Frontiers of Physics
https://doi.org/10.1007/s11467-020-0957-5
Front. Phys.
15(4), 43301 (2020)

REVIEW ARTICLE

# Fermion dynamical symmetry and strongly-correlated electrons: A comprehensive model of high-temperature superconductivity

**Mike Guidry**[1,†], **Yang Sun**[2,‡], **Lian-Ao Wu**[3,♯], **Cheng-Li Wu**[4,§]

[1] *Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996, USA*
[2] *School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*
[3] *IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain, and Department of Theoretical Physics and History of Science, Basque Country University (EHU/UPV), Post Office Box 644, 48080 Bilbao, Spain*
[4] *Department of Physics, Chung-Yuan Christian University, Chungli, Taiwan 320, China*
*E-mail:* †*guidry@utk.edu,* ‡*sunyang@sjtu.edu.cn,* ♯*lianaowu@gmail.com,* §*clwu@phys.cts.nthu.edu.tw*
*Received December 22, 2019; accepted March 9, 2020*

We review application of the SU(4) model of strongly-correlated electrons to cuprate and iron-based superconductors. A minimal self-consistent generalization of BCS theory to incorporate antiferromagnetism on an equal footing with pairing and strong Coulomb repulsion is found to account systematically for the major features of high-temperature superconductivity, with microscopic details of the parent compounds entering only parametrically. This provides a systematic procedure to separate essential from peripheral, suggesting that many features exhibited by the high-$T_c$ data set are of interest in their own right but are not central to the superconducting mechanism. More generally, we propose that the surprisingly broad range of conventional and unconventional superconducting and superfluid behavior observed across many fields of physics results from the systematic appearance of similar algebraic structures for the emergent effective Hamiltonians, even though the microscopic Hamiltonians of the corresponding parent states may differ radically from each other.

**Keywords** strongly-correlated electrons, SU(4) model, fermian dynamical symmetry, high-temperature superconductivity

Contents

1 Introduction 3
1.1 The adequacy of theoretical tools 3
1.2 Areas of some consensus 4
1.3 Fundamental issues with little consensus 5
1.3.1 Parent states and rapid onset of superconductivity 5
1.3.2 Nature of the pseudogap state 5
1.3.3 Spatial inhomogeneity but a universal phase diagram 6
1.4 Addressing these issues within a unified framework 7
2 Truncation of large Hilbert spaces 7
2.1 Truncation based on microscopic properties of the weakly-interacting system 7
2.2 Emergent-symmetry truncation 7
2.3 Spontaneously-broken symmetries 8
2.4 Examples of emergent symmetries 8
3 The dynamical symmetry method 8
3.1 Solution algorithm 8
3.2 Validity and utility of the approach 9
4 Strongly-correlated SU(4) electrons 9
4.1 Structure of the coherent pair basis 9
4.2 The collective operators 10
4.3 The SU(4) algebra and subalgebras 11
4.4 Collective subspace and associated Hamiltonian 12
5 The dynamical symmetry limits 13
5.1 The SO(4) dynamical-symmetry limit 14
5.2 The SU(2) dynamical-symmetry limit 14
5.3 The SO(5) dynamical-symmetry limit 15
6 Generalized SU(4) coherent states 15
6.1 Associating coherent states with Lie algebras 15
6.2 SU(4) coherent states 15
6.3 Generalized quasiparticle transformation 16
6.4 Temperature dependence 16
6.5 Energy gaps and gap equations 17
6.6 Relationship to ordinary BCS and Néel theory 18
6.7 Solution of the gap equations at zero temperature 18
6.7.1 The critical doping point 18
6.7.2 The all gaps finite solution for $T = 0$ 18

*arXiv: 2003.07994.

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

6.7.3 The pure singlet-pairing solution for $T = 0$ 18
6.7.4 The pure triplet-pairing solution for $T = 0$ 18
6.7.5 The pure AF solution for $T = 0$ 19
6.7.6 The uncorrelated solution for $T = 0$ 19
6.7.7 The zero-temperature ground state 19
6.8 Solution of the gap equations for finite temperature 19
6.8.1 The $\Delta_d + \Delta_q + \Delta_\pi$ finite-temperature case 19
6.8.2 The $\Delta_d$ finite-temperature case 20
6.8.3 The $\Delta_q$ finite-temperature case 20
6.9 Momentum-dependent SU(4) solutions 20
6.9.1 Momentum-dependent generators of the algebra 20
6.9.2 Zero-temperature, momentum-dependent gap equations 21
7 Global implications of SU(4) symmetry 21
7.1 Physical conditions for closure of the SU(4) algebra 21
7.1.1 Closure of the SU(4) algebra in momentum space 21
7.1.2 Closure of SU(4) on the real-space lattice 22
7.1.3 SU(4) symmetry and double occupancy of sites 22
7.1.4 Pair formfactors and closure of the algebra 23
7.2 Reduction from SO(8) to SU(4) symmetry 23
7.3 SU(4) symmetry and an upper doping limit for the superconducting state 25
7.4 The antiferromagnetic–superconducting transition 25
7.4.1 Competing antiferromagnetism and superconductivity 25
7.4.2 Analogies in nuclear structure physics 25
7.4.3 Analogies with graphene quantum-Hall physics 26
8 Ground-state energy surfaces 26
8.1 Energy surfaces in the SO(4) limit 27
8.2 Energy surfaces in the SU(2) limit 28
8.3 Energy surfaces in the SO(5) limit 28
8.4 Critical dynamical symmetries 28
8.5 Weakly-broken SO(5) symmetry 28
9 SU(4) energy gaps 29
9.1 Energy-ordering of gaps 29
9.2 Generic features of SU(4) gaps 29
9.3 The critical doping point 30
9.4 Comparison with gap data 30
9.5 Competing order and preformed pairs 31
9.6 The role of triplet pairs 31
10 SU(4) phase diagrams 31
10.1 The predicted phases 32
10.2 Comparison with data 32
11 Fundamental instabilities 33
11.1 Pairing instability with doping 34
11.1.1 Implications for cuprates at low hole doping 35
11.1.2 The generalized Cooper instability 35
11.1.3 Implications for resonating valence bond models 36
11.2 Critical dynamical symmetry and inhomogeneity 36
11.2.1 Dynamical criticality and sensitivity to perturbations 36
11.2.2 Spatial inhomogeneity induced by background perturbations 37
11.2.3 The role of charge 38
11.2.4 Self-organization versus dopant impurities 38
11.2.5 Amplification of proximity effects 38
11.2.6 Dynamical criticality, emergence, and complexity 38
11.2.7 Varied inhomogeneity but a universal phase diagram 39
11.2.8 Critical dynamical symmetries near magnetic vortices and magnetic impurities 39
11.2.9 Summary 39
12 The pseudogap and mean fields 39
13 Anisotropy of the pseudogap 40
13.1 Fermi arcs and magnetic quantum oscillations 40
13.2 Momentum-dependent SU(4) and the pseudogap 41
13.2.1 Implications for Fermi arcs 42
13.2.2 Temperature dependence of Fermi arcs 42
13.2.3 Discriminating among theories for Fermi arcs 43
13.2.4 Implications for Fermi surface pockets 43
13.3 Summary: Anisotropy, arcs, and pockets 43
14 The iron-based superconductors 43
14.1 Non-Abelian superconductors 44
14.2 Extending SU(4) to iron-based superconductors 44
14.2.1 Cuprate and Fe-based phenomenology 44
14.2.2 Mott insulator versus poor-metal parents 45
14.2.3 An SU(4) model for iron-based superconductivity 45
14.2.4 Multiple pairing gaps in the iron superconductors 47
14.3 Unified cuprate and Fe-based superconductivity 48
15 Relationship with other models 50
15.1 SU(4) and BCS models 50
15.2 SU(4) and Néel antiferromagnetism 50

15.3 SU(4) and Mott insulators 51
15.4 SU(4) and resonating valence bond states 51
15.5 SU(4) and the Zhang SO(5) model 51
15.6 SU(4) and the Hubbard and $t$–$J$ models 52
16 High critical temperatures 53
16.1 Unification of competing order 54
16.2 The generalized Cooper instability and high-$T_c$ 54
16.3 An information argument 54
16.4 The role of microscopic physics 55
17 Universality of superconducting and superfluid behavior 55
18 What is special about SU(4) symmetry? 56
18.1 The physical meaning of SU(4) symmetry 56
18.2 Intuitively correct limits 56
18.3 What SU(4) is not 57
18.4 Simple descriptions and complex phenomena 57
19 Summary and conclusions 57
Appendix A SU(4) subgroups and dynamical symmetries 60
Acknowledgements 60
References and notes 60

## 1 Introduction

High-temperature superconductivity (HTSC) was discovered in 1986 for the copper oxides [1] but its theoretical interpretation remains contentious [2]. In 2008, new high-temperature superconductivity was discovered in FeAs compounds [3, 4], and in 2010 in FeSe compounds [5]. This new iron-based superconductivity also is not well understood, with many open questions about the underlying mechanism and whether it has any relation to the mechanism for cuprate superconductivity [6–9]. Conventional superconductors are described well by the BCS (Bardeen–Cooper–Schrieffer) theory [10], and correspond to spin-singlet condensates of Cooper pairs [11] having phonon pair-binding and orbitally-symmetric ($s$-wave) pairing formfactors. It is generally thought that the cuprate and iron-based superconductors result from condensation of spin-singlet Cooper pairs, but that they have non-phonon pair binding with formfactors that differ from the conventional symmetric $s$-wave form; superconductors having such unconventional pairing are commonly called *unconventional superconductors.*

More generally, superconductivity (SC) exhibiting many similarities with that of cuprate and iron superconductors has been found in other condensed-matter systems, such as organics and heavy-fermion compounds [12, 13]. Superconductivity (or superfluidity) also is known to play a central role in nuclear structure [14], and is expected to occur for the neutrons and protons in neutron stars [15] and for the color degree of freedom in quark matter [16, 17]. These other forms of superconductivity or superfluidity are all thought to involve condensates of Cooper pairs, even though the underlying structure and interactions may differ fundamentally among these instances. The mechanism is often suspected to involve unconventional pairing, but is typically not well understood. There is even less understanding of how such a broad set of physical systems, ranging from diverse compounds in condensed matter to many different isotopes of atomic nuclei to a variety of neutron stars, should exhibit a similar Cooper instability leading to the superconducting or superfluid state.

### 1.1 The adequacy of theoretical tools

The lack of agreement concerning the mechanism for high-temperature superconductivity, and the limited understanding of how these mechanisms are connected to the various other occurrences of superconductivity noted above, have two ready explanations:

1) In each case the issues are complex but most models emphasize only limited aspects of the overall problem.

2) Often the superconducting mechanism is obscured by the complex behavior because the dearth of solvable models with broad physics content makes it difficult to separate essential features from secondary ones.

Thus, we believe that the primary issue is not that there are key measurements remaining to be made that will magically unravel the HTSC problem, but rather that most theoretical tools are inadequate to assess properly the implications of the quite sophisticated data sets already in existence. Stated concisely, theoretical models that deploy a sufficiently broad arsenal of physics tend to not be solvable, while those models that are solvable tend to be so at the cost of emphasizing certain aspects of the problem relative to others. This latter feature of solvable models then tends to produce strong either/or dichotomies around which various camps rally.

We propose here that a solvable model with an adequate range of physics for the HTSC problem is possible by applying a set of mathematical tools originating in

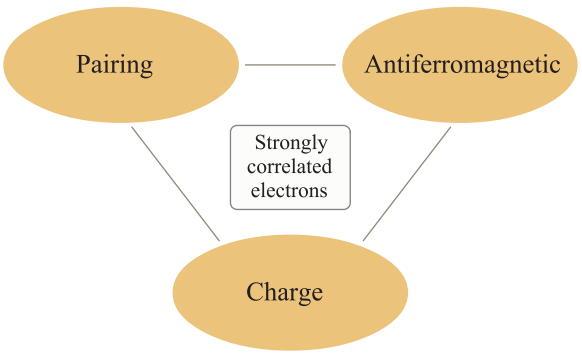



**Fig. 1** Cuprate superconductivity is thought to involve an interplay of pairing, antiferromagnetism, and charge in a strongly-correlated electron system.

the theory of Lie algebras and generalized coherent states. The methods that we shall use to accomplish this might be viewed as unconventional in that they have not found broad use in condensed matter physics prior to the work reviewed here. However, their validity is well established in strongly-correlated fermion and boson systems for fields such as nuclear [18, 19], elementary particle [20], molecular [21], and polymer physics [22], and—truth be told—they draw substantially on ideas that originated in condensed matter and related fields. The dynamical symmetry methods discussed here may be viewed as sophisticated generalizations of pseudospin models, which were introduced in application to the BCS model by Anderson [23], and later used extensively in both condensed matter and nuclear physics. Likewise, the method of generalized coherent states is a sophisticated extension of Glauber coherent states [24], which have been employed often in condensed matter contexts.

By making use of these powerful methods, we shall arrive at a physical picture that is surprisingly conventional relative to many proposed explanations of high-temperature superconductivity. We shall show that a theory unifying self-consistently superconductivity built on a BCS-like wavefunction (possibly with unconventional order) and the Néel model of antiferromagnetism in the presence of strong Coulomb repulsion leads, with few further assumptions, to physics very similar to that observed in actual high-temperature superconductors.

### 1.2 Areas of some consensus

It is useful to begin the discussion with some things that are not so contentious. Despite the absence of general agreement on the mechanism for high-temperature superconductivity in the cuprates and iron-based compounds, we believe that there *is* fairly broad consensus on four issues.

1) High-temperature superconductivity is related intimately to an interplay among pairing, antiferromagnetism, and charge degrees of freedom for strongly-correlated electrons, as illustrated schematically in Fig. 1.

2) The superconducting state is a condensate of spin-singlet Cooper pairs behaving in many respects as an ordinary BCS superconductor, but with important differences, particularly in states with low doping. The orbital pairing formfactor for the cuprate superconductors is of $d_{x^2-y^2}$ form and dominated by contributions from a single band. Figure 2 illustrates. For the iron-based superconductors there is less certainty but it is generally believed that multiple bands near the Fermi surface contribute and that there may be more than one pairing gap. The orbital symmetry of the gap often appears to be a modified *s*-wave form, but it is possible that the symmetry of the gap may vary with the compound being examined [25].

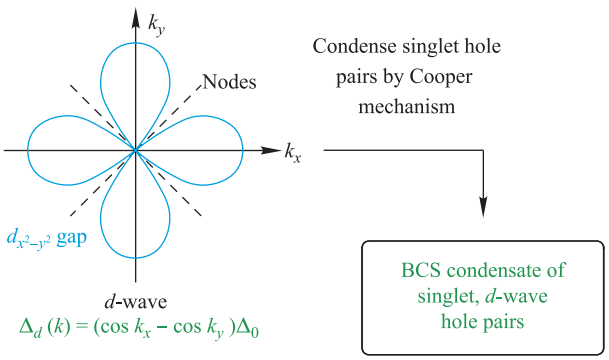



**Fig. 2** Cuprate superconductivity is thought to involve singlet pairing with a *d*-wave orbital formfactor.

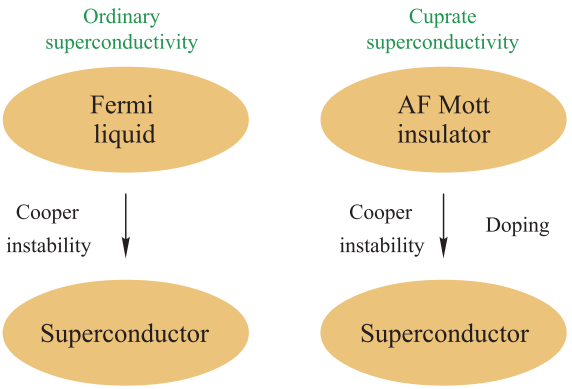



**Fig. 3** Normal superconductivity results from the Cooper instability for a Fermi liquid. Cuprate superconductivity also results from Cooper pairing, but it develops from a Mott insulator state through doping.

3) The precursor to the cuprate superconducting state is an antiferromagnetic (AF) Mott insulator (an insulating state in which the insulator properties derive specifically from strong onsite Coulomb repulsion), unlike for ordinary superconductors where the precursor is a normal Fermi liquid (an interacting system of fermions having excitations that can be put into one-to-one correspondence with those of a non-interacting fermion system). Figure 3 illustrates. The precursor to the iron-based superconductors is typically a poor AF metal that might (or might not) be near a Mott transition.

4) The pseudogap, corresponding to an observed partial gapping of states in the underdoped region above

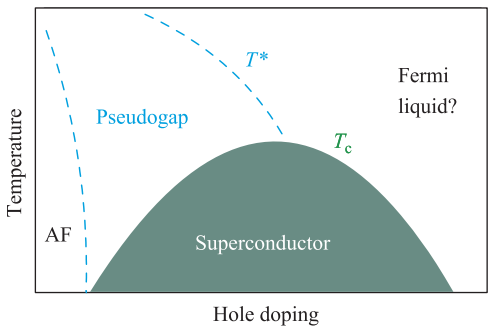



**Fig. 4** The pseudogap region in the cuprates.

the superconducting transition temperature [26, 27], is a well-established enigma for the cuprate superconductors (Fig. 4). Whether a pseudogap exists for the iron-based compounds is not settled (positive evidence for one is given in Refs. [28–32]). The pseudogap state is not a "normal" normal state, and it is not well described by traditional Fermi liquid concepts. There is a general feeling that understanding the pseudogap state is necessary to understanding HTSC, at least for the cuprates.

### 1.3 Fundamental issues with little consensus

In contrast to the areas enumerated in the previous section where there is relatively uniform agreement, there are a number of important issues for which there is little agreement. These share two common features, in our opinion: (i) Their resolution may be central to understanding the high-temperature superconducting mechanism. (ii) A resolution of these issues requires a multiphysics approach capable of integrating concepts on a similar footing that could be treated as approximately independent in many simpler problems.

#### *1.3.1 Parent states and rapid onset of superconductivity*

Band theory suggests that cuprates at half lattice filling should be metals, but they are in fact insulators with antiferromagnetic (AF) properties. This behavior is thought to result from a Mott-insulator normal state, where the insulator properties follow from strong onsite Coulomb repulsion. Doping the normal states with electron holes produces a rapid transition to a superconducting (SC) state, with a pairing gap typically appearing for about 3%–5% hole density per copper site in the copper–oxygen plane. Furthermore, at low to intermediate doping a partial energy gap appears at temperatures above the SC transition temperature $T_c$ that is termed a pseudogap (PG), with the SC gap and PG having opposite doping dependence at low doping [26, 27].

Parent states of normal superconductors are Fermi liquids and normal superconductors are described by BCS theory [10], which assumes the condensation of zero-spin, zero-momentum fermion pairs into a new collective state with long-range coherence of the wavefunction. The key to understanding normal superconductivity was the demonstration by Cooper [11] that normal Fermi liquids possess a fundamental instability: a zero-momentum electron pair above a filled Fermi sea can form a bound state for *vanishingly small attractive interaction*. In normal superconductors the attraction is provided by interactions with lattice phonons, which bind weakly over a limited frequency range because electrons and the lattice have different response times. However, it is the Cooper instability, not the source of the attractive interaction, that is most fundamental: a weak attraction alone cannot produce a superconducting state, but the Cooper instability can (in principle) produce a superconducting state for *any* weakly-attractive interaction between a spin-singlet, zero momentum pair.

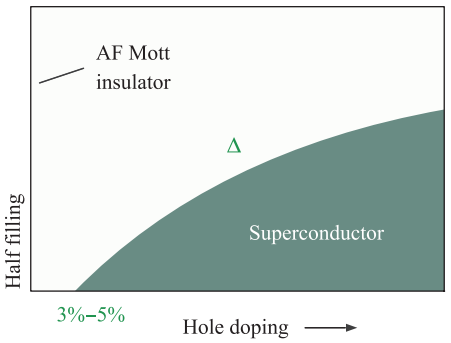



**Fig. 5** The rapid onset of cuprate superconductivity with doping.

The rapid onset of HTSC with hole doping in the cuprates (Fig. 5) suggests that the Mott insulator already contains within it a hidden propensity to superconductivity. Such behavior is indicative of a fundamental instability with respect to pair condensation, but this (as well as the origin of pseudogap states) is difficult to interpret within the standard BCS framework since the superconductor appears to derive from a Mott insulator, not a normal Fermi liquid.

As for normal superconductivity, we believe that the key to understanding HTSC is not the effective interaction leading to pair binding (a topic important in its own right), but rather the nature of the instability that produces the superconducting state. At larger doping the cuprate high-temperature superconducting state exhibits many properties of a normal ($d$-wave) BCS superconductor, so this instability must reduce to the classical Cooper instability at larger doping, but morph into something having more complex behavior at lower doping where the normal state approaches a Mott insulator and there is a pseudogap lying above the superconducting transition temperature.

#### *1.3.2 Nature of the pseudogap state*

As we have noted, there is broad agreement that an explanation of the pseudogap (PG) state may be central to understanding the superconductivity. However, there is extensive disagreement over what the appropriate explanation is. Two general ideas (illustrated in Fig. 6) have dominated conceptual understanding of the pseudogap.

1) The *preformed pairs picture* assumes that the pseudogap is associated with formation of correlated pairs at a pseudogap temperature $T^*$, but that these condense into a state having long-range pairing order—a superconductor—only at a lower temperature $T_c$ (for example, see Ref. [39]).

2) The *competing order picture* assumes that pseudogap properties result because another form of order com-

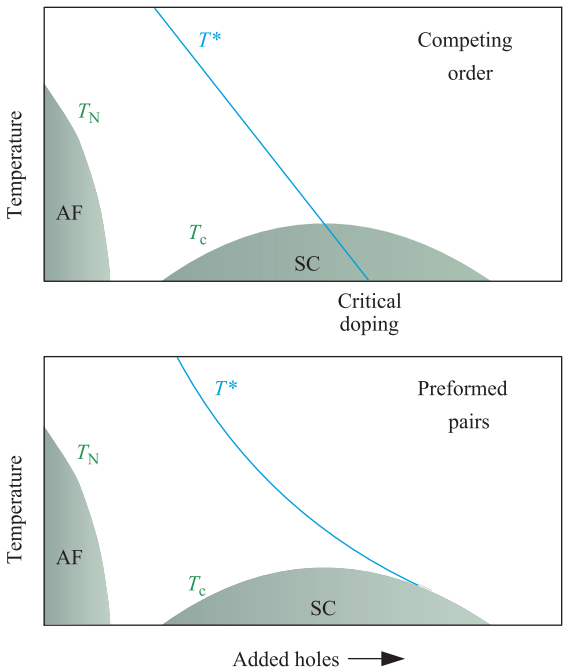



**Fig. 6** Two views of the nature of the pseudogap. In the *preformed pairs* picture, pairs form on a higher-energy scale than the scale on which they condense into long-range order. The preformed pairs *aid* the formation of the superconducting condensate since they are precursors to it. In the *competing order* picture, some other order *competes* with the superconductivity. It must be *suppressed* before superconductivity can set in fully.

petes with superconductivity for $T_c < T < T^*$, and the competing order must be suppressed before robust superconductivity can appear (for example, see Ref. [40]).

The order competing with superconductivity often is assumed to involve antiferromagnetism or charge degrees of freedom, and competing-order approaches often conjecture a quantum phase transition between states dominated by the two forms of order, with quantum-critical scaling that is assumed to account for many HTSC features, but in a way that is not often clearly elucidated.

Preformed pairs and competing order are viewed commonly as incompatible alternatives, and many papers claiming evidence for one or the other view may be found in the literature. These seemingly contradictory results raise the issue of whether the preformed pair and competing order pictures need be mutually exclusive. We shall argue that in a solvable theory of adequate complexity, explanation of the cuprate pseudogap is no longer an either/or choice between preformed pairs or competing order, but rather is deeply and essentially a consequence of both.

The question of pseudogap origin is related closely to another. Competing order, at least at a mean-field level, suggests an association of the pseudogap with a phase, but it has proven difficult to find any order parameters that characterize the PG states systematically across all compounds. Thus a common opinion is that the transition to the pseudogap state may be a crossover and not a true phase transition. Yet there is significant evidence for competing order in the cuprates, and mean-field pictures can account for many cuprate properties by treating the PG state as a phase characterized by some order parameter.

Hence, an important question that we will address is whether competing order is consistent with the scarcity of evidence for a clear phase associated with the pseudogap state. We shall show that in the pseudogap region the correlated many-body system becomes uniquely susceptible to quantum fluctuations in the pairing and antiferromagnetic degrees of freedom, and that these fluctuations can reconcile the successes of a mean-field approach to the pseudogap with the elusiveness of well-defined order parameters associated with the pseudogap.

### *1.3.3 Spatial inhomogeneity but a universal phase diagram*

There is evidence for a rather universal cuprate phase diagram (at least for hole-doped compounds), but there is at the same time strong indication of a broad variety of disorder in these same compounds. For example, cuprate high-temperature superconductors exhibit various spatial inhomogeneities such as stripes or checkerboards, particularly for lower hole doping and near magnetic vortex cores [33–37]. The behavior of the iron compounds is more varied, but there is a strong suggestion of universal properties in the phase diagram, and there is evidence for coexistence of AF and SC order both homogeneously (the same patch exhibiting AF and SC order) and inhomogeneously (AF and SC order in separate patches on the nanoscale). How does one reconcile evidence for a relatively universal phase diagram with evidence of a rich variety of inhomogeneity for individual compounds?

The relationship of such inhomogeneity to the unusual properties of these systems is not well established. Does it oppose superconductivity, does it enable superconductivity, or is it a distraction? Dopant atoms may favor superconductivity globally by enhancing charge carrier density, but may suppress superconductivity locally by inducing atomic-scale disorder. For example, strong disorder was found in atomically-resolved scanning tunneling microscope images of the superconducting gap for Bi-2212 [38], and it was concluded that this disorder derives primarily from dopant impurities. However, the charge variation between nanoregions was found to be small, implying that inhomogeneity may be tied to impurities but need not necessarily couple strongly to charge.

We shall show that such inhomogeneities follow generically from perturbations on the AF and SC correlations, largely independent of specifics and not necessarily coupled to charge variation. Further, we shall show that these properties are consistent with a global cuprate phase di-

agram, are directly related to the nature of pseudogap states, and imply a linkage among pseudogaps, inhomogeneity, and emergent behavior. Thus we shall propose a testable hypothesis for separating primary features from derivative features in the high-$T_c$ data set.

### 1.4 Addressing these issues within a unified framework

In this paper we use dynamical symmetries implemented in terms of Lie algebras and generalized coherent state to address all of these issues simultaneously. Specifically, we demonstrate in a solvable model motivated by cuprate phenomenology that the ground state of a minimal implementation of competing singlet pairing, antiferromagnetism, charge, and spin is an antiferromagnetic Mott insulator that is fundamentally unstable with respect to condensing electron hole pairs (and thus becoming a superconductor) at any finite hole doping. Furthermore, the same solution implies pseudogap states having many properties that are in quantitative agreement with data. We shall show that this ground state is unique near half filling and near optimal doping and beyond, but can become highly degenerate in the underdoped region. This degeneracy implies extreme sensitivity to perturbations and thus to a variety of induced inhomogeneity and other emergent behavior, but only in a narrow range of doping for underdoped compounds.

Then, we shall argue that the dynamical symmetry motivated by cuprate phenomenology is—despite superficial differences—also appropriate for a description of the Fe-based superconductors, thus providing a unified description of cuprate and iron superconductors. Indeed we shall argue that superficial differences between the cuprate and iron superconductors (for example, different orbital pairing symmetry) are not essential to a unified description at the broad level of understanding emergent superconductivity in these systems.

We shall then discuss the relationship of the theory derived in the present paper to various other proposals to explain high-temperature superconductors. We shall show, for example, that SU(4) coherent states contain much of the physics of resonating valence bond (RVB) models, but without some specifically RVB assumptions and with broader physics than RVB, that the present dynamical symmetry methodology provides a microscopic derivation of the Zhang SO(5) model as one of its approximate solutions, and that Hubbard or $t$–$J$ models and the present methodology are related by the adoption of fundamentally different (but equally valid) approaches to truncation of the Hilbert space for the full problem.

Finally, based on the experience with normal and unconventional superconductors discussed in this review and in the review of nuclear structure physics discussed in Ref. [18], we shall suggest an even more sweeping conjecture. Although superconductivity occurs in many forms in a variety of disciplines, we believe that all of these forms may represent a single basic mechanism involving a Cooper pairing instability that may occur in the presence of other strongly collective modes, and that has a common algebraic description across diverse systems in terms of fermion dynamical symmetries.

## 2 Truncation of large Hilbert spaces

In complicated many-body systems even the minimal Hilbert space is enormous and tractable theories must reduce this to a more manageable subspace. There are two common philosophies that may be followed in implementing such a truncation, which we shall term *microscopic-properties truncation* and *emergent-symmetry truncation.*

### 2.1 Truncation based on microscopic properties of the weakly-interacting system

Microscopic-properties truncation identifies key microscopic physical features of the idealized weakly-interacting system that are expected to be valid for the actual correlated many-body system, and uses that as a guide for truncating the full space. Typically Hubbard or $t$–$J$ model approaches are of this form. An assumption is made about the microscopic form of the important physical interactions and this is used to construct a simple Hamiltonian. In principle this implies no truncation of the configuration space, but in practice calculations are possible only if a small-enough subspace is chosen. The choice of this space is guided by microscopic physical insight and symmetries of the interactions (for example, a basis of spin states), often assuming that only states below some energy cutoff in the non-interacting basis contribute.

However, it is often not feasible to implement microscopic-properties truncation without drastic assumption. In high-$T_c$ superconductors, the correlations may be so strong that dynamics can no longer be given a meaningful description in terms of individual fundamental particles of the weakly-interacting system (see, for example, the discussion in Refs. [41–43]). The essential physics then tends to be governed by a few collective modes that are *emergent.* (An emergent mode is a new collective state that *emerges* because of interactions and not because it exists already in the microscopic constituents of the non-interacting system). Then a quite different kind of simplification is possible, based on what we shall term emergent-symmetry truncation.

### 2.2 Emergent-symmetry truncation

Truncation based on emergent symmetries is tailored for collective modes and long-range order. It identifies essential forms of the collective modes of interest and uses that as a guide to remove from the full space all states inconsistent with these forms, leaving a small collective subspace that is highly sympathetic to the relevant collective modes. In picturesque terms, we might also call this *Michelangelo truncation,* since the famous sculptor is said

to have replied to a query about how he made such beautiful statues that he looked at the block of stone, envisioned the statue trapped within it, and then chipped away everything that was not statue.

The most powerful systematic method of determining the essential form of collective modes is to identify a symmetry associated with them. We shall call this a symmetry of emergent modes, as opposed to symmetries of the weakly-interacting system. Note that if a mode is emergent there is no reason to expect its dynamical symmetries to have any direct relationship with the symmetries of the Hamiltonian for the weakly-interacting system in the full space before truncation.

### 2.3 Spontaneously-broken symmetries

This latter point is sufficiently important to merit further attention. A symmetry that is important at both the microscopic and collective levels is that associated with angular momentum (isotropy of space), which is conserved microscopically and also in the collective mode if it is treated exactly. But if we demand that angular momentum be conserved both microscopically and in the collective mode, the use of symmetry to truncate the space is limited to using the Clebsch–Gordan series to decompose the space into a finite number of subspaces that can be solved independently because each is labeled by a conserved angular momentum quantum number—in matrix language, the full matrix is transformed into one with block-diagonal form by a similarity transformation, permitting each block-diagonal submatrix to be solved independently.

A far more spectacular simplification results if we give up the requirement of angular momentum conservation for the collective state and treat it as geometrically deformed, thereby breaking rotational invariance. Then the symmetries of the emergent collective state do not include that of rotational invariance and the solution fails to conserve angular momentum. We then say that the emergent state breaks the symmetry of the true Hamiltonian for the system (which certainly does conserve angular momentum) spontaneously, because the Hamiltonian is symmetric but the wavefunction is not. Strictly, such a state is unphysical, because it breaks a symmetry observed by the exact solution. However, the broken-symmetry approximate solution can often be a very useful simplification because (i) the symmetry violation may not be very important for many physical properties of the system, and (ii) there are established techniques to restore the broken symmetry of the collective state by the use of projection integrals when it is important to do so.

### 2.4 Examples of emergent symmetries

The Bohr–Mottelson approach to nuclear structure physics utilizes as approximate solutions to the nuclear many-body problem deformed states that may break rotational invariance [44]. In the language employed here, these imply emergent symmetries. A second example of the emergent symmetries described above is the ordinary BCS superconducting state, which is separated by a phase transition from the normal state and cannot be reached by perturbation of the normal state. If we view the SC state as being described by a symmetry [pseudospin SU(2) in the case of a simple BCS superconductor], this symmetry is a symmetry of the collective many-body state, not of the underlying microscopic system. It is emergent. In this example, the BCS state may be viewed as a state that breaks rotational invariance in gauge space; more prosaically, it fails to conserve particle number [131].

The SU(4) symmetry described in this review may be understood as a very sophisticated pseudospin model embodying emergent symmetries associated with the collective modes of the system. In the symmetry limits it is an exact many-body solution, but in the coherent state approximation that we shall often employ it becomes a spontaneously-broken symmetry.

## 3 The dynamical symmetry method

This review is about emergent-symmetry truncation of a Hilbert space corresponding to the strongly-correlated electron problem. The approach that we shall use will rely upon the method of dynamical symmetries to identify the emergent collective subspace and implement the corresponding truncation. It begins with the following conjecture [18]:

**Conjecture 1** *Strongly correlated modes in fermion or boson many-body quantum systems imply a corresponding dynamical symmetry (a symmetry of the Hamiltonian or Lagrangian dynamics) described by a Lie algebra in the second-quantized operators representing the physical modes of the system.*

This is a conjecture, but there is a large amount of very strong circumstantial evidence to support its validity from various fields of many-body physics [18–22, 46].

### 3.1 Solution algorithm

Assuming the validity of the preceding conjecture, we may implement the following algorithm.

1) Identify a minimal set of emergent-state degrees of freedom thought to be physically relevant for the problem at hand, guided by phenomenology and theory.

2) Close a commutation algebra of manageable dimensionality on the second-quantized operators creating, annihilating, and counting the modes chosen in the first step, meaning that when all possible bilinear forms of this set of operators are commuted, the result is always a linear combination of the full operator set. This Lie algebra is

termed the *highest symmetry* of the problem, and may be specified completely in terms of the generators for emergent physical modes in the system, expressed as operators in second-quantized form.

3) Identify a collective subspace of the full Hilbert space by requiring that matrix elements of the operators found in the preceding step do not cause transitions out of the collective subspace. This (typically dramatic) reduction of the full space is termed *symmetry-dictated truncation.* The collective (emergent) states in this subspace will be of low energy but their wavefunction components are *selected by symmetry, not energy,* and may contain a mixture of both low and high energy pieces of the basis for the weakly interacting system.

4) Use standard methods to identify subalgebra chains of the highest algebraic structure that end in algebras for relevant conservation laws, such as those for charge and spin. Associated with each Lie algebra and subalgebra will be a corresponding *Lie group.* Each such subalgebra chain or corresponding subgroup chain defines a *dynamical symmetry* of the highest symmetry. Generally, more than one dynamical symmetry may be associated with a given highest symmetry.

5) Construct Hamiltonians that are polynomials in the Casimir invariants (*dynamical symmetry Hamiltonians*) for each chain. Each symmetry chain defines a wavefunction basis labeled by the eigenvalues of chain invariants (the Casimirs and the elements of the Cartan subalgebras), and a Hamiltonian that is diagonal in that basis because it is constructed explicitly from invariants. Thus, the Schödinger equation is solved analytically for each chain, by construction.

6) Calculate the physical implications of each of these dynamical symmetries by considering the diagonal and transitional matrix elements of physical relevance for the problem at hand. This is possible because of the eigenvalues and eigenvectors that were obtained in step 4, and because consistency of the symmetry requires that transition operators be related to group generators; otherwise transitions would mix irreducible multiplets and break the symmetry.

7) If the results of step (5) agree with experimental observables, indicating that a wise choice was made in step (1), construct the most general Hamiltonian in the model space, which is a linear combination of the terms in all the Hamiltonians for the symmetry group chains. The Casimir operators of different group chains do not generally commute with each other, so a Casimir invariant for one group chain may be a symmetry-breaking term for another group chain. Thus the competition between different dynamical symmetries and the corresponding (quantum) phase transitions may be studied.

8) Symmetry-limit solutions may be used as a starting point for more ambitious calculations that incorporate symmetry-breaking terms. Such more realistic approximations may be solved by (i) perturbation theory around the symmetry solutions (which are generally non-perturbative, so this corresponds to perturbation theory around a non-perturbative vacuum), (ii) by numerical diagonalization of symmetry-breaking terms, or (iii) by coherent-state or other approximations to the full Hamiltonian described above in step (6).

Representative application of these ideas for both fermion and boson systems may be found in Refs. [18–22, 46]. Our primary interest here is in strongly-correlated electron systems so we shall deal only with fermionic applications. Note also that the central concept employed here that symmetry can have dynamical and not just conservation-law implications, and that non-abelian symmetries imply theories that are at once richer and have fewer free parameters than abelian counterparts, are also key ingredients of local non-abelian gauge field theories in elementary particle physics, though the degrees of freedom and methodology are different there [45].

### 3.2 Validity and utility of the approach

The *only approximation* to the full quantum-mechanical problem in our approach is the space truncation. If all degrees of freedom are incorporated the resulting theory is microscopic and exact. Of course, practically only a few carefully-selected degrees of freedom can be included and the effect of the excluded space must be incorporated through renormalized (effective) interactions operating in the truncated space.

Thus, the utility of this approach depends on making a wise choice for the relevant collective degrees of freedom and on the availability of sufficient phenomenological or theoretical information to specify the effective interactions that operate within the truncated space. The validity of the resulting formalism then stands on whether predicted matrix elements agree with corresponding physical observables, once a small set of effective interaction parameters has been fixed by comparison with the global data set.

## 4 Strongly-correlated SU(4) electrons

Let us now introduce a formalism based on the approach described in Section 3 that is capable of dealing with the issues outlined in Section 1, by virtue of being complex enough to incorporate the essential physics and yet amenable to solutions that may be compared with data. We do so through a theory of strongly-correlated electrons that uses the power of Lie algebras, Lie groups, and generalized coherent states to truncate the Hilbert space to a manageable collective subspace [4, 18, 46–57].

### 4.1 Structure of the coherent pair basis

Motivated by phenomenology of the cuprates, our basic physical assumption is that the configuration space for a

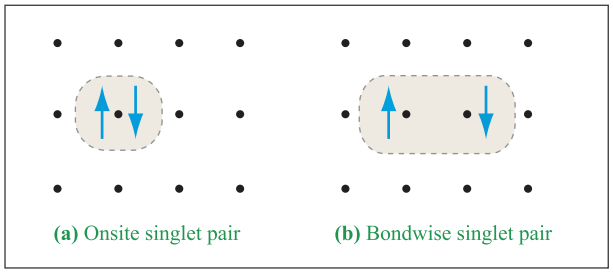



**Fig. 7** Onsite and nearest-neighbor bondwise singlet pairs.

minimal theory of high-temperature superconductivity is built from *coherent pairs* representing superpositions of particles or holes centered on different lattice sites [which we shall term *bondwise pairs*; see Fig. 7(b)]. In the interest of constructing a minimal theory, we shall neglect pairing between next nearest neighbors. We shall also consider an extended theory including in addition pairs of particles or holes defined on the same lattice sites [which we shall term *onsite pairs;* see Fig. 7(a)]. However, as will be discussed further in Section 7.2 below, such onsite pairing configurations are important in conventional superconductivity but are likely of less importance for the low-lying states in high-temperature superconductors. The next step is to use phenomenology as a guide to identify a set of operators that can be associated with the relevant physical degrees of freedom exhibited by this collective subspace.

### 4.2 The collective operators

We propose to solve for the doping and temperature dependence of observables in a theory that incorporates on an equal footing antiferromagnetism and (possibly unconventional) superconductivity. To construct a Hamiltonian embodying these degrees of freedom, and conservation laws for charge and spin, we employ the concept of a *complete set of quantum operators*, which may be defined in either physical or mathematical terms. (i) Physically, a complete set of quantum operators represents all degrees of freedom produced if an initial set of operators is allowed to undergo all possible (those not forbidden by fundamental principles) interactions among themselves. (ii) Mathematically, a complete set of quantum operators corresponds to a set of creation and annihilation operators closed under the operation of commutation, implying that the operators form a *Lie algebra*.

For the high temperature superconductor problem we require at a minimum three staggered magnetization operators $\boldsymbol{Q}$ to describe antiferromagnetism, creation and annihilation operators $p^\dagger$ and $p$ for bondwise singlet pairs plus a number operator $\hat{n}$ to describe superconductivity, and three spin operators $\boldsymbol{S}$ to describe electron spin. However, this set of nine operators is physically incomplete since scattering of singlet pairs (antiparallel spins on adjacent sites) from the AF particle–hole degrees of freedom can produce triplet pairs (parallel spins on adjacent sites), which are not part of the operator set. This is illustrated in Fig. 8.

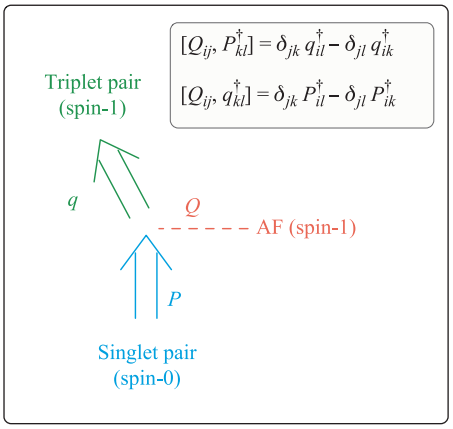



**Fig. 8** Scattering of singlet pairs by AF operators will necessarily produce triplet pairs, even if none existed before. Thus a physically and mathematically consistent Hilbert space must contain both kinds of pairs in the presence of antiferromagnetism.

The mathematical statement of this incompleteness is that the set $\{\boldsymbol{Q}, p^\dagger, p, \hat{n}, \boldsymbol{S}\}$ does not close a Lie algebra under commutation of set members, because commuting singlet-pair operators with antiferromagnetic operators produces triplet pair operators and commuting triplet-pair operators with antiferromagnetic operators produces singlet pair operators (inset to Fig. 8). The physical picture of Fig. 8 suggests that a self-consistent Hilbert space containing singlet pairs must contain triplet pairs also if antiferromagnetic interactions are present. As we shall now demonstrate, a minimal physically-complete operator set results if we add to the original nine operators three creation and three annihilation operators for bondwise spin-triplet pairs defined on adjacent lattice sites.

The operators that we shall use are particle–hole symmetric (though our predictions for observables generally will not be; see the discussion in note [58]). Unless specified explicitly in the following, we shall use "electrons" to reference either electrons or electron holes. We begin by introducing the following 16 operators:

$$\boldsymbol{S} = \left(\frac{S_{12}+S_{21}}{2},\ -\mathrm{i}\frac{S_{12}-S_{21}}{2},\ \frac{S_{11}-S_{22}}{2}\right), \tag{1a}$$

$$\boldsymbol{Q} = \left(\frac{Q_{12}+Q_{21}}{2}, -\mathrm{i}\frac{Q_{12}-Q_{21}}{2}, \frac{Q_{11}-Q_{22}}{2}\right), \tag{1b}$$

$$\boldsymbol{\pi}^\dagger = \left(\mathrm{i}\frac{q^\dagger_{11}-q^\dagger_{22}}{2},\ \frac{q^\dagger_{11}+q^\dagger_{22}}{2}, -\mathrm{i}\frac{q^\dagger_{12}+q^\dagger_{21}}{2}\right), \tag{1c}$$

$$\boldsymbol{\pi} = \left(-\mathrm{i}\frac{q_{11}-q_{22}}{2},\ \frac{q_{11}+q_{22}}{2},\ \mathrm{i}\frac{q_{12}+q_{21}}{2}\right), \tag{1d}$$

$$p^\dagger = p^\dagger_{12}, \quad p = p_{12}, \tag{1e}$$

$$\hat{n} = \sum_{k,i} c^\dagger_{k,i} c_{k,i} = S_{11} + S_{22} + \Omega, \tag{1f}$$

$$Q_+ = Q_{11} + Q_{22} = \sum_k (c^\dagger_{k+Q\uparrow} c_{k\uparrow} + c^\dagger_{k+Q\downarrow} c_{k\downarrow}), \tag{1g}$$

in which we define

$$p^\dagger = \sum_{\boldsymbol{k}bb'} g(\boldsymbol{k})\alpha_{\boldsymbol{k}b}\alpha_{-kb'}c^\dagger_{\boldsymbol{k}b\uparrow}c^\dagger_{-\boldsymbol{k}b'\downarrow}, \quad p = (p^\dagger)^\dagger, \tag{2a}$$

$$q^\dagger_{ij} = \sum_{\boldsymbol{k}bb'} g(\boldsymbol{k})\alpha_{\boldsymbol{k}+\boldsymbol{Q},b}\alpha_{-kb'}c^\dagger_{\boldsymbol{k}+\boldsymbol{Q},bi}c^\dagger_{-\boldsymbol{k},b'j}, \quad q = (q^\dagger)^\dagger, \tag{2b}$$

$$Q_{ij} = \sum_{\boldsymbol{k}bb'} \alpha_{\boldsymbol{k}+\boldsymbol{Q},b}\alpha^*_{\boldsymbol{k}b'}c^\dagger_{\boldsymbol{k}+\boldsymbol{Q},bi}c_{\boldsymbol{k}b'j}, \tag{2c}$$

$$S_{ij} = \sum_{\boldsymbol{k}bb'} \alpha_{\boldsymbol{k}b}\alpha^*_{\boldsymbol{k}b'}c^\dagger_{\boldsymbol{k},bi}c_{\boldsymbol{k},b'j} - \frac{1}{2}\Omega\delta_{ij}, \tag{2d}$$

where $\alpha_{\boldsymbol{k}b}$ is the amplitude to find an electron in a band labeled by $b$ with momentum $\boldsymbol{k}$, $c^\dagger_{\boldsymbol{k},b,i}$ creates a fermion of momentum $\boldsymbol{k}$ and spin projection $i, j = 1$ or $2 = \uparrow$ or $\downarrow$ in band $b$, $\boldsymbol{Q}$ is an AF ordering vector, $\Omega$ is the effective lattice degeneracy, which is the maximum allowed number of doped electrons that can form coherent SU(4) pairs (explained further below), and $g(\boldsymbol{k})$ is a pairing formfactor. We may attach a simple physical interpretation to the operators in Eqs. (1):

- The vector $\boldsymbol{S}$ is the electron spin operator.
- The vector $\boldsymbol{Q}$ is the staggered magnetization characterizing the antiferromagnetism.
- $\boldsymbol{\pi}^\dagger$ ($\boldsymbol{\pi}$) is a vector of creation (annihilation) operators for bondwise spin-triplet pairs.
- $p^\dagger$ ($p$) is a creation (annihilation) operator for bondwise singlet pairs.
- $\hat{n}$ is the electron number operator.
- $Q_+$ is a commensurate charge density wave operator.

It will sometimes prove useful to replace the number operator $\hat{n}$ in Eq. (1f) with

$$M = \frac{1}{2}(S_{11} + S_{22}) = \frac{1}{2}(\hat{n} - \Omega), \tag{3}$$

where we may interpret $M$ physically as the charge operator.

The preceding operators may receive contributions from more than one band. If we introduce effective one-band creation and annihilation operators through

$$a^\dagger_{ki} = \sum_b \alpha_{\boldsymbol{k}b}c^\dagger_{\boldsymbol{k}bi}, \quad a_{\boldsymbol{k}i} = (a^\dagger_{ki})^\dagger, \quad \sum_b |\alpha_{\boldsymbol{k}b}|^2 = 1, \tag{4}$$

then Eqs. (2) may we written as

$$p^\dagger = \sum_{\boldsymbol{k}} g(\boldsymbol{k})c^\dagger_{\boldsymbol{k}\uparrow}c^\dagger_{-\boldsymbol{k}\downarrow}, \quad p = (p^\dagger)^\dagger, \tag{5a}$$

$$q^\dagger_{ij} = \sum_{\boldsymbol{k}} g(\boldsymbol{k})c^\dagger_{\boldsymbol{k}+\boldsymbol{Q},i}c^\dagger_{-\boldsymbol{k},j}, \quad q = (q^\dagger)^\dagger, \tag{5b}$$

$$Q_{ij} = \sum_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k}+\boldsymbol{Q},i}c_{\boldsymbol{k},j}, \quad S_{ij} = \sum_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k},i}c_{\boldsymbol{k},j} - \frac{1}{2}\Omega\delta_{ij}, \tag{5c}$$

which is the form of the operators for the original single-band SU(4) model introduced in Ref. [46]. Although Eqs. (5) are an adequate starting point for discussion of cuprate superconductivity, the multiband expressions in Eqs. (2) are more appropriate for applications such as the iron-based superconductors where multiband pairing is important.

### 4.3 The SU(4) algebra and subalgebras

Inserting the AF ordering vector $\boldsymbol{Q} = (Q_x, Q_y) = (0, \pi)$ appropriate for the observed FeAs magnetic structure, or $\boldsymbol{Q} = (Q_x, Q_y) = (\pi, \pi)$ appropriate for the cuprates, and calculating all commutators for the 16 operators in Eqs. (1), the set is found to be closed under commutation if three conditions are satisfied by the pairing formfactor:

$$g(\boldsymbol{k}) = g(-\boldsymbol{k}), \quad g(\boldsymbol{k} + \boldsymbol{Q}) = \pm g(\boldsymbol{k}), \quad |g(\boldsymbol{k})| = 1. \tag{6}$$

(We shall elaborate on the physical meaning of these constraints in Section 7.1.) If these conditions are met, the operators defined in Eq. (1) close a U(4) $\supset$ U(1) $\times$ SU(4) Lie algebra, where the U(1) factor is generated by the commensurate charge density wave operator $Q_+$, which commutes with all other generators [46, 49]. Because of the direct-product structure, one can without loss of generality view the theory as an SU(4) theory describing superconductivity and antiferromagnetism, and global charge and spin conservation, with the U(1) charge-density wave sector treated independently. This SU(4) group has three independent subgroup chains

$$\begin{matrix} \curvearrowright \\ \mathrm{SU}(4) \\ \curvearrowright \end{matrix} \begin{matrix} \mathrm{SO}(4) \times \mathrm{U}(1) \supset \mathrm{SU}(2)_\mathrm{s} \times \mathrm{U}(1), \\ \supset \mathrm{SO}(5) \supset \mathrm{SU}(2)_\mathrm{s} \times \mathrm{U}(1), \\ \mathrm{SU}(2)_\mathrm{p} \times \mathrm{SU}(2)_\mathrm{s} \supset \mathrm{SU}(2)_\mathrm{s} \times \mathrm{U}(1), \end{matrix} \tag{7a, 7b, 7c}$$

ending in the subgroup $\mathrm{SU}(2)_\mathrm{s} \times \mathrm{U}(1)$ representing spin and charge conservation, with the U(1) subgroup being generated by the charge operator $M$ and the $\mathrm{SU}(2)_\mathrm{s}$ subgroup being generated by the three spin operators $S_1$, $S_2$, and $S_3$. Some important properties of this group and subgroup structure are summarized in Tables 1 and 2 of Appendix A.

As we shall discuss further below, the subgroup chains defined in Eqs. (7) imply three *fermion dynamical symmetries* [18] that will permit exact many-body solutions to be obtained for particular ratios of antiferromagnetic and pairing coupling strengths. The methodology is that of *symmetry-dictated truncation,* as outlined in Section 3. The subgroup chains of Eqs. (7) and the corresponding many-body solutions imply that charge and spin are conserved. In Section 6 we shall find it useful to introduce *approximate solutions* through coherent state methods that are generalizations of the BCS solution and lead to spontaneous symmetry breaking and to intrinsic states violating particle number conservation, but the starting point (7) conserves both charge and spin.

### 4.4 Collective subspace and associated Hamiltonian

The group SU(4) is rank-3 and the irreducible representations (irreps) may be labeled by three weight-space quantum numbers, $(\sigma_1, \sigma_2, \sigma_3)$ [59]. We assume a collective subspace illustrated in Fig. 9 that is spanned by the vectors

$$|S\rangle = |n_x n_y n_z n_s\rangle = (\pi_x^\dagger)^{n_x} (\pi_y^\dagger)^{n_y} (\pi_z^\dagger)^{n_z} (p^\dagger)^{n_s} |0\rangle \,. \quad (8)$$

If there are no unpaired particles, the collective subspace is associated with "fully-stretched", and therefore maximally collective, irreducible representations (irreps) of the form

$$(\sigma_1, \sigma_2, \sigma_3) = \left(\frac{\Omega}{2}, 0, 0\right). \quad (9)$$

(More general representations having broken pairs of particles are discussed in Ref. [50].) A subspace associated with such maximally-collective configurations is an obvious candidate for describing the lowest-energy states of the system. This wavefunction represents a coherent superposition of pairs and implies a rather rich structure, as illustrated for the real space in Fig. 10.

The commensurate charge density wave operator $Q_+$ defined in Eq. (1g) is the generator of the U(1) factor in $\mathrm{U}(4) \supset \mathrm{U}(1) \times \mathrm{SU}(4)$. It commutes with all generators so it annihilates the state $|S\rangle$ and

$$\langle S| Q_+ |S\rangle = 0, \quad (10)$$

which means that charge-density wave excitations are excluded in the symmetry limits for the low-lying collective subspace of the effective theory.

Thus we shall ignore the U(1) charge-density wave for much of the subsequent discussion and concentrate on the antiferromagnetic–superconductivity competition embodied in the SU(4) symmetry. However, we shall address the issue of possible charge degrees of freedom induced by perturbations of the SU(4) subspace later in Section 11.2.

The group SU(4) has a quadratic Casimir operator (see Table 1 of Appendix A),

$$C_{\mathrm{su}(4)} = \boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} + p^\dagger p + \boldsymbol{S} \cdot \boldsymbol{S} + \boldsymbol{Q} \cdot \boldsymbol{Q} + M(M-4), \quad (11)$$

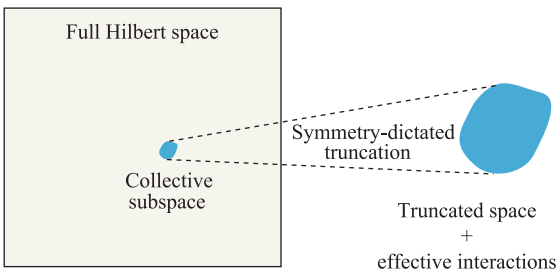



**Fig. 9** Symmetry-dictated truncation of the full Hilbert space to a small collective subspace. The actual collective subspace for the present discussion is miniscule compared with the full space.

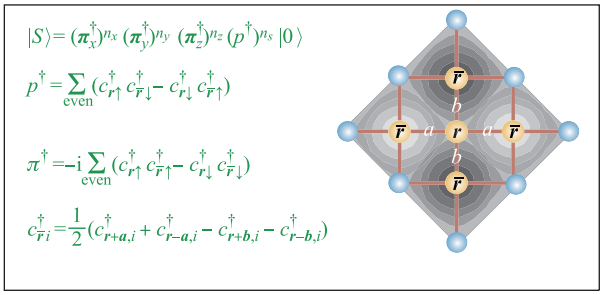



**Fig. 10** Linear superpositions in the coordinate space implied by the wavefunction (8). The inset figure is discussed more extensively in connection with Fig. 17.

and the corresponding expectation value of the SU(4) quadratic Casimir operator evaluated in the irreducible representations (9) is a constant,

$$\langle C_{\mathrm{su}(4)} \rangle = \frac{\Omega}{2}\left(\frac{\Omega}{2} + 4\right). \quad (12)$$

The most general 2-body Hamiltonian within the collective pair subspace consists of a linear combination of (lowest-order) Casimir operators $C_{\mathrm{g}}$ for all subgroups g (Table 1 of Appendix A) [60]. These are

$$C_{\mathrm{so}(5)} = \boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} + \boldsymbol{S} \cdot \boldsymbol{S} + M(M-3), \quad (13a)$$
$$C_{\mathrm{so}(4)} = \boldsymbol{Q} \cdot \boldsymbol{Q} + \boldsymbol{S} \cdot \boldsymbol{S}, \quad (13b)$$
$$C_{\mathrm{su}(2)_{\mathrm{p}}} = p^\dagger p + M(M-1), \quad (13c)$$
$$C_{\mathrm{su}(2)_{\mathrm{s}}} = \boldsymbol{S} \cdot \boldsymbol{S}, \quad (13d)$$
$$C_{\mathrm{u}(1)} = M \text{ and } M^2. \quad (13e)$$

We may utilize that the SU(4) Casimir expectation value (12) is constant to eliminate terms in $\boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi}$, and the most general SU(4) 2-body Hamiltonian can be expressed as [46, 57]

$$H = H_0 - \tilde{G}_0 \left[(1-\sigma) p^\dagger p + \sigma \boldsymbol{Q} \cdot \boldsymbol{Q}\right] + g' \boldsymbol{S} \cdot \boldsymbol{S}, \quad (14)$$

where $H_0$, $\tilde{G}_0$ and $g'$ are effective interaction parameters, $p^\dagger$ creates singlet pairs, $\boldsymbol{Q}$ is the staggered magnetization, $\boldsymbol{S}$ is spin, $\tilde{G}_0 = \chi(x) + G_0(x)$, and the parameter $\sigma$, given by

$$\sigma = \sigma(x) = \frac{\chi(x)}{\chi(x) + G_0(x)}, \quad (15)$$

with $G_0(x)$ and $\chi(x)$ the effective SC and AF coupling strengths, respectively, governs the relative strength of antiferromagnetic and pairing interactions in the subspace [51, 52]. In these expressions, doping is characterized by a parameter

$$x = 1 - \frac{n}{\Omega} \quad (16)$$

for an $n$-electron system, with $\Omega$ the maximum number of doped holes (or doped electrons for electron-doped compounds) that can form coherent pairs, assuming the normal state at half filling ($n = \Omega$, implying $M = 0$) to be

the vacuum. Since $\Omega - n$ is the hole number when $n < \Omega$, positive $x$ represents the case of hole-doping, with $x = 0$ corresponding to half filling (no doping) and $x = 1$ to maximal hole-doping. Negative $x$ ($n > \Omega$) is then the relative doping fraction for electron-doping.

The parameter $x$ is the effective doping concentration, corresponding to the ratio of the number of doped pairs $\frac{1}{2}(\Omega - n)$ to the pair degeneracy $\frac{1}{2}\Omega$ of the collective subspace. The physical doping fraction, defined as $P = (\Omega - n)/\Omega_e$, where $\Omega_e$ is the number of physical lattice sites, is related to $x$ through

$$P = x\,\frac{\Omega}{\Omega_e} = x\,P_{\rm f}, \quad P_{\rm f} \equiv \frac{\Omega}{\Omega_e}, \tag{17}$$

with $P > 0$ for hole doping and $P < 0$ for electron doping. $P_{\rm f}$ can be regarded as the maximum value of the physical doping fraction $P$ that is consistent with SU(4) symmetry, as we shall explain further below.

## 5 The dynamical symmetry limits

Each of the three dynamical symmetry limits in Eqs. (7) defines a basis in which the effective Hamiltonian (14) is diagonal. Thus, they have exact solutions [46–48] that may be constructed using the methods developed in Ref. [18]. These solutions result from special choices of the parameter $\sigma$ in the Hamiltonian (14) and are summarized in Table 3.

1) The SO(4) limit corresponds to choosing $\sigma = 1$ in Eq. (14), which eliminates the pairing terms in the Hamiltonian. It represents a collective Mott-insulator, AF state defined by the subgroup chain of Eq. (7a).

2) The $\mathrm{SU(2)}_{\rm p}$ limit [which we will term the SU(2) limit for brevity] corresponds to choosing $\sigma = 0$ in Eq. (14), which eliminates the antiferromagnetic terms in the Hamiltonian. It represents a collective $d$-wave singlet SC state defined by the subgroup chain of Eq. (7c).

3) The SO(5) limit corresponds to choosing $\sigma = \frac{1}{2}$ in Eq. (14), which leads to a Hamiltonian in which pairing and antiferromagnetism enter with equal strengths. The SO(5) limit represents a *critical dynamical symmetry* that interpolates dynamically between the SC and AF limits. It is defined by the subgroup chain of Eq. (7b).

**Table 3** Dynamical symmetries of the Hamiltonian (14).

| $\sigma$ | Symmetry | Physical interpretation |
|---|---|---|
| 0 | SO(4) | Antiferromagnetic Mott insulator |
| $\frac{1}{2}$ | SO(5) | Critical dynamical symmetry |
| 1 | SU(2) | $d$-wave singlet superconductor |

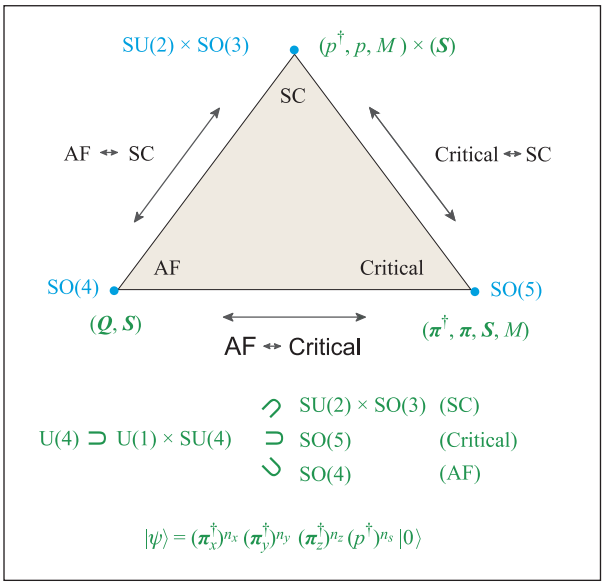



**Fig. 11** Relationships among the SU(4) dynamical symmetries.

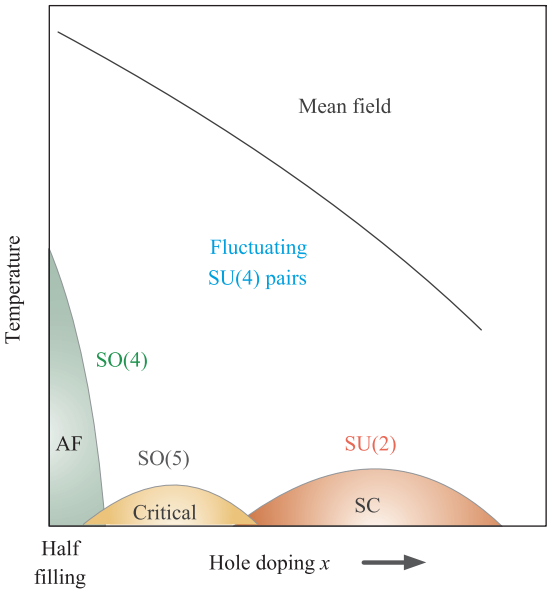



**Fig. 12** A schematic SU(4) phase diagram. The SO(4) symmetry corresponds to an antiferromagnetic Mott insulator, the SU(2) symmetry corresponds to a $d$-wave singlet superconductor, and the SO(5) symmetry corresponds to a critical dynamical symmetry that interpolates between the AF and SC phases. A more quantitative discussion of the expected phase structure will be given in Section 10.

The detailed mathematical properties of these symmetry limits are summarized in Table 2 of Appendix A and their schematic relationships are displayed in Fig. 11. A generic doping–temperature phase diagram corresponding to these dynamical symmetries is displayed in Fig. 12. We now discuss the justification for the above identifications and the important physical properties of the exact solutions in these symmetry limits.

### 5.1 The SO(4) dynamical-symmetry limit

The SU(4) subgroup chain

$$\mathrm{SU}(4) \supset \mathrm{SO}(4) \times \mathrm{U}(1) \supset \mathrm{SU}(2)_{\mathrm{s}} \times \mathrm{U}(1),$$

which we term the SO(4) *dynamical symmetry* for brevity, is the symmetry limit of Eq. (14) when $\sigma = 1$. As we now argue, it corresponds to a collective state having long-range antiferromagnetic order and Mott insulator character. The SO(4) subgroup is locally isomorphic to the product group $\mathrm{SU}(2)_F \times \mathrm{SU}(2)_G$ that is generated by the linear combinations

$$\boldsymbol{F} = \frac{1}{2}(\boldsymbol{Q} + \boldsymbol{S}), \quad \boldsymbol{G} = \frac{1}{2}(\boldsymbol{Q} - \boldsymbol{S}) \tag{18}$$

of the original SO(4) generators $\boldsymbol{Q}$ and $\boldsymbol{S}$. The new generators $\boldsymbol{F}$ and $\boldsymbol{G}$ may be interpreted physically by noting that if $Q_{ij}$ and $S_{ij}$ defined in momentum space in Eq. (2) are transformed to the physical coordinate lattice we obtain,

$$Q_{ij} = \sum_r (-)^r c^\dagger_{ri} c_{rj} = \sum_{r=\mathrm{even}} c^\dagger_{ri} c_{rj} - \sum_{r=\mathrm{odd}} c^\dagger_{ri} c_{rj}, \tag{19a}$$

$$S_{ij} = \sum_r c^\dagger_{ri} c_{rj} = \sum_{r=\mathrm{even}} c^\dagger_{ri} c_{rj} + \sum_{r=\mathrm{odd}} c^\dagger_{ri} c_{rj}, \tag{19b}$$

implying that $\boldsymbol{F}$ is the generator of total spin on even sites and $\boldsymbol{G}$ is the generator of total spin on odd sites, as illustrated schematically in Fig. 13.

Thus, SO(4) is generated by two independent spin operators: one that is the total spin on all sites and one that is the difference in spins on even and odd sites of the spatial lattice. This is clearly an algebraic version of the physical picture associated with antiferromagnetic long-range order.

The SO(4) Casimir operator may be expressed as

$$C_{\mathrm{SO}(4)} = 2(\boldsymbol{F}^2 + \boldsymbol{G}^2). \tag{20}$$

The SO(4) representations can be labeled by the spin-like quantum numbers $(F = \frac{1}{2}w, G = \frac{1}{2}w)$, where $w = N - \mu$ with $\mu = 0, 2, \ldots, N$. Eigenstates are labeled by $w$ and the spin $S$,

$$\psi(\mathrm{SO}(4)) = |N, w, S, m_s\rangle,$$

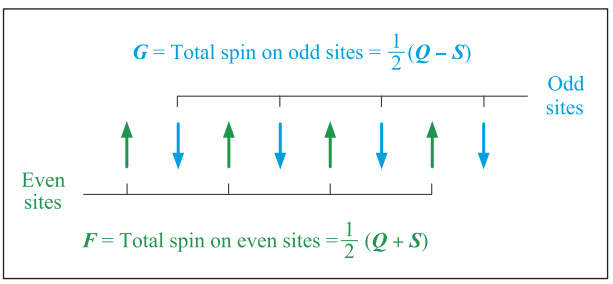



**Fig. 13** Geometrical interpretation of the SO(4) dynamical symmetry as an antiferromagnetic state.

and are of dimension $(w+1)^2$.

The ground state corresponds to $\omega = N$ and $S = 0$, and has $\frac{1}{2}n$ spin-up electrons on the even sites $(F = N/2)$ and $\frac{1}{2}n$ spin-down electrons on odd sites $(G = N/2)$, or vice versa. Thus it has maximal staggered magnetization,

$$Q = \frac{1}{2}\Omega(1-x) = \frac{1}{2}n, \tag{21}$$

and a large energy gap associated with the antiferromagnetic correlation $\boldsymbol{Q} \cdot \boldsymbol{Q}$,

$$\Delta E = 2\chi_{\mathrm{eff}}(1-x)\Omega. \tag{22}$$

In addition, the pairing gap

$$\Delta = \frac{1}{2}G^0_{\mathrm{eff}}\Omega\sqrt{x(1-x)} \tag{23}$$

is small near half filling $(x = 0)$, and we shall demonstrate below that the SU(4) symmetry requires that the lattice have no double occupancy at half filling and thus be a Mott insulator. We conclude that near half filling these SO(4) states are identified naturally with a collective, antiferromagnetic, Mott insulating state. This identification will be strengthened below by examination of the ground-state energy surface evaluated in this limit (illustrated in Fig. 23 and discussed in Section 8.1).

### 5.2 The SU(2) dynamical-symmetry limit

The SU(4)subgroup chain

$$\mathrm{SU}(4) \supset \mathrm{SU}(2)_{\mathrm{p}} \times \mathrm{SU}(2)_{\mathrm{s}} \supset \mathrm{SU}(2)_{\mathrm{s}} \times \mathrm{U}(1),$$

which we shall term the *SU(2) dynamical symmetry* for brevity, corresponds to the $\sigma = 0$ symmetry limit of Eq. (14). The eigenstates are labeled by $v$ and spin $S$,

$$\psi(\mathrm{SU}(2)) = |N, v, S, m_s\rangle,$$

and are of dimension $\frac{1}{2}(v+1)(v+2)$. The seniority-like quantum number $v$ is the number of particles that do not form singlet pairs (see Table 2 of Appendix A). The ground state has $v = 0$, implying that all electrons are singlet-paired. In addition, there exists a large pairing gap

$$\Delta E = G^{(0)}_{\mathrm{eff}}\Omega \tag{24}$$

(Table 2 of Appendix A), the pairing correlation is the largest among the three symmetry limits, and the staggered magnetization vanishes in the ground state:

$$\Delta = \frac{1}{2}G^0_{\mathrm{eff}}\Omega\sqrt{1-x^2}, \quad Q = 0. \tag{25}$$

Thus we interpret this dynamical symmetry as a *d*-wave pair condensate associated with a collective spin-singlet superconducting state. This identification will be strengthened below by examination of the ground-state energy surface evaluated in this limit (illustrated in Fig. 23 and discussed in Section 8.2).

### 5.3 The SO(5) dynamical-symmetry limit

The SU(4) subgroup chain

$$\mathrm{SU}(4) \supset \mathrm{SO}(5) \supset \mathrm{SU}(2)_s \times \mathrm{U}(1),$$

which we shall term the SO(5) *dynamical symmetry,* has the nature of a transitional or *critical dynamical symmetry.* This symmetry limit results when $\sigma = \frac{1}{2}$ in Eq. (14). The SO(5) irreps are labeled by a quantum number $\tau$ and the eigenstates may be labeled by $\tau$ and the spin $S$,

$$\psi(\mathrm{SO}(5)) = |N, \tau, S, m_s\rangle,$$

with $N = \Omega/2 - \tau + \lambda$, where $\lambda$ is the number of $\pi$ pairs. For given $N$ the irreducible representations are of dimensionality $\frac{1}{2}(\lambda+1)(\lambda+2)$ and the ground state has $\lambda = 0$ and $S = 0$.

The SO(5) dynamical symmetry has very unusual character. Although the expectation values of $\Delta$ and $Q$ for the ground state in this symmetry limit are the same as those of Eq. (25) for the SU(2) case, there exist many states with different values of $\lambda$ (the number of $\pi$ pairs) that can mix easily with the ground state when $x$ is small because the excitation energy in this symmetry limit is

$$\Delta E = \lambda G_{\mathrm{eff}}^{(0)} \Omega x \tag{26}$$

(see Table 2 of Appendix A). In particular, at half filling ($x = 0$) the ground state is highly degenerate with respect to $\lambda$ and mixing different numbers of $\pi$ pairs in the ground state costs no energy. The $\pi$ pairs must be responsible for the antiferromagnetism in this phase, since within the model space only $\pi$ pairs carry spin if there are no unbroken pairs. Thus the ground state in this symmetry limit has large-amplitude fluctuation in the AF (and SC) order. As such, it will be seen to play a role as a doorway between antiferromagnetic and superconducting order, as illustrated in Fig. 14. This identification will be strengthened below by examination of the ground-state energy surface evaluated in this limit (illustrated in Fig. 23 and discussed in Section 8.3).

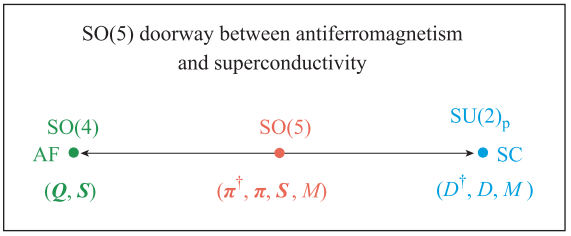



**Fig. 14** The SO(5) critical dynamical symmetry as a doorway connecting antiferromagnetism and superconductivity.

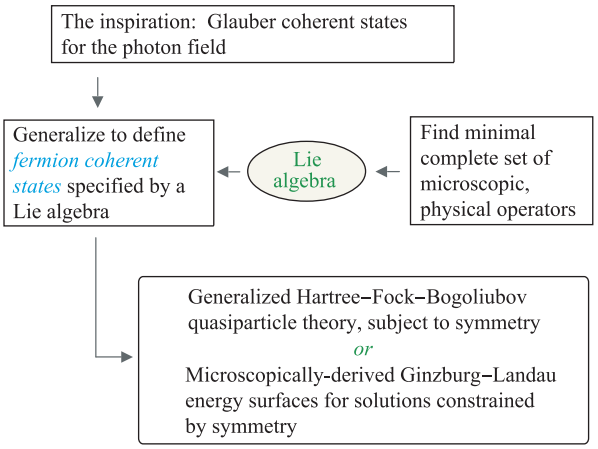



**Fig. 15** The method of generalized coherent states.

## 6 Generalized SU(4) coherent states

For values of the coupling-strength ratio $\sigma$ not equal to the special choices $\{0, \frac{1}{2}, 1\}$, the SU(4) symmetry model does not have an exact solution. However, an approximate solution can be obtained using the *generalized coherent-state method* [61, 62, 64, 65], which relates a many-body algebraic theory with unbroken symmetry to an approximation of that theory exhibiting spontaneously-broken symmetry. The methodology is outlined in Fig. 15 and explained more extensively below.

### 6.1 Associating coherent states with Lie algebras

The work of Gilmore [66, 67] and Perelomov [68] (see also Klauder [69]) showed that Glauber coherent states [24] of the electromagnetic field could be generalized to coherent states specified by an arbitrary Lie algebra. Specifically, it was found that the original Glauber theory for coherent photon states could be expressed in terms of an SU(2) Lie algebra by examining the commutation properties of the second-quantized operators of the theory, and then this formalism could be generalized to encompass a set of such operators closed under any Lie algebra. The resulting theory can apply to either fermion or boson fields, provided that the relevant creation and annihilation operators close a Lie algebra. This extension of the Glauber theory to arbitrary Lie algebras is termed the *generalized coherent state method.* A comprehensive review exists [61], so we proceed directly to application of the generalized coherent state method to fermions described by the SU(4) algebra of Section 4.3.

### 6.2 SU(4) coherent states

The SU(4) coherent state $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \mathcal{T} \mid 0^*\rangle, \tag{27}$$

where the operator $\mathcal{T}$ is given by

$$\mathcal{T} = \exp(\eta_{00} p_{12}^\dagger + \eta_{10} q_{12}^\dagger - \mathrm{h.c.}). \tag{28}$$

In Eq. (27), $|0^*\rangle$ is the physical vacuum (ground state) of the system, the real parameters $\eta_{00}$ and $\eta_{10}$ are symmetry-constrained variational parameters, and h.c.

stands for hermitian conjugation. The variational parameters weight the elementary excitation operators $p^\dagger_{12}$ and $q^\dagger_{12}$ in Eq. (28), so they represent collective state parameters for a pair subspace truncated under the SU(4) symmetry [70]. We then express a symmetry-constrained variational Hamiltonian as

$$H' = H - \lambda\hat{n}, \tag{29}$$

where $H$ is the Hamiltonian given in Eq. (14) and $\lambda$ is the chemical potential, determined by requiring that the average particle number be conserved. The parameters $\eta_{00}$ and $\eta_{10}$ in Eq. (28) are determined by the variational condition $\delta\langle H'\rangle = 0$, where

$$\langle H'\rangle \equiv \langle 0^*|H'|0^*\rangle \tag{30}$$

is the expectation value of $H'$ with respect to the ground state $|0^*\rangle$.

### 6.3 Generalized quasiparticle transformation

As shown in Appendix A of Ref. [50], it is convenient to evaluate the variation $\delta\langle H'\rangle = 0$ using a 4-dimensional matrix representation that was employed in Refs. [48, 62]. In this representation the unitary operator $\mathcal{T}$ implements a transformation from the original particle basis to a quasiparticle basis and the variational parameters $\eta_{00}$ and $\eta_{10}$ are replaced, respectively, by $u_\pm$ and $v_\pm$, subject to a unitarity condition

$$u_\pm^2 + v_\pm^2 = 1. \tag{31}$$

Under this transformation the physical vacuum state $|0^*\rangle$ is transformed to a quasiparticle vacuum state $|\psi\rangle$ and the basic fermion operators $c_{\boldsymbol{r}i}$ are converted to new quasiparticle operators $a_{\boldsymbol{r}i}$ through the transformation

$$\mathcal{T}\begin{pmatrix} c_{\boldsymbol{r}\uparrow} \\ c_{\boldsymbol{r}\downarrow} \\ c^\dagger_{\bar{\boldsymbol{r}}\uparrow} \\ c^\dagger_{\bar{\boldsymbol{r}}\downarrow} \end{pmatrix}|0^*\rangle = \begin{pmatrix} a_{\boldsymbol{r}\uparrow} \\ a_{\boldsymbol{r}\downarrow} \\ a^\dagger_{\bar{\boldsymbol{r}}\uparrow} \\ a^\dagger_{\bar{\boldsymbol{r}}\downarrow} \end{pmatrix}|\psi\rangle\,. \tag{32}$$

(see Appendix A of Ref. [50]), with

$$\begin{aligned}
\left(u_+c_{\boldsymbol{r}\uparrow} + v_+c^\dagger_{\bar{\boldsymbol{r}}\downarrow}\right)|0^*\rangle &= a_{\boldsymbol{r}\uparrow}|\psi\rangle, \\
\left(u_-c_{\boldsymbol{r}\downarrow} - v_-c^\dagger_{\bar{\boldsymbol{r}}\uparrow}\right)|0^*\rangle &= a_{\boldsymbol{r}\downarrow}|\psi\rangle, \\
\left(u_-c^\dagger_{\bar{\boldsymbol{r}}\uparrow} + v_-c_{\boldsymbol{r}\downarrow}\right)|0^*\rangle &= a^\dagger_{\bar{\boldsymbol{r}}\uparrow}|\psi\rangle, \\
\left(u_+c^\dagger_{\bar{\boldsymbol{r}}\downarrow} - v_+c_{\boldsymbol{r}\uparrow}\right)|0^*\rangle &= a^\dagger_{\bar{\boldsymbol{r}}\downarrow}|\psi\rangle.
\end{aligned}$$

Thus this is a transformation of the Bogoliubov form: each quasiparticle state is a mixture of a particle and a hole, and the coherent state $|\psi\rangle$ is a quasiparticle vacuum constrained to respect SU(4) symmetry. It follows that the generalized coherent state method described here is equivalent to the most general Hartree–Fock–Bogoliubov (HFB) variational method, but subject to an SU(4) symmetry constraint [46–48]. By using the 4-dimensional matrix representation, the expectation value for any operator $\hat{O}$ in the coherent state representation can be calculated through the transformation

$$\langle 0^*|\hat{O}|0^*\rangle = \langle\psi|\mathcal{T}\hat{O}\mathcal{T}^{-1}|\psi\rangle, \tag{33}$$

as detailed in Appendix A of Ref. [50].

### 6.4 Temperature dependence

At finite temperature $|\psi\rangle$ will generally no longer be a quasiparticle vacuum state and the quasiparticle annihilation operators acting on $|\psi\rangle$ do not necessarily give zero. In Appendix B of Ref. [50] a formalism is derived to deal with the finite-temperature case, by replacing the state $|\psi\rangle$ with a state $|\psi(T)\rangle$ in which the quasiparticles at temperature $T$ may be thermally excited.

For a temperature $T$ we assume that the single-particle levels $\varepsilon_{r\pm}$ [which are defined in Eq. (42) below] are degenerate and contain $\tilde{n}_{r+} + \tilde{n}_{r-}$ quasiparticles. The quasiparticle number densities are assumed given by the Fermi–Dirac distribution

$$\tilde{n}_\pm(T) = \frac{2}{\Omega}\sum_{r=\text{even}}\tilde{n}_{r\pm}(T) = \frac{2}{1+\exp(Re_\pm/(k_{\rm B}T))}, \tag{34}$$

where $e_\pm$ is the quasiparticle energy defined in Eq. (41) below and $R$ is an empirical scaling factor of order one that corrects on average for approximations that we shall make for the single-particle spectrum.

We may then evaluate the expectation value of one-body operators at finite temperature, with the result (see Appendix B of Ref. [50])

$$\begin{aligned}
\langle p^\dagger\rangle = \langle p\rangle &= -\frac{\Omega}{2}\left[P_+(T)u_+v_+ + P_-(T)u_-v_-\right], \\
\langle \pi_z^\dagger\rangle = \langle \pi_z\rangle &= -\frac{\Omega}{2}\left[P_+(T)u_+v_+ - P_-(T)u_-v_-\right], \\
\langle Q_z\rangle &= \frac{\Omega}{2}\left[P_+(T)v_+^2 - P_-(T)v_-^2\right], \\
\langle \hat{n}\rangle &= \frac{\Omega}{2}\left[P_+(T)(2v_+^2-1) + P_-(T)(2v_-^2-1) + 2\right], \\
\langle \pi_x\rangle = \langle \pi_y\rangle &= \langle \boldsymbol{S}\rangle = \langle Q_x\rangle = \langle Q_y\rangle = 0,
\end{aligned} \tag{35}$$

where in these expressions we have employed the definitions

$$P_\pm(T) \equiv 1 - \tilde{n}_\pm(T) = \tanh\left(\frac{Re_\pm}{2k_{\rm B}T}\right). \tag{36}$$

If a large-$\Omega$ approximation is invoked by ignoring terms of order $1/\Omega$ relative to the leading terms, the scalar products of these one-body operators reduce to products of the corresponding one-body ones,

$$\begin{aligned}
&\langle p^\dagger p\rangle = \langle p\rangle^2, \quad \langle \boldsymbol{\pi}^\dagger\cdot\boldsymbol{\pi}\rangle = \langle \pi_z\rangle^2, \\
&\langle \boldsymbol{Q}\cdot\boldsymbol{Q}\rangle = \langle Q_z\rangle^2, \quad \langle \boldsymbol{S}\cdot\boldsymbol{S}\rangle = \langle S\rangle^2 = 0.
\end{aligned} \tag{37}$$

If $T \to 0$, then $P_\pm(T) \to 1$ in Eq. (36) and Eqs. (35) and (37) reduce to the simpler forms

$$\langle p^\dagger \rangle = \langle p \rangle = -\frac{1}{2}\Omega(u_+v_+ + u_-v_-), \tag{38a}$$

$$\langle \pi_z^\dagger \rangle = \langle \pi_z \rangle = -\frac{1}{2}\Omega(u_+v_+ - u_-v_-), \tag{38b}$$

$$\langle Q_z \rangle = \frac{1}{2}\Omega(v_+^2 - v_-^2), \tag{38c}$$

$$\langle \hat{n} \rangle = \Omega(v_+^2 + v_-^2), \tag{38d}$$

$$\langle \pi_x \rangle = \langle \pi_y \rangle = \langle \boldsymbol{S} \rangle = \langle Q_x \rangle = \langle Q_y \rangle = 0, \tag{38e}$$

$$\langle p^\dagger p \rangle = \langle p \rangle^2 = \frac{1}{4}\Omega^2(u_+v_+ + u_-v_-)^2, \tag{38f}$$

$$\langle \boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} \rangle = \langle \pi_z \rangle^2 = \frac{1}{4}\Omega^2(u_+v_+ - u_-v_-)^2, \tag{38g}$$

$$\langle \boldsymbol{Q} \cdot \boldsymbol{Q} \rangle = \langle Q_z \rangle^2 = \frac{1}{4}\Omega^2(v_+^2 - v_-^2)^2, \tag{38h}$$

$$\langle \boldsymbol{S} \cdot \boldsymbol{S} \rangle = \langle S \rangle^2 = 0, \tag{38i}$$

$$\langle M^2 \rangle = \frac{1}{4}(n - \Omega)^2 + \frac{1}{2}\Omega[(u_+\nu_+)^2 + (u_-\nu_-)^2]. \tag{38j}$$

which are valid only for $T = 0$ (and large $\Omega$ for the 2-body terms).

### 6.5 Energy gaps and gap equations

The preceding results may be used to find the expectation value of the variational Hamiltonian $\langle H' \rangle$ in the coherent state representation. In terms of the energy gaps defined by

$$\Delta_d = G_0\sqrt{\langle p^\dagger p \rangle}, \tag{39a}$$

$$\Delta_\pi = G_1\sqrt{\langle \boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} \rangle}, \tag{39b}$$

$$\Delta_q = \chi\sqrt{\langle \boldsymbol{Q} \cdot \boldsymbol{Q} \rangle}, \tag{39c}$$

one obtains from Eqs. (29), (30), and (14),

$$\langle H' \rangle = (\varepsilon - \lambda)n - \left(\frac{\Delta_d^2}{G_0} + \frac{\Delta_\pi^2}{G_1} + \frac{\Delta_q^2}{\chi}\right),$$

where $\varepsilon$ is the single-particle energy. Variation of $\langle H' \rangle$ with respect to $u_\pm$ or $v_\pm$ yields

$$2u_\pm v_\pm(\varepsilon_\pm - \lambda) - \Delta_\pm(u_\pm^2 - v_\pm^2) = 0,$$

which is satisfied by

$$u_\pm^2 = \frac{1}{2}\left(1 + \frac{\varepsilon_\pm - \lambda}{e_\pm}\right), \quad v_\pm^2 = \frac{1}{2}\left(1 - \frac{\varepsilon_\pm - \lambda}{e_\pm}\right), \tag{40}$$

where

$$e_\pm = \sqrt{(\varepsilon_\pm - \lambda)^2 + {\Delta_\pm}^2} \tag{41}$$

and

$$\Delta_\pm = \Delta_d \pm \Delta_\pi, \quad \varepsilon_\pm = \varepsilon \mp \Delta_q. \tag{42}$$

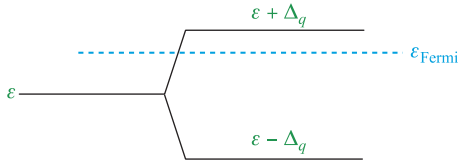



**Fig. 16** Modification of the quasiparticle energies by interaction with the antiferromagnetism in the SU(4) coherent-state gap equations.

Inserting Eq. (40) into Eqs. (35)–(37) and employing the gap definitions (39), one obtains the temperature-dependent gap equations

$$\Delta_d = \frac{G_0\Omega}{4}\left(w_+\Delta_+ + w_-\Delta_-\right), \tag{43a}$$

$$\Delta_\pi = \frac{G_1\Omega}{4}\left(w_+\Delta_+ - w_-\Delta_-\right), \tag{43b}$$

$$\frac{4\Delta_q}{\chi\Omega} = w_+(\Delta_q + \lambda') + w_-(\Delta_q - \lambda'), \tag{43c}$$

$$-2x = w_+(\Delta_q + \lambda') - w_-(\Delta_q - \lambda'), \tag{43d}$$

where we define

$$w_\pm \equiv \frac{P_\pm(T)}{e_\pm}, \quad \lambda' \equiv \lambda - \varepsilon. \tag{44}$$

Solution of the above algebraic equations yields all the gaps and the chemical potential $\lambda'$ (we shall discuss methods of solution in Sections 6.7 and 6.8 below). Then the total energy is

$$E = \langle H' \rangle + \lambda n = \varepsilon n - \left(\frac{\Delta_d^2}{G_0} + \frac{\Delta_\pi^2}{G_1} + \frac{\Delta_q^2}{\chi}\right).$$

To simplify the discussion, we shall ignore the single-particle energy in the above equation by setting $\varepsilon = 0$, since this term has been approximated as a state-independent constant and thus plays no role in the phase competition. The energy density $E/\Omega$ is then

$$\frac{E}{\Omega} = -\left(\frac{{\Delta_d}^2}{G_0\Omega} + \frac{{\Delta_\pi}^2}{G_1\Omega} + \frac{{\Delta_q}^2}{\chi\Omega}\right). \tag{45}$$

The three gaps $\Delta_d$, $\Delta_\pi$ and $\Delta_q$ in the above equations are defined in (39) and represent the characteristic energy scales of spin-singlet pairing correlations, spin-triplet pairing correlations, and antiferromagnetic correlations, respectively. Hence, these correlations determine the ground state energy. Once the gaps and the chemical potential $\lambda'$ are known from solution of the gap equations (43), the quasiparticle energies $e_\pm$ and the amplitudes $u_\pm$ and $v_\pm$ can all be determined through Eqs. (40)–(42), permitting other ground state properties to be calculated.

In the following sections we shall give analytical solutions for the gap equations (43), first for zero temperature and then for finite temperatures. As we shall see, a rich

phase structure emerges naturally in these solutions as a consequence of competition between the various energy scales.

### 6.6 Relationship to ordinary BCS and Néel theory

The preceding results are formally analogous to those of the BCS theory with $v_\pm^2$ the probability of single particle levels $\varepsilon_\pm$ being occupied, $\Delta_\pm$ the energy gaps, and $e_\pm$ the quasiparticle energies. The essential difference from normal BCS theory is that conventional pairing theories deal with one energy gap and one kind of quasiparticle; here we have two kinds of quasiparticles and several energy gaps, implying a large variety of new physics. As should be clear from the derivations leading to Eq. (45), this new physics arises as a natural consequence of a BCS-like theory in which the pairing may be unconventional, Coulomb repulsion plays a significant role, and antiferromagnetic correlations are allowed to enter on a footing equal with that of the pairing correlations.

In the formalism describing this more sophisticated pairing the quantities $e_\pm$ are energies for two kinds of quasiparticle excitation, corresponding to two sets of nondegenerate single particle energy spectra $\{\varepsilon_\pm\}$ separated by an energy $2\Delta_q$, as illustrated in Fig. 16. Each level can be occupied by only one electron of either up or down spin. The corresponding pairing gaps are $\Delta_\pm$, which are linear combinations of the two gaps $\Delta_d$ and $\Delta_\pi$. The probabilities for single-particle levels to be occupied or unoccupied are $v_\pm^2$ and $u_\pm^2$, respectively.

The resulting theory has the limits that we would expect for a symmetry that unites superconductivity and antiferromagnetism at a fundamental level. The gap equations reduce exactly to the gap equations of the ordinary BCS theory in the limit that the antiferromagnetic interactions can be neglected. Conversely, in the limit that the pairing interactions can be neglected we obtain the equations expected for a Néel antiferromagnet.

### 6.7 Solution of the gap equations at zero temperature

The three parameters, $\chi$, $G_0$, and $G_1$ in the algebraic equations (43) correspond to the three basic effective interactions in the SU(4) model:

1) The antiferromagnetic correlation, with strength $\chi$.
2) The spin-singlet pairing correlation, with strength $G_0$.
3) The spin-triplet pairing correlation, with strength $G_1$.

Experimental evidence suggests that these three interactions in cuprates are all attractive, and that their strengths are ordered

$$\chi > G_0 > G_1 > 0. \tag{46}$$

Solutions for the gap equations assuming this condition to be satisfied can be obtained as follows.

#### 6.7.1 The critical doping point

The gap solutions for Eqs. (43) at $T = 0$ can be written explicitly for two doping regimes separated by a special doping value given by

$$x_q = \sqrt{\frac{\chi - G_0}{\chi - G_1}}. \tag{47}$$

This value of $x_q$ is a *critical doping point* marking a quantum phase transition, because we shall find that at zero temperature the wavefunctions and physical properties of the two doping regions lying on either size of this point differ qualitatively. Specifically, one finds the following solutions (general derivations are given in Appendix C of Ref. [50]).

#### 6.7.2 The all gaps finite solution for $T = 0$

Equations (43) have a solution for all gaps nonzero that corresponds to:

$$\Delta_q = \frac{\chi\Omega}{2}\sqrt{(x_q^{-1} - x)(x_q - x)}, \tag{48a}$$

$$\Delta_d = \frac{G_0\Omega}{2}\sqrt{x(x_q^{-1} - x)}, \tag{48b}$$

$$\Delta_\pi = \frac{G_1\Omega}{2}\sqrt{x(x_q - x)}, \tag{48c}$$

$$\lambda' = -\frac{\chi\Omega}{2}x_q(1 - x_q x) - \frac{G_1\Omega}{2}x. \tag{48d}$$

This solution illustrates the central role of the critical doping point $x_q$ because *it exists only for the doping range* $x \leq x_q$. We shall find below that it corresponds to the ground state of the system at zero temperature if the doping lies in the range $0 < x \leq x_q$.

#### 6.7.3 The pure singlet-pairing solution for $T = 0$

Equations (43) also have a solution corresponding to $\Delta_q = \Delta_\pi = 0$ that is given by

$$\Delta_q = \Delta_\pi = 0, \tag{49a}$$

$$\Delta_0 \equiv \Delta_d = \frac{G_0\Omega}{2}\sqrt{1 - x^2}, \tag{49b}$$

$$\lambda' = -\frac{G_0\Omega}{2}x. \tag{49c}$$

This solution is valid for the *entire physical doping range* $0 \leq x \leq 1$, but we shall see later that it corresponds to the ground state at zero temperature only for $x_q < x < 1$.

#### 6.7.4 The pure triplet-pairing solution for $T = 0$

A solution to Eqs. (43) exists if all gaps except the triplet pairing vanish:

$$\Delta_q = \Delta_d = 0, \tag{50a}$$

$$\Delta_\pi = \frac{G_1\Omega}{2}\sqrt{1 - x^2}, \tag{50b}$$

$$\lambda' = -\frac{G_1\Omega}{2}x. \tag{50c}$$

This solution is valid for the entire range of physical doping.

#### 6.7.5 The pure AF solution for $T = 0$

A solution to Eqs. (43) that we shall term the *pure antiferromagnetic solution* exists if all gaps except the antiferromagnetic correlation vanish:

$$\Delta_d = \Delta_\pi = 0, \tag{51a}$$

$$\Delta_q = \frac{\chi\Omega}{2}(1-x), \tag{51b}$$

$$\lambda' = -\frac{\chi\Omega}{2}(1-x). \tag{51c}$$

This solution is valid for the entire range of physical doping.

#### 6.7.6 The uncorrelated solution for $T = 0$

If all gaps vanish, Eqs. (43) have a trivial solution that we shall term the *uncorrelated solution* in which all antiferromagnetic and pairing correlations are zero:

$$\Delta_q = \Delta_d = \Delta_\pi = 0, \tag{52a}$$

$$\lambda' = -2kT\ \mathrm{atanh}(x) = 0. \tag{52b}$$

This solution is valid for the entire range of physical doping.

#### 6.7.7 The zero-temperature ground state

Among the five sets of gap solutions for $T = 0$ presented above, we can determine which corresponds to the ground state at a given doping $x$ by inserting the gap solutions directly into Eq. (45) to ascertain which has the lowest energy. We shall discuss the resulting gap diagram in more detail below, but—assuming the ordering of coupling strengths given by Eq. (46)—the general result for $T = 0$ is that

- The all gaps finite solution is the ground state for the doping range $0 < x \leq x_q$.
- The pure singlet-pairing solution is the ground state for $x_q < x < 1$.
- The pure triplet pairing solution is never the ground state under any conditions.
- The uncorrelated solution is never the ground state if there are finite AF or pairing correlations.
- The AF correlation solution is never the ground state, except possibly for $x \sim 0$.

Thus, our primary interest at zero temperature will be in the all gaps finite and the pure singlet-pairing solutions for general values of $x$, and in the pure antiferromagnetic solution near $x = 0$.

### 6.8 Solution of the gap equations for finite temperature

The gap equations for $T > 0$ differ from those at $T = 0$ in that the terms $w_\pm$ depend on temperature. From Eqs. (44) and (36),

$$w_\pm = \frac{P_\pm(T)}{e_\pm} = \frac{\tanh(Re_\pm/(2k_\mathrm{B}T))}{e_\pm}. \tag{53}$$

For finite temperature the gap equations could have a variety of solutions, even for a fixed doping $x$. Some general solutions have been derived in Appendix C of Ref. [50] using the physically-motivated assumption (46) for the relative correlation strengths. By inserting these solutions in Eq. (45), one can then find the ones having the lowest energy and thus determine the physical ground-state solution for a given temperature $T$ and doping $x$. Several representative cases are examined below below.

#### 6.8.1 The $\Delta_d + \Delta_q + \Delta_\pi$ finite-temperature case

This case generalizes to finite $T$ the all-gap solution found at $T = 0$ for the doping range $0 \leq x \leq x_q$. The temperature and doping dependent gap solutions are derived in Appendix C of Ref. [50], with the results

$$\Delta_q = \frac{\chi\Omega}{2}\sqrt{(x_q^{-1}-y)(x_q-y)}\ \frac{x}{y}, \tag{54a}$$

$$\Delta_d = \frac{G_0\Omega}{2}\sqrt{x(x_q^{-1}-y)}\,g(y), \tag{54b}$$

$$\Delta_\pi = \frac{G_1\Omega}{2}\sqrt{x(x_q-y)}\,g(y), \tag{54c}$$

$$\lambda' = -\frac{(\chi-G_1)\Omega}{2}x_q\left(\frac{x}{y}-x_q x\right) - \frac{G_1\Omega}{2}x, \tag{54d}$$

where $y$ and $g(y)$ are defined through

$$y = \frac{x}{\sqrt{I_+(T)+I_-(T)\Gamma(y)}}, \tag{55a}$$

$$g(y) = \sqrt{\frac{x}{y}+\frac{I_+(T)-(x/y)^2}{2x(\bar{x}_q-y)}}, \tag{55b}$$

with

$$\Gamma(y) = \frac{\bar{x}_q-y}{\sqrt{(x_q-y)(x_q^{-1}-y)}}, \tag{56a}$$

$$I_\pm(T) = \frac{P_-^2(T)\pm P_+^2(T)}{2}, \tag{56b}$$

and $\bar{x}_q \equiv (x_q^{-1}+x_q)/2$. For a given value of doping $x$ and temperature $T$, the gaps and chemical potential may be found as follows. From Eqs. (53) and (41),

$$T = \frac{R\sqrt{(\Delta_q\pm\lambda')^2+{\Delta_\pm}^2}}{2k_\mathrm{B}\,\mathrm{atanh}(w_\pm e_\pm)}, \tag{57}$$

which implies that

$$\frac{\sqrt{(\Delta_q+\lambda')^2+{\Delta_+}^2}}{\mathrm{atanh}(w_+e_+)} = \frac{\sqrt{(\Delta_q-\lambda')^2+{\Delta_-}^2}}{\mathrm{atanh}(w_-e_-)}. \tag{58}$$

By solving Eqs. (43c) and (43d) directly, we obtain

$$w_{\pm} = \frac{[2\Delta_q/(\chi\Omega)] \mp x}{\Delta_q \pm \lambda'}.$$

Now from Eq. (C2) in Appendix C of Ref. [50], $\Delta_\pi$ and $\Delta_d$ are related by

$$\Delta_\pi = \frac{w_- - 2/(G_0\Omega)}{w_- - 2/(G_1\Omega)} \cdot \Delta_d, \tag{59}$$

and thus, utilizing the first of Eqs. (42), $\Delta_+$ and $\Delta_-$ can be related to $\Delta_d$ through

$$\Delta_\pm = \left[1 \pm \frac{w_- - 2/(G_0\Omega)}{w_- - 2/(G_1\Omega)}\right] \Delta_d. \tag{60}$$

Hence, for a given doping $x$, we can obtain a solution by the following algorithm:

1) Choose different values of $\Delta_q$. For each choice of $\Delta_q$ one can solve for $y$ from Eq. (54a), and then for $\lambda'$ from Eq. (54d).
2) The pairing gap $\Delta_d$ can then be obtained by substituting Eq. (60) into Eq. (58) and solving the resulting equation for $\Delta_d$.
3) With $\Delta_d$ determined, $\Delta_\pi$ can be found from Eq. (59) and the corresponding temperature $T$ follows from Eqs. (57) and (60).

By this procedure, we can find complete solutions for each choice of $x$ and $\Delta_q$.

#### 6.8.2 The $\Delta_d$ finite-temperature case

This solution generalizes the pure singlet-pairing solution at zero temperature. Following a similar procedure as in the $T = 0$ case, one obtains the temperature and doping dependent gap solutions

$$\Delta_q = \Delta_\pi = 0, \tag{61a}$$
$$\Delta_d = \frac{G_0\Omega}{2}\sqrt{I_+(T) - x^2}, \tag{61b}$$
$$\lambda' = -\frac{G_0\Omega}{2}x, \tag{61c}$$

where in the present case,

$$I_+(T) = P_+^2(T) = P_-^2(T).$$

By using Eqs. (41), (53), and (61c), Eq. (61b) may be expressed as

$$T = \frac{R\sqrt{\left(\frac{G_0\Omega}{2}x\right)^2 + {\Delta_d}^2}}{2k_{\mathrm{B}}\ \mathrm{atanh}\left[\sqrt{\left(\frac{2\Delta_d}{G_0\Omega}\right)^2 + x^2}\right]}. \tag{62}$$

Therefore, for given $x$ and $T$ the gap $\Delta_d$ can be obtained from Eq. (62), and $\lambda'$ follows from Eq. (61c).

#### 6.8.3 The $\Delta_q$ finite-temperature case

In this case both $\Delta_d$ and $\Delta_\pi$ are zero, corresponding to a solution that applies only for temperatures $T > T_c$:

$$\Delta_d = \Delta_\pi = 0, \tag{63a}$$
$$\Delta_q = \frac{\chi\Omega}{2}[P_-(T) - x]. \tag{63b}$$

These results can be derived from Eqs. (43c) and (43d), which in the present case reduce to

$$\frac{4\Delta_q}{\chi\Omega} = P_-(T) + P_+(T), \tag{64}$$
$$2x = P_-(T) - P_+(T). \tag{65}$$

By solving Eqs. (64) and (65) one obtains

$$T = \frac{R\Delta_q}{k_{\mathrm{B}}A_+}, \tag{66}$$
$$\lambda' = \frac{k_{\mathrm{B}}TA_-}{R}, \tag{67}$$

with

$$A_\pm \equiv \mathrm{atanh}\left(\frac{2\Delta_q}{\chi\Omega} - x\right) \pm \mathrm{atanh}\left(\frac{2\Delta_q}{\chi\Omega} + x\right).$$

Thus, for a temperature $T$, $\Delta_q$ can be obtained from Eq. (66) and $\lambda'$ can be obtained from Eq. (67).

### 6.9 Momentum-dependent SU(4) solutions

The SU(4) generators discussed to this point are summed over the momentum $\boldsymbol{k}$. They are appropriate for data that are not momentum-selected. However, some data, such as ARPES Fermi-arc measurements exhibit explicit dependence on $\boldsymbol{k}$.

#### 6.9.1 Momentum-dependent generators of the algebra

The SU(4) formalism has been extended [52] to deal with this case by viewing the individual $\boldsymbol{k}$ components of the operators defined in Eq. (1) as the symmetry generators $\mathcal{G}(\boldsymbol{k})$:

$$\{\mathcal{G}(\boldsymbol{k})\} \equiv \{p^\dagger(\boldsymbol{k}), p(\boldsymbol{k}), \boldsymbol{\pi}^\dagger(\boldsymbol{k}), \boldsymbol{\pi}(\boldsymbol{k}), \boldsymbol{Q}(\boldsymbol{k}), \boldsymbol{S}(\boldsymbol{k}), n_{\boldsymbol{k}}\} \tag{68}$$

where, for example, the singlet pair generator $p^\dagger$ in Eq. (1) is related to the $\boldsymbol{k}$-dependent generators $p^\dagger(\boldsymbol{k})$ by

$$p^\dagger = \sum_{\boldsymbol{k}>0} p^\dagger(\boldsymbol{k}),$$

and $\boldsymbol{k} > 0$ means either $k_x > 0$ or $k_y > 0$.

Instead of the global SU(4) symmetry generated by Eqs. (1), the symmetry now is a direct product of

$k$-dependent SU(4) groups, $\prod_{\boldsymbol{k}>0}\otimes\mathrm{SU}(4)_{\boldsymbol{k}}$. The corresponding Hamiltonian is

$$H=\sum_{\boldsymbol{k}>0}\epsilon_{\boldsymbol{k}}\,n_{\boldsymbol{k}}-\sum_{\boldsymbol{k},\boldsymbol{k}'>0}\{\chi_{\boldsymbol{k}\boldsymbol{k}'}\boldsymbol{Q}(\boldsymbol{k})\cdot\boldsymbol{Q}(\boldsymbol{k}') + G^{(0)}_{\boldsymbol{k}\boldsymbol{k}'}p^{\dagger}(\boldsymbol{k})\,p(\boldsymbol{k}')+G^{(1)}_{\boldsymbol{k}\boldsymbol{k}'}\boldsymbol{\pi}^{\dagger}(\boldsymbol{k})\cdot\boldsymbol{\pi}(\boldsymbol{k}')\},\tag{69}$$

where $\epsilon_{\boldsymbol{k}}$ and $n_{\boldsymbol{k}}$ are single-particle energies and occupation numbers, respectively, and the interaction strengths are

$$\chi_{\boldsymbol{k}\boldsymbol{k}'}=\chi^0|g(\boldsymbol{k})g(\boldsymbol{k}')|,\quad G^{(i)}_{\boldsymbol{k}\boldsymbol{k}'}=G^{(i)}|g(\boldsymbol{k})g(\boldsymbol{k}')|,\tag{70}$$

where the index $i$ takes the values 0 and 1.

#### 6.9.2 *Zero-temperature, momentum-dependent gap equations*

The formalism may now be developed in a manner parallel to that described in Refs. [46, 52]. For example, one obtains a set of $k$-dependent gap equations that generalize the BCS gap equations. To keep the discussion simple, we shall consider only the $T=0$ solutions here. Employing the critical doping point $x_q$ defined in Eq. (47) with the revised definitions

$$G_i\equiv G_i^0\bar{g}^2,\quad \chi\equiv\chi^0\bar{g}^2,$$

with $\bar{g}$ an averaged $g_k$, solutions for the gap equations at temperature $T=0$ and momentum $\boldsymbol{k}$ for doping $x\le x_q$ are found to be [54]

$$\Delta_{\mathrm{s}}(\boldsymbol{k})=\frac{\Omega}{2}G_0\frac{g(\boldsymbol{k})}{\bar{g}}\sqrt{x(x_q^{-1}-x)},\tag{71a}$$

$$\Delta_{\mathrm{t}}(\boldsymbol{k})=\frac{\Omega}{2}G_1\frac{g(\boldsymbol{k})}{\bar{g}}\sqrt{x(x_q-x)},\tag{71b}$$

$$\Delta_q(\boldsymbol{k})=\frac{\Omega}{2}\chi\frac{g(\boldsymbol{k})}{\bar{g}}\sqrt{x(x_q^{-1}-x)(x_q-x)},\tag{71c}$$

$$\lambda'_k=-\frac{\Omega}{2}\frac{g(\boldsymbol{k})}{\bar{g}}\left[(\chi-G_1)x_q(1-x_qx)+G_1x\right],\tag{71d}$$

and the corresponding solutions for $x>x_q$ are

$$\Delta_q(\boldsymbol{k})=\Delta_{\mathrm{t}}(\boldsymbol{k})=0,\tag{72a}$$

$$\Delta_{\mathrm{s}}(\boldsymbol{k})=\frac{\Omega}{2}G_0\frac{g(\boldsymbol{k})}{\bar{g}}\sqrt{1-x^2},\tag{72b}$$

$$\lambda'_k=-\frac{\Omega}{2}\frac{g(\boldsymbol{k})}{\bar{g}}G_0x.\tag{72c}$$

As before, $\Delta_{\mathrm{s}}$ and $\Delta_{\mathrm{t}}$ correspond to correlation energies for singlet and triplet pairing, respectively, $\Delta_q$ corresponds to correlation energy in the pseudogap state that is fluctuating AF in nature, and $\lambda'$ denotes the chemical potential. But now each of these quantities depends on the momentum.

These solutions may then be used to determine other physically important quantities using the methods described in Refs. [4, 46, 48–57]. For example, the superconducting transition temperature $T_{\mathrm{c}}$ is

$$T_{\mathrm{c}}(\boldsymbol{k})=G_0\frac{g(\boldsymbol{k})}{\bar{g}}\Omega\frac{Rx}{4k_{\mathrm{B}}\,\mathrm{atanh}(x)},\tag{73}$$

where the parameter $R$ is of order one and defined in Ref. [52], and the pseudogap temperature is

$$T^*(\boldsymbol{k})=\chi\frac{g(\boldsymbol{k})}{\bar{g}}\Omega\frac{R(1-x^2)}{4k_{\mathrm{B}}},\tag{74}$$

and is seen to depend on $k$, which we shall show below leads to a quantitative description of Fermi arcs.

Equations (71)–(74) define a $k$-dependent SU(4) model that can accommodate multiband physics. They are appropriate for comparison with experimental data that can resolve $\boldsymbol{k}$. If one averages these expressions over all momenta $\boldsymbol{k}$ near the Fermi surface, then the averaged factors $\langle g(\boldsymbol{k})/\bar{g}\rangle\to 1$ and Eqs. (71)–(74) reduce to the equations of the original SU(4) model. These are appropriate for comparison with experimental quantities that do not resolve $\boldsymbol{k}$.

## 7 Global implications of SU(4) symmetry

We shall discuss below in some detail that the SU(4) model permits various quantitative comparisons with data. However, because of the constraints implied by the non-abelian algebra and subalgebras, there are some important physical consequences of SU(4) symmetry that follow directly from the symmetry structure itself, without the need of significant calculation. We shall summarize some of those results in this section.

### 7.1 Physical conditions for closure of the SU(4) algebra

The preceding formalism is predicated on the operators of Eq. (1) closing an SU(4) Lie algebra, which requires that the pairing formfactors satisfy the conditions of Eq. (6). For purposes of illustration, let us discuss the closure condition for the specific case of cuprate superconductors, where there is essentially uniform agreement on the geometry of the pairing formfactor. (We shall address the corresponding situation for the iron superconductors in Section 14.)

#### 7.1.1 *Closure of the SU(4) algebra in momentum space*

For the cuprates, the pairing formfactor takes the $d$-wave form

$$g(k)=\cos k_x-\cos k_y.\tag{75}$$

However, $g(k)$ defined in this way does not satisfy the third closure criterion of Eq. (6). We may obtain closure

under commutation if we approximate the momentum-space formfactor by a step function,

$$g(k) = \cos k_x - \cos k_y \approx \mathrm{sgn}(\cos k_x - \cos k_y), \tag{76}$$

where sgn is the sign operator. The physics implied by this approximation becomes more transparent if we transform to coordinate space.

### 7.1.2 Closure of SU(4) on the real-space lattice

Using the *exact form* of Eq. (75), Eqs. (5) transformed to coordinate space take the form [49]

$$\begin{aligned} p_{12}^\dagger &= \sum_r c_{\boldsymbol{r}\uparrow}^\dagger c_{\bar{\boldsymbol{r}}\downarrow}^\dagger, & p_{12} &= \sum_r c_{\bar{\boldsymbol{r}}\downarrow} c_{\boldsymbol{r}\uparrow}, \\ q_{ij}^\dagger &= \sum_r (-)^r c_{\boldsymbol{r},i}^\dagger c_{\bar{\boldsymbol{r}},j}^\dagger, & q_{ij} &= \sum_r (-)^r c_{\bar{\boldsymbol{r}},j} c_{\boldsymbol{r},i}, \\ Q_{ij} &= \sum_r (-)^r c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j}, & S_{ij} &= \sum_r c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j} - \frac{1}{2}\Omega\delta_{ij}, \end{aligned} \tag{77}$$

where (see Fig. 17)

- $c_{\boldsymbol{r},i}^\dagger$ ($c_{\boldsymbol{r},i}$) creates (annihilates) an electron of spin $i$ located at $\boldsymbol{r}$, and
- $c_{\bar{\boldsymbol{r}},i}^\dagger$ ($c_{\bar{\boldsymbol{r}},i}$) creates (annihilates) an electron of spin $i$ at the four neighboring sites, $\boldsymbol{r} \pm \boldsymbol{a}$ and $\boldsymbol{r} \pm \boldsymbol{b}$,

with equal probabilities ($\boldsymbol{a}$ and $\boldsymbol{b}$ are lattice constants in the $\boldsymbol{x}$ and $\boldsymbol{y}$ directions, respectively),

$$c_{\bar{\boldsymbol{r}},i}^\dagger = \frac{1}{2}\left(c_{\boldsymbol{r}+\boldsymbol{a},i}^\dagger + c_{\boldsymbol{r}-\boldsymbol{a},i}^\dagger - c_{\boldsymbol{r}+\boldsymbol{b},i}^\dagger - c_{\boldsymbol{r}-\boldsymbol{b},i}^\dagger\right). \tag{78}$$

The factor $(-)^r$ in Eq. (77) is $(-)^{n_x+n_y}$ and $(n_x, n_y)$ are the coordinates of a lattice site on the copper oxide plane, $\boldsymbol{r} = n_x\boldsymbol{a} + n_y\boldsymbol{b}$, which is

- positive for even sites ($n_x + n_y$ = even),
- negative for the odd sites ($n_x + n_y$ = odd).

This factor originates from the assumption $\mathrm{e}^{\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{r}} \approx (-)^r$ and implies Mott insulator properties: the electrons are localized at lattice sites by strong Coulomb repulsion, with small overlap between orbitals of electrons on neighboring lattice sites.

From the coordinate representation (77) we see that spin-singlet and spin-triplet pairs are formed by holes on adjacent sites, as illustrated in Fig. 17. If one hole is located at $\boldsymbol{r}$, the other hole occupies the four adjacent sites ($\boldsymbol{r} \pm \boldsymbol{a}$ and $\boldsymbol{r} \pm \boldsymbol{b}$), each with equal probability. These pairs are highly coherent by virtue of the summation over $\boldsymbol{r}$ in the pair creation (annihilation) operators. It also can be seen that

$$\begin{aligned} \hat{n} &= \hat{n}^{(\mathrm{e})} + \hat{n}^{(\mathrm{o})}, \quad Q_+ = \hat{n}^{(\mathrm{e})} - \hat{n}^{(\mathrm{o})}, \\ S_{ij} &= S_{ij}^{(\mathrm{e})} + S_{ij}^{(\mathrm{o})}, \quad Q_{ij} = S_{ij}^{(\mathrm{e})} - S_{ij}^{(\mathrm{o})}, \end{aligned}$$

where $\hat{n}^{(\mathrm{e})}$ and $S_{ij}^{(\mathrm{e})}$ are the total electron number and spin operators at even sites, and $\hat{n}^{(\mathrm{o})}$ and $S_{ij}^{(\mathrm{o})}$ are the corresponding quantities at odd sites:

$$\begin{aligned} \hat{n}^{(\mathrm{e})} &= \sum_{i,r=\mathrm{even}} c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},i}, \quad \hat{n}^{(\mathrm{o})} = \sum_{i,r=\mathrm{odd}} c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},i}, \\ S_{ij}^{(\mathrm{e})} &= \sum_{r=\mathrm{even}} c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j}, \quad S_{ij}^{(\mathrm{o})} = \sum_{r=\mathrm{odd}} c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j}. \end{aligned}$$

Thus the collective operator $Q_+$ may be interpreted as the difference in total charge, and the collective operator $\boldsymbol{Q}$ may be interpreted as the difference in total spin, between even and odd sites on the lattice. From this we may conclude that in the U(4)⊃SU(4) algebra $Q_+$ will be associated with charge density waves and $\boldsymbol{Q}$ will be associated with antiferromagnetism.

### 7.1.3 SU(4) symmetry and double occupancy of sites

The real-space operator set that is defined in Eq. (77) will be closed under the operation of commutation for all operators in the set only if

$$\{c_{\bar{\boldsymbol{r}}',i}\ , c_{\bar{\boldsymbol{r}},j}^\dagger\} = \delta_{\boldsymbol{r}'r}\delta_{ij}, \quad \{c_{\bar{\boldsymbol{r}}',i}\ , c_{\bar{\boldsymbol{r}},j}\} = 0, \tag{79}$$

(that is, $c_{\bar{\boldsymbol{r}},i}^\dagger$ ($c_{\bar{\boldsymbol{r}},i}$) constitutes a basis for particles occupying sites adjacent to $\boldsymbol{r}$). This divides sites of the lattice into categories A and B, with

1) $r$ = even corresponding to the A sites, with operators $c_{\boldsymbol{r},i}^\dagger$ and $c_{\boldsymbol{r},i}$,
2) $r$ = odd corresponding to the B sites, with operators $c_{\bar{\boldsymbol{r}},i}^\dagger$ and $c_{\bar{\boldsymbol{r}},i}$

(or vice versa). Then Eq. (79) permits Eq. (77) to be written as

$$\begin{aligned} p_{12}^\dagger &= \sum_{\boldsymbol{r}\in A}\left(c_{\boldsymbol{r}\uparrow}^\dagger c_{\bar{\boldsymbol{r}}\downarrow}^\dagger \ - \ c_{\boldsymbol{r}\downarrow}^\dagger c_{\bar{\boldsymbol{r}}\uparrow}^\dagger\right), \\ q_{ij}^\dagger &= \pm\sum_{\boldsymbol{r}\in A}\left(c_{\boldsymbol{r},i}^\dagger c_{\bar{\boldsymbol{r}},j}^\dagger + c_{\boldsymbol{r},j}^\dagger c_{\bar{\boldsymbol{r}},i}^\dagger\right), \\ S_{ij} &= \sum_{\boldsymbol{r}\in A}\left(c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j} - c_{\bar{\boldsymbol{r}},j} c_{\bar{\boldsymbol{r}},i}^\dagger\right), \\ \tilde{Q}_{ij} &= \pm\sum_{\boldsymbol{r}\in A}\left(c_{\boldsymbol{r},i}^\dagger c_{\boldsymbol{r},j} + c_{\bar{\boldsymbol{r}},j} c_{\bar{\boldsymbol{r}},i}^\dagger\right), \\ p_{12} &= (p_{12}^\dagger)^\dagger, \quad q_{ij} = (q_{ij}^\dagger)^\dagger, \end{aligned} \tag{80}$$

with $\tilde{Q}_{ij} \equiv Q_{ij} + \frac{1}{2}\Omega\delta_{ij}$, where the sign is + (−) if A is chosen to be even (odd) sites. (Whether A sites are taken to be even or odd is a labeling choice and does not influence the physics.) Then by explicit commutation the operators (80) close an SU(4) algebra if Eq. (79) is satisfied. But by Eq. (78),

$$\{c_{\bar{\boldsymbol{r}}',i}\ , c_{\bar{\boldsymbol{r}},j}^\dagger\} = \delta_{\boldsymbol{r}'r}\delta_{ij} + \frac{1}{4}\delta_{ij}\sum_t g(\boldsymbol{t})\ \delta_{\boldsymbol{r}',r+t}, \tag{81a}$$

$$g(\boldsymbol{t}) = \begin{cases} +1 \text{ for } \boldsymbol{t} = \pm 2a, \pm 2b, \\ -1 \text{ for } \boldsymbol{t} = +a \pm b, -a \pm b, \end{cases} \tag{81b}$$

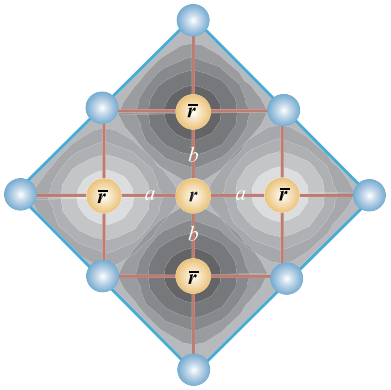



**Fig. 17** A schematic hole pair. The yellow-shaded balls are sites where electron holes form a pair: one hole at $\boldsymbol{r}$, the other with equal probability ($\frac{1}{4}$) at the four neighboring sites ($\bar{\boldsymbol{r}} = \boldsymbol{r} \pm \boldsymbol{a}$ and $\boldsymbol{r} \pm \boldsymbol{b}$). Balls on the outer boundary of the figure (connected by diagonal blue lines) are empty sites where the presence of a hole would imply double occupancy (see Fig. 21).

and Eq. (79) is generally *not* satisfied unless the second term on the right side of Eq. (81a) can be neglected. This term vanishes if we require that whenever there is a hole pair $c^\dagger_{\boldsymbol{r}i} c^\dagger_{\bar{\boldsymbol{r}}j}$ at $\boldsymbol{r}$ (see Fig. 17), no pair is permitted at $\boldsymbol{r}' = \boldsymbol{r} + t$, leaving nothing to be annihilated by $c_{\bar{\boldsymbol{r}}'i}$. This is equivalent to a *no-double-occupancy constraint,* because there is a finite amplitude for double occupation of lattice sites unless it is imposed. For example, if a pair is located at $\boldsymbol{r}' = \boldsymbol{r} + 2\boldsymbol{a}$ and a second pair is located at $\boldsymbol{r}$, there is a probability of 1/16 for two holes to be located at the site $\boldsymbol{r} + \boldsymbol{a}$ (see Fig. 17). We conclude that *closure of the SU(4) algebra is a direct consequence of no double occupancy* in the copper oxide conducting plane.

The validity of Eq. (79) actually follows from the more general condition that no pairs overlap spatially in the components of allowed configurations, which is a consistency requirement ensuring that the pair space and the pairing correlations be well defined. This no-pair-overlap constraint implies naturally that a pair centered at $\boldsymbol{r}$ precludes a pair being located at $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{t}$ with $\boldsymbol{t}$ given in Eq. (81b), and hence ensures that Eq. (79) holds.

#### *7.1.4 Pair formfactors and closure of the algebra*

In light of the preceding discussion, let us now revisit the momentum-space closure conditions for the SU(4) algebra given in Eq. (6). The first requirement $g(\boldsymbol{k}) = g(-\boldsymbol{k})$ is generally satisfied by physically-reasonable formfactors. As discussed in the preceding section and in Ref. [49], the condition $|g(\boldsymbol{k})| = 1$ necessary to close the algebra in momentum space may be interpreted as an occupation constraint on the *full formfactor without this condition* in the real space. Specifically, for cuprates the $d$-wave formfactor $g(\boldsymbol{k}) = \cos k_x - \cos k_y$ must be approximated by $\mathrm{sgn}(\cos k_x - \sin k_x)$ to close the algebra in momentum space. However, if the operators are Fourier transformed to the real space retaining the *full formfactor* $\cos k_x - \cos k_y$, the algebra closes, but only if the lattice is restricted to no double pair occupancy.

This suggests that $|g(\boldsymbol{k})| = 1$ is not an approximation but rather is a physically-necessary corollary in momentum space to no double occupancy (by pairs) for the collective wavefunction in the real space. Therefore, an alternative statement of the closure conditions (6) is the requirement of no double occupancy of the real-space lattice and

$$g(\boldsymbol{k} + \boldsymbol{Q}) = \pm g(\boldsymbol{k}), \quad g(\boldsymbol{k}) = g(-\boldsymbol{k}) \tag{82}$$

applied to the full momentum-space formfactor, *without* the condition $|g(\boldsymbol{k})| = 1$ of Eq. (6).

### 7.2 Reduction from SO(8) to SU(4) symmetry

If all possible bilinear particle–hole and particle–particle (pair) operators are taken as generators, the minimal closed algebra for an $N$-dimensional basis is SO($2N$) [71]. The simplest basis for cuprate superconductors may be regarded as 4-dimensional, since electrons can exist only in four basic states,

1) On A-sites with spin up.
2) On A-sites with spin down.
3) On B-sites with spin up.
4) On B-sites with spin down

Thus, in the absence of further constraints, the minimal Lie algebra for a set of generators that can describe high-$T_c$ superconductivity and antiferromagnetism simultaneously in a cuprate system is SO($2N$) = SO(8) and not its subgroup SU(4). The 28 generators of SO(8) are the 16 operators of Eq. (80), plus the 12 additional operators

$$\begin{aligned}
\bar{p}^\dagger_{12} &= \sum_{\boldsymbol{r}\in A} \left( c^\dagger_{\boldsymbol{r}\uparrow} c^\dagger_{\boldsymbol{r}\downarrow} - c^\dagger_{\bar{r}\downarrow} c^\dagger_{\bar{r}\uparrow} \right), \\
\bar{q}^\dagger_{12} &= \pm \sum_{\boldsymbol{r}\in A} \left( c^\dagger_{\boldsymbol{r}\uparrow} c^\dagger_{\boldsymbol{r}\downarrow} + c^\dagger_{\bar{r}\downarrow} c^\dagger_{\bar{r}\uparrow} \right), \\
\bar{S}_{ij} &= \sum_{\boldsymbol{r}\in A} \left( c^\dagger_{\boldsymbol{r},i} c_{\bar{r},j} - c_{\boldsymbol{r},j} c^\dagger_{\bar{r},i} \right), \\
\bar{Q}_{ij} &= \pm \sum_{\boldsymbol{r}\in A} \left( c^\dagger_{\boldsymbol{r},i} c_{\bar{r},j} + c_{\boldsymbol{r},j} c^\dagger_{\bar{r},i} \right), \\
\bar{p}_{12} &= (\bar{p}^\dagger_{12})^\dagger, \quad \bar{q}_{12} = (\bar{q}^\dagger_{12})^\dagger,
\end{aligned} \tag{83}$$

where the $\pm$ signs depend on the even–odd choice for A-sites; see Eq. (80). Equation (83) contains two new kinds of spin-singlet pairs created by $\bar{p}^\dagger_{12}$ and $\bar{q}^\dagger_{12}$, which we shall term $S$ and $S^*$ pairs, respectively. Unlike the bond-wise pairs associated with the SU(4) subalgebra within the SO(8) algebra, these new pairs are *onsite*, where the two electrons (or two holes) occupy the same site, with equal probability to appear anywhere in the lattice coherently. Figure 18 illustrates. The $S^*$ and $S$ pairs differ from each

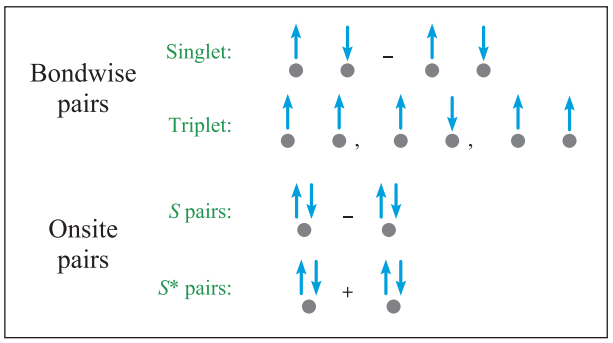



**Fig. 18** Schematic difference between bondwise and onsite airs in the SO(8) symmetry.

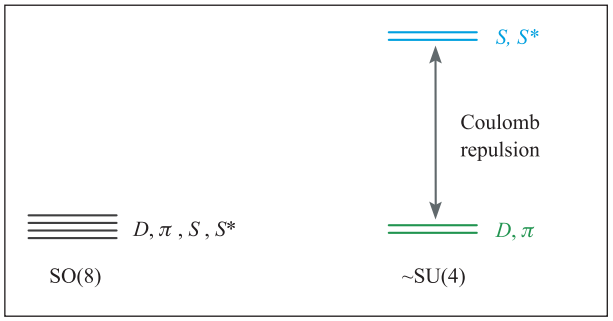



**Fig. 19** Separation of the onsite pair $(D, \pi)$ and bondwise pair $(S, S^*)$ energy scales by onsite Coulomb repulsion, which has the effect of reducing SO(8) symmetry to an effective SU(4) low-energy symmetry.

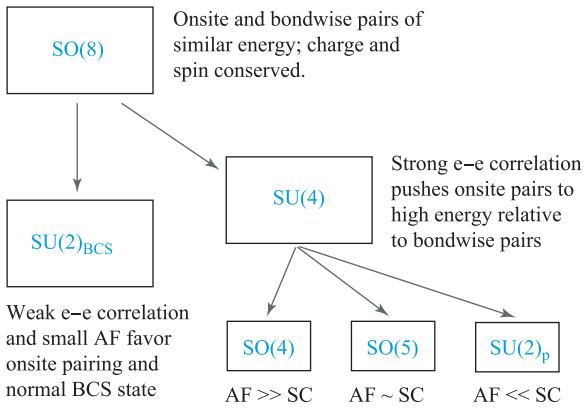



**Fig. 20** The general relationship among SU(4), SO(8), and the pseudospin SU(2) symmetry of a conventional BCS superconductor. Both SU(2) subgroups have pseudospin pair generators, but differ in the pairs being onsite for $SU(2)_{BCS}$ and bondwise for $SU(2)_p$, implying different orbital formfactors for pairing in the two cases. The symmetry $SU(2)_{BCS}$ with conventional formfactor is favored only if onsite Coulomb repulsion and any collective modes like AF competing with pairing can be neglected.

other only in phases. The operators $\bar{S}_{ij}$ are the hopping operators with and without spin flip, and $\bar{Q}_{ij}$ is the staggering of the hopping. These operators interchange singlet and triplet pairs with $S$ and $S^*$ pairs.

The SO(8) algebra reduces to the subalgebra SU(4) if the $S$ and $S^*$ pairs may be neglected. This will be a reasonable approximation if we assume onsite Coulomb repulsion pushing the $S$ and $S^*$ pairs to sufficiently high energy, as illustrated in Fig. 19. Thus, restriction to no double occupancy effectively allows the operators in Eq. (83) to be ignored and reduces SO(8) to its subalgebra SU(4). We conclude that the minimal Lie algebra that can describe antiferromagnetism and $d$-wave superconductivity in a cuprate system is in general SO(8), but under the constraint of no double occupancy the symmetry effectively reduces to SU(4). The assumption of an SU(4) symmetry in a cuprate system automatically implies the imposition of a *no-double-occupancy constraint* on the more general SO(8) symmetry in the copper–oxygen planes.

The general relationship between SO(8) and SU(4) symmetry, and the corresponding relationship of SO(8) to the symmetry of a conventional BCS superconductor, are illustrated in Fig. 20. If physically we assume weak enough electron–electron correlations and neglect antiferromagnetism, the SO(8) states favor a dynamical symmetry

$$\mathrm{SO}(8) \supset \ldots \supset \mathrm{SU}(2)_{\mathrm{BCS}} \tag{84}$$

that indicates a spontaneously-broken symmetry selecting the conventional superconductor direction in the SO(8) space (the ellipses in the subgroup chain denote possible intervening subgroups). The final SU(2) subgroup in this chain is that of normal pseudospin for onsite pairs (which are compatible with binding by the lattice phonon interaction characteristic of conventional BCS superconductivity).

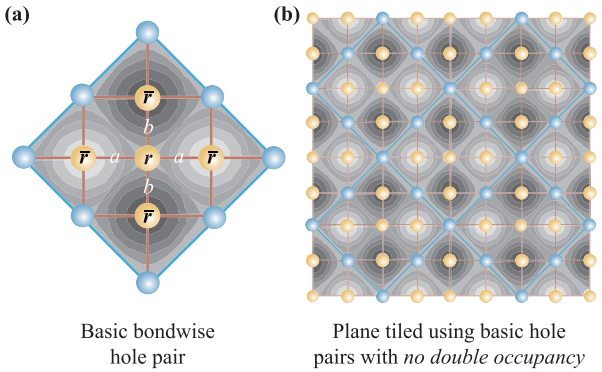



**Fig. 21** **(a)** The schematic hole pair of Fig. 17. **(b)** Tiling of the plane by the hole pairs of (a). Each diamond outlined by dashed boundaries corresponds to one unit pair from (a). Lighter-colored balls connected by solid horizontal and vertical lines are sites where the electron holes form pairs. Darker balls connected by diagonal dashed lines indicate sites where the presence of a hole would imply average double occupancy of some sites on the lattice, which would break SU(4) symmetry. By counting, the lattice can ensure no double occupancy [and thus preserves SU(4) symmetry] only if it is not more than $\frac{1}{4}$ occupied by electron holes.

### 7.3 SU(4) symmetry and an upper doping limit for the superconducting state

The implicit SU(4) occupancy constraint discussed in the preceding section dictates an upper limit for the doping fraction in SU(4)-conserving states [49]. This is illustrated in Fig. 21, which shows the spatial distribution of a representative configuration when the hole-pair number is maximal. By counting, the maximum number of holes consistent with SU(4) symmetry is $\Omega = \frac{1}{4}\Omega_e$, where $\Omega_e$ is the total number of lattice sites. Thus the largest doping fraction preserving SU(4) symmetry is $P_{\rm f} = \Omega/\Omega_e = \frac{1}{4}$. The maximum hole-doping fraction (0.23–0.27) that is seen experimentally for cuprate superconductivity may then be interpreted as a direct consequence of physical constraints on the realization of SU(4) symmetry.

### 7.4 The antiferromagnetic–superconducting transition

The SO(5) subgroup corresponds to a critical dynamical symmetry interpolating between AF and SC order for a range of intermediate doping parameters [48]. We now show that the emergent SU(4) symmetry implies a differing dependence on doping for SC and AF order, and that the dynamical symmetry structure itself controls the transition between the superconducting SU(2) symmetry and the antiferromagnetically ordered SO(4) symmetry. Thus, we shall show that the SU(4) symmetry has a natural propensity to favor antiferromagnetic Mott order at half-filling and singlet-pair superconductivity as the system is doped away from half-filling, for a broad range of Hamiltonian parameters.

#### *7.4.1 Competing antiferromagnetism and superconductivity*

We drop the common dependence of both phases on the spin and charge generators and consider the competition between SO(4) stabilization energy arising from $\boldsymbol{Q}\cdot\boldsymbol{Q}$ and SU(2) stabilization energy associated with the term $p^\dagger p$ in the Hamiltonian (14). These differ in their dependence on particle number and thus on doping. If we evaluate the AF correlation energy in the AF limit and the pairing correlation energy in the SC limit, we obtain as a function of doping $x$

$$\langle \chi \boldsymbol{Q}\cdot\boldsymbol{Q}\rangle = \frac{1}{4}\chi\Omega^2(1-x)^2, \tag{85a}$$

$$\langle G_0 p^\dagger p\rangle = \frac{1}{4}G_0\Omega^2(1-x^2), \tag{85b}$$

so that their ratio is

$$\frac{\langle G_0 p^\dagger p\rangle}{\langle \chi \boldsymbol{Q}\cdot\boldsymbol{Q}\rangle} = \frac{G_0}{\chi}\frac{(1+x)}{(1-x)}. \tag{86}$$

The competition between AF and pairing correlation energy is illustrated in Fig. 22. At half filling ($x = 0$), the Hamiltonian exhibits effective SO(4) symmetry if $\chi > G_0$

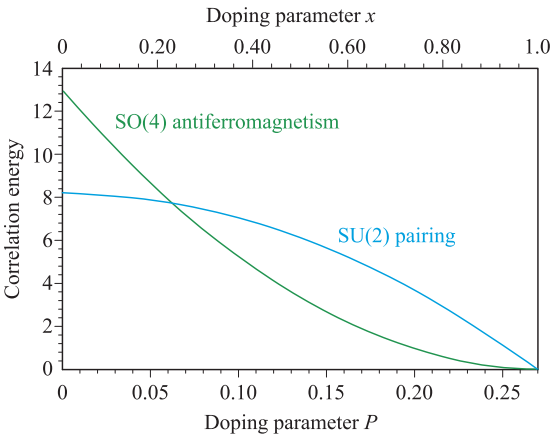



**Fig. 22** Competition between antiferromagnetic and pairing correlation energy as a function of doping assuming the coupling strength parameters used in Fig. 26.

because the SO(4) correlation energy (85a) is dominant. With increasing hole-doping $x$, the SO(4) correlation energy decreases more rapidly than the singlet pairing correlation energy (85b). Thus, the pairing correlation eventually dominates and the Hamiltonian exhibits effective SU(2) symmetry. This argument concerning the SC–AF competition is only qualitative, particularly in the vicinity of the crossing curves in Fig. 22, since the expressions (85) are evaluated in the respective symmetry limits, not in the actual physical ground state. We shall deal more quantitatively with the antiferromagnetic to superconducting transition in later sections but these qualitative arguments are useful to illustrate the general trend of the AF–SC competition with doping.

These features imply immediately that if $\chi/G_0 > 1$, antiferromagnetism tends to dominate at half-filling but pairing tends to dominate as holes are doped into the system. Thus, the AF ground state at half filling and the tendency to superconductivity as the system is doped away from half-filling follow directly from the dynamical symmetry structure of the Hamiltonian, independent of detailed parameter choices, and independent of detailed underlying microscopic physics [as long as it is consistent with emergent SU(4) symmetry and the condition $\chi/G_0 > 1$]. That is a significant conclusion but the implications of the SU(4) symmetry are even more dramatic than that. As we shall show in Section 11.1, because of the SU(4) dynamical symmetry structure the AF insulating state at half filling hides within it a superconductor that can emerge spontaneously with infinitesimal doping if there is a finite singlet pairing interaction.

#### *7.4.2 Analogies in nuclear structure physics*

The competition between antiferromagnetism and superconductivity has many parallels with the competition between spherical and deformed structure for nuclei that is a central paradigm of nuclear structure physics. The transi-

tion from spherical nuclei, which dominate the beginnings and endings of neutron and proton shells, to deformed nuclei, which tend to dominate the middle of shells, is governed by a microscopic competition between long-range quadrupole–quadrupole interactions favoring deformation and short-range monopole pairing interactions that favor spherical vibrational structure. This competition may be expressed algebraically as a competition between a dynamical symmetry that favors pairing (particle–particle) interactions and a dynamical symmetry that favors multipole (particle–hole) interactions [18].

The essential physics of the spherical–deformed transition in nuclear structure is determined by the dependence of the dynamical symmetries on particle number: nuclear pairing energy increases linearly with particle number from closed shells but the quadrupole deformation energy is approximately quadratic in particle number. Thus, the dynamical symmetry structure implies that spherical vibrational nuclei (which are favored by pairing energy) dominate the beginning and ends of shells and deformed nuclei (which are favored by the deformation energy) dominate the middle of shells [72].

This behavior is a close analog of the competition between AF dominating the half-filled lattice and SC dominating the hole-doped lattice that was discussed above, suggesting that these problems from different fields of physics may have a common dynamical symmetry basis. Although they involve different constituent particles, different basic interactions, and fundamentally different energy and length scales, they share the common general features of being strongly-correlated quantum systems of Fermi–Dirac particles with essential interactions in both particle–hole and particle–particle channels. The work presented here and in Refs. [18, 73, 74] suggests a unified picture of strongly correlated condensed matter and nuclear structure physics in terms of dynamical symmetries of their respective Hamiltonians, despite the obvious differences between the ingredients making up those Hamiltonians and the fundamentally different energy and length scales.

### *7.4.3 Analogies with graphene quantum-Hall physics*

Although we will not discuss it in any detail, the present methodology has been applied successfully to describing collective states for monolayer graphene in a strong magnetic field [75–77]. Remarkably, these graphene states also exhibit a dynamical symmetry with a Lie algebra similar to that described here, and the total energy surfaces for those graphene states were found to be almost identical to those found for high-temperature superconductors and nuclear structure physics, even though the underlying microscopic physics could hardly be more different among these cases. This suggests a universality of emergent behavior through Lie algebras that even goes beyond the universality of superconductivity and superfluidity proposed here [73, 74].

## 8 Ground-state energy surfaces

Let us now turn to more quantitative applications, beginning with ground-state energy surfaces. The total energy of the SU(4) state may be found in coherent state approximation by evaluating the expectation value of the Hamiltonian (14). There are only two independent variational parameters in the coherent state variational equations because of the unitarity condition (31). They may be chosen as either $\nu_+$ and $\nu_-$, or as $\alpha$ and $\beta$, using the definitions

$$\nu_+ \equiv \alpha + \beta, \quad \nu_- \equiv \alpha - \beta, \tag{87}$$

However, from Eq. (38d) the squares of $\nu_\pm$ (or of $\alpha$ and $\beta$) are constrained by

$$n = \langle \hat{n} \rangle = \Omega(\nu_+^2 + \nu_-^2) = 2\Omega(\alpha^2 + \beta^2). \tag{88}$$

Thus, for a fixed particle number $n$ we may evaluate matrix elements with only a single variational parameter, say $\beta$, which may in turn be related to standard order parameters by comparing matrix elements. For example, the $z$ component of the staggered magnetization is related to $\beta$ and $v_\pm$ by

$$\begin{aligned} Q \equiv \langle Q_z \rangle &= \frac{1}{2}\Omega(\nu_+^2 - \nu_-^2) \\ &= 2\Omega\beta[n/(2\Omega) - \beta^2]^{1/2}, \end{aligned} \tag{89}$$

and these measures of AF order are in turn related to the superconducting order parameter $\alpha$ through Eq. (88). From Eqs. (89) and (88), the ranges of $\beta$ and $\alpha$ are

$$0 \le \beta \le \sqrt{n/(4\Omega)}, \quad \sqrt{n/(4\Omega)} \le \alpha \le \sqrt{n/(2\Omega)}. \tag{90}$$

From Eqs. (88) and (89), one can show that

$$\nu_\pm^2 = \frac{n}{2\Omega} \pm \frac{Q}{\Omega}. \tag{91}$$

Equations (38a) and (38b) can be written as

$$\Delta \equiv \langle p^\dagger \rangle = \langle p \rangle = \sqrt{p^\dagger p} = \Delta_+ + \Delta_-, \tag{92a}$$

$$\Pi \equiv \langle \pi_z^\dagger \rangle = \langle \pi_z \rangle = \sqrt{\boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi}} = \Delta_+ - \Delta_-, \tag{92b}$$

where we define

$$\Delta_\pm \equiv \frac{\Omega}{2}\sqrt{\frac{1}{4} - \left(\frac{Q}{\Omega} \mp \frac{x}{2}\right)^2}, \tag{93}$$

and $x$ is the effective hole concentration that was introduced in Eq. (16). The quantities $\Delta$ and $\Pi$ present the spin-singlet and spin-triplet pairing correlations, respectively. The former is proportional to the singlet pairing gap and thus is directly related to the superconducting order. The latter is a measure of collectivity for triplet pairing, but it also measures the SO(5) correlation since from Table 1 of Appendix A,

$$C_{\mathrm{SO}(5)} = \boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} + \boldsymbol{S} \cdot \boldsymbol{S} + M(M-3) \tag{94}$$

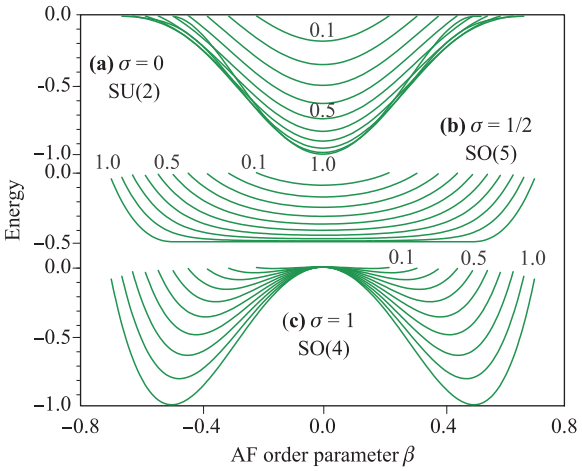



**Fig. 23** SU(4) Coherent-state energy surfaces for the three symmetry limits of the SU(4) Hamiltonian. The energy unit is $\frac{1}{4}\tilde{G}_0\Omega^2$ (see Eq. (96)). $H_0$ is taken as the energy zero point. Numbers on curves are the lattice occupation fractions $n/\Omega = 1 - x$, with $n/\Omega = 1$ corresponding to half filling and $0 < n/\Omega < 1$ to finite hole doping. The antiferromagnetic SO(4) symmetry corresponds to $\sigma = 1$, the critical SO(5) symmetry corresponds to $\sigma = \frac{1}{2}$, and the superconducting SU(2) symmetry corresponds to $\sigma = 0$ for the Hamiltonian (14). The allowed range of $\beta$ is $[-\frac{1}{2}(n/\Omega)^{1/2}, \frac{1}{2}(n/\Omega)^{1/2}]$, which depends on $n$. The order parameter $\beta$ is related to the order parameter $Q \equiv \langle Q_z \rangle$ (staggered magnetization) and the electron number $n$ through Eq. (89).

is the SO(5) Casimir operator, which from Eq. (92b) is proportional to $\Pi^2$ for fixed charge and spin.

Using Eqs. (38) and (87)–(89), one can then evaluate the energy surface as a function of the order parameters. For example, if we assume that $\langle \boldsymbol{S} \cdot \boldsymbol{S} \rangle = 0$, Eqs. (38f)–(38h) may be used to obtain a general expression for the SU(4) energy surface that takes the form [48]

$$E(Q) = \langle H \rangle - H_0 = -\tilde{G}_0 \left[ (1-\sigma)\Delta^2 + \sigma Q^2 \right], \tag{95}$$

in the limit $\Omega \to \infty$, where $\sigma$ is a parameter varying between 0 and 1. Equation (89) may be used to convert this to an energy surface as a function of the alternative AF order parameter $\beta$ at fixed particle number $n$,

$$E(\beta) = \langle H \rangle - H_0 = -\frac{\tilde{G}_0 \Omega^2}{4} \times \left\{ (24\sigma - 8)\beta^2 \left( \frac{n}{2\Omega} - \beta^2 \right) + 2(1-\sigma) \left[ \frac{n}{2\Omega} \left( 1 - \frac{n}{2\Omega} \right) + \left( \frac{n}{2\Omega} - 2\beta^2 \right) \sqrt{\left( 1 - \frac{n}{2\Omega} \right)^2 - 4\beta^2 \left( \frac{n}{2\Omega} - \beta^2 \right)} \right] \right\} \tag{96}$$

and this also may be expressed in terms of the superconducting order parameter $\alpha$ using Eq. (88).

The SU(4) coherent state solutions are valid for arbitrary ratios of the pairing and antiferromagnetic coupling strengths, so the corresponding ground-state total energy surfaces can be computed for arbitrary coupling strength ratio. However, it is instructive to first examine the energy surfaces in the dynamical symmetry limits of Section 5. Figures 23 and 24 illustrate.

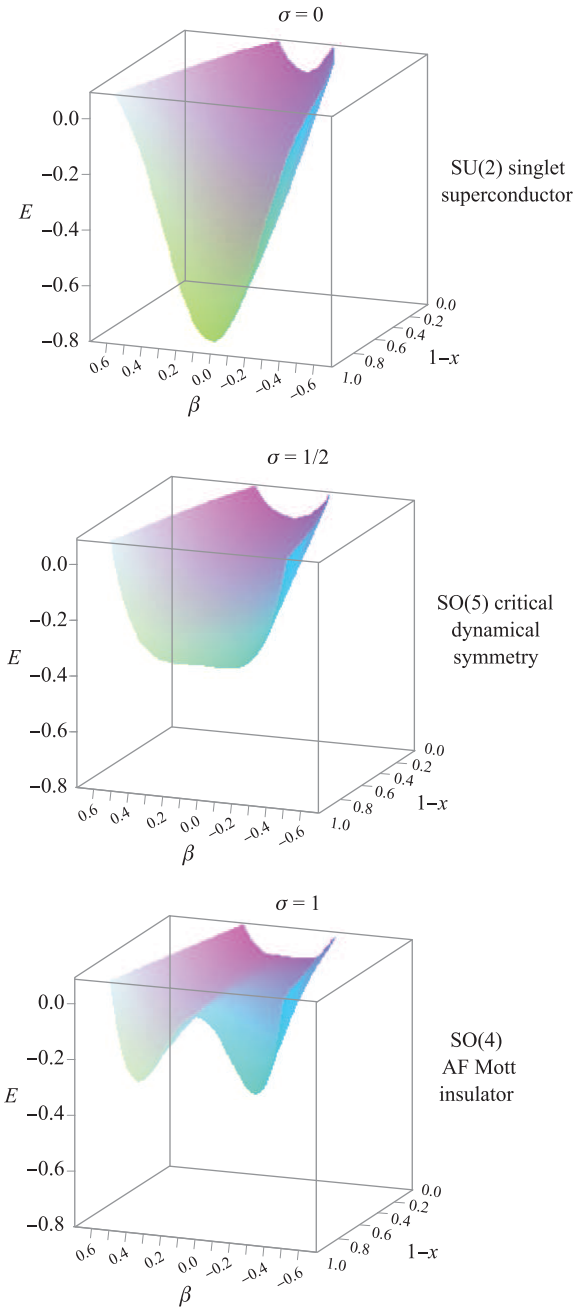



**Fig. 24** SU(4) energy surfaces as a function of the AF order parameter $\beta$ and the doping $x$.

## 8.1 Energy surfaces in the SO(4) limit

The identification of the SO(4) limit in Section 5.1 as an AF state is strengthened by examining the energy surfaces illustrated in Fig. 23. For $\sigma = 1$ [SO(4) limit; see Fig. 23(c)], $\beta = 0$ is an unstable point and an infinitesi-

mal fluctuation will drive the system to the energy minima at finite $\beta = \pm\frac{1}{2}(n/\Omega)^{1/2}$. Thus, $|Q|$ reaches its maximum value of $n/2$, indicating a spontaneously-broken symmetry and a state having AF order.

### 8.2 Energy surfaces in the SU(2) limit

Further insight into the SU(2) limit discussed in Section 5.2 follows by examining the ground-state energy surfaces illustrated in Fig. 23. For $\sigma = 0$ [SU(2) limit; see Fig. 23(a)], the minimum energy occurs at $\beta = 0$ (equivalently, $Q = 0$) for all values of $n$. Thus, $\Delta$ reaches its maximum value of

$$\Delta_{\max} = \frac{1}{2}\Omega\sqrt{1-x^2}, \tag{97}$$

indicating a state having superconducting order but vanishing antiferromagnetic order.

### 8.3 Energy surfaces in the SO(5) limit

The unusual nature of the SO(5) critical dynamical symmetry discussed in Section 5.3 is brought into sharp focus by examination of the ground-state energy surfaces that are illustrated in Fig. 23. From Fig. 23(b), the SO(5) dynamical symmetry is seen to have extremely interesting behavior: the minimum energy occurs at $\beta = 0$ for all values of $n$, as in the SU(2) case, but there are large-amplitude fluctuations in antiferromagnetic and superconducting order. In particular, when $n$ is near $\Omega$ (near half filling), the system has an energy surface almost flat for broad ranges of $\beta$ (or $Q$ or $\alpha$). This suggests a phase very soft against fluctuations in the AF and SC order parameters. However, as $n/\Omega$ decreases, fluctuations become smaller and the energy surface tends to the SU(2) (superconducting) limit.

### 8.4 Critical dynamical symmetries

Dynamical symmetries that interpolate between other dynamical symmetries are termed *critical dynamical symmetries* [64]. The SU(4) $\supset$ SO(5) symmetry exhibits such transitional properties. At half filling the energy surface is completely flat under variations of the antiferromagnetic order parameter $\beta$ [see the $n = 1.0$ curve of Fig. 23(b)], implying large fluctuations in the order parameters. But as hole doping is increased the SO(5) energy surface changes smoothly into one localized around $\beta = 0$ [see the $n = 0.1$ curve of Fig. 23(b)].

Under an exact SU(4) $\supset$ SO(5) symmetry the antiferromagnetic and superconducting states are degenerate at half filling, there is no barrier between AF and SC states, and one can fluctuate into the other at zero cost in energy [see the $n/\Omega = 1$ curve of Fig. 23(b)]. To see in more detail how in the SU(4) model an SO(5) symmetry can interpolate between AF and SC states as particle number varies, let us examine a case that is perturbed slightly away from the SO(5) limit of $\sigma = \frac{1}{2}$ in Eq. (14).

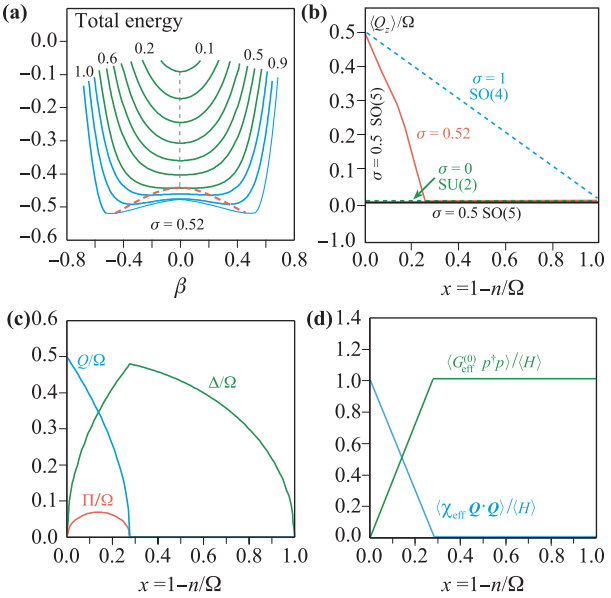



**Fig. 25** **(a)** As for Fig. 23, but for slightly perturbed SO(5) corresponding to $\sigma = 0.52$ [48]. The red dotted line indicates the location of the ground state in $\beta$ as $n$ varies. **(b)** Variation of the AF order parameter with effective occupation number for different values of $\sigma$. **(c)** Variation of the AF, SC and SO(5) correlations $Q$, $\Delta$, and $\Pi$ as functions of the effective hole concentration $x$. **(d)** Variation of the ratio of pairing and $\boldsymbol{Q}\cdot\boldsymbol{Q}$ interactions to the total energy of the system as functions of the effective hole concentration $x$.

### 8.5 Weakly-broken SO(5) symmetry

In Fig. 25(a), SU(4) coherent state energy surfaces as a function of the AF correlation $\beta$ are shown for the case $\sigma = 0.52$, which corresponds to SO(5) symmetry very weakly perturbed in the AF direction [48]. Numbers on curves are the lattice occupation fractions, with $n/\Omega = 1$ corresponding to half filling and $0 < n/\Omega < 1$ to finite hole doping. The corresponding variation of the AF correlation parameter $Q = \langle Q_z\rangle$ with hole doping $x$, and its comparison with the variation in various symmetry limits, are summarized in Fig. 25(b). The variations of the AF, SC and SO(5) correlations ($Q$, $\Delta$ and $\Pi$, respectively) with the hole doping $x$ are shown in Fig. 25(c), while the variations of the contributions of each term in the Hamiltonian to the total energy are shown in Fig. 25(d). From these results we see quite clearly the rapid evolution of the weakly-broken SO(5) symmetry from an energy surface that looks AF-like to one that looks SC-like, as the hole doping is increased from zero.

Thus, SO(5) is a critical dynamical symmetry that interpolates continuously between the SO(4) antiferromagnetic dynamical symmetry and the SU(2) superconducting dynamical symmetry. Such symmetries are well known in nuclear structure physics [61, 63–65] and the SO(5) critical dynamical symmetry discussed here in a condensed mat-

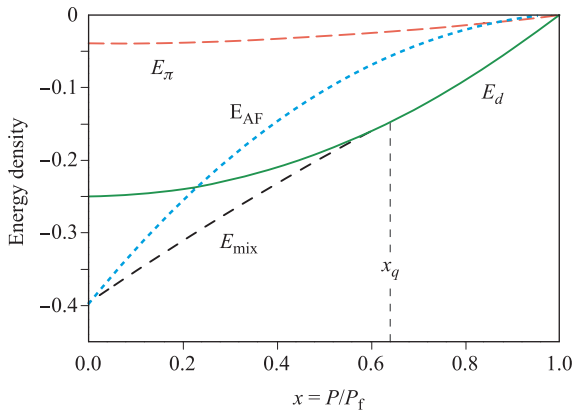



**Fig. 26** Total energy associated with gap solutions at $T = 0$. $E_{\rm mix}$ is the energy calculated with the all-gap solution (48), while $E_d$ [calculated from the solution in Eqs. (49)], $E_\pi$ [Eqs. (50)], and $E_{\rm AF}$ [Eqs. (51)] represent, respectively, the energy density of the spin-singlet pairing, the spin-triplet pairing, and the AF solutions. The energy of the uncorrelated solution is set to zero and taken as the energy reference. Interaction strengths are assumed constant, with $\chi = 13$, $G_0 = 8.2$, and $G_1 = 1.3$ (in an arbitrary energy unit), but reasonable values that satisfy the condition (46) will give similar results.

ter context has many formal similarities with critical dynamical symmetries of the (nuclear) Fermion Dynamical Symmetry Model [18]. We remark in passing that application of dynamical symmetry methods to quantum Hall states for monolayer graphene in a magnetic field exhibits also such critical dynamical symmetries [75–77]. These examples suggest that critical dynamical symmetries may be a fundamental organizing principle for strongly correlated fermionic systems in which an emergent order is found adjacent to another form of emergent order in the phase diagram.

## 9 SU(4) energy gaps

Let us now examine the correlation energies associated with SU(4) symmetry in more detail. We begin by analyzing the expected energy-gap structure as a function of doping, using parameters characteristic of the cuprate superconductors. The solutions of the SU(4) energy gap equations discussed in Section 6.7 imply a rich physics as a function of doping. Among the five sets of $T = 0$ gap solutions [Eqs. (48)–(52)], the one with the lowest energy at each doping corresponds to the physical ground state. We can calculate these energies by inserting the gap solutions directly into Eq. (45), and then investigate how these different sets of solutions compete with each other at $T = 0$. Some results are illustrated in Fig. 26.

### 9.1 Energy-ordering of gaps

In Fig. 26, we see that the all-gap solution $E_{\rm mix}$ is always lowest in energy and thus is the physical ground state for the doping range $x \leq x_q$. For $x > x_q$, the pure singlet pairing state becomes the ground state because $E_{\rm d}$ is the lowest energy for this doping range. All the other possible solutions lie higher in energy. They may be regarded as collective excited states but they cannot become the physical ground state at $T = 0$.

We shall find that for $T > 0$ the AF or the uncorrelated state could become the ground state in certain temperature and doping ranges. However, this can never happen for the pure spin-triplet pairing state, as long as $G_1$ is the weakest of the three coupling parameters in Eq. (46). Although spin-triplet pairing plays an important role as a component of the wavefunction in the SU(4) theory, the pure spin-triplet state is never found to be the ground state if the coupling-strength hierarchy of Eq. (46) is satisfied.

### 9.2 Generic features of SU(4) gaps

A generic gap diagram at $T = 0$ describing features of the energy gaps as functions of doping $x$ is shown in Fig. 27. The diagram is constructed using Eqs. (48) and (49). It is *generic* because the basic forms of the gaps are dictated entirely by the algebraic structure which, in turn, is determined by the physically-motivated choice of generators given in Eq. (1). The precise values of the coupling strengths affect only details. Four doping-dependent en-

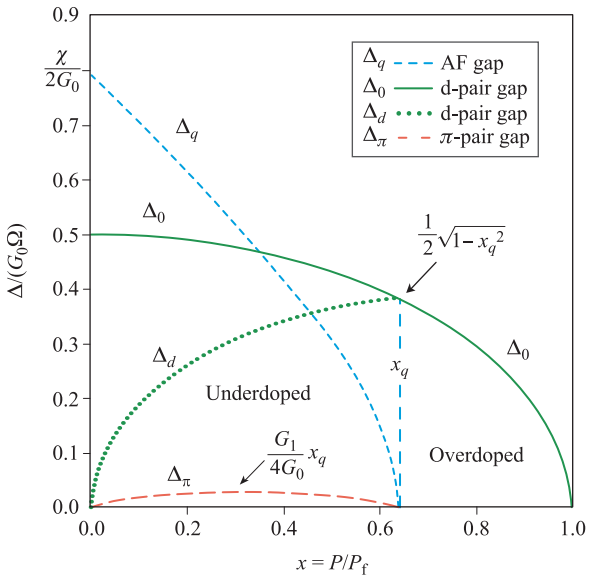



**Fig. 27** Generic diagram for energy gaps vs. doping, as predicted by the SU(4) model at $T = 0$. Energy gaps are scaled by $G_0\Omega$ and the doping parameter is scaled by the maximum doping $P_{\rm f}$ (consistent with data, we assume $P_{\rm f} = 1/4$ [49]). Interaction strengths are assumed to be independent of doping, with $\chi = 13$, $G_0 = 8.2$, and $G_1 = 1.3$ (arbitrary energy units), which, according to Eq. (47), requires the critical doping point to be $x_q = 0.64$, corresponding to a critical physical doping parameter $P_q = x_q P_{\rm f} \sim 0.16$.

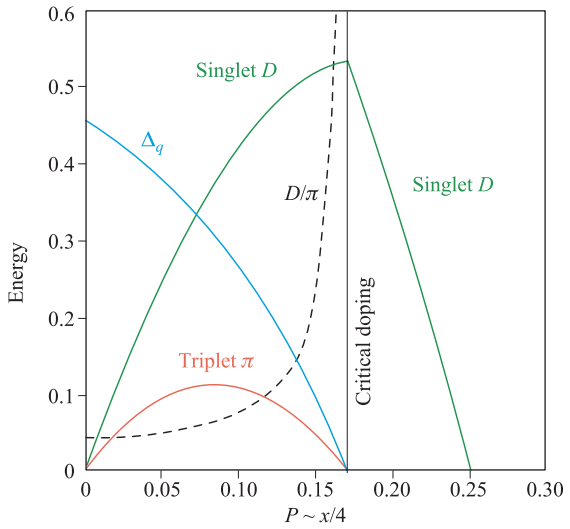



**Fig. 28** Behavior of various correlation energies with doping in the SU(4) model. The ratio of singlet to triplet pairing energy $D/\pi$ diverges as critical doping is approached.

ergy scales are predicted:

1) The gap $\Delta_q$ is defined in Eq. (48a) and measures AF correlations. It is maximal at $x = 0$, decreases rapidly to the region of the pairing gaps with increased doping, crosses the pairing gaps, and vanishes at the critical doping $x_q$.

2) The spin-singlet pairing gap $\Delta_d$ defined in Eq. (48b) is the superconducting gap for $x < x_q$.

3) The spin-singlet pairing gap $\Delta_0 = \Delta_d$ defined in Eq. (49b) is the superconducting gap for $x > x_q$, but is not the ground-state order parameter when $x \leq x_q$.

4) The spin-triplet pairing gap $\Delta_\pi$, is defined in Eq. (48c). Like $\Delta_q$, it exists only in the doping range $x \leq x_q$. It reaches its maximum value at $x_q/2$ and vanishes at $x = 0$ and $x = x_q$.

## 9.3 The critical doping point

The spin-singlet pairing gap exhibits qualitatively different behavior for doping less than or greater than $x_q$. For $x > x_q$, it corresponds to a monotonic curve labeled $\Delta_0$, but below $x_q$ the spin-singlet gap splits into two curves (labeled $\Delta_d$ and $\Delta_0$) having very different doping dependence, with the splitting increasing for decreasing doping.

The critical doping point and the splitting of the SC pairing gap result from competing SC pairing and AF correlation for $x < x_q$. At small doping the AF correlation dominates the SC pairing and a state with large AF correlations and suppressed pairing can become the ground state. Therefore, the superconducting gap $\Delta_d$ for the ground state is smaller and the larger pairing gap $\Delta_0$ is associated with an excited state in this doping range. However, as doping increases the pairing correlation grows quickly and the AF correlation decreases. The point $x_q$ marks the doping fraction at which the AF correlation is fully suppressed, leaving complete dominance of the superconducting correlations for $x_q < x < 1$.

The critical doping point $x_q$ defines a natural boundary between regions where the wavefunctions are qualitatively different, as illustrated further in Fig. 28. It separates a doping regime characterized by weak superconductivity and reduced pair condensation energy from a doping regime characterized by strong superconductivity and maximal pair condensation. The optimal doping point (maximum of the SC pairing gap) is often used empirically to demarcate underdoped and overdoped superconductors. It is the doping point where competition between the AF and SC correlations leads to the maximal $T_c$. Figure 28 suggests that optimal doping and the quantum phase transition at critical doping are closely associated.

## 9.4 Comparison with gap data

Figure 29 is reproduced from Ref. [51] and compares computed SU(4) gaps with measured ones for the hole-doped cuprates. The gap data, which are taken from Refs. [40, 78], appear to support the complex gap structure suggested by the SU(4) quasiparticle solutions. In particular, the predicted splitting of the singlet pairing gap and termination of the $\Delta_q$ gap at a critical doping $P_q \sim x_q/4$ are consistent with the data points plotted. A simple variation of the coupling-strength parameters (inset to Fig. 29) has been allowed to facilitate a more precise fit to data, but it was demonstrated in Ref. [51] (and will be discussed further below) that the basic features of the gap diagrams are reproduced even with values of the coupling strengths in the highly-renormalized subspace that are assumed (probably unphysically) to be independent of doping. This sug-

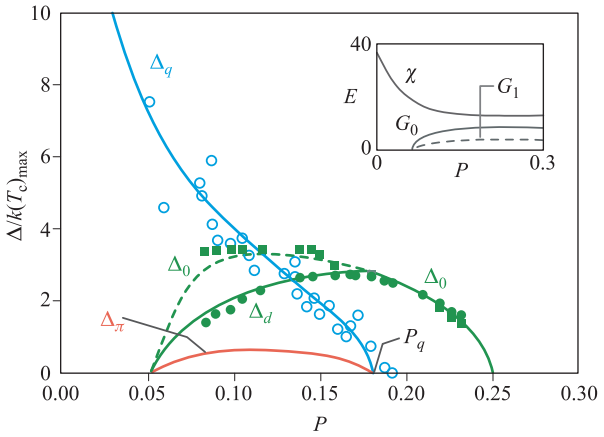



**Fig. 29** Comparison with data for energy gap diagrams at $T = 0$ as a function of doping in hole-doped cuprates. The doping rate is defined as $P = (\Omega - n)/\Omega_e$ with $\Omega_e$ and $\Omega$ being the number of lattice sites and the maximum allowed number of holes, respectively. Doping-dependent strengths indicated by the inset to the figure are used and $P_q = 0.18$ was assumed. Data were taken from Refs. [40] (open blue circles) and [78] (green squares and red filled circles).

gests that the gap structure observed in the cuprate superconductors is a generic feature of the SU(4) symmetry, independent of microscopic details.

### 9.5 Competing order and preformed pairs

In the *competing-order picture* [40], the pseudogap (PG) is an energy scale for an order that competes with superconductivity and vanishes at a critical doping point. From Fig. 29, $\Delta_q$ has precisely these properties. But the AF operators entering into $\Delta_q$ are generators of SU(4), so $\Delta_q$ *also* is the stabilization energy for a mixture of "preformed" singlet and triplet SU(4) pairs that condense into a strong superconductor only after AF and triplet-pairing fluctuations are suppressed by doping. This is a non-abelian generalization of the phase-fluctuation model [39] for *preformed pairs*. Preformed pair and competing order models have generally been viewed as mutually antagonistic explanations of the pseudogap. However, we see that the SU(4) PG state results from competing AF and SC order expressed in a basis of singlet and triplet fermion pairs, which may be viewed as a unification of the competing order and preformed pair pictures of the cuprate pseudogap state.

We conclude that the nature of the pseudogap state requires *both* competing order and preformed pairs, but the requisite preformed pairs are more sophisticated than those of a simple phase fluctuation model because of the strong and complex correlations that are present in the realistic system. BCS-like pairs correspond to an SU(2) subgroup of the full SU(4) algebra. Their phase fluctuations represent phase rotations about a single axis for an abelian U(1) subgroup. In contrast, the full SU(4) algebra has 15 generators and various non-abelian subgroups. Thus, phase fluctuations of SU(4) pairs involve (non-commuting) rotations around axes in a multidimensional space, implying a rich structure for the preformed pairs.

For example, we have seen that in the SO(5) limit at low doping there is a significant pairing correlation energy but no long-range order because there is no barrier to phase fluctuations between the SC and AF directions in the SU(4) space. This is a particularly clear example of states that have a significant pairing correlation energy (preformed pairs) but no long-range order of either SC or AF form, because of quantum fluctuations in the SU(4) solutions (competing order modified by quantum fluctuations).

### 9.6 The role of triplet pairs

Triplet pairs are an essential component of the SU(4) many-body wavefunction. The SU(4) algebra doesn't even close if the corresponding operators are omitted, implying that the operator set and corresponding Hilbert space are quantum-mechanically incomplete in the absence of triplet pairs (see Fig. 8 and the general discussion in Section 4.2). However, Fig. 29 indicates that the triplet-pair correlation energy is small for underdoped cuprates and vanishes for doping larger than the critical doping $x_q$.

The primary role of SU(4) triplet pairs in the hole-doped cuprates lies in fluctuations mediating the AF–SC competition at lower doping. This interpretation is supported by the observation that the triplet pair operators are fundamental generators of the SO(5) critical dynamical symmetry [see Table 1 of Appendix A and Eq. (94)], which acts physically as a doorway between AF and SC order (Fig. 14). Thus, although the superconductivity in the cuprates is implemented in terms of a condensate of singlet pairs and triplet pairs play no direct role in the charge transport of the superconducting state, the triplet pairs are central to the nature of cuprate superconductivity because they are key to mediating fluctuations and the rapid transition from AF to SC order with hole doping.

Hence, below the critical doping point $x_q$ and below the superconducting critical temperature $T_c$, the superconducting state is dominantly spin-singlet but it necessarily has a small spin-triplet admixture because the AF correlations are non-vanishing for $x < x_q$ and they scatter singlet pairs into triplet pairs and vice versa. Only for $x > x_q$ does the SU(4) symmetry force the AF correlations to vanish identically, leaving a condensate of pure-singlet Cooper pairs for $T < T_c$.

## 10 SU(4) phase diagrams

We may use the formalism developed in Section 6.8 to construct doping–temperature phase diagrams corresponding to SU(4) coherent-state solutions. In Figs. 30 and 31 we show typical phase diagrams resulting from solution of the

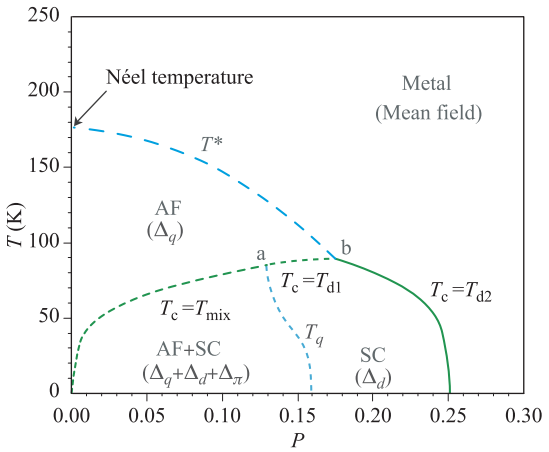



**Fig. 30** Phase diagram predicted by the SU(4) model with $R = 0.6$. Interaction strengths are the same as those used in Fig. 27, but in units of $k_B(T_c)_{\max}$, where $(T_c)_{\max}$ is taken to be 90 K. The critical doping point is $P_q = 0.16$ (corresponding to $x_q = 0.64$).

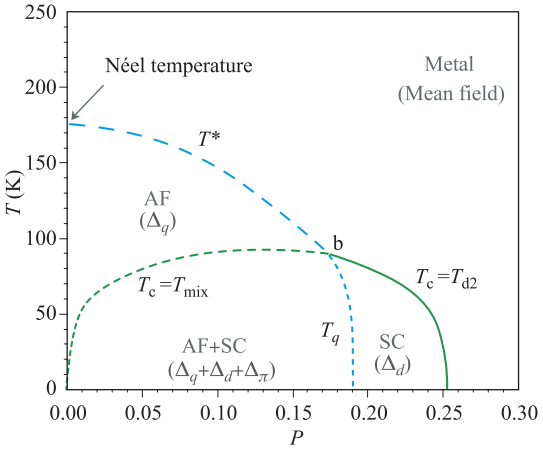



**Fig. 31** Phase diagram predicted by the SU(4) model for $P_q > P_b$. $P_q = 0.19$ (corresponding to $x_q = 0.76$) is chosen in this figure while $P_b = 0.175$ (corresponding to $x_q = 0.7$) is the same as that in Fig. 30. All parameters remain unchanged from Fig. 30, except that $G_1$ was increased from 1.3 to 4.7 [in units of $k_B(T_c)_{max}$], giving a larger value of $P_q$.

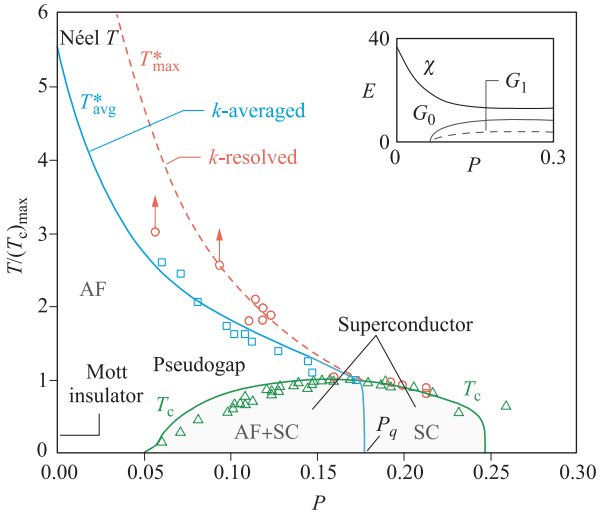



**Fig. 32** SU(4) cuprate phase diagram compared with data. Strengths of the AF and singlet pairing correlations were determined in Ref. [51] by global fits to cuprate data. The PG temperature is $T^*$ and the SC transition temperature is $T_c$. The AF correlations vanish, leaving a pure singlet $d$-wave condensate, above the critical doping $P_q$. Dominant correlations in each region are indicted by italic labels. Data in green (open triangles) and blue (open squares) are taken from Ref. [79], and those in red (open circles) from Ref. [80] (arrows indicate that the point is a lower limit). The two different curves for the theoretical pseudogap temperature are for experiments in which the momentum $k$ is either resolved ($T^*_{max}$) or not ($T^*_{avg}$). Data in blue do not resolve $k$; data in red resolve $k$.

SU(4) finite-temperature gap equation. Both figures represent possible physical solutions. They differ primarily in the assumed strength of triplet pairing, which is not a well-determined parameter. In the calculations the interaction strengths are kept constant, with the same values as those in the zero-temperature case discussed in Figs. 26 and 27. The only adjustable parameter is $R$ [appearing in Eq. (34)].

### 10.1 The predicted phases

Four distinct phases may be identified in Figs. 30 or 31:

1) An antiferromagnetic phase labeled AF.
2) A superconducting phase labeled SC.
3) A transitional phase with all three correlations present, which we label a mixed phase (marked as AF+SC).
4) A phase without net antiferromagnetic or superconducting order that is labeled metal (mean field).

The correlations (energy gaps) associated with each phase are indicated in parentheses. The doping-dependent transition temperatures $T_c$, $T^*$, and $T_q$ define the boundaries for these phases, and the points $a$ and $b$ mark the intersections of multiple phases.

### 10.2 Comparison with data

Figure 32 compares with cuprate data and indicates that the coherent state solution can give a quantitative description of both the superconducting transition temperature $T_c$ and the pseudogap transition temperature $T^*$. Two theoretical pseudogap temperatures are shown because the predicted value of $T^*$ depends on whether the experiment resolves the momentum $k$. The SU(4) operators that we have defined in Eqs. (1)–(5) average over the momentum $k$. Thus, they are appropriate for comparison with experiments that do not resolve $k$. We denote the PG temperature computed in this approximation $T^*_{avg}$. However, as discussed in depth in Ref. [52], for experiments that resolve $k$ the appropriate form of the SU(4) generators carries a momentum index and the pseudogap temperature computed in the SU(4) formalism for specified $k$, which we denote as $T^*_{max}$, is in general higher than the PG temperature computed in the $k$-averaged SU(4): $T^*_{max} > T^*_{avg}$.

The neutron-scattering data shown as blue squares in Fig. 32 do not resolve $k$. The corresponding SU(4) pseudogap temperature $T^*_{avg}$ was determined using the standard $k$-averaged version of the SU(4) model. The ARPES PG data shown in red circles in Fig. 32 do resolve $k$. The corresponding SU(4) pseudogap temperature $T^*_{max}$ was computed using the $k$-dependent SU(4) model described in Section 6.9 and in Refs. [52, 54], assuming a single band contributing to the pairing. Particularly in the underdoped region, this comparison of theory with data suggests that measurements on the *same system* but with different experimental methods having different $k$ resolutions may find different pseudogap temperatures $T^*$.

Figures 30–32 may explain the vortex-like Nernst signal [81] observed above $T_c$. In the present SU(4) picture the singlet pair gap vanishes there but the many-body SU(4)

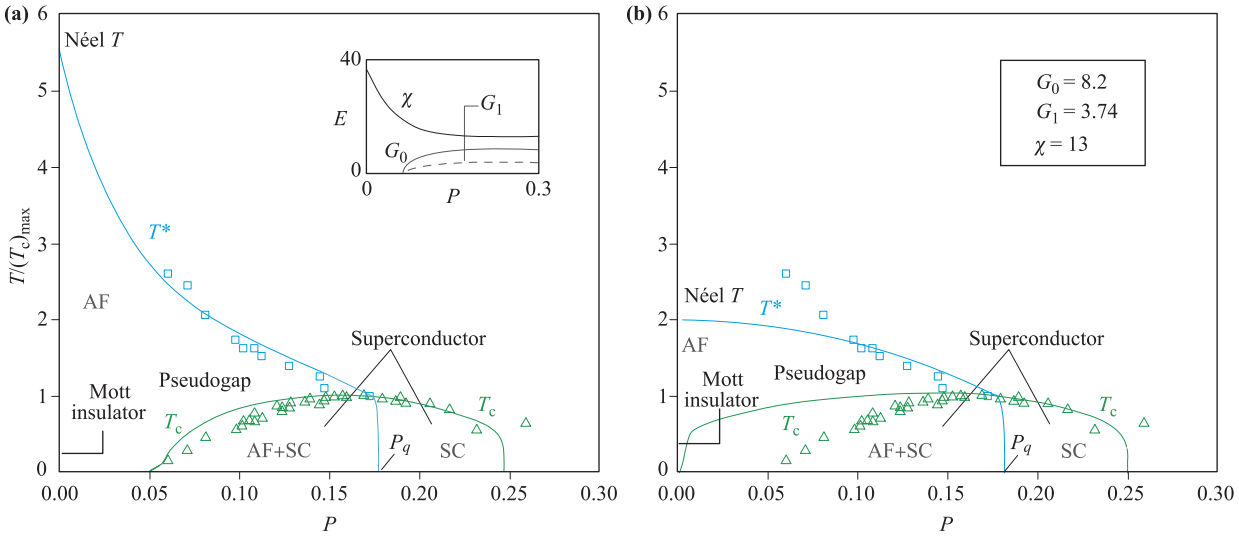



**Fig. 33** Camparison of cuprate phase diagram calculated **(a)** using an effective interaction varying smoothly with doping and **(b)** calculated using constant effective interaction strengths, independent of doping. Although an effective interaction that does not depend on doping is almost certainly not quantitatively realistic, the two diagrams are seen to be qualitatively similar. This suggests that the basic features of the cuprate phase diagram are a consequence of SU(4) dynamical symmetry.

wavefunction has finite pair content in the region between the curves $T^*$ and $T_c$ that decreases with increasing $T$ and decreasing doping. Thus, contours for pair fluctuations above $T_c$ may be expected to be similar to observed contours for Nernst signal strength, but a Meissner effect is expected only below $T_c$. Ong *et al.* [82] have concluded that a consistent explanation of pseudogap and Nernst data requires PG and SC pairing states that are distinct but related by symmetry, as proposed here.

Perhaps the most remarkable result obtained from phase calculations is illustrated in Fig. 33, where we compare a phase diagram calculated using an effective interaction varying smoothly with doping (as in Fig. 32) and one using a constant effective interaction. An effective interaction that does not depend on doping is almost certainly not realistic since the Hilbert space excluded by truncation is affected by the doping. Nevertheless, the two diagrams are *qualitatively similar*. This implies that the basic structure of the cuprate phase diagram is *determined entirely by the dynamical symmetry,* with parameters reflecting the underlying microscopic structure affecting only quantitative details in a smooth way. Specifically, we see that a quantititatively better description of data is obtained if the ratio of antiferromagnetic to singlet pairing coupling is larger at low doping, as might be expected on physical grounds, but that the qualitative properties of the phase diagram do not depend on such parameter adjustment.

## 11 Fundamental instabilities

The SU(4) symmetry implies two fundamental instabilities that may play a key role in understanding the properties of high-temperature superconductors. The first provides a natural explanation for the propensity of cuprate Mott insulator states to become superconductors with only modest hole doping. The second provides an explanation of how high-temperature superconductors can exhibit a rather universal phase diagram and at the same time display substantial local inhomogeneity, particularly at lower doping.

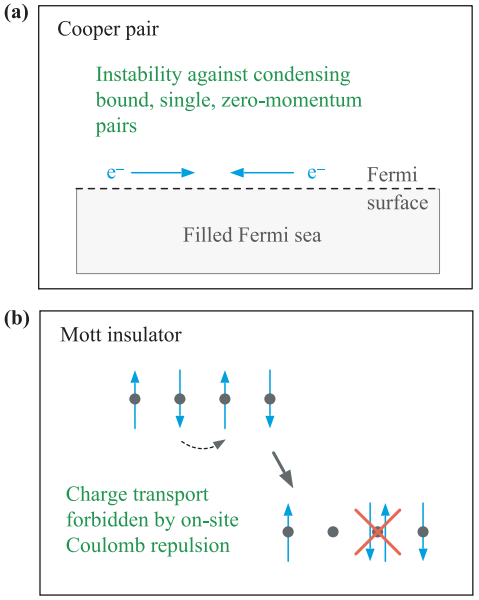



**Fig. 34** The Cooper instability and Mott insulators. **(a)** Basic properties of normal superconductors arise from an instability of the Fermi liquid normal state and are well described by the BCS formalism. **(b)** High-temperature superconductors call this picture into question because superconductivity in the cuprates derives from a Mott insulator normal state, not a Fermi liquid.

### 11.1 Pairing instability with doping

The pairing instability of cuprate superconductors has generally been viewed as difficult to understand. As illustrated in Fig. 34, normal superconductors develop from a Fermi liquid normal state, but cuprate superconductors appear to arise from a (non-Fermi-liquid) antiferromagnetic Mott insulator state. Furthermore, this Mott insulator state seems to harbor a secret propensity to superconductivity, since these compounds can be turned from AF Mott insulators into high-temperature superconductors by hole doping at a modest 3%–5% level. Why do the Mott insulator normal states for the cuprates become superconducting with only a small amount of hole doping? The emergent SU(4) symmetry provides a natural explanation. From the $T = 0$ solution for $\Delta$ given in Eq. (48b),

$$\left.\frac{\partial\Delta}{\partial x}\right|_{x=0} = \left.\frac{1}{4}\frac{x_q^{-1} - 2x}{[x(x_q^{-1} - x)]^{1/2}}\right|_{x=0} = \infty, \tag{98}$$

displaying explicitly a fundamental pairing instability at $x = 0$. Thus, the SU(4) symmetry implies that the ground state at half filling is an antiferromagnetic Mott insulator, but that this state is fundamentally unstable against condensing singlet hole pairs under infinitesimal hole doping if there is a finite attractive pairing interaction. We have termed this *precocious pairing* [56].

This instability of the SU(4) symmetry against condensing pairs as the system is doped away from half filling may also be illustrated graphically using the SU(4) coherent-state energy surfaces. Figure 35 shows the SU(4) total energy surface in coherent state approximation as a function of AF and SC order parameters

$$Q = \frac{1}{\Omega}\langle \boldsymbol{Q}\cdot\boldsymbol{Q}\rangle^{1/2}, \quad \Delta = \frac{1}{\Omega}\langle p^\dagger p\rangle^{1/2},$$

as doping $x \simeq 4P$ (for $P$ holes per copper lattice site) is varied. From Eq. (95), the explicit expression for the energy surface is given by

$$E = -\chi\Omega^2[(1 - x_q^2)\Delta^2 + Q^2], \tag{99a}$$

$$\Delta \equiv \frac{1}{2}\left[\frac{1}{4} - \left(Q - \frac{x}{2}\right)^2\right]^{1/2} + \frac{1}{2}\left[\frac{1}{4} - \left(Q + \frac{x}{2}\right)^2\right]^{1/2}, \tag{99b}$$

where $x_q$ is the critical doping (see Fig. 35 and Section 6.7.1). The vertical lines bounding the curves for different doping in Fig. 35 represent the constraints

$$|Q| \le \frac{1}{2}(1 - x), \quad (x(1 - x))^{1/2} \le 2\Delta \le (1 - x^2)^{1/2}$$

that result from SU(4) symmetry realized within a finite valence space. Specifically, $|Q|$ must lie between 0 and $\frac{1}{2}n$ (where $n$ is electron number) because of the finite number of spins available per lattice site, and SU(4) symmetry then relates this constraint on $Q$ to the one on $\Delta$.

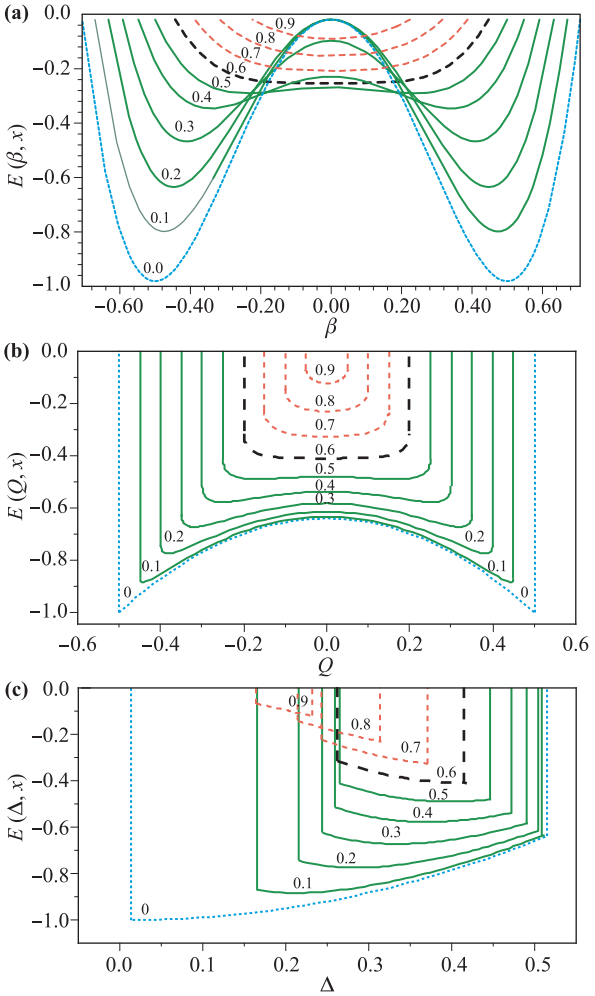



**Fig. 35** Total energy vs. AF order parameters $\beta$ **(a)** and $Q$ **(b)**, and SC order $\Delta$ **(c)**; curves labeled by hole-doping $x \simeq 4P$, where $P$ is the number of holes per copper site. Energy is in units of $\chi\Omega^2/4$, with $\chi$ the AF coupling strength. The heavy black dashed line indicates the critical doping $x = x_q \sim 0.6$ (see Ref. [50]); color coding indicates energy surfaces favoring SC (red, short-dash, from $x = 0.7$–$1.0$), AF (blue, dotted, $x \sim 0$), and AF + SC (green, solid, $x \sim 0.1$–$0.5$). The finite, doping-dependent ranges of the energy contours in the order parameters reflect the finite valence space (single band) of the microscopic model.

The expectation values of the order parameters $Q_0$ for antiferromagnetism and $\Delta_0$ for singlet pairing from the energy surfaces in Fig. 35 are illustrated in Fig. 36. From Fig. 35 and Fig. 36, the energy surface at half filling ($x = P = 0$) implies a $T = 0$ ground state with AF order but no pairing order ($Q_0 \neq 0$ and $\Delta_0 = 0$, where the subscript zero denotes the value at the minimum of the energy surface). The ground state for $x = 0$ may be interpreted as an antiferromagnetic Mott insulator [46, 48]. From Fig. 35(a), the energy surface retains strong AF character for small $x$ with $Q_0 \neq 0$, but from Fig. 35(b) the ground state differs qualitatively from that at half filling even for

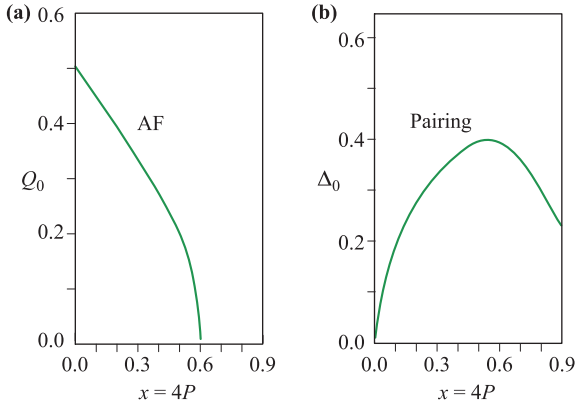



**Fig. 36** Expectation values of the order parameters **(a)** $Q_0$ for antiferromagnetism and **(b)** $\Delta_0$ for singlet pairing, as a function of hole doping. Evaluated from the energy surfaces in Fig. 35.

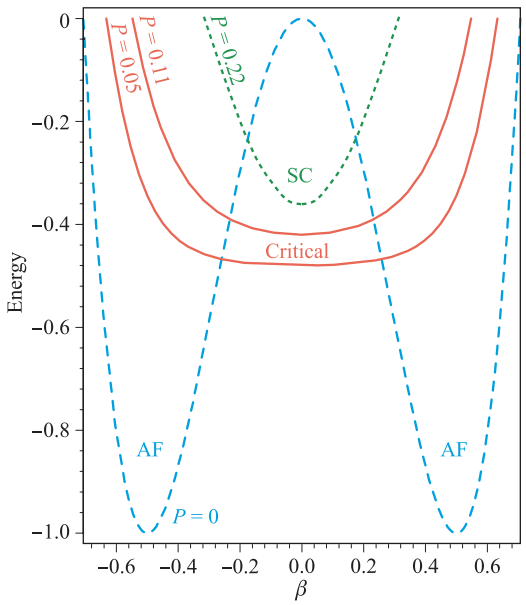



**Fig. 37** Three classes of SU(4) energy surfaces as a function of the AF order parameter $\beta$. Cases labeled by the hole-doping parameter $P$: superconducting (SC, dotted green), antiferrromagnetic (AF, dashed blue), and critical (solid red).

infinitesimal hole-doping. Specifically, for any non-zero attractive pairing strength a finite singlet $d$-wave pairing gap develops spontaneously for *any non-zero* $x$, and $\Delta_0$ has increased to half its value at optimal doping by $P \simeq 0.03$ ($x \simeq 0.12$).

Thus, at half filling the SU(4)-symmetric lattice is a Mott insulator with long-range AF order and no pairing order, but upon infinitesimal hole-doping a finite singlet pairing gap and a ground state representing strong competition between AF and SC order appears. This spontaneous development of a finite singlet pairing gap for infinitesimal hole-doping has been obtained in the coherent-state approximation subject to SU(4) symmetry. Doping dependence of the effective interactions, AF and SC fluctuations, or weak breaking of SU(4) symmetry could delay the onset of the instability from $P \sim 0$ to a small doping fraction, as observed. However, we propose that the pair-condensation instability of the SU(4) symmetry limit represents the essential physics governing the rapid emergence of superconductors from doped Mott insulators in the cuprates.

The remarkable conclusion that a weakly-doped antiferromagnetic Mott insulator can rearrange itself spontaneously into a superconductor follows directly from SU(4) invariance, which requires that for any SU(4)-symmetric solution,

$$Q^2 + \Delta^2 + \Pi^2 = \frac{1}{4}(1 - x^2), \tag{100}$$

where $\Pi = \langle \boldsymbol{\pi}^\dagger \boldsymbol{\pi} \rangle^{1/2}/\Omega$ is the triplet pair correlation [50]. But for the pure antiferromagnetic SU(4) solution, $Q^2 = \frac{1}{4}(1-x)^2$. Thus, comparing with (100), we see that even the antiferromagnetic limit has finite pairing gaps $\Delta^2 + \Pi^2$ unless $x$ vanishes identically.

### *11.1.1 Implications for cuprates at low hole doping*

Cuprate data for low doping suggest that normal compounds at half filling are AF Mott insulators, that a finite pairing gap develops when doping reaches $P \simeq 0.05$, and that a pseudogap develops in the underdoped region having doping dependence for the PG temperature $T^*$ opposite that of $T_c$ for the singlet pairing gap. These results are consistent with a Mott insulator state at half filling that evolves rapidly into a state with a finite singlet pairing gap at very low hole-doping. However, since at low doping both the singlet pairing and antiferromagnetic correlation energies are substantial, the AF fluctuations prevent development of full-strength superconductivity until the critical doping point $x_q$, where the zero-temperature AF correlations are completely suppressed at a quantum phase transition.

Below $T_c$ for doping less than the critical doping, this leads to a $d$-wave SC state weakened by antiferromagnetic fluctuations. For a range of temperatures above $T_c$ but for doping less than critical, the pairing gap vanishes but strong AF correlations in a basis of fermion pairs leads to a pseudogap that may be interpreted in terms of preformed pairs having a structure strongly influenced by competing AF and SC order. Finally, the AF competition weakens with hole doping until the pure superconductor emerges and the pseudogap disappears near the critical doping point (which is typically near optimal doping).

### *11.1.2 The generalized Cooper instability*

Our results show that an inherent instability against condensing Cooper pairs as doping is increased occurs naturally in a minimal model of singlet bond-wise pairing interacting with AF correlations on a lattice with no double occupancy. Thus, the rapid onset of superconductivity

with hole-doping in the cuprates results from a Cooper-like instability against condensing pairs for non-zero attractive pairing interaction, but for $d$-wave pairs in an AF Mott insulator. This solution reduces formally to ordinary $d$-wave BCS theory if the AF interaction vanishes, and to an antiferromagnetic Mott insulator if the pairing vanishes (see Sections 15.1 and 15.2, and Ref. [50]); thus it represents a minimal self-consistent generalization of the Cooper instability to doped Mott insulators. To summarize: there is nothing mysterious about the rapid onset of superconductivity upon doping a Mott insulator with holes. This is just the Cooper instability, but for a Fermi sea polarized by onsite Coulomb repulsion and antiferromagnetic correlations.

#### *11.1.3 Implications for resonating valence bond models*

The resonating valence bond (RVB) model [83] assumes that quantum antiferromagnets should exhibit superconductivity, at odds with the observation that cuprate ground states at half-filling appear to have long-range AF order and no superconductivity [84]. Thus, resonating valence bond models assume implicitly that the half-filled state is in some sense practically a spin liquid, though it looks like an AF state. Our results give independent support for a picture rather similar to this, but without explicit RVB assumptions [53]: the SU(4) ground state at half-filling is an antiferromagnetic Mott insulator, but its wavefunction can reorganize spontaneously into a superconductor when perturbed by a vanishingly-small hole doping in the presence of a non-zero pairing interaction. We shall have more to say about the relationship between the RVB and SU(4) models in Section 15.4.

### 11.2 Critical dynamical symmetry and inhomogeneity

We now demonstrate that SU(4) symmetry implies a second fundamental instability near the critical doping point $x_q$, and that this instability may also play a key role in the observed properties of cuprate superconductors. From the $T = 0$ solution for $Q$ given by Eqs. (48a) and (39c), we find

$$\left.\frac{\partial Q}{\partial x}\right|_{x=x_q} = -\frac{1}{4}\left.\frac{x_q + x_q^{-1} - 2x}{[(x_q - x)(x_q^{-1} - x)]^{1/2}}\right|_{x=x_q} = -\infty, \tag{101}$$

and a small change in doping will cause a large change in antiferromagnetic correlations near $x = x_q$. This is a consequence of SU(4) symmetry, which requires that $Q$ vanish for $x \geq x_q$, and be finite for $0 < x < x_q$. The physical implications of this instability are most easily demonstrated using the coherent-state energy surfaces discussed in Section 8.

The energy surfaces at constant doping in Figs. 23 and 24 fall into three general classes: AF+SC (e.g., $x = 0.1$), SC (e.g., $x = 0.9$), and critical (e.g., $x = 0.6$, which marks a quantum phase transition), as illustrated in Fig. 37. Curves in the AF+SC class have minima at finite and large $\beta_0$, and small but finite $\Delta_0$, where the subscript zero denotes the value of the order parameter at the minimum of the energy surface. Curves in the class SC are characterized by $\beta_0 = 0$ and finite $\Delta_0$. Of most interest in the present context are the surfaces that are near critical in Fig. 37, which correspond to broken SU(4) $\supset$ SO(5) dynamical symmetry [46, 48] and are rather *flat* over large regions of parameter space. This implies that there are many states lying near the ground state with very different values for $\beta$ and $\Delta$. Thus the surface is critically balanced between AF and SC order, and small perturbations can drive it strongly from one to the other. This defines a critical dynamical symmetry of the SU(4) algebra, as discussed in Sections 8.4 and 8.5, and Refs. [46, 48]; we shall term this situation *dynamical criticality*. Figure 35 suggests that underdoped cuprates have near-critical energy surfaces. Thus, critical dynamical symmetry may be central to the discussion of inhomogeneity and to the general issue of understanding pseudogap states in underdoped cuprates.

#### *11.2.1 Dynamical criticality and sensitivity to perturbations*

The extreme sensitivity of critical surfaces to perturbations is illustrated in Fig. 38. Each set of curves is associated with a fixed doping $x = 0.6$ ($P = 0.15$), with the

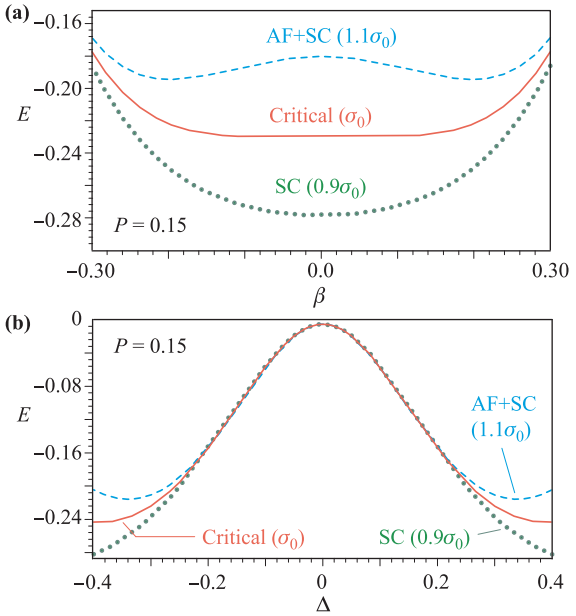



**Fig. 38** Energy surfaces as a function of **(a)** the AF order parameter $\beta$ and **(b)** the singlet pairing order parameter $\Delta$, for fixed doping $P = \frac{1}{4}x = 0.15$. The solid red curve corresponds to $\sigma = \sigma_0 = 0.6$, the upper dashed blue curve to a 10% increase in $\sigma$, and the lower dotted green curve to a 10% reduction in $\sigma$.

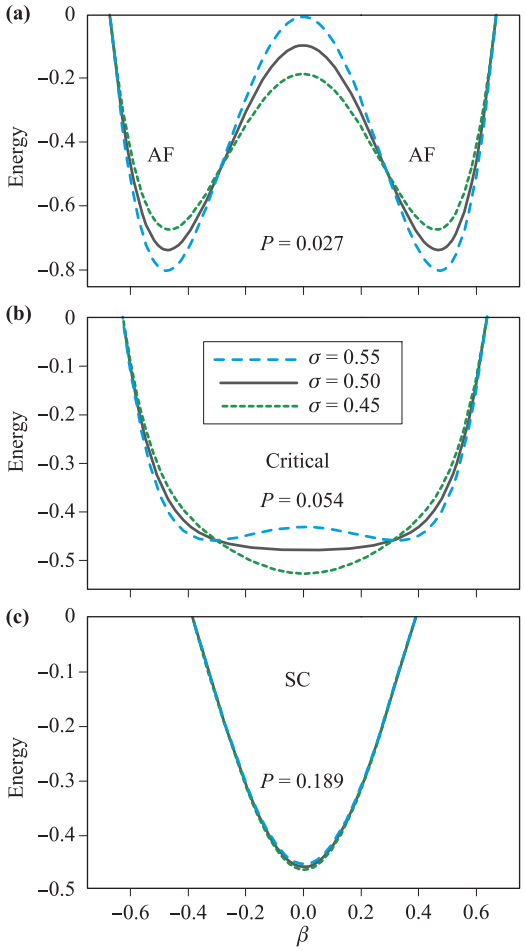



**Fig. 39** Effect of altering the ratio $\sigma$ of antiferromagnetic to superconducting coupling strengths for three values of hole doping in the cuprates. In **(b)** the perturbation changes the nature of the ground state, but in **(a)** and **(c)** it hardly alters the location of the deep energy minima, implying that the effects of dynamical criticality are specific to underdoped compounds.

solid line corresponding to $\sigma = 0.6$, the dashed line to a 10% increase in $\sigma$ (AF perturbation), and the dotted line to a 10% reduction in $\sigma$ (SC perturbation). In Fig. 38(a) we see that the perturbation changes the AF character of the ground state. The effect on $\Delta$ in Fi.g 38(b) is less dramatic: $\Delta_0$ is shifted, but remains finite in all three cases. We see that this small fluctuation in $\sigma$ can alter the energy surface between AF+SC (finite $\beta_0$ and $\Delta_0$) and SC ($\beta_0 = 0$ and finite $\Delta_0$). This sensitivity is specific to the critical [broken SO(5)] dynamical symmetry. The antiferromagnetic region near $x = 0$ and the superconducting region at larger hole doping (see Fig. 35) are very stable against such perturbations, as illustrated in Fig. 39.

Various mechanisms could alter the ratio of AF to SC coupling locally. For example, it was found in Ref. [38] that nanoscale disorder is tied to influence of dopant impurities, and a number of authors have discussed possible reasons for this in terms of mechanisms like lattice-distortion

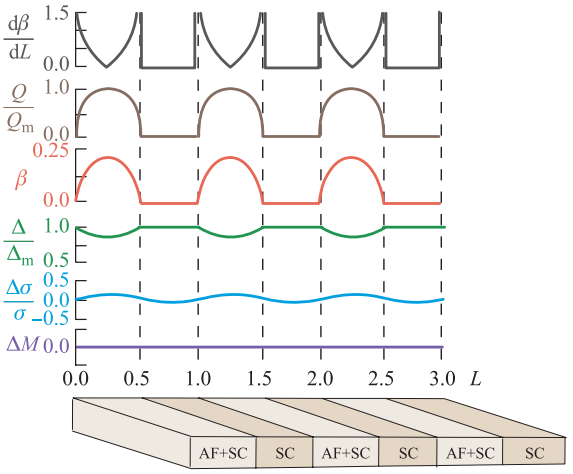



**Fig. 40** A small spatial variation $\Delta\sigma/\sigma$ in the AF and SC coupling can produce inhomogeneity without necessarily implying significant charge variation. The spatial scale is $L$, the charge is $M$, $\Delta$ is the SC order parameter, $\beta$ and $Q$ are AF order parameters, and a subscript m denotes maximum values.

leading to modification of electron–phonon coupling or superexchange [85–93].

### *11.2.2 Spatial inhomogeneity induced by background perturbations*

Figure 40 is constructed from the expectation value of Eq. (14) in coherent state approximation, assuming a periodic one-dimensional spatial perturbation of form $\Delta\sigma \propto \sin(2\pi L)$ around $\sigma = 0.6$. This figure illustrates schematically how a small (10%) periodic fluctuation in the antiferromagnetic and superconducting coupling for a critically-symmetric underdoped compound can lead to inhomogeneity. In this example, one-dimensional spatial variations of the coupling ratio $\sigma$ give fluctuations in order parameters leading to stripes in which AF+SC ($\sigma > 0.6$) and SC ($\sigma < 0.6$) are favored alternatively. Also shown are the responses of AF fluctuations $\mathrm{d}\beta/\mathrm{d}L$ to this variation in $\sigma$. (We do not intend this as a realistic model of a stripe phase, but rather as a cartoon indicating how such emergent structures can arise in the underdoped region because of background perturbations on energy surfaces that have been rendered critical by the SU(4) symmetry constraints.)

Figure 40 indicates that a small spatial variation in the coupling ratio $\sigma$ can produce regions having large AF and weaker SC order, interspersed with regions having significant SC correlations but no AF order. The nature of the underdoped energy surface dictates that the primary fluctuation between regions is in the AF order, which can jump between zero and the maximum allowed by the SU(4) model in adjacent stripes; the pairing order changes are smaller and pairing is finite in both the SC and SC+AF regions. Finally, the variation of $\mathrm{d}\beta/\mathrm{d}L$ indicates that appreciable softness in the AF and SC may

occur on the boundaries between regions, implying that such induced structure could be static or dynamic.

The minimal patch size that can support SU(4) coherent states is of interest in the present context. We have shown (for example, Ref. [18]) that dynamical symmetry can be realized in strongly-correlated fermion valence spaces having as few as several particles. This would be consistent with inhomogeneity on scales comparable to atomic dimensions, as required by data [38]. The preceding example employed a periodic modulation of $\sigma$ as a simple demonstration but it is clear that a similar mechanism can operate for non-periodic perturbations.

#### *11.2.3 The role of charge*

The expectation value of the charge $M$ is not a function of the coupling ratio $\sigma$ in the SU(4) coherent state, so the critical energy surface fluctuations responsible for alternating AF+SC and SC stripes in Fig. 40 cause no charge variation ($\Delta M = 0$). Data indicate that the relative charge variation for Bi-2212 surface nanoscale patches is less than 10%, implying heterogeneity that is not strongly coupled to charge [38]. This view is supported by analyses of heat capacity and NMR data on Bi-2212 and YBCO that find a universal phase behavior for cuprates, with little static charge modulation in evidence [94, 95].

Of course, the variation in $\sigma$ could be *caused* by a charge modulation. From the preceding discussion, we may expect that if a charge modulation occurs in either the antiferromagnetic region near half filling or the superconducting region at larger doping, its effect will be small because the energy surfaces are not critical there. However, if a charge modulation occurs in the underdoped region where energy surfaces are near critical, its effect may be amplified by dynamical criticality even if $\sigma$ is not altered significantly, since this is equivalent to a doping modulation.

Thus, we suggest a mechanism operating only in underdoped materials that could produce strong inhomogeneity without necessarily invoking charge fluctuations, but that could in particular cases be associated indirectly with a charge modulation. Such a mechanism could explain strong local inhomogeneity in the face of a universal overall phase diagram, and resolve competing experimental claims regarding the role of charge variation in producing inhomogeneity. In our view the central issue is spatial modulation of AF and SC coupling; spatial modulation of charge is a possible but not a necessary corollary.

#### *11.2.4 Self-organization versus dopant impurities*

In the literature, inhomogeneity caused by electronic self-organization is often contrasted with that caused by dopant impurities. Our discussion implicates both as sources of nanoscale structure. The immediate cause may be impurities that perturb the SU(4) energy surface but the criticality of that surface, which can greatly amplify the influence of the impurities, results from the self-organizing, doped Mott insulator encoded in the SU(4) algebra. Note that Ref. [86] suggests, from a different perspective, that intrinsic amplification of impurity effects is required to explain nanoscale structure in Bi-2212.

#### *11.2.5 Amplification of proximity effects*

Giant proximity effects are observed in the cuprates where non-superconducting copper oxide material sandwiched between superconducting material can carry a supercurrent, even for a thickness much larger than the coherence length [96]. Phenomenology indicates that pre-existing nanoscale SC patches can precipitate such effects [97]. We find similar possibilities here, but also suggest that the inhomogeneity need not pre-exist. Dynamical criticality renders even a homogeneous pseudogap phase unstable against fluctuations in the AF and SC order. Thus, proximity of superconducting material to pseudogap material, coupled with perturbations from background impurity fields, can trigger nucleation of nanoscale structure and giant proximity effects dynamically, even if no static inhomogeneity exists beforehand.

Because electronic disorder in cuprate superconductors can exist on a scale that is comparable to the coherence length ($\sim$ 15 Å), a proximity effect may be expected once static nanoscale disorder forms whereby SC character leaks between patches. This effect can be greatly enhanced if patches have energy surfaces that are critical, since a small perturbation that flips the energy surface of a patch has a non-perturbative influence far out of proportion to its size (a small tail wagging a large dog).

In experiments with one unit cell thickness $La_2CuO_4$ antiferromagnetic barrier layers between superconducting $La_{1.85}Sr_{0.15}CuO_4$ samples, it was found that the two phases did not mix, with the barrier layer completely blocking a supercurrent [98]. These results were interpreted to rule out models of high-temperature superconductors like the Zhang SO(5) model [99] in which SC and AF phases are nearly degenerate. The absence of a proximity effect between the antiferromagnetic and superconducting phases (but the presence of a strong proximity effect between pseudogap and SC material) is plausibly consistent with the SU(4) model. The SU(4) antiferromagnetic phase is not rotated directly into the SU(4) superconducting phase phase but instead evolves with increased doping into a competing SC and AF phase, which then is transformed into a pure superconducting state by a quantum phase transition at a critical doping point near optimal doping, where the antiferromagnetic correlations vanish identically [50].

#### *11.2.6 Dynamical criticality, emergence, and complexity*

As we have noted, the spontaneous appearance of properties in a many-body system that do not pre-exist in its elementary components is termed emergence; systems

with emergent behavior are said to exhibit complexity [100]. Complexity results when there are multiple potential ground states and the choice between them is sensitive to weak external perturbations. The amplification effect implied by SU(4) dynamical criticality can facilitate emergent behavior and associated complexity, as exemplified by giant proximity effects and the perturbatively-induced structure of Fig. 40. More generally, critical dynamical symmetry may be of fundamental importance for complexity in various strongly-correlated Fermi systems, not just those of condensed matter. For example, critical dynamical symmetries have long been known in nuclear structure physics [64, 65].

#### 11.2.7 *Varied inhomogeneity but a universal phase diagram*

The SU(4) coherent state method that yields the variational energy surfaces discussed here admits quasiparticle solutions that generalize the BCS equations, giving a rich, highly universal phase diagram in agreement with much available data [50]. Therefore, the propensity of the pseudogap state to a broad variety of inhomogeneity, and a quantitative model of the cuprate phase diagram (including pseudogaps) that exhibits highly universal character, both follow directly from a model that implements antiferromagnetic and $d$-wave superconductor competition in a doped Mott insulator. This natural coexistence of a universal phase diagram with a rich susceptibility to disorder in a limited region of its control-parameter space could reconcile many seemingly contradictory observations in the cuprate superconductors. For example, since the SU(4) pseudogap state has both pairing fluctuations and critical dynamical symmetry, Nernst vortex states could appear as perturbations on a homogeneous phase that is unstable against developing nanoscale disorder.

#### 11.2.8 *Critical dynamical symmetries near magnetic vortices and magnetic impurities*

Critical dynamics also may produce inhomogeneities in the vicinity of magnetic vortices and magnetic impurities. Because magnetic fields should suppress superconductivity, distance from a vortex $d$ may be expected to alter the average value of $\sigma$ in a similar way as changing the doping $P$ would. Therefore, regions near magnetic vortices or magnetic impurities may exist where the symmetry is critical and exhibits sensitivity to perturbations similar to that shown in Fig. 38, with distance $d$ from the vortex or impurity modulating $\sigma$ rather than $P$. This possibility is illustrated schematically in Fig. 41, which suggests that vortices and magnetic impurities may be surrounded by regions strongly susceptible to AF and SC fluctuations.

#### 11.2.9 *Summary*

The generality of the SU(4) solution discussed here implies that any realistic theory deriving superconductivity from a doped Mott insulator should contain features similar to those discussed here. Then the existence of complex inhomogeneities for compounds having universal phase diagrams suggests that (i) properties of superconductors in the underdoped region, and near magnetic vortices and impurities, are largely determined by critical dynamical symmetry, and (ii) inhomogeneity is a strong diagnostic for the mechanism of high-temperature superconductivity, but has little direct relationship with its cause.

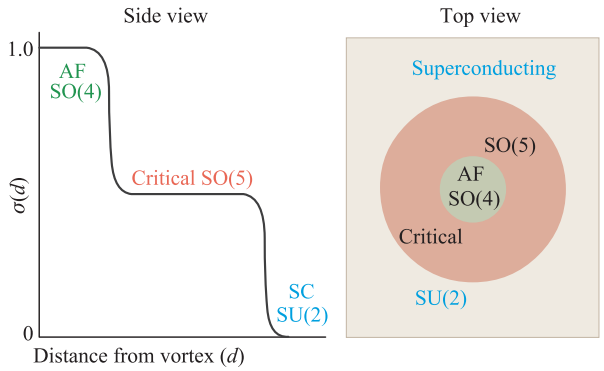



**Fig. 41** Illustration of critical dynamical symmetry in the vicinity of a magnetic vortex.

## 12 The pseudogap and mean fields

Let us now address whether the pseudogap is explicable within a mean-field context. In particular, can the pseudogap be described in terms of competing order when little conclusive experimental evidence supports a distinct phase associated systematically with the pseudogap? The coherent-state approximation is mean-field, in a similar sense as BCS. However, the corresponding mean field is quite rich because it is defined in terms of singlet and triplet fermion (not boson) pairs that are strongly influenced by antiferromagnetic and pairing correlations, and that are constrained to obey SU(4) symmetry. This leads to unusual properties in the resulting solutions. In addition, although the general coherent state solution is mean field, the SU(4) model has exact solutions in three (physically relevant) dynamical symmetry limits that do not correspond to mean-field approximations. These observations have the following implications.

1) The SU(4) theory is in its essentials a model of competing order, since its algebra closes only if there are operators corresponding to AF and pairing modes that enter the theory on an equivalent footing. However, because of the basis (8) of singlet and triplet fermion pairs in the collective subspace, the PG state may be interpreted also in terms of preformed pairs [50]: The Hamiltonian implements a competition between antiferromagnetism and pairing interactions, but the eigenstates in the PG region correspond to rich superpositions of fermion pairs with a

structure dictated by the AF–SC competition.

2) Because the SU(4) coherent state solution is mean field in nature, it does not include the effect of quantum fluctuations on the ground-state wavefunction. However, comparison of the coherent state solution with the three exact SU(4) solutions indicates that in all cases the coherent state energy of the ground state coincides with that of the exact solution, but the wavefunction of the coherent state solution provides only the expectation value of the energy, not the quantum fluctuations.

We may illustrate with two specific examples: (i) The solutions of the coherent state gap equations for a Néel state and for the PG state above $T_c$ are formally equivalent, but (as noted above) they lead to fundamentally different dependence of the corresponding energy surfaces on antiferromagnetism and pairing, implying that the full wavefunction of the two states (including quantum fluctuations) are very different, but that the exact solution still corresponds to a superposition of states described by mean fields having approximate SU(4) symmetry. (ii) The solutions for the coherent-state gap equations for the SU(2) superconducting limit and the SO(5) limit are formally equivalent, but again their energy surfaces are qualitatively different, implying very different roles for quantum fluctuations. This explains why the SU(4) coherent state solution is successful in describing quantities such as energy gaps and the critical temperatures $T_c$ and $T^*$ that are determined primarily by the ground-state energy.

3) The SU(4) energy surfaces provide a guide to the expected quantum fluctuations. The coherent state solutions imply that the ground-state in much of the pseudogap region is inherently unstable against both antiferromagnetic and pairing fluctuations, because many solutions having different combinations of antiferromagnetism and superconductivity become nearly degenerate. Thus, although the exact SU(4) coherent state contains no fluctuations, the form of the solution itself implies large fluctuations in both pairing and antiferromagnetism for states slightly perturbed away from the exact pseudogap solution. This instability (coinciding with broad, flat regions of the energy surface) occurs predominantly in the underdoped pseudogap region, implying unique features for the pseudogap state. For our Néel state solution at half filling and the superconducting solution in optimally doped and overdoped regions the energy surfaces have sharp minima in both the AF and pairing directions, suggesting that these define stable phases with small influence from quantum fluctuations (see Fig. 39). For the underdoped region below $T_c$ our solutions indicate an energy surface that is soft in the AF direction but localized around a finite value of superconducting order, giving a solution that is superconducting, but with the superconductivity modified significantly by AF fluctuations.

4) Thus, although the SU(4) coherent state is a mean field solution, the energy surfaces evaluated in this approximation (and compared with exact many-body solutions in the three symmetry limits) already give a strong guide to the expected role of quantum fluctuations around the mean-field solution. An extended theory including fluctuations associated with the underdoped instability described above may be expected to produce a PG state that is more complex than in the mean-field limit.

This mean field plus quantum fluctuations PG state would be a superposition of SU(4) states having expectation values near zero for SC and AF order parameters, but with large fluctuations in both, and strong susceptibility to external perturbations. The transition to this pseudogap state would be expected to exhibit a rapid evolution in properties instead of a clear phase transition. The resulting theory should retain the correct predictions of Fig. 32. However, quantum fluctuations of the PG state, coupled with the predicted extreme sensitivity to perturbations, would set the stage for a complex set of phenomena (Nernst vortex fluctuations, spatial inhomogeneity, giant proximity effects, and other emergent properties), all described within a framework unifying the competing order and preformed pairs pictures of the pseudogap.

## 13 Anisotropy of the pseudogap

For normal superconductors the Fermi surface is key to the understanding superconductivity. In optimally doped and overdoped high-$T_c$ compounds a "normal" Fermi surface exists and BCS theory utilizing $d$-wave singlet hole pairs seems applicable [2]. But in underdoped cuprates PG states (lying between $T^*$ and $T_c$) have anomalous Fermi surfaces, with indications that the large Fermi surface found in the overdoped region has been reduced to arc-like vestiges, or to small pockets in $k$-space.

### 13.1 Fermi arcs and magnetic quantum oscillations

Angle-resolved photoemission spectroscopy (ARPES) [101, 102] probes electronic properties [101, 103] of states in high-temperature superconductors. These data suggest decreased state density near the Fermi energy for temperature $T < T^*$ that is anisotropic in momentum and has a strong temperature dependence. A full Fermi surface is observed for $T > T^*$; as the temperature decreases below $T^*$, only arcs centered on the $d$-wave nodal lines survive [104–106], scaling as $T/T^*$ to zero length as $T$ tends to zero [106]. Thus gapping is anisotropic in momentum space, with ungapped Fermi surface surviving only as temperature-dependent *Fermi arcs*.

The validity of the Fermi arcs interpretation of ARPES measurements has been challenged by measurements of Shubnikov–de Haas and de Haas–van Alphen effects that probe the cuprate Fermi surface directly [107–111]. These magnetic quantum oscillation experiments indicate that the Fermi surface in the underdoped region consists, not of disconnected arcs, but of small closed regions ("pockets").

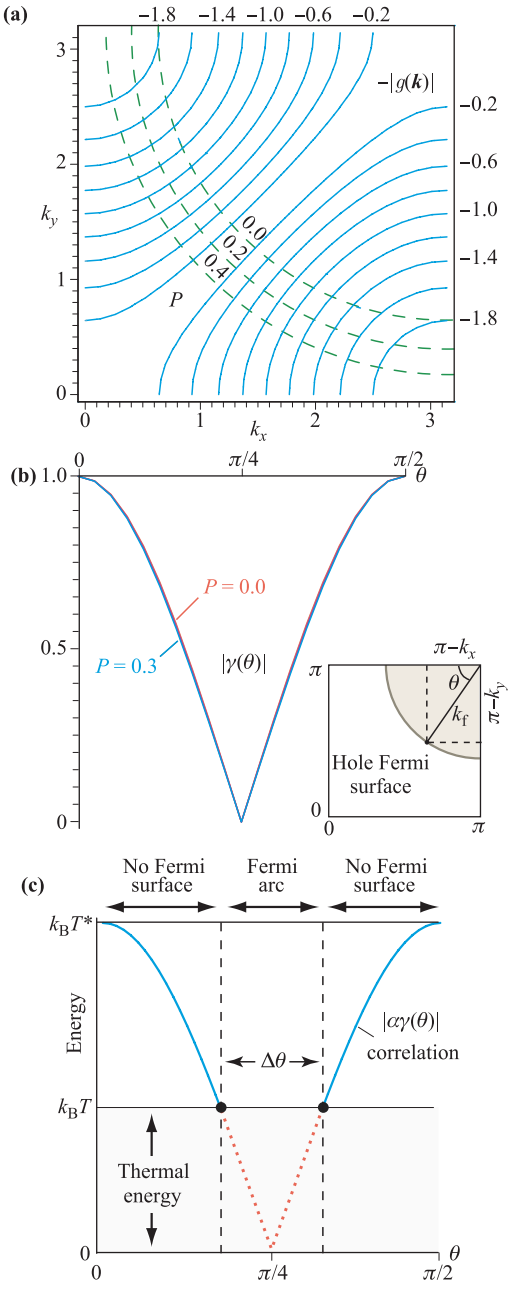



**Fig. 42** The $\boldsymbol{k}$-anisotropic factor, pseudogap correlation energy, and graphical Fermi-arc solution, assuming a hole-like Fermi surface centered on $(k_x, k_y) = (\pi, \pi)$. **(a)** The $\boldsymbol{k}$-anisotropic factor $|g(\boldsymbol{k})|$ (solid blue lines) and contours of equal hole doping $P$ (dashed green lines), in the first Brillouin zone. **(b)** Correlation energy $\alpha\gamma(\theta)$ for unit $\alpha = kT^*$ evaluated along the Fermi surface [curves of constant $P$ in part (a)] as a function of the angle $\theta$ (defined in inset). Curves corresponding to doping $P = 0 - 0.3$ are almost coincident, indicating that $\gamma(\theta)$ is insensitive to doping. **(c)** Graphical solution for Fermi arcs. The curve defines the PG correlation energy and horizontal lines correspond to constant thermal energy scales $k_BT$. Their intersections (black dots) represent points in the momentum space where the pseudogap is just closed by thermal fluctuations; these bracket arcs $\Delta\theta$ of surviving Fermi surface.

These measurements determine the area of the Fermi surface rather precisely and indicate that in the underdoped cuprates the Fermi surface is small, though they cannot determine the location of the Fermi surface in momentum space. This differs substantially from the large Fermi surface found for ordinary superconductors and for overdoped cuprate samples. It is generally expected that the state seen by quantum oscillation experiments under high magnetic field conditions corresponds to the normal state from which the superconductivity develops.

In this section we shall show that SU(4) symmetry for the cuprate superconductors implies that

1) The correlation energy associated with opening the pseudogap necessarily has a strong angular dependence in the $k$-space.

2) This momentum-space anisotropy leads to strong temperature-dependent and doping-dependent restrictions on regions of the Brillouin zone where ungapped Fermi surface can exist.

We shall demonstrate explicitly that this prescription yields a description of the length of Fermi arcs as a function of temperature that is in quantitative agreement with the ARPES results of Ref. [106], if one assumes as a starting point a full hole Fermi surface. Conversely, if one assumes the Fermi surface to correspond to the closed pockets suggested by quantum oscillation experiments, we shall argue that our results place strong restrictions on possible location and size of those pockets.

### 13.2 Momentum-dependent SU(4) and the pseudogap

The appropriate version of the SU(4) theory to address the questions of this section is the $SU(4)_k$ model discussed in Section 6.9 and Refs. [52, 55], which uses momentum-dependent generators of the algebra. All results of the $k$-independent SU(4) model [47, 49–51], including the $k$-averaged results in Fig. 32, are recovered as a special case of the $SU(4)_k$ model for observables dominated by contributions from near the Fermi surface ($\tilde{k} = k_f$) and averaged over all $\boldsymbol{k}$ directions. However, general solutions of the $SU(4)_k$ model give new $k$-anisotropic properties, one of which is that the temperature for the pseudogap closure $T^*(\boldsymbol{k})$ becomes anisotropic in $\boldsymbol{k}$:

$$T^*(\boldsymbol{k}) \equiv T^*(k_f, \theta) = \left| \frac{g(k_f, \theta)}{g_0(k_f)} \right| T^*, \tag{102}$$

where $T^*$ is the gap closure temperature measured by ARPES along the antinodal ($\theta = 0, \pi/2$) direction. See Ref. [52] for details. The physical reason for this temperature and doping dependence of the pseudogap closure temperature is that the pairs interacting by AF interactions in the SU(4) PG state each carry a $g(\boldsymbol{k})$ formfactor. This introduces a $\boldsymbol{k}$ dependence in the effective AF coupling, and thus in $T^*$. The behavior of $g(\boldsymbol{k})$ as a function of $(k_x, k_y)$ is illustrated in Fig. 42(a).

The strong dependence of the pseudogap temperature $T^*$ on $\boldsymbol{k}$ that is implied by Eq. (102) and Fig. 42(a) represents a theoretical prediction with potential observational implications that is independent of the experimental controversy over whether the Fermi surface of underdoped cuprates consists of small close pockets or disconnected arcs. Nevertheless, we now examine the possible implications of this finding for the competing Fermi arc and magnetic oscillation pictures of the cuprate Fermi surface at low hole doping.

### *13.2.1 Implications for Fermi arcs*

In $g(\boldsymbol{k})$ the components $(k_x, k_y)$ or $(k_\mathrm{f}, \theta)$ are constrained by the Fermi surface ($\tilde{k}^2 = k_\mathrm{f}^2$). A more realistic shape for the Fermi surface will be considered below but if we assume for illustration an isotropic hole surface centered around $(\pi, \pi)$ [dashed lines in Fig. 42(a) and inset to Fig. 42(b)], then

$$(\pi - k_x)^2 + (\pi - k_y)^2 = k_\mathrm{f}^2 = 2\pi(1 + P).$$

For a given doping $P$ (thus $k_\mathrm{f}$) and temperature $T$, with $T^*(\boldsymbol{k}) = T$ by virtue of Eq. (102) under the above constraint, we obtain the angles $\theta_1$ and $\theta_2$ at which the PG closes [the angle $\theta$ is defined in the inset to Fig. 42(b)], and the length of the surviving Fermi arc is $k_\mathrm{f}|\theta_2 - \theta_1| \equiv k_\mathrm{f}\Delta\theta$.

This result may be interpreted graphically. In Fig. 42(b) we show $\gamma(\theta) \equiv |g(k_\mathrm{f}, \theta)/g_0(k_\mathrm{f})|$ versus angle $\theta$ along different Fermi surfaces [dashed lines in Fig. 42(a)]. We see that $\gamma(\theta)$ is almost independent of doping $P$. Thus, solving Eq. (102) with the Fermi surface constraint is equivalent to solving

$$k_\mathrm{B}T^*(\boldsymbol{k}) = \alpha\gamma(\theta), \quad \alpha = k_\mathrm{B}T^*, \tag{103}$$

with $k_\mathrm{B}$ the Boltzmann constant. The PG correlation energy $\alpha\gamma(\theta)$ depends on the $\boldsymbol{k}$ direction $\theta$ [Fig. 42(c)]. The pseudogap closes when the thermal excitation energy $k_\mathrm{B}T$ is comparable to the pseudogap correlation energy. The intersections of horizontal lines of fixed $k_\mathrm{B}T$ with the correlation energy curve [heavy black dots in Fig. 42(c)] define Fermi-arc solutions $\theta_1$ and $\theta_2$. Outside those points (solid part of the curve), the correlation energy is larger than the energy of thermal fluctuations (shaded region), the pseudogap opens, and the Fermi surface is gapped. Inside these points (the dotted part of the curve), thermal energy exceeds the correlation energy, the PG closes, and the Fermi surface exists in an arc $\Delta\theta$ between the dots. When $T > T^*$, the PG is closed in all directions and there is a full Fermi surface since $k_\mathrm{B}T > \alpha$, and $\alpha$ is the maximum pseudogap correlation energy.

For a given temperature $T$, the arc solution that is exemplified graphically in Figs. 42(b) and (c), or algebraically in Eq. (102), implies an anisotropy of $\gamma(\theta)$ that partitions the $k$-space uniquely into regions that can have a Fermi surface [$T > T^*(k_\mathrm{f}, \theta)$] and regions that cannot [$T < T^*(k_\mathrm{f}, \theta)$]. As Fig. 42(c) suggests, arc lengths decrease with $T$ at fixed doping, with a full Fermi surface at $T = T^*$ but only the nodal points at $T = 0$. Doping dependence enters primarily through the maximum pseudogap temperature $T^*$ and the Fermi momentum $k_\mathrm{f}$; the temperature dependence enters through $T/T^*$. For fixed $T$, arcs shrink toward the nodal points with decreased doping because $T^*$ in $T/T^*$ increases at smaller doping (Fig. 32). However, if the length of Fermi arcs is measured by the fraction $\Delta\theta/(\pi/2)$ and the temperature is scaled by $T^*$, the weak doping dependence of $\gamma(\theta)$ ensures that $\Delta\theta/(\pi/2)$ versus $T/T^*$ is rather universal, almost independent of both doping and compound.

### *13.2.2 Temperature dependence of Fermi arcs*

In Fig. 42 we show experimentally-determined fractional arc lengths versus $T/T^*$ for Bi-2212 [106]. The theoretical solution for the fractional arc lengths [obtained from Fig. 42(c) or Eq. (102)] is the solid curve in Fig. 43. The agreement between theory and data is rather good, given that the theoretical curve has no adjustable parameters (it is determined completely by the parameters fixed previously in Fig. 32) and that the theory *predicts* the scale $T^*$ implicit in the data (Fig. 32). Notice that the rapid drop in arc length for decreasing $T \sim T^*$, transitioning to a linear decrease to zero arc-length at $T = 0$, is explained entirely by geometry in Fig. 42(c).

The dashed curve in Fig. 43 repeats the analysis using the flatter Fermi surface shown inset lower right. The similarity of dashed and solid curves indicates that solutions depend only weakly on differing curvature in nodal and antinodal regions. Absolute arc lengths depend on doping and temperature, but the scaled arc lengths of Fig. 43 *are*

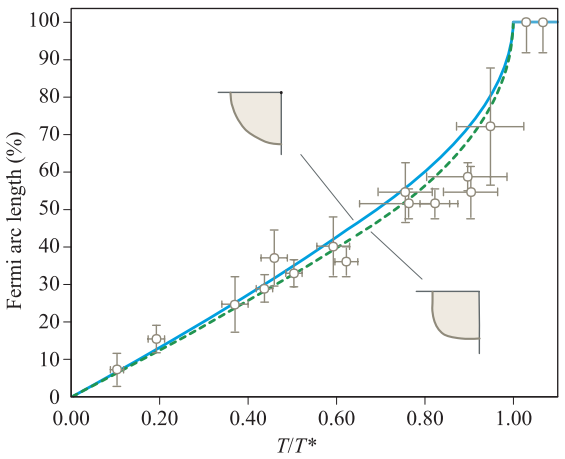



**Fig. 43** Fermi arc length vs. temperature. Experimental arc length is displayed as a percentage of full Fermi surface length vs. $T/T^*$ for underdoped Bi2212 [106]; the solid curve is our prediction for an isotropic Fermi surface (inset upper left); the dashed curve assumes a Fermi surface with flatter antinodal segments (inset lower right). No parameters (beyond those already fit to gaps in Fig. 32) were adjusted, and the curves are almost independent of doping.

*near-universal functions only of the ratio* $T/T^*$, largely independent of compound, doping, and Fermi surface details, as suggested by data.

#### *13.2.3 Discriminating among theories for Fermi arcs*

The preceding discussion shows that a viable theory of Fermi arcs must make two correct predictions: (1) the scale $T^*$ and its doping dependence, and (2) that $T^*(\boldsymbol{k}) \propto \gamma(\theta)$. *Any theory* having a $T^*(\boldsymbol{k})$ consistent with Eq. (103) can describe the scaled data of Fig. 43, if $T^*$ is taken from data. Scaled ARPES data (Fig. 43) test whether a pseudogap has a nodal structure similar to that of the superconducting state. But discriminating among different theories meeting this condition requires a quantitative, self-consistent description of Fermi-surface disappearance and doping dependence of the PG temperature scale $T^*$ (Fig. 43). Thus, only a highly-restricted set of models can be consistent with the aggregate properties of Fermi arcs.

#### *13.2.4 Implications for Fermi surface pockets*

Finally, let us consider the case where the Fermi surface for underdoped cuprates is interpreted in terms of small pockets instead of Fermi arcs, as is favored by magnetic quantum oscillation experiments. This discussion must necessarily be more qualitative, since quantum oscillation experiments cannot localize the inferred small pockets definitively in momentum-space.

We hypothesize that cuprate normal states deviate from Fermi liquid behavior primarily because of correlations producing the pseudogap, which in the present model represents fluctuating AF correlations in a singlet and triplet pair basis. Thus suppression of AF correlation in cuprates should lead to a normal Fermi surface and BCS-like superconductivity. There are no AF correlations beyond the critical doping $P_c \simeq 0.18$, consistent with experiments suggesting that SC is ($d$-wave) BCS-like with a Fermi liquid normal state. Below the critical doping the AF correlations compete with pairing in general, but the present analysis suggests the additional feature that this competition has strong angular localization in the $\boldsymbol{k}$-space.

The actual Fermi surface of the physical system depends on how interactions rearrange the occupied and unoccupied orbitals of the non-interacting system. However, the preceding arguments indicate that, independent of details, at low doping the momentum space available to form a normal Fermi surface becomes small and restricted to limited regions of $\boldsymbol{k}$-space by the angular formfactors displayed graphically in Fig. 42.

Thus, we expect large conventional Fermi surfaces near critical doping and beyond, but at low doping the strong anisotropy of the pseudogap correlations severely restricts the volume of $\boldsymbol{k}$-space available to produce Fermi pockets, or Fermi surfaces of any form. This argument also suggests that any ungapped Fermi surface is likely to lie near the nodal region in either the ARPES or quantum oscillation interpretations. This is consistent with the conjecture of Ref. [112] for location of the Fermi pockets. Hence, the most important implication of both Fermi arc and quantum oscillation data may be that they provide comprehensive evidence for localized, anisotropic correlations responsible for producing the pseudogap, independent of detailed interpretation.

### 13.3 Summary: Anisotropy, arcs, and pockets

We have demonstrated that an SU(4) solution for cuprate superconductivity implies a strong $\boldsymbol{k}$-space anisotropy and corresponding localization for correlations associated with the pseudogap, leading to a pronounced $\boldsymbol{k}$-dependence in the pseudogap temperature $T^*$. We believe this to reflect general properties expected for any realistic model of the cuprates, and to have significant consequences for gapping of the cuprate Fermi surface, particularly in the underdoped region.

If the Fermi surface in the underdoped region is interpreted in terms of Fermi arcs, the $SU(4)_k$ theory reproduces quantitatively the observed variation of the length of the arc segments with temperature, with no parameter adjustment. This suggests that ARPES experiments are measuring quantities related directly to the anisotropic structure of the pseudogap correlations. We can make less definitive statements about the Fermi surface pockets favored by quantum oscillation experiments because the proposed small pockets have not been localized in $\boldsymbol{k}$-space. However, our results suggest that an ungapped Fermi surface of *any type* is unlikely to survive the pseudogap correlations as the temperature is lowered below $T^*$, except in increasingly localized nodal regions of $\boldsymbol{k}$-space.

Thus, our analysis suggests that ARPES and quantum oscillation experiments are seeing (perhaps different aspects of) physics associated with strong $\boldsymbol{k}$-space anisotropy of the pseudogap correlations in underdoped cuprates. The essential physics of Fermi arcs and quantum oscillation results for underdoped cuprates may lie, not in the underlying Fermi surface itself, but rather in how that Fermi surface is modified by the anisotropic pseudogap correlations.

## 14 The iron-based superconductors

> "SU(4) symmetry, not $d$-wave pairing …is the ultimate cause of cuprate behavior, implying that systems could exist having non-$d$ pairing but cuprate-like dynamics …[This] prediction may be tested by searching experimentally for compounds having pairing structure other than $d_{x^2-y^2}$ that still satisfy the SU(4) algebra."
>
> *M. W. Guidry, Y. Sun, and C.-L. Wu (2004)* [49]

In Ref. [49] we proposed that cuprate superconductivity was characterized by more complex behavior than normal BCS superconductivity because the symmetry structure

associated with the superconductivity was non-abelian. Therefore, we termed this new form of superconductivity *non-abelian superconductivity,* and we predicted that new forms of superconductivity having many of the characteristics of the cuprates but not necessarily with *d*-wave formfactors were possible. In this section we shall provide evidence that the new class of high-temperature superconductors based on the iron arsenides and selenides that was discovered beginning in 2008 [3, 5, 113–118] represents the first examples of the new non-abelian high-temperature superconductors predicted in Ref. [49].

### 14.1 Non-Abelian superconductors

Non-abelian superconductivity differs from conventional superconductivity in the richness of the pair structure for condensed states and in the appearance of competing sources of long-range order [49]. The key issues for SU(4) non-abelian superconductivity are that coherent pairs are formed by electrons or holes on adjacent or nearby sites (rather than the same site), so that both singlet and triplet pair states can in principle contribute, and that alternative long-range order enters with the same standing as the superconductivity.

In contrast to BCS superconductivity, which is described by a single dynamical symmetry chain having only abelian subgroups [SU(2) ⊃ U(1)], the minimal symmetry consistent with cuprate data is SU(4), which has a much richer structure (three dynamical symmetries having non-abelian subgroups and differing fundamentally in their properties). We propose that the differences in observational characteristics for these two types of superconductivity originate in this difference in dynamical symmetry structure.

The important role of SU(4) symmetry in non-abelian cuprate superconductivity that we have documented in this review suggests that any pairing structure leading to the SU(4) algebra entails dynamics similar to that of cuprates. Therefore, *d*-wave symmetry of the pairs need not be critical to non-abelian superconductivity in general and SU(4) superconductivity in particular. Pairs with any internal symmetry could exhibit SU(4) superconductivity (or a similar form of superconductivity based on other non-abelian groups) if the no-double-occupancy constraint is valid and correlations can form bondwise (not on the same site) pairs.

For example, the operator $c^\dagger_{\bar{\boldsymbol{r}},i}$ defined in Eq. (78) specifically for cuprate *d*-wave pairs may be generalized to

$$c^\dagger_{\bar{\boldsymbol{r}},i} = \sum_{\boldsymbol{t}} g(\boldsymbol{t}) c^\dagger_{\boldsymbol{r}+\boldsymbol{t},i}, \quad \sum_{\boldsymbol{t}} |g(\boldsymbol{t})|^2 = 1, \tag{104}$$

where $\boldsymbol{t}$ is a few finite lattice displacements of $\boldsymbol{r}$ and $g(\boldsymbol{t})$ is the form factor. The structure (78) of the *d*-wave pairs is only a special case of (104) with

$$\boldsymbol{t} = \pm a, \pm b, \quad g(\pm \boldsymbol{a}) = \frac{1}{2}, \quad g(\pm \boldsymbol{b}) = -\frac{1}{2}.$$

Different internal symmetries of the pairs lead to different forms of $g(\boldsymbol{t})$, but they all can satisfy the condition (79) under no double occupancy and thus preserve the SU(4) algebra and the Hamiltonians implied by its dynamical symmetry chains.

### 14.2 Extending SU(4) to iron-based superconductors

SU(4) symmetry provides a comprehensive understanding of the cuprate superconductors. We shall now show that a quantitative extension of this approach to Fe-based SC requires only that (i) the relevant collective degrees of freedom are superconductivity and antiferromagnetism, and (ii) the superconductivity involves bondwise (not onsite) pairing. Let us begin by listing some of the similarities and differences between these two classes of high-temperature superconductors.

#### *14.2.1 Cuprate and Fe-based phenomenology*

There are many similarities between cuprate and Fe-based superconductors. (i) The superconductivity in both seems to be unconventional, involving strong electron correlations. (ii) The superconductivity in both seems to be in close proximity to antiferromagnetism and there is strong evidence for competing order. (iii) The superconducting states in both appear to involve singlet pairing.

On the other hand, there are some substantial differences between cuprate and Fe superconductors. For example, (i) The parent state in the cuprates is an AF Mott insulator; the parent state in the Fe-based compounds is a (poor) metal, though it may be near a Mott transition. (ii) The antiferromagnetism in the Fe-based compounds differs from that of the cuprates. (iii) There are multiple bands near the Fermi surface in Fe-based compounds, implying the possibility of more complex multiband physics for the gaps. (iv) Because As atoms are out of plane in Fe-based compounds, next-nearest neighbor interactions may be as important as nearest-neighbor ones. (v) The superconductivity in Fe-based compounds is more often influenced by pressure, and there is a larger variety of SC with both hole and particle doping in the Fe compounds than for the cuprates. (vi) Cuprate SC is highly 2-D; there is more evidence in Fe-based SC for 3D (*c*-axis) effects.

Prior to the discovery of Fe-based SC, many had thought that the Mott insulator parentage of cuprate superconductors was critical to the mechanism of high-temperature superconductivity. Difference (i) above calls this into question since parent compounds for the Fe-based high-temperature superconductors are typically metals (albeit poor ones), not Mott insulators. We now argue that the essential point is not Mott behavior but rather SU(4) symmetry, which is highly compatible with a Mott insulator parent state for the cuprates, but can also produce a superconductor from a parent state that need not be a Mott insulator, as for Fe-based SC.

### 14.2.2 Mott insulator versus poor-metal parents

Large onsite Coulomb repulsion leads to a Mott insulator normal state for cuprates. In Ref. [49] we demonstrated that suppression of double site occupancy for pairs in the real space is a sufficient condition to guarantee that the minimal closed algebra is SU(4). The situation in the Fe-based compounds is less clear. The onsite repulsion $U$ must lie in an intermediate range between no correlations and the strong onsite repulsion found for the cuprates, for if $U$ were too large the parent states would develop a charge gap and be good insulators and if it were too small the parent states would be good metals. Most are in fact poor metals, suggesting an intermediate range of $U$.

The presence of strongly-competing antiferromagnetism and superconductivity ensures that Fe-based superconductors correspond to a non-abelian symmetry [49], but the non-abelian algebra need not be the SU(4) algebra found for the cuprates. The minimal algebra for the iron arsenide superconductors depends on whether the pairing leading to the superconductivity involves onsite pairs. If we simplify by restricting attention to a single kind of nearest neighbor or next nearest neighbor bondwise pair, the arguments of Ref. [49] indicate that the minimal closed algebra is SU(4) if there are only bondwise (no onsite) pairs and they don't overlap spatially. (That is, the wavefunction is a superposition of bondwise pairs where no lattice site is occupied significantly by particles from two different pairs.)

For the cuprate superconductors, strong onsite repulsion opens a large energy gap between the bondwise and onsite pairs, implying that bondwise pairs dominate the ground state at low temperature (see Fig. 19). The iron arsenides have onsite repulsion of intermediate strength relative to the cuprates and the situation is less clear. However, the issue is not whether the onsite repulsion suppresses double occupancy in general (and thus produces a Mott insulator), but only whether the correlations are sufficient to push onsite collective pairs to substantially higher energy than bondwise collective pairs (which could be compatible with a poor-metal).

The key distinction is between the coherent pairs of the SU(4) symmetry-truncated basis (which are responsible for charge transport in the superconducting state) and additional valence particles that are not part of the coherent pairs and and may contribute to charge transport in the normal state at zero temperature. The SU(4) Lie algebra respects the Mott insulator characteristics of the cuprate normal states in that it represents optimal configurations for competing AF and SC in the presence of strong onsite repulsion. However, the SU(4) solution in general *need not correspond to an insulator,* since the (normal-state) charge transport properties may be strongly influenced by the properties of particles not in the collective pairs.

For example, the onsite repulsion could be sufficiently strong to make it energetically unfavorable either to form onsite collective pairs or to have double site occupancy by unpaired particles, in which case the symmetry is SU(4) and in addition the material would be expected to be insulating in the normal state. This is representative of the situation in the cuprates. But it could also be the case that somewhat weaker same-site repulsion strongly disfavors onsite pairs over bondwise pairs, but does not forbid some charge transport by unpaired particles in the normal state. The resulting superconducting material would again be described by SU(4), but now is expected to be a metal or poor metal in the normal state. This situation may be representative of the FeAs compounds.

If that is the case, a minimal low-energy theory may be constructed using only bondwise pairs and the algebra of that theory will be SU(4) if these pairs do not overlap on the spatial lattice [49]. Since numerical calculations [119, 120] indicate that the dominant pairing channels in the FeAs compounds involve nearest neighbor or next nearest neighbor *bondwise pairing*, we conclude that a minimal description of their superconductivity corresponds to the same SU(4) symmetry as was found for the cuprates, though values of the effective interaction parameters would likely differ from those of the cuprates.

### 14.2.3 An SU(4) model for iron-based superconductivity

The iron superconductors exhibit a different form of antiferromagnetism than that found in the cuprates, as illustrated in Fig. 44. Following the hint of a large amount of data, we shall assume that in the iron superconductors the superconductivity and antiferromagnetism are related, and that the former develops out of the latter through

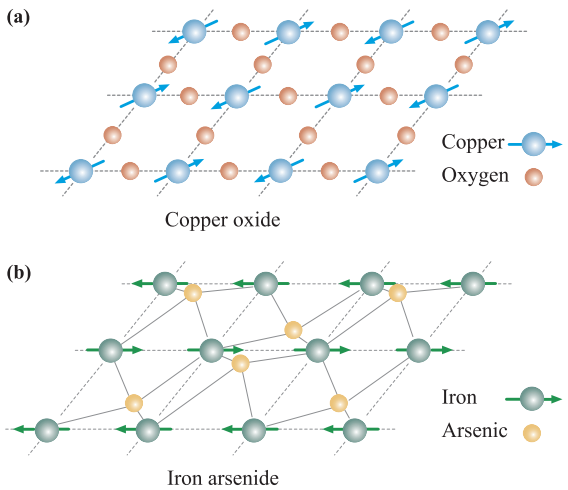



**Fig. 44** Schematic spin structure for **(a)** cuprate and **(b)** FeAs compounds. The undoped iron arsenides are characterized by a "stripe antiferromagnetism" magnetic structure that differs from the AF observed in the cuprates. The oxygen atoms are generally in the same plane as the copper atoms in the cuprates, but the arsenic atoms are located above and below the plane of the iron atoms in the iron arsenides.

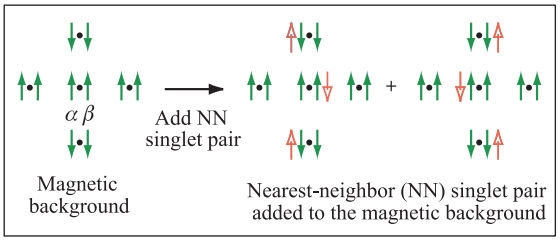



**Fig. 45** Possible SU(4) nearest neighbor (NN) spin-singlet pair structure in a 2-orbital model for a pair centered on a particular site. Arrows indicate spin-up and spin-down particles, with an arrow to the left of a lattice point signifying a particle in orbital $\alpha$ and an arrow to the right of a lattice point signifying a particle in orbital $\beta$. Solid green arrows represent the undoped magnetic background state. The open red arrows represent the added pair. The collective SU(4) NN pair would then correspond to a coherent sum over the lattice of such pairs centered on individual lattice sites.

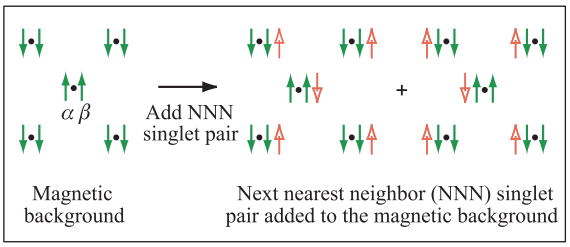



**Fig. 46** As for Fig. 45 but for a possible SU(4) next nearest neighbor (NNN) spin-singlet pair structure in a 2-orbital model for a pair centered on a particular site.

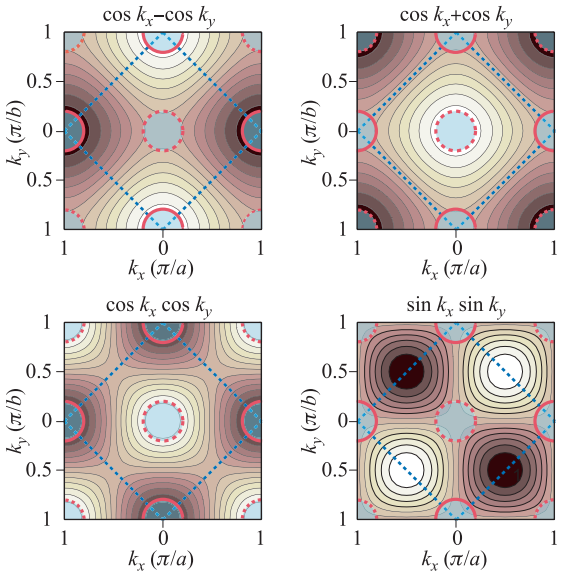



**Fig. 47** Momentum-space formfactors for some cases listed in Table 4. The outer black square for each diagram is the large Brillouin zone associated with the Fe-only real-space lattice. The dashed blue diamond inset in each box is the small Brillouin zone associated with the true real-space lattice. Very schematic locations for the small Fermi surface pockets obtained from typical calculations are sketched as heavy red curves: solid for regions with electron pockets and dashed for regions with hole pockets.

**Table 4** Some pairing gap orbital symmetries, whether they satisfy Eq. (82) and thus close the SU(4) algebra for Fe-based and cuprate compounds, and maximum doping fraction $P_f$ for allowed FeAs symmetries.

| $g(\boldsymbol{k})$ | Fe-based | Cuprate | $P_f$ |
|---|---|---|---|
| $s_{x^2+y^2} = \cos k_x + \cos k_y$ | No | Yes | – |
| $d_{x^2-y^2} = \cos k_x - \cos k_y$ | No | Yes | – |
| $s_{x^2y^2} = \cos k_x \cos k_y$ | Yes | Yes | 1/3 |
| $d_{xy} = \sin k_x \sin k_y$ | Yes | Yes | 1/3 |
| $s_{x^2+y^2} \pm d_{x^2-y^2}$ | Yes | Yes | 2/3 |
| $s_{x^2+y^2} \pm i d_{x^2-y^2}$ | No | Yes | – |

doping or pressure. Because in the FeAs superconducting compounds the arsenic atoms are out of the Fe plane, general arguments suggest that next nearest neighbor (NNN) lattice interactions may compete favorably with nearest neighbor (NN) interactions. Construction of NN singlet pairs by addition to the magnetic background corresponding to the FeAs compounds exhibited in Fig. 44 is illustrated schematically in Fig. 44, and the corresponding construction of an NNN singlet pair is illustrated in Fig. 46.

By our previous arguments, if we assume that the superconductivity in the Fe-based superconductors involves singlet Cooper pairs interacting with antiferromagnetism, charge, and spin, the corresponding emergent symmetry will be SU(4), just as in the cuprates, provided that (i) the pairs are bondwise and not onsite pairs, (ii) electron correlations are strong enough to suppress overlap of pairs (suppression of double site occupancy in the collective pairs), and (iii) the condition (82) is satisfied for the pairing formfactor. The requirement that $g(\boldsymbol{k})$ satisfy (82) places immediate constraints on its possible form. In Table 4 we apply these constraints to some gap symmetries that have been proposed for the FeAs superconductors, indicating whether the SU(4) algebra closes for each case. For reference, we carry out the same procedure for the cuprates, though in that case we already know experimentally that the gap symmetry is $d_{x^2-y^2}$. Some of the formfactors that are tested in Table 4 are also illustrated in Fig. 47.

From Table 4 we find that symmetries such as $s_{x^2y^2}$ and $d_{xy}$ can close the SU(4) algebra, but symmetries such as $s_{x^2+y^2}$ and $d_{x^2-y^2}$ do not. Thus, in the simplest symmetry-limit picture, neither $s_{x^2+y^2}$ nor $d_{x^2-y^2}$ are valid orbital symmetries for the Fe-based superconductors because their failure to close the SU(4) algebra indicates that pairing with that geometry is incompatible with the observed magnetic structure for the iron superconductors.

We conclude that a unified SU(4) model of cuprate and Fe-based superconductivity suggests that the corresponding orbital symmetry of the pair gap for Fe-based com-

pounds is likely to differ from the cuprate $d_{x^2-y^2}$ symmetry (in the symmetry limits of the theory). It is also seen from Table 4 that the more symmetric antiferromagnetism of the cuprates (see Fig. 44) is compatible with many possible pairing formfactors (though most seem not be realized physically), but the asymmetric AF of the iron arsenides provides stronger constraints on a compatible pairing structure.

The SU(4) symmetry allows a further constraint to be placed on the orbital formfactor. The coherent pair states for those formfactors allowed for Fe-based SC in Table 4 have a general structure

$$p^\dagger = \sum_{r=\{x,y\}} c^\dagger_{r\uparrow} c^\dagger_{\bar{r}\downarrow}$$

in the real space, where $c^\dagger_{ri}$ is the electron creation operator $a^\dagger_{\boldsymbol{k}i}$ in the coordinate representation. The $c^\dagger_{\bar{r}i}$ for

$$s_{x^2+y^2} + d_{x^2-y^2} = 2\cos k_x, \quad s_{x^2+y^2} - d_{x^2-y^2} = 2\cos k_y$$

pairs are formed from nearest-neighbors,

$$c^\dagger_{\bar{r}i} = 2^{-1/2}(c^\dagger_{(x+a,y)i} + c^\dagger_{(x-a,y)i}), \quad g(\boldsymbol{k}) = 2\cos k_x,$$
$$c^\dagger_{\bar{r}i} = 2^{-1/2}(c^\dagger_{(x,y+a)i} + c^\dagger_{(x,y-a)i}), \quad g(\boldsymbol{k}) = 2\cos k_y,$$

and for $\cos k_x \cos k_y$ or $\sin k_x \sin k_y$ pairs are formed from next-nearest neighbors,

$$c^\dagger_{\bar{r}\downarrow} = \frac{1}{2}(c^\dagger_{(x+a,y+b)\downarrow} + c^\dagger_{(x-a,y-b)\downarrow} \pm c^\dagger_{(x+a,y-b)\downarrow} \pm c^\dagger_{(x-a,y+b)\downarrow}), \tag{105}$$

with (+) corresponding to $\cos k_x \cos k_y$ and (−) to $\sin k_x \sin k_y$. These are illustrated in Fig. 48.

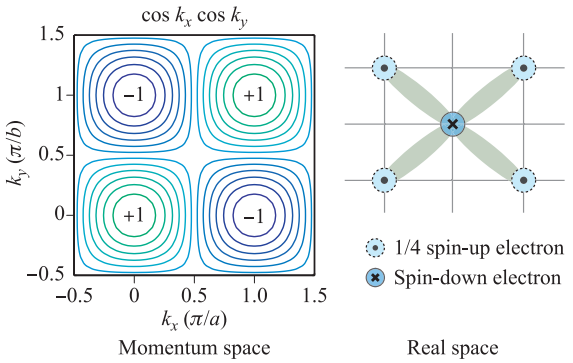



**Fig. 48** Pairing gap corresponding to a $\cos k_x \cos k_y$ formfactor in momentum space and the corresponding schematic real-space pair structure for a singlet electron pair. The shorthand notation "1/4 electron" indicates that the spin-up electron is distributed with equal probability on four next-nearest neighbor sites in the pair wavefunction of Eq. (105). The spatial pair structure for $\sin k_x \sin k_y$ is similar to that for $\cos k_x \cos k_y$, differing only in phases.

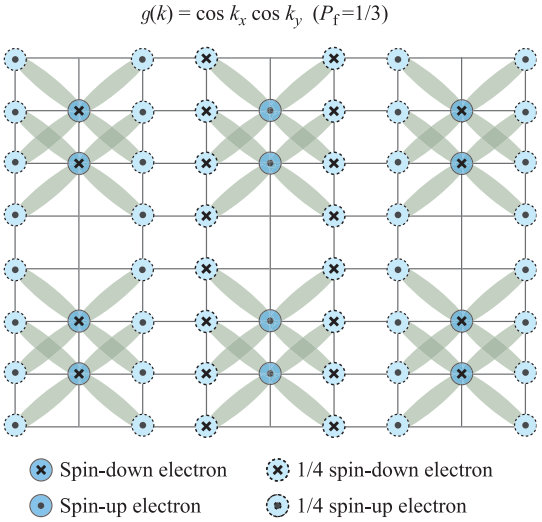



**Fig. 49** Schematic count of maximum pair density consistent with SU(4) symmetry assuming electron-doped material with a singlet $\cos k_x \cos k_y$ pair gap formfactor. For this segment of the lattice, no additional pairs of this structure can be added without causing a finite amplitude for double site occupancy by pairs, which would break SU(4) symmetry. By counting of occupied and unoccupied sites, the maximum fraction of lattice sites that can be occupied by $\cos k_x \cos k_y$ pairs without double occupancy is $\frac{1}{3}$. The realistic wavefunction will be a superposition of such configurations, each with a maximum pair occupancy of $\frac{1}{3}$. The spatial pair structure and maximum doping for $\sin k_x \sin k_y$ is the same as for $\cos k_x \cos k_y$, since they differ only in phases [see Eq. (105)].

As discussed for the cuprates in Section 7.3 and Ref. [49], the SU(4) requirement of no double occupation by pairs implies a lattice occupancy restriction for the superconducting state. Counting the maximum number of pairs that can be placed on the lattice without overlap, as illustrated in Fig. 49, indicates that the largest doping fraction consistent with SU(4) symmetry is $P_f = \frac{2}{3}$ for $\cos k_x$ and $P_f = \frac{1}{3}$ for $\cos k_x \cos k_y$ or $\sin k_x \sin k_y$. These are summarized in the last column of Table 4. Since current data suggest that the superconductivity does not extend much beyond $P_f = \frac{1}{3}$ for most FeAs compounds, this favors $\cos k_x \cos k_y$ or $\sin k_x \sin k_y$, among the allowed orbital symmetries for iron arsenides in Table 4.

### *14.2.4 Multiple pairing gaps in the iron superconductors*

The iron superconductors involve pairs that receive contributions from multiple Fe orbitals. Their Fermi surfaces often correspond to disconnected sheets in the Brillouin zone and this implies that their pairing properties are generally $k$-dependent. A generalization of the SU(4) coherent-state formalism to handle this situation has been described in Section 6.9, with the relevant gap equations given in Eqs. (71) and (72), and the predicted superconducting transition temperature as a function of doping in Eq. (73). We now apply this formalism to an analysis of iron super-

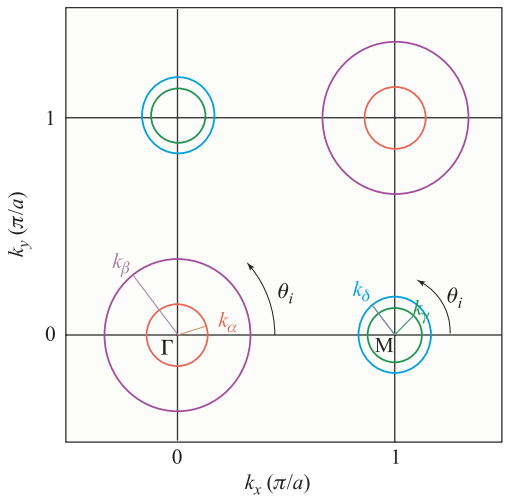



**Fig. 50** Approximate Fermi surfaces. The four Fermi surface pockets, two around the Γ point and two around the $X$ point, are approximated by circles.

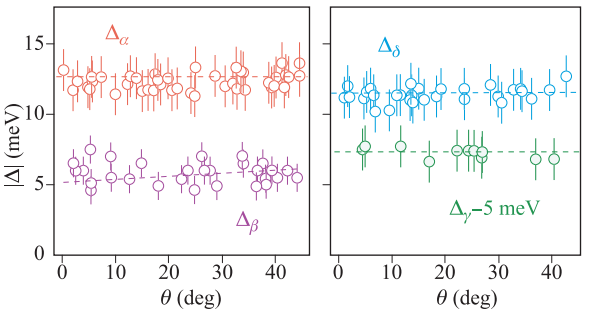



**Fig. 51** Pairing gaps on four sheets of the Fermi surface for $Ba_{0.6}K_{0.4}Fe_2As_2$. Note that $\Delta_\gamma$ is displaced by 5 meV for plotting purposes. Circles are data from Ref. [121] and dashed lines are theoretical using Eq. (106). Parameters $\Delta_0 = 13.5$ meV, $k_\alpha/\pi = 0.135$, $k_\beta/\pi = 0.370$, $k_\gamma/\pi = 0.141$, and $k_\delta/\pi = 0.181$ were determined by fitting to the data.

conductors [54].

In the preceding we argued that requiring (1) closure of the SU(4) algebra and (2) consistency of pairing with the observed antiferromagnetism favors a pairing formfactor $g(\boldsymbol{k}) = \cos k_x \cos k_y$ or $\sin k_x \sin k_y$. First consider $g(\boldsymbol{k}) = \cos k_x \cos k_y$. Assuming the doping $x$ to be less than the critical value $x_q$, Eq. (71a) gives for the singlet pairing gap

$$\Delta_s(\boldsymbol{k}) = \Delta_0 \cos k_x \cos k_y, \quad \Delta_0 \equiv \frac{G_0\Omega}{2\bar{g}}\sqrt{x(x_q^{-1} - x)}. \tag{106}$$

ARPES measurements on $Ba_{0.6}K_{0.4}Fe_2As_2$ find four sheets of Fermi surface within the Brillouin zone [121]. Let us introduce a simple model that assumes the four pockets of Fermi surface (labeled $\alpha$, $\beta$, $\gamma$, and $\delta$) to be spheres centered at the appropriate momentum, with the radii $k_\alpha$, $k_\beta$, $k_\gamma$, and $k_\delta$ determined by fits to ARPES data. Figure 50 illustrates. Then from Eq. (106) the pairing gaps on the four sheets of Fermi surface are given by

$$\Delta_i \equiv \Delta_s(k_i, \theta_i) = \Delta_0 \cos(k_i \cos\theta_i) \cos(k_i \sin\theta_i), \tag{107}$$

where $i = \alpha, \beta, \gamma, \delta$ labels the sheets and the polar angles $\theta_i$ are centered at the Γ and $M$ points (see Fig. 50). Writing this out explicitly for the four cases yields Table 5.

A fit of the parameters $\Delta_0$ and the $k_i$ to the data of Ref. [121] gives the description of the pairing gaps illustrated in Figs. 51–53, where in Fig. 51 circles indicate data and the dashed lines represent the gaps calculated from Eq. (107). We see that the ARPES measurements of Ref. [121] are at least approximately consistent with the $\cos k_x \cos k_y$ pairing gap formfactor that is deduced in the present paper by requiring self-consistency of antiferromagnetism and superconductivity within an SU(4) symmetry.

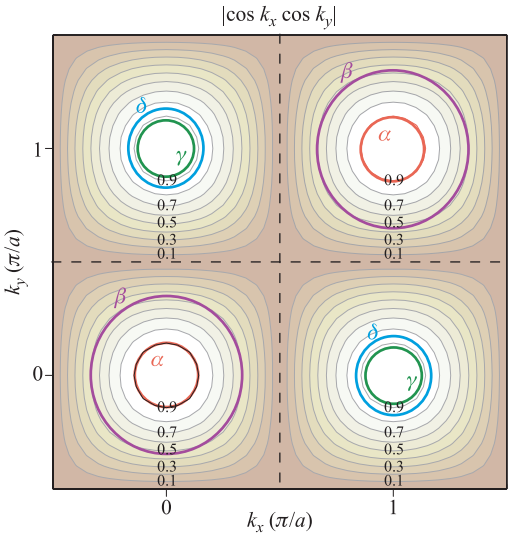



**Fig. 52** Fermi surfaces $\alpha$, $\beta$, $\gamma$, and $\delta$ (thick curves) superposed on contours of the pairing formfactor $|\cos k_x \cos k_y|$. Gap nodes are indicated by dashed lines.

The $\sin k_x \sin k_y$ formfactor compatible with the antiferromagnetism according to Table 4 would not be compatible with the data displayed in Fig. 51, since it would imply nodes on the Fermi surfaces (compare Figs. 47 and 52) that are not observed in the data. Consistency of the observed antiferromagnetism with the superconductivity in the FeAs compounds is possible with either $\cos k_x \cos k_y$ or $\sin k_x \sin k_y$ pairing formfactors, but we find that requiring in addition consistency with the ARPES data of Ref. [121] would restrict to the $\cos k_x \cos k_y$ choice.

### 14.3 Unified cuprate and Fe-based superconductivity

We have presented evidence that the Fe-based high-temperature superconductors represent the second example (after the cuprates) of the non-abelian superconductors proposed in Ref. [49]. The identification of a common non-abelian superconductivity in these two classes of com-

pounds permits a unified model of cuprate and Fe-based superconductors based on an SU(4) group and dynamical-symmetry subgroup chains corresponding to emergent degrees of freedom in the strongly-correlated electron system.

Requiring that the SU(4) algebra simultaneously close under commutation and be consistent with the magnetic structure inferred from neutron scattering experiments constrains the orbital symmetries for the pairing gap in FeAs compounds. We find that in the symmetry limits neither $s_{x^2+y^2}$ nor $d_{x^2-y^2}$ symmetries are compatible with the neutron scattering data but $s_{x^2y^2}$ or $d_{xy}$ could be, and comparing the predicted gaps with ARPES data restricts the choice uniquely to $d_{x^2y^2}$ (that is, $\cos k_x \cos k_y$). Thus, we find that a unified SU(4) model of iron-based and cuprate high temperature superconductivity is possible, but consistency with neutron scattering and ARPES data suggests that the pairing in the two cases corresponds to *different* orbital formfactors at the microscopic level.

There is widespread evidence from ARPES measurements for a relatively universal effective pairing formfactor of approximate $\cos k_x \cos k_y$ form in both 122 and 111 families of FeAs (iron-pnictide) superconductors [121–125]. This is puzzling from a microscopic point of view (see the discussion in Ref. [126]), since there is good reason to believe that all five $d$-orbitals of Fe will contribute in this region and there is little reason to believe that this contribution would be uniform across compounds. As we now discuss, the discovery of the FeSe (iron-chalcogenide) superconductors [5] makes this situation even more puzzling.

**Table 5** Pairing gaps on four sheets of the idealized Fermi surface.

| Label | Fermi surface | Pairing gap |
|---|---|---|
| $\Delta_\alpha$ | $k_x^2 + k_y^2 = k_\alpha^2$ | $\Delta_0 \cos(k_\alpha \cos\theta_\alpha) \cos(k_\alpha \sin\theta_\alpha)$ |
| $\Delta_\beta$ | $k_x^2 + k_y^2 = k_\beta^2$ | $\Delta_0 \cos(k_\beta \cos\theta_\beta) \cos(k_\beta \sin\theta_\beta)$ |
| $\Delta_\gamma$ | $k_x^2 + k_y^2 = k_\gamma^2$ | $\Delta_0 \cos(k_\gamma \cos\theta_\gamma) \cos(k_\gamma \sin\theta_\gamma)$ |
| $\Delta_\delta$ | $k_x^2 + k_y^2 = k_\delta^2$ | $\Delta_0 \cos(k_\delta \cos\theta_\delta) \cos(k_\delta \sin\theta_\delta)$ |

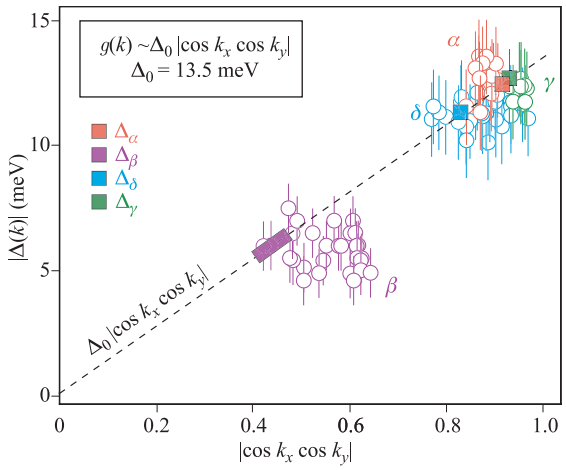



**Fig. 53** Pairing gaps on four sheets of the Fermi surface versus $|\cos k_x \cos k_y|$ for $Ba_{0.6}K_{0.4}Fe_2As_2$. Data from Ref. [121] and the squares and rectangles indicate theoretical values for gaps on the four sheets deduced from Fig. 52. The same parameters as for Fig. 51 were used.

Given the widespread belief that superconductivity in the iron-based superconductors is a consequence of pair binding by electron–electron correlations, an additional minus sign is required in the pairing matrix elements to turn the repulsive electron–electron interaction into an effective attractive one. A nodal gap function permitting pair scattering between nodes of different signs is one possibility, as is believed to be the case for the cuprates. In the FeAs compounds, instead data often suggest that the gap has no nodes, but that there are electron and hole pockets of fermi surface in the Brillouin zone separated approximately by lattice vectors. Thus, in that case it has been proposed that the attraction could come from scattering between these particle and hole pockets (which have gaps of opposite sign). This is the motivation for the extended $s$-wave gap symmetry assumed by many authors for the FeAs compounds [25].

However, data on the chalcogenides show high-temperature superconductivity for electron-doped compounds in which there appear to be only electron and not hole pockets of Fermi surface. This calls into question any mechanism based on scattering between electron and hole pockets as a general explanation of superconductivity in the iron-based compounds. We conclude that it is difficult to justify a unified picture of even the iron-based superconductors (much less unifying the iron superconductors with the cuprate superconductors) based on the standard microscopic approaches.

Unless we are content to assume that pnictide, chalcogenide, and cuprate superconductivity are all due to separate mechanisms, there must be broader symmetries at work than those manifest from the usual microscopic pictures of these compounds. As we have discussed here (and proposed originally in Refs. [49, 54]), any microscopic conditions that lead to the realization of emergent SU(4) symmetry will provide a unified picture of superconductivity across all of these compounds, independent of microscopic details (which change only the values of parameters for the emergent collective modes).

The preceding discussion is another example of a recurring theme of this review: understanding the origin of superconductivity, whether conventional or unconventional, and whether in various occurrences in condensed matter or in manifestations in a variety of other disciplines, lies in understanding the general conditions that lead to the Cooper instability in such a diverse set of systems, not in the detailed microscopic properties of each system. We believe that the iron-based superconductors have been extremely important in this regard because they—perhaps

more so than for any other class of superconductors—have shown that unconventional superconductivity seems to be details are consistent with emergence of superconductivity within a system exhibiting other forms of collective behavior, such as antiferromagnetism. That this should be the case is, of course, central to the point of view presented in this review.

## 15 Relationship with other models

The SU(4) model discussed in this review uses the mathematical tools of Lie algebras and Lie groups, which are in some respects rather different from the mathematical tools most commonly applied to the strongly-correlated electron problem. However, as we have emphasized in many contexts, the resulting physical picture has considerable resonance with many ideas found from application of other more standard condensed-matter methodologies. Formally it is always possible to make such identifications because either the present Lie algebra and Lie group based approach, or more conventional approaches for condensed matter, have as their common final results the calculated matrix elements corresponding to physical observables.

These matrix elements provide a common denominator for comparison, irrespective of the differences in mathematical techniques used to obtain them, as illustrated in Fig. 54. In this section we use such considerations to summarize the relationship of the SU(4) model of nonabelian superconductivity to some other methodologies that have been widely discussed for understanding superconductors.

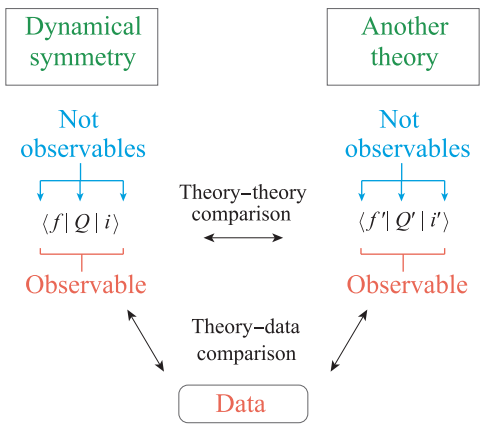



**Fig. 54** Comparison of matrix elements among different theories and data. Wavefunctions and operators are not observables. Only matrix elements are directly related to experimental data. Even though a dynamical symmetry theory and some other theory may use very different methodologies applied to a given problem, they both must produce matrix elements of observables as their physical output. Thus the valid comparisons between theories, and with data, are through matrix elements; wavefunctions and operators separately are relevant only in that they may be helpful padagogically, and that they are tools to produce matrix elements.

### 15.1 SU(4) and BCS models

As we have explained in some depth, the standard BCS description of conventional superconductors may be viewed as the limit of the methodology developed here when (i) other non-pairing order such as antiferromagnetism may be neglected, (ii) the electron–electron correlations are weak, and (iii) as a result of the preceding assumptions bondwise singlet pairs have no energetic advantage over onsite singlet pairs. In that case the SU(4) symmetry is enlarged to SO(8) with the addition of onsite pairs and their interactions, and the physical conditions then favor an SO(8) dynamical symmetry chain ending in the SU(2) pseudospin symmetry for onsite pairs, which is well known to describe a conventional ($s$-wave) BCS superconductor. Under these physical conditions the dynamical symmetry does not specify uniquely the underlying microscopic interaction, but is compatible with a weak phonon-based binding of the Cooper pairs.

However, for a different set of physical conditions the SU(4) formalism leads to another picture that is BCS in the form of the wavefunction, but differs from the preceding example in that the resulting superconductivity may be unconventional. In the limit of strong electron–electron correlations and significant antiferromagnetism the favored highest symmetry is SU(4) because the correlations favor bondwise pairs, and if the AF is sufficiently weakened by doping the favored dynamical symmetry becomes $\mathrm{SU(4)} \supset \mathrm{SU(2)_p}$. By comparison of physical matrix elements, we have shown that this dynamical symmetry has the properties of a singlet-pairing condensate having a BCS-like wavefunction, but a possibly unconventional order parameter since the formfactor $g(\boldsymbol{k})$ need not be conventional $s$-wave. Again, the dynamical symmetry does not specify uniquely the nature of the pair binding, but the context in which it arises suggests strongly correlated-electron rather than phonon binding for the Cooper pairs.

This second form of BCS superconductivity is also evident in the SU(4) coherent-state solutions, since we showed in Section 6.6 that the SU(4) gap equations reduce exactly to the BCS gap equations if the antiferromagnetic correlations can be neglected, but with a formfactor that is possibly unconventional. It is widely believed that the superconducting state in the overdoped cuprates has this BCS but with unconventional formfactor property, and it is clear that both data and the microscopic doping dependence of the SU(4) model favor the suppression of AF correlations and thus the $\mathrm{SU(4)} \supset \mathrm{SU(2)_p}$ dynamical symmetry in this region.

### 15.2 SU(4) and Néel antiferromagnetism

The $\mathrm{SU(4)} \supset \mathrm{SO(4)}$ dynamical symmetry has matrix elements corresponding to the usual picture of a Néel antiferromagnet (see Section 5.1), and for cuprates this state is favored at half filling by the intrinsic doping properties

of the SU(4) solutions. By virtue of the properties for the bondwise pairs of the parent SU(4) symmetry, this antiferromagnetic state is also a Mott insulator at half filling for cuprates, since SU(4) enforces no double occupancy of lattice sites by electrons or holes making up the SU(4) pairs. As we have discussed in Section 11.1, this AF Mott insulator state becomes strongly disfavored when doped away from half filling because it is unstable against condensing Cooper pairs if there is a non-zero effective pairing interaction in the collective subspace.

### 15.3 SU(4) and Mott insulators

It is known that the parent state of a cuprate superconductor corresponds to an antiferromagnetic state with no double occupancy of the lattice, implying Mott insulator character, and there is evidence that the parent states of some iron superconductors lie near a Mott transition. As noted above, the SU(4) symmetry itself implies that fundamentally the lattice is not doubly-occupied because closure of the SU(4) Lie algebra requires that sites contained in the pairs not have double occupancy. Physically, this is because the SU(4) symmetry results from the more general SO(8) symmetry under the assumption that onsite pairs are suppressed by Coulomb repulsion. Thus, Mott character for the parent states of high-temperature superconductors is a natural outcome of the SU(4) dynamical symmetry itself, and requires no additional constraints such as Gutzwiller projection.

Furthermore, one may argue that the Mott character of the cuprate and iron-based parent states is fundamental, in that the subgroup chains (collective states) corresponding to the three possible dynamical symmetries inherent the no double occupancy constraint from the parent SU(4) group. This is in contrast to a Gutzwiller projection of the unperturbed basis, which ensures no double occupancy of the basis, but does not ensure that emergent collective states (antiferromagnetic or superconducting) are described by a basis with no double occupancy. Hence, the SU(4) symmetry represents the consistent implementation of the role of Coulomb repulsion in suppressing double occupancy of lattice sites in highly-collective states. The parent state for a cuprate superconductor is then both an antiferromagnet and a Mott insulator because of *two separate properties* of the overall SU(4) symmetry: (i) the SU(4) symmetry is broken by double occupancy of lattice sites, and (ii) the SU(4) symmetry has an SO(4) subgroup with the matrix elements of a Néel antiferromagnetic state and a doping dependence that favors it as the ground state at zero doping.

### 15.4 SU(4) and resonating valence bond states

As we have already mentioned in Section 11.1.3, the properties of the SU(4) coherent state at low hole-doping presumably share many features with resonating valence bond (RVB) states. Spin-triplet pairs are essential for a complete set of operators in the minimal SU(4) model [for example, no double occupancy is enforced by the SU(4) Lie algebra, which fails to close without triplet pairing operators], and a mixture of singlet and triplet pairs is essential to describe the AF states at half filling in the highly-truncated SU(4) fermion basis. But the significance of triplet pairs relative to singlet pairs is small and decreases rapidly at higher doping, as illustrated in Fig. 28. Therefore, SU(4) ground states could have significant overlap with a singlet spin liquid.

The SU(4) coherent state justifies many features of RVB models, but it is richer than typical RVB applications because it accounts even-handedly for both AF and SC on a lattice with no double occupancy. As a consequence, the SU(4) variational wavefunction is more complex than that of a singlet spin liquid. Conversely, the SU(4) coherent-state model is simpler in many respects than RVB models because superconductivity and antiferromagnetism are accounted for quantitatively in a minimal theory having only (dressed) electron degrees of freedom: the theory requires no explicit introduction of pair bosons, gauge fields, or spinons and holons (which have formal justification in one dimension, but are less obviously justified in higher dimensions, and for which there is little direct evidence in cuprate superconductors).

The SU(4) coherent state represents a minimal extension of the BCS formalism to incorporate $d$-wave pairing in the presence of strong AF correlations and large effective onsite electron repulsion. It requires no Gutzwiller projection because the symmetry enforces no double occupancy on the lattice. It exhibits a type of spin–charge separation (see the discussion in Ref. [127]), but not through topological spinons and holons: in the fermion basis, charge is carried both by singlet fermion hole pairs having a spin of 0 and charge $-2$, and triplet fermion hole pairs having a spin of 1 and charge $-2$, but spin is carried solely by the triplet hole pairs. From this point of view, "spin–charge separation" in emergent degrees of freedom is a fairly mundane consequence of the complete set of operators argument in Section 4.2 and Fig. 8. Antiferromagnetic operators scatter spin-singlet pairs into spin-triplet pairs, and spin-triplet pairs into spin-singlet pairs. Thus, both types of excitations (singlet pairs carrying charge but no spin and triplet pairs carrying charge and spin) must be present as fundamental excitations in the realistic collective subspace if the system exhibits antiferromagnetic correlations.

### 15.5 SU(4) and the Zhang SO(5) model

Ideas having some similarity to those discussed in this review have been proposed by S. C. Zhang and collaborators [128]. To distinguish from the SO(5) dynamical symmetry of the SU(4) model discussed in this review, we shall term this the Zhang SO(5) model. In the Zhang SO(5)

model the AF and SC order parameters are assembled into a 5-dimensional vector order parameter that is rotated between AF and SC order by SO(5) generators. The methodology described here is different, starting instead from identification of a closed algebra associated with a general set of fermion pairing and particle–hole operators defined on a periodic lattice, with order parameters arising as matrix elements of various bilinear forms for these operators rather than being introduced as fundamental entities, and a corresponding truncation of the full Hilbert space to a symmetry-dictated collective subspace. Nevertheless, we find that we recover Zhang's SO(5) symmetry as a subgroup of a more general SU(4) symmetry if various approximations are made in the full SU(4) theory.

Our SO(5) subgroup is embedded in a larger SU(4) group defined microscopically in the fermion degrees of freedom, which implies constraints on the SO(5) subgroup. Our SU(4) model and Zhang's SO(5) model start from the same building blocks (the operator set (1), but see note [129]). However, we implement the full quantum dynamics (the commutator algebra) of these operators exactly, while in Ref. [128] a subset of 10 of the operators acts as a rotation on the remaining 5 operators $\{p^\dagger, p, \boldsymbol{Q}\}$, which are treated phenomenologically as 5 independent components of an order-parameter vector. Thus only 10 of the 15 generators of our SU(4) group are treated dynamically in the Zhang SO(5).

The embedding of SO(5) as a subgroup in our larger SU(4) group has various physical consequences that do not appear if an SO(5) group is considered in isolation.

1) A transition from antiferromagnetism to superconductivity at zero temperature that is controlled by the doping emerges naturally and microscopically from the SU(4) symmetry. The corresponding behavior in the Zhang SO(5) model requires that a symmetry-breaking term proportional to a chemical potential be introduced by hand.

2) The full SU(4) dynamics show that the SO(5) subgroup is only one of the symmetries relevant to the cuprate problem. It is a transitional symmetry that links AF to SC behavior, suggesting that it is most useful for the underdoped region. The AF phases at half filling and the optimally doped superconductors are more economically described by our SO(4) and $SU(2)_\eta$ symmetries, respectively.

3) SU(4) symmetry leads naturally to pseudogap behavior, with the SO(5) subgroup being central to this property.

4) The methodology of the SU(4) dynamical symmetry approach shows that the SO(5) subgroup is an effective symmetry operating in a severely truncated space. It should be interpreted, not in terms of an approximate symmetry of a Hubbard or $t$–$J$ Hamiltonian, but in terms of an exact (emergent) *dynamical symmetry*. Thus, its microscopic validity—as for that of its parent SU(4) symmetry—must be judged by the physical correctness of the matrix elements evaluated in that truncated model space, not by whether a particular Hamiltonian thought to have some relevance for the full space possesses such a symmetry (see the discussion in Section 2).

For exact SO(5) symmetry, antiferromagnetic and superconducting states are degenerate and there is no barrier between them at half filling [see the $n/\Omega = 1$ curve of Fig. 23(b)]. But this is inconsistent with observed Mott insulating behavior at half-filling in the cuprates, because the symmetric Zhang SO(5) model predicts no charge gap at half filling of the lattice. Thus, for antiferromagnetic insulator properties to exist at half filling, it is necessary to break SO(5) symmetry [128]. As we have discussed in Section 8.5, this symmetry breaking is implicit in the SU(4) model, occurring naturally if $\sigma \neq \frac{1}{2}$ in the Hamiltonian (14). Furthermore, SU(4) symmetry implies the constraint

$$\langle p^\dagger p + \boldsymbol{Q}\cdot\boldsymbol{Q} + \boldsymbol{\pi}^\dagger\cdot\boldsymbol{\pi}\rangle = \frac{1-x^2}{4}\Omega^2. \tag{108}$$

This ensures a doping dependence in the solutions that describes the transition from AF to SC in the cuprates, as discussed in Section 7.4.

Hence, the SU(4) coherent state analysis indicates that the phenomenologically-required SO(5) symmetry breaking, and the doping dependence in the solutions, occur as natural consequences in the SU(4) model. They need not be introduced empirically, as proposed in the original Zhang SO(5) model. A projected Zhang SO(5) model was introduced in Ref. [130] that uses Gutzwiller projection to satisfy the large-$U$ Hubbard (no double occupancy) constraint. There is no need to introduce such a projection if SO(5) is treated as a subgroup of SU(4), because the SU(4) symmetry itself already implies a no double occupancy constraint [49] .

We conclude that cuprate high-temperature superconductivity in the underdoped region may be described by a Hamiltonian that conserves SU(4) but breaks SO(5) explicitly in a manner favoring AF order over SC order, as has been discussed in Section 8.5. The weakly-broken SO(5) symmetry acts as a critical dynamical symmetry mediating the transition between superconducting states, described by the $SU(2)_\eta$ dynamical symmetry, and antiferromagnetic states, described by the SO(4) dynamical symmetry. Thus it dominates the behavior in the underdoped region.

### 15.6 SU(4) and the Hubbard and $t$–$J$ models

As discussed more extensively in Section 2, a Hubbard or $t$–$J$ model and the dynamical symmetry approach applied here are alternative ways to simplify a strongly-correlated electron system. In the Hubbard or $t$–$J$ models a greatly simplified Hamiltonian is chosen but no specific configuration-space truncation is assumed (though practically a truncation is required). In contrast, our only approximation in the theory discussed here is the space truncation, since the symmetry-dictated Hamiltonian includes

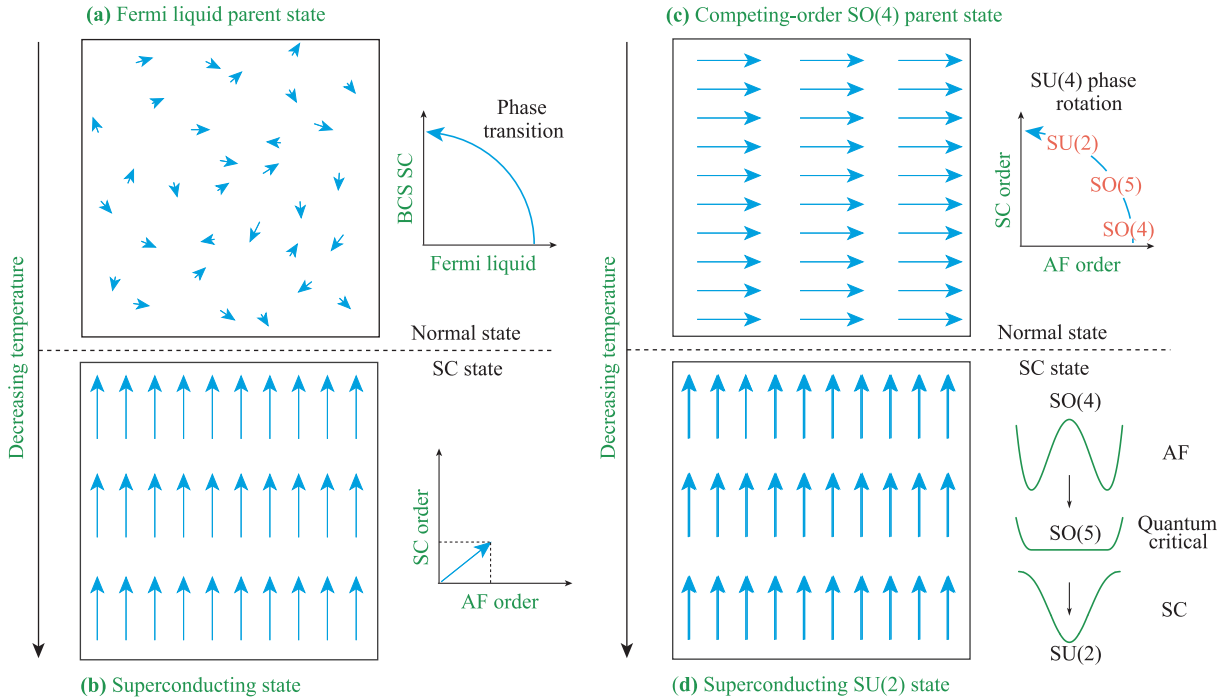



**Fig. 55** (Left) Formation of a normal (BCS) superconducting condensate. The vertical projection of an arrow represents the amount of pairing order; the horizontal projection of an arrow represents the amount of some competing order (such as antiferromagnetism). **(a)** Since we assume no net order in the parent state, the arrows are short (indicating matrix elements of non-collective strength) and randomly oriented (indicating no SC or competing order in the bulk). **(b)** Producing a superconducting state requires imposing order (aligning each randomly oriented arrow vertically) on a high-entropy initial state, which can occur only a very low temperature and implies a small value of $T_c$. (Right) Formation of a superconducting condensate in a system having an order like antiferromagnetism that competes with superconductivity as its normal ground state. **(c)** This requires imposing order on a state that is not superconducting, but is already highly ordered and thus of low entropy. **(d)** If the SC and the competing order are *both generators of some higher symmetry* like SU(4), then the transition from the competing order to the SC state is a collective rotation in the group space of the higher symmetry, which requires little change in entropy and can occur at a high value of $T_c$. In essence the SC state already exists in the competing-order ground state; it only has to be pointed in the right direction by a rotational nudge in the group space. For the cuprates doping can provide the required nudge, as indicated in (b).

all possible interactions in the truncated space, with the effect of the excluded space absorbed into the effective interactions of the truncated space. The validity of this approach depends entirely on validity of the choice of truncated space and its effective interactions, which may be tested by comparing calculated SU(4) matrix elements with data.

The Hamiltonian and wavefunctions for effective low-energy theories of the kind described here need not (likely should not!) resemble those of a Hubbard or $t$–$J$ model. Quantum mechanically only matrix elements are related to observables, not operators or wavefunctions separately, as illustrated in Fig. 54. Thus, a direct comparison of Hamiltonians or wavefunctions between two theories is valid *only if the two theories are defined within the same space*. If they are defined in different spaces, the only comparison that quantum mechanics permits is that of matrix elements evaluated in the two spaces; a comparison of Hamiltonians or wavefunctions separately has no physical content.

We may view emergent dynamical symmetries as operating in a truncated collective subspace in which the truncation has been implemented primarily by symmetry considerations. Thus, if the Hubbard or $t$–$J$ models and the SU(4) model are both valid descriptions of high-temperature SC, their physical matrix elements must be similar, making it highly unlikely that their Hamiltonians or wavefunctions separately would be similar, since they are defined in very different spaces.

## 16 High critical temperatures

Why are the critical temperatures for transition to the superconducting state unusually high for cuprates, iron-based superconductors, and many other unconventional superconductors (when measured in appropriate units for each case)? The unusual properties, including high $T_c$,

of unconventional superconductors are not because they have unconventional (not $s$-wave) pairing formfactors. The unconventional formfactors are not causes but rather are *symptoms* of a deeper and more important issue. The essential point is proximity of (one or more) other collective modes to SC in the phase diagram, suggesting that SC and the other mode are both possible ground states. If the two competing modes are related to each other, they may compete for the same Hilbert space. If they do compete for the same Hilbert space, this competition will tend naturally to produce unconventional formfactors but it can do something much more fundamental.

### 16.1 Unification of competing order

If the formfactor is unconventional it is a likely sign of competing order, and if the competing order is *related to the SC* in the right way [both being generators of a higher symmetry like SU(4), implying that they compete for the same Hilbert subspace] the competing order parent state can "precondition" the system for the SC phase transition. This allows it to occur at a higher value of $T_c$ because the competing-order ground state is a low-entropy state that can be rotated collectively into the SC state, as illustrated in Fig. 55.

In Fig. 55, superconducting order is indicated schematically by an up arrow and a competing order (assume it to be antiferromagnetism for discussion) by a right arrow. In the normal BCS case of Fig. 55(a) there is no net SC or AF in the initial state, so all arrows must individually be lengthened and ordered vertically in the superconducting phase transition, as indicated in Fig. 55(b). This transition from a high-entropy initial state to a highly-ordered final state requires a correspondingly low temperature to implement. In the competing order case of Fig. 55(c) the initial state is already ordered in a way such that a simple collective rotation in the group space produces the SC state, as in Fig. 55(d). This is generally the case when superconductivity and the order competing with it are unified in a higher symmetry like SU(4).

### 16.2 The generalized Cooper instability and high-$T_c$

The preceding argument is a cartoon version of the proof in Section 11.1 that the antiferromagnetic insulator ground state at half filling in the cuprates is inherently unstable against condensing Cooper pairs with doping, implying that the AF Mott state contains hidden within it a superconductor that can appear spontaneously with a slight disturbance. The spontaneous appearance of a significant pair gap $\Delta$ with infinitessimal doping in Fig. 35(c) and Fig. 36(b) is a quantitative implementation of the phase transition illustrated schematically in Fig. 55. However, the realistic case in Fig. 36 is more complex than the schematic picture in Fig. 55 in that a pairing gap appears spontaneously for doping $x \neq 0$ but AF correlations decrease with increased doping but remain finite until the rotation from antiferromagnetism to pure superconductivity is complete at the critical doping point $x = x_c$; see Fig. 27.

The SC transition between Figs. 55(c) and (d) can occur spontaneously if there is no barrier to the SU(4) rotation. The SU(4) $\supset$ SO(5) critical dynamical symmetry limit exhibits such a property. At low doping the energy surface implies degenerate AF and SC ground states with effectively no energy barrier separating them [see the curves in Fig. 23(b) for $n/\Omega = (1-x) \sim 1$]. This suggests that the AF and SC phases can be connected by a sequence of infinitesimal SU(4) rotations through intermediate SO(5) states having different mixtures of AF and SC order that are nearly degenerate in energy with the pure antiferromagnetic and superconducting states.

### 16.3 An information argument

The preceding entropy arguments may be expressed as an information argument. Figure 55(d) is obtained from Fig. 55(c) by collectively rotating all vectors. This can be specified in terms of a single rotation angle applied to all vectors, which requires minimal information. Conversely, in Fig. 55(a) there is no order in the parent state and each arrow must be lengthened and oriented separately to give Fig. 55(b). This requires supplying a much larger amount of information. Thus the reduction in entropy necessary to condense the superconducting state from the parent state is much greater in Figs. 55(a) and (b) than in Figs. 55(c) and (d).

The information argument also highlights the fundamental distinction between competing collective modes that are independent and those that are related by a higher symmetry. If the competing modes are independent, a large amount of information is required to change the competing-order state into the SC state because they are not fundamentally related. Microscopically, one collective mode must first be broken up and then reassembled into the other collective mode, which will hinder onset of superconductivity and decrease $T_c$. In the case that competing modes are related by a higher symmetry, the higher symmetry *already encodes the relationship between the two modes.* Hence only a small amount of additional information is required to produce the superconducting state from the competing-order state, because they arise from the same collective Hilbert subspace spanned by generators of the higher symmetry, and correspond to subgroups of the same highest symmetry.

This difference may be illustrated further by considering the competion of charge degrees of freedom with SC in the cuprate superconductors. We have seen that the minimal symmetry that can describe the cuprate superconductors is U(1)×SU(4), where SU(4) describes the competion of AF and SC, and U(1) is associated with a commensurate charge density wave. The direct product between U(1) and

SU(4) implies physically that the commensurate charge density wave is independent of the SU(4) description of AF and SC. Thus it cannot lower $T_c$ by the mechanism described above.

A more complex charge density wave would require a symmetry having more generators [the smallest compact Lie groups with with more generators than than the 15 of SU(4) are SO(7) or Sp(6) with 21 generators, SU(5) with 24 generators, and SO(8) with 28 generators]. Thus, for the minimal SU(4) model charge density waves would not be rotated into superconductivity by group generators, and by our argument they should not facilitate high values of $T_c$. We conclude that in a minimal model charge density waves represent a competing order lacking a structural relationship with AF and SC that at best has little effect, and at worst hinders, the formation of a superconducting state. As discussed in Section 11.2, charge density waves can perturb SU(4) symmetry for underdoped compounds where the energy surfaces may become critical. This can enable inhomogeneities such as stripes in a narrow range of doping, but is not likely to influence $T_c$ significantly.

### 16.4 The role of microscopic physics

The values of effective interaction parameters (influenced by underlying microscopic physics) will modulate exactly how large $T_c$ is in a specific compound, but the generic reason for abnormally high-$T_c$ in HTSC compounds—and unconventional superconductors in general—is the preconditioning implied by the symmetry relationship between the competing-order ground state and the superconductivity that emerges from it, as illustrated in Fig. 56. That is, the cuprate superconductors exhibit systematically high values of $T_c$ because of a fundamental scale set by SU(4) symmetry modulated by the strength of the effective interactions, but $T_c$ can vary from compound to compound within that scale because of microscopic physics, which influences the effective interaction parameters, and from variation of control parameters (doping in Fig. 56) that reflect the possibility of quantum phase transitions within the highest symmetry arising from competing multiple orders (subgroups of the highest symmetry). Succinctly, the underlying microscopic physics affects the properties of superconductivity such as $T_c$ in unconventional SC, but to first order it does so only parametrically.

## 17 Universality of superconducting and superfluid behavior

Let us use the terms "normal" to describe standard BCS superconductivity and "unusual" to denote superconductivity that results from Cooper pairing, but with some essential properties different from than of normal superconductivity. (Thus unconventional superconductors fall in our class labeled "unusual".) Based on our discussion of cuprate and iron-based superconductors in this review, on the general understanding that other forms of SC such as that in the heavy fermion and organic superconductors may have close resemblance to these, and the systematic application of dynamical symmetries to superconducting and superfluid behavior in nuclear structure physics, we are led to a sweeping conjecture [73, 74].

**Conjecture 2** *All superconductivity and superfluidity in all fields can be understood in terms of a generalized Cooper instability realized in terms of an operator algebra that is abelian in the simplest cases, but non-abelian in the most interesting cases. The large and fundamental differences in underlying microscopic physics across these fields are important only parametrically to the superconducting mechanism.*

Normal superconductivity corresponds to a U(1) subgroup in the dynamical symmetry chain SU(2) ⊃ U(1), where the SU(2) group is generated by pseudospin and charge operators. Physically, it is realized when the pairing collectivity can be described largely independent of other collective modes and the pairing is not modfied by effects such as strong on-site Coulomb repulsion. It corresponds to the standard BCS model. Unusual superconductivity corresponds to a physical situation where the pairing collectivity cannot be cleanly decoupled from other collective modes in the system. It corresponds to non-abelian

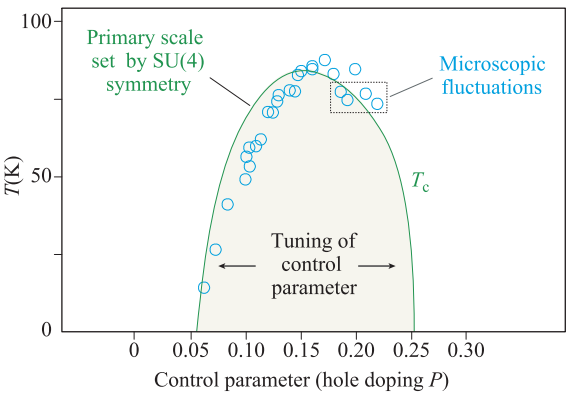



**Fig. 56** Primary factors affecting the temperature dependence of the superconducting transition temperature $T_c$ in an unconventional superconductor. Cuprate data (open circles) and an SU(4) calculation (solid curve) were adapted from Fig. 32. Error bars have been omitted but they are less than or equal to the sizes of the data points in most cases. The control parameter in this example is hole doping. The location and shape of the curve is determined by SU(4) symmetry. Its height is set by the symmetry modulated by the strength of the effective interaction in the truncated space, which depends on the average microphysics of the space excluded by the truncation. The fluctuations of data around the curve presumably reflect varying microphysics between compounds with different doping fractions that are not accounted for in this simple calculation.

subgroups of the algebra. High-temperature superconductivity in the cuprates and iron compounds are examples, corresponding to the dynamical symmetries of Eq. (7). Superconductivity and superfluidity in nuclear structure, which often occurs in the presence of collectivity associated with strong quadrupole deformation, is another [18].

Superficially, unusual superconductivity may not look like BCS superconductivity, but that is deceiving. The complex behavior is not capricious but rather is related systematically to normal BCS superconductivity by replacement of an abelian operator algebra with a non-abelian algebra that couples superconductivity strongly and non-linearly to other emergent collective modes. This, as discussed more extensively in Refs. [73, 74], permits the Cooper instability to be realized across very different physical systems that may exhibit complex and varied detailed behavior but clearly recognizable general features.

## 18 What is special about SU(4) symmetry?

In this review we have presented a substantial amount of formalism, expressed often in the language of Lie algebras and Lie groups. It is important that this not obscure the underlying motivation for this formalism, which is grounded directly in the phenomenology of high-$T_c$ superconductivity, and can be given an intuitive physical meaning largely separate from the mathematics employed.

### 18.1 The physical meaning of SU(4) symmetry

SU(4) symmetry is concise mathematical shorthand for a minimal physical model involving bondwise pairing and antiferromagnetism that conserves charge and spin, and implements no double occupancy of sites by components of the collective pairs. It is favored under physical conditions where superconductivity and antiferromagnetism lie near each other in the phase diagram and strong electron–electron correlations disfavor double occupation of the lattice sites. These conditions are fulfilled very well in the cuprates and at least approximately so in the iron superconductors, and similar ones may hold in a variety of other unconventional superconductors such as the heavy fermion and organic superconductors.

To the question "why SU(4)", we may then give a simple answer by asking a slightly more precise question. If we have a condensed matter system with strong electron–electron correlations and a tendency toward both magnetic and superconducting order, what is the simplest arrangement of these complex and partially antagonistic ingredients that minimizes the energy of the ground state? The surprisingly concise answer is that we must arrange the spin, charge, magnetic, and pairing operators so that they satisfy a set of physical constraints corresponding mathematically to closure of a non-abelian Lie algebra under commutation. The *simplest possibility* for those degrees of freedom is SU(4).

This relationship is no more surprising than that between the physical observation that angular momentum is always conserved and the statement of that physics in terms of an SU(2) Lie algebra. In both cases we are expressing a physical observation in concise mathematical terms using a symmetry implied by Lie algebras. But, there are important differences.

1) The SU(4) theory represents *dynamical symmetries* relating the interactions of different physical degrees of freedom in the Hamiltonian, while angular momentum SU(2) symmetry implies only a *conservation law,* not dynamical constraints. There is a parallel with gauge field theories, where *global gauge invariance* implies only a conservation law, but *local gauge invariance* is a much deeper statement about dynamics. SU(4) symmetry for strongly-correlated fermions is a statement about dynamics, not just conservation laws.

2) As a consequence, the symmetry associated with SU(4) is more difficult to uncover because it implies a more abstract and complex pattern in the observables than that associated with conservation of angular momentum.

3) Angular momentum SU(2) is an exact symmetry for any closed system but the SU(4) theory is based on an approximate symmetry that is expected to be realized only when a certain set of physical conditions is satisfied (a strongly-correlated electron system with emergent antiferromagnetism and superconductivity proximate in the phase diagram).

### 18.2 Intuitively correct limits

The physical validity of the SU(4) dynamical symmetry prescription is confirmed by detailed quantitative comparison of prediction with data, but it is also reinforced by the observation that it has intuitively-correct physical limits. As we have demonstrated in this review:

1) In the limit that the antiferromagnetic interactions of the effective Hamiltonian may be neglected, we recover from the SU(4) coherent state the standard BCS gap equations, but for pairs exhibiting a $g(\boldsymbol{k})$ formfactor that could be conventional $s$-wave or could be unconventional, depending on the underlying microscopic physics.

2) In the limit that the pairing interactions of the effective Hamiltonian may be neglected, we recover a theory with matrix elements corresponding to those of a Néel antiferromagnetic state.

3) This Néel antiferromagnetic state has suppressed double site occupancy because overlap of pairs breaks SU(4) symmetry. As we have explained, this is consistent with Mott insulator character (in the cuprates) or poor metal behavior (in the iron superconductors) for the normal state.

4) In the limit that both the pairing and antiferromagnetic interactions of the effective Hamiltonian may be neglected, we still recover a state with correlated electrons

and thus a tendency to suppress onsite interactions in the ground state, but without significant AF or SC order.

5) In the limit that electron–electron repulsion is weak, bondwise pairs are not favored energetically over onsite pairs and the SU(4) symmetry is enlarged to its parent group SO(8), which adds to the SU(4) generators onsite pairs and associated interactions. If physically we assume very weak electron–electron correlations and negligible antiferromagnetism, the SO(8) states favor a dynamical symmetry that corresponds to normal BCS superconductivity for onsite pairs. Thus the dynamical symmetries deriving from SO(8) and SO(8) $\supset$ SU(4) may be capable of describing the entire range of observed superconductivity in condensed matter within a unified framework.

These observations make it less surprising that our application of non-traditional mathematical methods to the strongly-correlated electron problem yields results that admit an interpretation of superconductivity, unconventional or otherwise, having a highly-traditional look and feel.

### 18.3 What SU(4) is not

Clarifying what SU(4) symmetry is not is as important as clarifying what it is. It is *not* generally expected to be a symmetry of the underlying weakly-interacting microscopic system, or of a Hamiltonian appropriate for the microscopic system such as for the Hubbard model. SU(4) is an *emergent symmetry* that is realized only when the dynamics of the system favor the emergence of the corresponding collective modes (antiferromagnetism, superconductivity, and their interaction in a system exhibiting strong electron–electron correlations). As we have emphasized, emergent dynamical symmetries generally would not be expected to have direct connections with any symmetries of the underlying microscopic system, because emergent phenomena generally cannot be derived perturbatively from the non-interacting constituent system.

### 18.4 Simple descriptions and complex phenomena

Finally, let us counter a possible philosophical criticism of our dynamical symmetry approach: the potential ground states in any strongly-correlated electron system are so complex microscopically that we might be tempted to doubt that a simple model like the current one would have any validity. But this argument ignores the quite obvious point that Nature has managed to construct a stable ground state for high-temperature superconductors having well-defined, collective properties that vary in a controlled manner according to a few well-chosen parameters. Thus Nature is waving a red flag indicating that the phenomenon in question is actually simple, if we will but change our perspective to view it in a more natural basis.

From experience in many fields of many-body physics, this is a clear physical signal that the superconductor is described mathematically by a *small effective subspace with renormalized interactions*, and governed by a dynamical symmetry structure of relatively small dimensionality acting within that subspace. Thus, if an approach like the one proposed here gives correct results for highly non-trivial phenomenology like the doping dependence of various observable quantities, one must take seriously the possibility that the corresponding small symmetry-dictated subspace may have relevance to the effective behavior of real physical systems.

## 19 Summary and conclusions

In summary, we have developed a solvable microscopic theory of unconventional superconductivity and antiferromagnetism on a lattice with suppressed double site occupancy. This permits an exact many-body solution for a minimal model having charge, spin, pairing, and antiferromagnetism for special ratios of the coupling parameters, and an approximate generalized coherent-state solution for arbitrary coupling strengths. Superconductivity and antiferromagnetism enter on an equal footing. The three dynamical symmetries of the model, SU(2), SO(4), and SO(5), yield exact solutions that correspond respectively to states exhibiting singlet-pair superconductivity, antiferromagnetism, and a critical dynamical symmetry interpolating between superconductivity and antiferromagnetism. The competing AF and SC order imply an SU(4) symmetry that embodies many essential features of high-temperature superconductivity:

1) The SU(4) symmetry imposes no double occupancy of lattice sites for the electrons of the correlated pairs.

2) If the particles near the Fermi surface all contribute uniformly to the correlated pairs, the preceding point implies that at half filling the system favors antiferromagnetic Mott insulator behavior (as observed in the cuprates).

3) If the particles near the Fermi surface contribute non-uniformly to the correlated pairs the normal state at half filling is antiferromagnetic, but could be a poor metal in the normal state (as observed in the pnictides).

4) Doping of the AF Mott insulator ground state at half filling leads to the spontaneous emergence of a singlet superconductor ground state.

5) Pseudogap states emerge naturally in the model at intermediate doping.

The ground state of this theory exhibits two fundamental instabilities:

1) The antiferromagnetic state at half filling (which for cuprates is a Mott insulator) is unstable against condensing singlet pairs in the presence of infinitesimal

doping unless the pairing interaction vanishes. Only competing AF fluctuations, vanishing pair correlations, or breaking of SU(4) symmetry prevent the immediate formation of a pure singlet-pair condensate with doping. Thus we have generalized the Cooper instability to strongly-correlated electron systems with possible competing order.

2) In the underdoped region, the ground state is unstable against fluctuations in both antiferromagnetic and superconducting order. This enables all manner of emergent behavior as the nature of the ground state can be altered qualitatively by small perturbations (stripes, checkerboards, …) The model suggests that such behavior is not fundamental to the high-$T_c$ mechanism but rather is opportunistic, exploiting a fundamental softness of the SU(4) ground state that is predicted by the symmetry to occur only in a narrow range of doping.

We propose that most cuprate phenomenology may be understood in terms of these two fundamental instabilities: the generalized Cooper instability accounts for the rapid appearance of superconductivity when the Mott insulator is doped with holes and the antiferromagnetic instability accounts largely for the pseudogap and its properties.

The model introduces, as a consequence of the highest symmetry and it subgroups, multiple energy scales. These lead to a rich variety of gaps that could play a physical role. As a result, we find a phase diagram that is in quantitative agreement with that observed experimentally for cuprate superconductors. The basic phase diagram is a direct consequence of the symmetry, with parameter adjustment influencing only details.

The cuprate pseudogap state has fluctuating antiferromagnetic and pairing character in the SU(4) description, and terminates at a quantum phase transition marked by a critical doping $P \simeq 0.18$; it is distinct from the superconducting state but related to it by a non-abelian symmetry. The pseudogap may be interpreted in terms of both SC–AF competition and preformed SU(4) pairs that condense into a singlet $d$-wave superconductor as hole doping suppresses fluctuations. We account quantitatively for the doping dependence of the pseudogap temperature $T^*$ and, because of the fluctuating nature of the state, we conclude that this PG state is most likely to be observed as a crossover rather than a distinct phase.

The structure of the SU(4) pseudogap state leads to a simultaneous quantitative description of the pseudogap temperature scale $T^*$ and Fermi arcs in ARPES experiments, including the origin of $T^*$ scaling. We have shown that requiring a quantitative description simultaneously of ARPES Fermi arcs and the doping dependence of the $T^*$ scale upon which they depend places extremely strong constraints on an acceptable theory of high-temperature superconductivity. If instead the Fermi surface for underdoped cuprates is interpreted in terms of small closed pockets (as suggested by quantum oscillation experiments), the anisotropic pseudogap correlations implied by the SU(4) model place strong constraints on where those pockets could be.

We have applied the method of generalized SU(4) coherent states, which provides a systematic procedure to relate a many-body theory to its approximate broken-symmetry solutions. This approach may be viewed as a standardized technology for constructing energy surfaces of many-body theories defined in terms of the algebra of their second-quantized operators, which provides a microscopic connection to Ginzburg–Landau methods. Equivalently, it may be viewed as implementing the most general Hartree–Fock–Bogoliubov theory, subject to a symmetry constraint on the Hamiltonian of the system. Thus the coherent-state solution for SU(4) allows us to express results in language familiar in condensed matter: spontaneously-broken symmetries, gap equations for quasiparticles, and variational energy surfaces.

We have shown that competing antiferromagnetism and superconductivity, constrained by SU(4) symmetry which imposes no double lattice occupancy, leads on general grounds to energy surfaces in hole underdoped cuprates corresponding to weakly-broken SU(4) $\supset$ SO(5) symmetry that may be critically balanced between antiferromagnetic and superconducting order. These surfaces can be flipped between dominance of one order or the other by small fluctuations in the ratio of the antiferromagnetic to pairing strength. Therefore weak perturbations in the underdoped region, or near vortex cores or magnetic impurities, can produce amplified inhomogeneity having the spatial dependence of the perturbation but the intrinsic character of an SU(4) symmetry. (The symmetry defines the possible states; the perturbation selects among them.) Our results show that such effects can, but need not necessarily, involve spatial modulation of charge. More generally, we have suggested that critical dynamical symmetry may be a fundamental organizing principle for emergent behavior in correlated fermion systems, and that it provides a natural explanation for observed rich inhomogeneity in underdoped compounds with (paradoxically) near-universal overall phase diagrams in cuprate superconductors.

These considerations suggest that stripes, checkerboards and related inhomogeneities are secondary issues in understanding high-$T_c$ superconductivity. They are perturbative (around a non-perturbative vacuum) consequences of the superconducting physics, not its cause. It is important to emphasize that the pseudogap and this sensitivity to spatial inhomogeneity in the underdoped region derive from the same fundamental physics of the underlying SU(4) symmetry.

The simplest charge-density wave decouples from the pairing–AF subspace in lowest order, suggesting that commensurate charge-density waves are not central to the high-$T_c$ mechanism. However, they can exploit the underdoped instability described above, producing a variety of

induced structure for underdoped compounds.

We have argued that quantitative extension of this SU(4) approach to Fe-based superconductivity requires only that (i) the relevant collective degrees of freedom are SC and AF, and (ii) the SC involves bondwise (not onsite) pairing. Thus, evidence has been presented that the Fe-based high-temperature superconductors represent the second example (after the cuprates) of the non-abelian superconductors that we proposed in 2004 [49]. The identification of non-abelian superconductivity in these two classes of compounds permits a unified model of cuprate and Fe-based superconductors to be constructed based on an SU(4) group (and subgroups) generated by emergent degrees of freedom, despite the obvious physical and microscopic differences between these classes of compounds. However, consistency with neutron scattering and ARPES data places strong constraints on possible FeAs orbital pairing formfactors, and closure of the SU(4) algebra suggests generally that the pairing in the FeAs case could correspond to different orbital formfactors at the microscopic level than is required for the cuprates.

The SU(4) model presented here has connections to a number of other theoretical approaches that have been applied to the high-temperature superconductor problem:

1) Our results provide some support for the assumption of resonating valence bond models that the state with AF order at half filling would really like to be a state with many characteristics of a spin-singlet liquid. However, the SU(4) variational coherent state is simpler to implement and yet contains a broader range of physics than a spin-singlet liquid, and accounts for many cuprate properties across the entire physical doping range without introducing spinons, holons, or related concepts having marginal experimental support in HTSC compounds.

2) The Zhang SO(5) model is recovered as one symmetry-limit approximation of the SU(4) theory. However, the present approach differs fundamentally from that of Zhang and derives the SO(5) subgroup by approximation from a richer theory with broader physical implications.

3) The SU(4) emergent-symmetry approach differs in spirit from that of approaches such as the Hubbard or $t$—$J$ models. However, these approaches need not be antagonistic. Because the SU(4) theory is defined in a highly-truncated subspace, it makes no sense to compare operators or wavefunctions directly with other theories, but it is legitimate to compare matrix elements. As we have demonstrated, the SU(4) model correctly describes the matrix elements corresponding to many fundamental properties of high-temperature superconductors. It is less certain that Hubbard or $t$–$J$ models can make the same claim, because they are difficult to solve in spaces large enough to give definitive results. For example, there are differing opinions on whether these approaches actually lead to a robust superconducting state for realistic systems.

The properties in the phase diagram discussed here for cuprate and iron-based high-temperature superconductors have similarities with properties observed in other materials. In heavy-fermion compounds and some organic superconductors there is evidence for significant electron–electron correlation and a superconducting phase appears near the boundary of an AF phase. As a second example, the manganites have strong correlations and complex competing phases, some bearing a resemblance to those that we have discussed. Therefore, the formalism developed here to describe multiple competing low-temperatures phases should be applicable to a much broader range of strongly-correlated electron systems, with doping replaced or supplemented by additional control parameters such as pressure or strength of a magnetic field.

Hence, by employing techniques that are well-established in general many-body physics, we conclude that the properties of high-temperature superconductors, including the rapid development of a superconductor from a Mott insulator in the cuprates, the properties of pseudogap states, and rich disorder localized within an otherwise universal phase diagram, are understandable in terms of a minimal generalization of traditional BCS theory and the Cooper instability to include self-consistently the role of antiferromagnetism and onsite Coulomb repulsion. Our results represent a minimal variational solution of competing antiferromagnetism and singlet superconductivity on a fermionic lattice with no double occupancy for the pair components. Therefore, we believe that the general gap and phase structure presented here will be a necessary consequence of any realistic theory that takes a lattice with strongly-correlated electrons and competing pairing and antiferromagnetism as the basis for describing high-temperature superconductivity.

The preceding observations, coupled with the deep algebraic analogies noted between AF–SC competition in condensed matter and deformation–superconductivity competition in nuclear physics, suggests a fundamental relationship between the forms of superconductivity observed in many fields of science. The robustness and similarities of superconductivity across so many subfields that deal with matter having very different length and energy scales, and very different physical environments, indicates that the Cooper-pair superconducting mechanism *cannot* depend essentially on microscopic details in any one field. Indeed, it is suggested that the opposite is true: superconductivity must correspond to a mechanism that is extremely robust and compatible with a very broad range of underlying microscopic details. These details influence the theory parametrically (for example, determining the exact value of the superconducting transition temperature and the range of parameters over which SC is found) but must have little power to determine the general properties of superconductivity in the system, except to favor generic conditions that permit it to emerge.

A common algebraic structure for the relevant quantum

operators is one of the few things that could be similar and independent of detailed microscopic structure across the superconductivity observed or expected in such diverse physical systems. This leads us to conjecture that all superconductivity and superfluidity in all fields can be understood in terms of a generalized Cooper instability that is realized in terms of similar abelian or non-abelian operator algebras.

Finally, let us draw attention to the irony that this discussion may seem at first blush to be unconventional because of methodology, yet it leads to the most conventional and conservative of conclusions. The high-temperature superconductors are described at all dopings by a BCS formalism generalized self-consistently to incorporate antiferromagnetism, pairing, and on-site Coulomb repulsion on an equivalent footing. Microscopic details such as dimensionality, gap orbital symmetry, pair binding mechanism, microscopic structure of the magnetism, the crystal structure, presence or absence of disorder, and so on are important in their own right, but their influence on the superconductor is primarily to set the value of coefficients in equations whose form has largely been determined by emergent dynamical symmetry, independent of those microscopic details. The emergent collective properties defining the essence of the superconducting state that are so easily recognized across many physical systems with very different microscopic structure require only that the microscopic conditions permit realization of emergent dynamical symmetries of the Hamiltonian like the SU(4) symmetry described here, largely independent of further microscopic details.

## Appendix A SU(4) subgroups and dynamical symmetries

The basic properties of SU(4) and its dynamical subgroups that conserve charge and spin are summarized in Table 1 and Table 2.

**Table 1** Properties of SU(4) and its subgroups (assuming no broken pairs).

| Group | Generators | Quantum numbers | Casimir operator | Casimir eigenvalue |
|---|---|---|---|---|
| SU(4) | $\boldsymbol{S}$, $\boldsymbol{Q}$, $\boldsymbol{\pi}^\dagger$, $\boldsymbol{\pi}$, $p^\dagger$, $p$, $M$ | $\sigma_1 = \frac{\Omega}{2}$, $(\sigma_2 = \sigma_3 = 0)$ | $\boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} + p^\dagger p + \boldsymbol{S} \cdot \boldsymbol{S} + \boldsymbol{Q} \cdot \boldsymbol{Q} + M(M-4)$ | $\frac{\Omega}{2}\left(\frac{\Omega}{2}+4\right)$ |
| SO(4) | $\boldsymbol{Q}$, $\boldsymbol{S}$ | $w$, $S$ | $\boldsymbol{Q} \cdot \boldsymbol{Q} + \boldsymbol{S} \cdot \boldsymbol{S}$ | $w(w+2) + S(S+1)$ |
| $\mathrm{SU(2)_p}$ | $p^\dagger$, $p$, $M$ | $N$, $\nu$ | $p^\dagger p + M(M-1)$ | $\frac{1}{4}(N-\nu)(2\Omega - N + 2)$ |
| $\mathrm{SU(2)_s}$ | $\boldsymbol{S}$ | $S$ | $\boldsymbol{S} \cdot \boldsymbol{S}$ | $S(S+1)$ |
| SO(5) | $\boldsymbol{S}$, $\boldsymbol{\pi}^\dagger$, $\boldsymbol{\pi}$, $M$ | $\tau$ $(\tau_2 = 0)$ | $\boldsymbol{\pi}^\dagger \cdot \boldsymbol{\pi} + \boldsymbol{S} \cdot \boldsymbol{S}$ | $\tau(\tau+3)$ |

**Table 2** The Hamiltonian, eigenstates and spectra in three dynamical symmetry limits of the SU(4) model. $E_{\mathrm{g.s.}}$ is the ground state energy, $\Delta E$ the excitation energy, $N = \frac{1}{2}n$ is the pair number, $x = 1 - n/\Omega$, and $\kappa_{\mathrm{so(4)}} = \kappa_{\mathrm{eff}} + \chi_{\mathrm{eff}}$.

| SU(2) limit: $\lvert\psi(\mathrm{SU(2)})\rangle = \lvert N, v, S, m_S\rangle$ | SO(4) limit: $\lvert\psi(\mathrm{SO(4)})\rangle = \lvert N, w, S, m_S\rangle$ | SO(5) limit: $\lvert\psi(\mathrm{SO(5)})\rangle = \lvert\tau, N, S, m_S\rangle$ |
|---|---|---|
| $\langle C_{\mathrm{su(2)}}\rangle = \frac{1}{4}(\Omega - v)(\Omega - v + 2)$ | $\langle C_{\mathrm{so(4)}}\rangle = w(w+2), \quad w = N - \mu$ | $\langle C_{\mathrm{so(5)}}\rangle = \tau(\tau+3), \quad \Omega/2 - \tau = N - \lambda$ |
| $H = H_0 + \kappa_{\mathrm{eff}}\,\boldsymbol{S}\cdot\boldsymbol{S}$ $-G^{(0)}_{\mathrm{eff}}\left[C_{\mathrm{su(2)}_p} - M(M-1)\right]$ | $H = H_0 + \kappa_{\mathrm{so(4)}}\,\boldsymbol{S}\cdot\boldsymbol{S} - \chi_{\mathrm{eff}} C_{\mathrm{so(4)}}$ | $H = H_0 + \kappa_{\mathrm{eff}}\,\boldsymbol{S}\cdot\boldsymbol{S}$ $-G^{(0)}_{\mathrm{eff}}\left[C_{\mathrm{su(4)}} + M - C_{\mathrm{so(5)}}\right]$ |
| $E_{\mathrm{g.s.}} = H_0 - \frac{1}{4}G^{(0)}_{\mathrm{eff}}\Omega^2(1-x^2)$ | $E_{\mathrm{g.s.}} = H_0 - \frac{1}{4}\chi_{\mathrm{eff}}\Omega^2(1-x)^2$ | $E_{\mathrm{g.s.}} = H_0 - \frac{1}{4}\chi_{\mathrm{eff}}\Omega^2(1-x)^2$ |
| $\Delta E = \nu G^{(0)}_{\mathrm{eff}}\Omega + \kappa_{\mathrm{eff}}\ S(S+1), \quad \nu = v/2$ | $\Delta E = \mu\chi_{\mathrm{eff}}(1-x)\Omega + \kappa_{\mathrm{so(4)}}\ S(S+1)$ | $\Delta E = \lambda x G^{(0)}_{\mathrm{eff}}\Omega + \kappa_{\mathrm{eff}}\ S(S+1)$ |
| $\nu = N, N-1, \ldots, 0;\ S = \nu, \nu-2, \ldots, 0$ or 1 | $\mu = N, N-2, \ldots, 0$ or 1; $S = w, w-1, \ldots, 0$ | $\lambda = N, N-1, \ldots, 0$ or 1; $S = \lambda, \lambda-2, \ldots, 0$ |

**Acknowledgements** We would like to thank Pengcheng Dai, Elbio Dagotto, Adriana Moreo, Takeshi Egami, John Quinn, Hai-Hu Wen, and Wei Ku for discussions and advice that have greatly enhanced our understanding of strongly correlated electron systems. This work was partially supported by the National Key Program for S&T Research and Development (Grant No. 2016YFA0400501). L. W. acknowledges grant support from the Basque Government (Grant No. IT986-16) and PGC2018-101355-B-I00 (MCIU/AEI/FEDER, UE). This work was partially supported by LightCone Interactive LLC.

## References and notes

1. J. G. Bednorz and K. A. M uller, Possible high $T_c$ superconductivity in the Ba–La–Cu–O system, *Z. Phys. B* 64(2), 189 (1986)

2. See: D. A. Bonn, Are high-temperature superconductors exotic? *Nat. Phys. 2*, 159 (2006); M. R. Norman and C. Pépin, The electronic nature of high temperature cuprate superconductors, *Rep. Prog. Phys.* 66(10), 1547 (2003) (and Refs. therein)
3. Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, Iron-based layered superconductor $LaO_{1-x}F_xFeAs$ ($x$ = 0.05–0.12) with $T_c$ = 26 K, *J. Am. Chem. Soc.* 130(11), 3296 (2008)
4. Y. Sun, M. W. Guidry, and C. L. Wu, A new family of high $T_c$ compounds—Stepping stones toward understanding unconventional superconductivity, *Chin. Sci. Bull.* 53, 1617 (2009)
5. J. Guo, S. Jin, G. Wang, S. Wang, K. Zhu, T. Zhou, M. He, and X. Chen, Superconductivity in the iron selenide $K_xFe_2Se_2$, *Phys. Rev. B* 82(18), 180520 (2010)
6. D. C. Johnston, The puzzle of high temperature superconductivity in layered iron pnictides and chalcogenides, *Adv. Phys.* 59(6), 803 (2010)
7. J. Paglione and R. L. Greene, High-temperature superconductivity in iron-based materials, *Nat. Phys.* 6(9), 645 (2010)
8. D. Mou, L. Zhao, and X. Zhou, Structural, magnetic and electronic properties of the iron-chalcogenide $A_xFe_{2-y}Se_2$ (A=K, Cs, Rb, and Tl, etc.) superconductors, *Front. Phys.* 6(4), 410 (2011)
9. H. Oh, J. Moon, D. Shin, C. Moon, and H. J. Choi, Brief review on iron-based superconductors: Are there clues for unconventional superconductivity? *Progress in Superconductivity* 13, 65 (2011)
10. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Theory of superconductivity, *Phys. Rev.* 108(5), 1175 (1957)
11. L. N. Cooper, Bound electron pairs in a degenerate Fermi gas, *Phys. Rev.* 104(4), 1189 (1956)
12. For example, see M. R. Norman, The challenge of unconventional superconductivity, *Science* 332(6026), 196 (2011) (and references cited there)
13. D. Jérome, Organic superconductors: When correlations and magnetism walk in, arXiv: 1201.5796 (2012)
14. See, for example, P. Ring, and P. Schuck, The Nuclear Many-Body Problem, Springer–Verlag, 1980
15. D. Page, M. Prakash, J. M. Lattimer, and A. W. Steiner, Superfluid Neutrons in the core of the neutron star in cassiopeia A, arXiv: 1110.5116 (2011)
16. T. Noda, M. Hashimoto, N. Yasutake, T. Maruyama, T. Tatsumi, and M. Fujimoto, Cooling of compact stars with color superconducting phase in quark hadron mixed phase, *Astrophys. J.* 765(1), 1 (2012)
17. M. G. Alford, A.Schmitt, K. Rajagopal, and T. Schäfer, Color superconductivity in dense quark matter, *Rev. Mod. Phys.* 80(4), 1455 (2008)
18. C. L. Wu, D. H. Feng, and M. W. Guidry, The fermion dynamical symmetry model, *Adv. Nucl. Phys.* 21, 227 (1994)
19. F. Iachello and A.Arima, The Interacting Boson Model, Cambridge University Press, Cambridge, 1987
20. R. Bijker, F. Iachello, and A. Leviatan, Algebraic models of hadron structure (I): Nonstrange baryons, *Ann. Phys.* 236(1), 69 (1994)
21. F. Iachello and R. D. Levine, Algebraic Theory of Molecules, Oxford University Press, Oxford, 1995
22. F. Iachello and P. Truini, Algebraic model of anharmonic polymer chains, *Ann. Phys.* 276(1), 120 (1999)
23. P. W. Anderson, Random-phase approximation in the theory of superconductivity, *Phys. Rev.* 112(6), 1900 (1958)
24. R. J. Glauber, Photon correlations, *Phys. Rev. Lett.* 10(3), 277 (1963)
25. For a review, see, Ref. [6] and P. J. Hirschfeld, M. M. Korshunov, and I. I. Mazin, Gap symmetry and structure of Fe-based superconductors, *Rep. Prog. Phys.* 74, 124508 (2011)
26. T. Timusk and B. Statt, The pseudogap in high-temperature superconductors: An experimental survey, *Rep. Prog. Phys.* 62(1), 61 (1999)
27. M. Norman, D. Pines, and C. Kollin, The pseudogap: Friend or foe of high $T_c$? *Adv. Phys.* 54(8), 715 (2005)
28. G. Sheet, M. Mehta, D. A. Dikin, S. Lee, C. W. Bark, J. Jiang, J. D. Weiss, E. E. Hellstrom, M. S. Rzchowski, C. B. Eom, and V. Chandrasekhar, Phase-incoherent superconducting pairs in the normal state of $Ba(Fe_{1-x}Co_x)_2As_2$, *Phys. Rev. Lett.* 105(16), 167003 (2010)
29. T. Mertelj, V. V. Kabanov, C. Gadermaier, N. D. Zhigadlo, S. Katrych, J. Karpinski, and D. Mihailovic, Distinct pseudogap and quasiparticle relaxation dynamics in the superconducting state of nearly optimally doped $SmFeAsO_{0.8}F_{0.2}$ single crystals, *Phys. Rev. Lett.* 102(11), 117002 (2009)
30. K. Ahilan, F. L. Ning, T. Imai, A. S. Sefat, R. Jin, M. A. McGuire, B. C. Sales, and D. Mandrus, $^{19}F$ NMR investigation of the iron pnictide superconductor $LaFeAsO_{0.89}F_{0.11}$, *Phys. Rev. B* 78, 100501(R) (2008)
31. T. Sato, K. Nakayama, Y. Sekiba, T. Arakane, K. Terashima, S. Souma, T. Takahashi, Y. Kamihara, M. Hirano, and H. Hosono, Doping dependence of pseudogap in $LaFeAsO_{1-x}F_x$, *J. Phys. Soc. Jpn. C* 77 (Suppl.), 65 (2008)
32. F. Ning, K. Ahilan, T. Imai, A. S. Sefat, R. Jin, M. A. McGuire, B. C. Sales, and D. Mandrus, $^{59}Co$ and $^{75}As$ NMR investigation of electron-doped high $T_c$ superconductor $BaFe_{1.8}Co_{0.2}As_2$ ($T_c$ = 22 K), *J. Phys. Soc. Jpn.* 77(10), 103705 (2008)
33. S. M. Hayden, H. A. Mook, P. Dai, T. G. Perring, and F. Doğan, The structure of the high-energy spin excitations in a high-transition-temperature superconductor, *Nature* 429 (6991), 531 (2004)
34. J. M. Tranquada, H. Woo, T. G. Perring, H. Goka, G. D. Gu, G. Xu, M. Fujita, and K. Yamada, Quantum magnetic excitations from stripes in copper oxide superconductors, *Nature* 429(6991), 534 (2004)

35. J. E. Hoffman, E. W. Hudson, K. M. Lang, V. Madhavan, H. Eisaki, S. Uchida, and J. C. Davis, A four unit cell periodic pattern of quasi-particle states surrounding vortex cores in $Bi_2Sr_2CaCu_2O_{8+\delta}$, *Science* 295 (5554), 466 (2002)
36. M. Vershinin, S. Misra, S. Ono, Y. Abe, Y. Ando, and A. Yazdani, Local ordering in the pseudogap state of the high-$T_c$ superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, *Science* 303 (5666), 1995 (2004)
37. T. Hanaguri, C. Lupien, Y. Kohsaka, D. H. Lee, M. Azuma, M. Takano, H. Takagi, and J. C.Davis, A "checkerboard" electronic crystal state in lightly hole-doped $Ca_{2-x}Na_xCuO_2Cl_2$, *Nature* 430 (7003), 1001 (2004)
38. K. McElroy, J. Lee, J. A. Slezak, D.-H. Lee, H. Eisaki, S. Uchida, and J. C. Davis, Atomic-scale sources and mechanism of nanoscale electronic disorder in $Bi_2Sr_2CaCu_2O_{8+\delta}$, *Science* 309(5737), 1048 (2005)
39. V. J. Emery and S. A. Kivelson, Importance of phase fluctuations in superconductors with small superfluid density, *Nature* 374(6521), 434(1995)
40. J. L. Tallon and J. W. Loram, The doping dependence of $T^*$–What is the real high-$T_c$ phase diagram? *Physica* C349(1-2), 53(2001)
41. P. W. Anderson, Spin-charge separation is the key to the high $T_c$ cuprates, *Physica C* 341-348, 9(2000)
42. D. Pines, Quantum protectorates in the cuprate superconductors, *Physica C* 341–348, 59(2000)
43. R. B. Laughlin and D. Pines, The theory of everything, *Proc. Natl. Acad. Sci. USA* 97(1), 28(2000)
44. A. Bohr and B. R. Mottelson, Nuclear Structure, Vols. I and II, W. A. Benjamin, 1969 and 1975
45. However the standard methodologies in the elementary particle physics case differ from the ones used here, partially because a relativistic quantum field theory is required there but non-relativistic fields are adequate for the present discussion. An accessible introduction to non-Abelian gauge fields may be found in Gauge Field Theories, An Introduction with Applications, Mike Guidry, Wiley, 1992.
46. M. W. Guidry, L. A. Wu, Y. Sun, and C. L. Wu, SU(4) model of high-temperature superconductivity and antiferromagnetism, *Phys. Rev. B*63(13), 134516(2001)
47. M. W. Guidry, A fermion dynamical symmetry model of high-temperature superconductivity and antiferromagnetic order, *Rev. Mex. Fís.* 45(S2), 132 (1999)
48. L. A. Wu, M. W. Guidry, Y. Sun, and C. L. Wu, SO(5) as a critical dynamical symmetry in the SU(4) model of high-temperature superconductivity, *Phys. Rev. B* 67(1), 014515 (2003)
49. M. W. Guidry, Y. Sun, and C. L. Wu, Mott insulators, no-double-occupancy, and non-Abelian superconductivity, *Phys. Rev. B* 70(18), 184501 (2004)
50. Y. Sun, M. W. Guidry, and C. L. Wu, Temperature-dependent gap equations and their solutions in the $SU(4)$ model of high-temperature superconductivity, *Phys. Rev. B* 73(13), 134519 (2006)
51. Y. Sun, M. W. Guidry, and C. L. Wu, Pairing Gaps, Pseudogaps, and Phase Diagrams for Cuprate Superconductors, *Phys. Rev. B* 75(13), 134511 (2007)
52. Y. Sun, M. W. Guidry, and C. L. Wu, k-dependent $SU(4)$ model of high-temperature superconductivity and its coherent state solutions, *Phys. Rev. B* 78(17), 174524 (2008)
53. M. W. Guidry, Y. Sun, and C. L. Wu, Instabilities of doped Mott insulators and the properties of high-temperature superconductors, published in *Nuclei and Mesoscopic Physics*, p. 160, P. Danielewicz, P. Piecuch, and V. Zelevinsky (Eds.), AIP Conference Proceedings, 2008.
54. M. W. Guidry, Y. Sun, and C. L. Wu, A unified description of cuprate and iron arsenide superconductors, *Front. Phys. China* 4(4), 433 (2009)
55. M. W. Guidry, Y. Sun, and C. L. Wu, Strong anisotropy of cupratepseudogap correlations: Implications for Fermi arcs and Fermi pockets, *New J. Phys.* 11(12), 123023 (2009)
56. M. Guidry, Y. Sun, and C. L. Wu, Generalizing the Cooper pair instability to doped Mott insulators, *Front. Phys. China* 5(2), 171 (2010)
57. M. W. Guidry, Y. Sun, and C. L. Wu, Inhomogeneity, dynamical symmetry, and complexity in high-temperature superconductors: Reconciling a universal phase diagram with rich local disorder, *Chin. Sci. Bull.* 56(4–5), 367 (2011)
58. The particle–hole symmetry intrinsic to these models does not mean that hole-doped and electron-doped compounds are expected to behave in the same manner. Although the operators and basis states of the model are particle–hole symmetric, the interactions entering the effective Hamiltonian would not be expected to be the same for hole-doped and particle-doped compounds. Thus, the physical properties of hole-doped and electron-doped compounds could differ substantially.
59. We employ an isomorphism between the groups SU(4) and SO(6) to label irreducible representations using SO(6) quantum numbers. The representation structure and relationship of SU(4) and SO(6) is discussed in: J. N. Ginocchio, *Ann. Phys.* 126, 234 (1980)
60. Groups generally may have more than one Casimir invariant. We shall use the term "Casimir" to refer loosely to the lowest-order such invariants (which are generally quadratic in the group generators). In the context of the present discussion, quadratic Casimirs are associated with 2-body interactions at the microscopic level. Higher-order Casimirs are then generally associated with 3-body and higher interactions. The restriction of our Hamiltonians to polynomials of order 2 in the Casimirs is then a physical restriction to consideration of only 1-body and 2-body interactions.
61. W. M. Zhang, D. H. Feng, and R. Gilmore, Coherent states: Theory and some applications, *Rev. Mod. Phys.* 62(4), 867(1990)
62. W. M. Zhang, C. L. Wu, D. H. Feng, J. N. Ginocchio, and M. W. Guidry, Geometrical structure and critical phenomena in the fermion dynamical symmetry model: $Sp(6)$, *Phys. Rev. C* 38(3), 1475 (1988)

63. W. M. Zhang, D. H. Feng, C. L. Wu, H. Wu, and J. N. Ginocchio, Symmetry constrained Hartree–Fock–Bogoliubov theory with applications to the fermion dynamical symmetry model, *Nucl. Phys. A* 505(1), 7 (1989)

64. W. M. Zhang, D. H. Feng, and J. N. Ginocchio, Geometrical interpretation of SO(7): A critical dynamical symmetry, *Phys. Rev. Lett.* 59(18), 2032(1987)

65. W. M. Zhang, D. H. Feng, and J. N. Ginocchio, Geometrical structure and critical phenomena in the fermion dynamical symmetry model: SO(8), *Phys. Rev. C* 37(3), 1281 (1988)

66. R. Gilmore, Geometry of symmetrized states, *Ann. Phys.* 74(2), 391 (1972)

67. R. Gilmore, On the properties of coherent states, *Rev. Mex. Fis.* 23(1–2), 143 (1974)

68. A. M. Perelomov, Coherent states for arbitrary Lie group, *Commun. Math. Phys.* 26(3), 222 (1972)

69. J. R. Klauder, Continuous-representation theory: Postulates of continuous-representation theory, *J. Math. Phys.* 4, 1055 (1963); Continuous-representation theory (II): Postulates of continuous-representation theory, *J. Math. Phys.* 4, 1058 (1963)

70. Thus the most general SU(4) coherent state depends on eight real variables. The reduction of the coherent state parameters to only two in Eq. (28) follows from requiring time reversal symmetry and assuming conservation of spin projection $S_z$ for the wave function.

71. B. R.Judd, Operator Techniques in Atomic Spectroscopy, McGraw-Hill, 1963

72. The competition between dynamical symmetries governing the transition between spherical and deformed nuclei is discussed in §4.5 (in particular, §4.5.4) of Ref. [18].

73. M. W. Guidry and Y. Sun, Superconductivity and superfluidity as universal emergent phenomena in diverse fermionic systems, *Front. Phys.* 10(4), 1 (2015)

74. M. W. Guidry, Universality of emergent states in diverse physical systems, *AIP Conf. Proc.* 1912, 020005 (2017)

75. L. A. Wu and M. W. Guidry, The ground state of monolayer graphene in a strong magnetic field, *Sci. Rep.* 6(1), 22423 (2016)

76. L. A. Wu, M. Murphy, and M. W. Guidry, $SO(8)$ fermion dynamical symmetry and strongly correlated quantum Hall states in monolayer graphene, *Phys. Rev. B* 95(11), 115117 (2017)

77. M. W. Guidry, $SO(8)$ fermion dynamical symmetry and quantum hall states for graphene in a strong magnetic field, *Fortschr. Phys.* 65(6–8), 1600057 (2017)

78. J. L. Tallon, J. W. Loram, J. R. Cooper, C. Panagopoulos, and C. Bernhard, Superfluid density in cuprate high-$T_c$ superconductors: A new paradigm, *Phys. Rev. B* 68(18), 180501 (2003)

79. P. Dai, H. A. Mook, S. M. Hayden, G. Aeppli, T. G. Perring, R. D. Hunt, and F. Doğan, The magnetic excitation spectrum and thermodynamics of high-$T_c$ superconductors, *Science* 284(5418), 1344 (1999)

80. J. C. Campuzano, H. Ding, M. R. Norman, H. M. Fretwell, M. Randeria, A. Kaminski, J. Mesot, T. Takeuchi, T. Sato, T. Yokoya, T. Takahashi, T. Mochiku, K. Kadowaki, P. Guptasarma, D. G. Hinks, Z. Konstantinovic, Z. Z. Li, and H. Raffy, Electronic spectra and their relation to the $(\pi, \pi)$ collective mode in high-$T_c$ superconductors, *Phys. Rev. Lett.* 83(18), 3709 (1999)

81. Z. A. Xu, N. P. Ong, Y. Wang, T. Kakeshita, and S. Uchida, Vortex-like excitations and the onset of superconducting phase fluctuation in underdoped $La_{2-x}Sr_xCuO_4$, *Nature* 406(6795), 486 (2000)

82. N. P. Ong, Y. Wang, S. Ono, Y. Ando, and S. Uchida, Vorticity and the Nernst effect in cuprate superconductors, *Ann. Phys.* 13(12), 9 (2004)

83. P. W. Anderson, The resonating valence bond state in $La_2CuO_4$ and superconductivity, *Science* 235(4793), 1196 (1987)

84. D. Vaknin, S. K. Sinha, D. E. Moncton, D. C. Johnston, J. M. Newsam, C. R. Safinya, and H. E. King, Antiferromagnetism in $La_2CuO_{4-y}$, *Phys. Rev. Lett.* 58(26), 2802 (1987)

85. T. S. Nunner, B. M. Anderson, A. Melikyan, and P. J. Hirschfeld, Dopant-modulated pair interaction in cuprate superconductors, *Phys. Rev. Lett.* 95(17), 177003 (2005)

86. A. C. Fang, L. Capriotti, D. J. Scalapino, S. A. Kivelson, N. Kaneko, M. Greven, and A. Kapitulnik, Gap-inhomogeneity-induced electronic states in superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$, *Phys. Rev. Lett.* 96(1), 017007 (2006)

87. Y. He, T. S. Nunner, P. J. Hirschfeld, and H. P. Cheng, Local electronic structure of $Bi_2Sr_2CaCu_2O_8$ near oxygen dopants: A window on the high-$T_c$ pairing mechanism, *Phys. Rev. Lett.* 96(19), 197002 (2006)

88. M. M. Maśka, Z. Sledz, K. Czajka, and M. Mierzejewski, Inhomogeneity-induced enhancement of the pairing interaction in cuprate superconductors, *Phys. Rev. Lett.* 99(14), 147006 (2007)

89. S. Petit and M. B. Lepetit, Real-space fluctuations of effective exchange integrals in high-$T_c$ cuprates, *Europhys. Lett.* 87(6), 67005 (2009)

90. K. Foyevtsova, R. Valentı, and P. J. Hirschfeld, Effect of dopant atoms on local superexchange in cuprate superconductors: A perturbative treatment, *Phys. Rev. B* 79(14), 144424 (2009)

91. S. Johnston, F. Vernay, and T. P. Devereaux, Impact of an oxygen dopant in $Bi_2Sr_2CaCu_2O_{8+\delta}$, *Europhys. Lett.* 86(3), 37007 (2009)

92. S. Okamoto and T. A. Maier, Microscopic inhomogeneity and superconducting properties of a two-dimensional Hubbard model for high-$T_c$ cuprates, *Phys. Rev. B* 81(21), 214525 (2010)

93. G. Khaliullin, M. Mori, T. Tohyama, and S. Maekawa, Enhanced pairing correlations near oxygen dopants in cuprate superconductors, *Phys. Rev. Lett.* 105(25), 257005 (2010)

94. J. W. Loram, J. L. Tallon, and W. Y. Liang, Absence of gross static inhomogeneity in cuprate superconductors, *Phys. Rev. B* 69, 060502(R) (2004)

95. J. Bobroff, H. Alloul, S. Ouazi, P. Mendels, A. Mahajan, N. Blanchard, G. Collin, V. Guillen, and J. F. Marucco, Absence of static phase separation in the high $T_c$ cuprate $YBa_2Cu_3O_{6+y}$, *Phys. Rev. Lett.* 89(15), 157002 (2002)

96. I. Bozovic, G. Logvenov, M. A. J. Verhoeven, P. Caputo, E. Goldobin, and M. R. Beasley, Giant proximity effect in cuprate superconductors, *Phys. Rev. Lett.* 93(15), 157002 (2004)

97. G. Alvarez, M. Mayr, A. Moreo, and E. Dagotto, Areas of superconductivity and giant proximity effects in underdoped cuprates, *Phys. Rev. B* 71(1), 014514 (2005)

98. I. Bozovic, G. Logvenov, M. A. J. Verhoeven, P. Caputo, E. Goldobin, and T. H. Geballe, No mixing of superconductivity and antiferromagnetism in a high-temperature superconductor, *Nature* 422(6934), 873 (2003)

99. E. Demler, A. J. Berlinsky, C. Kallin, G. B. Arnold, and M. R. Beasley, Proximity effect and Josephson coupling in the SO(5) theory of high-$T_c$ superconductivity, *Phys. Rev. Lett.* 80(13), 2917 (1998)

100. E. Dagotto, Complexity in strongly correlated electronic systems, *Science* 309(5732), 257 (2005)

101. M. R. Norman and C. Pépin, The electronic nature of high temperature cuprate superconductors, *Rep. Prog. Phys.* 66(10), 1547 (2003)

102. A. Damascelli, A. Hussain, and Z. X. Shen, Angle-resolved photoemission studies of the cuprate superconductors, *Rev. Mod. Phys.* 75(2), 473 (2003)

103. E. Dagotto, Correlated electrons in high-temperature superconductors, *Rev. Mod. Phys.* 66(3), 763 (1994)

104. M. R. Norman, H. Ding, M. Randeria, J. C. Campuzano, T. Yokoya, T. Takeuchi, T. Takahashi, T. Mochiku, K. Kadowaki, P. Guptasarma, and D. G. Hinks, Destruction of the Fermi surface in underdoped high-$T_c$ superconductors, *Nature* 392(6672), 157 (1998)

105. X. J. Zhou, T. Yoshida, D. H. Lee, W. L. Yang, V. Brouet, F. Zhou, W. X. Ti, J. W. Xiong, Z. X. Zhao, T. Sasagawa, T. Kakeshita, H. Eisaki, S. Uchida, A. Fujimori, Z. Hussain, and Z. X. Shen, Dichotomy between nodal and antinodal quasiparticles in underdoped $(La_{2-x}Sr_x)CuO_4$ superconductor, *Phys. Rev. Lett.* 92(18), 187001 (2004)

106. A. Kanigel, M. R. Norman, M. Randeria, U. Chatterjee, S. Souma, A. Kaminski, H. M. Fretwell, S. Rosenkranz, M. Shi, T. Sato, T. Takahashi, Z. Z. Li, H. Raffy, K. Kadowaki, D. Hinks, L. Ozyuzer, and J. C. Campuzano, Evolution of the pseudogap from Fermi arcs to the nodal liquid, *Nat. Phys.* 2(7), 447 (2006)

107. N. Doiron–-Leyraud, C. Proust, D. LeBoeuf, J. Levallois, J. B. Bonnemaison, R. Liang, D. A. Bonn, W. N. Hardy, and L. Taillefer, Quantum oscillations and the Fermi surface in an underdoped high-$T_c$ superconductor, *Nature* 447(7144), 565 (2007)

108. D. LeBoeuf, N. Doiron-Leyraud, J. Levallois, R. Daou, J. B. Bonnemaison, N. E. Hussey, L. Balicas, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy, S. Adachi, C. Proust, and L. Taillefer, Electron pockets in the Fermi surface of hole-doped high-$T_c$ superconductors, *Nature* 450(7169), 533 (2007)

109. E. A. Yelland, J. Singleton, C. H. Mielke, N. Harrison, F. F. Balakirev, B. Dabrowski, and J. R. Cooper, Quantum oscillations in the underdopedcuprate $YBa_2Cu_4O_8$, *Phys. Rev. Lett.* 100(4), 047003 (2008)

110. A. F. Bangura, J. D. Fletcher, A. Carrington, J. Levallois, M. Nardone, B. Vignolle, P. J. Heard, N. Doiron-Leyraud, D. LeBoeuf, L. Taillefer, S. Adachi, C. Proust, and N. E. Hussey, Small Fermi surface pockets in underdopedhigh temperature superconductors: Observation of Shubnikov–de Haas oscillations in $YBa_2Cu_4O_8$, *Phys. Rev. Lett.* 100(4), 047004 (2008)

111. C. Jaudet, D. Vignolles, A. Audouard, J. Levallois, D. LeBoeuf, N. Doiron-Leyraud, B. Vignolle, M. Nardone, A. Zitouni, R. Liang, D. A. Bonn, W. N. Hardy, L. Taillefer, and C. Proust, de Haas–van Alphen oscillations in the underdoped high-temperature superconductor $YBa_2Cu_3O_{6.5}$, *Phys. Rev. Lett.* 100(18), 187005 (2008)

112. S. R. Julian and M. R. Norman, Local pairs and small surfaces, *Nature* 447(7144), 537 (2007)

113. G. F. Chen, Z. Li, G. Li, J. Zhou, D. Wu, J. Dong, W. Z. Hu, P. Zheng, Z. J. Chen, H. Q. Yuan, J. Singleton, J. L. Luo, and N. L. Wang, Superconducting properties of the Fe-based layered superconductor $LaFeAsO_{0.9}F_{0.1-\delta}$, *Phys. Rev. Lett.* 101(5), 057007 (2008)

114. H. H. Wen, G. Mu, L. Fang, H. Yang, and X. Zhu, Superconductivity at 25 K in hole-doped $(La_{1-x}Sr_x)OFeAs$, *Europhys. Lett.* 82(1), 17009 (2008)

115. X. H. Chen, T. Wu, G. Wu, R. H. Liu, H. Chen, and D. F. Fang, Superconductivity at 43 K in $SmFeAsO_{1-x}F_x$, *Nature* 453(7196), 761 (2008)

116. G. F. Chen, Z. Li, D. Wu, G. Li, W. Z. Hu, J. Dong, P. Zheng, J. L. Luo, and N. L. Wang, Superconductivity at 41 K and its competition with spin-density-wave instability in layered $CeO_{1-x}F_xFeAs$, *Phys. Rev. Lett.* 100(24), 247002 (2008)

117. Z. A. Ren, J. Yang, W. Lu, W. Yi, X. L. Shen, Z. C. Li, G. C. Che, X. L. Dong, L. L. Sun, F. Zhou, and Z. X. Zhao, Superconductivity in the iron-based $F$-doped layered quaternary compound $Nd[O_{1-x}F_x]FeAs$, *Europhys. Lett.* 82(5), 57002 (2008)

118. G. F. Chen, Z. Li, D. Wu, J. Dong, G. Li, W. Z. Hu, P. Zheng, J. L. Luo, and N. L. Wang, Element substitution effect in transition metal oxypnictide $Re(O_{1-x}F_x)TAs$ (Re = rare earth, T = transition metal), *Chin. Phys. Lett.* 25(6), 2235 (2008)

119. M. Daghofer, A. Moreo, J. A. Riera, E. Arrigoni, D. J. Scalapino, and E. Dagotto, Model for the magnetic order and pairing channels in Fe pnictide superconductors, *Phys. Rev. Lett.* 101(23), 237004 (2008)

120. A. Moreo, M. Daghofer, J. A. Riera, and E. Dagotto, Properties of a two-orbital model for oxypnictide superconductors: Magnetic order, $B_{2g}$ spin-singlet pairing channel, and its nodal structure, *Phys. Rev. B* 79(13), 134502 (2009)

121. K. Nakayama, T. Sato, P. Richard, Y. M. Xu, Y. Sekiba, S. Souma, G. F. Chen, J. L. Luo, N. L. Wang, H. Ding, and T. Takahashi, Superconducting gap symmetry of $Ba_{0.6}K_{0.4}Fe_2As_2$ studied by angle-resolved photoemission spectroscopy, *Europhys. Lett.* 85(6), 67002 (2009)

122. H. Ding, P. Richard, K. Nakayama, K. Sugawara, T. Arakane, Y. Sekiba, A. Takayama, S. Souma, T. Sato, T. Takahashi, Z. Wang, X. Dai, Z. Fang, G. F. Chen, J. L. Luo, and N. L. Wang, Observation of Fermi-surface-dependent nodeless superconducting gaps in $Ba_{0.6}K_{0.4}Fe_2As_2$, *Europhys. Lett.* 83(4), 47001 (2008)

123. L. Zhao, H. Y. Liu, W. T. Zhang, J. Q. Meng, X. W. Jia, G. D. Liu, X. L. Dong, G. F. Chen, J. L. Luo, N. L. Wang, W. Lu, G. L. Wang, Y. Zhou, Y. Zhu, X. Y. Wang, Z. Y. Xu, C. T. Chen, and X. J. Zhou, Multiple nodeless superconducting gaps in $(Ba_{0.6}K_{0.4})Fe_2As_2$ superconductor from angle resolved photoemission spectroscopy, *Chin. Phys. Lett.* 25(12), 4402 (2008)

124. K. Umezawa, Y. Li, H. Miao, K. Nakayama, Z. H. Liu, P. Richard, T. Sato, J. B. He, D. M. Wang, G. F. Chen, H. Ding, T. Takahashi, and S. C. Wang, Unconventional anisotropic *s*-wave superconducting gaps of LiFeAs iron-pnictide superconductor, *Phys. Rev. Lett.* 108(3), 037002 (2012)

125. Z. H. Liu, P. Richard, K. Nakayama, G. F. Chen, S. Dong, J. B. He, D. M. Wang, T. L. Xia, K. Umezawa, T. Kawahara, S. Souma, T. Sato, T. Takahashi, T. Qian, Y. Huang, N. Xu, Y. Shi, H. Ding, and S. C. Wang, Unconventional superconducting gap in $NaFe_{0.95}Co_{0.05}As$ observed by angle-resolved photoemission spectroscopy, *Phys. Rev. B* 84(6), 064519 (2011)

126. J. Hu and N. Hao, $S_4$ symmetric microscopic model for iron-based superconductors, *Phys. Rev. X* 2(2), 021009 (2012)

127. P. W. Anderson, When the electron falls apart, *Phys. Today* 50(10), 42 (1997)

128. S. C. Zhang, A unified theory based on SO(5) symmetry of superconductivity and antiferromagnetism, *Science* 275(5303), 1089 (1997)

129. There also is the U(1) generator of charge density waves in our full $U(4)\supset U(1)\times SU(4)$ algebra that does not appear in the Zhang SO(5) algebra.

130. S. C. Zhang, J. P. Hu, E. Arrigoni, W. Hanke, and A. Auerbach, Projected SO(5) models, *Phys. Rev. B* 60(18), 13070 (1999)

131. The broken particle number symmetry can be restored by particle-number projection, but in practice this procedure may not be necessary as we are dealing with a system having a very large number of fermions.