arXiv:cond-mat/0410753 v1 29 Oct 2004

Quantum Mott Transition and Multi-Furcating Criticality

Masatoshi Imada[1,2]

[1] *Institute for Solid State Physics, University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8581*

[2] *PRESTO, Japan Science and Technology Agency*

Phenomenological theory of the Mott transition is presented. When the critical temperature of the Mott transition is much higher than the quantum degeneracy temperature, the transition is essentially described by the Ising universality class. Below the critical temperature, phase separation or first-order transition occurs. However, if the critical point is involved in the Fermi degeneracy region, a marginal quantum critical point appears at zero temperature. The originally single Mott critical point generates subsequent many unstable fixed points through various Fermi surface instabilities induced by the Mott criticality characterized by the diverging charge susceptibility or doublon susceptibility. This occurs in marginal quantum-critical region. Charge, magnetic and superconducting instabilitites compete severely under these critical charge fluctuations. The quantum Mott transition triggers multi-furcating criticality, which goes beyond the conventional concept of multicriticality in quantum phase transitions. Near the quantum Mott transition, the criticality generically drives growth of inhomogeneous structure in the momentum space with singular points of flat dispersion on the Fermi surface. The singular points determine the quantum dynamics of the Mott transition by the dynamical exponent $z = 4$. We argue that many of filling-control Mott transitions are classified to this category. Recent numerical results as well as experimental results on strongly correlated systems including transition metal oxides, organic materials and ^{3}He layer adsorbed on a substrate are consistently analyzed especially in two-dimensional systems. The mechanism of cuprate high-T_c superconductivity is also discussed in the light of the present insight and interpreted from the multi-furcation instability.

KEYWORDS: Ginzburg-Landau theory, quantum phase transition, Mott transition, multi-furcation, two-dimensional Hubbard model, high-Tc superconductivity, quantum critical phenomena, marginal critical point, high-Tc superconductivity, phase separation

1. Introduction

Recently, enormous number of experimental results have been accumulated on properties of metals near correlated insulators such as the Mott insulator.[1] In these results, the nature shows diverse properties depending on subtle differences in systems and ranging from strongly renormalized Fermi liquid, charge order, magnetic order to superconductivity at low temperatures. More complicated and entangled phase diagrams and very sensitive responses to external perturbations are also observed. In some cases, severe competitions of multiple

orders are found. Inhomogeneous structure at the Fermi surface is also widely observed particularly in the high-T_c cuprates. This includes an emergence of flat dispersion in special region of the Brillouin zone as in the $(\pi, 0)$ and $(0, \pi)$ regions in the high-T_c cuprates.[1–3] Spatially inhomogeneous structure is also widely recognized in the cuprates[4] as well as in manganaites.[5,6] Such dramatic competitions and self-organized structure formations in real and momentum spaces have not been expected in weakly correlated electron systems.

Intensive studies on theoretical models of strongly correlated electrons have also revealed subtlety of the results on stabilities of phases and controversies depending on approaches, approximations and the models themselves.[1] These controversies and diversities both in theories and experiments consistently show not necessarily the failure of the present theoretical treatments for each problem but the emergence of an underlying new class of phenomena and necessity for a new concept in physical systems in a region of strong electron correlation to reproduce the diversity and sensitivity.

In this paper, a phenomenological theory of the Mott physics emerging near the Mott insulator is studied and insights from recent experimental results together with numerical results of theoretical models are analyzed. We show that the Mott transition of Fermion systems and its criticality provides a new type of quantum phase transitions, where a classical Ising-type transition is transformed to a quantum one in a unique way. This unique feature naturally accounts for diversity of phenomena with underlying strong competition of orders, emergence of structure in real and momentum spaces.

When the critical temperature of the Mott transition is much higher than the energy scale of the quantum degeneracy, the transition at the critical point retains a classical nature, which is analogous to the gas-liquid transition. Around the critical point, the diverging and critical density fluctuation is expected. At low temperatures, the transition is of the first order and strong quantum fluctuations do not appear.

The gas-liquid type transition (or phase separation) in quantum systems was studied by Blume, Emery and Griffiths[7] in the context to understand the phase diagram of ^{3}He-^{4}He mixture with the superfluid transition of ^{4}He being involved. For Fermion systems, Castellani *et al.*[8] extended this study to the Fermion Hubbard model by including magnetic degrees of freedom and the doubly occupied sites as well. The metal-to-insulator transition may be treated by the analogy to the gas-liquid transition in this framework and Kotliar *et al.*[9,10] indeed derived the Ising-type transition within the framework of the dynamical mean field theory.[11,12] The critical temperature was high enough so that it was recognized to be equivalent to the classical Ising-type transition and quantum effects are rather irrelevant. A fundamental open problem still remains in the issue how quantum effects alter the criticality of the conventional gas-liquid or bainary-alloy transitions of the Ising type.

In Fermion systems, if the critical temperature at the termination point of the first-order

transition is suppressed, the diverging density fluctuations inherent at the critical point of the gas-liquid transition become involved in the quantum Fermi degeneracy region. The Fermi degeneracy by itself generates various instabilities called the Fermi surface effects. Then the coexistence of these two becomes an intriguing issue.

The purpose of this paper is to show that the coexistence of the Fermi degeneracy and the critical density fluctuations yields a new type of quantum criticality. The Mott transition offers a salient example of this issue. At this quantum criticality, the mother transition, namely the Mott transition triggers various daughter instabilities and quantum criticalities such as charge, magnetic, and superconducting transitions. In other words, at the unstable fixed point of the root transition, namely the Mott transition, because of its criticality itself, it generates many other unstable fixed points simultaneously if it coexists with a quantum degeneracy. The coexistence also induces inhomogeneity at the Fermi surface in the momentum space as well as in real space. As far as the author knows, this type of *multi-furcating* or cascading quantum criticality originating from one quantum phase transition has never been considered before and offers a new type of physics beyond the conventional theory of multicritical phenomena in quantum phase transitions as we discuss later. The critical region may not be characterized by simple critical exponents but by possible cascading and hierarchical structure formations. The phenomena must first be understood as a whole; not by individual competitions and transitions but from a synergetic viewpoint. Quantum critical phenomena have been intensively studied in case of the magnetic critical point, where the vanishing magnetic critical temperature in metals causes non-Fermi liquid behavior and large fluctuations in the critical region.[?,?,1] In the present case, quantum criticality of the Mott transition generates a quite different type through diverging charge fluctuations, where the critical temperature as the end point of a first-order transition of the electronic (or excitonic) density is suppressed and a marginal critical point appears at zero temperature.

The applicability of this new class of quantum criticality is not confined to the Mott transition but may also be extended to other type of first-order transitions such as the charge order transitions when their critical points are involved in the Fermi degeneracy region. Such examples may also be found near the first-order chage of the valence in heavy fermioin compounds. Valence fluctuations in heavy fermion compounds and density fluctuations for charge orders in organic and transition metal compounds offer such examples.

In Sec. 2, we summarize the general scheme of the Ginzburg-Landau (GL) expansion within the classical framework. In Sec. 3, two-parameter expansion for the Mott transition is formulated. In Sec. 4, we discuss comparison with results of recent numerical and microscopic calculations as well as comparison with experimental results. We consider quantum effects in Sec. 5. A two-fluid model of the Mott transition is presented in Sec. 6. Section 7 is devoted to a consequence of the coexistence of the Fermi degeneracy and the Mott criticality, where

multi-furcation instability is considered. In Sec.8, insights into experimental indications are discussed. A mechanism of high-temperature superconductivity is also considered.

2. Ginzburg-Landau (GL) Expansion

2.1 Classical GL Picture

At finite temperatures, a metal and an insulator may be separated in the phase diagram if a first-order transition takes place. The first-order transition should terminate at the critical end point at a finite temperature T_c. To understand the origin of the first-order transition, the equivalence with the gas-liquid transitions or the binary-alloy phase separation has been identified.[9,10]

In the mean-field Ginzburg-Landau (GL) expansion of the conventional gas-liquid transition, the free energy is expressed by coarse-grained variables of the natural order parameter, namely, the density n as

$$F = \frac{1}{2}a(T - T_c)n^2 + \frac{1}{4}bn^4 \tag{1}$$

with a and b being constants. Here the density is measured from the critical density. At the critical temperature $T = T_c$, in this mean-field theory, the charge susceptibility (compressibility) $\chi_c \equiv (d^2F/dn^2)^{-1}$ diverges as $\chi_c = 1/(3bn^2)$ at $n = 0$, which results in diverging and critical density fluctuations. As a function of temperature, the charge susceptibility also diverges at the critical density as $\chi_c \propto (T-T_c)^{-\gamma}$ with the mean-field exponent $\gamma = 1$. Below T_c, the first-order transition between gas and liquid occurs between $n = \sqrt{a/b}$ and $n = -\sqrt{a/b}$.

The Mott transition has been considered in the GL expansion in the dynamical mean-field theory.[9,10] Essentially, the dynamical mean-field theory reproduces the Ising-type transition. The natural order parameter is identified as the single-particle local Green's function itself, which represents the transition between localized and itinerant nature of electrons at zero frequency. The transition is well defined at finite temperatures when the structure of the low-frequency Green's function shows a jump as the first-order transition.

2.2 Two Routes of Mott Transition

Microscopically, the driving force of the Mott transition arises from the competition of the kinetic energy and the interaction energy of electrons. The N-site Hubbard model defined by

$$\mathcal{H} = \mathcal{H}_t + \sum_i H_{Ui} - \mu MN \tag{2}$$

$$\mathcal{H}_t = -\sum_{\langle ij\rangle} t_{ij}(c^\dagger_{i\sigma}c_{j\sigma} + h.c.) \tag{3}$$

$$\tag{4}$$

and

$$\mathcal{H}_{Ui} = U(n_{i\uparrow} - \frac{1}{2})(n_{i\downarrow} - \frac{1}{2}), \tag{5}$$

represents this competition in a simplest way. Here, $M \equiv \sum_{i\sigma} n_{i\sigma}/N$ and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ with the creation (annihilation) operator $c^\dagger_{i\sigma}(c_{i\sigma})$ of an electron at the site i with the spin σ. The chemical potential is μ and U is the onsite Coulomb repulsion. The competition of the kinetic and interaction energies can be controlled by two independent parameters, one, the bandwidth relative to the interaction and the other, the chemical potential μ. Then the first route of the Mott transition is realized by changing the ratio t_{ij}/U. The first route, the bandwidth-control transition may also be driven by changing U itself. The other route is actually realized in many cases by changing the filling, which is the conjugate quantity to the control parameter μ. The natural order parameter of the transition is given by the quantity conjugate to the control parameter. In the Hubbard model Hamiltonian, the conjugate quantity to the control parameter U is nothing but the avearged double occupancy defined by the "doublon density", $n_d \equiv \langle n_{i\uparrow} n_{i\downarrow} \rangle$ as we see from Eq.(5). Therefore, n_d or the averaged empty site defined by the "holon density", $n_h \equiv \langle (1-n_{i\uparrow})(1-n_{i\downarrow}) \rangle$ may be taken as a natural order parameter of phenomenological theory for the bandwidth-control transition at half filling. At half filling, we have the constraint $n_d = n_h$. The natural order parameter of the second route is given by the filling $n \equiv \langle M \rangle$ or the doping concentration $\delta = 1 - n$, which is conjugate to the chemical potential μ. Then the GL expansion may be taken by the double expansion of two independent parameters, $\nu_s = n_d + n_h$ and $\delta_s = n_h - n_d$.

3. Two-Parameter GL Expansion for Mott Transition

3.1 Framework

The GL expansion of the free energy near the critical point may be phenomenologically constructed from the expansion in terms of two natural order parameters $\nu = \nu_s - \nu_0$ and $\delta = \delta_s - \delta_c$ measured from the critical points, $\nu = \nu_0$ and $\delta = \delta_c$. Here, the control parameter is shifted from ν_s to ν and measured from the critical value ν_0. Near the critical point of the bandwidth-control transition, the free energy is expanded as

$$\begin{aligned} F &= \frac{1}{2} A_{h0}(T - T_{ch0})\delta^2 + \frac{1}{2} A_d (T - T_{cd})\nu^2 \\ &+ \frac{1}{4} B_{h0}\delta^4 + \frac{1}{2} B_{hd}\delta^2\nu^2 + \frac{1}{4} B_d \nu^4, \end{aligned} \tag{6}$$

where $A_{h0}, A_d, B_{h0}, B_d,\ B_{hd}, T_{ch0}$ and T_{cd} are positive constants. Odd order terms do not appear because of the symmetry of the critical point. In this coarse grained form, only ν and δ are retained as variables and this form should be regarded as the result after tracing out other degrees of freedom including spins. Since the two control parameters are independent, the critical temperatures T_{ch0} and T_{cd} are not necessarily the same in this formulation. From the minimization with respect to ν, the double occupancy at the free energy minimum is given

by

$$\nu = \nu_m \equiv \pm\sqrt{\frac{A_d(T_{cd} - T) - B_{hd}\delta^2}{B_d}}. \tag{7}$$

for $T < T_{cd}$ and $|\delta| \leq \delta_w \equiv \sqrt{A_d(T_{cd} - T)/B_{hd}}$. For $|\delta| > \delta_w$, F is minimized at $\nu = \nu_m = 0$. One consequence of this form of free energy is that the first-order transition occurs even from metal to metal when the bandwidth is controlled in the region $|\delta| < \delta_w$ below T_{cd}, if this density region can be realized.

By taking $\nu = \nu_m$ and eliminating ν, the free energy is expressed by the single parameter δ as

$$F = \frac{1}{2}A_h(T - T_{ch})\delta^2 + \frac{1}{4}B_h\delta^4 + C, \tag{8}$$

where

$$A_h = A_{h0} - \frac{A_d B_{hd}}{B_d}, \tag{9}$$

$$T_{ch} = \frac{1}{A_h}(A_{h0}T_{ch0} - \frac{A_d B_{hd} T_{cd}}{B_d}) \tag{10}$$

and

$$B_h = B_{h0} - \frac{B_{hd}^2}{B_d} \tag{11}$$

for $|\delta| \leq \delta_w$, while $\nu_m = 0$ leads to $A_h = A_{h0}$, $T_{ch} = T_{ch0}$ and $B_h = B_{h0}$ for $|\delta| > \delta_w$. Here C is a constant. This free energy form requires B_h to be a positive constant. This mean-field expansion is justified only near $T = T_{cd}$ and $\nu = \nu_s$, while the degeneracy structure is qualitatively correct along the first-order line below T_{cd}.

3.2 Bandwidth-Control Transition

Below T_{cd} at $\delta = 0$, the first-order transition occurs through the jump of ν from $-\nu_m$ to ν_m. At T_{cd}, the "doublon susceptibility" defined by $\chi_d \equiv [\partial^2 F/\partial\nu^2]^{-1}$ diverges, which means that electron-hole or excitonic density fluctuation diverges. The mean-field exponent is given by $\chi_d = (T - T_{cd})^{-\gamma_d}$ with $\gamma_d = 1$ and $\chi_d \propto \nu^{-\zeta_d}$ with $\zeta_d = 2$. The correct exponents within the classical Ising universality is given by $\gamma_d = 7/4$ and $\zeta_d = 14$ in two dimensions and $\gamma_d \sim 1.24$ and $\zeta_d = 3.8$ in three dimensions.[15] At $T = 0$, the first-order bandwidth-control transition at $U = U_c$ is connected to the first-order line by putting $T = 0$ in Eq.(6). Namely, the jump of ν at $U = U_c$ and $T = 0$ is $\Delta\nu = 2\sqrt{A_d T_{cd}/B_d}$, although the quantitative aspect of the GL expansion becomes questionable if $\Delta\nu$ is not small.

3.3 Filling-Control Transition

When the filling is controlled, the charge susceptibility defined by $\chi_c \equiv dn/d\mu = [\partial^2 F/\partial\delta^2]^{-1}$ diverges at the critical point $T = T_{ch}$.

The filling-control transition occurs at $U > U_c$, where the free energy can be expanded

around the minimum $-\nu_m$ representing the Mott insulator for ν. Since ν does not have criticality away from U_c, the free energy can be expanded solely by δ after eliminating ν. Then the free energy has essentially the same form as eq.(8). The critical exponents are the same as the case of the bandwidth control. Namely, we obtain the mean-field exponent $\chi_c = (T - T_{ch})^{-\gamma_c}$ with $\gamma_c = 1$ and $\chi_c \propto \nu^{-\zeta_c}$ with $\zeta_c = 2$, while the correct exponents within the classical Ising universality is given by $\gamma_c = 7/4$ and $\zeta_c = 14$ in two dimensions and $\gamma_c \sim 1.24$ and $\zeta_c = 3.8$ in three dimensions.

Below $T = T_{ch}$, the phase separation generates a miscibility gap in the density, where the regions of the spinodal decomposition and nucleation are contained. The phase separation occurs because the total particle density is conserved. However, the phase separation has a subtlety because the electrons have charge. The real macroscopic phase separation should be suppressed because of the long-ranged part of the Coulomb interaction, which is beyond the scope of the Hubbard model. In both the nucleation and spinodal regions, the real phase separation is suppressed and may be frozen at a nanoscale regime of the phase separation as if it is frozen at a nonequilibrium state in neutral systems. A related attempt has been made for micro phase separation of weakly charged polymer.[13,56] Nucleation and spinodal regions may show distinct frozen patterns. In the spinodal region, typical entangled stripped pattern at an early stage of the spinodal decomposition is expected to freeze while droplet-like pattern may be seen in the nucleation region. In the nucleation region, the phase separation would be completely suppressed when the critical nuclei radius is larger than the radius allowed from the electrostatic condition. However, if the inhomogeneity in the momentum space occurs as we discuss later, the suppression by the long-ranged Coulomb force may be relaxed to some extent by the formation of inhomogeneous Fermi surface structure. A classification of filling-control Mott transition is possible with the criterion of $T_{ch} > 0$ or formally $T_{ch} \leq 0$.[14]

3.4 Mott Transition and Gas-Liquid Transition, Origin of Attraction

If the transition between the Mott insulator and metals occur through the mechanism similar to the gas-liquid transition or the binary alloy transition, the above phenomenological and classical GL free energy should capture the essence. In the gas-liquid-type transition, the presence of the attractive interaction of particles is known to be crucial. In fact, the liquid phase does not exist without the attractive interaction. When the particle-particle interaction is attractive, we have a negative coefficient for the n^2 term of the free-energy expansion with respect to the density n in Eq.(1). This results in the negative curvature in the free energy as a function of n and this necessarily causes the phase separation (or two-phase coexistence) and the first-order transition, because the negative curvature (convex curve) makes it possible to draw a common tangent between two phase-separated densities.

In case of the transition to the band insulator, the carrier interaction is primarily repulsive and the picture of the gas-liquid transition does not apply straightforwardly. On the other

hand, the Mott insulator and charge order has a stability because of the commensurability effect of the electrons with underlying periodic lattice. This commensurability extraordinarily lowers the energy of the insulating phase as compared to the metallic phase. This lowering makes the kink structure for the energy at commensurate fillings as a function of the filling. The question whether the attractive interaction exists or not as in the classical liquid phase is a highly nontrivial issue.

However, in case of the bandwidth-control transition, the existence of the attractive interaction is rather obvious. The Mott insulating phase is characterized by the doubly occupied site (doublon) and the empty site (holon) forming the bound state while the metallic phase is characterized by the unbinding of the doublon and the holon.[1] The attractive interaction of a doublon and a holon clearly exists if U is large, because the single doublon and holon approximately cost the energy $2U$ while the cost vanishes when they annihilate in pair. The attractive interaction E_b is $2U$ if the kinetic energy is ignored, and can be substantially less if the kinetic energy effect is considered. Then the term proportional to $n_d n_h$ has a negative coefficient in the free energy at $T = 0$. If the filling is fixed at half filling, we have a constraint $n_d = n_h$. After the transformation of the variables from n_d and n_h to ν and δ, the negative curvature of the free energy as a function of ν appears under the constraint of $\delta = 0$. This yields $A_d T_{cd} = E_b$. If quantum effects can be ignored, the critical point thus obtained at $T = T_{cd}$ represents the classical Ising nature of the transition.

The doublon-holon attractive interaction is enlarged by a stronger antiferromagnetic correlation because the doublon and holon disturb the antiferromagnetic background and the energy is lowered if they less disturb by the binding. The order of the transition should be determined from the ratio of the holon-doublon interaction to the formation energy of the holon or doublon, which is basically scaled by U. This is reminiscent of the conversion of the Berezinsky-Kosterlitz-Thouless tansition of the continuous type to the first-order transition when the vortex core energy is reduced relative to the vortex-antivortex attraction.[16] Weaker first-order transition observed for larger geometrical frustration in a numerical study[17] of the Hubbard model on an anisotropic triangular lattice shown in Fig. 1B is explained from the weaker doublon-holon attraction because of the weaker antiferromagnetic background.

In case of the filling-control transition, it is much more subtle as we further discuss later. The filling-control transition is realized by controlling the doping concentration, δ_c or δ. The interaction between two holons is a nontrivial issue and should be quite different from the doublon-holon interaction. As we mentioned above, the stability of the Mott insulator is ascribed to the fact that the electron filling is one per lattice site, which is lost if the filling deviates from the Mott insulator. The energy at half filling is extraordinarily lowered due to the commensurability. If the states deviating from half filling remain uncorrelated metal, it necessarily causes a jump of the energy because of the loss of the energy due to the U term, and

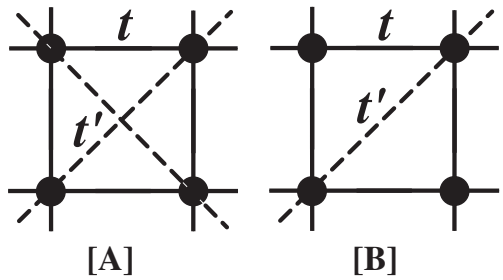

Fig. 1. Lattice structure of geometrically frustrated lattices [A] on a square lattice and [B] on an anisotropic triangular lattice. The nearest- and next-nearest-neighbor transfers are denoted by t (solid bonds) and t' (dashed bonds), respectively.

it trivially leads to the phase separation if the filling deviates from commensurate filling. This means the attractive holon-holon interaction. In the underdoped region, the real metallic state is realized by compromising the correlation effect, and the energy in this region becomes much lower than that calculated from the free fermion state. Then the jump in energy as a function of the filling is transformed to a continuous function with a kink and the negative curveture is much reduced or even may vanish. The origin of the attractive holon-holon interaction, however, if it exists, ultimately traces back to this commensurability force. In any case, at least the holon-holon attractive interaction should be much weaker than the doublon-holon attraction, which explains the continuous character of the filling-control transition.

The present observation suggests that the attractive interaction of holons or holons and doublons may be controlled to some extent. By this control, it may be possible to drive the system to the quantum criticality. At the quantum criticality, the critical end point of the first-order transition (or phase separation) occurs just at zero temperature.

4. Comparison with Numerical and Experimental Results

4.1 Numerical Results

The insight from the GL expansion can be further analyzed with the help of recent numerical results of the Hubbard model. In case of the bandwidth-control transition, the dynamical mean-field theory of the Hubbard model indicates that the transition is of the first order below a finite critical temperature. The critical point is, moreover, characterized by the mean field exponent of the Ising transition.[9,10] Therefore, the GL mean-field picture seems to apply as the mean-field picture of the Mott transition.

In the two-dimensional Hubbard model, the path-integral renormalization group (PIRG) study[18,19] clearly shows that the first-order transition is retained in two dimensions with a jump of the double occupancy n_d at $T = 0$.[17,20] Such bandwidth-control transitions in two dimensions are realized by introducing next-nearest neighbor transfers in addition to the nearest-neighbor transfer in the Hubbard model as is illustrated in Fig. 1. The correlator

projection method,[21] in which the momentum dependence of the self-energy is taken into account in the dynamical mean-field theory, also consistently shows the first-order transition with a jump of the double occupancy and the finite temperature critical point in the same model.[21] As is mentioned above, the first-order transition for the bandwidth control is a natural consequence of strong doublon-holon attraction. The first-order jump decreases with the decrease in the antiferromagnetic correlation by introducing frustration effects. This is also consistent with the picture in the previous section.

In case of the filling-control transition, the numerical results suggest that the order of the transition is more subtle in accordance with the analysis in the previous section. The dynamical mean-field theory appears to show a presumable phase separation (only in the region of very small doping less than 1%) at finite temperature while the single-band model calculation shows the continuous transition at $T = 0$. A small overestimate of the holon-holon interaction by some approximation easily alters the transition from continuous to the first order because an infinitesimally small overestimate transforms the diverging charge susceptibility at $\delta = \delta_s = 0$ to that at a nonzero δ and leads to the phase separation. This issue will be discussed later. It is well known that the mean-field theory may have a tendency to give a fictitious first-order transition because of the ignorance of the fluctuation effects near the transition. Small cluster studies on two-dimensional lattice appeared to show a phase separation,[22] but it turned out to be an artifact and result from overlooking the shell effect seen in the finite size clusters.[23,24] In the quantum Monte Carlo[23–25] and PIRG[26] studies on the two-dimensional lattices, the charge susceptibility divergence appears to occur at zero temperature at the transition point $\delta = \delta_c = 0$. The divergence occurs as $\chi_c \propto \delta^{-1}$. The diverging charge susceptibility at $T = 0$ in fact means that $T_{ch} = 0$ is just satisfied in Eq.(8). This does not exclude very small but nonzero T_{ch} and a small region of the phase separation if it is beyond the numerical accuracy. In addition, the diverging charge susceptibility at $\delta = 0$ suggests that the critical temperature could be easily driven to a small but positive and nonzero value depending on the models, approximations and dimensionality. At least within the Hubbard model, however, the filling-control transition in finite dimensional systems shows no evidence for the phase separation within the numerical accuracy. Instead the filling-control transition in two dimensions shows continuous character with diverging compressibility at vanishing doping concentration δ. The numerical results indicate a general agreement that the bandwidth-control transition is of the first order while the filling-control transition shows a continuous nature with T_c being nearly vanishing.

Recently, it has been clarified that this contrasted behavior between bandwidth-control and filling-control transitions is consistent with the peculiar V-shape structure of the phase boundary between the Mott insulator and metal in the plane of U/t and the chemical potential.[26]

4.2 Comparison with Experimental Results

In strongly correlated electron systems as in transition metal oxides and organic compounds, particularly near their Mott insulating phases, various competing orders and their fluctuations are universally observed and the subject of intensive studies in the recent two decades.[1] We will clarify later that such severe competitions at least partially originate from the tendency toward the underlying criticality of the Mott transition mentioned above. To understand the striking complexity of these competing orders and their fluctuations as compared to weakly correlated systems, it is important to understand the underlying mechanism and physics of the criticality.

Very recently the first-order Mott transition with a critical point at nonzero T in V_2O_3 has been reanalyzed in detail.[27] This bandwidth-control transition is consistent with the above GL expansion and numerical results.

In addition to enormous number of studies in transition metal compounds, we note two interesting recent experimental studies, one on the organic compounds and the other on the adsorbed monolayer ^{3}He on graphite.[28,29] Fournier *et al.*[30] and Kagawa et al.[31] have shown that a first-order transition between the Mott insulator and a metal (at low temperatures, a superconductor) takes place in an organic material, $\kappa-$ET compound, and its first-order boundary starting from zero temperature extends up to the critical end point at around 34K in the plane of pressure and temperature. In this study, the pressure controls the relative bandwidth to the interaction strength while the electron filling is fixed, and therefore, a bandwidth-control transition is realized. In fact, according to the extended Hückel calculation for κ-type ET compound,[32] a minimal model may be the single-band Hubbard model at half filling on an anisotropic triangular lattice defined in Eq.(2) on the lattice structure given in Fig. 1[B]. As we already discussed in the previous subsection, this first-order transition with the critical end point was predicted[17,21] for this lattice structure of the Hubbard model before these experiments.

On the other hand, the ^{3}He monolayer adsorbed on graphite substrate shows a typical filling-control transition to the Mott insulating state by changing the ^{3}He density. The Mott insulating state is realized at the commensurate density with the periodic potential formed by the underlying layer. When the density approaches the Mott insulating state in the liquid phase, it shows striking increase and critical divergence of the effective mass probed by the specific heat coefficient and the magnetic susceptibility.[28,29] The liquid phase nicely obeys the Fermi liquid criterion. Although a possible weak first-order character (namely phase coexistence or phase separation) cannot be excluded, this critical divergence implies the basic continuous character of the filling-control transition in contrast with the bandwidth-control transition observed in the organic compounds.

Of course, it is questionable that the ^{3}He system can be described by the Hubbard model

and more realistic model with hard-core repulsion should be examined for a quantitative analysis. It is also conceivable that the order of the transition may depend on the relevant model as we mentioned above. However, it is also important to realize that the continuous-type transition for the filling control can occur with a critical divergence of fluctuations.

These two different experiments performed on relatively clean and simple systems as compared to transition metal oxides show a similar contrast between bandwidth-control and filling-control transitions to numerical studies which we mentioned above.

5. Quantum Effect

5.1 Mott Transition at T=0

Insulators are characterized by vanishing conductivity at zero temperature. Among various types of insulators, band insulators and Mott insulators are both distinguished from metals by vanishing Drude weight and vanishing charge susceptibility (compressibility) at zero temperature.[1] As compared with phase transitions with symmetry breakings such as magnetic orders, the transitions from metals to the band insulators and the Mott insulators have been considered to be associated with no symmetry breaking by themselves. However, the Drude weight and the charge susceptibility are regarded as the stiffnesses to twisted boundary conditions of the wavefunction phases.[1,33,34] The Drude weight is the stiffness to the twist of the phase of the wavefunction in the spatial direction, while the charge susceptibility is the stiffness to the temporal direction in the path-integral formalism. In this sense, metals in a perfect crystal may be interpreted by the symmetry breaking of the phase of the spatially extended electron's wavefunction as compared to the localized and incoherent wavefunction in the insulator. At zero temperature, insulators cannot be adiabatically continued to a metal and the distinction of metals and insulators is well defined. Then two phases have to be clearly separated by a phase boundary. On the other hand, such distinction is not clear at finite temperatures, because, strictly speaking, the wavefunction coherence is immediately destroyed away from $T = 0$.

5.2 Coexistence with Fermi Degeneracy

The Mott transition discussed so far from the GL mean-field picture is essentially described as a classical one by ignoring quantum effects. The first-order metal-insulator transition can occur even without quantum degeneracy. If the Fermi degeneracy temperature near the first-order transition is much lower than the critical temperature T_{ch} or T_{cd}, the quantum effects can be ignored. In the opposite case, however, the quantum effects must be considered beyond the above framework. The quantum effects have to be seriously considered when the Fermi degeneracy temperature becomes comparable or higher than the critical temperature of the Mott transition. Even when the bare Fermi temperature is high, the effective Fermi temperature is suppressed near the Mott critical point because approaching the continuous transition

to the insulator should suppress the emergence of the Fermi degeneracy. However, the Fermi degeneracy may still coexist with the critical fluctuation in the metallic side near T_c. This should be particularly eminent when T_c is suppressed to zero.

In the Ising-transition picture,[15] the charge susceptibility χ_c diverges at the critical temperature as $\chi_c \propto (T - T_c)^{-\gamma}$ and $\chi_c \propto \delta^{-\zeta}$ with the exponents γ and ζ given above. The transition is characterized by these simple exponents with the hyperscaling assumption being satisfied below the upper critical dimension $d_u = 4$. The system has a single length scale ξ which diverges at $T = T_c$. In the present context, ξ expresses the density correlation length or doublon density correlation length.

In the quantum region, however, we have to consider quantum dynamics. This can be done by considering the path-integral formalism, where the imaginary time direction must be additionally considered in addition to the real spatial dimension. The time scale ω^{-1} diverges as $\omega^{-1} \propto \xi^{-z}$ in addition to the divergence of the spatial correlation length ξ. The quantum dynamics is characterized by the dynamical exponent z.

In the mean-field level, the local quantum effect can be taken into account by the dynamical mean-field theory. The critical point of the metal-insulator transition is signalled by $\partial^2 F/\partial\Delta^2 = 0$, where the hybridization function Δ is the conjugate variable to the single-particle local Green function G and defined from $\Delta = \partial F/\partial G$.

As mentioned above, however, the mean-field theory may overestimate the tendency to the first-order transition, which may mask the quantum effects because of a resultant high critical temperature. The order of the transition and the amplitude of the critical temperature may depend on the details of models, and materials. When the critical temperature T_c is high, as we mentioned, the Mott transition is essentially classical and the mapping to the Ising transition holds. One can suppress T_c by controlling a parameter through the interaction of holons (or holons and doublons). This corresponds to control T_{ch} or T_{cd} to zero. One might think that the critical temperature could become formally negative when one controls the parameter further. This literally means that in the GL expansion of Eqs.(6) and (8), the metal-insulator transition disappears in this region. However, this cannot happen because metals and insulators have to be clearly separated at zero temperature and the phase boundary cannot terminate as we have mentioned above. This necessarily leads to the breakdown of the classical GL expansion or in other words, the transition always exists and the critical temperature stays just at zero temperature through the further parameter control. As we see in Fig.2, this generates a $T_c = 0$ line, where the transition would be far from the gas-liquid picture and the transition stays as the continuous quantum phase transition. In this sense, this quantum criticality is different from the conventional quantum critical point. In the conventional case with a symmetry breaking at finite temperatures, the continuous critical point at $T \neq 0$ terminates at the quantum critical point without first-order transitions and without a $T_c = 0$ line. In the

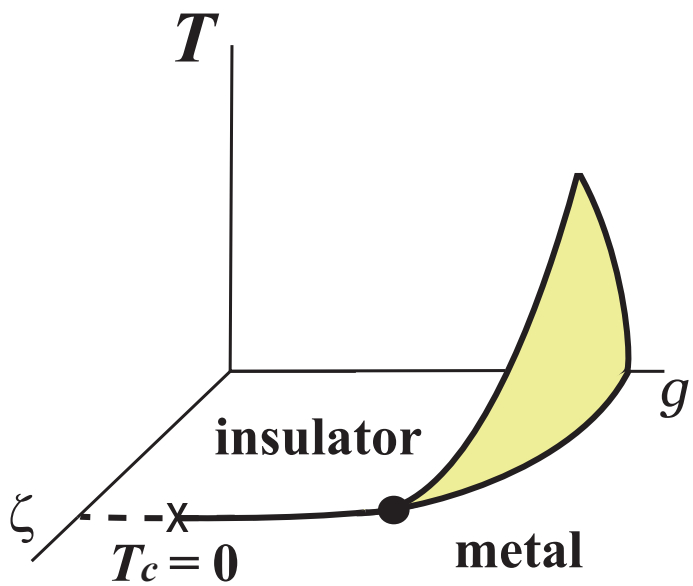

Fig. 2. Schematic phase diagram of metals and insulators in the plane of a control parameter ζ such as U or μ, and g, which additionally controls the interactions of holons and doublons. The continuous transition curve T_c meets zero temperature at the solid circle and then T_c stays at zero with further decrease in g. We call the solid circle point *marginal quantum critical point*. The shaded surface represents the first-order transition. Along the $Tc = 0$ curve, the coefficient of the quadratic dispersion, a, vanishes in the solid part, while it remains positive in the dashed part. Thus the compressibility diverges in the solid part of the T_c curve. We call the region between the cross point and the solid circle (the marginal quantum critical point), *marginal quantum critical region*.

present case, at $T = 0$, the $T_c = 0$ line terminates at *marginal quantum critical point*, where the surface of the first-order transition starts opening with the nonzero critical temperature for the larger g. Hereafter, our focus is largely on the nature around this marginal quantum critical point.

Here, we discuss how the quantum effect alters the transition by assuming the region that the critical temperature is low or even zero. This corresponds to the region of vanishing effective interaction for holon or doublon in the GL expansion. When the critical temperature becomes low, the GL expansion tells that the charge fluctuation becomes diverging accompanied by the quantum degeneracy, which becomes beyond the scope of the form (6) and (8) . The numerical results of the filling-control transition in two dimensions discussed above indeed suggest a continuous transition at zero temperature with the diverging charge susceptibility, which is consistent with $T_{ch} = 0$ and vanishing effective interaction between holons. Therefore, this is certainly a realistic possibility.

5.3 Phenomenological Construction

Here we formulate a phenomenological theory of the Mott transition in the quantum region. When the path integral formalism with the imaginary time τ is employed, the free energy is formally expressed by using the Hubbard model as

$$F = -\ln Z/\beta N \tag{12}$$

$$Z = \int \prod_i \mathcal{D}\phi_i(\tau)\mathcal{D}\phi_i^*(\tau)\mathrm{e}^{-S/\hbar}, \tag{13}$$

$$S = \int_0^{\beta\hbar} \mathrm{d}\tau \sum_i \phi_i^* \hbar\partial_\tau \phi_i + \int_0^{\beta\hbar} \mathrm{d}\tau H(\phi^*, \phi), \tag{14}$$

$$\tag{15}$$

where the Hubbard hamiltonian H for N site system is rewritten by using the Grassmann variables ϕ_i and ϕ_i^* at the site i.

In the metallic phase except in one dimension, the adiabatic continuity of the Fermi liquid allows the description of the free energy by the renormalized single particle at low energies, where the higher-order terms are renormalized to the single-particle coefficient. In the insulating side, the action may also be given from a quasiparticle description with a gap Δ_c as

$$S = \sum_{i,n} \phi^*(-i\omega_n + \Delta_c + E(q,k))\phi \tag{16}$$

with the Matsubara frequency ω_n, and $E(q,k)$ is assumed to satisfy $E(q,k) \geq 0$ and vanishes at the gap edge. Around the gap edge, we take the expansion in terms of k as

$$E(q,k) = a(q)k^2 + b(q)k^4 + ..., \tag{17}$$

where k is the momentum coordinate perpendicular to the locus of $E(q,k) = 0$ and q denotes that parallel to the $E(q,k) = 0$ surface. The gap edge is not necessarily an isolated point in the momentum space, but may be a line or a surface. Just at the transition point, Δ_c vanishes.

The coefficients a and b are regarded as renormalized values after the renormalization group treatment of the higher order terms in the expansion with respect to ϕ. In the presence of the a term, the dynamical exponent z characterized by the dispersion is given by $z = 2$ as in the usual transition to the band insulator.[35] For example in case of the filling-control transition, at $T = 0$, the singular part of the free energy is expanded in terms of the density as $F = \delta^{(d+z)/d}$ as we see later, while the form Eq.(8) is justified at $T > 0$. However, the hyperscaling assumption is still satisfied[35] at $T = 0$. The transition between the band insulator and metals may also be interpreted from Eq.(8) that T_{ch} is kept formally negative for spatial dimensions $d \geq 2$ at $T > 0$, while the singular form beyond Eq.(8) appears only at $T = 0$ as $F = \delta^{(d+2)/d}$.

In this formalism, the attractive holon-holon (or holon-doublon) interaction should be

renormalized and result in the flat dispersion in Eq.(17) at least in a part of the Brillouin zone when the phase separation or the first-order transition starts, because the localization by the phase separation (or diverging doublon susceptibility) should give the vanishing dispersion. In other words, the coexisting phase or phase separation region is characterized by the region of diverging charge susceptibility or doublon susceptibility, which have to generate a diverging density of states at the gap edge in spatial dimension $d \geq 2$. This necessarily leads to $a \to 0$. Toward the marginal quantum critical point, the real part of the lowest order term a continuously vanishes. This alters the dynamical exponent from 2 to 4, because the quadratic term proportional to a vanishes and quartic term proportional to b still remains in general. We note that in the GL description, this marginal behavior with $z = 4$ is realized at the critical point, T_{ch} or T_{cd}. Strictly speaking, however, this argument is justified when the Fermi degeneracy is well developed and the quasiparticle description in the form (16) is valid. This is satisfied around the marginal quantum critical point.

Although the quasi-particle picture does not hold, it has numerically been shown that the dynamical exponent indeed becomes $z = 4$ at the marginal quantum critical point in the one-dimensional Hubbard model with next-nearest-neighbor transfers[37]

If the gap edge given by $E(q, k) = 0$ is composed of a line or a surface in the momentum space, the value a in reality depends on the momentum along this line or surface. When the system is metallized, this line or surface is expected to become the Fermi surface which satisfies the contained Luttinger volume. Generically, the parameter a depends on the momentum. Therefore, when the parameter is controlled to the critical point, it is unlikely that a vanishes simultaneously at all the points along the $E(q, k) = 0$ surface in the metallic side. On the contrary, generically the point with the smallest amplitude of a becomes zero first as a special point of the $E = 0$ surface. Then a quartic dispersion appears at this special point of the $E = 0$ surface when the system becomes marginal, which results in $z = 4$. Therefore, the Mott critical point is also characterized by the evolution of an electron differentiation if the large Fermi surface is involved in the metallic side. The singular differentiation generates a quartic dispersion at particular points of the Fermi surface coexisting with dispersive generic part.

This large dynamical exponent $z = 4$ is indeed supported in a number of independent numerical calculations for the filling-control transition of the Hubbard model in two dimensions at $T = 0$,[23,24,36,38–41] which suggests that T_{ch} is zero or very small in this case.

We now consider the filling-control transition in more detail. In the quantum region, the expansion in the form (6) and (8) is not justified any more. However, the singular part of the free energy may still in principle be expanded at zero temperature as a function of the doping concentration δ. When the quadratic part of the dispersion proportional to a around a point q_1 in eq.(17) remains, the lowest order term of the total energy measured from the insulator

in this quasiparticle picture is given at $T = 0$ from

$$F = Ca\delta^{(d+2)/d} \tag{18}$$

with a constant C and therefore the charge susceptibility shows the scaling $\chi_c \equiv (\partial^2 F/\partial\delta^2)^{-1} \propto \delta^{1-2/d}$, which is the same as the transition to the band insulator. This is because

$$F \propto \int_0^{k_F} k^{d-1} k^2 \mathrm{d}k \tag{19}$$

and

$$\delta \propto \int_0^{k_F} k^{d-1} \mathrm{d}k, \tag{20}$$

where we have integrated around q_1. If a large Fermi surface which satisfies the Luttinger theorem appears immediately upon doping, one has to take that Fermi surface as the locus $E(q,k) = 0$ and one gets

$$F = C\delta^3 \tag{21}$$

because

$$F \propto \int_0^{k_F} k^2 \mathrm{d}k \tag{22}$$

and

$$\delta \propto \int_0^{k_F} \mathrm{d}k, \tag{23}$$

where the integrations are performed in the region around the locus $E = 0$. However, in the marginal region near the marginal critical point with vanishing a term at particular points q_0, we have the lowest order term

$$F = Cb\delta^{(d+z)/d}, \tag{24}$$

with $z = 4$, which yields

$$\chi_c \propto \delta^{1-z/d}. \tag{25}$$

The free energy near the Fermi ground state is therefore expressed as

$$F = C(a\delta^{(d+2)/d} + b\delta^{(d+4)/d}), \tag{26}$$

Even in the case of the bandwidth-control transition, when one can regard the closing of the gap by hole doping around a point q_{0h} and simultaneous particle doping around q_{0p} with the constraint of keeping the electron density $n = 1$, the above relation Eq.(26) may be replaced with ν as

$$F = a\nu^{(d+2)/d} + b\nu^{(d+4)/d}, \tag{27}$$

because the "doublon" and "holon" concentration is nothing but the above self-doping concentration of particles and holes. However, the validity of this picture is limited because the

large Fermi surface which satisfies the Luttinger theorem does not appear in this way. In fact, the bandwidth-control transition seems to occur as the first-order transition at $T = 0$ and the critical behavior may not exist.

A crucial point is that, at the marginal quantum critical point, the diverging charge susceptibility (or doublon susceptibility) driven by $z = 4$ in the Fermi-degenerate region causes a multifurcating criticality as we discuss later. This instability is much stronger in the lower dimensions than in three dimensions as we see from Eq.(25). In two dimensions, $\chi_c \propto \delta^{-1}$, while $\chi_c \propto \delta^{-1/3}$ in three dimensions.

6. Two-Fluid Model

Near the critical point of the Mott transition, $a(q) \geq 0$ has momentum dependence along the locus $E(q,k) = 0$ and the marginal critical region is characterized by the vanishing $a(q)$ at particular points of the locus. If it happens, the Fermi surface roughly consists of the two parts in the metallic side. One is the part where $a(q)$ shows rather large value, while the other is particular points where $a(q)$ gets smaller and smaller around such particular points $q = q_0$ as $a(q) = a_0\varepsilon + (q - q_0)^2$, where ε measures the distance from the critical point in the control parameter, say g. The contribution around the points q_1 where $a_1 = a(q_1)$ is large gives the ground state energy as

$$F = a_1 \zeta^{(d+2)/d}, \tag{28}$$

where ζ is either δ or ν depending on the filling or bandwidth control. The contribution from the area around q_0 is

$$F = a_0 \varepsilon \zeta^{(d+2)/d} + b\zeta^{(d+4)/d}. \tag{29}$$

When the filling is controlled from a heavily doped region toward the marginal quantum critical point, there exists a crossover from the normal behavior, where the free enrgy is insensitive to δ, to the critical proximity region, where

$$F = b\delta^{(d+4)/d}. \tag{30}$$

In the normal region, the specific heat has a normal γ value, which is essentially δ insensitive. The specific heat may have a broad peak around a normal effective Fermi energy. With the decreasing doping concentration, this high-temperature weight is transferred to the low-temperature region, where the critical proximity from Eq.(30) yields a quite different γ value. In this low-temperature structure, the T-linear behavior becomes limited to lower and lower temperatures with decreasing doping concentration. When a_0 vanishes, γ diverges as $\gamma \propto 1/(a_0\varepsilon)$ while this T-linear behavior holds only in the region proportional to $a_0\varepsilon\delta^{2/d} + b\delta^{4/d}$ because the effective Fermi energy is scaled in this way. This low-temperature structure may continuously merge to the bahavior of the insulating phase, where the specific heat is determined from the spin entropy only and may have a peak structure at a characteristic spin-

exchange energy. In fact, this spin entropy contribution persists in the doped region and the region of the linear-T specific heat should be limited to the lower energy than the peak of the spin contribution, when the doping concentration is low enough.

Upon approaching the Mott critical point, the dispersion around q_0 becomes flatter and flatter and the damping effect may also become large. In addition, scattering of quasiparticles in this momentum region may be responsible for the spin correlation in the insulating phase. This flattness is the origin of the charge susceptibility divergence in Eq.(25). Although the doped carriers are largely accepted in this region upon doping because of the flattness, the quasiparticle structure may not be clear because of the damping. On the contrary, other part of the Fermi surface with dispersive quasiparticles accepts lower concentration of carriers while it shows more coherent quasiparticle structure. Because of this electron differentiation depending on the momentum on the Fermi surface, the DC transport contributed from the dispersive region and the spin dynamics mainly arising from the flat region may behave as if they are separately evolving. This is reminiscent of the phenomenological argument of "hot" and "cold" spots[42] in the high-T_c cuprates.

When T_c becomes positive, the phase separation starts at $T < T_c$. This phase separation starts from this flat region of the Fermi surface, while it can coexist with the metallic carriers in the dispersive region, if the carriers are also doped in the dispersive region. If carriers are not doped into the dispersive part, the phase separation must be frozen at an early stage of the spinodal decomposition or the nucleation, because the long-ranged Coulomb repulsion prohibits further evolution of the phase separation. When the holes are doped also into the dispersive part in addition to the flat part, the screening by the dispersive carriers may relax the constraint from the Coulomb repulsion. The characteristic frozen domain size ξ is $\xi \sim \sqrt{\epsilon \Delta E}/(\delta_0 e)$, where δ_0 is the density difference between two phase separated area, ΔE is the free energy barrier between two minima, and ϵ is the effective dielectric constant. The coexistence with the dispersive carrier increases ϵ through screening effects. In a region of the patched real space, the carriers from the flat dispersion region are more dense and in the other region, coherent carriers from the dispersive region are more densely populated.

The two-fluid model is a simplified picture for the coexistence of the dispersive and flat regions near the Mott critical point. For more quantitative analyses, one has to consider the multi-fluid model that contains more than two components because the flat and dispersive regions are continuously interpolated.

7. Multi-furcation Instability

Near the marginal quantum critical point $T_{ch} = 0$, or more precisely in the region of $a = 0$, the charge susceptibility diverges at $T = 0$, while its coexistence with the Fermi degeneracy generates a drastic effect. If these two coexist, the instabilities to various orders including magnetic, charge and superconducting phases are undoubtedly largely enhanced as

compared to the existence of only one. The antiferromagnetism, ferromagnetism, charge order and the superconducting order severely compete, because all of them easily diverge because of enhanced density of states arising from the proximity to the diverging charge susceptibility. The susceptibility at $T = 0$ for the order parameter u is given by

$$\chi_u = \int \mathrm{d}E \rho(E) \frac{|\langle 0|u^\dagger|\varphi_E\rangle|^2}{E-\omega}, \tag{31}$$

which diverges because of the diverging density of states. Here, $|\varphi_E\rangle$ is the eigenstate with the energy E, $|0\rangle$ is the ground state and $\rho(E)$ is the density of states. The above orders are also competing with the tendency towards the phase separation. It turns out that such instability is enhanced in spatial dimensions lower than three because of Eq.(25), namely $\chi_c \propto \delta^{1-4/d}$. Because the symmetry breaking is severely suppressed in one dimension, the systems with strong two-dimensional anisotropy with weak three-dimensional coupling may have the most drastic and largest energy scale of the severe competitions and instabilities.

Near the Mott critical point, the instability towards various symmetry breakings occurs mainly at the particular points p_0 with the flattened dispersion. The amplitude of the instability for different orders sensitively depends on the details of the lattice structure, Fermi surface shape and band structure, because the position of p_0 may depend on details of systems. We do not discuss individual differences of instabilities depending on the detailed distinction of systems. An important point is that such multi-furcating instability generically starts developing and very sensitive dependence of fluctuation enhancement appears when the system approaches the marginal quantum critical point of the Mott transition. At least the diversity and complexity of the phenomena in experiments in this region[1] are understood from the underlying combination of the *Fermi degeneracy* and the *compressibility divergence* caused by a unique quantum criticality of the Mott transition.

It is certainly possible that, under critical charge fluctuations of the Mott transition, some symmetry breaking occurs at finite temperatures before the real Mott critical point is reached. For example, the antiferromagnetic order or superconducting order may appear before the real Mott critical point. In this case, the transition to the Mott insulator occurs from such symmetry broken states and then the transition is ultimately rather analogous to the transition to the band insulator at $T = 0$.[43,44] For example, the transition between an antiferromagnetic insulator and an antiferromagnetic metal show similar behavior with the transition to the band insulator, because the folded Brillouin zone generated by the antiferromagnetic order yields the vanishing Fermi pocket at the metal-insulator transition point. Such transitions between the symmetry broken phases are characterized by the vanishing number of carriers with a noncritical effective effective mass. Instead, if the symmetry breaking does not occur, the real Mott criticality can be observed at $T = T_c$ and the diverging charge fluctuations cannnot be described by a mapping to a transition to the band insulator. Thus a classification

scheme of two different types of transitions to the Mott insulators is useful.[43,44] One is the case with the Mott critical point observable and in the other, the critical point is masked by the symmetry breaking transition induced by the Mott criticality.

In this context of multi-furcation, instabilities of the metals near the Mott insulator have been examined from various standpoints, although the fundamental and unique structure of the Mott criticality with the divergent κ has not necessarily been fully considered. In fact, the controversies[1] in the literature all indicate that the stable phase near the Mott insulator very sensitively depends on models and approximations. The divergent κ as $\delta \to 0$ easily induces a real phase separation if a small hole-hole attraction is induced either by a real additional perturbation or by an artifact of the approximation. This means that T_{ch} may be controlled to a substantially positive value if some additional interaction can be generated. Then the phase separation can occur even at the classical level without the Fermi degeneracy. This may happen, for example, when the antiferromagnetic order is overestimated than the real one because it introduces an additional attraction of holes by hand. A similar situation may appear when one takes the t-J model with a rather large J. The interaction of the holons should be determined in a self-consistent fashion with very subtle balance in the realistic electronic structure near half filling while the J term controlled by hand very easily destroys this requirement.

A more intriguing possibility among the candidates of the furcation is the superconductivity, where the coexistence of the quantum degeneracy is crucially important. There exist a large number of works on the possibility of superconducting phase without paying full attention on the unique character of the Mott transition discussed in this paper. Among them we briefly discuss some of the studies in relation to the present scope. The enhancement of the d-wave superconducting correlations are signaled by tuning the level structure near the Fermi surface in numerical studies,[45] which also suggests that the pairing is sensitively enhanced by emphasizing the degeneracy near the Fermi level. The FLEX studies[46] show an enhanced pairing induced by the effective attractive interaction coming from enhanced and overestimated antiferromagnetic fluctuations. The stabilization of the superconducting phase is also shown by taking a relatively large J in the t-J model in variational or mean-field studies.[47] The J term in the t-J model explicitly enhances an attraction of two singlet electrons by hand, while an overestimate of the role of the antiferromagnetic correlations may also overestimate the stability of superconductivity.

The stabilization of the superconducting state is also seen by taking a small additional perturbation of correlated hopping in numerically well controlled calculations.[48,49] The correlated hopping term is in fact derived in the strong-coupling expansion of multi-band systems and can be significant because of the enhanced compressibility near the Mott transition.[39,50] This superconductivity is driven by the kinetic energy gain, which is entirely different from

the BCS mechanism.[50] The kinetic energy gain arises because the quantum Mott criticality suppresses the kinetic energy of single particle through $z = 4$. The attractive interaction of two pairs, each propagating coherently, is much reduced as compared to the attractive interaction of two holes, each coming from q_0 and $-q_0$. In other words, the marginal quantum critical region for paired electrons shifts from that of single particle so that the $T_c = 0$ line extends for paired electrons than for unpaired electrons. Because of this shift, the dispersion of pairs is not severely suppressed as compared with the single-particle dispersion. An extreme sensitivity toward the superconducting instability may in reality be again a consequence of the diverging density of states near the Mott insulator. In fact, when the effective interaction of the *mobile* holes becomes attractive, it is very natural to have a stabilization of the superconducting phase before the real phase separation (or spatial inhomogeneity) takes place when the Fermi degeneracy coexists. The kinetic pairing mechanism is expressed by the following effective Hamiltonian:

$$\mathcal{H} = \sum_{q,\sigma} E_1(q) c^\dagger_{q\sigma} c_{q\sigma} + \sum_q E_2(q) \Delta(q)^\dagger \Delta(q) \tag{32}$$

with the pairing order parameter $\Delta(q) = f(q) c_{q\uparrow} c_{-q\downarrow}$. In (32), $E_1(q)$ has vanishing dispersion around q_0 only with the quartic term as $E_1(q) = (q - q_0)^4$, while $E_2(q)$ has a normal dispersion with the quadratic term. This drives the pairing of this form. The form factor f should be chosen so that the dispersion $E_2(q)$ becomes the largest. Because of this two-particle dispersion, dispersive pairing states are formed within the single-particle Mott gap.

It should also be noted that if $a(q)$ becomes vanishing at four points, say, $q_0, -q_0, q_1$ and $-q_1$ on the Fermi surface in the Brillouin zone, it additionally favors the superconducting instability by the pairing of

$$P_{q_0,q_1} = \langle c_{q_0\uparrow} c_{-q_0\downarrow} - c_{q_1\uparrow} c_{-q_1\downarrow} \rangle, \tag{33}$$

where the exchange interaction may work as attractive because of the negative phase factor in (33). This situation indeed occurs in the d-wave pairing realized in the high-T_c cuprates at around $(\pi, 0)$ and $(0, \pi)$.

The analysis in the present paper adds a new insight into the controversies for the determination of the ground state. Most of the examined instabilities, including magnetic, charge, and pairing ones, are singularly enhanced by the coexistence of the Fermi degeneracy and the Mott criticality. All these instabilities induced by a small perturbation or crudeness of the approximation together with their controversies depending on the models and approximations certainly indicate the significance of underlying "multi-furcating criticality". The multi-unstable nature is also tightly connected with unusually suppressed coherence of quasiparticles in the flat-dispersion region around q_0, which is also a consequence of a large dynamical exponent $z = 4$ derived in the underlying Mott transition. The system becomes *almost* unstable to

various directions in this case of the quantum criticality.

Then we realize that we have to be very careful in concluding which type of the instabilities eventually wins since calculations suggest a strong degeneracy of various different phases under quantum fluctuations. As we mentioned, if the critical point is suppressed to lower and lower temperatures as in the filling-control case, the instability is more enhanced because of the interplay with the Fermi degeneracy. Near the critical points of the Mott transition, whichever at $T_c = 0$ and $T_c \neq 0$, the electronic state may have strong instabilities and they even further generate new subsequent instability mediated by enhanced fluctuations. Without clarifying this subtle multi-furcating instabilities and competitions under the proximity to the compressibility divergence, the final result at zero temperature would not be reliable.

We again note that the diverging charge susceptibility at the Mott critical point is connected with particular points of the Fermi surface, where the dispersion becomes flat. This causes the "differentiation" of electrons depending on the position of the Fermi surface. It should be noted that this differentiation has a positive feedback effect. The region of the Fermi surface, which has initially relatively stronger correlation effects, due either to the initial Fermi surface anisotropy or to anisotropic growth of fluctuations, allows stronger and stronger correlation effects. This positive feedback appears because the dressed and slower quasipaticles with stronger damping capture the interaction effects more sensitively and strongly. Such positive feedback effects have numerically been seen in the flat dispersion around $(\pi, 0)$ and $(0, \pi)$ points near the Fermi surface[40] of the Hubbard model and also discussed by the numerical renormalization group.[51] The "differentiation" in the momentum space may have a further possibility of the higher order structure at lower energy level, which results in a typical complex phenomena with a residual entropy retained.

Moreover, it opens a new field of critical phenomena, where the originally single unstable fixed point generates "daughters", namely, a multi-furcation of the subsequent and different unstable fixed points and the competitions of these fluctuations generate further lower-energy structure through a transient inhomogeneity in real and momentum space. This is beyond the conventional scheme of the critical phenomena. In fact, the analogous GL bifurcation point in the classical system generates simply a diverging susceptibility at the critical temperature, for example, in the gas-liquid and binary alloy mixture transitions. At the classical level, this at most generates a nucleation and a spinodal decomposition regimes below the critical temperature and the physics is rather simple. However, in the case of the Mott transition, the apparently similar problem turns out to be rich due to the interplay with the Fermi degeneracy, whose consequence may be regarded as a quantum emergence from the interplay of these two elements. Under a careful parameter control, an emergence of a complicated hierarchy is expected with lowering energy after severe competitions of various fluctuations before the ground state is reached.

An attempt for searching a more enhanced instability towards such as superconductivity is also the subject of recent intensive studies, because the instability may depend on details of the models and a number of degrees of freedom.[52,53] Near the Mott critical point, emergence of the flat dispersion around q_0 points is a key element when one wishes to design an enhancement of desired orders, since the emergence of q_0 points may be properly controlled by system parameters. By considering the above viewpoint of new hierarchy generation, this is a significant issue to be studied further.

We have discussed only for the region near the Mott insulator at half filling of the Hubbard model. However, in principle, a similar structure appears near any simple commensurate fillings of particle density. It is known in many correlated electron systems that the Mott insulating state called the charge order is universally observed at a simple commensurate fillings such as $n = 1/3$ and $1/4$, while it quantum mechanically melts away from such simple fractional densities.[1] With the long-ranged Coulomb interaction, the melting is also driven by decreasing m^*/ϵ with m^* and ϵ being the effective mass and the dielectric constant. With increasing m^*/ϵ, charge orderings are stabilized at more and more complicated fractional fillings and a structure of the devil's staircase appears as clarified recently by the PIRG method.[54] In principle, a similar structure with divergent compressibility to the half-filled Mott insulator considered in this article may appear around each charge ordered states, which generates another complexity and hierarchy structure. The order of the charge order transition may again be determined from the ratio of the formation energy of the defect structure (charged defects) to the amplitude of defect-defect interaction. An important difference from the quantum Mott transition is that the critical point is actually the marginal critical point in case of the charge order transition. This is because that the charge-order transition necessarily accompanies the translational symmetry breaking. Therefore, the critical point as the termination point of the first-order transition must be the end point of the critical line.

8. Insight into High-Tc Cuprates and Other Strongly Correlated Materials

It is useful to discuss the above consequences in relation to the observed experimental results. Various experimental results suggest that the Mott transitions in many of the transition metal compounds are rather close to the marginal quantum critical region especially for the filling-control transition. ^{3}He adsorbed on a substrate also shows a critical behavior of the mass enhancement when the filling is controlled close to the Mott insulator (or registered phase).

For the case of the bandwidth-control transition, the κ-ET type organic material[30,31] and V_2O_3[27] beautifully show the existence of the finite temperature critical point with the first-order transition below it.

For the filling-control transition, in particular high-T_c cuprates show various indications which are consistent with the present picture. In the high-Tc cuprates, flat dispersions are

observed near $(\pi, 0)$ and $(0, \pi)$ points. It is widely recognized that the flatness is beyond the conventional expectation from the van Hove singularity. Such flattening is a natural consequence if the Mott critical point at $T = T_{ch}$ is located at low temperatures or at $T = 0$ with an emergence of the point q_0 with vanishing quadratic dispersion, given by $a = 0$ as we clarified in the previous sections. The actual q_0 positions responsible for the Mott criticality may depend on materials and models. For example, Y and Bi compounds of the high-T_c cuprates seem to be well modelled by a larger t' than La based high-T_c compounds when one takes the Hubbard model (2).[55] Then the singular points q_0 may deviate from $(\pi, 0)$ and $(0, \pi)$ points for larger t'. Width of the critical region may be influenced by the amplitude of t', while the existence of the criticality itself appears to be universal. In the electron doped cuprates, the critical behavior does not seem to exist because the doping within the antiferromagnetic ordered phase is equivalent to the doping into the band insulator.

If the phase separation or spatial inhomogeneity of charge occurs, the critical point of the diverging compressibility shifts to a finite temperature as the critical termination point of the phase separation. If this critical temperature is still in the region of the Fermi degeneracy in the metallic side, the same physics of the multi-furcating criticality we discuss in this article survives. With the existence of the long-ranged Coulomb force, the phase separation is of course converted either to the charge order (or stripe) with a finite-length periodicity or to an inhomogeneity like the quenched and transient spinodal pattern, which has indeed been observed on the surface of the high-T_c cuprates by the scanning tunnel microscope spectroscopy(STM and STS).[4] The observed spatial pattern of the inhomogeneity on the surface of the cuprate superconductors is very similar to the spinodal pattern widely observed in binary alloy systems.[56,57] The freezing of the spinodal pattern without the phase separation may be well explained by the long-ranged Coulomb repulsion.

The tendency for the phase separation is enhanced when itinerant electrons are coupled to classical degrees of freedom and thus the Mott critical temperature T_c is enhanced. This happens in perovskite manganites, where the itinerant e_g electrons are strongly coupled to localized t_{2g} degrees of freedom through Hund's rule coupling.[5,6] It also couples to lattice degrees of freedom.

The experimental results suggest that the stripe or inhomogeneity is actually rather easily stabilized if a corrugation potential from the lattice distortion or effects of random impurity potential, easily introduced by the doping, are present. This is again the consequence of the "almost phase separated" nature realized by the criticality of the Mott transition with diverging κ at the critical point. If other symmetry breakings can be excluded, it is clear that the enhanced compressibility causes such inhomogeneous charge patterns. Although it is not clear whether the charge inhomogeneity is an intrinsic property of the high-Tc cuprates or it is stabilized just by the randomness and/or surface effects, the observed STM image strongly

suggests that they are close to the spinodal or nucleation regions of the phase separation, because such inhomogeneous patterns have not been observed neither in heavily doped region of the cuprates nor in weakly correlated conventional metals even when the randomness is present. Though the charge susceptibility $\chi_c(q,\omega)$ is suppressed strictly at $q = 0$ and $\omega = 0$, the experimental results suggests a substantial enhancement in the length scale smaller than the size of inhomogeneous pattern (~ 10 nm).

It should be noted that ^{3}He system[28,29] does not have charge in contrast to the electronic systems. He system seems to show continuous transition without a clear phase separation. The present analysis suggests that T_{ch} is very low in this case. It would be intriguing to analyze the Mott transition of ^{3}He system on the substrate by the two-fluid or multi-fluid models discussed in Sec.6, as we can examine a wide temperature range.

The proximity of the Mott criticality around the marginal quantum critical point reveals, first, enhancement of the energy scale of the instabilities to various orders with strong competitions, and second, emergence of the electron differentiation in the momentum space together with a tendency for the inhomogeneity in the real space. The electron diffferentiation generates flat bands around particular points of the Fermi surface while it strongly favors the anisotropic superconductivity. All of these unusual properties are indeed observed in the high-T_c cuprates. The competing orders are a consequence of the mother criticality of the Mott transition, which generates many daughter instabilities. Although various mechanisms of the superconductivity has been discussed in the literature, the criticality of the quantum Mott transition including the divergence of the charge susceptibility and charge fluctuations has not been fully studied. From the present insight, the primary mechanism of the cuprate superconductors is ascribed to the mother criticality of the quantum Mott transition, with the diverging charge fluctuations. Secondary support of the pairing can be found in various fluctuations induced by the competitors such as the antiferromanetic fluctuations, while it does not drive the energy scale as high as in the Mott critical region.

Further studies along this line would certainly be an intriguing issue. More quantitative analysis of the multi-furcation due to the Mott criticality would be desired not only for searching superconductivity but also for other instabilities such as ferromagnetism and ferroelectricity.

Acknowledgments

The author would like to thank H. Nakano for useful discussions on unpublished results prior to publication.

References

1) For a review, see M. Imada, A. Fujimori and Y. Tokura: Rev. Mod. Phys. **70** (1998) 1039.
2) Z.-X. Shen and D. S. Dessau: Phys. Rep. **253** (1995) 1.
3) A. Ino et al.: Phys. Rev. B **62** (2000) 4137.
4) Z.Q. Wang, J.R. Engekbrecht, S.C. Wang, H. Ding and S.H.H. Pan: Phys. Rev. B **65** (2002) 064509;S.H. Pan, J.P. O'Neal, R.L. Badzey, C. Chamon, H. Ding, J.R. Engelbrecht, Z. Wang, H. Eisaki , S. Uchida, A.K. Guptak, K.W. Ng, E.W. Hudson, K.M. Lang, J.C.Y. Davis:Nature **413** (2001) 282; Y. Kohsaka, T. Hanaguri, K. Kitazawa, M. Azuma, M. Takano and H. Takagi: J. Low Temp. Phys. **131** (2003) 299.
5) For example, M. Uehara, S. Mori, C.H. Chen and S.-W. Cheong: Nature **399**(1999) 560
6) For reviews, Y. Tokura: in *Colossal Magnetoresistance Oxides* ed. by Tokura, Gordon and Breach, New York, 2000; E. Dagotto, T. Hotta and A. Moreo: Phys. Rep.
7) M. Blume, V.J. Emery and R.B. Griffiths: Phys. Rev. A**4** (1971) 1071
8) C. Castellani, C. Di Castro, D. Feinberg and J. Ranninger: Phys. Rev. Lett **43** 1(1979) 1957
9) G. Kotliar, S. Murthy and M.J. Rosenberg: Phys. Rev. Lett. **89**(2002) 046401.
10) G. Kotliar, E. Lange and M.J. Rozenberg: Phys. Rev. Lett. **84** (2000) 5180.
11) W. Metzner and D. Vollhardt: Phys. Rev. Lett. **62** (1989) 324
12) A. Georges *et al.*: Rev. Mod. Phys. **68** (1996) 13.
13) For example, V.Y. Borue and I.Y. Erukhimovich: Macromolecules **21** (1988) 3240.
14) D.-H. Lee and S. A. Kivelson: Phys. Rev. B **67** (2003) 24506.
15) For example, see N. G. Goldenfeld: *Lectures on Phase Transitions and Renormalization Group* (Addison-Wesley, 1992).
16) Y. Saito: Phys. Rev. B **26** (1982) 6239.
17) H. Morita, S. Watanabe and M. Imada: J. Phys. Soc. Jpn. **71** (2001) 2109.
18) T. Kashima and M. Imada: J. Phys. Soc. Jpn. **70** (2001) 2287.
19) M. Imada and T. Kashima: J. Phys. Soc. Jpn. **69** (2000) 2723.
20) T. Kashima *et al.*: J. Phys. Soc. Jpn. **70** (2001) 3052.
21) S. Onoda and M. Imada: Phys. Rev. B **67** (2003) 161102.
22) V.J. Emery and S.A. Kivelson: Nature (London) **374** (1995) 434.
23) N.Furukawa and M. Imada: J. Phys. Soc. Jpn. **61** (1992) 3331.
24) N. Furukawa and M. Imada: J. Phys. Soc. Jpn. **63** (1993) 2557.
25) M. Kohno: Phys Rev. B **55** (1997) 1435.
26) S. Watanabe and M. Imada: cond-mat/0312671
27) P. Limelette, A. Georges, D. Jerome, P. Wzietek, P. Metcalf and J.M. Honig: Science **302** (2003) 89.
28) A. Casey, H. Patel, J. Nyeki, B.P. Cowan, J. Saunders: Phys. Rev. Lett.**90** (2003) 115301.
29) K. Ishida, M. Morishita, K. Yawata and H. Fukuyama: Phys. Rev. Lett.**79** (1997) 3451.
30) D. Fournier, M. Poirier, M. Castonguay, et al.: Phys. Rev. Lett.**90** (2003)127002.
31) F. Kagawa, T. Itou, K. Miyagawa and K. Kanoda: Phys. Rev. B **69** (2004) 1.
32) K. Oshima et al.: Phys. Rev. B **38** (1998) 938.
33) W. Kohn: Phys. Rev. A**133** (1964) 171.
34) M. P. A. Fisher, P.B. Weichman, G. Grinstein and D.S. Fisher: Phys Rev. B **40** (1989) 546.

35) See Chap.II.F in Ref.1.
36) M. Imada: J. Phys. Soc. Jpn. **64** (1995) 2954.
37) H. Nakano and Y. Takahashi: unpublished
38) F.F. Assaad, and M. Imada: Phys. Rev. Lett.**76** (1996) 3176.
39) H. Tsunetsugu and M.Imada: J. Phys. Soc. Jpn. **67** (1998) 1864.
40) F. F. Assaad and M. Imada: Eur. Phys. J. B **10** (1999) 595.
41) H. Nakano and M.Imada: J. Phys. Soc. Jpn. **68** (1999) 1458.
42) L.B. Ioffe and A.J. Millis: Phys. Rev. B **58** (1999) 11631.
43) M. Imada: J. Phys. Soc. Jpn. **62** (1993) 1105.
44) M. Imada: J. Phys. Soc. Jpn. **63** (1994) 3059.
45) K. Kuroki and H. Aoki: Phys. Rev. B **56** (1997) R14287.
46) B.Z. Bickers, D.J. Scalapino and S.R. White: Phys. Rev. Lett. **62** (1989) 961.
47) G. Baskaran, Z. Zou and P.W. Anderson: Solid State Commun. **63** (1987) 973.
48) F. F. Assaad, M. Imada and D. J. Scalapino, Phys. Rev. Lett. **77** (1996) 4592.
49) H. Tsunetsugu and M. Imada, J. Phys. Soc. Jpn. **68** (1999) 3162.
50) M. Imada and S. Onoda, *"Open Problems in Strongly Correlated Electron Systems"* eds. J. Bonca et al. (Kluwer Academic Publishers 2001) p.69-80.
51) C. Honerkamp, M. Salmhofer, N. Furukawa and T.M. Rice, Phys.Rev. B **63** (2001) 035109.
52) R. Arita, K. Kuroki and H. Aoki: J. Phys. Soc. Jpn. **69** (2000) 1181.
53) M. Imada and M. Kohno: Phys. Rev. Lett. **84** (2000) 143.
54) Y. Noda and M. Imada: Phys. Rev. Lett. **89** (2002) 176803.
55) E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, and O. K. Andersen: Phys. Rev. Lett. **87** (2001) 47003.
56) See for example, A. Onuki: Phase Transition Dynamics (Cambridge University Press, 2002).
57) J.W. Cahn: in Critical Phenomena in Alloys, Magnets and Superconductors, ed. by R.I.Jafee et al. (McGraw-Hill, New York,1971) p.41.