd-integer chains are gapless. Early attempts at confirming this numerically for $s = 1$ were somewhat confused by the poor finite size convergence of the gapless $s = \frac{1}{2}$ case although eventually the numerical evidence for a gap became quite good [7]. Bethe *ansatz* integrable chains [8] with integer spin and no gap (albeit with special generalised interactions) seemed to contradict Haldane's claim. The rather esoteric and non-rigorous nature of the original arguments for the difference between integer and half-odd-integer spin, involving topological terms in quantum-field theories, made this issue rather impenetrable.

Several more recent developments have helped to clarify the situation in several respects. A rigorous proof of zero gap, by Lieb, Shultz and Mattis (LSM) [9] for spin $s = \frac{1}{2}$, was extended to arbitrary half-integer s, but shown to fail for integer s [10]. A different

0953-8984/89/193047 + 26 $02.50

class of solvable (valence-bond) models with generalised interactions was discovered and a non-zero gap was proven rigorously in the $s = 1$ case [11]. The exact ground state could be written down in a simple way (unlike the Bethe ground state) and provided a good variational ground state for the Heisenberg Hamiltonian. Furthermore these models distinguished integer and half-integer s in a simple way. Perhaps most importantly, experimental evidence for Haldane gaps has recently been observed in $CsNiCl_3$ [12–14] and $Ni(C_2H_8N_2)_2NO_2ClO_4$ [15].

Richard Feynman once said [16] that if a physicist gives several arguments to make his point then he must not have one *really good* argument. In this review, I will give several more or less independent arguments for the difference between integer and half-integer s. I will begin, in § 2, by briefly reviewing the failure of spin-wave theory in $d = 1$. This discussion makes natural the existence of a finite correlation length, and hence gap. I will then, in § 3, discuss the large-s mapping onto the O(3) non-linear σ-model, emphasising the connections with standard spin-wave theory. Without any reference to topology, the gap for integer s can now be understood by appealing to the analogous problem of the classical, finite-temperature, $d = 2$ ferromagnet, and to the renormalisation group. From this perspective what is more difficult to understand is what was previously taken for granted: the *non-existence* of a gap for half-integer s. In § 4, I will discuss the LSM theorem, which proves this rigorously, using the basic peculiar property of half-integer s: wavefunctions change sign under 2π rotation. In § 5 I will discuss the solvable valence-bond models. In § 6, I will use topological arguments, involving vortex condensation, to explain the difference. Section 7 briefly reviews the numerical, finite-chain work. Section 8 reviews the experimental tests of the Haldane gap. Sections 4, 5, 7 and 8 do not depend in an essential way on § 3 and do not use any field-theory techniques. While § 3 *does* involve some field theory, I have endeavoured to make it as elementary and self-contained as possible. Section 6, on the other hand, requires the most baggage and contains the most hand-waving.

2. Spin-wave theory for the Heisenberg model [17]

The Heisenberg Hamiltonian describes the nearest-neighbour interactions of localised quantum spins:

$$H = J \sum_{\langle i,j \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j$$

where the $\boldsymbol{S}_i$ are quantum spins obeying

$$[S^a, S^b] = i\varepsilon^{abc} S^c \qquad S^2 = s(s+1).$$

(I will let $\langle i, j \rangle$ denote a pair of nearest-neighbour sites on some regular lattice. The sum is over each pair of nearest-neighbour sites once. Repeated spin indices are summed.)

If the $\boldsymbol{S}_i$ were *classical* spins, i.e. fixed-length classical vectors, the ground state of H would be the state with all spins parallel for the ferromagnetic case or neighbouring spins antiparallel for the antiferromagnetic case. This classical antiferromagnetic ground state is known as the Néel state. It is illuminating to rewrite H in terms of raising and lowering operators:

$$S^{\pm} = S^x \pm iS^y \qquad H = J \sum_{\langle i,j \rangle} [\tfrac{1}{2}(S_i^+ S_j^- + S_i^- S_j^+) + S_i^z S_j^z].$$

The classical ground states obey

$$S_i^z |\psi\rangle = \pm s |\psi\rangle.$$

Note that the ferromagnetic state is indeed the quantum ground state, but the Néel state is not an eigenstate since acting with H we may raise and lower a pair of neighbouring spins.

Since the Néel state is not the exact ground state for the antiferromagnetic Heisenberg model, we would like to know whether or not, and under what circumstances, it is a qualitatively correct approximation to the ground state. In particular, we would like to know whether or not long-range antiferromagnetic order

$$\langle S^z(\boldsymbol{x})S^z(\boldsymbol{0})\rangle = \pm m$$

occurs. (The + or − sign occurs depending on whether the points $\boldsymbol{0}$ and $\boldsymbol{x}$ are on the same or opposite sublattice.) To study this question it is convenient to perturb away from a limit in which the Néel state becomes exact. This occurs for a spin-s Heisenberg antiferromagnet in the limit $s \to \infty$. In this limit, quantum spins become classical because the commutator is much smaller than the square of the spin variables:

$$[S^a, S^b] = \mathrm{i}\varepsilon^{abc}S^c = \mathrm{O}(s) \ll \mathrm{O}(s^2).$$

For large s we expect S^z to make only relatively small deviations away from $\pm s$ (the + or − sign occurs for one or the other sublattice). A nice way of formalising this and studying the corrections to the Néel state is to use the Holstein–Primakov transformation:

$$S^z = s - a^+a \qquad S^- = (2s)^{1/2}a^+(1 - a^+a/2s)^{1/2}.$$

This produces the exact commutation relations and the constraint $S^2 = s(s+1)$. The state with maximal S^z, $S^z = s$, corresponds to the state with no bosons present. Acting on this state with S^- creates a boson and hence lowers S^z by one. The state of minimal S^z, $S^z = -s$, is the state with $2s$ bosons present. This state is annihilated by S^-, as it should be since $1 - a^+a/2s = 0$ acting on this state. For antiferromagnets we make the above substitution on one sublattice, the A sublattice, but on the other B lattice it is convenient to represent the spins in terms of a boson, b:

$$S^z = -s + b^+b \qquad S^- = (2s)^{1/2}(1 - b^+b/2s)^{1/2}b.$$

Again the state with no bosons corresponds to the Néel state. In this case acting with S^+ creates bosons. Using this representation for the Heisenberg model we will find that a^+a is O(1) so that S^z deviates from s by an amount of $\mathrm{O}(1) \ll s$.

We may develop a systematic expansion in $1/s$ by expanding the square root in the definition of S^-:

$$S^- \simeq (2s)^{1/2}a^+.$$

To quadratic order the Hamiltonian is

$$H = J\sum_{\langle x,y\rangle}[-s^2 + s(a_x^+a_x + b_y^+b_y + a_xb_y + a_x^+b_y^+) + \ldots].$$

Here we sum over nearest neighbours with x, y in the a and b sublattice respectively. The . . . represent higher-order terms in the boson operators. These terms are down by powers of s relative to the terms we kept, and hence can be treated in perturbation theory in $1/s$. To leading order we must diagonalise the quadratic Hamiltonian. We do this by

going to momentum space and then diagonalising by a Bogoliubov transformation. In momentum space

$$H = Jsz \sum_k [\gamma_k (a_k b_{-k} + a_k^+ b_{-k}^+) + (a_k^+ a_k + b_k^+ b_k)]$$

where

$$\gamma_k \equiv \sum_\delta e^{ik\cdot\delta}/z = \gamma_{-k}$$

and z is the number of nearest neighbours. Here $\boldsymbol{\delta}$ represents the translation vectors to the neighbouring sites. We see that we must make a Bogoliubov transformation mixing a_k with b_k. We define new boson operators by the canonical transformation

$$c_k = u_k a_k - v_k b_{-k}^+ \qquad d_k = u_k b_k - v_k a_{-k}^+.$$

The correct normalisation of the commutators of c with c^+ and d with d^+ imply

$$|u|^2 - |v|^2 = 1.$$

Inverting the transformation and plugging into H, we find the condition that makes the off-diagonal terms vanish:

$$\gamma(u^2 + v^2) + 2uv = 0.$$

Solving for u and v we find the diagonal Hamiltonian:

$$H = Jsz \sum_k (1 - \gamma_k^2)^{1/2} (c_k^+ c_k + d_k^+ d_k).$$

The excitations created by c and d are known as spin waves. They correspond to infinitesimal deviations of the spins away from the Néel state. The spin-wave dispersion relation is

$$E(\boldsymbol{k}) = Jsz(1 - \gamma_k^2)^{1/2}.$$

Note that $E(\boldsymbol{k})$ vanishes linearly at $\boldsymbol{k} = 0$:

$$E(\boldsymbol{k}) \rightarrow v|\boldsymbol{k}|.$$

For a d-dimensional hypercubic lattice, we have

$$v = 2d^{1/2} Js.$$

This is a relativistic dispersion relation, with v the effective 'velocity of light'. We will shortly discover a relativistic quantum-field theory description of the low-energy theory. The vanishing of the spin-wave gap is a consequence of Goldstone's theorem. There are two Goldstone modes, c and d, corresponding to the breaking of SO(3) down to SO(2), rotations about the z axis. Long-wavelength spin waves correspond to states where all regions are locally in some Néel ground state but the direction of the sublattice magnetisation vector makes long-wavelength rotations. Note that the two different Goldstone modes, c and d, have spin $S^z = \pm 1$, respectively. They correspond to raising S^z one sublattice or lowering it on the other.

We may use the spin-wave representation to calculate the reduction in the sublattice magnetisation due to quantum fluctuations:

$$\langle S_z(x)\rangle = s - \int [\mathrm{d}^d k/(2\pi)^d]\tfrac{1}{2}[(1-\gamma_k^2)^{-1/2} - 1].$$

Essentially the quantum spins have zero-point motion, just like a quantum harmonic oscillator. The zero-point reduction is independent of s, to leading order, whereas the classical value is s. Thus the quantum correction is relatively small for large s. However, for $d = 1$ we find that the 'correction' is divergent at small wavenumbers:

$$\Delta\langle S_z(x)\rangle \simeq -\int [\mathrm{d}k/(2\pi)](1/2k) = -\infty.$$

This indicates that the Néel state is destabilised by quantum fluctuations in $d = 1$, no matter how large s is.

It is even more illuminating to calculate the long-wavelength behaviour of the correlation function to next order in $1/s$:

$$\langle \boldsymbol{S}_0 \cdot \boldsymbol{S}_x\rangle \simeq \pm s^2[1 - (1/\pi s)\ln x/a + \mathrm{O}(1/s^2)]$$

where a is a distance of order the lattice spacing. This formula indicates that, for large s, the system is ordered at short wavelengths, the disorder occurring at exponentially large scales

$$\xi \simeq a\mathrm{e}^{\pi s}.$$

3. Large-s mapping onto the O(3) non-linear σ-model [6, 18]

We now will try to derive the low-energy, continuum limit of antiferromagnetic quantum chains. The basic idea is to keep only the long-wavelength, low-energy modes. Formally, we may 'integrate out' the high-energy modes to obtain an effective Hamiltonian. This treatment can be justified in the large-s limit, although many of the conclusions should remain true for arbitrary s.

We saw in § 2 that there are low-energy modes at wavevectors near 0 and also near the ordering wavevector, π (we doubled the unit cell). We wish to keep the Fourier modes of $\boldsymbol{S}_i$ near these two values of k. Thus we write

$$\boldsymbol{S}_i \simeq \pm s\boldsymbol{\varphi}_i + \boldsymbol{l}_i$$

where $\boldsymbol{\varphi}$ and $\boldsymbol{l}$ are slowly varying on the scale of the lattice. We expect the $k \simeq 0$ component to be O(1) and the $k \simeq \pi$ component to be O(s), based on the classical or Holstein–Primakoff analysis. To set this up in a careful, quantum-mechanical way, it is convenient to define the $\boldsymbol{\varphi}$ and $\boldsymbol{l}$ variables as linear combinations of a few nearby sites.

We make the arbitrary choice of combining each spin on an even site, $2i$, with the spin to its right, at $(2i+1)$ as shown in figure 1:

$$\boldsymbol{\varphi}(2i+\tfrac{1}{2}) = (\boldsymbol{S}_{2i+1} - \boldsymbol{S}_{2i})/2s \qquad \boldsymbol{l}(2i+\tfrac{1}{2}) = (\boldsymbol{S}_{2i} + \boldsymbol{S}_{2i+1})/2. \tag{1}$$

The advantage of this approach is that the commutators are simple:

$$[l^a(x), l^b(y)] = \mathrm{i}\varepsilon^{abc} l^c \delta(x-y)$$

$$[l^a(x), \varphi^b(y)] = \mathrm{i}\varepsilon^{abc}\varphi^c \delta(x-y)$$

$$[\varphi^a(x), \varphi^b(y)] = \mathrm{i}\varepsilon^{abc} l^c \delta(x-y)/s^2 \to 0.$$

Here the Dirac δ-function is defined from the continuum limit of the Kronecker δ-function as

$$\delta(x-y) = \lim \delta_{x,y}/a$$

where, in this case, the lattice spacing between the points where the fields $\boldsymbol{\varphi}$ and $\boldsymbol{l}$ are defined is $a = 2$. These commutation relations at $s \to \infty$ are of the type that occur in quantum-field theory. If we have a triplet of free fields $\boldsymbol{\varphi}$ and conjugate momenta Π, obeying

$$[\varphi_i(x), \Pi_j(y)] = \mathrm{i}\delta_{ij}\delta(x-y)$$

then the above commutation relations occur if we define

$$\boldsymbol{l} \equiv \boldsymbol{\varphi} \times \boldsymbol{\Pi}.$$

$\int \mathrm{d}x\, \boldsymbol{l}$ is the generator of rotations in the field theory. It is conserved if the Hamiltonian is rotationally invariant. As defined by (1), $\boldsymbol{\varphi}$ and $\boldsymbol{l}$ obey the constraints $\boldsymbol{\varphi} \cdot \boldsymbol{l} = 0$, and

$$\boldsymbol{\varphi}^2 = 1 + 1/s - \boldsymbol{l}^2/s^2 \to 1.$$

We now assume that $\boldsymbol{\varphi}$ and $\boldsymbol{l}$ are slowly varying on the lattice scale and calculate the Hamiltonian in a gradient expansion. We keep terms up to $O(\boldsymbol{\varphi}'^2)$ and $O(\boldsymbol{l}^2)$ only since $\boldsymbol{l}$ effectively contains a time derivative. Somewhat surprisingly, we will pick up a term *linear* in $(\mathrm{d}/\mathrm{d}x)$ which appears to break parity. This is a consequence of the parity-breaking method of passing to the continuum limit introduced above. Using

$$\boldsymbol{S}_{2i} \cdot \boldsymbol{S}_{2i+1} = 2\boldsymbol{l}(2i + \tfrac{1}{2})^2 + \text{constant}$$

$$\begin{aligned}\boldsymbol{S}_{2i} \cdot \boldsymbol{S}_{2i-1} &= -s^2 \boldsymbol{\varphi}(2i+\tfrac{1}{2}) \cdot \boldsymbol{\varphi}(2i-\tfrac{3}{2}) + s[\boldsymbol{l}(2i-\tfrac{3}{2}) \cdot \boldsymbol{\varphi}(2i+\tfrac{1}{2}) - \boldsymbol{\varphi}(2i-\tfrac{3}{2}) \cdot \boldsymbol{l}(2i+\tfrac{1}{2})] \\ &\quad + \boldsymbol{l}(2i+\tfrac{1}{2}) \cdot \boldsymbol{l}(2i-\tfrac{3}{2}) \\ &\simeq 2s^2 \boldsymbol{\varphi}'^2 - s(\boldsymbol{l} \cdot \boldsymbol{\varphi}' + \boldsymbol{\varphi}' \cdot \boldsymbol{l}) + 2\boldsymbol{l}^2\end{aligned}$$

the Hamiltonian then becomes

$$H = J \int (\mathrm{d}x/2)[4\boldsymbol{l}^2 + 2s^2 \boldsymbol{\varphi}'^2 - s(\boldsymbol{l} \cdot \boldsymbol{\varphi}' + \boldsymbol{\varphi}' \cdot \boldsymbol{l})].$$

We rewrite this as

$$\mathcal{H} = (v/2)\{g^2[\boldsymbol{l} - (\theta/4\pi)\boldsymbol{\varphi}']^2 + \boldsymbol{\varphi}'^2/g^2\} \tag{2}$$

with velocity, coupling constant and topological angle

$$v = 2Js \qquad g = 2/s \qquad \theta = 2\pi s.$$

To understand better what this Hamiltonian represents, we observe that it follows from the Lagrangian

$$\mathcal{L} = (1/2g)\partial_\mu \boldsymbol{\varphi} \cdot \partial^\mu \boldsymbol{\varphi} + (\theta/8\pi)\varepsilon^{\mu\nu} \boldsymbol{\varphi} \cdot (\partial_\mu \boldsymbol{\varphi} \times \partial_\nu \boldsymbol{\varphi})$$

with the constraint $\boldsymbol{\varphi}^2 = 1$. (We set $v = 1$ for this discussion.) The simplest way of handling the constraint is to introduce coordinates for the sphere:

$$\boldsymbol{\varphi} = (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)$$

$$\mathcal{L} = (1/2g)[(\partial_\mu \alpha)^2 + \sin^2\alpha\,(\partial_\mu \beta)^2] + (\theta/8\pi)\sin\alpha\,\varepsilon^{\mu\nu}\partial_\mu \alpha \partial_\nu \beta.$$

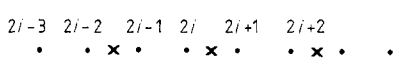

Figure 1. The pairing of sites to define the continuum variables, $\boldsymbol{\varphi}$ and $\boldsymbol{l}$.

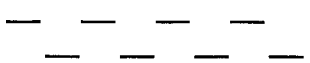

Figure 2. The dimer ground states for an $s = \frac{1}{2}$ chain.

We now see that the θ-term is a total derivative. On the compactified Euclidean space (i.e. with $\boldsymbol{\varphi}(\boldsymbol{x}) \to$ constant at $|\boldsymbol{x}| \to \infty$) both $\boldsymbol{\varphi}$ and $\boldsymbol{x}$ can be represented by points on a sphere, S^2. The integral

$$(1/8\pi) \int \mathrm{d}^2 \boldsymbol{x}\, \varepsilon^{\mu\nu} \boldsymbol{\varphi} \cdot (\partial_\mu \boldsymbol{\varphi} \times \partial_\nu \boldsymbol{\varphi})$$

measures the winding number of the sphere onto the sphere, i.e. $\Pi^2(S^2)$. It is simply the Jacobian for the change of variables from polar coordinates in $\boldsymbol{x}$-space (with $\boldsymbol{x}$ the spatial coordinate on a sphere) to polar coordinates in $\boldsymbol{\varphi}$-space. This is an integer for any smooth (finite-action) field configuration. Since the θ-term is a total derivative it has no effect on the classical equations of motion, and no effect in perturbation theory. It does, however, lead to a change in the Hamiltonian. The conjugate momenta are

$$\Pi_\alpha = \dot{\alpha}/g + (\theta/4\pi)\beta' \sin\alpha$$

$$\Pi_\beta = (\dot{\beta} \sin^2\alpha)/g - (\theta/4\pi)\alpha' \sin\alpha$$

and the Hamiltonian density is

$$\mathcal{H} = (g/2)\{[\Pi_\alpha - (\theta/4\pi)\beta' \sin\alpha]^2 + [\Pi_\beta + (\theta/4\pi)\alpha' \sin\alpha]^2/\sin^2\alpha\}$$
$$+ (1/2g)[\alpha'^2 + \beta'^2 \sin^2\alpha].$$

Note that the effect of the topological term is to redefine the conjugate momenta by a canonical transformation

$$\Pi_a \to \exp\left(\mathrm{i}(\theta/4\pi) \int \mathrm{d}x\, \beta' \cos\alpha\right) \Pi_a \exp\left(-\mathrm{i}(\theta/4\pi) \int \mathrm{d}x\, \beta' \cos\alpha\right) \qquad (a = \alpha \text{ or } \beta).$$

Thus the θ-dependence of the Hamiltonian may be removed by making this transformation. However, it is expected that the Hilbert space breaks up into decoupled sectors labelled by θ. Making the canonical transformation on the *states* maps between the different sectors, or θ-vacua.

Finally we may rewrite this Hamiltonian in a coordinate-independent way by introducing the generator of rotations:

$$\boldsymbol{l} \equiv \boldsymbol{\varphi} \times [(\partial\boldsymbol{\varphi}/\partial\alpha)\Pi_\alpha + (1/\sin^2\alpha)(\partial\boldsymbol{\varphi}/\partial\beta)\Pi_\beta].$$

The Hamiltonian then takes the form of equation (2).

We now comment on the implications of Hamiltonian (2) for the quantum spin chain. Since the action has the form

$$S = S_0 + \mathrm{i}\theta Q$$

where Q is the integer-valued topological charge, e^{-S} is a periodic function of θ for any smooth field configuration. Therefore we expect all physical properties of the model (which can be derived from the path integral $\int [\mathrm{d}\boldsymbol{\varphi}]\mathrm{e}^{-S}$) to be periodic in θ. Thus we see that for integer-s chains we effectively have $\theta = 0$. This case is by far the easiest to understand. The action is real and corresponds to a classical two-dimensional ferromagnet at finite temperature, $T = g$. On the other hand, for half-integer s we have $\theta = \pi$. The action is now complex so there is no correspondence to a standard classical two-dimensional problem. We will give a hand-waving explanation of the behaviour of the $\theta = \pi$ model in § 6. Here we would like to concentrate on the simpler $\theta = 0$ case, corresponding to integer s.

First note that a semi-classical analysis of (2) reproduces the results of lowest order (in $1/s$) spin-wave theory. Semi-classically we assume spontaneous symmetry breaking and write

$$\boldsymbol{\varphi} = (\varphi_1, \varphi_2, (1 - \varphi_1\varphi_1 - \varphi_2\varphi_2)^{1/2}) \simeq (\varphi_1, \varphi_2, 1).$$

We then obtain the approximate free Lagrangian density

$$\mathcal{L} \simeq (1/2g) \sum_{i=1}^{2} (\partial_\mu \varphi_i)^2.$$

Thus we obtain two free massless Goldstone bosons. This is the same result obtained using spin-wave theory. We also obtain the same velocity parameter, $v = 2Js$. The sublattice magnetisation is given by

$$s\langle\varphi_3\rangle \simeq s\left(1 - \left\langle \sum_{i=1}^{2} \varphi_i^2 \right\rangle\right) = s\left(1 - 2g \int [\mathrm{d}^2\boldsymbol{k}/(2\pi)^2]\boldsymbol{k}^2\right).$$

We obtain the same infrared divergence as before, telling us that the symmetry is in fact *not* spontaneously broken, since we are in (1 + 1) dimensions.

We apply the renormalisation group to obtain a deeper understanding of this infrared divergence. The effective coupling at length scale L behaves as

$$\mathrm{d}g/\mathrm{d} \ln L = g^2/2\pi.$$

The effective coupling grows large at large length scales

$$g(L) \simeq g_0/[1 - (g_0/2\pi) \ln L].$$

Initially the bare coupling at the lattice scale is $g_0 = 2/s$. $g(L)$ becomes O(1) at length scales

$$\xi \simeq \mathrm{e}^{2\pi/g_0} = \mathrm{e}^{\pi s}.$$

This implies that for large s the spin chain has short-range order, but the spins make very long-wavelength, $O(\xi)$, rotations which destroy the order. We normally expect that a field theory with this type of renormalisation group behaviour develops a mass gap due to the infrared fluctuations, with a mass $m \simeq \xi^{-1}$. This discussion of perturbation theory applies equally well for $\theta = 0$ or π. It is only the non-perturbative effects, discussed in § 6, which distinguish the two cases. The $\theta = \pi$ model is rather abnormal in that it does not develop a mass gap, despite having a characteristic energy scale ξ^{-1}.

The generation of a mass, for $\theta = 0$, can be seen, for example, in the large-n limit of the O(n) σ-mode [19]. This is defined by the Lagrangian

$$\mathcal{L} = (n/2g)(\partial_\mu \boldsymbol{\varphi})^2 \qquad \boldsymbol{\varphi}^2 = 1$$

with $\boldsymbol{\varphi}$ an n-component vector. It is convenient to enforce the constraint using the Fourier representation of the δ-function:

$$\mathcal{L} = (n/2g)[(\partial_\mu \boldsymbol{\varphi})^2 + \mathrm{i}\lambda(\boldsymbol{\varphi}^2 - 1)].$$

We now integrate out $\boldsymbol{\varphi}$, which is no longer constrained, obtaining an effective action for λ:

$$S^{\mathrm{eff}}(\lambda) = (n/2)\left(-\int \mathrm{d}^2\boldsymbol{x}\,(\mathrm{i}\lambda/g) + \mathrm{tr}\ln(-\partial^2 + \mathrm{i}\lambda)\right).$$

Because there is a factor of n in S^{eff} we may ignore the fluctuations of λ and evaluate it

at the lowest action saddle point. This is obtained by deforming the functional integration contour of λ into the complex plane. A saddle point is found at constant, imaginary λ. Defining $i\lambda_0 = m^2$ (the saddle-point value), we have the saddle-point equation

$$(1/g) = \int [d^2\boldsymbol{k}/(2\pi)^2](\boldsymbol{k}^2 + m^2) = (1/2\pi) \ln \Lambda/m$$

where we have introduced an ultraviolet cut-off Λ (essentially the lattice spacing in the quantum spin chain). This equation determines the mass parameter m in terms of the cut-off and bare coupling g:

$$m = \Lambda e^{-2\pi/g}.$$

To lowest order in $1/n$, the $\boldsymbol{\varphi}$ Green functions are just those of n free bosons of mass m.

This picture of the O(3) σ-model is reinforced by the exact S-matrix result [20]. An exact S-matrix has been found obeying all the requirements of a quantum-field theory S-matrix (Lorentz invariance, unitarity, etc.) in which the only single-particle states are a triplet of massive bosons. This is believed to be the S-matrix of the O(3) σ-model, a conjecture which can be checked in perturbation theory.

Another type of evidence about the behaviour of the O(3) σ-model is based on the Euclidean lattice version of the theory:

$$S = -(1/T) \sum_{\langle i,j \rangle} \boldsymbol{\varphi}_i \cdot \boldsymbol{\varphi}_j.$$

This is simply the two-dimensional *classical* Heisenberg ferromagnet (or antiferromagnet; it makes no difference at the classical level). Monte Carlo simulations and experiments on magnetic systems indicate that the system is in a disordered phase with a finite correlation length for all T. The $T = 0, (1 + 1)$-dimensional quantum system has an effective temperature $T_{\text{eff}} = 2/s$. The disorder due to thermal fluctuations in the classical model corresponds to the quantum disorder in the quantum model.

4. The Lieb–Shultz–Mattis theorem

We saw above that there is a difference between integer and half-integer spin; in the large-s limit they map onto different field theories—with $\theta = 0$ or π respectively. While the behaviour of the $\theta = 0$ theory seems to be very well understood, the $\theta = \pi$ case is much less so. An important clue about the behaviour of half-integer-s isotropic spin chains comes from a rigorous theorem, first proved for $s = \frac{1}{2}$ by Lieb *et al* [9] and then extended to arbitrary half-integer s in [10]. It is of special significance that the theorem only holds for half-integer s.

The theorem proves that a half-integer-s spin chain with essentially *any* reasonably local Hamiltonian respecting translational symmetry and rotational symmetry either has zero gap (i.e. 'mass') or else has degenerate ground states, corresponding to spontaneously broken parity (i.e. reflection about a site). We will see later that either possibility can be realised for various spin-chain Hamiltonians.

The theorem, though rigorous, is elementary, We begin with a finite chain of even length L and periodic boundary conditions, and a ground state, $|\psi_0\rangle$, respecting parity

and rotational invariance. We now construct a state $|\psi_1\rangle$ which is orthogonal to $|\psi_0\rangle$ and has low energy. Explicitly

$$\langle\psi_1|(H - E_0)|\psi_1\rangle = O(1/L)$$

where E_0 is the ground-state energy. While $|\psi_1\rangle$ is not itself an energy eigenstate, this construction proves that a low-energy eigenstate exists. The state $|\psi_1\rangle$ is constructed by acting with a unitary operator U: $|\psi_1\rangle = U|\psi_0\rangle$. Numbering the sites from $-L/2$ to $L/2 - 1$, U is defined by

$$U = \exp\left((i\pi/l)\sum_{j=-l}^{l}(j + l)S_j^z\right).$$

Thus we are twisting the jth spin by $\pi(j + l)/l$ about the z axis. This twist varies from 0 to 2π as we travel along the chain, from $-l$ to l. The sites with $|i| > l$ are untwisted. l is chosen to be some number of $O(L)$; for instance, we may choose, $2l = L/2 - 1$ or $L/2 - 2$. We must prove two things: first that $|\psi_1\rangle$ has low energy, and secondly that it is orthogonal to the ground state. The first fact follows from the fact that the relative twist of two neighbouring sites is $\pi/l \ll 1$. Since a uniform twist is a symmetry operation and costs no energy at all, a slow twist costs only a small energy. Explicitly,

$$U^+ S_i^+ S_{i+1}^- U = e^{-i(\pi/l)} S_i^+ S_{i+1}^-.$$

Thus, for the Heisenberg Hamiltonian

$$\delta E \equiv \langle\psi_1|(H - E_0)|\psi_1\rangle = (J/2)\sum_i \langle\psi_0|(e^{-i\pi/l} - 1)S_i^+ S_{i+1}^- + (e^{i\pi/l} - 1)S_i^- S_{i+1}^+|\psi_0\rangle.$$

The assumed symmetry of the ground state allows us to rewrite this as

$$\delta E = J[\cos(\pi/l) - 1]\sum_i \langle\psi_0|S_i^+ S_{i+1}^-|\psi_0\rangle = \tfrac{2}{3}e_0[\cos(\pi/l) - 1](2l + 2) = O(1/l)$$

(e_0 is the ground-state energy per link). We have only given the proof for the Heinsenberg Hamiltonian but it generalises easily to essentially any symmetric local Hamiltonian. δE can always be expressed as the expectation value of some local operator times a quantity of $O(1/l)$.

Nothing in this part of the proof distinguishes integer and half-integer s. However, we must still prove that $\langle\psi_1|\psi_0\rangle = 0$. Otherwise there is danger that $|\psi_1\rangle$ will approach $|\psi_0\rangle$ as $l \to \infty$, and this may not imply anything about the spectrum in the infinite l limit. To prove orthogonality we simply note that $|\psi_1\rangle$ has the opposite parity to $|\psi_0\rangle$. Actually, we use a product of parity and a rotation about the y axis by π. The combined operation takes $S_i^z \to -S_{-i}^z$ and hence $U \to U\exp(-2i\pi\Sigma_{j=-l}^{l} S_j^z)$. But since the sum contains an odd number of spins

$$\exp\left(-2i\pi\sum_{j=-l}^{l} S_j^z\right) = \begin{cases} -1 & \text{(half-integer } s\text{)} \\ +1 & \text{(integer } s\text{).} \end{cases}$$

Thus $|\psi_1\rangle$ is parity even or odd for s integer or half-integer, respectively. (Here we are measuring the parity relative to that of the ground state.) In the latter case $|\psi_1\rangle$ is necessarily orthogonal to the ground state.

There are certainly two possibilities left open by this result. At $L \to \infty$, there may be a gapless, particle-like excitation with odd parity. Alternatively, parity may be spontaneously broken with two ground states and a gap above each. An example of

simple candidate ground states with broken parity are the simple dimer configurations shown in figure 2. Neighbouring pairs of spins are contracted to form singlets:

$$|\psi\rangle = |\alpha_1\alpha_2\alpha_3\alpha_4 \ldots\rangle\varepsilon^{\alpha_1\alpha_2}\varepsilon^{\alpha_3\alpha_3}\ldots.$$

An exactly solvable Hamiltonian with this ground state is discussed in § 5. Normally the ground state is unique for a finite-length chain, in this situation. The two approximate ground states have some overlap, which goes to zero like $e^{-\text{const}\cdot L}$. This mixing produces a symmetric ground state (essentially the symmetric linear combination of the two) and a low-lying ($\delta E = O(e^{-\text{const}\cdot L})$) antisymmetric excited state. This is completely analogous to the mixing of the two approximate ground states in a double-well potential in one-dimensional quantum mechanics. The mixing is proportional to $e^{-S/\hbar}$, where S is the action of the tunnelling configuration.

Although the above arguments constitute a 'physicist's proof' that either parity is spontaneously broken or else there are gapless excitations of odd parity, it is non-trivial to make this argument completely rigorous. The needed argument [10] starts with a fixed value of l and lets $L \to \infty$ first. The fact that U is a local operator, although acting on an infinite vector space, is essential. If it were only possible to create a low-energy state by spreading it over the whole chain then there is a possible danger that it might become unobservable, and yet not connected with ground-state degeneracy at $L \to \infty$.

We argued in § 3 that integer-s Heisenberg chains have a unique ground state with a finite gap, at least for large s. The above rigorous proof shows that this is not possible for half-integer s. Either the ground state is degenerate or the gap vanishes. In the next section we will see examples of Hamiltonians where each of these possibilities is realised. For the simplest Heisenberg interaction we expect that the gapless option occurs, for all half-integer s.

5. Solvable models

Spin-wave theory and the mapping onto the non-linear σ-model are not rigorous. The implications of the LSM theorem are ambiguous. Thus it is very useful to have some solvable models of quantum spin chains, even though, in general, these Hamiltonians are not realistic.

There are two known classes of solvable models. The first and by far the oldest are the models solvable by the Bethe *ansatz*. These include the nearest-neighbour $s = \frac{1}{2}$ Heisenberg chain. In general there is a solvable model of this type for all s, with polynomial nearest-neighbour interactions [8]. Introducing P^j, the projection operator onto spin j, the Bethe *ansatz* soluble model of spin s can be written

$$H_s = \sum_i \sum_{j=1}^{2s} a_j P^{(j)}(\boldsymbol{S}_i + \boldsymbol{S}_{j+1})$$

where

$$a_j \equiv \sum_{k=1}^{j} 1/k.$$

For the $s = 1$ case, this can be written

$$H^1 \propto \sum_i [\boldsymbol{S}_j \cdot \boldsymbol{S}_{i+1} - (\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1})^2 + \text{constant}].$$

These models are seen from the Bethe *ansatz* to have a unique ground state and no gap for all s. This seems to contradict our previous arguments for a gap in the integer-s case.

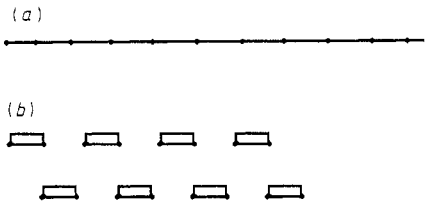

Figure 3. Valence-bond ground states for $s = 1$ chains: (*a*) the VBS; (*b*) the dimerised state.

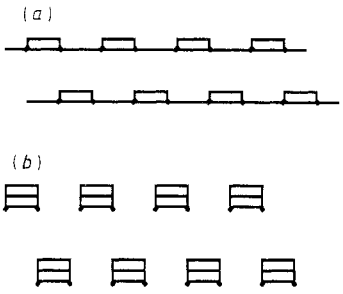

Figure 4. Valence-bond ground states for $s = \frac{3}{2}$ chains: (*a*) partially dimerised; (*b*) completely dimerised.

Fortunately, there is another class of solvable models whose exact ground states can be simply written in terms of 'valence bonds'. The first model of this type was discovered by Majumdar and Ghosh [21]. The spin is $\frac{1}{2}$ and the Hamiltonian is

$$H = \sum_i (\boldsymbol{S}_i + \boldsymbol{S}_{i+1} + \boldsymbol{S}_{i+2})^2 = \sum_i (J_1 \boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1} + J_2 \boldsymbol{S}_i \cdot \boldsymbol{S}_{i+2} + \text{constant})$$

with $J_2/J_1 = \frac{1}{2}$. Since the sum of three spin-$\frac{1}{2}$ variables has total spin $\frac{3}{2}$ or $\frac{1}{2}$, we see that ground states have the sum of any three neighbouring spins in an $s = \frac{1}{2}$ state. We can easily write down two such states for an even periodic chain. We do this by contracting every second pair of spins to form a singlet, i.e.

$$|\psi\rangle = |\alpha_1 \alpha_2 \alpha_3 \alpha_4 \ldots\rangle \varepsilon^{\alpha_1 \alpha_2} \varepsilon^{\alpha_3 \alpha_4} \ldots .$$

(Here α runs over 1, 2, representing $S^z = \pm\frac{1}{2}$ and repeated indices are summed.) It is convenient to represent such a state by a diagram with a line connecting the pair of sites contracted to form a singlet. The two ground states are shown in figure 2. Note that the spin-correlation function in these states is extremely short-range, i.e.

$$\langle \boldsymbol{S}_i \cdot \boldsymbol{S}_j \rangle = 0 \qquad \text{for } |i - j| > 1.$$

Thus, it is natural to expect the existence of a finite gap. In fact it was recently proven that these two ground states are the only ones and that there is a finite gap above them [11]. Parity is broken in these ground states, so the LSM theorem is obeyed.

A rather different type of solvable model, for integer s, was found in [11]. Again the Hamiltonian involves a projection operator onto a state of definite spin, and the ground state can be written simply in terms of valence bonds. As a preliminary, we need to generalise the notion of valence bonds to higher s. This can be done by regarding a spin-s operator as a symmetrised product of $2s$ spin-$\frac{1}{2}$ operators. Thus, for $s = 1$, we imagine two $s = \frac{1}{2}$ variables per site. We can construct singlet states by contracting these $s = \frac{1}{2}$ variables as above. We must never contract two spins on the same site with each other, since this would antisymmetrise, i.e. project out $s = 0$ instead of $s = 1$. Furthermore, any state that we obtain in terms of valence bonds must be symmetrised with respect to the two spins on each site. A convenient notation, then, is not to distinguish the two spins per site, but simply draw two lines emanating from each site, both of which must terminate on other sites. With this notation, a simple candidate ground state is almost obvious. We draw one line between each pair of neighbouring sites, as in figure 3(*a*). Since the structure of valence bonds mimics

the lattice structure, we refer to this as a valence-bond solid (VBS) state. The Hamiltonian for which this is the ground state is

$$H = \sum_i P^{(2)}(\boldsymbol{S}_i + \boldsymbol{S}_{i+1}) \propto \sum_i [\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1} + \tfrac{1}{3}(\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1})^2] + \text{constant}.$$

To see that the VBS state is indeed a ground state, simply observe that of the four $s = \frac{1}{2}$ variables associated with any pair of neighbouring sites, two of them are contracted to form a singlet. Symmetrising or antisymmetrising the other two gives spin 1 or 0, but never 2. Thus the ground-state energy is 0.

A very elegant algebraic representation of the VBS state was given by Arovas *et al* [22]. They used the Weyl representation of SU(2) involving polynomials of a unit vector (u, v) with complex components:

$$S^+ = u\partial/\partial v \qquad S^- = v\partial/\partial u \qquad S^z = (u\partial/\partial u - v\partial/\partial v)/2 \qquad |u|^2 + |v|^2 = 1.$$

Spin-s wavefunctions are polynomials of degree $2s$ in u and v. Such a unit vector can be represented by a point, $\boldsymbol{\Omega}$, on a sphere with polar coordinates θ, φ given by

$$u = \mathrm{e}^{\mathrm{i}\varphi/2}\cos(\theta/2) \qquad v = \mathrm{e}^{-\mathrm{i}\varphi/2}\sin(\theta/2).$$

The inner product of two wavefunctions can be written as the integral over the unit sphere. The VBS wavefunction becomes

$$\Psi = \Pi_i \sqrt{2}(u_i v_{i+1} - v_i u_{i+1}).$$

$|\Psi|^2$ can be written in terms of the corresponding points on the sphere:

$$|\Psi|^2 = \Pi_i(1 - \boldsymbol{\Omega}_i \cdot \boldsymbol{\Omega}_{i+1}).$$

It can be checked explicitly that in the space of polynomials of degree 2 in u and v, matrix elements of the spin operators are the same as those of

$$S^+ \to 4uv^* \qquad S^- \to 4vu^* \qquad S^z \to 2(v^*u - u^*v).$$

Thus we may replace

$$\boldsymbol{S}_i \cdot \boldsymbol{S}_j \to 8|u_i^* u_j + v_i^* v_j|^2 - 4(|u_i|^2 + |v_i|^2)(|u_j|^2 + |v_j|^2) = 4\boldsymbol{\Omega}_i \cdot \boldsymbol{\Omega}_j.$$

For a chain of N sites with periodic boundary conditions, the ground-state normalisation is

$$\int (\Pi_i \mathrm{d}^2\boldsymbol{\Omega}_i/4\pi)|\Psi|^2 = 1 + (-\tfrac{1}{3})^N$$

where we have used

$$\int (\mathrm{d}^2\boldsymbol{\Omega}_i/4\pi)\Omega_i^a \boldsymbol{\Omega}_i \cdot \boldsymbol{\Omega}_j = \Omega_j^a/3.$$

Ψ is normalised to unity at $N \to \infty$. Repeated use of this equation gives the spin-correlation function

$$\langle \boldsymbol{S}_0 \cdot \boldsymbol{S}_r \rangle = 4\langle \boldsymbol{\Omega}_0 \cdot \boldsymbol{\Omega}_r \rangle = (-1)^r 4 \times 3^{-r}.$$

This has the alternating sign characteristic of an antiferromagnet, but drops off exponentially with correlation length, $\xi = 1/\ln 3 \simeq 0.9$. A rigorous proof of a finite gap has also been given for this model [11].

Let us now consider the general bilinear–quadratic Hamiltonian

$$H = \sum_i [\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1} - \beta(\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1})^2].$$

At $\beta = 1$ the gap vanishes and at $\beta = -\frac{1}{3}$ it is finite. We would like to know what happens in between at $\beta = 0$ or, more generally, we would like to understand the phase diagram for arbitrary β. A hint can be obtained by considering the other simple ground state that can be formed out of valence bonds for an $s = 1$ chain: the dimerised state, with two valence bonds connecting every second pair of sites (figure 3(*b*)). It is again not hard to write down a Hamiltonian for which this is the ground state:

$$H = -\Sigma P^{(1)}(\boldsymbol{S}_i + \boldsymbol{S}_{i+1} + \boldsymbol{S}_{i+2}).$$

It seems likely that the dimer states are the only ground states, although this has not been proved. The correlation function is short-range and again it seems likely, in analogy to the $s = \frac{1}{2}$ Majumdar–Ghosh model, that there is a finite gap. These two simple valence-bond states, the VBS and dimerised states, suggest two possible phases for an $s = 1$ antiferromagnet: one with a unique ground state and one with two ground states and spontaneously broken translational symmetry. Both phases appear to have a finite correlation length and finite gap. The dimerised phase has dimer long-range order:

$$\langle(\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1})(\boldsymbol{S}_j \cdot \boldsymbol{S}_{j+1})\rangle \rightarrow (-1)^{i-j} \cdot \text{constant} + \text{constant}$$

although $\langle \boldsymbol{S}_i \cdot \boldsymbol{S}_j \rangle$ drops off exponentially. It has been argued [11, 23] that the Bethe *ansatz* integrable point, $\beta = 1$, corresponds to the transition point between the VBS and dimerised phases for the general nearest-neighbour bilinear–biquadratic model. Thus it represents a critical point and it is natural that the correlation length and inverse gap should diverge there.

A simple argument can be given [11] for this picture by using the VBS and dimer states as variational ground states for the Hamiltonian with arbitrary β. The variational energies of the VBS and dimerised states are respectively

$$E_{\text{VBS}} = -\tfrac{4}{3} - 2\beta \qquad E_{\text{dim}} = -1 - 8\beta/3.$$

For comparison, we note that the energy of the Neél state is

$$E_{\text{Néel}} = -1 - 2\beta.$$

Thus we see that the VBS state has lowest energy for $\beta < \frac{1}{2}$, and the dimerised state is lowest for $\beta > \frac{1}{2}$. This suggests that a transition between VBS and dimerised phases should occur. Of course, improving these variational states would change the estimate of the critical value of β. Since the Bethe *ansatz* tells us that the gap vanishes at $\beta = 1$, we hypothesised that this is the actual location of the critical point.

Note that the phase diagrams for the $s = \frac{1}{2}$ chain with first- and nearest-neighbour coupling and the $s = 1$ chain with bilinear and biquadratic interactions are rather similar. In both cases there are dimerised and undimerised phases. For $s = 1$ the gap vanishes only at the critical point whereas for $s = \frac{1}{2}$ it vanishes in the entire undimerised phase. The former behaviour is more normal, from the point of view of general critical phenomena. However, the latter behaviour must occur for half-integer s due to the LSM theorem. Critical theories of the phase transition have been developed in both cases [23, 24]. For $s = \frac{1}{2}$ the critical theory is the $k = 1$ Wess–Zumino–Witten non-linear σ-model (WZW model). (This corresponds to a special point in the set of free-

boson critical theories; it is equivalent to T_c in the Kosterlitz–Thouless phase diagram for the two-dimensional xy model.) The second nearest-neighbour interaction is a marginal operator, so that the entire undimerised phase is attracted to the fixed point and is massless. For $s = 1$ the critical theory is the $k = 2$ WZW model. In this case, the bilinear interaction is a relevant operator and the fixed point is repulsive, as for example, in a standard Ising transition. (The critical theory is *not* equivalent to an Ising critical point, however.)

We may attempt to understand in an intuitive manner the difference between the undimerised phases for $s = 1$ and $s = \frac{1}{2}$ in terms of valence-bond configurations, which form a complete set of singlet states for the chain. The massive, undimerised state for $s = 1$ may be thought of, in general, as consisting of the VBS state with an admixture of more complicated valence-bond arrangements, such that the probability of obtaining a long valence bond drops off exponentially. In the $s = \frac{1}{2}$ case, it is instructive to consider what happens as we reduce J_2/J_1 below $\frac{1}{2}$. More and more, longer and longer valence-bond configurations get mixed into the dimerised ground state until eventually the translational symmetry is restored. In this undimerised 'resonating valence-bond' state, the probability of obtaining long valence bonds only drops off as a power of the length. If only finite-length valence bonds were permitted it would be impossible to destroy the valence-bond order.

We may also ask to what extent the VBS state is a good approximation to the true ground state for the purely bilinear $s = 1$ chain. The bilinear Hamiltonian may be written

$$H = \sum_i [-2 + P^{(1)}(\boldsymbol{S}_i + \boldsymbol{S}_{i+1}) + 3P^{(2)}(\boldsymbol{S}_i + \boldsymbol{S}_{i+1})].$$

Insofar as the coefficient of $P^{(1)}$ is much less than that of $P^{(2)}$, this resembles the VBS model [22]. The ground-state energy for the Heisenberg Hamiltonian has been estimated from finite chain extrapolation as -1.40, whereas the VBS, dimerised and Néel energies for $\beta = 0$ are $-\frac{4}{3}$, -1 and -1, respectively [11]. Thus the VBS state has an energy which is only 5% higher than that of the true ground state. However, a more sensitive comparison of the two ground states is obtained by considering the correlation function at large distances. Numerical and experimental results, reviewed in §§ 7 and 8, give $\xi \simeq 5$ for the Heisenberg model. On the other hand, the VBS model has $\xi = 1/\ln 3 \simeq 0.9$. Furthermore, the mapping onto the O(3) non-linear σ-model discussed in § 3 implies that the correlation function should behave like that of a massive relativistic theory in $(1 + 1)$ dimensions:

$$G(\boldsymbol{r}) = \int [\mathrm{d}^2\boldsymbol{k}\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}/(2\pi)^2](\boldsymbol{k}^2 + m^2) \rightarrow_{r\rightarrow\infty} \text{constant} \cdot \mathrm{e}^{-mr}/r^{1/2}$$

whereas the exact result for the VBS model gives

$$G(r) = 4\mathrm{e}^{-r/\xi}$$

corresponding to $(1 + 0)$-dimensional behaviour. Haldane has argued that the former behaviour is generic to the undimerised phase and the latter only occurs at the solvable point, $\beta = \frac{1}{3}$, which therefore represents a type of singular point. The first and second moments of the dynamical structure function have been calculated for the VBS model and do not seem to agree very well with the experimental data, as discussed in § 8. Thus, while the VBS model provides an intuitive picture of what the massive undimer-

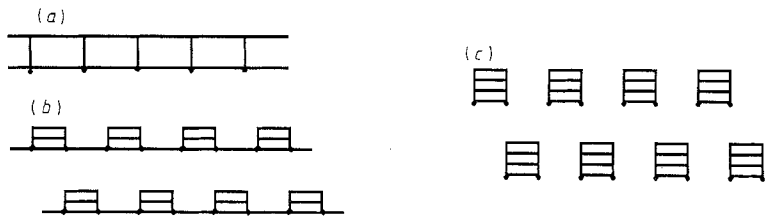

Figure 5. Valence-bond ground states for $s = 2$ chains: (*a*) VBS; (*b*) partially dimerised; (*c*) completely dimerised.

ised ground state might look like, it may not be very useful for more detailed calculations.

The distinction between higher integer and half-integer s can also be understood in terms of valence-bond states [11, 25]. If we consider $s = \frac{3}{2}$, we find four simple valence-bond states, two of which are completely dimerised, and two of which are partially dimerised (see figure 4). On the other hand, for $s = 2$ there are five states: two fully dimerised, two partially dimerised and the VBS state with unbroken translational symmetry and two valence bonds between each pair of neighbouring sites (see figure 5). In general, the VBS state, i.e. a nearest-neighbour valence-bond configuration with unbroken translational symmetry, only exists for integer s, whereas broken symmetry states exist for all s. Thus a translationally invariant state for half-integer s necessarily involves valence-bond 'resonation' whereas for integer s it does not. The correlation function decays exponentially in the VBS state for all s, and there is presumably a gap. However, the LSM theorem implies, in the half-integer-s case, that if the bonds resonate enough to restore the translational symmetry the gap necessarily vanishes.

6. Merons

I want to give a hand-waving argument that suggests how it is that $\theta = \pi$ might tend to suppress the mass which is present at $\theta = 0$ in the non-linear σ-model [26]. To make this argument it is necessary to establish some connection between topological effects and the generation of a mass at $\theta = 0$. This is actually a very controversial connection, both in this model and also in (3 + 1)-dimensional Yang–Mills theory. In my opinion, it is misleading to say that the mass is 'generated by instantons' as opposed to being generated by perturbative effects, as seen from the β-function. In particular, a mass is expected in the O(n) model for all n (and can be explicitly calculated at large n) but instantons only exist for $n = 3$. However, I will approach the O(3) model through a sequence of models in which the infrared divergences of perturbation theory are cut off, and the mass is clearly generated by topological effects. We may regard the O(3) model as the limit of a sequence of models in which the mass is generated by topological effects. We will then see that $\theta = \pi$ tends to suppress this topological mass-generation mechanism.

The infrared cut-off is provided by an anisotropic coupling. In the spin-chain language this is

$$H = J \sum_i (\boldsymbol{S}_i \cdot \boldsymbol{S}_{i+1} - bS_i^z S_{i-1}^z).$$

In the large-s limit we obtain the non-linear σ-model with a potential energy term

$$V = m^2(\varphi^z)^2/2g$$

(where $m^2 = -2b$ and I have set the velocity to 1). We wish to consider the case of

planar anisotropy, $b < 0$, $m^2 > 0$. The spins tend to lie in the xy plane. Semi-classically, the remaining U(1) symmetry gets spontaneously broken and some ordering direction is picked out on the plane. However, this cannot really occur in (1 + 1) dimensions, just as it cannot in the O(3) case, due to the infrared divergences connected with the resulting Goldstone bosons.

What *can* occur in (1 + 1) dimensions is the Kosterlitz–Thouless phase with no broken symmetry, but power-law correlations and a massless boson which is *not* a Goldstone boson. By going to polar coordinates we can see that this is what occurs at weak coupling:

$$\mathcal{L} = (1/2g)[(\partial_\mu \alpha)^2 + \sin^2 \alpha (\partial_\mu \beta)^2 + m^2 \cos^2 \alpha]. \tag{3}$$

Expanding around the minimum, $\alpha = \pi/2 + \alpha'$, we obtain

$$\mathcal{L} \simeq (1/2g)[(\partial_\mu \alpha')^2 + m^2 \alpha'^2 + (\partial_\mu \beta)^2].$$

Here α' is a massive boson, but β is massless. However, the symmetry is not spontaneously broken. To see this note that

$$\langle S_i^+ S_j^- \rangle (-1)^{i-j} \propto \langle (\varphi_1 + i\varphi_2)(x)(\varphi_1 - i\varphi_2)(y) \rangle / 2 \simeq \langle e^{i\beta(x)} e^{-i\beta(y)} \rangle \propto e^{gD(x-y)}$$

where D is the free massless boson propagator. In a higher space-time dimension, $D(x - y) \to 0$ at ∞ so we have long-range order. However, in (1 + 1) dimensions

$$D(x - y) = (-1/2\pi) \ln |x - y| + \text{const.}$$

so the correlation function has power-law decay

$$\langle e^{i\beta(x)} e^{-i\beta(y)} \rangle \propto |x - y|^{-g/2\pi} \to 0.$$

Thus the symmetry is unbroken. How weak must the coupling be for us to trust this conclusion? In the isotropic model ($m = 0$), no matter how weak the coupling is at small scales, it always grows large eventually so we can never trust perturbation theory regarding the existence of massless particles (the ultimate long-distance question). However for $m \neq 0$, the infrared divergences are cut off by the scale m. We may integrate out the massive α' boson to obtain an effective Lagrangian for β. This effective Lagrangian cannot contain any infrared dangerous interactions, simply because no renormalisable interactions are permitted by the U(1) symmetry $\beta \to \beta +$ constant. The real expansion parameter for this calculation is

$$g_{\text{eff}}(m) \simeq g/[1 + (g/2\pi) \ln ma]$$

where a is the ultraviolet cut-off, essentially the lattice spacing of the spin chain. However, if we decrease m to zero (with fixed bare coupling, g) we eventually expect a transition to a massive phase, since the isotropic $m = 0$ model is in such a phase. We wish to argue that this phase transition, which occurs at finite m, for fixed g, can be regarded as a consequence of topological effects, and may therefore be altered by the presence of a topological term. The transition can actually be induced by lowering m at fixed g, or alternatively by increasing g at fixed m. Either way $g_{\text{eff}}(m)$, the important dimensionless parameter, is increased. The expected phase diagram is shown in figure 6. It is actually simplest to think about the transition as a function of g at very large m (i.e. $ma \gg 1$).

I have been implicitly discussing the theory in terms of bare parameters with some assumed ultraviolet cut-off, a. While the initial physical motivation was the one-dimensional quantum spin chain, with a being its lattice spacing, it is actually convenient

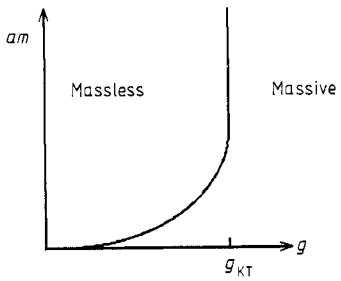

Figure 6. Phase diagram for the ferromagnet with easy plane anisotropy.

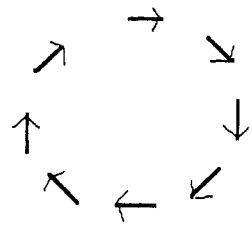

Figure 7. A vortex.

to consider a Euclidean space regularisation of the σ-model instead. Thus we consider a two-dimensional classical ferromagnet, at temperature g:

$$S = (1/g)\left(-\sum_{\langle i,j\rangle} \boldsymbol{\varphi}_i \cdot \boldsymbol{\varphi}_j + (am)^2 \sum_i (\varphi^z)^2/2\right)$$

on a lattice of spacing, a.

In the limit $am \to \infty$, the spins $\boldsymbol{\varphi}$ lie strictly in the xy plane, and s may be rewritten in terms of β:

$$S \to -(1/g)\sum_{\langle i,j\rangle} \cos(\beta_i - \beta_j).$$

This is the famous classical 2D xy model. In the continuum limit it just becomes a free field theory plus non-renormalisable (irrelevant) interactions:

$$S \to (1/2g)\int \mathrm{d}^2\boldsymbol{x}\,(\nabla\beta)^2. \tag{4}$$

However, it is believed that the boson obtains a mass for sufficiently strong g due to topological effects [4]. The topology in this case is vortices, configurations where the spin at ∞ rotates with the spatial direction as in figure 7. Such a configuration has a clear tendency to disorder the system since the spins at infinity are pointing in an indefinite direction. However, the vortex has an infrared-divergent action

$$S \simeq (1/2g)\int \mathrm{d}^2\boldsymbol{x}\,(\nabla\beta)^2 \simeq (1/2g)\int \mathrm{d}^2\boldsymbol{x}\,(1/r^2) \simeq (1/2g)(\text{constant} + 2\pi \ln La)$$

where L is the length of the system. The continuum action of equation (4) would also give an *ultraviolet* divergence, but this is cut off by the finite lattice spacing, a. We get a finite action for a vortex–antivortex pair, or more generally a collection of vortices with net vorticity zero. The vorticity, v, of a region is defined by the line integral

$$v = (1/2\pi)\int \mathrm{d}\boldsymbol{s}\cdot\nabla\beta.$$

(The vector $\mathrm{d}\boldsymbol{s}$ is tangent to the curve.) The minimum action, or 'classical', field configuration, for a specified set of vortices and antivortices at locations $\boldsymbol{x}_i$, is obtained

by minimising S or solving Laplace's equation, $\nabla^2\beta = 0$, with singularities at the vortices. This gives

$$\beta_c(\boldsymbol{x}) = \sum_i v_i \tan^{-1}[(y - y_i)/(x - x_i)]$$

where $v_i = \pm 1$, for a vortex or antivortex. Far from all vortices, β_c goes to a constant if the net vorticity is zero.

It is actually more convenient to introduce the dual field $\tilde{\beta}$ defined by

$$\varepsilon^{\mu\nu}\partial_\nu\tilde{\beta} = \partial_\mu\beta.$$

$\tilde{\beta}_c$ also obeys Laplace's equation away from the vortex cores and the vorticity of a region becomes

$$v = -(1/2\pi)\int \mathrm{d}\boldsymbol{S}\cdot\nabla\tilde{\beta}$$

where $\mathrm{d}\boldsymbol{S}$ is *normal* to the curve. This can be rewritten

$$v = -(1/2\pi)\int \mathrm{d}^2\boldsymbol{x}\,\nabla^2\tilde{\beta}$$

and hence $\tilde{\beta}_c$ obeys Poisson's equation

$$\nabla^2\tilde{\beta}_c = -2\pi\rho(\boldsymbol{x}) = -2\pi\sum_i v_i\delta^2(\boldsymbol{x} - \boldsymbol{x}_i).$$

The action for a classical field configuration β_c can be written in terms of $\tilde{\beta}_c$, since $(\nabla\beta)^2 = (\nabla\tilde{\beta})^2$.

A standard, if not completely justified, approximation is to restrict the functional integral over $\beta(\boldsymbol{x})$ to a dilute gas of vortices plus small fluctuations. We assume that the vortices are far apart compared to the lattice spacing, so we may treat them as point-like, with the lattice providing an ultraviolet cut-off. Thus we write

$$\beta = \beta_c + \beta'$$

where β_c is the classical field configuration and β' represents the small (single-valued) fluctuation. The action becomes

$$S = (1/2g)\int \mathrm{d}^2\boldsymbol{x}\,[(\nabla\beta_c)^2 + (\nabla\beta')^2]. \tag{5}$$

It is convenient to replace β_c in S by the vortex density, $\rho(\boldsymbol{x})$:

$$S = (1/2g)\int \mathrm{d}^2\boldsymbol{x}\,[(\nabla\tilde{\beta})^2 + 4\pi\mathrm{i}\rho\tilde{\beta}]).$$

Shifting $\tilde{\beta}$ by $\tilde{\beta}_c$ gives the same action as in equation (5) with $\tilde{\beta} = \tilde{\beta}_c + \tilde{\beta}'$.

The next step is to evaluate the sum over all vortex configurations *before* integrating over $\tilde{\beta}$. We use

$$\mathrm{e}^{-S} = \exp\left(-(1/2g)\int \mathrm{d}^2\boldsymbol{x}\,(\nabla\tilde{\beta})^2\right)(1/n!)\prod_{i=1}^{n}\int \mathrm{d}^2\boldsymbol{x}_j\exp[-(2\pi\mathrm{i}/g)\tilde{\beta}(\boldsymbol{x}_j)]$$

$$\times\,(1/m!)\prod_{k=1}^{m}\int \mathrm{d}^2\boldsymbol{x}_k\exp[(2\pi\mathrm{i}/g)\tilde{\beta}(\boldsymbol{x}_k)].$$

Here j and k label the vortices and antivortices. The integrals over the vortex locations now factorise, and the sum over the numbers of vortices and antivortices exponentiates to give

$$S(\tilde{\beta}) = \int \mathrm{d}^2 x \{(1/2g)(\nabla\tilde{\beta})^2 - \gamma[\exp(-2\pi i\tilde{\beta}(x)/g) + \exp(2\pi i\tilde{\beta}(x)/g]/2a^2\}.$$

Here, γ is a regularisation-dependent constant. We have obtained the sine–Gordon model with Lagrangian

$$\mathcal{L} = (1/2g)(\nabla\tilde{\beta})^2 - (\gamma/a^2)\cos 2\pi\tilde{\beta}(x)/g.$$

At weak coupling, the interaction can be ignored, but at a critical value of g it produces a phase transition. Physically, we can think of the vortices as coming in dilute tightly bound pairs at small g which do not disorder the system very much. At a critical value of β the vortex gas becomes dense and unbound, producing disorder. This can be understood from the renormalisation group as applied to the sine–Gordon model. We need the scaling dimension of the interaction term. The free field theory two-point function is

$$\langle \exp[2\pi i\tilde{\beta}(x)/g] \exp[-2\pi i\tilde{\beta}(y)/g]\rangle = \exp[(2\pi)^2 D(x-y)/g]$$

$$= \exp[-(2\pi/g)\ln r/a] \propto (a/r)^{2\pi/g}.$$

The interaction has scaling dimension, π/g. Thus this becomes relevant at $g_c = \pi/2$. For smaller values of g perturbation theory in γ is infrared-finite so we may simply ignore the vortices, and hence the sine–Gordon interaction at large distances. Or, putting this another way, if we rescale the lattice spacing, $a \to a'$, then the operator $\exp[2\pi i\tilde{\beta}(x)/g]$ is rescaled by $(a'/a)^{\pi/g}$. Thus the coupling constant γ is rescaled by

$$\gamma(a) = (a'/a)^{\pi/g-2}\gamma(a').$$

Thus $\gamma(a)$ grows small at large distance scales for $g < g_c$. This is the massless phase. The correlation functions have power-law decay, corresponding to 'quasi-long-range order'. If we rescale $\tilde{\beta}$ by $\sqrt{g}$, we have, at large g

$$\mathcal{L} = \tfrac{1}{2}(\nabla\tilde{\beta})^2 + (2\pi^2\gamma/a^2 g)\tilde{\beta}^2 + \mathrm{O}(1/g^{3/2})$$

i.e. $\tilde{\beta}$ is a weakly coupled massive field. We now have exponential decay.

Thus we see that in the large-m limit, the O(3) σ-model (with an ultraviolet regularisation which permits vortices) has a finite-coupling phase transition between massless and massive phases due to topological effects. We now wish to ask what happens as the mass is decreased relative to the lattice spacing. The essential feature is that the field φ is no longer forced to lie in the plane. This implies the existence of ultraviolet finite vortices in the continuum limit. While the purely planar action

$$S = (1/2g)\int \mathrm{d}^2 x\,(\nabla\beta)^2$$

necessarily has a singularity for a vortex configuration, the action of equation (3) need not. We simply require that $\alpha \to 0$ or π at the centre of the vortex, i.e. the field, φ, which lies on the equator at $x \to \infty$, moves to the north or south pole in the vortex

core. This vortex core has a size $\simeq m^{-1} \gg a$ in the small-m limit and the action is of the form

$$S = (1/g)(\text{constant} + 2\pi \ln mL).$$

As mentioned above, we may integrate out the α-field to obtain an effective Lagrangian for the massless β-field. This contains an effective coupling $g(m)$. We expect the vortex condensation transition to occur at $g(m) \simeq \pi/2$ as before. Now let us consider the limit of a small but fixed bare coupling, g, and a varying m. As we reduce m the transition to the massive phase (i.e. the vortex condensation) occurs when $ma \simeq O(e^{-2\pi/g})$. Thus we obtain a phase diagram as in figure 6. At larger values of m perturbation theory applies, and we have the massless phase. At smaller m perturbation theory breaks down, $g(m)$ becomes large, and vortex condensation occurs. Thus it is probably correct to say that the mass gap in the theory for m slightly less than the critical value is caused by topological effects. However, as we reduce m right down to zero, the isotropic point, this description loses its meaning.

So far we have not discussed the effects of the θ-term in the σ-model action. In the large-m limit where the field $\boldsymbol{\varphi}$ always lies in the plane and is discontinuous in the vortex core at the lattice scale, the θ-term does not play any role. (In fact it is not very well defined.) However, for $ma \ll 1$, where the vortex becomes a continuous configuration and $\boldsymbol{\varphi}$ lifts off the plane, the θ-term becomes very important. We now must distinguish two types of continuum vortices; those which have $\boldsymbol{\varphi} \to \pm\hat{z}$ at their cores. The topological charge of a vortex (not to be confused with its vortex number) is given by

$$q = (1/4\pi) \int \mathrm{d}\boldsymbol{s} \cdot (\boldsymbol{\nabla}\beta) \cos \alpha.$$

An instanton configuration has $\alpha \to 0,\pi$ at $r \to 0,\infty$, and q gets equal contributions of $\frac{1}{2}$ from the line integrals at the origin and at ∞. However, the continuum vortex has $\alpha = 0,\pi/2$ at $r \to 0,\infty$ and thus only the line integral at $r \to 0$ contributes to q, giving $q = \frac{1}{2}$. Thus a vortex is half an instanton and is sometimes called a meron. Now, when we sum over the vortex gas we must sum over the two types of vortex and two types of antivortex ($v = \pm 1$, $q = \pm 1$). The two types of vortex are weighted by a phase $e^{\pm i\theta/2}$ so that the sine–Gordon interaction now consists of four terms from the four objects:

$$S(\tilde{\beta}) = \int \mathrm{d}^2x \{(1/2g)(\nabla\tilde{\beta})^2 - \gamma(e^{i\theta/2} + e^{-i\theta/2})[\exp(-2\pi i\tilde{\beta}(x)/g) + \exp(2\pi i\tilde{\beta}(x)/g)]/4\}.$$

The effect of the topological term is simply to rescale the (regularisation-dependent) parameter γ by

$$\gamma \to \gamma \cos \theta/2.$$

With increasing θ the vortices become less and less effective. (The correlation length in the condensed phase increases.) Finally, at $\theta = \pi$ we seem to find that the vortices are never effective. The contributions of the two types of vortices (with $q = \pm 1$) are weighted by factors of $\pm i$ and precisely cancel each other.

We might be tempted to conclude from this that the theory will remain in the massless phase all the way to the isotropic point, $m = 0$. However, this is a bit naive

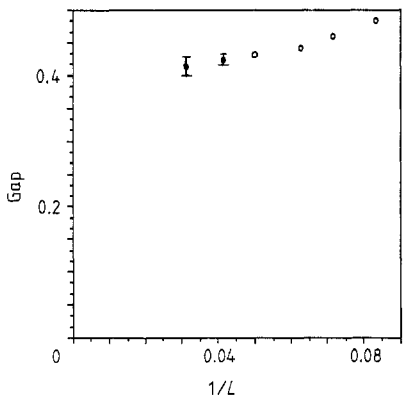

Figure 8. Gap versus length for $s = 1$ Heisenberg model (from [24]).

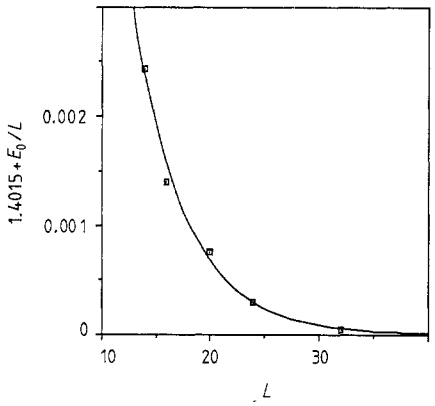

Figure 9. Ground-state energy, E_0, versus length for the $s = 1$ Heisenberg chain. The points are data from exact [23] and Monte Carlo [24] diagonalisation. The curve is $0.0448e^{-L/\xi}$, with $\xi = 4.78$.

for several reasons. There are double vortices with $v = \pm 2$ and $q = \pm 2$ or 0. These do not cancel each other. However, since v is twice as big, this raises the critical value of $g(m)$ by roughly a factor of 4. Whether or not $g(m)$ ever gets this big for any $m > 0$ is not obvious. In any event we can conclude that the critical value of m (for fixed g) is smaller at $\theta = \pi$.

It is also interesting to consider what happens for negative m. For large negative m where $g(|m|) \ll 1$, perturbation theory is again valid, and we have $\langle \varphi \rangle = \pm \hat{z}$, and the spectrum consists of two massive bosons. This Ising transition should occur at a value of $m \simeq -M$ in the $\theta = 0$ case, so there are two phase transitions at $m \simeq \pm M$. The LSM theorem suggests that, for $\theta = \pi$, there is a single transition at the isotropic point separating the massless and ordered phases.

7. Numerical results

The gap has been calculated using exact diagonalisation [27] for chains of length $L \leqslant 14$ and Monte Carlo techniques [28] for L as large as 32 (see figure 8). While the exact results by themselves are ambiguous, together with the Monte Carlo results they rather strongly suggest a gap of about 0.41 J. It is also possible to estimate the correlation length by fitting the ground-state energy to

$$E_0/L \simeq e_0 + Ae^{-L/\xi}.$$

This gives an estimate $\xi \simeq 4.8$ (see figure 9), in good agreement with the relativistic result $\Delta = v/\xi$, using the lowest-order spin-wave velocity, $v = 2$ J.

Exact diagonalisation results are also available [29] for the gap to the lowest singlet excited state (which has the same parity as the ground state but momentum π relative to it), for various values of the biquadratic coupling constant, β, and $L \leqslant 16$. While the scaling with L is not very conclusive, the dependence on β lends support to the conjecture that, with increasing β, the gap vanishes for $\beta \geqslant 1$. Recall that this low energy state for a finite chain is essentially the antisymmetric linear combination of

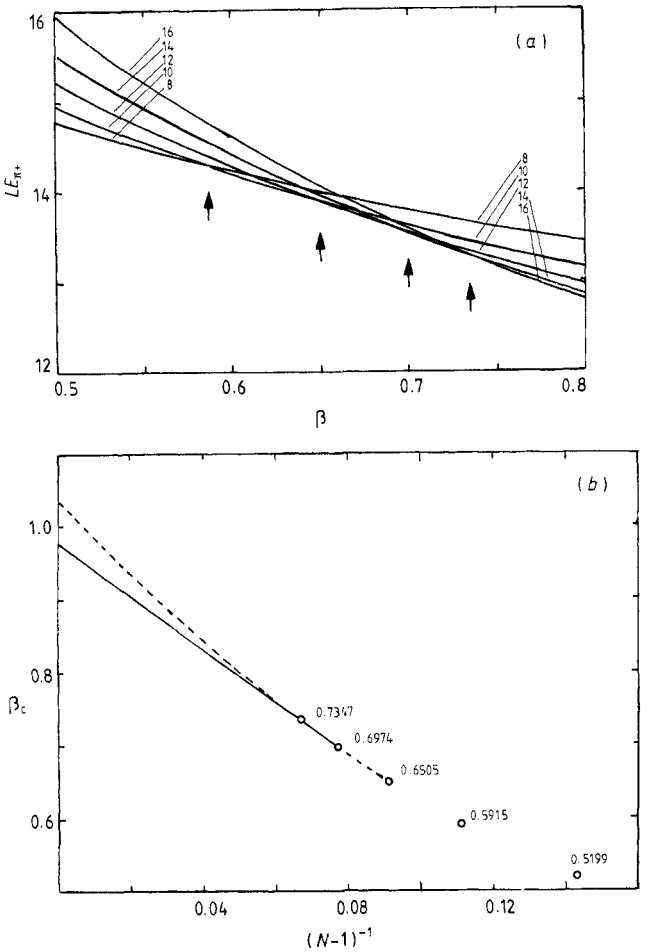

Figure 10. (*a*) Scaled singlet excitation gap, $L(E_s - E_0)$, versus β for the general bilinear–biquadratic $s = 1$ model, for various values of L. (*b*) The crossing points of the L and $L + 2$ curves plotted versus $1/(L + 1)$. Linear and quadratic extrapolation are indicated by the full line and broken curve, respectively.

the two infinite-chain ground states, the finite-chain ground state being the symmetric combination. It is found that the scaled gap curves, $L\Delta E(\beta)$, for successive even values of L, cross at various values of $\beta < 1$. As $L \to \infty$ the crossing point should approach the critical value of β, since $L\Delta E \to$ constant for $\beta < \beta_c$ and $L\Delta E \to L\mathrm{e}^{-\text{constant}\cdot L}$ for $\beta > \beta_c$. Extrapolating the crossing points leads to an estimate of β_c of 1.01 ± 0.03 (see figure 10).

A third approach [30] is to study the finite size scaling as a function of the anisotropy parameters, J_z and D:

$$H = \sum_i [S_i^x S_{i+1}^x + S_i^y S_{i+1}^y + J_z S_i^z S_{i+1}^z + D(S_i^z)^2].$$

The Haldane phase with a finite gap and no broken symmetries was found to exist (for $D = 0$) for a range of J_z between approximately 0 and 1.2. At small negative J_z a gapless planar phase exists and at $J_z > 1.2$ a massive Ising antiferromagnetic phase

occurs, essentially of the type predicted by ordinary spin-wave theory. (Note that this is possible, even in one dimension since only a discrete symmetry, $S^z \rightarrow -S^z$ is spontaneously broken.) For $J_z = 1$, the Haldane phase persists for arbitrary positive D, except for one special point ($D \simeq 1$) where the gap vanishes. At large D this result is more or less trivial because the ground state has $S_i^z = 0$ for all sites, and the gap is approximately D. For negative D a transition to the Ising ordered phase occurs at $D \simeq -0.2$. Note that these numbers are directly relevant to the question of whether various materials have a gap due to anisotropy or due to the Haldane mechanism.

8. Experimental tests of the Haldane conjecture

Two different materials have been studied which represent good candidates for experimental realisation of the Haldane conjecture, $CsNiCl_3$ and $Ni(C_2H_8N_2)_2NO_2ClO_4$. They are both quasi-one-dimensional spin-1 antiferromagnets.

The magnetic Ni ions in $CsNiCl_3$ form chains with a strong Ni^{2+}–Cl^-–Ni^{2+} super-exchange coupling, $J \simeq 33$ K (note that our convention for J differs from that in [12] by a factor of 2). The chains form a triangular lattice and the weaker Ni^{2+}–Cl^-–Cl^-–Ni^{2+} super-exchange coupling between Ni ions on neighbouring chains is estimated as $J'/J \simeq 0.018$. There is a crystal-field anisotropy, modelled by a term in the Hamiltonian:

$$H \rightarrow H + D \sum_i (S^z)^2.$$

$CsNiCl_3$ appears to undergo two Néel ordering transitions, at $T = 4.84$ K and 4.40 K. The ground state for an isotropic classical antiferromagnet on a triangular lattice has tripartite order with the spins on the three different sublattices in a given plane pointing in directions differing by $2\pi/3$. The spin directions alternate in between planes. Essentially this type of order is observed in the lowest-temperature phase. In the intermediate phase only the z-components of the spins are ordered. This also occurs [31] in classical 'soft-spin' models with $D < 0$. D was estimated from the spin-flop field [30] as -0.12 K, or from the angle between spin ordering directions [32], 121°, as -0.05 K.

The main evidence for the Haldane conjecture comes from neutron scattering experiments [12] in the high-T phase, with $T \simeq 10$ K. A narrow peak is seen at wavectors with $k_z = \pi$, and gaps of about 0.4 J. However there is a large transverse dispersion, the gap ranging from 0.29 J to 0.65 J. Polarised neutron scattering experiments [13, 14] show that the excitation is, to within experimental resolution, a spin triplet, as predicted by Haldane's theory. A naive application of spin-wave theory to decoupled chains would also predict a finite gap for $D < 0$ but it would have xy polarisation and, using the above estimates of D, would be at least nine times too small [12]. (The observed neutron scattering intensity [14] for $\pi/5 < k_z < \pi/2$, seems to disagree with the exact first moment calculated [22] in the VBS model, suggesting that this model is not a very good approximation to the Heisenberg Hamiltonian.)

$Ni(C_2H_8N_2)_2NO_2ClO_4$ has a super-exchange coupling between Ni^{2+} atoms of $J \simeq 55$ K. The inter-chain coupling is estimated to be $J'/J \simeq 4 \times 10^{-4}$. No indications of Néel order are seen down to at least 1.2 K. Again a finite gap is measured [15] at $k_z = \pi$. Actually two different peaks are seen with excitation energies of 0.27 J and 0.53 J, the average being 0.4 J. The transverse dispersion of the gaps is small (about 15%). If we assume that the two branches are the triplet, split by anisotropy, then we must

conclude that the effective anisotropy is quite large, again somewhat obscuring the interpretation of the data in terms of the Haldane gap.

Recently, a detailed model for inter-chain couplings in quasi-one-dimensional integer-s antiferromagnets has been proposed, which contains the Haldane gap [33]. This offers a possible resolution of inconsistencies between the low-T ordered-phase neutron scattering data on $CsNiCl_3$ and conventional spin-wave theory and otherwise reinforces the interpretation of the higher-T data as evidence for the Haldane gap.

Acknowledgment

The research reported here was partially supported by NSERC of Canada.

References

[1] Bethe H A 1931 *Z. Phys.* **71** 205
[2] Anderson P W 1952 *Phys. Rev.* B **86** 694
[3] Kubo R 1952 *Phys. Rev.* **87** 568
[4] Kosterlitz J M and Thouless D 1973 *J. Phys. C: Solid State Phys.* **6** 1181
Kosterlitz J M 1974 *J. Phys. C: Solid State Phys.* **7** 1046
[5] Luther A and Peschel I 1975 *Phys. Rev.* B **12** 3908
[6] Haldane F D M 1983 *Phys. Lett.* **93A** 464; 1983 *Phys. Rev. Lett.* **50** 1153; 1985 *J. Appl. Phys.* **57** 3359
[7] Botet R and Jullien R 1983 *Phys. Rev.* B **27** 613
Kolb M, Botet R and Jullien R 1983 *J. Phys. A: Math. Gen.* **16** L673
Botet R, Jullien R and Kolb M 1984 *Phys. Rev. B* **30** 215
Parkinson J B and Bonner J C 1985 *Phys. Rev.* B **32** 4703
Nightingale M P and Blote H W J 1986 *Phys. Rev.* B **33** 659
[8] Kulish P and Sklyanin E 1982 *Lecture Notes in Physics* vol 151 (Berlin: Springer) p 61
Kulish P, Reshetikhin N Yu and Sklyanin E 1981 *Lett. Math. Phys.* **5** 393
Takhtajan L 1982 *Phys. Lett.* **87A** 479
Babudjian J 1982 *Phys. Lett.* **90A** 479; 1983 *Nucl. Phys.* B **215** 317
[9] Lieb E H, Schultz T and Mattis D J 1961 *Ann. Phys., NY* **16** 407
[10] Affleck I and Lieb E H 1986 *Lett. Math. Phys.* **12** 57
[11] Affleck I, Kennedy T, Lieb E H and Tasaki H 1987 *Phys. Rev. Lett.* **59** 799; 1988 *Commun. Math. Phys.* **115** 477
[12] Buyers W J L, Morra R M, Armstrong R L, Gerlach P and Hirakawa K 1986 *Phys. Rev. Lett.* **56** 371
Morra R M, Buyers W J L, Armstrong R L and Hirakawa K 1988 *Phys. Rev.* B **38** 543
[13] Steiner M, Kakurai K, Kjems J K, Petitgrand D and Pynn R 1987 *J. Appl. Phys.* **61** 3953
[14] Tun Z, Buyers W J L, Armstrong R L, Hallman E D and Arovas D P 1989 *Proc. Int. Conf. Magnetism (Paris) 1988; J. Physique Coll.* to appear
[15] Renard J P, Verdaguer M, Regnault L P, Erkelens W A C, Rossat-Mignod J and Stirling W G 1987 *Europhys. Lett.* **3** 945
[16] Sieberg N private communication
[17] Kittel C 1963 *Quantum Theory of Solids* (New York: Wiley)
[18] Affleck I 1985 *Nucl. Phys.* B **257** 397
[19] Brezin E and Zinn-Justin J 1976 *Phys. Rev.* B **14** 3110
[20] Zamalodchikov A B and Zamalodchikov A B 1979 *Ann. Phys., NY* **120** 253
[21] Majumdar C K and Ghosh D K 1969 *J. Math. Phys.* **10** 1388
[22] Arovas D P, Auerbach A and Haldane F D M 1988 *Phys. Rev. Lett.* **60** 531
[23] Affleck I and Haldane F D M 1987 *Phys. Rev.* B **36** 5291
[24] Affleck I 1985 *Phys. Rev. Lett.* **55** 1355; 1986 *Nucl. Phys.* B **265** 409
[25] Affleck I 1985 *Phys. Rev. Lett.* **54** 966
[26] Nelson D R and Pelcovits R A 1977 *Phys. Rev.* B **16** 2191
Haldane F D M unpublished
Affleck I 1986 *Phys. Rev. Lett.* **56** 408

[27] Parkinson J B and Bonner J C 1985 *Phys. Rev.* B **32** 4703
[28] Chang K, Affleck I, Hayden G and Soos Z 1989 *J. Phys.: Condens. Matter* **1** 153
[29] Botet R and Jullien R 1983 *Phys. Rev.* B **27** 613
Schulz H J and Ziman T 1986 *Phys. Rev.* B **33** 6545
[30] Kadowaki H, Ubukoshi K and Hirakawa K 1987 *J. Phys. Soc. Japan* **56** 751
[31] Plumer M L and Caille A to be published
Plumer M L and Caille A 1988 *Phys. Rev. Lett.* **60** 45
[32] Yelon W B and Cox D E 1973 *Phys. Rev. B* **7** 2024
[33] Affleck I 1989 *Phys. Rev. Lett.* **62** 474