Dynamical slave-boson mean-field study of the Mott transition in the Hubbard model in the large-z limit

Sen Zhou, Long Liang, and Ziqiang Wang

Dynamical slave-boson mean-field study of the Mott transition in the Hubbard model in the large-z limit

Sen Zhou[1,2], Long Liang[3], and Ziqiang Wang[4]

[1] *CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics,*
Chinese Academy of Sciences, Beijing 100190, China
[2] *School of Physical Sciences & CAS Center for Excellence in Topological Quantum Computation,*
University of Chinese Academy of Sciences, Beijing 100049, China
[3] *COMP Centre of Excellence, Department of Applied Physics,*
Aalto University School of Science, FI-00076 Aalto, Finland and
[4] *Department of Physics, Boston College, Chestnut Hill, MA 02467, USA*
(Dated: November 18, 2019)

The Mott metal-insulator transition in the Hubbard model is studied by constructing a dynamical slave-boson mean-field theory in the limit of large lattice coordination number z that incorporates the binding between doubly occupied (doublon) and empty (holon) sites. On the Mott insulating side where all doublons and holons bond in real space into excitonic pairs leading to the charge gap, the theory simplifies considerably to leading order in $1/\sqrt{z}$, and becomes exact on the infinite-z Bethe lattice. An asymptotic solution is obtained for a continuous Mott transition associated with the closing of the charge gap at a critical value of the Hubbard U_c and the corresponding doublon density n_d^c, hopping χ_d^c and doublon-holon pairing Δ_d^c amplitudes. We find $U_c = U_{\mathrm{BR}}[1 - 2n_d^c - \sqrt{z}(\chi_d^c + \Delta_d^c)] \simeq 0.8U_{\mathrm{BR}}$, where U_{BR} is the critical value for the Brinkman-Rice transition in the Gutzwiller approximation captured in the static mean-field solution of the slave-boson formulation of Kotliar and Ruckenstein. Thus, the Mott transition can be viewed as the quantum correction to the Brinkman-Rice transition due to doublon-holon binding. Quantitative comparisons are made to the results of the dynamical mean-field theory, showing good agreement. In the absence of magnetic order, the Mott insulator is a $U(1)$ quantum spin liquid with nonzero intersite spinon hopping that survives the large-z limit and lifts the 2^N-fold degeneracy of the local moments. We show that the spinons are coupled to the doublons/holons by a dissipative compact $U(1)$ gauge field in the deconfined phase, realizing the spin-charge separated gapless spin liquid Mott insulator.

PACS numbers: 71.10.-w, 71.10.Fd, 71.27.+a, 74.70.-b

I. INTRODUCTION

A Mott insulator is a fundamental quantum electronic state driven by large Coulomb repulsion[1–3]. It is protected by a nonzero energy gap for charge excitations, but not associated with any symmetry breaking. A Mott insulator differs from the other class of correlation-driven insulators (e.g., magnets), better termed as Landau insulators, whose origins require symmetry breaking order parameters produced by the residual quasiparticle (QP) interactions from a parent Fermi liquid state. The most striking feature of a Mott insulator is the separation of charge and spin degrees of freedom of an electron that completely destroys the coherent QP excitations. A ubiquitous example of Mott insulator is the quantum spin liquid (QSL) where the spins are short-range correlated but do not exhibit any symmetry-breaking long-range order[4–7]. Overwhelming evidence for QSLs has been observed in the κ-organics *near* the Mott metal-insulator transition[8–11] and in frustrated quantum magnets[12–14] that are deep in the Mott insulating state.

The Mott insulator and the Mott transition are at the heart of the strong correlation physics since it is conceivable that the Mott insulator is the ultimate parent phase of strong correlation from which many novel quantum states can emerge. Indeed, strong correlation often results in an insulating ground state with anti-ferromagnetic long-range order, where the low-energy physics is described by the Heisenberg type of spin models, which can be viewed as instabilities of the spin liquids due to the condensation of low energy spin excitations in the Mott insulator. In addition to QSLs and magnetic ordered states, doping a Mott insulator can lead to the pseudogap phenomenon and unconventional high-T_c superconductivity[15–20].

The prototypical model for the Mott physics is the half-filled single-band Hubbard model with purely onsite Coulomb repulsion U. The Hilbert space is thus a product of the local Hilbert space on a single lattice site that consists of the doubly occupied (doublon), empty (holon), and singly occupied (spinon) states. The excitonic binding between the oppositely charged doublons (D) and holons (H) is believed to play an essential role in describing the Mott insulator and the Mott transition in strongly correlated Mott-Hubbard systems. This idea was advocated sometime ago[1,21–23] and studied in the context of improved variational Gutzwiller wave functions[24,25]. More recently, the idea has been made more explicit in the field theory description[26], improved saddle-point solution[27] of the slave-boson functional integral formulation of the Hubbard model[28], and other numerical[29,30] approaches.

The slave-boson representation by Kotliar-Ruckenstein (KR) is an exact functional integral formulation of

the Hubbard model. By ignoring all intersite correlations and condensing all bosons, the static saddle-point solution[28] correctly captures the result of Gutzwiller approximation[31] and gives rise to the Brinkman-Rice (BR) metal-insulator transition at U_{BR}[28,32,33] where the renormalized mass of the QPs diverges. This is, however, a rather crude approximation of the Mott transition, since the interactions between the doublons and holons as well as the incoherent excitations have been ignored. In an effort to go beyond the Gutzwiller approximation and the BR picture, an improved saddle point solution incorporating intersite correlations was constructed in Ref. 27 on two-dimensional bipartite lattices. It was elucidated that the doublon-holon (D-H) binding governs the incoherent excitations and plays a key role in the Mott transition. On the Mott insulator side at large U, although the D/H condensate vanishes, together with the disappearance of the coherent QPs, the D/H density remains nonzero, but with all the doublons bond to the holons. With decreasing U, the D/H density increases and D-H binding energy decreases. At a critical value U_c, the D-H excitation gap closes and a D/H single-particle condensate starts to develop, marking the onset of the Mott transition. Despite the success in capturing the essential Mott physics, the intersite correlations such as the D-H binding were incorporated in a uncontrolled way due to the lack of a small parameter. The results obtained in Ref. 27 are thus quantitatively unreliable.

In this work, we construct a dynamical slave-boson mean-field (SBMF) theory in which the D-H binding and hence the dynamical fluctuations can be studied systematically in a controlled manner (Section II). It turns out that the theory simplifies considerably on the Mott insulating side to leading order in $1/\sqrt{z}$ and becomes exact in the large-z limit. We therefore study the Mott transition from the large-U Mott insulating side in the large-z limit (Section III.A). The asymptotic solution obtained on the infinite-z Bethe lattice exhibits a continuous Mott transition from an insulating QSL to a correlated metal, where the closing of the Mott gap and the onset of the QP coherence coincide at the same U_c. We demonstrate that in the presence of D-H binding captured by the dynamical SBMF theory, the BR transition is preempted by the Mott transition since $U_c < U_{\mathrm{BR}}$, and the Mott insulator is characterized by the incoherent upper and lower Hubbard bands separated by the Mott gap. A key feature of the asymptotic solution is that on the insulating side of the Mott transition, quantum spin fluctuations via the intersite spinon correlations remains and survives in the large-z limit. Various physical quantities are calculated from simple and transparent expressions analytically at the transition point U_c, as well as to leading order in the Mott insulator at $U > U_c$. The results are quantitatively compared to and found to agree well with those obtained from the dynamical mean-field theory[34] (DMFT) with various numerical quantum impurity solvers, which is exact in the large-z limit. In Section III.B, we derive the effective action for the compact gauge field in the large-z

limit and show that the emergent dissipative dynamics drives the gauge field to the deconfinement phase where the spin-charge separated $U(1)$ spin liquid is stable with respect to the gauge fluctuations. The summary and discussions are presented in Section IV.

II. MODEL AND LARGE-z THEORY

We start with the half-filled Hubbard model given by

$$H = -\frac{t}{\sqrt{z}} \sum_{\langle ij \rangle, \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \mathrm{h.c.} + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (1)$$

where the t-term describes electron hopping on a lattice with z nearest neighbor (NN) bonds, and the U-term is the on-site Coulomb repulsion. When the quantum states are spatially extended, the NN single-particle correlator scales with the coordination number as $\langle c_{i\sigma}^\dagger c_{j\sigma} \rangle \sim 1/\sqrt{z}$. As a result, the $1/\sqrt{z}$-scaling for the hopping t in Eq. (1) is necessary in order to maintain a finite kinetic energy in the large-z limit[35]. This rescaling is used in the DMFT[34].

To construct a strong-coupling theory that is nonperturbative in U, Kotliar and Ruckenstein[28] introduced a spin-1/2 fermion f_σ and four slave bosons e (holon), d (doublon), and p_σ to represent the local Hilbert space for the empty, doubly-occupied, and singly occupied sites respectively: $|0\rangle = e^\dagger |\mathrm{vac}\rangle$, $|\uparrow\downarrow\rangle = d^\dagger f_\downarrow^\dagger f_\uparrow^\dagger |\mathrm{vac}\rangle$, and $|\sigma\rangle = p_\sigma^\dagger f_\sigma^\dagger |\mathrm{vac}\rangle$. The physical Hilbert space is obtained under the local constraints for the completeness

$$e_i^\dagger e_i + \sum_\sigma p_{i\sigma}^\dagger p_{i\sigma} + d_i^\dagger d_i = 1, \quad (2)$$

and the consistency

$$f_{i\sigma}^\dagger f_{i\sigma} = p_{i\sigma}^\dagger p_{i\sigma} + d_i^\dagger d_i. \quad (3)$$

The Hubbard model is thus faithfully represented by

$$H = -\frac{t}{\sqrt{z}} \sum_{\langle ij \rangle, \sigma} Z_{i\sigma}^\dagger Z_{j\sigma} f_{i\sigma}^\dagger f_{j\sigma} + \mathrm{h.c.} + U \sum_i d_i^\dagger d_i, \quad (4)$$

where the composite bosonic operator

$$Z_{i\sigma} = L_{i\sigma}^{-1/2} (p_{i\bar\sigma}^\dagger d_i + e_i^\dagger p_{i\sigma}) R_{i\bar\sigma}^{-1/2}, \quad (5)$$

is introduced for the bookkeeping purpose, so as to make the following derivations more transparent. The operators $L_{i\sigma} = 1 - d_i^\dagger d_i - p_{i\sigma}^\dagger p_{i\sigma}$ and $R_{i\bar\sigma} = 1 - e_i^\dagger e_i - p_{i\bar\sigma}^\dagger p_{i\bar\sigma}$ should be understood as projection operators for hard-core bosons with unit eigenvalues, and the choice of the $-1/2$ power in Eq. (5) reproduces the Gutzwiller approximation at the level of the static saddle point[28].

Unlike fermions that subject to Pauli exclusion principle, the bosons have a remarkable property: a macroscopically large number of them can condense into a single quantum state. Thus the $Z_{i\sigma}$-boson can be decomposed into a single-particle condensate part $Z_{i\sigma,0}$ and

an uncondensed "normal" or fluctuating part $\tilde{Z}_{i\sigma}$, i.e., $Z_{i\sigma}^{(\dagger)} = Z_{i\sigma,0} + \tilde{Z}_{i\sigma}^{(\dagger)}$, and similarly for the slave bosons. Consequently, the single-particle correlator of Z-bosons has two contributions

$$\langle Z_{i\sigma}^{\dagger} Z_{j\sigma} \rangle = Z_{i\sigma,0} Z_{j\sigma,0} + \langle \tilde{Z}_{i\sigma}^{\dagger} \tilde{Z}_{j\sigma} \rangle, \tag{6}$$

where the first term comes from the single-boson condensate and does not scale with the coordination number z or the distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between site i and j. The second term $\langle \tilde{Z}_{i\sigma}^{\dagger} \tilde{Z}_{j\sigma} \rangle$ comes from the uncondensed fluctuating bosons and decreases with increasing z and r_{ij} according to $z^{-r_{ij}/2}$. More explicitly, on the NN bonds, $Z_{i\sigma,0} Z_{j\sigma,0} \sim 1$, whereas $\langle \tilde{Z}_{i\sigma}^{\dagger} \tilde{Z}_{j\sigma} \rangle \sim 1/\sqrt{z}$. As a result, the kinetic hopping energy of uncondensed bosons is of higher order by $1/\sqrt{z}$ than that of the condensed bosons, and thus negligible in the large-z limit and the Hamiltonian in Eq. (4) reduces to the static saddle point solution characterized by the single-particle condensation of all the slave bosons[28]. Thus, taking the large-z limit this way results in the static SBMF theory which is equivalent to the Gutzwiller approximation. The BR metal-insulator transition takes place when the condensate density of Z-boson is driven zero by the vanishing of the D/H density for $U > U_{BR}$, which is equivalent to having a divergent effective mass for the QPs.

In order to include the effects of the fluctuating Z-bosons, it is necessary to treat the contributions from the condensed and uncondensed bosons on equal footing. As pointed out explicitly in the formulation of the *bosonic* DMFT[36], this can be achieved by different rescalings of the bosonic hopping amplitudes for the condensate and the fluctuating parts. Utilizing this rescaling for the bosons, the slave-boson formulation of the Hubbard model in Eq. (4) is rewritten as

$$H = -\frac{t}{\sqrt{z}} \sum_{\langle ij \rangle, \sigma} \left(Z_{i\sigma,0} Z_{j\sigma,0} + \sqrt{z} \tilde{Z}_{i\sigma}^{\dagger} \tilde{Z}_{j\sigma} \right) f_{i\sigma}^{\dagger} f_{j\sigma}$$
$$+ \text{h.c.} + U \sum_i d_i^{\dagger} d_i. \tag{7}$$

Since the NN correlator for the fluctuating bosons $\langle \tilde{Z}_{i\sigma}^{\dagger} \tilde{Z}_{j\sigma} \rangle \sim 1/\sqrt{z}$, it contributes to the hopping integral on equal footing as the condensed part. On account of the fermion correlator $\langle f_{i\sigma}^{\dagger} f_{j\sigma} \rangle \sim 1/\sqrt{z}$, this lead to a finite kinetic energy coming from the uncondensed and fluctuating bosons beyond the Gutzwiller approximation or the static SBMF theory in the large-z limit.

The rescaled Hamiltonian Eq. (7) has a remarkable property that it simplifies considerably on the Mott insulator side where the D/H single-particle condensate vanishes, i.e. $d_{i0} = e_{i0} = 0$, which implies $Z_{i0} = 0$ by Eq. (5). As a result, the kinetic energy solely comes from the uncondensed bosons accompanied by the back-flow of the fermions between the neighboring sites, which is a signature of electron fractionalization in the Mott insulating state. To see that the electrons must be incoherent, it is instructive to note that since the correlators

of the f-fermion and the Z-bosons both scale as $1/\sqrt{z}$, the electron intersite correlator $\langle c_{i\sigma}^{\dagger} c_{j\sigma} \rangle \sim 1/z$, which is completely different from that of the coherent QP hopping behavior on the metallic side, yet contributes to a finite kinetic energy in this large z-limit.

We thus study the Mott transition from the Mott insulating side at large $U > U_c$. The absence of D/H condensate leads to $\tilde{Z}_{i\sigma} = L_{i\sigma}^{-1/2}(p_{i\bar{\sigma},0} d_i + e_i^{\dagger} p_{i\sigma,0}) R_{i\bar{\sigma}}^{-1/2}$, where d_i and e_i are the fluctuating D/H having a nonzero density $n_d = n_e = \langle d_i^{\dagger} d_i \rangle = \langle e_i^{\dagger} e_i \rangle \neq 0$. The $p_{i\sigma}$ bosons representing single-particle occupation condense into c-numbers with $p_{i\sigma}^{(\dagger)} = p_0$ in the absence of magnetism. Furthermore, the operators $L_{i\sigma}$ and $R_{i\sigma}$ contained in $\tilde{Z}_{i\sigma}$ should not introduce additional intersite correlations to leading order in $1/\sqrt{z}$, in contrast to uncontrolled saddle point approximations with D-H binding in two dimensions[27]. They can thus be written in terms of the *local* densities as $L_{i\sigma} = 1 - n_i^d - n_{i\sigma}^p$ and $R_{i\sigma} = 1 - n_i^e - n_{i\bar{\sigma}}^p$, with $n_{i\sigma}^p = \langle p_{i\sigma}^{\dagger} p_{i\sigma} \rangle = p_0^2$. The local constraints Eqs (2-3) thus ensure $L_{i\sigma} = R_{i\sigma} = 1/2$ in paramagnetic phases at half-filling. Hence, on the Mott insulating side,

$$\tilde{Z}_{i\sigma} = 2p_0(d_i + e_i^{\dagger}), \tag{8}$$

to leading order in $1/\sqrt{z}$ and the Hamiltonian in Eq. (7) becomes

$$H = -4p_0^2 t \sum_{\langle ij \rangle} (d_i^{\dagger} d_j + e_j^{\dagger} e_i + e_i d_j + d_i^{\dagger} e_j^{\dagger}) f_{i\sigma}^{\dagger} f_{j\sigma}$$
$$+ \text{h.c.} + U \sum_i d_i^{\dagger} d_i. \tag{9}$$

It is straightforward to write down the path integral of the model, with the detailed derivation provided in the Appendix. The condensation of the p_{σ} bosons collapses two of the operator constraints in Eqs (2-3) into consistency equations for particle densities $n_d + p_0^2 = n_{\sigma}^f$. The remaining one can be written as

$$e_i^{\dagger} e_i - d_i^{\dagger} d_i + \sum_{\sigma} f_{i\sigma}^{\dagger} f_{i\sigma} = 1, \tag{10}$$

which corresponds to the unbroken internal $U(1)$ gauge symmetry and specifies the gauge charges of the particles. Eq. (10) shows that increasing the spinon number by one must be accompanied by either destroying a holon or creating a doublon at the same site. The partition function can be written down as an imaginary-time path integral

$$Z = \int \mathcal{D}[f^{\dagger}, f] \mathcal{D}[d^{\dagger}, d] \mathcal{D}[e^{\dagger}, e] \mathcal{D}[a_0, a] \mathcal{D}\lambda e^{-\int_0^{\beta} \mathcal{L} d\tau}, \tag{11}$$

with the Lagrangian

$$\mathcal{L} = \sum_i \left[d_i^{\dagger}(\partial_{\tau} - ia_0) d_i + e_i^{\dagger}(\partial_{\tau} + ia_0) e_i \right.$$
$$\left. + f_{i\sigma}^{\dagger}(\partial_{\tau} + ia_0) f_{i\sigma} \right] - H_f - H_b$$
$$+ \sum_i \left[i\lambda_i(d_i^{\dagger} d_i + e_i^{\dagger} e_i + 2p_0^2 - 1) - U/2 \right], \tag{12}$$

where λ_i is a Lagrangian multiplier. The decoupled fermion and boson Hamiltonian[19] are given by

$$H_f = -\frac{t_f}{\sqrt{z}} \sum_{\langle i,j \rangle} e^{ia_{ij}} f_{i\sigma}^\dagger f_{j\sigma} + \text{h.c.} + \frac{U}{2} \sum_{i\sigma} f_{i\sigma}^\dagger f_{i\sigma}, \quad (13)$$

$$H_b = -\frac{t_b}{\sqrt{z}} \sum_{\langle i,j \rangle} e^{-ia_{ij}} (e_j^\dagger e_i + d_i^\dagger d_j + e_i d_j + d_i^\dagger e_j^\dagger)$$
$$+ \text{h.c.} + \frac{U}{2} \sum_i (d_i^\dagger d_i + e_i^\dagger e_i), \quad (14)$$

with

$$t_f = 8t p_0^2 \sqrt{z} (\chi_d + \Delta_d), \quad t_b = 8t p_0^2 \sqrt{z} \chi_f. \quad (15)$$

In a stationary state, $\chi_d = \langle d_i^\dagger d_j \rangle = \langle e_j^\dagger e_i \rangle$ is the quantum average of the D/H nearest neighbor hopping, $\chi_f = \langle f_{i\sigma}^\dagger f_{j\sigma} \rangle$ the spinon hopping per spin, and $\Delta_d = \langle d_i^\dagger e_j^\dagger \rangle = \langle e_i d_j \rangle$ is the D-H binding order parameter. In Eqs (12-14), the spinons and the D/H are coupled by the emergent $U(1)$ gauge fields a_0 and a_{ij} associated with the constraint in Eq. (10). Physically, the instantons of this compact gauge field correspond to the tunneling events where the spinons and D/H tunnel in and out of the lattice sites[37].

III.   MOTT TRANSITION AND SPIN LIQUID MOTT INSULATOR

A.   Asymptotic solution for Mott transition

We will first obtain the stationary state solution with $a_0 = a_{ij} = 0$, and then study the properties of the gauge field fluctuations. Eq. (13) shows that the spinon hopping amplitude is $t_f/\sqrt{z}$ where t_f defined in Eq. (15) is proportional to the D/H intersite correlations. We will show that the latter leads to a renormalized narrow spinon band with a bandwidth on the order of the exchange coupling $J \sim t^2/U$ in the large U limit. The spinon kinetic energy per site is $K_f = (t_f/t)K_0$ where $K_0 = 2 \int_0^D \rho_0(\omega)\omega d\omega$ is that for noninteracting electrons with hopping $t/\sqrt{z}$ and ρ_0 is the corresponding semicircle density of states $\rho_0(\omega) = \frac{2}{\pi D}\sqrt{1 - (\omega/D)^2}$ on the infinite-z Bethe lattice[38] with a half-bandwidth $D = 2t$. Note that both D and t are order one quantities, since the $1/\sqrt{z}$ factors in Eqs (13) and (14) are dynamically generated by the NN intersite correlators $(\chi_d, \Delta_d, \chi_f) \sim 1/\sqrt{z}$ in Eq. (15) where t_b and t_f are of order one. Thus, $K_0 = 8t/3\pi = 4D/3\pi$ and $K_f = 8t_f/3\pi$. Since K_f can also be written as $K_f = 4t_f\sqrt{z}\chi_f$, we obtain readily $\chi_f = \frac{1}{\sqrt{z}}\frac{2}{3\pi}$, independent of U. The boson hopping parameter t_b in Eq. (15) is thus given by $t_b = 16p_0^2 t/3\pi$, which is on the order of t. Hence, the spectrum of charge excitations residing in the D/H sector has a bandwidth on the order of the bare electron bandwidth, giving rise to
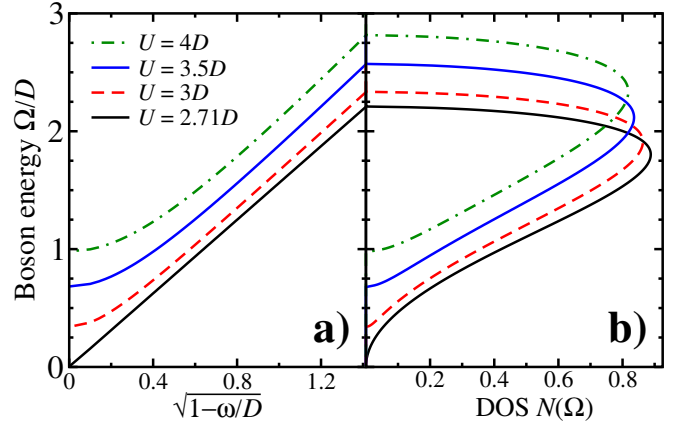

FIG. 1: (color online) The doublon/holon energy spectrum (a) and the corresponding spectral density of states (b) for different U.

the broad incoherent spectral weight induced by strong correlation.

From Eqs (12) and (14), the stationary state bosonic Hamiltonian in the D/H sector is

$$H_{\text{D/H}} = \int_{-D}^{D} d\omega \rho_0(\omega) \begin{bmatrix} d_\omega^\dagger, e_\omega \end{bmatrix} \begin{bmatrix} \varepsilon_\omega & -\Delta_\omega \\ -\Delta_\omega & \varepsilon_\omega \end{bmatrix} \begin{bmatrix} d_\omega \\ e_\omega^\dagger \end{bmatrix}, \quad (16)$$

where $\varepsilon_\omega = \frac{U}{2} + \lambda - \frac{t_b}{t}\omega$, $\Delta_\omega = \frac{t_b}{t}\omega$ are the D/H kinetic and pairing energies; $\lambda = \langle i\lambda \rangle$. Diagonalizing $H_{\text{D/H}}$ by Bogoliubov transformation produces two degenerate branches for the D/H excitations,

$$\Omega_\omega = \sqrt{\varepsilon_\omega^2 - \Delta_\omega^2}. \quad (17)$$

The Mott insulator is thus an excitonic insulator and the Mott gap is given by the charge gap in Ω_ω,

$$G_{\text{Mott}}(U) = 2\Omega_D = 2\sqrt{\left(\frac{U}{2} + \lambda\right)\left(\frac{U}{2} + \lambda - 4t_b\right)}. \quad (18)$$

The physical condition for a real Ω requires $U \geq 8t_b - 2\lambda$ and the equal sign determines the critical U_c for the Mott transition where $G_{\text{Mott}}(U_c) = 0$.

Minimizing the energy leads to the self-consistent equations, $p_0^2 = \frac{1}{2} - n_d$, $\lambda = 4K_0\sqrt{z}(\chi_d + \Delta_d)$, and

$$n_d = \frac{1}{2}\int_{-D}^{D}\left(\frac{\varepsilon_\omega}{\Omega_\omega} - 1\right)\rho_0(\omega)d\omega, \quad (19)$$

$$\sqrt{z}\chi_d = \frac{1}{2D}\int_{-D}^{D}\frac{\varepsilon_\omega}{\Omega_\omega}\omega\rho_0(\omega)d\omega, \quad (20)$$

$$\sqrt{z}\Delta_d = \frac{1}{2D}\int_{-D}^{D}\frac{\Delta_\omega}{\Omega_\omega}\omega\rho_0(\omega)d\omega. \quad (21)$$

Eq. (19) shows that the nonzero D/H density is entirely due to the quantum fluctuations above the Mott gap in Ω_ω for $U > U_c$. Lowering U toward U_c, G_{Mott} must

reduce to host the increased D/H density until $G_{\mathrm{Mott}} = 0$ at $U = U_c$ where the D/H condensation emerges and, as we shall shown, the continuous Mott transition takes place.

Solving these equations self-consistently, we obtain the properties of the Mott insulator and the Mott transition. The D/H excitation spectrum is plotted in Fig. 1a, showing the closing of the Mott gap as U is reduced toward U_c. Note that the calculated spectral density of states $N^{\mathrm{D/H}}(\Omega)$ shown in Fig. 1b vanishes quadratically upon gap closing, which ensures that the Mott transition is continuous at zero temperature.

1. Critical properties at Mott transition

Remarkably, the critical properties of the transition can be determined analytically. First, setting the Mott gap $G_{\mathrm{Mott}}(U_c) = 0$ gives

$$U_c = 8t_b^c - 2\lambda^c, \tag{22}$$

where the script c denotes the critical values of the corresponding quantity at the transition point. Next, using the expressions for t_b in Eq. (15) and λ given above Eq. (19), we obtain

$$U_c = U_{\mathrm{BR}}[1 - 2n_d^c - \sqrt{z}(\chi_d^c + \Delta_d^c)], \tag{23}$$

where $U_{\mathrm{BR}} = 8K_0 = 32D/3\pi$ is the critical value for the BR transition on the Bethe lattice and $(n_d^c, \chi_d^c, \Delta_d^c)$ are the critical values of the doublon density, doublon hopping, and the D-H binding, respectively. Eq. (23) reveals the much desirable connection between the Mott transition and the BR transition. It shows that the Mott transition can be viewed as the quantum correction to the BR transition due to D-H binding. Since $U_c < U_{\mathrm{BR}}$, the BR transition is preempted by the Mott transition and unobservable in the Hubbard model.

At $U = U_c$, it is straightforward to calculate the D/H kinetic and pairing energies in Eq. (16) to obtain $(\varepsilon_\omega^c, \Delta_\omega^c) = (1 - 2n_d^c)\frac{8D}{3\pi}(2 - \frac{\omega}{D}, \frac{\omega}{D})$, such that the critical D/H excitation spectrum in Eq. (17) becomes

$$\Omega_\omega^c = (1 - 2n_d^c)\frac{16D}{3\pi}\sqrt{1 - \frac{\omega}{D}}. \tag{24}$$

The spectrum is independent of χ_d and Δ_d and agrees with the one shown in Fig. 1a at $U = 2.71D$. Furthermore, the ratios $\varepsilon_\omega^c/\Omega_\omega^c$ and $\Delta_\omega^c/\Omega_\omega^c$ that enter Eqs (19-21) are simple universal functions such that these integrals can be evaluated analytically to obtain the critical quantities at the Mott transition,

$$n_d^c = \frac{12\sqrt{2} - 5\pi}{10\pi} \simeq 0.040 \tag{25}$$

$$\sqrt{z}\chi_d^c = \frac{2}{35\pi}\sqrt{2} \simeq 0.026 \tag{26}$$

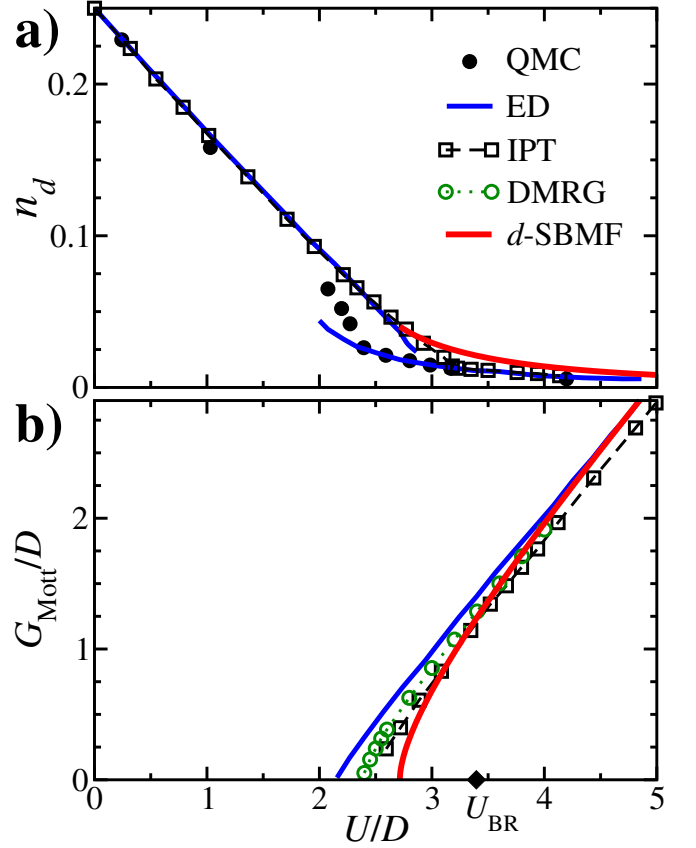$$\sqrt{z}\Delta_d^c = \frac{22}{105\pi}\sqrt{2} \simeq 0.094. \tag{27}$$

FIG. 2: (color online) Mott insulator in the large-z limit of the dynamical SBMF theory (d-SBMF). (a) The doublon density as a function of U. (b) The Mott gap in the charge sector as a function of U. The DMFT results obtained using different impurity solvers are also shown for comparison (data from Ref.38 and 39): quantum monte carlo (QMC), exact diagonalization (ED), iterative perturbation theory (IPT), and dynamical density matrix renormalization group (DMRG).

Inserting these values into Eq. (23), we obtain the critical Hubbard interaction for the Mott transition,

$$U_c \simeq 0.80 \cdot U_{\mathrm{BR}} \simeq 2.71D, \tag{28}$$

at which the charge gap closes and the QP coherence emerges with the D/H condensate simultaneously. It is interesting to note that U_c is close in value to the one obtained in an improved Gutzwiller projected wavefunction approach[41].

2. Doublon density and Mott gap for $U > U_c$

In Figs 2a and 2b, the calculated doublon density and Mott gap are plotted in red solid lines as a function of U/D on the insulating side of the Mott transition. Various single-site DMFT results[38–40] are also plotted in Fig. 2 for comparison solely for the purpose of benchmarking the results in the charge sector, despite the different large-z limit and the continuous Mott

transition to a spin liquid at a single U_c. These zero temperature ground state properties are known to be difficult to obtain reliably in the DMFT and near the Mott transition, as reflected in the discrepancies between the results obtained using different quantum impurity solvers[38]. Fig. 2a shows that the doublon density decays algebraically with increasing U. Indeed, Eqs (19-21) can be solved analytically to obtain the large U behaviors

$$n_d = \left(\frac{8}{3\pi}\right)^2 \frac{t^2}{U^2} + \mathcal{O}\left(\frac{t}{U}\right)^4, \qquad (29)$$

$$\sqrt{z}\chi_d = 4\left(\frac{8}{3\pi}\right)^3 \frac{t^3}{U^3} + \mathcal{O}\left(\frac{t}{U}\right)^5, \qquad (30)$$

$$\sqrt{z}\Delta_d = \frac{4}{3\pi}\frac{t}{U} + \mathcal{O}\left(\frac{t}{U}\right)^3. \qquad (31)$$

The kinetic and potential energies deep in the Mott insulator can be readily obtained from the above equations, with $E_{\mathrm{kin}} = -2(\frac{8}{3\pi})^2\frac{t^2}{U} + \mathcal{O}(\frac{t^4}{U^3})$ and $E_{\mathrm{pot}} = (\frac{8}{3\pi})^2\frac{t^2}{U} + \mathcal{O}(\frac{t^4}{U^3})$. Therefore, the energy of the Mott insulator $E = -(\frac{8}{3\pi})^2\frac{t^2}{U} + \mathcal{O}(\frac{t^4}{U^3})$, for $U > U_c$. In Fig. 3, the evolution of $(n_d, \sqrt{z}\chi_d, \sqrt{z}\Delta_d)$ as a function of D/U is shown on a log-log plot. It can be seen that the general self-consistent solutions of Eqs (19-21) represented by the solid lines merges with the corresponding dashed-lines describing the asymptotic large-U behaviors given in Eqs (29-31). Thus, the holons and doublons are always present at any U. The binding of the opposite charges on the energy scale of the Mott gap $G_{\mathrm{Mott}}(U)$ makes it possible to treat them as localized quantum defects in the Heisenberg model description of the physics on the energy scale of the exchange coupling J, provided that U is large enough such that $G_{\mathrm{Mott}}(U) \gg J$. Note that $\sqrt{z}\Delta_d \gg n_d \gg \sqrt{z}\chi_d$ in the large-U regime. As a consequence, the large-U physics of the spin-liquid Mott insulator is controlled by D-H binding. We will come back to the physical significance of the latter shortly.

As U is reduced towards U_c, the calculated doublon density n_d in Fig. 2a approaches and merges at U_c smoothly with those obtained for $U < U_c$ by the DMFT using the *zero-temperature* iterative perturbation theory (IPT) and exact diagonalization (ED) impurity solvers. This is reassuring since the large-z limit used in the DMFT is both natural and appropriate for the Hubbard model on the metallic side of the Mott transition. The critical behavior of the Mott gap near U_c can also be obtained analytically from Eq. (18), $G_{\mathrm{Mott}}(U) = \alpha\sqrt{U - U_c}$, $\alpha = 2\sqrt{2t_b^c} \simeq 2.61\sqrt{t}$, where the square-root singularity is clearly seen in Fig. 2b. The Mott gap increases with U and approaches that obtained in the DMFT using ED impurity solver and exhibit the asymptotic behavior in the large-U limit $G_{\mathrm{Mott}}(U \gg D) = U$ seen from Eq. (18).
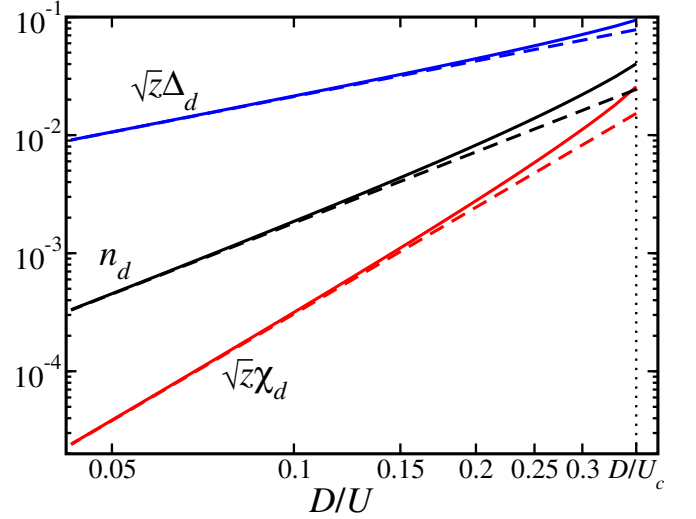

FIG. 3: (color online) Evolution of the doublon density n_d, D/H hopping $\sqrt{z}\chi_d$, and D-H binding $\sqrt{z}\Delta_d$ as a function of D/U on a log-log plot in the Mott insulating state. Solid lines: fully self-consistent solutions of Eqs (19 -21). Dashed lines: asymptotic solutions in the large-U limit given in Eqs (29-31). The vertical dotted line indicates the critical $D/U_c \simeq 0.37$.

3. Spectroscopy of spin liquid Mott insulator

Figs 4a and 4b show the spectroscopic properties on the Mott insulating side with comparison to the corresponding DMFT results. They are obtained by calculating the local electron Green's function

$$G_\sigma(\tau) = -\langle T_\tau c_{i\sigma}(\tau)c_{i\sigma}^\dagger(0)\rangle = G_\sigma^f(\tau)G_Z(\tau), \qquad (32)$$

where G_σ^f and G_Z are the corresponding local Green's functions of the spinon and the Z-boson (linear combinations of the D/H). In Matsubara frequency space, Eq. (32) amounts to a convolution

$$G_\sigma(i\omega_n) = \sum_{i\nu_n} G_\sigma^f(i\omega_n - i\nu_n)G_Z(i\nu_n) \qquad (33)$$

of the spinon and the D/H local Green's functions[27]

$$G_\sigma^f(i\omega_n) = \int d\epsilon\rho_0(\epsilon)G_\sigma^f(\epsilon, i\omega_n), \qquad (34)$$

$$G_Z(i\nu_n) = \int d\epsilon\rho_0(\epsilon)G_Z(\epsilon, i\nu_n). \qquad (35)$$

The electron spectral density is given by $N_\sigma(\omega) = -\frac{1}{\pi}\mathrm{Im}G_\sigma(i\omega_n \to \omega + i0^+)$. Fig. 4a shows $N_\sigma(\omega)$ obtained at $U = 4D$, exhibiting the upper and the lower Hubbard bands separated by the Mott gap, in broad semi-quantitative agreement with the DMFT results obtained by IPT and the more recent dynamical density matrix renormalization group (DMRG) impurity solvers[38,39]. The spectral density of the spinons $N_\sigma^f(\omega)$ also shown in Fig. 4a is, on the other hand, gapless and contributes
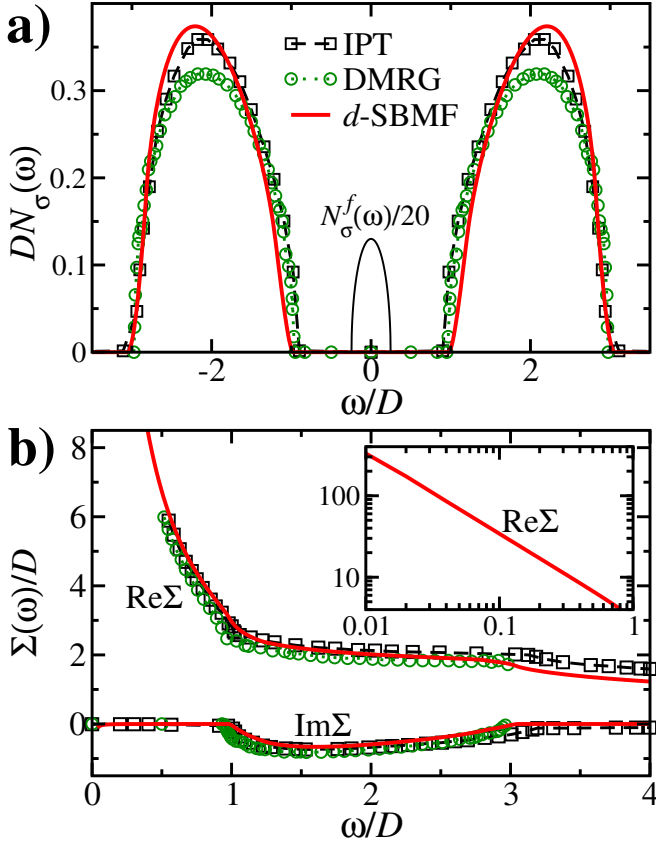
FIG. 4: (color online) Spectroscopy of the Mott insulator in the large-z limit of the dynamical SBMF theory (red lines) at $U = 4D$. (a) The spectral density of states. Thin solid line: spinon density of states. (b) The real and imaginary parts of the electron self-energy. Inset: Real part of self energy on log-log plot, showing the $1/\omega$ dependence. The DMFT results obtained at $U = 4D$ using the zero temperature IPT (open squares)[38] and the dynamical DMRG (open circles)[39] impurity solvers are also shown for comparison.

to the thermodynamic properties of the spin liquid at low temperatures. The spinon half bandwidth is $D_f = 2t_f$ in the large-z limit, where t_f is the spinon hopping integral given in Eq. (15). Let's consider the physics when U is large. In this case, t_f can be readily evaluated using the solutions for $(n_d, \sqrt{z}\chi_d, \sqrt{z}\Delta_d)$ obtained in Eq. (29-31). Note that since $\sqrt{z}\Delta_d \propto t/U \gg \sqrt{z}\chi_d$ in the large-U limit, it dominates the contributions to t_f and leads to $t_f = \frac{4}{3\pi}\frac{4t^2}{U}$. Thus, the spinon hopping amplitude and bandwidth are controlled by the exchange coupling J, capturing the physics of the gapless $U(1)$ spin liquid phase in the effective Heisenberg model. Moreover, the analysis shows that the origin of the exchange coupling J on the insulating side of the Mott transition is intimately connected to the physics of D-H binding in the Hubbard model. We note in passing that these properties of the Mott transition and Mott insulator are inaccessible to Gaussian fluctuations around the Kotliar-Ruckenstein saddle point governing the putative BR transition at

large U[42,43].

The central quantity in the large-z limit is the electron local self-energy $\Sigma(\omega)$. It can be extracted by casting the local *electron* Green's function in Eq. (33) into the standard form in terms of the self-energy,

$$G_\sigma(\omega) = \int_{-D}^{D} d\epsilon\, \rho_0(\epsilon) \frac{1}{\omega - \epsilon - \Sigma(\omega)}. \tag{36}$$

The calculated real part ($\mathrm{Re}\Sigma$) and imaginary part ($\mathrm{Im}\Sigma$) of the *electron* self-energy in the current dynamical SBMF theory are plotted in Fig. 4b as a function of ω at $U = 4D$. For comparison, the DMFT results obtained using the zero temperature IPT and the dynamical DMRG impurity solvers are also shown at the same value of $U = 4D$[38,39]. Remarkably close agreement can be seen between both the real and the imaginary part of the self-energies. Moreover, inside the Mott gap, the real part of the self-energy shows the scaling behavior $\mathrm{Re}\Sigma(\omega) \propto 1/\omega$ as shown on the log-log scale in the inset of Fig. 4b in agreement with the DMFT[38].

We note that, while the theory in the large-z limit amounts to two static mean-field theories for the spin-1/2 fermion $f_{i\sigma}$ and the slave bosons, it captures the incoherent excitations of the physical electrons due to the dynamical fluctuations, since the Green's function of the physical electron $c_{i\sigma}$ involves the convolution of those of the $f_{i\sigma}$-fermions and the slave bosons, as shown in Eqs. (33-35), resulting in a dynamical self-energy in Eq. (36). For these reasons, the theory is termed as the dynamical SBMF theory in this paper.

B. Gauge field dynamics and deconfinement

The emergence of the spin-liquid Mott insulator with gapless spinon excitations requires spin-charge separation and is stable only if the gauge field that couples them in Eqs (12-14) is deconfining. To derive the gauge field action, we integrate out the matter fields using the hopping expansion[44]. To leading order in $1/z$, the low energy effective gauge field action is given by[45],

$$\begin{aligned}
S_{\mathrm{eff}} &= -\frac{\eta}{z\pi^2} \sum_{\langle i,j \rangle} \int_0^\beta \mathrm{d}\tau_1 \int_0^\beta \mathrm{d}\tau_2 \frac{\cos\left(a_{ij} - a_{ij}(\tau_2)\right)}{(\tau_1 - \tau_2)^2} \\
&\quad + \frac{1}{zC} \sum_{\langle i,j \rangle} \int_0^\beta \mathrm{d}\tau (\partial_\tau a_{ij})^2,
\end{aligned} \tag{37}$$

where the second term comes from integrating out the gapped D/H and corresponds to charging with C the "charging energy" of a link. In the large-U limit, $C \propto U^3/t^2$. In the language of the $U(1)$ gauge theory, it describes the electric field action and causes the confinement of the gauge charges. The first term, which is nonlocal in imaginary time and corresponds to dissipation, comes from the contribution from the gapless fermion spinons. In our case, $\eta = 1$, but we will keep it as a

parameter for the dissipation strength in the following discussion. This term is periodic in the gauge field, reflecting its compact nature.

Thus the gauge field action is dissipative. It has been argued under various settings that a large enough dissipation η can drive the compact $U(1)$ gauge field to the deconfinement phase at zero temperature[46–48]. In the large-z limit, Eq. (37) shows that spatial fluctuations of the link gauge field are suppressed and the dissipative gauge field theory becomes local, i.e. $a_{ij}(\tau) = a(\tau)$. As a result, the action becomes identical to the dissipative tunneling action derived by Ambegaokar, Eckern, and Schön[49] for a quantum dot coupled to metallic leads, or a shunted Josephson junction with QP tunneling[50]. The 2π-periodicity of the compact gauge field requires $a(\tau) = \tilde{a}(\tau) + 2\pi n\tau/\beta$ where $\tilde{a}(\tau)$ is single-valued and satisfies $\tilde{a}(0) = \tilde{a}(\beta)$, and n is an integer winding number associated with charge quantization, i.e. the instantons in the electric field when charges tunnel in and out of the link. If the temporal fluctuation of the winding number n is strong, the periodicity of the $a(\tau)$ is important and the gauge field is in the confinement phase. Otherwise, its compactness is irrelevant and the gauge field is in the deconfinement phase. For a 2D array of dissipative tunnel junctions, it has been shown that there exists a confinement-deconfinement (C-DC) transition of the winding number at a critical $\eta_c^{2D} \simeq 0.45$[51]. Using the Villain transformation[52], one can show that the instanton action is described by a dissipative sine-Gordon model, exhibiting a C-DC transition at a critical dissipation $\eta_c = 1/4$. In our case, $\eta > \eta_c$, and the temporal proliferation of the instantons is suppressed by dissipation[45]. Thus, the gauge electric field is deconfining and the gapless $U(1)$ spin-liquid in the Mott insulating phase is indeed stable with respect to gauge fluctuations. When time-reversal symmetry breaking is allowed, however, the Mott insulating state is expected to develop antiferromagnetic long-range order on bipartite lattices.

IV. SUMMARIES AND DISCUSSIONS

In summary , we have provided an asymptotic solution of the Hubbard model in the large-z limit to capture the most essential Mott physics, i.e., the excitonic binding between oppositely charged doublons and holons[23–27,30]. In the Mott insulator, where the theory simplifies considerably as all doublons and holons are bound in real space into excitonic pairs, the motion of the QP must involve breaking the D/H pairs and thus amounts entirely to incoherent excitations above the charge gap set by the D-H binding energy. We construct a dynamical SBMF theory in the large-z limit and find a continuous Mott transition, where the opening of the Mott gap and the vanishing of the QP coherence coincide at the same U_c. The BR transition is preempted by quantum fluctuations and replaced by the Mott transition. A key feature of our asymptotic solution is that on the insulating side of the
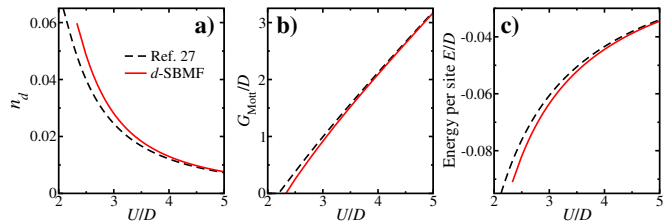

FIG. 5: (color online) Mott transition on two-dimensional square lattice obtained in Ref. 27 and the present dynamical SBMF theory. (a) Doublon density, (b) charge gap, and (c) state energy per site as a function of Hubbard U.

Mott transition, quantum spin fluctuations via the intersite spinon corrlation remain and survive the large-z limit. The coherent hopping of the spinons gives rise to a gapless QSL by lifting the ground state degeneracy. The obtained results is in quantitatively agreement with the DMFT with various numerical quantum impurity solvers. The derived effective action for the compact gauge field in the large-z limit show that the emergent dissipative dynamics drives the gauge field to the deconfinement phase where the spin-charge separated $U(1)$ spin liquid is stable with respect to the gauge fluctuations.

It is instructive to apply the dynamical SBMF theory reported in this paper to the two-dimensional square lattice and compare to the results obtained in the uncontrolled method in Ref. 27. The comparison is shown in Fig. 5, where the doublon density, the Mott gap, and the state energy per site are plotted as a function of U. Although these two methods are consistent deep in the Mott insulating phase for $U \gg U_c$, there are significant differences near the Mott transition. The critical Hubbard interaction U_c is increased from $\sim 2.2D$ in Ref. 27 to $\sim 2.33D$ in the dynamical SBMF theory. The doublon density close to the transition is also much larger. Remarkably, the energy of the Mott insulator obtained in the controlled meanfield theory is significantly lower than that obtained in Ref. 27 near the transition, indicating the importance to treat the D-H binding and the quantum fluctuations systematically in a controlled manner. We expect the controlled treatment of the D-H binding to be also important for the study of magnetism.

To end this paper, we compare the electron spectral function obtained in the present theory to that obtained in other scenarios of the Mott transition. Focusing exclusively on the coherent QP, Gutzwiller variational wave function approaches[31] obtained a strongly correlated Fermi liquid[32] that undergoes a BR transition[33] to a localized state with vanishing QP bandwidth and vanishing doublon D/H density (Fig. 6a). The single-site DMFT maps the lattice Hubbard model to a quantum impurity embedded in a self-consistent bath[38,40]. The mapping is exact in the well-defined large-z limit. The obtained $T = 0$ Mott transition shown in Fig. 6b shows that the opening of the Mott gap at U_{c1} and the disappearance of the QP coherence at U_{c2} do not coin-
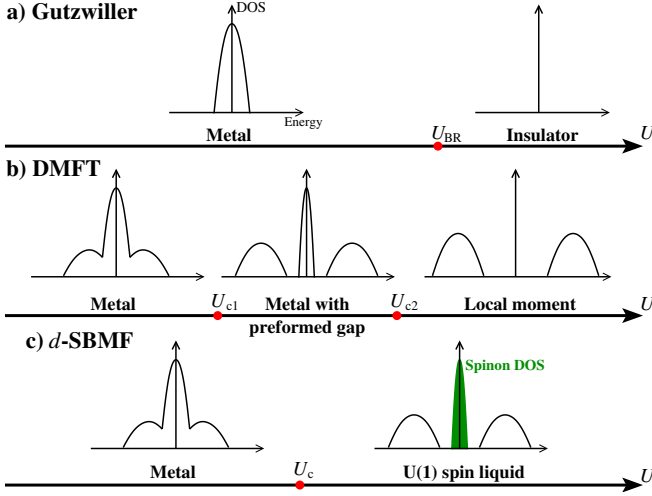
a) Gutzwiller

b) DMFT

c) d-SBMF

FIG. 6: (color online) Schematic diagram of the Mott transition. (a) Gutzwiller with $U_{\rm BR} \simeq 3.40D$; (b) DMFT with $U_{c1} \simeq 2.38D$ and $U_{c2} \simeq 3.04D$[38–40]; and (c) present dynamical SBMF theory with $U_c \simeq 0.8U_{\rm BR} \simeq 2.71D$. D: half bandwidth.

cide such that the QP states in the metallic state for $U_{c1} < U < U_{c2}$ are separated from the incoherent spectrum by a preformed gap. This peculiar property[53–55] was shown to be correct[56] for the large-z limit taken in the DMFT where the spin-exchange interaction $J \sim t^2/U$ scales with $1/z$ and forces the paramagnetic insulating state to be in a local moment phase with 2^N-fold degeneracy, i.e., a quantum paramagnet. In contrast, the current dynamical SBMF theory finds a continuous Mott transition shown in Fig. 6c from a correlated metal to an insulating QSL, where the opening of the Mott gap and the vanishing of the QP coherence coincide at the same U_c.

Appendix A: Slave-boson path integral in the large-z limit

To leading order in $1/\sqrt{z}$, the Hamiltonian in the paramagnetic Mott insulating phase is given by Eq. (9), under the two local constraints in Eqs (2-3). It is straightforward to write down the Lagrangian as

$$
\begin{aligned}
\mathcal{L} = & \sum_i (d_i^\dagger \partial_\tau d_i + e_i^\dagger \partial_\tau e_i + \sum_\sigma p_{i\sigma}^\dagger \partial_\tau p_{i\sigma} + f_{i\sigma}^\dagger \partial_\tau f_{i\sigma}) \\
& -4tp_0^2 \sum_{\langle i,j \rangle, \sigma} [(d_i^\dagger d_j + e_j^\dagger e_i + d_i^\dagger e_j^\dagger + e_i d_j) f_{i\sigma}^\dagger f_{j\sigma} + \text{h.c.}] \\
& +U \sum_i d_i^\dagger d_i + i \sum_i \tilde{\lambda}_i (e_i^\dagger e_i + d_i^\dagger d_i + 2p_0^2 - 1) \\
& +i \sum_{i\sigma} \lambda'_{i\sigma} (f_{i\sigma}^\dagger f_{i\sigma} - d_i^\dagger d_i - p_0^2).
\end{aligned} \tag{A1}
$$

where Lagrangian multipliers $\tilde{\lambda}_i$ and $\lambda'_{i\sigma}$ are introduced to enforce the local constraints. We may introduce phase fluctuations of p-boson condensates and write $p_{i\sigma} = p_0 e^{i\phi^p_{i\sigma}}$. However, the phase fluctuations can be absorbed by a field redefinition $d_i = d_i e^{-i\phi^p_{i\uparrow} - i\phi^p_{i\downarrow}}$, $f_{i\sigma} = f_{i\sigma} e^{i\phi^p_{i\sigma}}$, and $\lambda'_{i\sigma} = \lambda'_{i\sigma} - \partial_\tau \phi^p_{i\sigma}$. Thus, we can simply take $p_{i\sigma} = p_0$.

The above Lagrangian, Eq. (A1), possesses a $U(1)$ gauge symmetry, i.e., it is invariant under the local transformation

$$
\begin{aligned}
d_i \to d_i e^{i\theta_i}, \quad e_i \to e_i e^{-i\theta_i}, \quad f_{i\sigma} \to f_{i\sigma} e^{-i\theta_i}, \\
\tilde{\lambda}_i \to \tilde{\lambda}_i + \partial_\tau \theta_i, \quad \lambda'_{i\sigma} \to \lambda'_{i\sigma} + \partial_\tau \theta_i.
\end{aligned} \tag{A2}
$$

The gauge charges of the fields are determined by the gauge transformation and it is easily deduced that the doublon field carries the opposite charge as that of the holon and spinon fields. The gauge invariance is a result of the redundancy in the slave-boson representation of the physical degrees of freedom.

We then decouple the Bose-Fermi mixed term (the hopping term) by a Hubbard-Stratonovich transformation. It is important that the gauge invariance must be retained[57], by introducing explicitly the gauge fields a_0 and a_{ij} that describe the temporal and spatial gauge fluctuations, respectively. The resulting Lagrangian is

$$
\begin{aligned}
\mathcal{L} = & \sum_i \big[d_i^\dagger (\partial_\tau - ia_0) d_i + e_i^\dagger (\partial_\tau + ia_0) e_i \\
& + f_{i\sigma}^\dagger (\partial_\tau + ia_0) f_{i\sigma} \big] - \tilde{H}_f - \tilde{H}_b \\
& +i \sum_i \tilde{\lambda}_i (d_i^\dagger d_i + e_i^\dagger e_i + 2p_0^2 - 1),
\end{aligned} \tag{A3}
$$

with decoupled fermion and boson Hamiltonians given

by

$$\tilde{H}_f = -\frac{t_f}{\sqrt{z}} \sum_{\langle i,j \rangle} e^{ia_{ij}} f_{i\sigma}^\dagger f_{j\sigma} + \text{h.c.} + \frac{U}{2} \sum_{i\sigma} f_{i\sigma}^\dagger f_{i\sigma}, \quad (A4)$$

$$\tilde{H}_b = -\frac{t_b}{\sqrt{z}} \sum_{\langle i,j \rangle} e^{-ia_{ij}} (e_j^\dagger e_i + d_i^\dagger d_j + e_i d_j + d_i^\dagger e_j^\dagger)$$

$$+\text{h.c.} - U \sum_i p_0^2, \quad (A5)$$

where the hopping amplitudes t_f and t_b take their saddle point values given in Eq. (15). Here, the Lagrangian multiplier $i\lambda'_{i\sigma}$ takes its saddle-point value of $U/2$ to ensure the symmetry between doublons and holons at half-filling. In order to make the discussions in the main text more transparent, we define $i\tilde{\lambda}_i = i\lambda_i + U/2$. Consequently, the Lagrangian in Eq. (A3-A5) can be rewritten

as Eq. (12-14) in the main text.

As a final remark, note that under the $U(1)$ gauge transformation, the fields transform as

$$d_i \to d_i e^{i\theta_i}, \quad e_i \to e_i e^{-i\theta_i}, \quad f_{i\sigma} \to f_{i\sigma} e^{-i\theta_i},$$

$$a_{ij} \to a_{ij} + \theta_j - \theta_i, \quad a_{0,i} \to a_{0,i} + \partial_\tau \theta_i, \quad (A6)$$

such that the theory is indeed gauge invariant. The gauge invariance has a simple physical interpretation. From Eqs (11) and (12), it is clearly seen that integrating out the scalar gauge field a_0 enforces the local constraint in Eq. (10), just like how a Lagrange multiplier works. Furthermore, integrating over the vector gauge fields enforces a physical constraint on the flow of the currents, i.e. the Ioffe-Larkin rule[37], which states that the fermion current is always accompanied by the counter-flow of the boson current.

[1] N.F. Mott, Proc. Phys. Soc. (London) A**62**, 416 (1949).
[2] N.F. Mott, Rev. Mod. Phys. **40**, 677 (1968).
[3] N. F. Mott, *Metal-Insulator Transitions*, 2nd Ed. Taylor & Francis, London 1990.
[4] P. W. Anderson, Mater. Res. Bull. **8**, 153 (1973).
[5] X.-G. Wen, Phys. Rev. B **44**, 2664 (1991).
[6] P. A. Lee, Science **321**, 1306 (2008).
[7] L. Balents, Nature **464**, 199 (2010).
[8] Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, Phys. Rev. Lett. **91**, 107001 (2003).
[9] Y. Kurosaki, Y. Shimizu, K. Miyagawa, K. Kanoda, and G. Saito, Phys. Rev. Lett. **95**, 177001 (2005).
[10] S. Yamashita, Y. Nakazawa, M. Oguni, Y. Oshima, H. Nojiri, Y. Shimizu, K. Miyagawa, and K. Kanoda, Nat. Phys. **4**, 459 (2008).
[11] M. Yamashita, N. Nakata, Y. Kasahara, T. Sasaki, N. Yoneyama, N. Kobayashi, S. Fujimoto, T. Shibauchi, and Y. Matsuda, Nat. Phys. **5**, 44 (2008).
[12] T.-H. Han, J. S. Helton, S. Chu, D. G. Nocera, J. A. Rodriguez-Rivera, C. Broholm, and Y. S. Lee, Nature **492**, 406 (2012).
[13] Y. Li, G. Chen, W. Tong, L. Pi, J. Liu, Z. Yang, X. Wang, and Q. Zhang, Phys. Rev. Lett. **115**, 167203 (2015).
[14] Y. Shen, Y.-D. Li, H. Wo, Y. Li, S. Shen, B. Pan, Q. Wang, H. C. Walker, P. Steffens, M. Boehm, Y. Hao, D. L. Quintero-Castro, L. W. Harriger, M. D. Frontzek, L. Hao, S. Meng, Q. Zhang, and J. Zhao, Nature **540** 559 (2016).
[15] P. W. Anderson, Science **235**, 1196 (1987).
[16] S. A. Kivelson, D. S. Rokhsar, J. P. Sethna, Phys. Rev. B **35**, 8865 (1987).
[17] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998).
[18] P. Phillips, Ann. of Phys. **321**, 1634 (2006).
[19] P. A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78** 17, (2006).
[20] Z.-Y. Weng, New J. Phys. **13**, 103039 (2011).
[21] W. Kohn, Phys. Rev. **133**, A171 (1964).
[22] G. Kemeny and L. G. Caron, Rev. Mod. Phys. **40**, 790 (1968).
[23] T. A. Kaplan, P. Horsch, and P. Fulde, Phys. Rev. Lett.

[23] **49**, 889 (1982).
[24] H. Yokoyama and H. Shiba, J. phys. Soc. J. **59**, 3669 (1990); H. Yokoyama, M. Ogata, and Y. Tanaka, J. Phys. Soc. Jpn. **75**, 114706 (2006).
[25] M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti, Phys. Rev. Lett. **94**, 026406 (2005).
[26] R. G. Leigh and P. Phillips, Phys. Rev. B **79**, 245120 (2009); P. Phillips, Rev. Mod. Phys. **82**, 1719 (2010).
[27] S. Zhou, Y. Wang, and Z. Wang, Phys. Rev. B **89**, 195119 (2014).
[28] G. Kotliar and A.E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986).
[29] T. Sato and H. Tsunetsugu, Phys. Rev. B **90**, 115114 (2014).
[30] P. Prelovšek, J. Kokalj, Z. Lenarčič, and R. H. McKenzie, Phys. Rev. B **92**, 235155 (2015).
[31] M. C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963).
[32] For a review, see D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984).
[33] W. Brinkman and T. M. Rice, Phys. Rev. B **2**, 4302 (1970).
[34] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
[35] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[36] K. Byczuk and D. Vollhardt, Phys. Rev. B **77**, 235106 (2008).
[37] L. B. Ioffe and A. I. Larkin, Phys. Rev. B **39**, 8988 (1989).
[38] A. Georges, G. Kotliar, W. Krauth, M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[39] M. Karski, C. Raas, and G. S. Uhrig, Phys. Rev. B**77**, 075116 (2008).
[40] D. Vollhardt, Annalen der Physik **524**, 1 (2012).
[41] M. M. Wysokiński and M. Fabrizio, Phys. Rev. B **95**, 161106(R) (2017).
[42] C. Castellani, G. Kotliar, R. Raimondi, M. Grilli, Z. Wang, and M. Rozenberg, Phys. Rev. Lett. **69**, 2009 (1992).
[43] R. Raimondi and C. Castellani, Phys. Rev. B **48**, 11453 (1993).
[44] I. Ichinose and T. Matsui, Phys. Rev. B **51**, 11860 (1995).
[45] L. Liang, S. Zhou, and Z. Wang, to be published.
[46] N. Nagaosa, Phys. Rev. Lett. **71**, 4210 (1993).

[47] Z. Wang, Phys. Rev. Lett. **94**, 176804 (2005).

[48] K.-S. Kim, Phys. Rev. B **72**, 245106 (2005).

[49] V. Ambegaokar, U. Eckern, and G. Schon, Phys. Rev. Lett. **48**, 1745 (1982).

[50] A. Kampf and G. Schön, Phys. Rev. B **36**, 3651 (1987).

[51] J. E. Mooij, B. J. van Wees, L. J. Geerligs, M. Peters, R. Fazio, and G. Schon, Phys. Rev. Lett. **65**, 645 (1990); R. Fazio and G. Schön, Phys. Rev. B **43**, 5307 (1991).

[52] A. Golub, Physica B **179**, 94 (1992).

[53] Ph. Noziéres, Eur. Phys. J. B**6**, 447 (1998).

[54] S. Kehrein, Phys. Rev. Lett. **81**, 3912 (1998).

[55] R. M. Noack and F. Gebhard, Phys. Rev. Lett. **82**, 1915 (1999).

[56] G. Kotliar, Eur. Phys. J. B**11**, 27 (1999).

[57] X.-G. Wen, *Quantum Field Theory of Many-body Systems: From the Origin of Sound to an Origin of Light and Electrons*, Oxford University Press, Oxford, U. K., (2004).