# A Unified Theory Based on *SO*(5) Symmetry of Superconductivity and Antiferromagnetism

Shou-Cheng Zhang

The complex phase diagram of high–critical temperature ($T_c$) superconductors can be deduced from an *SO*(5) symmetry principle that unifies antiferromagnetism and *d*-wave superconductivity. The approximate *SO*(5) symmetry has been derived from the microscopic Hamiltonian, and it becomes exact under renormalization group flow toward a bicritical point. This symmetry enables the construction of a *SO*(5) quantum nonlinear σ model that describes the phase diagram and the effective low-energy dynamics of the system. This model naturally explains the basic phenomenology of the high-$T_c$ superconductors from the insulating to the underdoped and the optimally doped region.

The high-$T_c$ superconductors are among the most complex systems studied in condensed-matter physics. Anderson (*1*) and Zhang and Rice (*2*) argued rather successfully that a good starting point for modeling the strong correlation effects in the oxides should be the Hubbard, or the *t-J* models close to half-filling (in these models, the repulsion energy $U$ and the antiferro-magnetic exchange constant $J$ are related as $J = t^2/U$, where $t$ is the site-hopping matrix element). Unfortunately, the exact solutions to these models in one dimension shed very little light on how to solve them in higher dimensions.

Even within the much simplified Hubbard or *t-J* models near half-filling, there are two different energy scales in the problem. A "high-energy physics" is responsible for forming local singlet pairs that basically originate from the $J$ term in the Hamiltonian $H$. A mean-field temperature $T_{MF}$ can be associated with this energy scale, that can be between 0 and 1000 K, depending on the filling factor. For conventional BCS (Bardeen-Cooper-Schrieffer) superconductors, the energy scale of the pair formation is the same as the true phase-transition temperature into a superconducting (SC) state. However, for the oxides, the high-energy physics does not reveal the true ground state of the system.

In 1987, Lee and Read (*3*) asked a thought-provoking question, "Why is the $T_c$ so low?" If the singlet formation occurs at the SC transition, the natural transition temperature would be $T_{MF}$, which is much higher than the observed $T_c$. The reason that $T_c$ is low is because there is also other "low-energy physics," valid below $T_{MF}$, that governs the fate of the singlet pairs. The singlet pairs can: (i) form a spatially ordered state [an antiferromagnetic (AF) state], (ii) condense to form a *d*-wave superconductor, (iii) arrange themselves into a spatially non-uniform state [phase separation (*4*)], (iv) form a spatially homogenous mixed state of coexisting AF and SC order [a "spin-bag phase" (*5*)], or (v) disorder if the effects of quantum fluctuation are strong enough [a "resonating valence-bond (RVB) phase" (*1*)]. In order to distinguish the different competing ground states, an effective $H$ is needed that describes the low-energy physics sector of the *t-J* model below $T_{MF}$. The form of $H$ should be simple so that it can be solved analytically, and yet the complexity of the possible phases should emerge from the unity of the model.

The different competing orders are not separated by distinct energy scales in the remaining problem, so it is not obvious how one can systematically apply the renormalization group idea to integrate out the irrelevant degrees of freedom. Fortunately, the strongly interacting low-energy degrees of freedom are related by symmetry properties. The main approach here is to identify the symmetries of the microscopic $H$, such as the *t-J* model, and use them to constrain the possible form of the low-energy effective $H$.

What symmetries are known? Chakravaty, Halperin, and Nelson (*6*), showed that the low-energy effective $H$ of an antiferromagnet is the *SO*(3) quantum nonlinear σ model. More recently, Pines and coworkers (*7, 8*), and Chubukov, Sachdev, and Ye (*9*) argued that this model is applicable in the underdoped regime as well. The low-energy effective $H$ of a superconductor is described by a *U*(1) quantum nonlinear σ model, sometimes called the *XY* model. Doniach and Inui (*10*) attempted to describe the metal-insulator transition in terms of a quantum *XY* model, and recently Emery and Kivelson (*11*) presented evidence that the SC transition on the underdoped side of the oxides can be described by a renormalized classical *XY* model. Both the *SO*(3) spin rotation and the *U*(1) charge rotation symmetries are obvious symmetries of the microscopic *t-J* model and constrain any new, unified low-energy theory.

The simplest way to construct a unified theory of AF and SC is to introduce a concept that I call superspin. It is a five-dimensional (5D) vector, $n_a = (n_1, n_2, n_3, n_4, n_5)$. The first and the fifth components are the superconducting components of the superspin, identified with the two *d*-wave SC-order parameters (*8, 12*):

$$n_1 = \Delta^\dagger + \Delta \tag{1a}$$

$$n_5 = -i(\Delta^\dagger - \Delta) \tag{1b}$$

where

$$\Delta^\dagger = \tfrac{1}{2}\sum_p g(p) c^\dagger_{p\uparrow} c^\dagger_{-p\downarrow} \tag{1c}$$

and

$$g(p) = \cos p_x - \cos p_y \tag{1d}$$

($c^\dagger_p$ creates electron with momentum $p$). The remaining three components are the spin components of the superspin, identified with the AF order parameter:

$$n_2 = \sum_p c^\dagger_{p+Q,i} \sigma^x_{ij} c_{p,j} \tag{2a}$$

$$n_3 = \sum_p c^\dagger_{p+Q,i} \sigma^y_{ij} c_{p,j} \tag{2b}$$

$$n_4 = \sum_p c^\dagger_{p+Q,i} \sigma^z_{ij} c_{p,j} \tag{2c}$$

where $\sigma^\alpha$'s are the Pauli spin matrices and $\vec{Q} = (\pi, \pi, \pi)$ is the AF ordering vector. The concept of the superspin is similar to that of the pseudospin (*13, 14*), with a crucial difference that pseudospin is really "pseudo" in the sense that it has no real spin component. There is an *SO*(3) spin symmetry acting on the $(n_2, n_3, n_4)$ subspace, with the total spin $S_\alpha$ being the generator of the rotation, and there is a *U*(1) charge symmetry acting on the $(n_1, n_5)$ subspace, with the total charge $Q$ being the rotation generator:

$$S_\alpha = \sum_p c^\dagger_{p,i} \sigma^\alpha_{ij} c_{p,j}, \qquad Q = \tfrac{1}{2}(N - M) \tag{3}$$

Here $N$ is the number of electrons and $M$ is the number of lattice sites. Of course, the concept of superspin is only useful if we can enlarge the known *SO*(3) × *U*(1) symmetry group to include orthogonal transformations that can rotate the AF order parameters into the SC ones. The minimal group to accomplish this would be an *SO*(5) symmetry group. If such orthogonal transformations exist, and if they commute with the microscopic $H$, then the concept of the

Department of Physics, Stanford University, Stanford, CA 94305, USA.

superspin would be very useful. Because orthogonal transformations preserve the magnitude of the vector, we can impose the constraint:

$$n_a^2 = 1 \tag{4}$$

and really think of the superspin as a vector of a fixed magnitude in a 5D space, where the AF and SC order appear as 3D and 2D projections, respectively. Ten different "Euler angles" are needed to describe a general rotation in a 5D space, and so the $SO(5)$ group has 10 symmetry generators. [In general, the $SO(n)$ group has $n(n-1)/2$ symmetry generators.] The known $SO(3) \times U(1)$ symmetry supplies us with four such generators.

**The $\pi$ operators and the $SO(5)$ algebra.** The remaining six symmetry generators have already been discovered by Demler and Zhang (*15*) in their recent theory of resonant neutron scattering of high-$T_c$ superconductors. These operators carry charge 2, spin 1, and total momentum $(\pi, \pi, \pi)$. These $\pi$ operators are defined as follows:

$$\pi_\alpha^\dagger = \sum_p g(p) c_{p+Q,i}^\dagger (\sigma^\alpha \sigma^y)_{ij} c_{-p,j}^\dagger, \quad \pi_\alpha = (\pi_\alpha^\dagger)^\dagger \tag{5}$$

and we see that there obviously are six of them.

The $SO(5)$ Lie algebra has 10 generators $L_{ab} = -L_{ba}$ satisfying the following commutation relations:

$$[L_{ab}, L_{cd}] = i(\delta_{ac} L_{bd} - \delta_{ad} L_{bc} - \delta_{bc} L_{ad} + \delta_{bd} L_{ac}) \tag{6}$$

Let us identify the 10 operators $(Q, S_\alpha, \pi_\alpha, \pi_\alpha^\dagger)$ with the generators of the $SO(5)$, $L_{ab}$, in the following way, where $L_{ab}$ is:

$$\begin{bmatrix} 0 & & & & \\ \pi_x^\dagger + \pi_x & 0 & & & \\ \pi_y^\dagger + \pi_y & -S_z & 0 & & \\ \pi_z^\dagger + \pi_z & S_y & -S_x & 0 & \\ Q & -i(\pi_x^\dagger - \pi_x) & -i(\pi_y^\dagger - \pi_y) & -i(\pi_z^\dagger - \pi_z) & 0 \end{bmatrix} \tag{7}$$

where the value of the matrix elements on the upper right triangle are determined by antisymmetry, $L_{ab} = -L_{ba}$. The fermionic representations of the $(Q, S_\alpha, \pi_\alpha, \pi_\alpha^\dagger)$ operators in Eqs. 3 and 5 and the above identification can be used to check explicitly that the $SO(5)$ commutation relations (Eq. 6) are indeed satisfied in the continuum approximation:

$$g^2(p) \approx 1 \tag{8}$$

This approximation is valid in the long wavelength limit where the only important information about the internal structure of a composite operator is its symmetry, not its detailed shape. Because $g^2(p)$ has $s$-wave symmetry, it can be approximated by its average value.

With the aid of the $SO(5)$ algebra, the concept of the superspin introduced earlier can be defined precisely. The superspin $n_a$ is defined as a vector representation of the $SO(5)$ Lie algebra that satisfies the following commutation relation:

$$[L_{ab}, n_c] = -i\delta_{bc} n_a + i\delta_{ac} n_b \tag{9}$$

Substitution for $L_{ab}$ and $n_c$ of the microscopic definitions in terms of the electron operators in Eqs. 1, 2, and 7, shows that they satisfy Eq. 9 in the continuum approximation. In particular, the $\pi$ operators rotate the AF order parameter into SC order parameter and vice versa.

With the basic low-energy variables assembled and their kinematic relations specified, dynamic properties can be addressed and the commutation relations between the symmetry generators $L_{ab}$ and the microscopic $t$-$J$ $H$ can be discussed. The total spin operators $L_{23}$, $L_{24}$, and $L_{34}$ and the total charge operator $L_{15}$ certainly commute exactly with the $t$-$J$ $H$. What about the remaining six generators? Demler and Zhang (*15*) have shown that although they do not commute exactly with the $t$-$J$ $H$, they are approximate eigenoperators in the sense that:

$$[\mathcal{H}, \pi_\alpha^\dagger] = \omega_0 \pi_\alpha^\dagger \tag{10}$$

where $\omega_0 = J(1-n)/2 - 2\mu$, and $\mu$ is the chemical potential measured from half-filling. This relation is obtained by the standard $t$ matrix approximation that is similar to the random-phase-approximation for particle-hole operators. The reason that this relation holds is rather simple. The $\pi_\alpha^\dagger$ operator is a particle-particle operator that describes a pair of electrons with a center-of-mass momentum $\vec{Q} = (\pi, \pi, \pi)$. However, for tight-binding models with nearest-neighbor hopping, the two-particle continuum vanishes at $\vec{Q}$. Therefore, a repulsive interaction, such as $J$ in the triplet channel, naturally leads to an antibound state. Thus, $\omega_0$ in Eq. 10 is nothing but the energy eigenvalue of this antibound state. More sophisticated calculations beyond the $t$ matrix approximation have also been carried out recently (*16, 17*). Equation 10 is fundamental in that it provides a bridge from the microscopic $t$-$J$ $H$ to the effective nonlinear $\sigma$ model theory. One can also define a $\pi'$ operator by using $g(p) = \cos p_x + \cos p_y$ in Eq. 5. Such an operator would transform an AF state to $s$-wave SC state. However, it was found (*15*) that these are in general not eigenoperators. Therefore, our theory naturally predicts $d$-pairing symmetry.

There are two reasons why the $SO(5)$ symmetry is only an approximate one. One is the approximation used in deriving Eq. 10. Therefore, this relation should be tested in exact diagonalization studies of the Hubbard or the $t$-$J$ models. This can be done by calculating the dynamical correlation functions of the $\pi$ operators and identifying isolated poles in their spectral function. Fortunately, such a study has been carried out recently by Meixner and Hanke (*18*), and a sharp collective mode in the $\pi$ channel was found. The second reason why the $SO(5)$ symmetry is broken explicitly is that even if Eq. 10 holds exactly, $H$ does not commute with all generators of $SO(5)$. However, this source of explicit symmetry breaking is easy to handle and has interesting physical consequences to be explored below. The eigenoperator relation (Eq. 10) is almost as good as the vanishing of the commutator. The reason is that the Casimir operator $C = \sum_{a<b} L_{ab}^2$ of the $SO(5)$ algebra commutes with $H$, and therefore, we can still classify all of the eigenstates of $H$ according to their $SO(5)$ quantum numbers. The situation is analogous to the problem of a spin precessing in a magnetic field—although the full spin-rotation symmetry is broken by the magnetic field, the magnitude of the spin is preserved.

The present work is a generalization of the concept of the exact $SO(4)$ symmetry (*19*) of the Hubbard model and its application to the negative $U$ Hubbard model (*14, 20, 21*). Using the $\eta$ operators constructed earlier by Yang (*22*), Yang and Zhang (*19*) pointed out that the Hubbard model has, in addition to the usual $SO(3)$ spin rotation symmetry, an $SO(3)$ pseudospin symmetry generated by the following operator:

$$\eta^\dagger = \sum_p c_{p+Q\uparrow}^\dagger c_{-p\downarrow}^\dagger, \qquad \eta = (\eta^\dagger)^\dagger, \qquad \eta_0 = Q \tag{11}$$

Although this symmetry is valid both for the positive and the negative $U$ Hubbard model, the simplest vector multiplet contains $s$-wave SC and charge-density wave (CDW) order parameters (*14, 20*):

$$\Delta_s^\dagger = \sum_p c_{p\uparrow}^\dagger c_{-p\downarrow}^\dagger, \qquad \Delta_0 = \sum_{p\alpha} c_{p+Q,\alpha}^\dagger c_{p,\alpha}, \qquad \Delta_s = (\Delta_s^\dagger)^\dagger \tag{12}$$

and is therefore only useful for the negative $U$ model. The simplest generalization of the pseudospin to the positive $U$ or the $t$-$J$ model is the concept of the superspin introduced in this article. The precise correspondence between these two symmetries are summarized in Table 1.

**Construction of the $SO(5)$ quantum nonlinear $\sigma$ model.** With the basic kinematic and dynamic constraints assembled, a low-energy effective $H$ can be constructed. If the $SO(5)$ symmetry is exact, one can ask what the form of the low-energy $H$ would

be. As argued above, below $T_{MF}$, the system tends to form local singlets, but could form either the AF or SC ground states. Therefore, $T_{MF}$ can be thought of as a mean-field transition below which the superspin acquires a fixed magnitude that leaves its orientation as a low-energy degree of freedom. Near $T_{MF}$, the anisotropy terms in the superspin space are not very important, and this transition can be simply described by a standard Landau free energy functional:

$$F = a|\vec{n}|^2 + b|\vec{n}|^4 \qquad (13)$$

$T_{MF}$ is the temperature at which the coefficient of the quadratic term $a = a'(T - T_{MF})$ changes sign.

Below $T_{MF}$, the magnitude of the superspin is fixed, and it can always be rescaled to satisfy the constraint of Eq. 4. The AF and SC phases correspond to fixed directions of the superspin and thus break the $SO(5)$ symmetry spontaneously. The low-energy dynamics of such a system is determined completely in terms of the Goldstone bosons and their nonlinear interactions as specified by the $SO(5)$ symmetry. The kinetic energy of the system is simply that of a $SO(5)$ rigid rotor, given by $1/2\chi \sum_{a<b} L_{ab}^2(x)$, where $\chi$ is simply the moment of inertia of the rigid rotor. In the long-wavelength limit, a gradient expansion can be performed to obtain a term $\rho/2 \sum_a [\partial_k n_a(x)]^2$ to leading order. Therefore, the resulting $H$ density is given by:

$$H_s = \frac{1}{2\chi} \sum_{a<b} L_{ab}^2(x) + \frac{\rho}{2} \sum_a [\partial_k n_a(x)]^2 \qquad (14)$$

$H_s$ is a low-energy $H$ constrained by the $SO(5)$ symmetry. However, as this symmetry is not exact, some weak symmetry-breaking perturbations should be allowed such that the subgroup $SO(3) \times U(1)$ of the spin and charge symmetry is still exact. The asymmetric $H$ is given by:

$$H_a = \sum_{a<b} \frac{1}{2\chi_{ab}} L_{ab}(x)^2 + \sum_{a<b} \frac{\rho_{ab}}{2} [v_{ab}^k(x)]^2 + V(n) \qquad (15)$$

where $\chi_{15} = \chi_c$, $\chi_{23} = \chi_{24} = \chi_{34} = \chi_s$, and $\chi_{1(2,3,4)} = \chi_{(2,3,4)5} = \chi_\pi$ are the charge, spin, and the newly introduced "$\pi$" susceptibility. A generalized velocity field $v_{ab}^k = n_a \partial_k n_b - n_b \partial_k n_a$ and the corresponding stiffness in the charge ($\rho_{15} = \rho_c$), spin ($\rho_{23} = \rho_{24} = \rho_{34} = \rho_s$) and the $\pi$ ($\rho_{1(2,3,4)} = \rho_{(2,3,4)5} = \rho_\pi$) channel, have also been introduced. In the presence of explicit symmetry breaking, a quadratic symmetry-breaking term of the form:

$$V(n) = -\frac{g}{2}(n_2^2 + n_3^2 + n_4^2) \qquad (16)$$

is also allowed.

$H_s$ and $H_a$ describe the system at half-filling only. In order to go away from half-filling, one simply has to add a $\mu$ term $-2\mu Q = -2\mu L_{15}$ (where $Q$ is a number operator). This yields $\mathcal{H}_s = H_s - 2\mu L_{15}$ and $\mathcal{H}_a = H_a - 2\mu L_{15}$, respectively. Both $\mathcal{H}_a$ and $\mathcal{H}_s$ can be quantized by using the commutation relations (Eqs. 6 and 9), and a complete set of equations of motion can be determined.

Given $\mathcal{H}_a$, one can simply perform a Legendre transformation and obtain the corresponding Lagrangian. After a simple Wick rotation $t \rightarrow i\tau$, the Lagrangian density becomes:

$$\mathcal{L}_a = \sum_{a<b} \frac{\chi_{ab}}{2} \omega_{ab}(x)^2 + \sum_{a<b} \frac{\rho_{ab}}{2} [v_{ab}^k(x)]^2 + V(n) \qquad (17)$$

where:

$$\omega_{ab} = n_a(\partial_\tau n_b - iB_{bc}n_c) - (a \rightarrow b) \qquad (18)$$

is the angular velocity, where a set of generalized external potentials $B_{ab}$ are coupled to $L_{ab}$. In the current problem, the only nonvanishing component is $B_{15} = -B_{51} = 2\mu$. The partition function of the system is given by:

$$\mathcal{Z} = \int [dn_a] \delta(n^2 - 1) e^{-\int_0^\beta d\tau \int d^d x \mathcal{L}_a} \qquad (19)$$

The constants appearing in the model can now be defined. $\chi_s$ and $\chi_c$ are the familiar uniform spin susceptibility and charge compressibility of the system, and $\rho_s$ and $\rho_c$ are the spin stiffness and charge stiffness, respectively ($\rho_s$ is not to be confused with the superfluid density that sometimes also uses this notation). $\chi_\pi$ and $\rho_\pi$ are two new constants introduced in this article. They describe the time and length scale over which an AF region can be converted to a SC region and vice versa. In order for the $SO(5)$ symmetry to be approximately valid, one would require that these constants are close in value. The last remaining constant of our model is the anisotropy constant $g$, which selects either an "easy plane" in the SC space ($n_1$, $n_5$), or an "easy sphere" in the AF space ($n_2$, $n_3$, $n_4$), depending on the sign of $g$. As argued before, $\mu = 0$ defines the model at half-filling, where we know the system is AF. Therefore, we fix $g > 0$ to match this fact. At this point, the values of the constants (especially the sign of $g$) of the model are not deduced from any first-principles calculations. However, once these values are fixed at half-filling, they are not allowed to vary in an arbitrary fashion. As shown below, the richness of phase diagram comes entirely from variation of $\mu$.

**Origin of superconductivity.** At half-filling, $g > 0$ is chosen so that the superspin prefers to lie in an easy sphere of ($n_2$, $n_3$, $n_4$). Away from half-filling, the only new term appearing in $\mathcal{H}$ is just the $\mu$ term. In considering $\mathcal{L}_a$, the $\mu$ term appears as a "gauge coupling" in the imaginary time direction. A constant $\mu$ term is a "pure gauge" and could therefore be "gauged away." A naive expectation would be that such a term has no dynamic consequences. However, there is a cost involved—although a constant $\mu$ term can be gauged away in the bulk, it reappears as a twisted boundary condition in the imaginary time direction and could have nontrivial consequences.

It is more direct to investigate the $\mathcal{L}_a$ (Eq. 17) with the periodic boundary condition in the imaginary time direction. Because of the periodic boundary condition, the classical path extremizing the path integral are the static solutions. The only nonvanishing contribution of the kinetic energy for these static solutions is the $\mu$ term, which gives an effective potential energy:

$$V_{eff} = V(n) - \frac{(2\mu)^2}{2}(n_1^2 + n_5^2) \times [\chi_c(n_1^2 + n_5^2) + \chi_\pi(n_2^2 + n_3^2 + n_4^2)] \qquad (20)$$

Let us first consider the $SO(5)$ symmetric case where $\chi_\pi = \chi_c = \chi$, in which case the terms in the square bracket reduce to $\chi$. If we were dealing with an abelian $XY$ model, such a term would still not have any dy-

**Table 1.** Comparison of three models (see text).

| Parameter | *t-J* Model | Negative *U* Hubbard model | Antiferromagnet in a *B* field |
|---|---|---|---|
| Symmetry | *SO*(5) | *SO*(4) | *SO*(3) |
| Symmetry generators | $S_\alpha$, $Q$, $\pi_\alpha$, $\pi_\alpha^\dagger$ | $S_\alpha$, $Q$, $\eta$, $\eta^\dagger$ | $S_\alpha$ |
| Order parameter | Superspin | Pseudospin | Néel vector |
| Symmetry breaking | $\mu$ | $\mu$ | $B$ |
| Phase transition | Superspin flop (from AF to *d*-wave SC) | Pseudospin flop (from CDW to *s*-wave SC) | Spin flop (from easy axis to easy plane) |
| Collective modes | 1 Massless phase and 3 massive AF modes | 1 Massless phase and 1 massive CDW mode | 1 Massless *XY* and 1 massive *Z* mode |

namical consequences. Because $n_1^2 + n_5^2 = 1$ in this case, the above term would reduce to a number and give only a trivial shift of the ground-state energy and could not lead to any nontrivial phase transitions. However, if the *XY* symmetry is embedded into a higher symmetry group, as in the present case, this term has a profound dynamic consequence. Whereas the $g$ term with $g > 0$ favors an AF easy sphere ($n_2$, $n_3$, $n_4$), the $\mu$ term favors a SC easy plane ($n_1$, $n_5$). This competition leads to a first-order phase transition (for more detailed discussion of the order of the transition, see below) when:

$$\mu = \mu_c = \tfrac{1}{2}\sqrt{g/\chi} \qquad (21)$$

For $\mu < \mu_c$, there is an AF ground state, and the SC state has a finite energy. This energy decreases gradually as one increases $\mu$ until levels cross at $\mu_c$. For $\mu > \mu_c$, the situation is reversed; there is a SC ground state, and the AF state has a finite energy.

For $\chi_\pi \neq \chi_c$, the situation is a bit more complex. In the parameter regime $\chi_c > \chi_\pi$, there is always a direct first-order transition from the AF to the SC state at $\mu = \mu_c = \frac{1}{2}\sqrt{g/\chi_c}$. However, in the parameter regime $\chi_c < \chi_\pi < 2\chi_c$, and for:

$$\frac{g}{\chi_\pi} < (2\mu)^2 < \frac{g}{2}\frac{1}{\chi_c - \chi_\pi/2} \qquad (22)$$

there exists an intermediate spin bag phase (*5*) with coexisting AF and SC order. Schrieffer, Wen, and Zhang (*5*) describe such a phase in terms of pairing the eigenstates of the AF background. When expressed back in terms of the original electron operators, their order parameter is a mixture of the AF and the *d*-wave SC order parameters and the $\pi$ operators.

The above discussion based on the $\mathcal{L}_a$ may appear rather formal, and an intuitive physical example would be useful. Consider an antiferromagnet with easy-axis anisotropy (say the $z$ axis). Below the Néel transition, the Néel vector prefers to point along the easy axis. A uniform magnetic field **B** pointed along the easy axis creates an easy $xy$ plane. At a critical value of the **B** field, there is a "spin-flop" transition where the Néel vector changes its orientation from the $z$ axis to the $xy$ plane. Once the underlying *SO*(5) symmetry is revealed, the physics of the high-$T_c$ superconductivity is as simple as the spin-flop transition. We see that Eq. 18 describes a precession of the superspin about a "fictitious **B** field," namely, $\mu$. The easy axis of the AF state translates into the easy sphere ($n_2$, $n_3$, $n_4$), whereas the easy plane of the AF state translates into the easy SC plane ($n_1$, $n_5$). The transition from an AF ground state to a SC one is a "superspin-flop transition." In fact, the phase diagram of an easy-axis AF in the $T$ and $B$ plane is similar to that of a high-$T_c$ superconductor in the $T$ and $\mu$ plane. The complete analogy between these two systems is summarized in Table 1.

This analogy also helps in understanding the origin of the superspin-flop transition in the $\mathcal{H}$ formulation of the problem. In the $\mathcal{H}$ formulation, the $\mu$ term appears as a coupling to the symmetry-generator $L_{15}$, the number operator $Q$, but not to the superspin $n_a$ directly. It is not immediately obvious why this term would select any particular direction in the superspin space. Of course, the same problem occurs in the AF analog, where the uniform **B** field couples to the total spin vector rather than the Néel vector. If one tries to visualize the Néel $\mathcal{H}$ in a semiclassical fashion, one must remember that the total spin vector is orthogonal to the Néel vector. The externally applied uniform **B** field leads to a finite total spin in the $z$ direction. In order to satisfy the orthogonality condition, the Néel vector therefore has to lie in the plane orthogonal to $z$. In the present model [restricted to the *SO*(5) symmetric case to simplify the algebra], the orthogonality constraint takes the form:

$$\epsilon^{abcde} n_c L_{de} = 0 \qquad (23)$$

which can be simply proved by expressing the angular momenta in terms of the angular velocities $L_{ab} = \chi(n_a\dot{n}_b - n_b\dot{n}_a)$. The $\mu$ term leads to doping, or a finite value of $L_{15}$. The constraint gives $n_2 = n_3 = n_4 = 0$ if other $L_{ab}$ generators have no ground-state expectation value, which is the case since $\chi > 0$.

The question of the origin of the SC state is not a single question, but rather two related questions separated by an energy scale. The high-energy mechanism leads to binding of electrons into singlet pairs. The origin of this binding is rather obvious in the *t-J* model, because the $J$ term favors electrons on near-neighbor sites having opposite spins. The problem is that this same $J$ could also lead to an AF state. In this new model, pair binding gives rise to a finite magnitude of the superspin without fixing its orientation, in marked contrast to the BCS theory, where pair binding is equated with the onset of the SC state, and also different from the large phase-fluctuation model (*11*), where the only fluctuation of the order parameter is its phase. In addition to the high-energy pair-binding mechanism, a low-energy mechanism selects an orientation of the superspin and distinguishes the AF from the SC state. The selection of the different possible ground states is described by *SO*(5) nonlinear $\sigma$ model and the mechanism for favoring the SC state is the superspin-flop mechanism discussed above. Thus, superconductivity is an inevitable consequence of the *SO*(5) symmetry and an AF state at half-filling. From this point of view, an AF state can in some sense be thought of as a solid formed by Cooper pairs, and the superspin-flop transition is a first-order melting transition from the solid into a superfluid of the Cooper pairs. In this framework, the spin fluctuation exchange calculations (*23, 8, 12*) should be interpreted as a mean-field theory of $T_{MF}$. Such calculations lead to the important prediction of *d*-wave superconductivity, but they are not complete because they do not address the actual phase transition and the competition between AF and SC states. The two different mechanisms complement each other in their respective energy regime and together form a complete picture of the origin of superconductivity in the oxides.

**Theory of collective modes and their nonlinear interactions.** The superspin model gives a natural description of the collective modes in the ordered phase. Near the transition between the AF and the SC ground states, the ordered phases are not conventional, but their low-energy excitations reflect the competition between the two different kinds of ordering. These low-energy excitations can be studied systematically by applying symmetry principles and have profound experimental consequences.

Spontaneous breaking of a continuous symmetry naturally leads to Goldstone bosons. The number of the Goldstone bosons is the number of the broken symmetry generators. For the *SO*(5) group, four Goldstone bosons correspond to the four broken symmetry generators.

If the superspin vector lies in the SC plane, for example, along the $n_1$ direction, the mode corresponding to the broken rotation generator $L_{15}$ is just the usual Goldstone mode, which describes the phase degrees of freedom of the SC order parameter. The three other modes correspond to the broken generators $L_{12}$, $L_{13}$, and $L_{14}$. These modes form a triplet representation of the unbroken *SO*(3) spin symmetry. However, if the superspin vector lies in the AF sphere, say along the $n_2$ direction, the usual spin waves correspond to the broken generators $L_{23}$ and $L_{24}$. In addition to the usual spin waves, the current theory predicts two additional Goldstone modes that correspond to the broken generators $L_{21}$ and $L_{25}$. They form a doublet representation of the charge *U*(1) symmetry group.

This symmetry analysis shows what happens if the *SO*(5) symmetry is explicitly broken to $SO(3) \times U(1)$. In the presence of a quadratic symmetry-breaking term $g_{eff} = g - \chi(2\mu)^2$, the spin-triplet Goldstone mode of the SC state would become massive if $g_{eff} < 0$ and the charge-doublet mode of the AF state would become massive if $g_{eff}$

$> 0$. The usual spin-wave modes and the phase mode remain massless because of the $SO(3) \times U(1)$ symmetry. Goldstone bosons that become massive because of explicit symmetry-breaking terms are sometimes called pseudo-Goldstone bosons.

This general analysis can be easily checked by a explicit calculation starting from the $\mathcal{H}_a$. The equations of motion derived from $\mathcal{H}_a$ take the form:

$$\dot{L}_{ab}(x) = \frac{1}{2}\sum_c (\chi_{ac}^{-1} - \chi_{bc}^{-1})\{L_{ac}(x), L_{bc}(x)\} + \sum_c (\rho_{ac} - \rho_{bc})v_{ac}^k(x)v_{bc}^k(x) + \partial_k \sum_c [\rho_{ac}n_b n_c v_{ac}^k(x) - \rho_{bc}n_a n_c v_{bc}^k(x)] + \sum_c [B_{ac}L_{bc}(x) - B_{bc}L_{ac}(x)] + g\sum_{\alpha=2,3,4}(n_a\delta_{b\alpha}n_\alpha - n_b\delta_{a\alpha}n_\alpha) \quad (24)$$

$$L_{ab}(x) = \chi_{ab}\omega_{ab} \quad (25)$$

where $\omega_{ab}$ is given by the Wick rotation of Eq. 18. For $Q(x)$ and $S_\alpha(x)$, only the third term is nonvanishing, expressing the continuity of these exactly conserved charges. This set of equations is a nonabelian generalization of the familiar Josephson equations (*24*) in conventional superconductors. The first equation is a generalization of the Josephson current relation. If we take $L_{ab}(x)$ to be $L_{15}(x)$ in the first equation, the only surviving term is the third one, which reduces to the standard Josephson current expression $\vec{J} = \rho_c\vec{\nabla}\theta$ in the isotropic limit. The second equation is a generalization of the Josephson acceleration equation. If we take $L_{ab}(x)$ to be $L_{15}(x)$ in this equation, the familiar ac Josephson equation $\mu = d\theta/dt$ is recovered. This set of equations for $L_{23}$, $L_{24}$, and $L_{34}$ are generalization of the Landau-Lifshitz equation of antiferromagnets (*25*).

These equations unify AF with SC states and describe the linear spectrum as well as the nonlinear interactions of the collective modes in AF and SC phases. The new physics contained in these equations can be easily analyzed in the linearized approximation. For $g_{\text{eff}} > 0$, there is an AF ground state, and the equations can be linearized around the Néel vector, say $n_2$:

$$\chi_\pi\partial_t^2 n_1 = \rho_\pi\partial_k^2 n_1 - [g - (2\mu)^2\chi_\pi]n_1 - 4\mu\chi_\pi\partial_t n_5$$
$$\chi_s\partial_t^2 n_\alpha = \rho_s\partial_k^2 n_\alpha, \quad \alpha = 3,4$$
$$\chi_\pi\partial_t^2 n_5 = \rho_\pi\partial_k^2 n_5 - [g - (2\mu)^2\chi_\pi]n_5 + 4\mu\chi_\pi\partial_t n_1 \quad (26)$$

The second equation describes the two spin-wave modes of the AF state, while the first and the last equations predict a new, massive doublet-pairing mode. In the long-wavelength limit, the energies of these two modes are given by $\omega_\pm = \sqrt{g/\chi_\pi} \pm 2\mu$. This mode is the precursor of superconductivity in the AF phase and makes it plausible that one can think of an antiferromagnet as a quantum solid of Cooper pairs. A quantum solid has two types of excitations, the gapless phonon modes and a gaped mode corresponding to extracting an atom from its position (*24*). The $\omega_-$ mode described in the above equation roughly corresponds to the second type of excitation of a solid if the Cooper pair is identified as an "atom" of the solid. As $\mu$ is increased, this pairing mode lowers its energy, the quantum solid becomes softer, and eventually the solid melts completely to form a superfluid of Cooper pairs.

It is commonly believed that at half-filling, spin excitations are gapless, whereas all charge excitations have finite energy gaps. The above analysis shows that the charge excitations have two distinct branches corresponding to adding and removing two particles. For $\mu$ close to $\mu_c$, the system is still at half-filling, but the particle-hole symmetry is broken. In this case, the $\omega_-$ mode is nearly gapless but the $\omega_+$ mode has a large gap. The conventional definition of a charge-excitation gap is given by the sum of $\omega_+$ and $\omega_-$ and could remain large. Therefore, near $\mu_c$, the half-filled Hubbard model has both gapless spin and nearly gapless charge excitations, and it is not surprising that an approximate $SO(5)$ symmetry could be valid in this regime.

For $g_{\text{eff}} < 0$, there is an SC ground state, and the equations can be linearized around the $n_1$ vector to obtain:

$$\chi_c\partial_t^2 n_5 = \rho_c\partial_k^2 n_5$$
$$\chi_\pi\partial_t^2 n_\alpha = \rho_\pi\partial_k^2 n_\alpha - [\chi_\pi(2\mu)^2 - g]n_\alpha, \quad \alpha = 2, 3, 4 \quad (27)$$

The first equation describes the usual Goldstone mode (sound mode) of a superconductor, whereas the second equation describes the triplet of massive magnetic modes predicted by Demler and Zhang (*15*). This mode is a precursor of antiferromagnetism in the SC phase. It can be roughly thought of as the roton mode in a superfluid, because both of them reflect the "diagonal short-range order" in a superfluid. As $\mu$ is decreased, this mode lowers its energy until the superfluid eventually "solidifies" to form an AF state.

Both the doublet-pairing mode in the AF phase and the triplet magnetic mode in the SC phase owe their existence to the kinetic energy terms of the $\pi$ operators and are called the $\pi$ doublet and the $\pi$ triplet modes, respectively. Both classes of new collective modes have important experimental consequences. The $\pi$ triplet mode has been used by Demler and Zhang (*15*) to explain the recent resonant neutron-scattering experiments on yttrium-barium-copper oxide (YBCO) superconductors. Below $T_c$, a collective excitation peak appears in the neutron-scattering cross section between 25 and 41 meV (*26–30*) depending on the doping level, and the scatterings are observed in the spin-triplet channel at the commensurate momentum $(\pi, \pi, \pi)$, in agreement with the quantum numbers found in my theory. These modes vanish at $T_c$ (*28, 30*), consistent with the interpretation that they are the pseudo-Goldstone modes associated with the $U(1)$ symmetry breaking. [The resonant neutron-scattering peak vanished above $T_c$ in the $T_c = 92$ K (*28*) and $T_c = 62$ K (*30*) materials. In the $T_c = 52$ K material, the peak intensity drops continuously across $T_c$ (*29*).] The precise correlation between the energy of the neutron peak and $T_c$ will be discussed below.

The above discussion assumes that quantum fluctuations are not strong enough to destroy the long-range order. In the opposite limit, when $\chi^{-1} \gg \rho$, the kinetic terms in $\mathcal{H}_s$ and $\mathcal{H}_a$ should be diagonalized first and the potential energy terms treated as a perturbation. The kinetic terms can be diagonalized easily by using the representation theory of $SO(n)$. They are classified by an integer $l$: the Casimir operator has the value $l(l + n - 2)$ in this representation. The ground state is a $SO(5)$ singlet with $l = 0$, and the lowest energy excitation is a massive vector multiplet with $l = 1$. This multiplet is split by the anisotropy in the kinetic coefficients. The massive pairing doublet has energy $(1/2\chi_c) + (3/2\chi_\pi)$, whereas the massive magnetic triplet has energy $(2/2\chi_s) + (2/2\chi_\pi)$.

**The global phase diagram of high-$T_c$ superconductors.** The phase diagram of the $SO(5)$ nonlinear $\sigma$ model can be determined and compared with experiment. The topology of the phase diagram takes different forms depending on the strength of the quantum fluctuation. The case of $d = 3$ is discussed below unless otherwise stated. The model does not contain disorder and therefore precludes any disorder-induced phases such as the spin-glass phase.

I first discuss the so-called renormalized classical regime (*6*), where quantum fluctuations merely renormalize the coupling constants of the model but are not strong enough to destroy the various types of ordering. In this case, at half-filling, $\mu = 0$, and $g > 0$, the superspin lies in the easy sphere $(n_2, n_3, n_4)$ and there is an AF ground state. There is a phase transition to a paramagnetic phase at $T_N$ as the temperature is raised, and this phase transition is in

the same universality class as the classical $SO(3)$ model in three dimensions. As $\mu$ is increased from 0, the system is still half-filled, but the gap toward one of the $\pi$ doublet mode decreases. Because it is easier for the superspin to fluctuate into other directions, the effective spin stiffness decreases and with it the Néel temperature. At $\mu_c$, the superspin has no preferential easy directions and the system is described by an isotropic $SO(5)$ nonlinear $\sigma$ model. This model has its own critical temperature $T_{bc}$, which is in general smaller than the Néel temperature, because the critical temperature of the $SO(N)$ model scales inversely with $N$. If the $\mu = \mu_c$ line is reached when $T < T_{bc}$, the system is ordered and the superspin defined in this limiting procedure still lies within AF easy sphere. When $\mu$ is increased beyond the critical value, the superspin "flops" into the SC plane. Because the direction of the superspin changes abruptly at the transition and the correlation length in the Josephson sense is finite at the transition point, the superspin flop transition is in general first order for $T < T_{bc}$. Beyond the superspin flop transition, the system crosses from a $SO(5)$ critical behavior over to an $XY$ critical behavior, and the SC transition temperature would increase with increasing $\mu$, because the mass of the $\pi$ triplet mode increases. The topology of the phase diagram in this case is depicted in Fig. 1A.

The line of the first-order superspin flop transition is in general not exactly vertical because the susceptibilities and the stiffnesses in different superspin directions may not be the same. At zero temperature and for $\mu < \mu_c$, $Q = 0$ and the density of holes vanish. For $\mu > \mu_c$, the density of holes is finite. The discontinuous jump of the density is given by $2\chi\mu_c$. Therefore, if the phase diagram is plotted with the hole density rather than $\mu$ as a variable, the phase diagram will contain a two-phase region where $0 < Q < 2\chi\mu_c$. The possibility of phase separation was first pointed out by Emery, Kivelson, and Lin (*4*) in the context of the *t-J* model. However, in their work the nature of the hole-deficient region is not clear, while in the present model I argue that it is in the SC phase. In the two-phase region, the long-range Coulomb interaction can lead to formation of a "stripe order" with alternating AF and SC stripes.

The first-order superspin flop line terminates at a critical point $T_{bc}$, from which two second-order lines, one corresponding to the Néel transition and the other corresponding to the $XY$ transition, emerge. For this reason, this critical point is called a bicritical point. There are two relevant operators near the bicritical point, the reduced temperature $t_{bc} = (T - T_{bc})/T_{bc}$ and $g_{eff}$. The scaling behavior near this point is well studied (*31–33*). One of the most important results is that even if the model defined at the short length scales is not exactly $SO(N)$ invariant, that is, the susceptibility and stiffness in different directions may not be the same, the difference among them scales to zero under the renormalization group flow when the bicritical point is reached. A simple version of this result is shown by Pelcovits and Nelson (*32*), who considered a $SO(N)$ model where the $N$th component has a different stiffness than the others. The difference operator has negative scaling dimension in the $2 + \epsilon$ expansion of the nonlinear $\sigma$ model and is therefore irrelevant near $T_{bc}$. A more general result was derived by Friedan (*34*), who considered a nonlinear $\sigma$ model on an arbitrary Riemanian manifold with metric $g_{ab}$. Friedan showed that the fixed point of the renormalization group flow is an Einstein manifold with $g_{ab} = [2\pi(d-2)]^{-1}R_{ab}$, where $R_{ab}$ is the Ricci tensor. In particular, if one starts with a distorted sphere $S_n$, only the average curvature is a relevant variable, whereas the distortions preserving the average curvature are irrelevant. Therefore, if one has a microscopic model without the full spherical symmetry, a perfect sphere with constant curvature is produced by the large-scale fluctuations near the critical point. This result provides a strong theoretical justification for describing the high-$T_c$ superconductors by a $SO(5)$ nonlinear $\sigma$ model: Even if the symmetry is only approximately valid in the microscopic models such as the Hubbard or the *t-J* model, near the bicritical point, where the most interesting transition from an AF to a SC state occurs, this symmetry becomes exact.

Near the bicritical point, thermodynamic quantities obey exact scaling relations. As mentioned above, there are two relevant operators. Finite $t_{bc}$ gives a temperature correlation length according to the scaling law $\xi_{t_{bc}} \sim t_{bc}^{-\nu}$. The other relevant parameter is the biquadratic symmetry-breaking term $g_{eff}$, which gives a correlation length $\xi_g \sim g_{eff}^{-\nu_g}$. Near the bicritical point, the singular part of the free energy scales according to:

$$f(t_{bc}, g_{eff}) \sim t_{bc}^{2-\alpha}\tilde{f}(Ag_{eff}/t_{bc}^{\phi}) \qquad (28)$$

where $\alpha$ is the free energy exponent of the $SO(5)$ model and $\phi$ is a crossover exponent given by $\phi = \nu/\nu_g$. In Eq. 28, $A$ is a nonuniversal constant and $\tilde{f}$ is a universal scaling function of a single argument. Other singular thermodynamic quantities are determined in a similar way.

The crossover exponent $\phi$ determines the way in which the two second-order lines merge into the first-order line. The universal scaling function $\tilde{f}$ diverges at two arguments, $x_+ > 0$ and $x_- < 0$. The two second-order lines are determined by:

$$g_{eff} = (x_+/A)t_{bc}^{\phi}, \quad g_{eff} = (x_-/A)t_{bc}^{\phi} \qquad (29)$$

Explicit calculations of the exponents can be carried out within both the $\epsilon$ expansion (*32*) and the large $N$ approximation (*35*). To leading order, $\phi = 4/3$ and $\phi = 2$ respectively. The two second-order lines merge into the first-order line tangentially (Fig. 1A).

There exist parameter regimes where AF

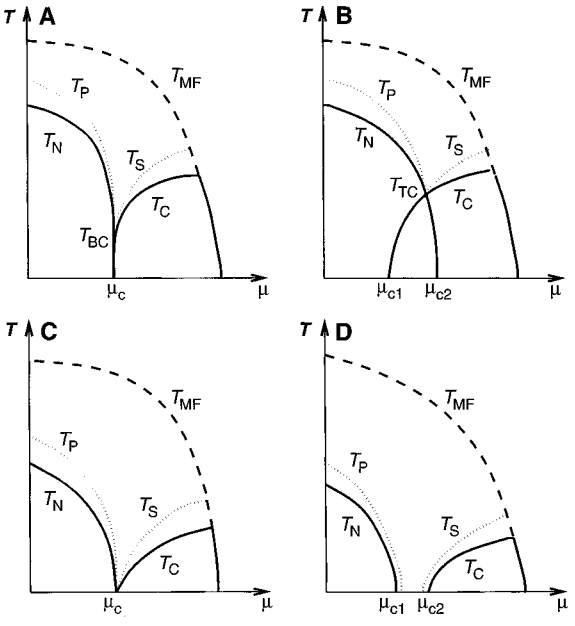


**Fig. 1.** Possible phase diagrams of our theoretical model: $T_{MF}$ is the temperature below which electrons bind to form singlet pairs, in the current model, it corresponds to a finite magnitude of the superspin, and $T_N$ and $T_c$ are the Néel and SC transition temperatures respectively. There are four possible types of transitions from the AF to SC state. (**A**) There is a direct first-order transition that terminates at a bicritical point $T_{bc}$ that can be described as a superspin-flop transition. (**B**) There are two second-order phase transitions with an intermediate spin-bag phase. The four second-order lines merge at a tetracritical point $T_{tc}$. (**C**) There is a single second-order phase transition at a quantum critical point. (**D**) There are two second-order quantum phase transitions with an intermediate quantum-disordered phase. In each type of the phase diagrams, there are two crossover temperatures $T_s$ and $T_p$. The spin-gap temperature $T_s$ corresponds to the temperature below which the superspin lies within the SC plane, and pairing-gap temperature $T_p$ corresponds to the temperature below which the superspin lies within AF sphere.

and SC phases can coexist. In this case, the topology of the phase diagram must be modified to include the spin bag phase (5), which is conceptually similar to the conjectured supersolid phase in $^4H_e$. In this case, all or part of the entire first-order superspin-flop lines separates into two second-order lines enclosing the spin bag phase (Fig. 1B). The four second-order lines intersect at a tetracritical point $T_{tc}$.

All of the phase transitions discussed so far are classical, in the sense that quantum-mechanical fluctuations only renormalize the coupling constants in the problem. When quantum fluctuations are gradually increased, ordering in some phases can be destroyed. Because the *SO*(5) isotropic point has the largest quantum fluctuation, one would expect $T_{bc}$ to be driven to zero first. In this case, $T_{bc}$ is a quantum critical point where the AF state goes into the SC state through a direct second-order phase transition. (Fig. 1C).

When the quantum fluctuations are increased further, an entire region between the AF phase and the SC phase becomes quantum disordered (Fig. 1D). In this region, the properties of the model are better described by first diagonalizing the kinetic energy term of $\mathcal{H}_a$. Both the triplet magnetic excitations and the doublet-pairing excitations have finite energy gaps. The transition from the quantum-disordered phase into the AF phase is second order, the doublet-pairing mode of the quantum-disordered phase naturally continues into the $\pi$ doublet mode in the AF phase, whereas the triplet magnetic excitation of the quantum-disordered phase has evolved into the two gapless Goldstone modes of the AF phase. The transition from the quantum-disordered phase into the SC phase is second order as well, and the triplet magnetic excitation of the quantum-disordered phase naturally continues into the $\pi$ triplet mode of the SC phase, whereas the doublet-pairing mode becomes the gapless Goldstone mode of the superconductor.

The *SO*(5) quantum-disordered phase differs from the *SO*(3) quantum-disordered phase of the quantum antiferromagnet (6) and the quantum *XY* disordered phase (*10*) because it has both a massive triplet magnetic excitation and a massive doublet of pairing excitation. It is tempting to identify it with the RVB phase introduced by Anderson. In fact, phase diagram in Fig. 1D looks similar to the original RVB idea proposed by Anderson (*1*). The basic idea is that with low dimensionality and increased doping, quantum fluctuations will first "melt" the AF state to form an RVB liquid of spin singlet pairs, and these singlet pairs further condense into a superfluid state. However, the *SO*(5) quantum-disordered phase differs from Anderson's RVB phase in terms of the excitation spectrum. In the RVB phase, spinons and holons are the elementary excitations, but in the *SO*(5) quantum-disordered phase, even the electrons are "bound" into the collective coordinates. In some sense, the *SO*(5) quantum-disordered phase describes an "incompressible liquid" of singlet pairs, similar to the Laughlin liquid in the fractional quantum Hall effect.

The global features of the phase diagram deduced from the *SO*(5) quantum nonlinear $\sigma$ model agree reasonably well with the general topology of the experimentally observed phase diagram of high-$T_c$ superconductors. Within this model, the depression of the Néel and the SC temperature have a common origin: A region in the phase diagram where the *SO*(5) fluctuations are maximal. This theory makes precise predictions about the scaling behavior in the crossover region. These predictions can be tested experimentally in sufficiently clean systems where a direct first-order transition becomes possible. I believe that phase diagrams in Fig. 1, A and C, are reasonably close to reality in the high-$T_c$ superconductors. Figure 1B may be useful to understand the heavy fermion materials, and Fig. 1D could be realized in the quasi-1D ladder systems where quantum fluctuations are strong.

**Theory of the spin gap, the SC transition, and their relation.** Underdoped high-$T_c$ materials exhibit so-called pseudogap behavior. Initially, this behavior was observed in the $1/T_1$ nuclear relaxation experiment, where the relaxation rate starts to drop rapidly below a temperature $T^*$ that is higher than $T_c$. Similar effects were then observed in the optical absorption spectrum and, more recently, in the photoemission experiments. A "pseudogap" in both spin and charge sector of the electronic spectrum opens up at $T^*$ (*36*). The name "spin gap," sometimes used for this behavior, is actually a misnomer. Within our current theoretical model, $T^*$ should simply be identified with $T_{MF}$ where the superspin amplitude forms and an electronic pseudogap opens as a result of that. In our model, $T_{MF}$ increases with decreasing doping, consistent with experimental findings.

Another type of nuclear spin relaxation rate, $1/T_{2G}$ (*38*), and neutron-scattering experiments (*26–30*) can measure the spin-spin correlation function directly. Above a certain temperature $T_s$, the magnetic correlation length increases with decreasing $T$, while it saturates below $T_s > T_c$. I shall refer to this temperature $T_s$ as the spin-gap temperature, because it clearly distinguishes the spin and the charge response of the system.

This spin-gap phenomenon has a natural explanation within our current theoretical model. There are three different temperature scales in the current model. When the temperature is lowered below $T_{MF}$, the superspin vector acquires a finite magnitude. However, for a finite range of temperature below $T_{MF}$, say $T_s < T < T_{MF}$, (see Fig. 1), the temperature scale is still high enough that the superspin does not "notice" the anisotropy in its orientational degrees of freedom. In this temperature range, the model is essentially *SO*(5) symmetric, and the AF correlation length increases together with the pairing correlation length as temperature is lowered. For $T < T_s$, the thermal energy becomes low compared to the anisotropy energy in the superspin space, and the superspin vector lies preferably in the SC plane. Therefore, below $T_s$, the AF correlation length saturates to a finite value. Finally, at $T_c$, the superspin vector picks a particular direction within the SC plane, therefore breaking the *U*(1) gauge symmetry, and the system becomes superconducting. A similar picture applies to the AF side of the transition. There is a "pairing gap" temperature $T_p$, above which the pairing correlation length increases together with the AF correlation length, while it saturates below $T_p$.

Both $T_s$ and $T_c$ can be calculated quantitatively within the *SO*(5) nonlinear $\sigma$ model (*37*). Assuming that $T_{bc}$ is much smaller than the maximal $T_c$ at optimal doping, one can approximate the transition region between the AF and SC phase by a quantum critical point. Near the quantum critical point, a finite temperature $T$ introduces a thermal correlation length $\hbar c/kT$ (where $\hbar$ is Planck's constant divided by $2\pi$, $c$ is the speed of light, and $k$ is Boltzmann's constant) and the energy of the $\pi$ triplet mode $\omega_0^2 = (2\mu)^2 - (2\mu_c)^2$ introduces a spatial correlation length $c/\omega_0$. In this region, the critical SC transition temperature is set by the equality of these two length scales and leads to a linear dependence of $kT_c$ and $\omega_0$. This linear dependence agrees with the recent resonant neutron-scattering experiment in the underdoped and overdoped regime of YBCO superconductors (*29*).

From the fit of the above mentioned proportionality, the spin-gap temperature $T_s$ can be determined, and $T_s/T_c \approx 1.37$. There are some experimental uncertainties about the value of $T_s$: The most accurate measurement has been carried out in the $T_c = 62$ K material $YBa_2Cu_3O_{6.63}$. The $1/T_{2G}$ measurement shows that the AF correlation length saturates at $T_s \approx 100$ K (*38*). This gives a $T_s/T_c$ ratio of 1.61, about 15% greater than the theoretically predicted ratio.

**Attempt at a synthesis.** The most im-

portant message from this work is that AF and SC are complementary. A *SO*(5) symmetry determines the competition between these phases as well as the low-energy dynamics of the high-$T_c$ superconductors.

The mechanism of superconductivity is divided into a high-energy piece, identified with the pair binding or the formation of a superspin amplitude, and a low-energy piece that involves a superspin-flop mechanism for selecting the orientation of the superspin and resolving the competition between the SC and AF states. These two different energy scales have different doping dependences. With increasing doping, $T_{MF}$ decreases because the effective *J* decreases, but $T_c$ increases because both the thermal and the quantum *SO*(5) fluctuations decrease. Recent photoemission experiments (*40*) found that the electronic gap decreases with increasing $T_c$, which is inconsistent with the weak-coupling BCS theory but consistent with the current model. This work also answers the question of why $T_c$ is low compared with the energy scales of the pair formation. Emery and Kivelson (*11*) rightfully argued that this happens because the superfluid density is low, but did not offer an explanation of why this is so. The current work explains this in terms of increased quantum and classical *SO*(5) fluctuations near the isotropic point $\mu_c$, which suppress both the Néel temperature and the SC transition temperature. The spin-gap phenomenon is naturally interpreted in terms of the competition between entropy and anisotropy energy in superspin space. The most direct experimental evidence in support of the approximate *SO*(5) symmetry is the resonant neutron-scattering peaks in the underdoped and optimally doped region. These modes are naturally interpreted in terms of the pseudo-Goldstone bosons associated with the spontaneous *SO*(5) symmetry breaking. Notably absent in this discussion is the transport properties of the high-$T_c$ oxides. Below $T_{MF}$, transport properties could be addressed by considering fermions coupled to the *SO*(5) nonlinear $\sigma$ model degrees of freedom. However, quantitative details have not yet been fully worked out.

The current theory unifies a number of seeming divergent theoretical approaches to the high-$T_c$ problem. It outlines a strategy to systematically extract the low-energy content of the *t-J* model by constructing low-energy field theory constrained by the microscopic symmetry. The *SO*(5) symmetry can be used as a basic principle to organize the various theoretical proposals. Most directly, it unifies the *SO*(3) nonlinear $\sigma$ model theory (*6*, *8*, *9*) with the *U*(1) nonlinear $\sigma$ model theory (*10*, *11*). The "spin fluctuation exchange" (*23*, *8*, *12*) approach should be interpreted as a theory of $T_{MF}$. Phase separation (*4*) occurs in the *t-J* model because the superspin-flop transition is first order, therefore giving rise to a "forbidden density region." The spin-bag phase (*5*) can emerge in the phase diagram as a result of increased $\pi$ fluctuation, or more precisely when $\chi_\pi > \chi_c$. The *SO*(5) quantum-disordered phase could in some sense be associated with the RVB phase in Anderson's (*1*) original proposal, and occurs as a result of increased *SO*(5) quantum fluctuation.

The *SO*(5) theory makes a number of new experimental predictions. The most direct prediction is that when the materials are sufficiently clean, there could be a direct first-order transition between the AF and the SC states. Measurements near the bicritical point can be compared quantitatively with the theoretical predictions and can serve as an important test of the theory. There are other predictions for which I shall only outline the basic ideas. In the conventional Landau-Ginzburg theory of superconductivity, the order parameter is constrained to lie in a plane. Near the core of a SC vortex, a mathematical singularity is unavoidable. However, in the superspin model, because of the 5D-order parameter, the superspin could lie in the SC plane far away from a vortex, but flip into the AF sphere inside a vortex. Such a topological configuration is called a "meron" in field theory, meaning half of a Skyrmion. Therefore, the current theory predicts that the core of a vortex in underdoped superconductors is not filled with normal electrons but is antiferromagnetic instead. Such an effect could be observed directly by studying the elastic satellite peaks in neutron-scattering experiments of the vortex lattice. The current theory can also be used to predict the static and dynamic properties of the interface between the AF and the SC phases.

## REFERENCES AND NOTES


1. P. W. Anderson, *Science* **235**, 1196 (1987).
2. F. C. Zhang and T. M. Rice, *Phys. Rev. B* **37**, 3759 (1988).
3. P. A. Lee and N. Read, *Phys. Rev. Lett.* **58**, 2891 (1987).
4. V. J. Emery, S. Kivelson, H. Q. Lin, *ibid.* **64**, 475 (1990).
5. J. R. Schrieffer, X. G. Wen, S. C. Zhang, *Phys. Rev. B* **39**, 11663 (1989).
6. S. Chakravarty, B. I. Halperin, D. R. Nelson, *ibid*., p. 2344.
7. V. Barzykin and D. Pines, *ibid.* **52**, 13585 (1995).
8. D. Pines, *Physica C* **235**, 113 (1994).
9. A. V. Chubukov, S. Sachdev, J. Ye, *Phys. Rev. B* **49**, 11919 (1994).
10. S. Doniach and M. Inui, *Phys. Rev. B* **41**, 6668 (1990); *J. Phys. Chem. Solids* **57**, 506 (1996).
11. V. J. Emery and S. Kivelson, *Nature* **374**, 434 (1995).
12. D. Scalapino, *Phys. Rep.* **250**, 329 (1995).
13. P. W. Anderson, *Phys. Rev.* **112**, 1900 (1958).
14. S. C. Zhang, *Phys. Rev. Lett.* **65**, 120 (1990).
15. E. Demler and S. C. Zhang, *ibid.* **75**, 4126 (1995).
16. H. Kohno, B. Normad, H. Fukuyama, *Proceedings of the 10th Anniversary HTS Workshop* (World Scientific, Singapore, 1996).
17. E. Demler and S. C. Zhang, unpublished results.
18. S. Meixner, W. Hanke, E. Demler, S. C. Zhang, preprint, cond nat/9701217.
19. C. N. Yang and S. C. Zhang, *Mod. Phys. Lett. B* **4**, 759 (1990).
20. S. C. Zhang, *Int. J. Mod. Phys. B* **5**, 153 (1991).
21. E. Demler, S. C. Zhang, N. Bulut, D. Scalapino, *ibid.* **10**, 2137 (1996).
22. C. N. Yang, *Phys. Rev. Lett.* **63**, 2144 (1989).
23. N. E. Bickers, D. J. Scalapino, S. R. White, *ibid.* **62**, 961 (1989).
24. P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin/Cummings, Menlo Park, CA, 1983).
25. B. I. Halperin and P. C. Hohenberg, *Phys. Rev.* **188**, 898 (1969).
26. J. Rossat-Mignod *et al*., *Physica C* **185**, 86 (1991).
27. H. Mook *et al*., *Phys. Rev. Lett.* **70**, 3490 (1994).
28. H. F. Fong *et al*., *ibid.* **75**, 316 (1995).
29. H. F. Fong *et al*., *ibid.* **78**, 713 (1997).
30. P. Dai *et al*., *ibid.* **77**, 5425 (1996).
31. M. E. Fisher and D. R. Nelson, *ibid.* **32**, 1350 (1974).
32. R. A. Pelcovits and D. R. Nelson, *Phys. Lett. A* **57**, 23 (1976).
33. D. R. Nelson and R. A. Pelcovits, *Phys. Rev. B* **16**, 2191 (1977); D. S. Fisher, *ibid.* **39**, 11783 (1989).
34. D. H. Friedan, *Ann. Phys.* **163**, 318 (1985).
35. S. Hikami and R. Abe, *Prog. Theor. Phys.* **52**, 369 (1974).
36. B. Batlogg and V. J. Emery, *Nature* **382**, 20 (1996).
37. Detailed calculations appear as supplementary material available to *Science Online* subscribers at http://www.sciencemag.org/science/feature/data/zhang.
38. M. Takigawa, *Phys. Rev. B* **49**, 4158 (1994).
39. A. G. Loeser *et al*., *Science* **273**, 325 (1996).
40. J. M. Harris *et al*., preprint.


41. I thank R. B. Laughlin for a stimulating discussion that helped me form the superspin concept, and T. M. Rice for a stimulating discussion that led to the identification of the triplet magnetic excitation of the quantum-disordered phase with the $\pi$ triplet mode of the SC phase. I thank J. R. Schrieffer, S. Kivelson, R. B. Laughlin, and D. Scalapino for patiently teaching me about various aspects of the high-$T_c$ problem throughout the years and for sharing with me their deep insights and enthusiasm and C. P. Burgess and C. Lütken for tutoring me about the nonlinear $\sigma$ model theory of QCD. Inspiring discussions with N. Bulut, E. Demler, H. Fukuyama, M. Greiter, W. Hanke, B. Keimer, H. Kohno, S. Meixner, N. Nagaosa, and Z. X. Shen are also gratefully acknowledged. Last but not least, I thank C. N. Yang for collaborating on an earlier work which inspired the current one, and for impressing me with the power of symmetry. Supported by the NSF under grants DMR-9400372 and DMR-9522915. Part of this work was carried out while visiting the IBM Almaden Research Center and I thank B. A. Jones for the hospitality extended to me during my visit.

15 October 1996; accepted 13 December 1996