Towards full simulations of high-temperature superconductors

T.A. Maier, J.B. White III, T.C. Schulthess
Center for Computational Sciences and Computer Science and Mathematics Division, Oak Ridge National Laboratory

M. Jarrell
University of Cincinnati

Abstract

The Cray X1 in the Center for Computational Sciences at Oak Ridge National Laboratory is enabling significant new science in the simulation of high-temperature "cuprate" superconductors. We describe the method of dynamic cluster approximation with quantum Monte Carlo, along with its computational requirements. We then show the unique capabilities of the X1 for supporting this method, porting experiences, performance, and the resulting new scientific results.

1 Introduction

Despite years of active research, the understanding of superconductivity in the high-temperature "cuprate" superconductors (HTSC) remains one of the most important outstanding problems in materials science. A complete theoretical understanding of cuprate superconductors could lead to the ability to design and synthesize room-temperature superconductors, which would have tremendous technological implications. In the superconducting state of a material, electrons form so-called Cooper pairs, allowing them to condense into a coherent macroscopic quantum state in which they conduct electricity without resistance. In conventional superconductors, pairing results from an attractive interaction between electrons that is mediated by lattice vibrations (phonon-mediated pairing).

The consensus today is that the pairing mechanism in high-temperature superconductors is of an entirely different nature and is probably related to strong correlations between electrons, a feature that distinguishes these materials from conventional superconductors. To address the problem theoretically, one must solve the quantum many-body problem for a macroscopic number of electrons without being limited to the typical single-particle approximations, such as Hartree-Fock or the local density approximation to density functional theory. A recent concurrence of new algorithmic developments and significant improvements in computational capability has opened a clear path to solving the quantum many-body problem for high-temperature superconductors.

2 Hubbard model

The characteristic feature of all HTSC is a strongly anisotropic layered perovskite-like crystal structure with conducting CuO_2-planes separated by insulating layers of other elements (see right part of Fig. 1). Superconductivity takes place within the two-dimensional CuO_2 layers with the insulating barriers only providing charge carriers, usually holes to the layers and thus controlling the doping of CuO_2 planes.

First-principles calculations for HTSC compounds provide evidence that the band which crosses the Fermi surface has mainly CuO_2 character (see e.g. [1] and references therein). To reduce the complexity of the problem it thus seems reasonable to restrict calculations to a two-dimensional model with electrons moving in a single CuO_2 layer. Justified by the strong in-plane CuO bonds, the complexity may be further reduced by constructing a model that treats a whole CuO_2 plaquette as a single site. The resulting two-dimensional Hubbard model [2] is believed to capture the essential physics of HTSC [3, 4, 5]. A schematic of its Hamiltonian,

$$H = -t \sum_{\langle ij \rangle} c^{\dagger}_{i\sigma} c_{\jmath\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow} \qquad (1)$$

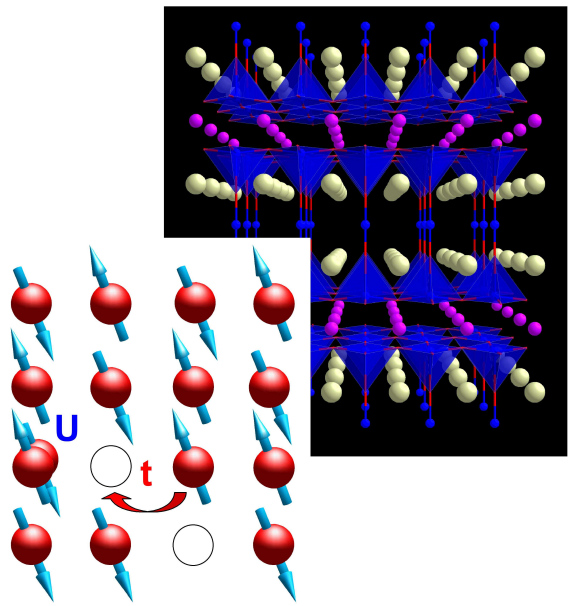

Figure 1: Crystal structure of $YBa_2Cu_3O_5$ and two-dimensional Hubbard model of the hole doped CuO_2 planes, with nearest neighbor hopping integral t and on-site Coulomb interaction U.

is illustrated in Fig. 1. The fermionic operator $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) creates (destroys) an electron on site i with spin σ, and $n_{i\sigma} = c^\dagger_{i\sigma} c_{\imath\sigma}$ is the corresponding number operator. The first term describes the hybridization between sites with amplitude t, and the second term the Coulomb repulsion between two electrons residing on the same site. Because of screening, the magnitude of longer-ranged interactions is believed to be small compared to the on-site interaction.

Despite decades of intensive studies, this model remains unsolved except in one or infinite dimensions. Analytical methods based on a perturbative approaches suffer from the large magnitude of U, which renders these calculations at least questionable. Many theorists have turned to numerical approaches to close the gap between the model defined by its Hamiltonian and its properties. A large body of work has been devoted to a direct (numerically) exact solution of finite-size systems using exact diagonalization or Quantum Monte Carlo (QMC) methods (for a review see [6]). Exact diagonalization, however, is severely limited by the exponential growth of computational effort with system size, while QMC methods suffer from what is known as "the sign problem" at low temperatures. Another difficulty of these methods arises from their strong finite-size effects, often ruling out the reliable extraction of low-energy scales, which are important to capture the competition between different ground states often present in correlated electron systems.

3 Dynamical Cluster Approximation

Mean-field theories are defined in the thermodynamic limit and therefore do not face the finite-size problems. Generally, mean-field theories divide the infinite number of degrees of freedom into two sets. A small set of degrees of freedom is treated explicitly, while the effects of the remaining degrees of freedom are summarized as a mean field acting on the first set. The Dynamical Mean-Field Theory (DMFT) [7, 8] (for a review see [9]) for itinerant correlated systems (such as the HTSC or systems described by the model Eq. (1)) is analogous to the coherent potential approximation for disordered systems [10, 11, 12]. It retains the dynamics of local degrees of freedom by mapping the lattice onto an impurity self-consistently embedded in a dynamical mean-field host.

Despite its success in the description of many correlated phenomena such as the Mott-Hubbard transition, the DMFT and CPA share the critical flaw of neglecting the effects of non-local fluctuations. Thus the DMFT is unable to capture the effects of *e.g.* spin-waves in magnetic systems, localization in disordered systems, or spin-liquid physics in correlated electron systems. Furthermore it cannot capture phase transitions to states with non-local order parameters, such as the d-wave superconducting phase in the HTSC. Non-local corrections are required to treat even the initial effects of these phenomena.

Here we use the Dynamical Cluster Approximation (DCA) [13, 14, 15, 16] (for a review see [17]) to study the properties of the Hubbard model, Eq. (1). The DCA extends the DMFT by non-local correlations. Instead of mapping the lattice onto a single impurity, the system is mapped onto a periodic cluster of size N_c coupled to a mean-field host representing the remaining degrees of freedom (see Fig. 2). As a result, dynamical correlations up to a range limited by the cluster size are treated accurately, while the physics on longer length scales is described on the mean-field level. Translational invariance of the original system assures that the quantity describ-

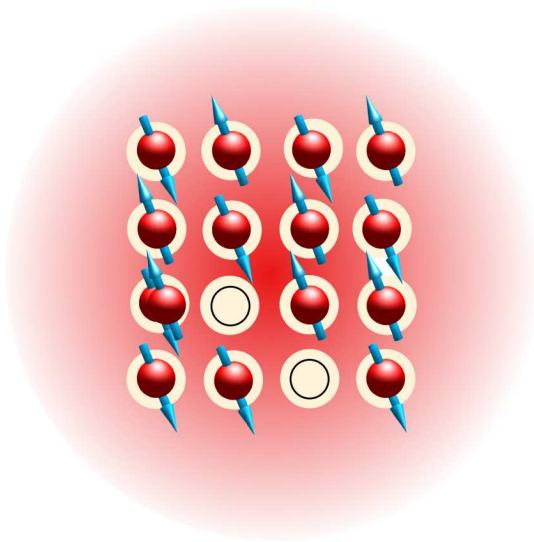

Figure 2: Schematic illustration of the DCA formalism. The model is mapped onto a finite-size cluster self-consistently coupled to a mean-field host. Correlations within the cluster are treated accurately while the physics on length scales beyond the cluster size is described on the mean-field level.

ing the mean-field host can be self-consistently determined from the solution of the cluster problem. The complexity of the original problem with an infinite number of degrees of freedom is thus reduced to a self-consistent finite-size cluster problem with N_c degrees of freedom. The remaining cluster problem may then be solved numerically by a number of techniques including the QMC method [16] used here.

4 Small Clusters

Computations with a cluster of only four sites, the smallest cluster that can capture superconductivity with a d-wave order parameter, on the IBM p690 at the Center for Computational Sciences (CCS) show very good general agreement with HTSC. These results are summarized in the temperature-doping phase-diagram shown in Fig. 3 (see also [18, 19]). At low doping, δ, the system is an antiferromagnetic insulator below the Neél temperature $T_{\rm N}$. At finite doping, $\delta \leq 0.3$, an instability is found at the critical temperature T_c to a superconducting state described by a $d_{x^2-y^2}$-wave order parameter. In the normal state, low-energy spin excitations be-

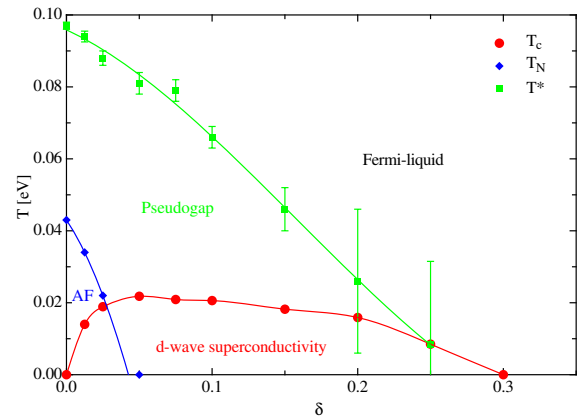

Figure 3: DCA/QMC Temperature-doping phase diagram of the two-dimensional Hubbard model when $t = 0.25$ eV, $U = 2$ eV for a 4-site cluster. Consistent with experiments on HTSC, regions of antiferromagnetism, d-wave superconductivity and pseudogap behavior are found.

come suppressed below the crossover temperature T^*. Simultaneously the electronic excitation spectrum displays a pseudogap, *i.e.* a partial suppression of low-energy spectral weight. Consistent with optical experiments, computations for a four-site cluster show that the superconducting transition is accompanied by a lowering of the electronic kinetic energy [20]. This result further shows the unconventional character of superconductivity in these systems. It is fundamentally different from the BCS theory for conventional superconductors [21], where pairing occurs through a reduction of the electronic potential energy accompanied by a slight increase in kinetic energy.

The apparent violation of the Mermin-Wagner theorem [22], according to which no phases with conventional long-range order can occur at finite temperatures in the two-dimensional Hubbard model, is a consequence of the small cluster size, and hence large mean-field character, in these simulations. In the case of antiferromagnetism, the Mermin-Wagner theorem thus necessarily translates to $T_{\rm N} = 0$ for the two-dimensional system. Superconductivity however can exist even at finite temperatures as topological order below the Kosterlitz-Thouless transition temperature [23]. Therefore, larger-cluster-size studies are needed to see if the simulations recover the Mermin-Wagner theorem and if superconductivity survives as topological order in the infinite cluster size limit where the DCA becomes exact.

In the HTSC, on the other hand, a small but finite coupling between the two-dimensional CuO_2 layers induces long-range order at finite temperatures.

5 Porting and Performance

The central quantity of the DCA code is the single-particle cluster Green function $\mathbf{G}_c$, which is a matrix of size $N \times N$ [16]. Here $N = N_c \times N_l$ where N_l is the number of "time-slices" in the time direction. The majority of the CPU time is spent in the inner loop of the QMC simulation, which updates the Green function matrix according to the vector outer product

$$\mathbf{G}' = \mathbf{G} + \mathbf{a} * \mathbf{b}^T \,, \tag{2}$$

where $\mathbf{a}$ and $\mathbf{b}$ are two vectors of dimension N. This computation is handled by the BLAS [24] call DGER, which performs a double-precision rank-one matrix update representing $O(N^2)$ operations. Each iteration requires N such calls, however, resulting in $O(N^3)$ operations.

Another CPU-intensive task is the evaluation of two-particle correlation functions. In the QMC technique this reduces to evaluating products of Green functions and thus to computing matrix products. This is done by using the BLAS call CGEMM, which performs single-precision complex matrix-matrix multiplication, and one call again is $O(N^3)$.

Porting and tuning the DCA implementation on the Cray X1 was straightforward. The port required no changes beyond the "`Makefile`", and tuning involved performance profiling and adding "`concurrent`" directives to one file. This file contains a number of nested loops using indirect addressing, or index arrays. The bulk of the tuning effort was in determining which loops did and did not iterate over repeated indices.

The Cray X1 has a number of advantages over general-purpose systems in performing DCA computations, particularly with increasing cluster size. This advantage is demonstrated in Fig. 4, which compares runtimes of some early DCA runs on the X1 and the IBM p690 in the CCS, using 8 and 32 processors (MSPs) on each.The figure shows runtimes for production runs with a fixed value of $N_c = 64$ and increasing values of N_l and thus N, where the value of N is shown. Eight X1 MSPs easily outperform thirty-two 1.3-GHz Power4 processors for the larger problem sizes.

As discussed above, the DCA implementation includes two $O(N^3)$ computations built on the BLAS calls CGEMM and DGER. CGEMM is a BLAS3 call, which implies that it can be blocked effectively for cache memory, and many modern general-purpose processors can perform the operations near their peak. The X1 processors can also, but they have the added benefit of a very high peak rate augmented by the ability to perform single-precision operations at twice the rate of double-precision.

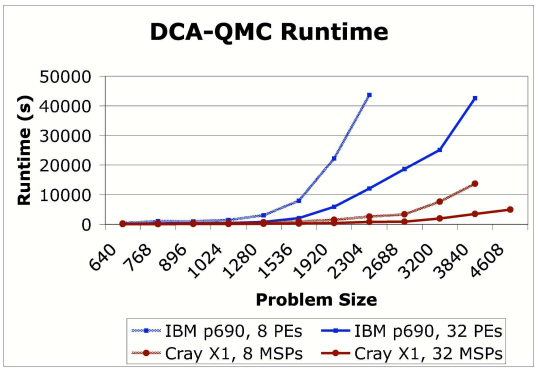

Figure 4: Runtimes for a series of DCA production runs. Each run is indicated by its value of $N = N_c N_l$. The lines connecting the data points are only guides to the eye.

The X1 has a more significant advantage over the prevailing cache-dependent architectures in the DGER operations. Each call depends on the results of the previous call, so the operations cannot be interleaved. DGER is a BLAS2 call, which implies that it does much fewer computations per memory access than CGEMM, and thus is typically limited by memory bandwidth.

We conducted separate DGER benchmarks to measure the advantage of the X1 in this operation, and results for the CCS Cray X1, SGI Altix (1.5 GHz Itanium2), and IBM p690 (1.3 GHz Power4) are in Fig. 5. The vendor-optimized DGER was used for each system. The figure shows the performance of DGER for a matrix of size $N = 64 \times 70 = 4480$, which is representative of large DCA runs. Separate DGER instances were run concurrently across increasing numbers of processors (MSPs), mimicking the processes of a Monte-Carlo simulation. The X1 memory system is able to maintain performance and efficiency with added processors, while the p690 steadily degrades. The Altix degrades going from one to two processors because memory bandwidth is

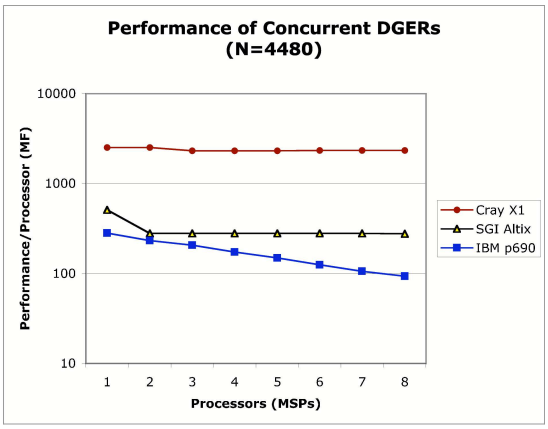

Figure 5: Per-processor performance of concurrent DGER calls using $N = 4480$ matrices.

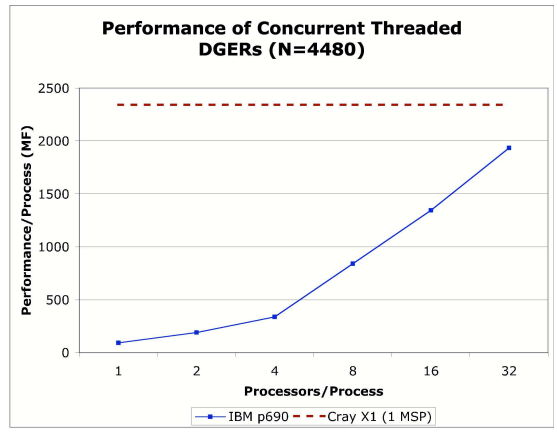

Figure 6: DGER performance of a fully loaded IBM p690 using different numbers of processors per threaded process. The dotted line is the per-MSP performance of loaded X1 nodes.

shared between processor pairs. The X1 maintains 8–25 times the performance and 4–10 times the efficiency of the other systems.

Despite the Monte-Carlo nature of the DCA algorithm, the X1 also has an important scalability advantage over systems with weaker processors. Each DCA process has a significant fixed start-up cost, which favors splitting the Monte-Carlo iterations across fewer, faster processors.

Another option would be to multithread each Monte-Carlo process, effectively using an SMP as a large single "processor". We explore this possibility in Fig. 6, which shows the performance of IBM's multithreaded DGER on a p690, again using a matrix size of $N = 4480$. Fig. 6. The dashed line shows the per-MSP performance of an X1 performing concurrent DGER operations on 32 MSPs, thus simulating a loaded system. The solid line shows the performance of a 32-processor p690 loaded with concurrent DGER computations, but using different numbers of processors per DGER process.

The left-most point thus shows the performance of a single processor when all 32 processors of the p690 are performing independent DGER operations, while the right-most point shows the aggregate performance of dedicating all 32 processors to a single DGER. The figure indicates that dedicating a full IBM p690 to each DGER does not match the performance of a single X1 MSP. No threaded version of vendor-optimized DGER was available for the Altix or the X1 at the time of this test. Tests of untuned DGER implemented with Fortran loops and OpenMP showed little improvement on the X1 for matrices of size 4480, and the Fortran/OpenMP implementation on the Altix was not competitive with the single-threaded vendor-optimized DGER.

The significant performance advantage of the X1 for DCA computations, as illustrated by its dominance in DGER performance, has allowed us to perform simulations that are out of the reach of other systems, all without having to resort to hybrid parallelization. In particular, the X1 has provided the capability needed to perform DCA computation with much-larger cluster sizes.

6 Larger Clusters

As discussed in Sec. 4, the DCA retains a large mean-field character at small cluster sizes and consequently yields long-range order at finite temperatures. Long wave-length modes which destroy long-range order at finite temperatures in two-dimensional systems are neglected. With increasing cluster size, however, the DCA progressively includes these longer-ranged fluctuations. These are expected to drive the Neél temperature systematically to zero and thus recover the Mermin-Wagner theorem in the infinite cluster size limit where the DCA becomes exact.

Fig. 7 displays the DCA results for the Neél temperature T_{N} as a function of the inverse of the linear cluster size $L_c = \sqrt{N_c}$. With increasing cluster size, T_{N} rapidly decreases and extrapolates to $T_{\mathrm{N}} = 0$

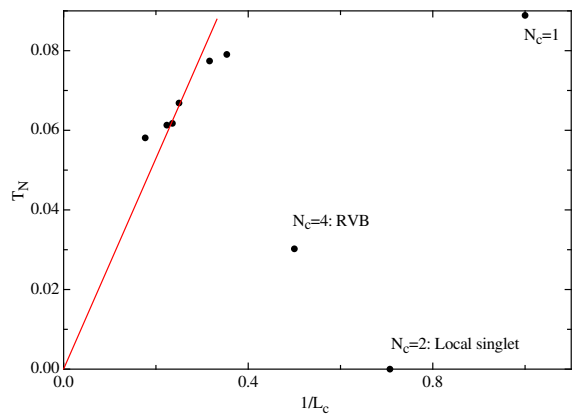

Figure 7: Neél temperature at 5% doping as a function of the inverse linear cluster size $1/L_c = 1/\sqrt{N_c}$.

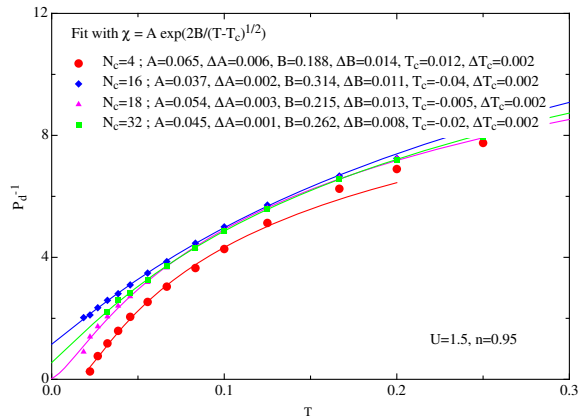

Figure 8: DCA/QMC result for the inverse d-wave pair-field susceptibility as a function of temperature for different cluster sizes at 5% doping when $t = 0.25$ and $U = 1.5$. Superconductivity is suppressed at cluster sizes $N_c > 4$.

in the infinite cluster size limit consistent with the Mermin-Wagner theorem. The data points scatter about a curve linear in $1/L_c$, except for the "special" cluster sizes $N_c = 1, 2$ and 4. For $N_c = 2$ a local singlet is formed on the cluster. When $N_c = 4$ the ground state of the periodic cluster is a resonating valence-bond state [25] with fluctuating singlet bonds between the cluster sites. Hence antiferromagnetic order is suppressed for these cluster sizes, and the results thus do not fall on the curve.

As discussed in Sec. 4, superconductivity may persist in the infinite cluster size limit as topological Kosterlitz-Thouless order, although no conventional long-range order is allowed.

The transition to a superconducting state with d-wave symmetry is indicated by the divergence of the pair-field susceptibility P_d, or equivalently by the node of P_d^{-1}. The DCA result for this quantity at 5% doping is plotted in Fig. 8 for different cluster sizes N_c. At finite cluster sizes N_c, the critical behavior found in the DCA at temperatures close to the transition temperature T_c defined by the node in P_d^{-1} has to be mean-field like, *i.e.* $P_d^{-1} \propto |T - T_c|$, since the long-ranged physics is treated on the mean-field level. At higher temperatures however, where the correlation length is within the cluster size, the true critical behavior may be observed in the DCA. Therefore we fit the DCA results at intermediate temperatures with the function $\chi = A \exp(1B/(T - T_c)^{0.5})$, the critical behavior expected for a Kosterlitz-Thouless transition [26]. At lower temperatures we expect the linear dependence to connect smoothly to this behavior.

When $N_c = 4$, the DCA predicts a transition to a d-wave superconducting state at a finite temperature T_c[1]. When $N_c > 4$, however, the results seem to indicate the absence of a finite temperature transition to a superconducting state at 5% doping.

Fig. 9 shows the results for the 15% doped system. Again, for $N_c = 4$ a superconducting transition is obtained at a finite temperature. In contrast to the 5% doped case, the results are almost converged, *i.e.* independent of cluster size, when $N_c > 4$. This is a clear indication that at this doping correlations are short-ranged and do not extend beyond the cluster size at the temperatures studied. Clearly, the results are incompatible with superconductivity at finite temperatures at 15% doping, even if the lowest temperatures data points are extrapolated linearly.

Based on these results we infer that, despite its tendency to exhibit d-wave pairing, the purely two-dimensional Hubbard model is not enough to describe high-temperature superconductivity. We conclude that either a coupling to the third dimension, a more realistic modeling of the electronic structure, the additional inclusion of lattice degrees of freedom or even a combination of these extensions is necessary to stabilize superconductivity in the infinite cluster size limit. Work along these lines is in progress.

[1] Note however that the fit function for temperatures close to T_c changes curvature and therefore underestimates the actual T_c predicted by the DCA

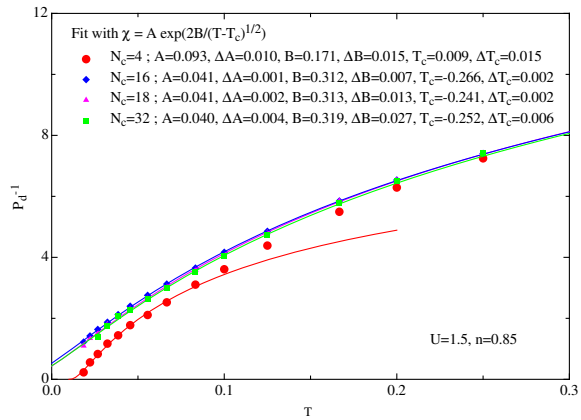

Figure 9: DCA/QMC result for the inverse d-wave pair-field susceptibility as a function of temperature for different cluster sizes at 15% doping when $t = 0.25$ and $U = 1.5$. The results are almost converged for $N_c > 4$, where superconductivity is strongly suppressed

7 Summary and Conclusions

The Cray X1 in the Center for Computational Sciences at Oak Ridge National Laboratory has enabled significant new progress in the understanding of HTSC within a minimal microscopic model, the two-dimensional Hubbard model. DCA/QMC simulations at small cluster size $N_c = 4$ show very good general agreement with HTSC, including superconductivity at high temperatures. Due to the small cluster size however, the results violate the Mermin-Wagner theorem, according to which no long-range order is allowed at finite temperatures in the two-dimensional model. The significant performance advantage of the X1 for the DCA/QMC computations has provided the capability to study much larger cluster sizes. Recent runs on the Cray X1 show that, with larger clusters, relevant longer-ranged fluctuations are captured, and the Mermin-Wagner theorem is recovered. Furthermore, the results show the absence of a finite-temperature superconducting transition, and thus are incompatible with a possible Kosterlitz-Thouless transition to a phase with topological order.

These problems in the description of HTSC may be overcome by carrying out fully three-dimensional calculations with an infinite set of Hubbard planes coupled along the third dimension. To eventually enable the design of new and optimized superconductors, we further plan to parameterize the model with *ab-initio* electronic-structure calculations and to generalize the method to include multiple bands.

8 Acknowledgments

This research used resources of the Center for Computational Sciences and was sponsored in part by the offices of Advance Scientific Computing Research and Basic Energy Sciences, U.S. Department of Energy. The work was performed at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725, and where TM is a Eugene P. Wigner Fellow. Work at Cincinnati was supported by the NSF Grant No. DMR-0113574.

References

[1] P. Fulde, *Electron Correlations in Molecules and Solids*, Springer Series in Solid-State Sciences (Springer Verlag, Berlin/Heidelberg/New York, 1991).

[2] J. Hubbard, *Proc. Royal. Soc. London* **276**, 238 (1963).

[3] F. Zhang, T. Rice, *Phys. Rev. B* **37**, 3759 (1988).

[4] F. Zhang, T. Rice, *Phys. Rev. B* **41**, 7243 (1990).

[5] P. W. Anderson, *The Theory of Superconductivity in the High-T_c Cuprates* (Princeton University Press, Princeton, NJ, 1997).

[6] E. Dagotto, *Rev. Mod. Phys* **66**, 763 (1994).

[7] W. Metzner, D. Vollhardt, *Phys. Rev. Lett.* **62**, 324 (1989).

[8] E. Müller-Hartmann, *Z. Phys.* **B 74**, 507 (1989).

[9] A. Georges, G. Kotliar, W. Krauth, M. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996).

[10] P. Soven, *Phys. Rev.* **156**, 809 (1967).

[11] D. Taylor, *Phys. Rev.* **156**, 1017 (1967).

[12] H. Shiba, *Prog. Theo. Phys.* **46**, 77 (1971).

[13] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, H. R. Krishnamurthy, *Phys. Rev. B* **58**, R7475 (1998).

[14] M. H. Hettler, M. Mukherjee, M. Jarrell, H. R. Krishnamurthy, *Phys. Rev. B* **61**, 12739 (2000).

[15] T. Maier, M. Jarrell, T. Pruschke, J. Keller, *Eur. Phys. J B* **13**, 613 (2000).

[16] M. Jarrell, T. Maier, C. Huscroft, S. Moukouri, *Phys. Rev. B* **64**, 195130 (2001).

[17] T. Maier, M. Jarrell, T. Pruschke, M. Hettler, *preprint cond-mat/0404055* (2004).

[18] T. Maier, M. Jarrell, T. Pruschke, J. Keller, *Phys. Rev. Lett.* **85**, 1524 (2000).

[19] M. Jarrell, T. Maier, M. H. Hettler, A. N. Tahvildarzadeh, *Europhys. Lett.* **56**, 563 (2001).

[20] T. Maier, M. Jarrell, A. Macridin, C. Slezak, *Phys. Rev. Lett.* **92**, 027005 (2004).

[21] J. Schrieffer, *Theory of Superconductivity* (Addison Wesley, Reading, MA, 1993).

[22] N. Mermin, H. Wagner, *Phys. Rev. Lett.* **22**, 1133 (1966).

[23] J. Kosterlitz, D. Thouless, *J. Phys. C* **6**, 1181 (1973).

[24] J. J. Dongarra, J. D. Croz, I. S. Duff, S. Hammarling, *ACM Trans. Math. Soft.* **16**, 1 (1990).

[25] P. W. Anderson, *Science* **235**, 1196 (1987).

[26] J. Kosterlitz, *J. Phys. C* **7**, 1046 (1973).