Scaling theory of the Mott-Hubbard metal-insulator transition in one dimension

C. A. Stafford
Department of Physics, University of Maryland, College Park, Maryland 20742

A. J. Millis
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 27 January 1993)

We use the Bethe ansatz equations to calculate the charge stiffness $D_c = (L/2)d^2E_0/d\Phi_c^2|_{\Phi_c=0}$ of the one-dimensional repulsive-interaction Hubbard model for electron densities close to the Mott insulating value of one electron per site ($n = 1$), where E_0 is the ground-state energy, L is the circumference of the system (assumed to have periodic boundary conditions), and $(\hbar c/e)\Phi_c$ is the magnetic flux enclosed. We obtain an exact result for the asymptotic form of $D_c(L)$ as $L \to \infty$ at $n = 1$, which defines and yields an analytic expression for the correlation length ξ in the Mott insulating phase of the model as a function of the on-site repulsion U. In the vicinity of the zero-temperature critical point $U = 0$, $n = 1$, we show that the charge stiffness has the hyperscaling form $D_c(n, L, U) = Y_+(\xi\delta, \xi/L)$, where $\delta = |1 - n|$ and Y_+ is a universal scaling function which we calculate. The physical significance of ξ in the metallic phase of the model is that it defines the characteristic size of the charge-carrying solitons, or *holons*. We construct an explicit mapping for arbitrary U and $\xi\delta \ll 1$ of the holons onto weakly interacting spinless fermions, and use this mapping to obtain an asymptotically exact expression for the low-temperature thermopower near the metal-insulator transition, which is a generalization to arbitrary U of a result previously obtained using a weak-coupling approximation, and implies holelike transport for $0 < 1 - n \ll \xi^{-1}$.

I. INTRODUCTION

The Mott-Hubbard metal-insulator transition[1] is one of the fundamental problems in electronic condensed-matter physics. The assertion by Anderson[2] that the high-T_c CuO_2 superconductors should be viewed as lightly doped Mott insulators has stimulated much work on the problem. In this paper, we use the Bethe ansatz equations[3,4] to calculate the zero-temperature conductivity and low-temperature thermopower of the one-dimensional (1D) repulsive-interaction Hubbard model for electron densities close to the Mott insulating value of one electron per site.

One way to characterize interacting fermion systems is via the zero-temperature frequency-dependent conductivity $\tilde{\sigma}(\omega)$, which is the linear response of the system to a spatially uniform, time-dependent electric field. In general, $\tilde{\sigma}(\omega)$ has the form

$$\tilde{\sigma}(\omega) = 2D_c\left(\frac{i}{\omega} + \pi\delta(\omega)\right) + \tilde{\sigma}_{reg}(\omega), \tag{1.1}$$

where $\lim_{\omega\to 0}\omega\tilde{\sigma}_{reg}(\omega) = 0$. The coefficient D_c is the charge stiffness. In a Galilean-invariant system with n electrons per unit volume of charge e and mass m interacting via velocity-independent forces, $D_c = ne^2/2m$ and $\tilde{\sigma}_{reg}(\omega) = 0$, independent of the interactions. For electrons moving in an external potential, D_c may depend on the interactions and on the potential. Kohn has argued[5] that for spatially infinite systems with a discrete translational invariance two cases are possible: either $D_c = 0$ and $\lim_{\omega\to 0}\mathrm{Re}[\tilde{\sigma}_{reg}(\omega)] = 0$, or $D_c > 0$. If $D_c = 0$, the material is insulating; if $D_c > 0$, the material has the infinite dc conductivity and perfect diamagnetism expected of a perfect metal. D_c is thus an order parameter[5] for the Mott-Hubbard metal-insulator transition which may occur in a nondisordered system as the electron concentration and interaction strength are varied.

In a system of finite length L with periodic boundary conditions, D_c is in general nonvanishing.[6] In a previous paper,[7] we showed numerically and via a large-U approximation that in the Mott insulating phase of the 1D Hubbard model, which occurs at $n = 1$ for repulsive interactions, the large-L behavior of D_c is

$$D_c(L) \sim L^{1/2}\exp(-L/\xi). \tag{1.2}$$

This equation defines the correlation length ξ of the Mott insulator, which is a function of the on-site repulsion U. In this paper, we derive Eq. (1.2) analytically for arbitrary U, and show that the correlation length so defined is important also in the metallic phase of the model. In particular, we show that in the vicinity of the zero-temperature critical point $U = 0$, $n = 1$, the charge stiffness assumes the hyperscaling form:

$$D_c(n, L, U) = Y(\xi\delta, \xi/L), \tag{1.3}$$

where $\delta = |1 - n|$ and Y is a (presumably universal[8]) scaling function which we calculate.

The physical significance of ξ in the metallic phase of the model is that it defines the characteristic size of the charge-carrying solitons, or *holons*. We show that Woynarovich's reformulation[9] of the Bethe ansatz equations

of the 1D Hubbard model in terms of the parameters of the charge excitations only is formally equivalent to an asymptotic (large L) Bethe ansatz for *holons*, and use this asymptotic Bethe ansatz to construct an explicit mapping for arbitrary U and $\xi\delta \ll 1$ of the low-lying charge degrees of freedom of the model onto weakly interacting spinless fermions. Using this mapping, we obtain asymptotically exact expressions for the charge stiffness and low-temperature thermopower near the metal-insulator transition. Our result for the thermopower extends and makes more rigorous previous work of Schulz[10] which was based on a weak-coupling approximation, and implies that the transport is holelike for $0 < 1-n \ll \xi^{-1}$.

The paper is organized as follows: In Sec. II, we review the linear-response arguments and Bethe ansatz equations used to calculate D_c in 1D. In Sec. III, we present an exact result for the charge stiffness at $n = 1$ in the large-L limit, verifying Eq. (1.2), and show that the correlation length so defined also governs the exponential decay of the equal-time single-particle Green's function at $n = 1$ in both the weak- and strong-coupling limits, as well as the pairing correlations in the ground state of the attractive 1D Hubbard model at $n = 1$. In Sec. IV, we determine the scaling behavior of D_c near the metal-insulator critical point. The total optical spectral weight $\pi N_{\rm tot} \equiv \int_0^\infty {\rm Re}\,[\tilde\sigma(\omega)]d\omega$ is also discussed. In Sec. V, we calculate the finite-size corrections to D_c and $N_{\rm tot}$ in the regime $L > \xi$, and discuss the implications for the interpretation of numerical calculations of $\tilde\sigma(\omega)$ in interacting fermion systems on small clusters. In Sec. VI, we determine the charge excitations, charge stiffness, and low-temperature thermopower near the metal-insulator transition. We summarize our results in Sec. VII, and comment briefly on their relevance for the physics of high-T_c superconducting materials. The details of the calculations of D_c at and near $n = 1$ in the large-L limit are given in Appendixes A and B, respectively.

II. FORMALISM

We consider the Hubbard model of spin-1/2 fermions hopping with matrix element t between nearest-neighbor sites of a one-dimensional lattice with unit lattice constant and subject to a repulsive interaction U when two fermions (of opposite spin) occupy the same lattice site. We impose periodic boundary conditions and thread the system with a (dimensionless) spin-dependent flux Φ_σ, which we represent by a uniform spin-dependent vector potential $A_\sigma = (\hbar c/e)\Phi_\sigma/L$, where L is the circumference of the system. The Hamiltonian may be written

$$H = -t\sum_{l,\sigma}\left[e^{i\Phi_\sigma/L}c^\dagger_{l+1\sigma}c_{l\sigma} + {\rm H.c.}\right] + U\sum_l n_{l\uparrow}n_{l\downarrow}, \tag{2.1}$$

where $n_{l\sigma} = c^\dagger_{l\sigma}c_{l\sigma}$ and the sums run from $l = 1$ to L and $\sigma = \uparrow,\downarrow$. In the following, we set $e = \hbar = t = 1$ unless explicitly stated.

The numbers M and M' of up- and down-spin electrons are constants of the motion; we denote the minimum-energy eigenvalue in a given sector (M, M') by $E(M, M', U, \Phi_\uparrow, \Phi_\downarrow)$. The sign of the interaction U in Eq. (2.1) can be reversed by a particle-hole transformation on the down-spin electrons,[11]

$$\begin{aligned} c^\dagger_{l\downarrow} &\to (-1)^l c_{l\downarrow}, \quad & c^\dagger_{l\uparrow} &\to c^\dagger_{l\uparrow}, \\ c_{l\downarrow} &\to (-1)^l c^\dagger_{l\downarrow}, \quad & c_{l\uparrow} &\to c_{l\uparrow}, \end{aligned} \tag{2.2}$$

leading to the following relation:

$$E(M, M', U, \Phi_\uparrow, \Phi_\downarrow) = E(M, L-M', -U, \Phi_\uparrow, -\Phi_\downarrow) + MU, \tag{2.3}$$

which we will exploit in Secs. III and VII.

To simplify matters, we consider only systems where the total number of electrons $N = M + M'$ is even. Then the ground state is a singlet[12] and the ground-state energy is $E_0 = E(N/2, N/2, U, \Phi_\uparrow, \Phi_\downarrow)$. The charge stiffness D_c is defined by[5,4]

$$D_c = \frac{1}{2}\frac{d^2(E_0/L)}{d(\Phi_c/L)^2}\bigg|_{\Phi_\uparrow=\Phi_\downarrow=0}, \tag{2.4}$$

where $\Phi_c = (\Phi_\uparrow + \Phi_\downarrow)/2$. Applying second-order perturbation theory in Φ_c to Eq. (2.1) gives[4,13]

$$D_c = \frac{1}{L}\left(\frac{1}{2}\langle -T\rangle - \sum_{\nu\neq 0}\frac{|\langle\nu|J_p|0\rangle|^2}{E_\nu - E_0}\right), \tag{2.5}$$

where $T = -\sum(c^\dagger_{l+1\sigma}c_{l\sigma} + {\rm H.c.})$ is the kinetic-energy operator, $J_p = -i\sum(c^\dagger_{l+1\sigma}c_{l\sigma} - {\rm H.c.})$ is the paramagnetic current operator, $\langle\ \rangle$ denotes the expectation value in the ground state, and all quantities are evaluated at $\Phi_\uparrow = \Phi_\downarrow = 0$.

We next consider threading the Hubbard ring with a time-dependent flux $\Phi_\uparrow(t) = \Phi_\downarrow(t) = \Phi_c(0)\exp(-i\omega t)$, which leads to a uniform time-dependent electric field by Faraday's law. Standard linear-response arguments applied to Eq. (2.1) give the frequency-dependent conductivity at zero temperature,[5,4,13]

$$\tilde\sigma(\omega) = \frac{i}{\omega + i0^+}\left(\frac{\langle -T\rangle}{L} + \frac{2}{L}\sum_{\nu\neq 0}\frac{|\langle\nu|J_p|0\rangle|^2}{\omega - E_\nu + E_0 + i0^+}\right). \tag{2.6}$$

The conductivity thus has the form (1.1) with D_c given by Eq. (2.5). The high-frequency behavior of ${\rm Im}\,[\tilde\sigma(\omega)]$ leads, via the Kramers-Kronig relations which link the real and imaginary parts of $\tilde\sigma(\omega)$, to an f-sum rule for the total optical spectral weight,[5,14]

$$\pi N_{\rm tot} \equiv \int_0^\infty {\rm Re}\,[\tilde\sigma(\omega)]d\omega = \frac{\pi}{2L}\langle -T\rangle. \tag{2.7}$$

The Hellman-Feynman theorem gives $\langle T\rangle = E_0 - U\partial E_0/\partial U$. Thus both $N_{\rm tot}$ and D_c may be obtained as derivatives of the ground-state energy.

The energy and momentum of the eigenstates of Eq. (2.1) can be expressed via a generalization of Bethe's ansatz as[3,4]

$$E = -2\sum_{n=1}^{N} \cos k_n, \tag{2.8}$$

$$P = \sum_{n=1}^{N} k_n, \tag{2.9}$$

where the *pseudomomenta* k_n are a set of distinct numbers determined by[4]

$$Lk_n = 2\pi I_n + \Phi_\downarrow - \sum_{\alpha=1}^{M} 2\tan^{-1}\frac{\sin k_n - \Lambda_\alpha}{U/4}, \tag{2.10}$$

and the *spin rapidities* Λ_α are a set of distinct numbers related to the k_n by[4]

$$\sum_{n=1}^{N} 2\tan^{-1}\frac{\Lambda_\alpha - \sin k_n}{U/4}$$

$$= 2\pi J_\alpha + \Phi_\uparrow - \Phi_\downarrow + \sum_{\beta=1}^{M} 2\tan^{-1}\frac{\Lambda_\alpha - \Lambda_\beta}{U/2}. \tag{2.11}$$

Here $\{I_n; n = 1,\ldots,N\}$ and $\{J_\alpha; \alpha = 1,\ldots,M\}$ are the quantum numbers which specify the state of the system:[15] the I_n describe the charge degrees of freedom[16,17,9] and are distinct integers (half-integers) for M even (odd),[3] while the J_α describe the spin degrees of freedom,[16,17,9] and are integers (half-integers) for $N-M$ odd (even).[3] For the ground state, the quantum numbers are consecutive integers (or half-integers) centered about the origin.[3]

In the following, we solve the Bethe ansatz equations (2.10) and (2.11) analytically and numerically in various limits to obtain D_c and $N_{\rm tot}$ as a function of n, L, and U. The knowledge of the eigenstates of H also allows us to calculate the low-temperature thermopower near half filling.

III. THE CORRELATION LENGTH OF THE MOTT INSULATOR

A. Definition of ξ

The 1D Hubbard model for $n = 1$ and $U > 0$ is known to have an insulating ground state,[3] so that the charge stiffness D_c vanishes for an infinite system. In a previous paper,[7] we showed numerically and via a large-U approximation that for a large finite ring of circumference L at $n = 1$

$$D_c(L) = (-1)^{L/2+1} L^{1/2} D(U) \exp[-L/\xi(U)], \qquad L \to \infty \tag{3.1}$$

(L even), where $D(U)$ is a positive U-dependent number. Equation (3.1) serves to define the correlation length $\xi(U)$ in the Mott insulating phase of the 1D Hubbard model. We have subsequently derived Eq. (3.1) analytically for arbitrary U. The details are given in Appendix A. The result for the correlation length $\xi(U)$ is

$$1/\xi(U) = \frac{4}{U}\int_1^\infty dy \frac{\ln(y+\sqrt{y^2-1})}{\cosh(2\pi y/U)}. \tag{3.2}$$

The result for $D(U)$ is given in Eq. (A28). The limiting behavior of $D(U)$ is

$$D(U) = \begin{cases} (2/\pi\xi)^{1/2}, & U \to 0 \\ AU, & U \to \infty, \end{cases} \tag{3.3}$$

where $A \simeq 0.147\,376$.

The negative sign of D_c when L is a multiple of 4 indicates that the persistent current of the ring is paramagnetic, as discussed in Refs. 7 and 18. Orbital paramagnetism is a generic feature of half-filled single-band $4n$ electron systems, and was first observed many years ago[19] in NMR spectra of [16] annulene and [24] annulene, larger analogs of benzene. The full flux dependence of the ground-state energy of the Mott insulator is

$$E_0(\Phi_c) - E_0(0) = \frac{2D_c(L)}{L}(1-\cos\Phi_c), \qquad L \to \infty. \tag{3.4}$$

By contrast, in the metallic phase at $n \neq 1$, the flux dependence of the ground-state energy in the large-L limit is $E_0(\Phi_c) - E_0(0) = D_c\Phi_c^2/L$ for $|\Phi_c| \ll 2\pi L$, and the periodicity $E_0(\Phi_c) = E_0(\Phi_c + 2\pi)$ required by gauge invariance[20] is restored by level crossings.

B. Relation of ξ to the single-particle Green's function

The zero-temperature equal-time single-particle Green's function is defined as

$$G_{\sigma\sigma'}(|x-x'|) = \langle 0|c^\dagger_{x'\sigma'}c_{x\sigma} + c^\dagger_{x\sigma}c_{x'\sigma'}|0\rangle. \tag{3.5}$$

From Eq. (2.1) it follows that $G_{\sigma\sigma'}(|x-x'|) = G_{\sigma\sigma}(|x-x'|)\delta_{\sigma,\sigma'}$. We wish to show that the Green's function of the Mott insulator has the form (at $L = \infty$)

$$G_{\sigma\sigma}(|x|) \sim \exp(-|x|/\xi) \quad \text{as} \quad |x| \to \infty. \tag{3.6}$$

In order to do so, it is useful first to consider some limiting cases.

The weak-coupling limit of ξ may be obtained from an asymptotic expansion of Eq. (3.2) around $U = 0$. We obtain

$$\lim_{U\to 0}\xi = \frac{\pi}{2}(t/U)^{1/2}e^{2\pi t/U}(1 + U/16\pi t + \cdots)$$

$$= \frac{2t + U/2\pi + \cdots}{\Delta(U,t)}, \tag{3.7}$$

where we have made the hopping matrix element t explicit, and

$$2\Delta(U,t) = \frac{16t^2}{U}\int_1^\infty \frac{(y^2-1)^{1/2}dy}{\sinh(2\pi t y/U)} \tag{3.8}$$

is the Lieb-Wu charge gap.[3,9,16] The low-energy (continuum limit) physics of the 1D Hubbard model has been studied extensively by bosonization techniques.[21–25] In the low-energy sector, the effective Hamiltonian separates into independent terms describing the charge and spin degrees of freedom.[21–25] Near $n = 1$ and in the weak-coupling limit, the Hamiltonian for the charge degrees of

freedom is of the relativistic sine-Gordon form.[21,25] The quantity which plays the role of the speed of light c in the Hamiltonian is given to leading order in U at $n = 1$ by[21,25]

$$c = 2t + U/2\pi + \cdots . \tag{3.9}$$

The elementary objects in the sine-Gordon model are solitons[26] carrying charge e and obeying the relativistic dispersion relation $E_k = (c^2k^2 + \Delta^2)^{1/2}$.[27,28] The *quantum soliton length* $\xi_s = c/\Delta$ defines the characteristic size of the solitons,[29] so we see from Eqs. (3.7) and (3.9) that in the weak-coupling limit the correlation length is equal to the quantum soliton length.

The mapping onto the weak-coupling Hubbard model fixes the number density of solitons as $n_s = 1 - n$.[25] The operator $c_{x\sigma}$ involves creating at least one soliton (plus spin excitations). Thus, if we evaluate Eq. (3.5) at a timelike separation at $n = 1$, we find $G_{\sigma\sigma}(t - t') \sim \exp[-i\Delta(t - t')]$ + terms involving higher energies. Using the Lorentz invariance of the weak-coupling field theory to rotate back to a spacelike separation, we obtain $G_{\sigma\sigma}(|x-x'|) \sim \exp(-\Delta|x-x'|/c)$.[30] The gapless spin degrees of freedom correct this expression only by a power of $|x - x'|$. We have thus established Eq. (3.6) in the weak-coupling limit.

Now let us consider the strong-coupling limit. From Eq. (3.2) it follows[31] that

$$\xi^{-1} = \ln(U/at), \qquad U \to \infty, \tag{3.10}$$

where $a = [\Gamma(1/4)/\sqrt{2\pi}]^4 \simeq 4.377$ (this result was obtained previously by us in Ref. 7). Equations (3.1) and (3.3) then imply that $D_c(L) = AUL^{1/2}(at/U)^L$, where $A \simeq 0.147\,376$. The correlation length may thus be obtained in the large-U limit from an expansion of the ground-state energy in powers of t/U. The leading contribution to $D_c(L)$ comes from the lowest-order term which is flux dependent and is[7]

$$\lim_{U\to\infty} \frac{2D_c(L)}{L}$$

$$= \frac{t^L}{U^{L-1}} \left[\sum_{P\in\mathcal{S}_L} \langle 0_\infty | v_{P_L}^\dagger \prod_{i=1}^{L-1} \left(\mathbf{D}^{-1} v_{P_i}^\dagger \right) |0_\infty\rangle + \text{H.c.} \right], \tag{3.11}$$

where $v_i^\dagger = \sum_\sigma c_{i+1\sigma}^\dagger c_{i\sigma}$, $\mathbf{D}^{-1} = 1/\sum_{i=1}^L n_{i\uparrow} n_{i\downarrow}$, $|0_\infty\rangle = \lim_{U\to\infty} |0\rangle$, and $\mathcal{S}_L$ is the group of permutations of L objects. This expression for D_c is a sum over all processes which transport one unit of charge around the ring in the minimum number of steps L. In the large-U limit, the Green's function $G_{\sigma\sigma}(L-1)$ may be evaluated by expanding the ground state $|0\rangle$ in Eq. (3.5) to $\mathcal{O}((t/U)^{L-1})$. The resulting expression for the leading contribution to $t[G_{\uparrow\uparrow}(L-1)+G_{\downarrow\downarrow}(L-1)]$ is identical to that for $2D_c(L)/L$ in Eq. (3.11), except that $|0_\infty\rangle$ now refers to the infinite system, not the periodic system of length L.[32] One can fix the phase factors such that $G_{\uparrow\uparrow} = G_{\downarrow\downarrow}$. Then the only difference between $tG_{\sigma\sigma}(L-1)$ and $D_c(L)/L$ in the large-U limit stems from the difference between the spin correlations in the infinite Heisenberg model and those in the periodic one of size L, which can at most cause $tG_{\sigma\sigma}(L-1)$ to differ from $D_c(L)/L$ by an algebraic factor in L. The *correlation lengths* defined by the two functions must be the same in the large-U limit.

We have thus established Eq. (3.6) in both the weak- and strong-coupling limits. We now give a physical argument which suggests that G has this asymptotic form for arbitrary U.

C. Singlet pairing for $U < 0$

The modulus $|G_{\sigma\sigma}(|x|)|$ of the equal-time single-particle Green's function is invariant under $U \to -U$ at half filling, as may be seen by applying the transformation (2.2) to Eq. (3.5). Let us therefore consider Eq. (3.6) for the case $U < 0$. The $U < 0$ Hubbard model is a superconductor (or more precisely, in one dimension, has divergent superconducting fluctuations[33]). Superconductors are characterized by a correlation length ξ which one may think of as the size of a Cooper pair, and G decays exponentially with characteristic length ξ. The superconducting correlation length can also be determined via the periodicity of the ground-state energy of a superconducting ring with respect to the magnetic flux enclosed. Gauge invariance implies that E_0 must be periodic in Φ_c with period 2π (one flux quantum);[20] however, in a system with superconducting correlations, the period will be π (half a flux quantum) when the system is large compared to the correlation length.[20] In this subsection, we show that the deviations from a periodicity of half a flux quantum in the ground-state energy of the $U < 0$ 1D Hubbard model scale as $\exp[-L/\xi(|U|)]$ at $n = 1$. We therefore identify $\xi(|U|)$ as the superconducting correlation length; this makes it plausible that G decays exponentially with this length for arbitrary U, as asserted in Eq. (3.6). The analogous electron-hole correlations in the ground state of the Mott insulator were discussed by Krishnamurthy *et al.*[34] in a slightly different context.

The Bethe ansatz description of the ground state of the attractive 1D Hubbard model involves complex pseudomometa,[35] and is unwieldy for a finite system with periodic boundary conditions. However, we can use the identity (2.3) to obtain the ground-state energy of the half-filled attractive 1D Hubbard model as a function of Φ_c from the ground-state energy of the half-filled repulsive 1D Hubbard model as a function of $\Phi_s = (\Phi_\uparrow - \Phi_\downarrow)/2$. The latter may be obtained from a solution of Eqs. (2.10) and (2.11) with the k's and Λ's strictly real (alternatively, one can think of this as a Bethe ansatz for $U < 0$ in a representation where the down-spin electrons are treated as holes). In 1D, the variation of the ground-state energy with respect to flux for $U < 0$ is $E_0(\Phi_c) \simeq E_0(0) + D_c(-|U|)\Phi_c^2/L$ when L is large. However, there are level crossings with states whose energy parabolas have local minima at $\Phi_c = \pm\pi, \pm 2\pi$, etc. Each parabola corresponds to a particular choice of the Bethe ansatz quantum numbers $\{I_n\}$ and $\{J_\alpha\}$ (in the repulsive representation). The quantum numbers $\{I_n^0\}$ and $\{J_\alpha^0\}$ of the ground state at

$\Phi_c = 0$ are given in Eqs. (A1) and (A2) in Appendix A. The ground state at $\Phi_c = 2\pi$ has quantum numbers satisfying $I_n = I_n^0 + 1$, $J_\alpha = J_\alpha^0 - 2$, and is equivalent to the above state, but shifted by one flux quantum. The equivalence of these two states follows from gauge invariance.[20] There is a level crossing with a third, inequivalent state whose energy minimum lies at $\Phi_c = \pi$, and whose quantum numbers satisfy $I_n = I_n^0$, $J_\alpha = J_\alpha^0 - 1$. The energies of these three states as a function of Φ_c are plotted as solid curves in Fig. 1 for $U = -4$ and $N = L = 6$ and 28 [note that $\xi(4) \approx 4.06$]. The ground-state energy for the repulsive case (shifted by $-N|U|/2$) is plotted as a dashed line for comparison. Equations (2.10), (2.11), and (3.4) imply that in the large-L limit the local energy minimum at $\Phi_c = \pi$ differs from that at $\Phi_c = 0$ by

$$E_0(\pi) - E_0(0) = \frac{4D_c(|U|)}{L} \sim L^{-1/2} \exp[-L/\xi(|U|)]. \tag{3.12}$$

It is evident from Fig. 1 that as L increases and the local energy minima at $\Phi_c = 0$ and $\Phi_c = \pi$ become degenerate, the ground-state energy becomes periodic with period π (half a flux quantum).

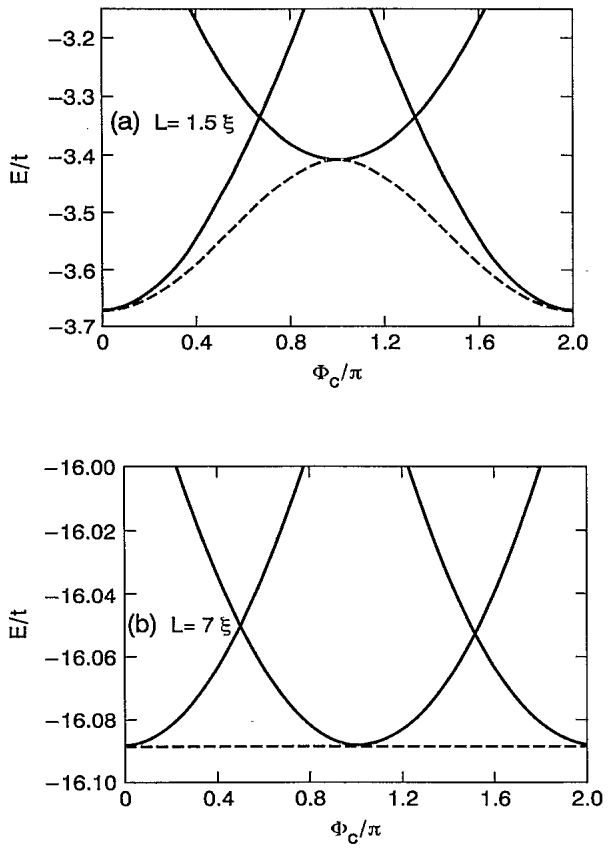

FIG. 1. Ground-state energy of a half-filled Hubbard ring vs magnetic flux $(\hbar c/e)\Phi_c$ from $\Phi_c = 0$ to 2π. Solid curves, $E_0(\Phi_c) + N|U|/2$ for $U = -4$; dashed curves, $E_0(\Phi_c)$ for $U = 4$. (a) $N = L = 6 \approx 1.5\xi$; (b) $N = L = 28 \approx 7\xi$, where $\xi(4) \simeq 4.06$ is the correlation length at half filling for $|U| = 4$.

IV. SCALING BEHAVIOR OF D_C NEAR THE METAL-INSULATOR CRITICAL POINT

We have shown that the charge response of the half-filled ($n = 1$) $U > 0$ 1D Hubbard model may be characterized by a length ξ which diverges as $U \to 0$. In this section, we show that the point $U = 0$, $n = 1$ is a conventional quantum critical point in the sense that the singular behavior of $D_c(n, L, U)$ in the vicinity of this point is given by a (presumably universal[8]) scaling function $Y(\xi|1-n|, \xi/L)$, which we calculate numerically and, in various limits, analytically. Our results confirm the applicability of the hyperscaling ansatz to this system. We also discuss the total optical spectral weight $\pi N_{\rm tot}$.

The hyperscaling ansatz for a $T = 0$ phase transition is that the singular part of the ground-state energy per correlation volume scales as the inverse of the correlation time.[8] Since the correlation time scales as ξ^z,[36] it follows that

$$E_0^{\rm sing}/L^d \sim \xi^{-(d+z)}, \tag{4.1}$$

where d is the spatial dimension and z is the dynamic critical exponent. In our problem $d = 1$, and the Lorentz invariance of the critical point implies $z = 1$. From Eq. (2.4), we see that D_c involves two derivatives of E_0/L with respect to Φ_c/L; the singular part of D_c should therefore have scaling dimension $2 - d - z = 0$. We thus postulate that as $U \to 0$, $n \to 1$, and $L \to \infty$, the singular part of D_c is given by

$$D_c^{\rm sing}(n, L, U) = Y_\pm(\xi\delta, \xi/L), \tag{4.2}$$

where $\delta = |1 - n|$ and $Y_\pm$ are dimensionless scaling functions, universal up to a metric factor fixing the units of length and energy. The subscripts $\pm$ on Y refer to $U > 0$ and $U < 0$. The charge response at $U = 0^-$ is that of noninteracting electrons with nearest-neighbor hopping, so $Y_-(0, 0) = 2/\pi$. One consequence of Eq. (4.2), therefore, is that the charge stiffness at $n = 1$ and $L = \infty$ has a universal jump of $2/\pi$ as U goes from 0^+ to 0^-.

Equations (3.1) and (3.3) imply that

$$Y_+(0, y) = (2/\pi y)^{1/2} \exp(-1/y), \quad y \to 0. \tag{4.3}$$

To verify the scaling law (4.2) for the general case, we have calculated $D_c(n, L, U > 0)$ by solving the Bethe ansatz equations (2.10) and (2.11) numerically for systems with $N \leq 200$ and $L \geq N$. We have considered only systems with an even number of electrons N. In order to obtain for D_c a smooth function of n and L, we impose antiperiodic boundary conditions (i.e., $\Phi_\uparrow = \Phi_\downarrow = \pi$) when $N \bmod 4 = 0$, so that the ground state is always a singlet.[37,38] Because $\delta \equiv |1 - N/L| = 0, 1/L, 2/L, \ldots$, the scaling function $Y(x, y)$ is only defined on a countable set of lines in the x, y plane, namely on $x = my$; $m = 0, 1, 2, \ldots, \infty$. In Fig. 2, we plot πD_c versus ξ/L along the lines $\delta = 0, 1/L, 2/L$, and $4/L$ for systems with $N = 60$, 80, and 100 electrons. In each case, the results for $N = 60$, 80, and 100 fall onto a smooth curve over the entire range $0.3 \leq \xi/L \leq 10\,000$, verifying Eq. (4.2). (The limit $L \gg \xi$ is examined more fully in Sec. V.) For

comparison, Fig. 2(b) is replotted with ξ as the abscissa in Fig. 2(e). It is evident from Fig. 2 that as $\xi/L \to \infty$, $D_c \to 2/\pi$ (the value of the universal jump at the critical point), which we may express as

$$\lim_{y\to\infty} Y_+(x,y) = \frac{2}{\pi}. \tag{4.4}$$

The charge stiffness in the limit $L \to \infty$ ($y \to 0$) has been calculated previously[24,39,40] from a numerical solution of the Bethe ansatz integral equations. Here we merely point out that because Y is universal, one can extract $Y_+(x,0)$ from Haldane's analytical result[29] for the sine-Gordon model at $\beta^2 = 8\pi^-$. The mapping between the two models is uniquely determined by equating the energy gap Δ to the soliton rest energy $m_s c^2$, equating the correlation length ξ to the quantum soliton length, and equating the doping $\delta = |1-n|$ to the mean number of solitons per unit length, as discussed in Sec. III B. The result for the scaling function is[29]

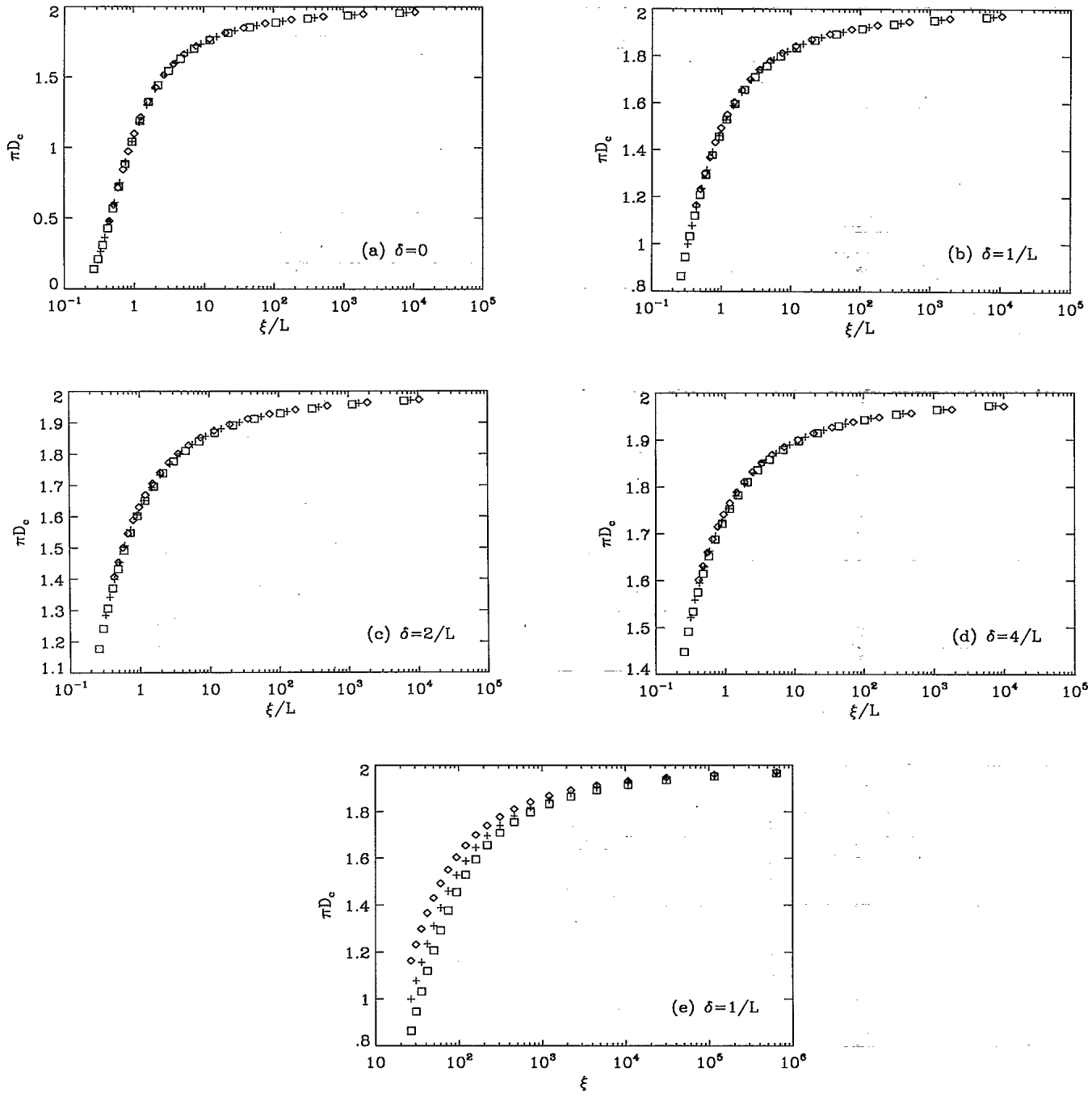

FIG. 2. Plots of πD_c vs ξ/L for (a) $\delta = 0$, (b) $\delta = 1/L$, (c) $\delta = 2/L$, and (d) $\delta = 4/L$ for systems with $N = 60$ (diamonds), $N = 80$ (crosses), and $N = 100$ (squares), and $0.5 \le U \le 2$, illustrating the scaling law, Eq. (4.2). Note the different vertical scales. (e) For comparison, (b) is replotted with ξ as the abscissa.

$$Y_+(x,0) = x\left(1 - \frac{1}{2}(\pi x)^2 + \frac{8\ln 2}{3\pi^2}(\pi x)^3 + \cdots\right), \quad x \to 0, \tag{4.5}$$

$$Y_+(x,0) = \frac{2}{\pi}\left(1 - \frac{1}{2}[\ln(\pi x)]^{-1} + \cdots\right), \quad x \to \infty. \tag{4.6}$$

Near the critical point, therefore, the charge stiffness grows rapidly as one dopes the Mott insulator, saturating near its maximum value $2/\pi$ when $x = \xi\delta \sim 1$. In Sec. VI, we compute the small-x behavior of the scaling function directly from the Bethe ansatz equations, verifying Eq. (4.5).

The total optical spectral weight $\pi N_{\rm tot}$ involves information at high frequencies, but is related via the f-sum rule (2.7) to the expectation value of the kinetic energy in the ground state; thus $N_{\rm tot}$ may also be expected to exhibit universal behavior near the critical point. In the critical regime ξ, L, $\delta^{-1} \gg 1$, we find that

$$N_{\rm tot} \simeq \frac{2}{\pi}, \tag{4.7}$$

approximately independent of δ, L, and ξ. This result was obtained previously for the case $L = \infty$.[24] The approximate δ independence of $N_{\rm tot}$ near the critical point is in agreement with the intuitive picture that when $\xi \gg 1$ the kinetic energy of the system should be little affected by the transition to an insulating state. At $\delta = 0$, there is an energy gap 2Δ for charge excitations, and all the optical spectral weight is at frequencies $\omega \geq 2\Delta$ when $L = \infty$. We refer to the states at $\omega \geq 2\Delta$ as being in the "upper Hubbard band" (UHB) (see Sec. VI). Although for $\delta > 0$ there is no true gap in $\sigma(\omega) \equiv \mathrm{Re}[\tilde{\sigma}(\omega)]$, the structure in the UHB region $\omega \geq 2\Delta$ must persist for sufficiently small δ by continuity. However, the rapid growth of D_c with δ and the approximate δ independence of $N_{\rm tot}$ imply that near the critical point the primary effect of doping on $\sigma(\omega)$ is to transfer spectral weight from the UHB region into the Drude peak at $\omega = 0$, so that the UHB structure in $\sigma(\omega)$ is essentially destroyed for $\delta > \xi^{-1}$. A similar shift of spectral weight from the UHB region into the Drude peak with decreasing system size is implied by Fig. 2, the UHB structure being essentially destroyed for $L < \xi$. In the language of the continuum field theory,[41] the approximate equality $N_{\rm tot} \simeq D_c \simeq 2/\pi$, which holds in the weak-coupling limit both for $\xi\delta \gg 1$ and for $\xi \gg L$, arises because the effects of umklapp scattering are largely smoothed out in those limits, so that $\tilde{\sigma}(\omega)$ approaches the result expected in the absence of umklapp scattering,[25] with $N_{\rm tot} = D_c = (2/\pi)\sin(n\pi/2)$.

V. FINITE-SIZE CORRECTIONS TO THE OPTICAL SPECTRAL WEIGHTS

In this section, we calculate the finite-size corrections (for $L > \xi$) to the charge stiffness D_c and the total optical spectral weight $\pi N_{\rm tot}$ in the metallic phase of the $U > 0$ 1D Hubbard model. Since the scaling function for the charge stiffness is expected to be universal,[8] the finite-size corrections we obtain should be relevant for other, nonsoluble models as well.

We find that the finite-size corrections to both D_c and $N_{\rm tot}$ are positive and depend smoothly on L, provided the boundary conditions are chosen so that the ground state with an even number of electrons is a singlet: periodic boundary conditions when $N \bmod 4 = 2$, antiperiodic when $N \bmod 4 = 0$.[42] We employ these boundary conditions throughout this section. In Fig. 3, D_c and $N_{\rm tot}$ are displayed as a function of L for $\delta = 0.2$ and $U = 6$ ($\xi \approx 2.04$). Note that the finite-size corrections to D_c are larger than those to $N_{\rm tot}$; this is a generic feature of the finite-size corrections and indicates that the finite-size enhancement of D_c is in part due to a transfer of spectral weight from higher frequencies, as discussed above. We find that the leading finite-size corrections to both D_c and $N_{\rm tot}$ are $\mathcal{O}(L^{-2})$ in the metallic phase,[43] in accordance with the system-size dependence of the ground-state energy.[44] The solid curves in Fig. 3 are extrapolations of the L^{-2} behavior to small L, indicating that the asymptotic form of the finite-size corrections gives a reasonable approximation even for fairly small systems, provided $L > \xi$.

The coefficients $A(\delta,U)$ and $B(\delta,U)$ of the leading finite-size corrections in the metallic phase, defined by $D_c(L) = D_c(\infty)[1 + A(\delta,U)/L^2 + \cdots]$ and $N_{\rm tot}(L) = N_{\rm tot}(\infty)[1 + B(\delta,U)/L^2 + \cdots]$ are plotted versus U for several values of δ in Fig. 4 (note the different vertical scales). Both for $U \to 0$ and far from the critical point, $A(\delta,U)$, $B(\delta,U) \to \pi^2/6$, the value expected for noninteracting fermions (with or without spin) with nearest-neighbor hopping. While the relative finite-size corrections to the total optical spectral weight are always of order $\pi^2/6L^2$, the relative finite-size corrections to D_c can be much larger, particularly near the critical point. The curves in Fig. 4(a) display increasingly sharp maxima as δ is decreased, which occur at values of U such that $\xi(U)\delta \approx 0.4$, and at which point $A(\delta,U) \approx 0.8\xi(U)^2$.

In the critical regime, we expect the leading finite-size

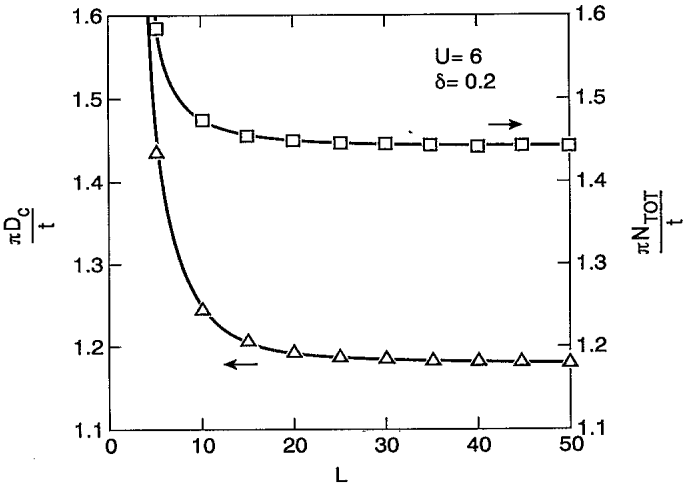

FIG. 3. Total spectral weight $\pi N_{\rm tot}$ (squares) and Drude weight πD_c (triangles) as a function of system size L for $U = 6$ and $\delta = 0.2$. The solid curves indicate L^{-2} behavior extrapolated to small L.

corrections to D_c to scale as

$$D_c(L) = D_c(\infty)[1 + \pi^2/6L^2 + f_+(\xi\delta)(\xi/L)^2 + \cdots], \tag{5.1}$$

where $f_+(x) = d^2Y_+(x,y)/dy^2|_{y=0}/2Y_+(x,0)$, and we have included the regular term $\pi^2/6L^2$. The small-x behavior of $f_+(x)$ can be obtained by noting that to $\mathcal{O}(\xi\delta)^3$, the charge stiffness in the weak-coupling limit is equivalent to that of noninteracting relativistic spinless fermions[29] [c.f. Eq. (4.5)], whence

$$f_+(x) = \frac{\pi^2}{2} - \pi^4 x^2 + \cdots, \quad x \to 0. \tag{5.2}$$

In Fig. 5, the function $f_+ \equiv [A(\delta,U) - \pi^2/6]/\xi(U)^2$ is plotted versus $\xi(U)\delta$; the solid curve was obtained by fixing δ and varying U, while the triangles were obtained by fixing U and varying δ. Note that the numerical data connect smoothly to Eq. (5.2) (dotted curve) as $\xi\delta \to 0$. Figure 5 provides further proof of the scaling law (4.2) in the regime $L \gg \xi$.

Though the finite-size corrections to D_c near the critical point are much larger than expected in a noninteracting system, they should not be a great obstacle to obtaining information about optical properties of 1D models from numerical calculations on small lattices, except at $\delta = 0$, due to the impossibility of simultaneously having small L and small but finite δ ($\min\delta = 1/L$). Roughly speaking, the maximum value of the L^{-2} correction in the critical regime is

$$\frac{D_c(L) - D_c(\infty)}{D_c(\infty)} \approx \frac{0.13}{(\delta L)^2}, \tag{5.3}$$

though of course we have seen in Fig. 3 that the additional term $\pi^2/6L^2$ can lead to slightly larger corrections ($\sim$20%) for small systems.

A further comment on the choice of boundary conditions is appropriate at this point. One practice which has been employed to reduce finite-size effects is to average over periodic and antiperiodic boundary conditions. While this may result in cancellations for some values of n and U due to the fact that the finite-size corrections sometimes have different signs, it is ineffective near the critical point, where the finite-size corrections are largest, since the corrections are always positive and essentially independent of boundary conditions in that regime.[45] Furthermore, depending on U and n, any of the three possibilities—either periodic or antiperiodic boundary conditions or an average over the two—can result in the smallest finite-size corrections. One is better off choosing boundary conditions such that the ground state with an even number of electrons is a singlet; then the finite-size corrections to D_c and $N_{\rm tot}$ are always positive, with magnitudes as given above.

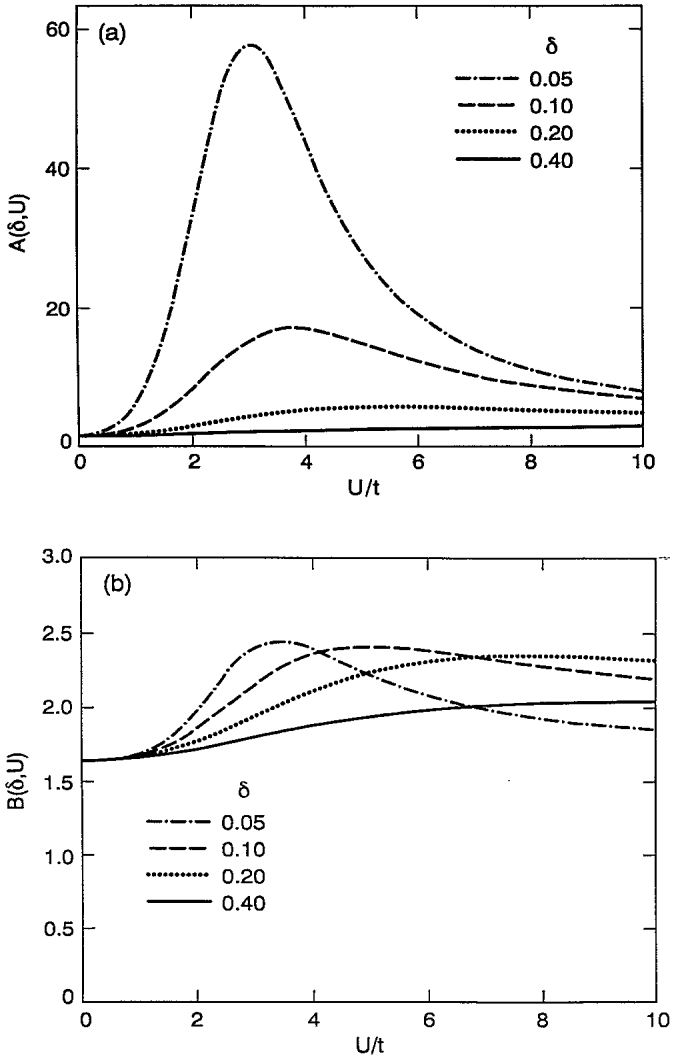

FIG. 4. Plots of (a) $A(\delta,U)$ and (b) $B(\delta,U)$ as a function of U for $\delta = 0.05$, 0.1, 0.2, and 0.4 (top to bottom), where $D_c(L) = D_c(\infty)[1 + A(\delta,U)/L^2 + \cdots]$ and $N_{\rm tot}(L) = N_{\rm tot}(\infty)[1 + B(\delta,U)/L^2 + \cdots]$. Note the different vertical scales.

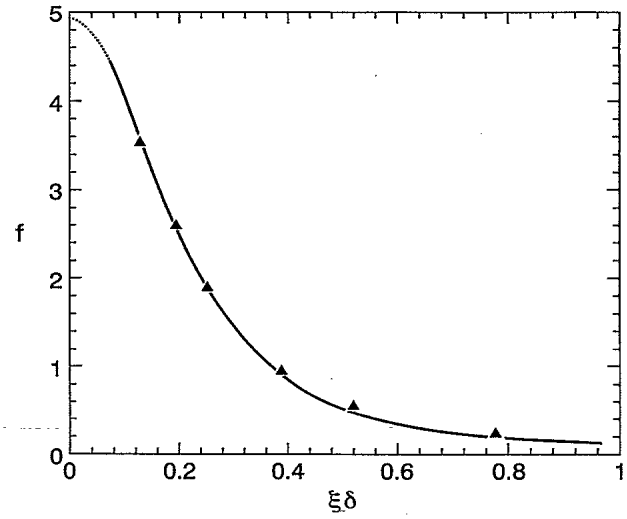

FIG. 5. Plot of the function $f_+ \equiv [A(\delta,U) - \pi^2/6]/\xi(U)^2$ vs $\xi(U)\delta$. Solid curve: $\delta^{-1} = 41$; $1.8 \le U \le 4.7$. Triangles: $U = 3$; $\delta^{-1} = 10,15,20,30,40,60$. Dotted curve: $f_+(x) = \pi^2/2 - \pi^4 x^2$, the analytic result in the small-x limit.

VI. CHARGE EXCITATIONS AND THERMOPOWER NEAR THE METAL-INSULATOR TRANSITION

The mapping of the charge sector of the weak-coupling 1D Hubbard model near $n = 1$ onto the sine-Gordon model suggests that the charge excitations near half filling can be described in terms of the states of $L-N$ solitonlike charge carriers which behave like weakly interacting spinless fermions in the low-density limit $L-N \ll L$. In this section, we use a reformulation of the Bethe ansatz equations due to Woynarovich[9] to construct an explicit mapping for arbitrary U and $\xi\delta \ll 1$ of the low-lying charge degrees of freedom of the 1D repulsive-interaction Hubbard model onto spinless fermions whose mutual interactions vanish as $\xi\delta \to 0$, and which obey a quadratic energy-momentum dispersion relation with a U-dependent effective mass. We use this mapping to obtain asymptotically exact expressions for the charge stiffness and low-temperature thermopower near the metal-insulator transition. Our calculation of the thermopower extends and makes more rigorous previous work of Schulz[10] which was based on a weak-coupling approximation and implies that the transport is holelike for $0 < 1-n \ll \xi^{-1}$.

The energy eigenstates of a $U > 0$ Hubbard chain of length L containing $L-H$ electrons can be obtained by solving Eqs. (2.10) and (2.11), with an appropriate choice of charge quantum numbers $\{I_n\}$ and spin quantum numbers $\{J_\alpha\}$. The charge quantum numbers of the ground state are $L-H$ consecutive integers (or half-integers), centered about the origin, and the spin quantum numbers of the ground state are $(L-H)/2$ consecutive integers (or half-integers), also centered about the origin (as discussed in Sec. II). We wish to consider the low-lying charge excitations of the system, so we restrict our attention to those states whose spin quantum numbers $\{J_\alpha\}$ are those of the ground state, and in which all of the pseudomomenta are real (states with complex pseudomomenta are separated from these by an energy gap[9]). A hole I_h in the distribution of the charge quantum numbers (i.e., $I_{n+1} - I_n = 2$, $I_h \equiv I_n + 1$) corresponds to a charge excitation which, following Anderson,[46] we refer to as a *holon*. We wish to consider the case $H \ll L$; since the I_n are only defined mod(L), we can characterize the low-lying charge excitations of the system by the H holes $\{I_h; h = 1, \ldots, H\}$ in the distribution of the charge quantum numbers rather than working with the larger set $\{I_n\}$. Formally, we shall consider these holes in the distribution of the charge quantum numbers to exist even in the ground state of the system, where they will be consecutive integers (or half-integers) centered about $L/2$ (for convenience, we restrict I_h to the region $1 \le I_h \le L$). We denote the corresponding holes in the distribution of the pseudomomenta by the set $\{k_h\}$. Woynarovich[9] has derived a reduced form of the Bethe ansatz equations, valid in the large-L limit, which deals only with the parameters $\{I_h\}$ and $\{k_h\}$ of these excitations rather than the parameters of all the electrons. We extend the results of Ref. 9 to include a magnetic flux $\Phi_\uparrow = \Phi_\downarrow = \Phi_c$ through the ring, and point out that the reduced set of Bethe ansatz equations derived in Ref. 9 can be interpreted as an asymptotic Bethe ansatz for *holons*, which consequently is expected to become exact in the limit $L \to \infty$ for arbitrary H. We use this asymptotic Bethe ansatz to investigate in detail the limit $H, L \to \infty$, with $\delta = H/L$ small but finite.

In the large-L limit, the energy of such a state is given in terms of the set $\{k_h\}$ by[9]

$$E(L-H) = E_0(L) - \sum_{h=1}^{H} \varepsilon_c(k_h), \tag{6.1}$$

where $E_0(L)$ is the ground-state energy at half filling,[3]

$$E_0(L) = -4L \int_0^\infty \frac{J_0(\omega)J_1(\omega)\,d\omega}{\omega[1+\exp(\omega U/2)]}, \tag{6.2}$$

and[9]

$$\varepsilon_c(k) = -2\cos k - 4\int_0^\infty \frac{J_1(\omega)\cos(\omega \sin k)\,d\omega}{\omega[1+\exp(\omega U/2)]}. \tag{6.3}$$

The momentum of the state, defined mod(2π), is[9]

$$P = \Phi_c - \sum_{h=1}^{H} p_c(k_h) + \pi(L - N/2 + 1), \tag{6.4}$$

where

$$p_c(k) = k + 2\int_0^\infty \frac{J_0(\omega)\sin(\omega \sin k)\,d\omega}{\omega[1+\exp(\omega U/2)]}. \tag{6.5}$$

Following Anderson,[46] we interpret $-\varepsilon_c(k_h)$ and $-p_c(k_h)$ as the energy and momentum of a *holon*. The holons cannot be regarded as noninteracting quasiparticles, however, since the k_h are not free parameters, but are related to the set $\{I_h\}$ by the equations[9]

$$Lp_c(k_h) = 2\pi I_h + \Phi_c + \sum_{h'=1}^{H} \Theta(k_h, k_{h'}), \tag{6.6}$$

where

$$\Theta(k,k') = 2\int_0^\infty \frac{\sin[\omega(\sin k - \sin k')]\,d\omega}{\omega[1+\exp(\omega U/2)]}. \tag{6.7}$$

We point out that the set of equations (6.6) is formally equivalent to an *asymptotic Bethe ansatz* for the holons, $\Theta(k,k')$ being the effective holon-holon scattering phase shift.[47] An asymptotic Bethe ansatz[35] can be used to obtain the asymptotic form of the many-body wave function of a 1D system with nondiffractive scattering in the limit where the particles are widely separated—even if the exact wave function of the system is not of Bethe ansatz form. Moreover, the asymptotic form of the wave function is sufficient to determine the energy eigenvalues of the system *exactly* in the limit $L \to \infty$, even at finite particle density.[35] Consequently, we expect Eqs. (6.1), (6.4), and (6.6) to become exact in the limit $L \to \infty$ for arbitrary H. We emphasize that this conclusion follows *a posteriori* from the form of the equations and is by no means obvious from the method[9] of their derivation. As a test of this conjecture, we use Eqs. (6.1) and (6.6) to calculate the energy of the empty lattice, which in this

formalism is treated as the ground state of a system of L holons. In the limit $H = L \to \infty$, Eq. (6.6) implies that the holon pseudomomenta $\{k_h\}$ are equally spaced with density $L/2\pi$, so that the sum over k_h in Eq. (6.1) can be replaced by $L\int_{-\pi}^{\pi} dk/2\pi$. The resulting integral is equal to $E_0(L)$, yielding $E = 0$, as expected for an empty lattice.

Equations (6.3) and (6.5) implicitly define an energy band $\varepsilon_c(k(p))$ for charge excitations, $k(p)$ being the inverse of the function $p_c(k)$ defined in Eq. (6.5). The holons are just holes in this energy band, which is full at $N = L$. Since $d\varepsilon_c(k(p))/dp|_{p=\pi} = d^3\varepsilon_c(k(p))/dp^3|_{p=\pi} = 0$, the holon energy near the zone boundary $p = k = \pi$ can be written

$$\varepsilon_c(k(p)) = \mu_- - \frac{(p-\pi)^2}{2|m^*|} + \frac{\lambda(p-\pi)^4}{4!} + \cdots, \tag{6.8}$$

where $\mu_- = \varepsilon_c(\pi)$ is the chemical potential in the limit $n \to 1^-$,[3]

$$\frac{1}{m^*} \equiv \left.\frac{d^2\varepsilon_c(k(p))}{dp^2}\right|_{p=\pi} = \frac{\varepsilon_c''(\pi)}{[p_c'(\pi)]^2} < 0 \tag{6.9}$$

and

$$\lambda \equiv \left.\frac{d^4\varepsilon_c(k(p))}{dp^4}\right|_{p=\pi} = \frac{\varepsilon_c^{(4)}(\pi)}{[p_c'(\pi)]^4} - \frac{4\varepsilon_c''(\pi)p_c^{(3)}(\pi)}{[p_c'(\pi)]^5}. \tag{6.10}$$

Equation (6.9) defines the effective mass m^* of the holons near the zone boundary,[48–50] which is negative, as is appropriate for a holelike charge carrier. Explicit expressions for $p_c'(\pi)$, $\varepsilon_c''(\pi)$, $p_c^{(3)}(\pi)$, and $\varepsilon_c^{(4)}(\pi)$ are given in Eqs. (B9)–(B12) in Appendix B.

Equation (6.6) implies that the holon momenta differ from those of noninteracting spinless fermions by a term which vanishes as $\delta \to 0$; we write $p_h \equiv p_c(k_h) = 2\pi I_h/L + \Phi_c/L + \delta p_h$, where $\delta p_h = L^{-1}\sum_{h'=1}^{H}\Theta(k_h, k_{h'})$. Equation (6.7) implies that

$$\lim_{k_h, k_{h'} \to \pi} \Theta(k_h, k_{h'}) = -\frac{4\ln 2}{Up_c'(\pi)}\left[p_c(k_h) - p_c(k_{h'})\right]. \tag{6.11}$$

For small $\delta = H/L$ and low temperatures [$k_BT \ll E_F \simeq (\pi\delta)^2/2|m^*|$], we can use this expression for Θ to calculate δp_h: since the thermal average of $2\pi I_{h'}/L$ is π, the thermal average of δp_h (for fixed I_h) is

$$\langle\langle \delta p_h \rangle\rangle = -\frac{4\ln 2\,\delta}{Up_c'(\pi)}\left(\frac{2\pi I_h}{L} - \pi\right) + \mathcal{O}(\delta^4). \tag{6.12}$$

Inserting this result into Eq. (6.8), one finds that the shift of the holon momentum $-\delta p_h$ caused by holon-holon scattering leads to a fractional shift in the holon energy, measured relative to the chemical potential at half filling, given by

$$\frac{-\delta\varepsilon_c}{\mu_- - \varepsilon_c} \simeq -\frac{8\ln 2\,\delta}{Up_c'(\pi)} = \begin{cases} -4\ln 2\,\xi\delta/\pi, & U \to 0 \\ -8\ln 2\,\delta/U, & U \to \infty. \end{cases} \tag{6.13}$$

The effective holon-holon interaction, which should be considered as a correction to the hard-core repulsion of free spinless fermions, is thus attractive, and is negligible for $\xi\delta \ll 1$ (or $U \gg t$). While the explicit form of Eq. (6.13) is only valid for $\xi\delta \ll 1$ and small excitation energies, one can show quite generally that $|\delta p_h| < 8\ln 2\,\delta/U$, so that holon-holon interactions are also negligible at high temperatures when $\delta/U \ll 1$.

The ground state of the system is obtained by choosing the set $\{I_h\}$ to be consecutive integers (or half-integers) centered about $L/2$, i.e., by placing the holons near the energy minimum at $p = \pi$ (with our convention for I_h, p_h is restricted to the interval $0 \le p_h \le 2\pi$). Since $\delta p_h/p_h \sim \mathcal{O}(\xi\delta)$ as $\delta \to 0$, the holon momenta are approximately equally spaced in the ground state when $\xi\delta \ll 1$. We can thus use Eqs. (6.1), (6.6), and (6.8) to obtain an analytic expression for the charge stiffness near half filling,[51]

$$D_c = \frac{\delta}{2|m^*|} - \frac{\pi^2\lambda\delta^3}{12} + \frac{2\pi^2\ln 2\,\delta^4}{3Up_c'(\pi)}\left(\lambda + \frac{p_c'(\pi) + p_c^{(3)}(\pi)}{|m^*|[p_c'(\pi)]^3}\right) + \mathcal{O}(\delta^5). \tag{6.14}$$

To $\mathcal{O}(\delta^3)$, the charge stiffness is equivalent to that of noninteracting spinless fermions in the energy band (6.8), reflecting the approximate Galilean invariance which holds at low holon densities. The term proportional to δ^4 comes from holon-holon interactions. The term linear in δ in Eq. (6.14) may also be obtained by a more conventional technique, which we describe in Appendix B. Equation (6.14) is consistent with previous exact numerical calculations of the charge stiffness.[24,39,40]

It is instructive to consider the weak- and strong-coupling limits of Eq. (6.14). Making an asymptotic expansion of the U-dependent coefficients in Eq. (6.14) about $U = 0$, and using Eqs. (3.7) and (3.9), we obtain

$$\lim_{U\to 0} D_c = \frac{c\xi\delta}{2}\left(1 - \frac{1}{2}(\pi\xi\delta)^2 + \frac{8\ln 2}{3\pi^2}(\pi\xi\delta)^3 + \mathcal{O}(\xi\delta)^4\right), \tag{6.15}$$

which is equivalent to Eq. (4.5), apart from the nonsingular multiplicative factor $c/2t = 1 + U/4\pi t$. Note that the weak-coupling limit of the holon effective mass can be written as $\lim_{U\to 0}|m^*| = \Delta/c^2$, which is just the rest mass of the soliton in the equivalent sine-Gordon model. In the limit $U \to \infty$, Eq. (6.14) becomes

$$\lim_{U\to\infty} D_c = \frac{t}{\pi}\left(\pi\delta - \frac{(\pi\delta)^3}{3!} + \mathcal{O}(\delta^5)\right), \tag{6.16}$$

where we have made the hopping matrix element t explicit. This is just the Taylor series for $(t/\pi)\sin\pi\delta$, the charge stiffness of noninteracting spinless fermions with dispersion $\varepsilon_c(p) = -2t\cos p$, in accord with the well-known behavior of the holons in the limit $U \gg t$.[9,37]

The low-lying charge excitation eigenstates of a Hubbard ring with $L + H'$ electrons follow from the above results and the particle-hole symmetry present at half filling.[9] The energy and momentum of such a state are

given by Eqs. (6.1) and (6.4) with $-\varepsilon_c(k_h) \to U - \varepsilon_c(k_h)$, and the k_h are determined by Eq. (6.6) with $\Phi_c \to -\Phi_c$. For $H' \ll L$, and for small excitation energies, these states can be described as the states of H' noninteracting spinless fermions (antiholons) near the bottom of the parabolic energy band,

$$\varepsilon_c^{\rm UHB}(p) = U - \varepsilon_c(k(p)) = U - \mu_- + \frac{(p-\pi)^2}{2|m^*|} + \cdots . \tag{6.17}$$

The antiholons are thus characterized by a positive effective mass near $p = \pi$. Note that the gap between the maximum of the lower Hubbard band (6.8) and the minimum of the upper Hubbard band (6.17) is $U - 2\mu_- = 2\Delta$, the Lieb-Wu charge gap.[3,9] The result (6.14) also holds for $n = 1 + \delta$. We can thus write

$$\frac{\partial D_c}{\partial n} = \frac{1}{2m^*}, \quad |1-n| \ll 1, \tag{6.18}$$

where it is understood that $m^* > 0$ (< 0) for $n > 1$ (< 1). This expression emphasizes that the conductivity is "holelike" when $m^* < 0$, in that the Drude weight *decreases* when an electron is added to the system, and conversely, that it is "electronlike" when $m^* > 0$.[52]

The physical significance of the effective mass m^* is revealed in the dynamics of a wave packet. The *charge velocity* v_c is defined as the the group velocity of a wave packet composed of low-energy charge excitations. For $\delta \ll 1$, we obtain

$$v_c = \left. \frac{d\varepsilon_c(k(p))}{dp} \right|_{p=\pi(1-\delta)} \simeq \frac{\pi\delta}{|m^*|}. \tag{6.19}$$

If we apply an electric field $\mathbf{E}$ to the system, the wave packet will be accelerated:[53]

$$\frac{dv_c}{dt} = \frac{d^2\varepsilon_c}{dp^2}\left(\frac{dp}{dt}\right) = \frac{1}{m^*}(-e\mathbf{E}). \tag{6.20}$$

The holons, having $m^* < 0$, will be accelerated parallel to an applied electric field, and are thus holelike, while antiholons are electronlike.[54]

The transport coefficients which are conventionally used to determine the sign of the carriers—the thermopower and the Hall coefficient—couple to electrons, not holons, however, so the above picture is not the whole story. In 1D, only the thermopower (Seebeck coefficient) S is available, which is defined in terms of the open circuit electric field $\mathbf{E}$ produced by a temperature gradient ∇T across the sample:

$$\mathbf{E} = S\nabla T. \tag{6.21}$$

S is the entropy carried per unit charge by an electric current,[55] and is ordinarily negative for electron conduction and positive for holes. In general, both the charge and spin entropies will contribute to S. However, the holon density of states diverges ($v_c \to 0$) as $\delta \to 0$. Close to the metal-insulator transition, therefore, the entropy of the holons will be much greater than the spin entropy at low temperatures and will dominate the thermopower.[10,56] Furthermore, in this limit the holons can be treated as noninteracting spinless fermions in the energy bands (6.8) and (6.17). So far, we have proven this only for the case where the spin wave function is in its ground state and there are no charge excitations into the upper Hubbard band [i.e., no excitations from the band (6.8) into the band (6.17)]. Such charge excitations correspond to umklapp processes, and lead to a finite conductivity for $T > 0$;[57] however, their effect on the thermopower is negligible when $k_B T \ll 2\Delta$. The interaction of the holons with a thermal population of spin excitations can be described by adding a term $L\,\delta p_h^s(T)$ to the right-hand side of Eq. (6.6); in Ref. 56, we show that for small δ and low temperatures $\delta p_h^s(T) \sim \mathcal{O}(T\delta)$, so the holons are still effectively noninteracting when δ and T are small. A standard formula for the low-temperature thermopower[58] can therefore be used to obtain

$$S = -\frac{k_B^2 T}{3|e|}\frac{m^*}{\delta^2}, \tag{6.22}$$

which is valid for $\xi\delta \ll 1$ and $k_B T \ll E_F \simeq (\pi\delta)^2/2|m^*|$ in the absence of impurity scattering. The low-temperature thermopower thus becomes large and positive as the metal-insulator transition is approached from $n < 1$ and has the opposite sign for $n > 1$. The small-U behavior of Eq. (6.22) is in agreement with an earlier result[10] obtained using a weak-coupling approximation. The entropy carried by spin excitations is important at higher temperatures and/or dopings, and is discussed in Ref. 56.

VII. CONCLUSIONS

We have obtained an exact result for the charge stiffness of a Hubbard ring with $U > 0$ and $n = 1$ in the large-circumference limit, which defines and yields an analytic expression for the correlation length $\xi(U)$ in the Mott insulating phase of the 1D Hubbard model. We have shown that this correlation length also governs the exponential decay of the equal-time single-particle Green's function at $n = 1$ in both the weak- and strong-coupling limits, as well as the pairing correlations in the ground state of the attractive 1D Hubbard model at $n = 1$. We remark that the strong-coupling expansion [Eq. (3.11)], which gave $D_c(L)$, $G_{\sigma\sigma}(L) \sim \exp(-L/\xi)$, with $\xi^{-1} = \ln(U/t) + \text{const}$, must hold for the half-filled Hubbard model in the large-U limit in any spatial dimension.

In the vicinity of the zero-temperature critical point $U = 0$, $n = 1$, we have shown that the doping and system-size dependence of the charge stiffness scale with the correlation length ξ. The scaling function for the charge stiffness is expected to be universal,[8] and thus should be applicable to other 1D metal-insulator transitions. In addition, this scaling function appears to describe a certain class of 1D magnetic phase transitions which do not involve broken symmetry and which are characterized by a transition from a state with algebraic correlations to a state with a gap. In particular, Eqs. (2.3) and (4.2) imply

that the spin stiffness $D_s = (L/2)d^2E_0/d\Phi_s^2|_{\Phi_s=0}$ [where $\Phi_s = (\Phi_\uparrow - \Phi_\downarrow)/2$] of the 1D attractive-interaction Hubbard model with $n = 1$ and magnetization $\sigma = 2S^z/L$ has the form

$$D_s(S^z, L, -U) = Y_\pm(\xi\sigma, \xi/L) \tag{7.1}$$

in the vicinity of the magnetic critical point $U = 0$, $\sigma = 0$. The spin stiffness is related to the magnetic susceptibility χ by $D_s = (1/2\pi^2)\chi^{-1}$ (see Ref. 4). There is a similar mapping between the metal-insulator transition occurring at $n = 1/2$ for 1D lattice spinless fermions with nearest-neighbor repulsion and the magnetic transition occurring at the isotropic point in the antiferromagnetic Heisenberg-Ising spin chain. Because of the universality of the scaling function for the metal-insulator transition, this magnetic transition is also expected to be described by the scaling function Y.

One important consequence of scaling is that the finite-size corrections to D_c are enhanced near the critical point. While the scaling form (5.1) for the finite-size corrections can only be expected to hold for 1D systems, it is plausible that the finite-size corrections to D_c may be enhanced near half filling for smaller values of U in higher-dimensional systems as well, suggesting that caution is required in interpreting numerical calculations of $\tilde{\sigma}(\omega)$ on small clusters in this regime. Enhanced finite-size effects of this type could explain anamolous negative values for the Drude weight of the 2D Hubbard model on a 4×4 lattice with 14 electrons when $U = 4$ and 8.[59] We do not expect similar enhancements of the finite-size corrections in the $t - J$ model, since the constraint of no double occupancy fixes $\xi = 0$.

In the metallic phase of the model, the physical significance of the correlation length is that it defines the characteristic size of the charge-carrying solitons, or *holons*. We have shown that a reformulation[9] of the Bethe ansatz equations of the 1D Hubbard model in terms of the parameters of the charge excitations only is formally equivalent to an asymptotic Bethe ansatz for holons, and have used this asymptotic Bethe ansatz to show explicitly for arbitrary U that the holons are equivalent to spinless fermions with mutual interactions that vanish in the low-density limit $\xi\delta \to 0$. We have used this mapping to obtain an expression for the low-temperature thermopower near the metal-insulator transition, which implies hole-like transport for $0 < 1 - n \ll \xi^{-1}$.

It has been argued in a series of papers by Anderson[60] that the physics of doped Mott insulators is similar in one and two dimensions. It is therefore interesting to compare the metal-insulator transition in the 1D Hubbard model with that observed in the cuprate materials in which high-temperature superconductivity occurs, which are widely regarded to be quasi-two-dimensional doped Mott insulators.[2] The thermopower of the doped cuprates is generically positive for hole doping[61–63] and negative for electron doping,[64–66] with a magnitude which increases drastically as the nominal concentration of doped carriers goes to zero, in qualitative agreement with Eq. (6.22). However, linear-T thermopower has not been clearly identified in metallic samples.[67] For larger dopings, the thermopower has an unusual temperature dependence,[61–66,68,69] and will be discussed in more detail elsewhere.[56]

The optical conductivity near the metal-insulator transition in the doped cuprates[70–72] also shows qualitative similarities to that near the Mott-Hubbard metal-insulator transition in 1D.[24,39,50] In particular, a linear growth of the Drude weight with doping is observed in the cuprates for small dopings,[70–72] while the total integrated spectral weight in the optical conductivity up to 3 eV, which includes both the low-frequency response and the charge-transfer band (or upper Hubbard band) is roughly constant from the insulating antiferromagnet all the way through the superconducting phase.[70,71] The primary effect of doping in these compounds is thus to transfer spectral weight from high to low frequencies. This transfer of spectral weight is quite rapid, being essentially complete for a doping of approximately 0.25.[70,71] Similarly, near the critical point of the Mott-Hubbard metal-insulator transition in 1D, we find that the total optical spectral weight $\pi N_{\rm tot}$ is approximately independent of doping, while the Drude weight πD_c grows linearly for small dopings, the transfer of spectral weight from high to low frequencies being essentially complete when $\xi\delta \sim 1$. Of course, the fraction of the low-frequency spectral weight which is collapsed into the Drude peak at $\omega = 0$ is much greater in our 1D model for several reasons, among which are the absence of an impurity potential and the kinematic constraints which limit carrier-carrier scattering in 1D.

Note added in proof. After this manuscript was accepted we became aware of work by M. Continentino [Phys. Rev. B **45**, 11 312 (1992)] in which it is argued that the charge stiffness ought, in general, to obey scaling near a second-order Mott transition, but that in the particular case $d = 1$, $n = 1$, $U \to 0$ scaling does not apply. The work presented here proves that the charge stiffness scales in the expected way near the Mott transition in $d = 1$, even in the case $n = 1$, $U \to 0$. The exponents obtained here agree with those of Continentino; the definition of ξ and the calculation of the scaling function presented here are different.

ACKNOWLEDGMENTS

C.A.S. acknowledges support from AT&T Bell Laboratories and NSF Grants Nos. DMR-91-04873 (at Princeton) and DMR-91-23577 (at Maryland), and thanks P. W. Anderson, T. Giamarchi, F. D. M. Haldane, and B. S. Shastry for valuable discussions.

APPENDIX A: CALCULATION OF $\xi(U)$

Here we derive the result (3.1) for the asymptotic form of $D_c(L)$ as $L \to \infty$ at $n = 1$ using a technique[73] previously employed for computing finite-size corrections to energy eigenvalues in Bethe ansatz solvable models.

We use Eq. (2.4) to calculate the charge stiffness from the dependence of the ground state energy on Φ_c. $E_0(\Phi_c)$ may be obtained by solving the Bethe ansatz equations (2.10) and (2.11), and using Eq. (2.8). We consider only even N $(= L)$, so that the ground state is a singlet $(M =$

$N/2$) and nondegenerate. The charge quantum numbers are then

$$I_n = \left\{-\frac{L}{2}+1-s, -\frac{L}{2}+2-s, \ldots, \frac{L}{2}-s\right\}, \tag{A1}$$

where $s = \frac{1}{4}(L \bmod 4)$, and the spin quantum numbers are

$$J_\alpha = \left\{-\frac{M-1}{2}, -\frac{M-3}{2}, \ldots, \frac{M-1}{2}\right\}. \tag{A2}$$

In order to facilitate a solution of Eqs. (2.10) and (2.11) in the large-L limit, let us define the following functions:[73]

$$p_L(k) \equiv k - \frac{1}{L}\sum_\beta 2\tan^{-1}\frac{\Lambda_\beta - \sin k}{U/4}, \tag{A3}$$

$$z_L(\Lambda) \equiv \frac{1}{L}\sum_n 2\tan^{-1}\frac{\Lambda - \sin k_n}{U/4} - \frac{1}{L}\sum_\beta 2\tan^{-1}\frac{\Lambda-\Lambda_\beta}{U/2}. \tag{A4}$$

With these definitions, the Bethe ansatz equations (2.10) and (2.11) become, for $\Phi_\uparrow = \Phi_\downarrow = \Phi_c$,

$$p_L(k_n) = \frac{2\pi I_n}{L} + \frac{\Phi_c}{L}, \tag{A5}$$

$$z_L(\Lambda_\alpha) = \frac{2\pi J_\alpha}{L}. \tag{A6}$$

We also introduce the functions[73]

$$\rho_L(k) \equiv \frac{1}{2\pi}\frac{dp_L(k)}{dk}, \tag{A7}$$

$$\mu_L(\Lambda) \equiv \frac{1}{2\pi}\frac{dz_L(\Lambda)}{d\Lambda}. \tag{A8}$$

In the limit $L \to \infty$ with $N/L \equiv n$ kept finite, the pseudomomenta $\{k_n\}$ in the ground state are distributed continuously on the real axis between the pseudo-Fermi points Q_- and Q_+ with density $\rho(k) \equiv \lim_{L\to\infty}\rho_L(k)$, and the spin rapidities $\{\Lambda_\alpha\}$ are distributed continuously between $-\infty$ and ∞ with density $\mu(\Lambda) \equiv \lim_{L\to\infty}\mu_L(\Lambda)$. Equations (2.10) and (2.11) then lead to the following coupled integral equations for $\rho(k)$ and $\mu(\Lambda)$:[74]

$$2\pi\rho(k) = 1 + \cos k \int_{-\infty}^{\infty} d\Lambda \frac{8U\mu(\Lambda)}{U^2 + 16(\Lambda - \sin k)^2}, \tag{A9}$$

$$2\pi\mu(\Lambda) + \int_{-\infty}^{\infty} d\Lambda' \frac{4U\mu(\Lambda')}{U^2 + 4(\Lambda-\Lambda')^2} = \int_{Q_-}^{Q_+} dk \frac{8U\rho(k)}{U^2 + 16(\Lambda - \sin k)^2}, \tag{A10}$$

where Q_- and Q_+ are determined by the conditions

$$\int_{Q_-}^{Q_+} dk\,\rho(k) = n, \qquad \int_{Q_-}^{Q_+} dk\,k\,\rho(k) = n\Phi_c/L, \tag{A11}$$

and μ is normalized to

$$\int_{-\infty}^{\infty} d\Lambda\,\mu(\Lambda) = n/2. \tag{A12}$$

The ground-state energy is then

$$E_0 = -2L\int_{Q_-}^{Q_+} dk\,\cos k\,\rho(k). \tag{A13}$$

E_0 is independent of Φ_c at $n = 1$ in the limit $L \to \infty$, as may be seen from Eqs. (A9)–(A13): at half filling, the pseudomomenta span the entire Brillouin zone $(-\pi, \pi)$, so the shift in the pseudo-Fermi points Q_- and Q_+ caused by the flux Φ_c merely takes the Brillouin zone into itself, so that $\rho(k)$, $\mu(\Lambda)$, and E_0 are all independent of Φ_c at $n = 1$. To obtain an expression for the ground-state energy of the finite system suitable for an asymptotic expansion about $L = \infty$, we express the sum over pseudomomenta in Eq. (2.8) as an integral by means of Eqs. (A1), (A5), and the Poisson summation formula, with the result

$$E_0(\Phi_c) = -2L\int_{-\pi}^{\pi} dk\cos k\,\rho_L(k) \times \sum_{m=-\infty}^{\infty} \exp\{im[Lp_L(k) - \Phi_c + 2\pi s]\}. \tag{A14}$$

In the large-L limit, the dominant Φ_c dependence of E_0 comes from the terms with $m = 0, \pm 1$. Moreover, for large L, $p_L(k)$ may be approximated by $p_c(k) \equiv \lim_{L\to\infty} p_L(k)$ in the terms with $m \neq 0$. $p_c(k)$ may be obtained by integrating the expression for $\rho(k)$ in Ref. 3, and was given in Eq. (6.5). Thus, using Eq. (2.4), the charge stiffness may be written in the large-L limit as $D_c(L) = D_c^{(\delta\rho)} + D_c^{(\delta k)}$, where

$$D_c^{(\delta\rho)} = -L^2\int_{-\pi}^{\pi} dk\cos k\,\left.\frac{\delta^2\rho_L(k)}{\delta\Phi_c^2}\right|_{\Phi_c=0}, \tag{A15}$$

$$D_c^{(\delta k)} = \frac{L}{\pi}\int_{-\pi}^{\pi} dk\sin k\,\sin[Lp_c(k) + 2\pi s]. \tag{A16}$$

$D_c^{(\delta\rho)}$ describes the flux dependence of the ground-state energy due to the change in the distribution of pseudomomenta as a function of Φ_c, while $D_c^{(\delta k)}$ describes that due to a uniform shift of the pseudomomenta $\delta k_n = \Phi_c/L$. Both vanish as $L \to \infty$.

The integral in Eq. (A16) is dominated by the saddle point

$$k_0 = \pi + i\sinh^{-1}(U/4) \tag{A17}$$

and may be evaluated by the method of stationary phase, yielding

$$D_c^{(\delta k)} = \frac{(-1)^{L/2+1}L^{1/2}(U/4)}{|\pi p_c''(k_0)/2|^{1/2}}\exp[-L/\xi(U)], \qquad L\to\infty \tag{A18}$$

(L even), where

$$1/\xi(U) \equiv -ip_c(k_0) = \frac{4}{U}\int_1^\infty dy\,\frac{\ln(y+\sqrt{y^2-1})}{\cosh(2\pi y/U)}. \tag{A19}$$

We next consider Eq. (A15) for $D_c^{(\delta\rho)}$. To find $\delta^2\rho_L(k)/\delta\Phi_c^2|_{\Phi_c=0}$, we first introduce coupled integral equations for $\rho_L(k)$ and $\mu_L(\Lambda)$. Combining Eqs. (A3), (A7), and (A9), we obtain

$$2\pi[\rho_L(k)-\rho(k)] = \cos k\int_{-\infty}^{\infty} d\Lambda\,\frac{8U[\mu_L(\Lambda)-\mu(\Lambda)]}{U^2+16(\Lambda-\sin k)^2} + \cos k\int_{-\infty}^{\infty} d\Lambda\,\frac{8U}{U^2+16(\Lambda-\sin k)^2}\frac{1}{L}\sum_{\beta=1}^{M}[\delta(\Lambda-\Lambda_\beta)-\mu_L(\Lambda)]. \tag{A20}$$

The Poisson summation formula and Eqs. (A2) and (A6) can be used to rewrite Eq. (A20) as

$$2\pi[\rho_L(k)-\rho(k)] = \cos k\int_{-\infty}^{\infty} d\Lambda\,\frac{8U[\mu_L(\Lambda)-\mu(\Lambda)]}{U^2+16(\Lambda-\sin k)^2} + \cos k\int_{-\infty}^{\infty} d\Lambda\,\frac{8U\mu_L(\Lambda)}{U^2+16(\Lambda-\sin k)^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\exp\{im[Lz_L(\Lambda)+2\pi s']\}, \tag{A21}$$

where $s' = \frac{1}{2}[(M-1) \bmod 2]$. After similar manipulations, Eqs. (A4), (A8), and (A10) lead to

$$\begin{aligned}2\pi[\mu_L(\Lambda)-\mu(\Lambda)] = {}& \int_{-\pi}^{\pi} dk\,\frac{8U[\rho_L(k)-\rho(k)]}{U^2+16(\Lambda-\sin k)^2} - \int_{-\infty}^{\infty} d\Lambda'\,\frac{4U[\mu_L(\Lambda')-\mu(\Lambda')]}{U^2+4(\Lambda-\Lambda')^2}\\ &+\int_{-\pi}^{\pi} dk\,\frac{8U\rho_L(k)}{U^2+16(\Lambda-\sin k)^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\exp\{im[Lp_L(k)-\Phi_c+2\pi s]\}\\ &-\int_{-\infty}^{\infty} d\Lambda'\,\frac{4U\mu_L(\Lambda')}{U^2+4(\Lambda-\Lambda')^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\exp\{im[Lz_L(\Lambda')+2\pi s']\}.\end{aligned} \tag{A22}$$

Differentiating Eqs. (A21) and (A22) twice with respect to Φ_c and keeping only the leading-order terms, we obtain the following coupled integral equations for $\delta^2\rho_L(k)/\delta\Phi_c^2|_{\Phi_c=0}$ and $\delta^2\mu_L(\Lambda)/\delta\Phi_c^2|_{\Phi_c=0}$, valid in the large-$L$ limit:

$$2\pi\left.\frac{\delta^2\rho_L(k)}{\delta\Phi_c^2}\right|_{\Phi_c=0} = \cos k\int_{-\infty}^{\infty} d\Lambda\,\frac{8U}{U^2+16(\Lambda-\sin k)^2}\left.\frac{\delta^2\mu_L(\Lambda)}{\delta\Phi_c^2}\right|_{\Phi_c=0}, \tag{A23}$$

$$\begin{aligned}&2\pi\left.\frac{\delta^2\mu_L(\Lambda)}{\delta\Phi_c^2}\right|_{\Phi_c=0} + \int_{-\infty}^{\infty} d\Lambda'\,\frac{4U}{U^2+4(\Lambda-\Lambda')^2}\left.\frac{\delta^2\mu_L(\Lambda')}{\delta\Phi_c^2}\right|_{\Phi_c=0}\\ &\qquad = \int_{-\pi}^{\pi} dk\,\frac{8U}{U^2+16(\Lambda-\sin k)^2}\left.\frac{\delta^2\rho_L(k)}{\delta\Phi_c^2}\right|_{\Phi_c=0} - \int_{-\pi}^{\pi} dk\,\frac{16U\,\rho_\infty(k)}{U^2+16(\Lambda-\sin k)^2}\cos[Lp_c(k)+2\pi s].\end{aligned} \tag{A24}$$

These equations can be solved by Fourier transforms, with the result

$$\left.\frac{\delta^2\mu_L(\Lambda)}{\delta\Phi_c^2}\right|_{\Phi_c=0} = -\int_{-\pi}^{\pi}\frac{dk}{2\pi}\rho(k)\cos[Lp_c(k)+2\pi s]\int_{-\infty}^{\infty}\frac{\exp[i\omega(\Lambda-\sin k)]\,d\omega}{2\cosh(\omega U/4)}, \tag{A25}$$

$$\left.\frac{\delta^2\rho_L(k)}{\delta\Phi_c^2}\right|_{\Phi_c=0} = -\int_{-\pi}^{\pi}\frac{dk'}{2\pi}\rho(k')\cos[Lp_c(k')+2\pi s]\frac{\partial\Theta(k,k')}{\partial k}, \tag{A26}$$

where $\Theta(k,k')$ was defined in Eq. (6.7). Inserting Eq. (A26) into Eq. (A15), and evaluating the integral over k' by the method of stationary phase in the large-L limit, we obtain

$$D_c^{(\delta\rho)} = \frac{(-1)^{L/2}L^{1/2}[1+(4/U)^2]^{1/2}\exp[-L/\xi(U)]}{|\pi p_c''(k_0)/2|^{1/2}}\int_0^\infty dx\,e^{-x}\tanh(x)J_1(4x/U), \quad L\to\infty \tag{A27}$$

(L even). The sign of $D_c^{(\delta\rho)}$ is opposite that of $D_c^{(\delta k)}$. As noted in Ref. 7, $D_c^{(\delta\rho)}$ is negligible compared to $D_c^{(\delta k)}$ in the large-U limit. In the small-U limit they are comparable in magnitude, but $|D_c^{(\delta k)}| > |D_c^{(\delta\rho)}| \ \forall\, U$. Combining Eqs. (A18) and (A27) yields Eq. (3.1), with

$$D(U) = \frac{(U/4)^2 - [1+(U/4)^2]^{1/2}\int_0^\infty dx\, e^{-x}\tanh(x)J_1(4x/U)}{\left(\frac{\pi}{2}\int_0^\infty dx\, e^{-x}J_0(4x/U)\{(U/4)^2 + [1+(U/4)^2]x\tanh(x)\}\right)^{1/2}}. \tag{A28}$$

By applying arguments similar to those above directly to Eq. (A14), one can readily derive Eq. (3.4).

APPENDIX B: CALCULATION OF D_c NEAR HALF FILLING

Here we use a technique due to Haldane[75] to calculate the charge stiffness near half filling, verifying the term linear in δ in Eq. (6.14). Eliminating $\mu(\Lambda)$ from Eqs. (A9) and (A10) yields an integral equation for $\rho(k)$ alone:[76]

$$2\pi\rho(k) = 1 + \int_{Q_-}^{Q_+} dk'\,\frac{\partial\Theta(k,k')}{\partial k}\rho(k'), \tag{B1}$$

where $\Theta(k,k')$ was defined in Eq. (6.7). We note that Eqs. (A11), (A13), and (B1), which determine the dependence of the ground-state energy on Φ_c in the limit $L\to\infty$, are identical in form to those considered by Haldane in his treatment of the spinless 1D quantum fluid.[75] The result for the charge stiffness obtained in Ref. 75 is therefore applicable here, and is

$$D_c = \frac{e^{2\phi}v_c}{2\pi}, \tag{B2}$$

where $e^\phi = 1-\omega(Q)-\omega(-Q)$ is a positive number which parametrizes the electron-electron interactions,[22] and

$$v_c = \frac{1}{2\pi\rho(Q)}\left(\varepsilon'(Q) - \int_{-Q}^{Q} dk\,\varepsilon'(k)\tau(k)\right) \tag{B3}$$

is the charge velocity. Here $\varepsilon(k) = -2t\cos k$, and $\omega(k)$ and $\tau(k)$ are solutions of the following integral equations:

$$2\pi\omega(k) = \Theta(k,Q) - \int_{-Q}^{Q} dk'\,\frac{\partial\Theta(k,k')}{\partial k'}\omega(k'), \tag{B4}$$

$$2\pi\tau(k) = \frac{\partial\Theta(k,Q)}{\partial Q} - \int_{-Q}^{Q} dk'\,\frac{\partial\Theta(k,k')}{\partial k'}\tau(k'). \tag{B5}$$

Using Eq. (6.7) for $\Theta(k,k')$ and the fact that $Q=\pi$ at half filling, we see that v_c vanishes at half filling because $\varepsilon'(\pi)=0$ and $\tau(-k)=\tau(k)$ when $Q=\pi$. Also, $\omega(\pi) = \omega(-\pi) = 0$, so $e^\phi = 1$ at half filling. Thus, for a small doping δ away from half filling,

$$D_c \simeq -\delta\left.\frac{\partial D_c}{\partial n}\right|_{n=1^-} = -\delta\left.\frac{\partial D_c}{\partial Q}\right|_{Q=\pi}\left.\frac{\partial Q}{\partial n}\right|_{n=1^-}. \tag{B6}$$

From Ref. 75 $\partial Q/\partial n = [1-\omega(Q)+\omega(-Q)]/2\rho(Q)$, so $(\partial Q/\partial n)|_{n=1^-} = 1/2\rho(\pi)$. After some manipulation, Eq. (B6) becomes

$$D_c \simeq \frac{t\delta}{[2\pi\rho(\pi)]^2}\left(1 + \frac{1}{2\pi}\int_{-\pi}^{\pi}\sin k\left.\frac{\partial^2\Theta(k,Q)}{\partial Q^2}\right|_{Q=\pi} dk\right). \tag{B7}$$

Inserting the result for $\rho(\pi)$ from Ref. 3 into Eq. (B7) and using Eq. (6.7), we obtain[51]

$$D_c \simeq t\delta\frac{1 - 2\int_0^\infty d\omega\,\omega J_1(\omega)/[1+\exp(\omega U/2t)]}{\left(1 - 2\int_0^\infty d\omega J_0(\omega)/[1+\exp(\omega U/2t)]\right)^2}. \tag{B8}$$

Note that the above expression for D_c may be rewritten using Eqs. (6.9), (B9), and (B10) as $D_c \simeq \delta/|2m^*|$, thus confirming the term linear in δ in Eq. (6.14). Note also that $v_c \simeq \pi\delta/|m^*|$ for $\delta \ll 1$, in agreement with Eq. (6.19).

For completeness, we give below the explicit expressions for $p_c'(\pi)$, $\varepsilon_c''(\pi)$, $p_c^{(3)}(\pi)$, and $\varepsilon_c^{(4)}(\pi)$:

$$p_c'(\pi) = 1 - 2\int_0^\infty \frac{J_0(\omega)\,d\omega}{1+\exp(\omega U/2t)}, \tag{B9}$$

$$\varepsilon_c''(\pi) = -2t\left(1 - 2\int_0^\infty \frac{\omega J_1(\omega)\,d\omega}{1+\exp(\omega U/2t)}\right), \tag{B10}$$

$$p_c^{(3)}(\pi) = 2\int_0^\infty \frac{(1+\omega^2)J_0(\omega)\,d\omega}{1+\exp(\omega U/2t)}, \tag{B11}$$

$$\varepsilon_c^{(4)}(\pi) = 2t\left(1 - 2\int_0^\infty \frac{(\omega^3+4\omega)J_1(\omega)\,d\omega}{1+\exp(\omega U/2t)}\right). \tag{B12}$$

Using the Laplace transforms of J_0 and J_1, and the Sommerfeld-Watson transformation, Eq. (B8) may be rewritten in a form more suitable for an asymptotic expansion about $U=0$ as follows:

$$D_c \simeq t\delta\left(\frac{\pi^2 t}{U}\int_1^\infty dy\,\frac{(y^2-1)^{1/2}[2\coth^2(2\pi t y/U)-1]}{\sinh(2\pi t y/U)}\right) \times\left(\int_1^\infty dy\,\frac{(y^2-1)^{-1/2}}{\sinh(2\pi t y/U)}\right)^{-2}. \tag{B13}$$

[1]N. F. Mott, *Metal-Insulator Transitions* (Taylor and Francis, London, 1974).
[2]P. W. Anderson, Science **235**, 1196 (1987).
[3]E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968).
[4]B. S. Shastry and B. Sutherland, Phys. Rev. Lett. **65**, 243 (1990).
[5]W. Kohn, Phys. Rev. **133**, A171 (1964).
[6]This is true even in disordered systems, and is closely related to the existence of persistent currents in mesoscopic rings [see M. Büttiker, Y. Imry, and R. Landauer, Phys. Lett. A **96**, 365 (1983)].
[7]C. A. Stafford, A. J. Millis, and B. S. Shastry, Phys. Rev. B **43**, 13 660 (1991).
[8]K. Kim and P. B. Weichman, Phys. Rev. B **43**, 13 583 (1991).
[9]F. Woynarovich, J. Phys. C **15**, 85 (1982); **15**, 97 (1982); **16**, 5293 (1983).
[10]H. J. Schulz, Int. J. Mod. Phys. B **5**, 57 (1991).
[11]A. Robaszkiewicz, R. Micnas, and K. A. Chao, Phys. Rev. B **23**, 1447 (1981), and references therein, see also Ref. 3.
[12]The minimum energy eigenvalue for fixed N, L, and U is always obtained in the singlet state and may occur at either $\Phi_\uparrow = \Phi_\downarrow = 0$ or $\Phi_\uparrow = \Phi_\downarrow = \pi$. However, when $N \neq L$ and $U > 0$ there is a level crossing with the triplet state (see Refs. 37 and 38).
[13]A. J. Millis and S. N. Coppersmith, Phys. Rev. B **42**, 10 807 (1990).
[14]P. F. Maldague, Phys. Rev. B **16**, 2437 (1977).
[15]Actually, the Bethe ansatz states are not a complete set of energy eigenstates. However, the remaining eigenstates are related to these by symmetry [see F. H. L. Essler, V. E. Korepin, and K. Schoutens, Phys. Rev. Lett. **67**, 3848 (1991)].
[16]A. A. Ovchinnikov, Zh. Eksp. Teor. Fiz. **57**, 2137 (1969) [Sov. Phys.—JETP **30**, 1160 (1970)].
[17]C. F. Coll, Phys. Rev. B **9**, 2150 (1974).
[18]R. M. Fye, M. J. Martins, D. J. Scalapino, J. Wagner, and W. Hanke, Phys. Rev. B **44**, 6909 (1991).
[19]G. Schröder and J. F. M. Oth, Tetrahedron Lett. **1966**, 4083 (1966); I. C. Calder and F. Sondheimer, Chem. Commun. **24**, 904 (1966).
[20]N. Byers and C. N. Yang, Phys. Rev. Lett. **7**, 46 (1961).
[21]V. J. Emery, in *Highly Conducting One-Dimensional Solids*, edited by J. T. De Vreese, R. P. Evrard, and V. E. Van Doren (Plenum, New York, 1979); J. Solyom, Adv. Phys. **28**, 209 (1979), and references therein.
[22]F. D. M. Haldane, J. Phys. C **12**, 4791 (1979); **14**, 2585 (1981).
[23]R. Heindenreich, R. Seiler, and D. A. Uhlenbrock, J. Stat. Phys. **22**, 27 (1980).
[24]H. J. Schulz, Phys. Rev. Lett. **64**, 2831 (1990).
[25]T. Giamarchi, Phys. Rev. B **44**, 2905 (1991).
[26]S. Coleman, Phys. Rev. D **11**, 2088 (1975).
[27]V. J. Emery, A. Luther, and I. Peschel, Phys. Rev. B **13**, 1272 (1976).
[28]H. J. Schulz, Phys. Rev. B **22**, 5274 (1980).
[29]F. D. M. Haldane, J. Phys. A **15**, 507 (1982).
[30]We thank F. D. M. Haldane for pointing this argument out to us.
[31]To facilitate an asymptotic expansion of $\xi(U)$ about $U = \infty$, one can rewrite Eq. (3.2) as $1/\xi(U) = \sinh^{-1}(U/4) - \int_0^\infty dx\, x^{-1} e^{-x} \tanh x\, J_0(4x/U)$.
[32]This may be seen most easily in Brillouin-Wigner perturbation theory, where the energy denominators need only be kept to $\mathcal{O}(U)$ to obtain the leading-order expression for G.
[33]The existence of pairing correlations in the ground state of the attractive 1D Hubbard model was inferred from the power-law decay of the pair-pair correlation function by N. M. Bogoliubov and V. E. Korepin, Int. J. Mod. Phys. B **3**, 427 (1989) (see also Ref. 39). Note, however, that the expression of Bogoliubov and Korepin for the correlation length of the equal-time single-particle Green's function does not agree with our result at half filling, Eq. (3.2).
[34]H. R. Krishnamurthy, C. Jayaprakash, S. Sarker, and W. Wenzel, Phys. Rev. Lett. **64**, 950 (1990).
[35]B. Sutherland, in *Exactly Solvable Problems in Condensed Matter and Relativistic Field Theory*, edited by B. S. Shastry, S. S. Jha, and V. Singh (Springer-Verlag, Berlin, 1985).
[36]J. A. Hertz, Phys. Rev. B **14**, 1165 (1976).
[37]M. Ogata and H. Shiba, Phys. Rev. B **41**, 2326 (1990).
[38]Under periodic boundary conditions the ground state of the system is a singlet when $N \bmod 4 = 2$, but when $N \bmod 4 = 0$ the ground state is a singlet if $N = L$ and a *triplet* if $N \neq L$. This does not violate the Lieb-Mattis theorem [E. H. Lieb and D. C. Mattis, Phys. Rev. **125**, 164 (1962)], which implies that the ground state of a system of spin-1/2 fermions on a lattice, interacting via a velocity-independent potential, must be a singlet when $N \bmod 2 = 0$ given *open* boundary conditions. It is straightforward to apply their argument to the case of *periodic* boundary conditions, in which case one finds that when $N \bmod 4 = 2$ the ground state is a singlet, but when $N \bmod 4 = 0$ the ground state may be either singlet or triplet. Under *antiperiodic* boundary conditions, the theorem implies that the ground state is always a singlet when $N \bmod 4 = 0$, but may be either singlet or triplet when $N \bmod 4 = 2$.
[39]N. Kawakami and S.-K. Yang, Phys. Rev. B **44**, 7844 (1991).
[40]J. M. P. Carmelo, P. Horsch, and A. A. Ovchinnikov, Phys. Rev. B **46**, 14 728 (1992).
[41]V. J. Emery, Phys. Rev. Lett. **65**, 1076 (1990).
[42]The finite-size corrections are more complicated for other choices of boundary conditions, and may be either positive or negative, depending on N, L, and U.
[43]In the insulating phase, the finite-size corrections to D_c are of course exponential in L, but those to $N_{\rm tot}$ are still $\mathcal{O}(L^{-2})$ (see Ref. 7).
[44]F. Woynarovich, J. Phys. A **22**, 4243 (1989).
[45]The case $n = 1$ is an exception, where the finite-size corrections to D_c near the critical point have opposite signs and vastly different magnitudes for periodic and antiperiodic boundary conditions when $L < \xi$, as discussed in Refs. 7 and 18.
[46]P. W. Anderson, in *Frontiers and Borderlines in Many-Particle Physics*, edited by R. A. Broglia and J. R. Schrieffer (North-Holland, New York, 1988); Phys. Rep. **184**, 195 (1989).
[47]Note that $\Theta(k, k')$ is the same effective two-body scattering phase shift which appears in the Bethe ansatz integral equation for the ground-state distribution of pseudomomenta in the infinite system, Eq. (B1).
[48]N. Kawakami and A. Okiji, Phys. Rev. B **40**, 7066 (1989).
[49]J. Carmelo, P. Horsch, P. A. Bares, and A. A. Ovchinnikov, Phys. Rev. B **44**, 9967 (1991).
[50]C. A. Stafford, Ph.D. thesis, Princeton University (University Microfilms International, Ann Arbor, MI, 1992).
[51]We remark that the term linear in δ in Eq. (6.14) [see also Eqs. (B8) and (B13)] differs from that obtained by N. Kawakami and S.-K. Yang, Phys. Rev. Lett. **65**, 3063

(1990) (see also Ref. 39), by a factor of 4. They obtain their result by use of a relation which follows from the conformal invariance of the fixed point, combined with their previous result for the compressibility, Eq. (2.11) in Ref. 48. We believe Eq. (2.11) in Ref. 48 is incorrect by a factor of 4; we have traced the discrepancy to the expression for $\rho(\pi)$ in Ref. 48, Eq. (2.6b), which differs from the correct result (Ref. 3) by a factor of 2.

[52]We remark that Eq. (6.18) remains valid when terms of $\mathcal{O}(\delta^2)$ are included, provided we replace $1/m^*$ by $d^2\varepsilon_c/dp^2|_{p=\pi(1-\delta)}$ for $n < 1$ and by $d^2\varepsilon_c^{\rm UHB}/dp^2|_{p=\pi(1-\delta)}$ for $n > 1$. The relation $2\partial D_c/\partial n = \varepsilon_c''(p_F)$ is generically true for noninteracting fermions in a 1D energy band $\varepsilon_c(p)$, provided there are only two Fermi points.

[53]Here we have used the fact that the holons are effectively noninteracting near half filling, so that the effective charge of a charge excitation, as defined in Ref. 39, is unity. [See also F. Woynarovich, J. Phys. A **22**, 4243 (1989); H. Frahm and V. E. Korepin, Phys. Rev. B **42**, 10 553 (1990); J. Carmelo and P. Horsch, Phys. Rev. Lett. **68**, 871 (1992).]

[54]We note that neither Eq. (6.18) nor Eq. (6.20) holds for arbitrary electron density. However, we have evaluated the second derivative of the charge excitation energy with respect to momentum for large, finite systems ($L > N \approx 100$) using a finite-difference formula and a numerical solution of Eqs. (2.10) and (2.11), and we find that the effective mass $m^* = (d^2\varepsilon_c/dp^2)^{-1}$ of a wave packet composed of low-lying charge excitations changes sign precisely where D_c has its maximum. Our conclusions are restricted to the region $U \geq 2$, where we are not limited by finite-size corrections to $D_c(n)$. The two definitions of the *sign* of the charge carriers implicit in Eqs. (6.18) and (6.20) thus appear to be consistent for arbitrary electron density.

[55]See, e.g., H. B. Callen, *Thermodynamics*, 2nd ed. (Wiley, New York, 1985), pp. 320–322.

[56]C. A. Stafford (unpublished).

[57]T. Giamarchi and A. J. Millis, Phys. Rev. B **46**, 9325 (1992).

[58]P. M. Chaikin, R. L. Greene, S. Etemad, and E. Engler, Phys. Rev. B **13**, 1627 (1976).

[59]E. Dagotto, A. Moreo, F. Ortolani, D. Poilblanc, J. Riera, and D. Scalapino, Phys. Rev. B **45**, 10 741 (1992).

[60]P. W. Anderson, Phys. Rep. **184**, 195 (1989); Phys. Rev. Lett. **64**, 1839 (1990); **65**, 2306 (1990); **66**, 3226 (1991); **67**, 2092 (1991).

[61]H. Ishii, H. Sato, N. Kanazawa, H. Takagi, S. Uchida, K. Kitazawa, K. Kishio, K. Fueki, and S. Tanaka, Physica **148B**, 419 (1987).

[62]Y. Ando, M. Sera, S. Yamagata, S. Kondoh, M. Onoda, and M. Sato, Solid State Commun. **70**, 303 (1989).

[63]A. B. Kaiser and C. Uher, in *Studies of High Temperature Superconductors*, edited by A. V. Narlikar (Nova Science, New York, 1990), Vol. 7, and references therein.

[64]S. J. Hagen, X. Q. Xu, W. Jiang, J. L. Peng, Z. Y. Li, and R. L. Greene, Phys. Rev. B **45**, 515 (1992).

[65]X. Q. Xu, S. J. Hagen, W. Jiang, J. L. Peng, Z. Y. Li, and R. L. Greene, Phys. Rev. B **45**, 7356 (1992).

[66]J. Sugiyama, K. Matsuura, M. Kosuge, H. Yamauchi, and S. Tanaka, Phys. Rev. B **45**, 9951 (1992).

[67]The experimental situation is complicated by the fact that the impurity potential introduced by the dopants localizes the charge carriers at small dopant concentrations. Thus the doping regime which Eq. (6.22) might be expected to describe may not be accessible experimentally. At larger dopings, the thermopower of the 1D Hubbard model has a contribution from the entropy carried by the spin excitations, which leads to an unusual temperature dependence (see Ref. 56).

[68]G. H. Chen *et al.*, Mod. Phys. Lett. B **3**, 295 (1989).

[69]J. L. Cohn, E. F. Skelton, S. A. Wolf, and J. Z. Liu, Phys. Rev. B **45**, 13 140 (1992).

[70]S. L. Cooper *et al.*, Phys. Rev. B **41**, 11 605 (1990).

[71]S. Uchida, T. Ido, H. Takagi, T. Arima, Y. Tokura, and S. Tajima, Phys. Rev. B **43**, 7942 (1991).

[72]G. A. Thomas, D. H. Rapkine, S. L. Cooper, S.-W. Cheong, L. F. Schneemeyer, and J. V. Waszczak, Phys. Rev. B **45**, 2474 (1992).

[73]H. J. de Vega and F. Woynarovich, Nucl. Phys. B **251**, 439 (1985); F. Woynarovich and H. P. Eckle, J. Phys. A. **20**, L443 (1987).

[74]B. S. Shastry (private communication). Equations (A9) and (A10) are identical to those in Ref. 3, except for the shift in the pseudo-Fermi points Q_- and Q_+ due to the flux Φ_c.

[75]F. D. M. Haldane, Phys. Lett. **81A**, 153 (1981).

[76]B. S. Shastry (private communication). Equation (B1) differs from that derived by H. Shiba, Phys. Rev. B **6**, 930 (1972), only by the shift of the pseudo-Fermi points from $-Q, Q$ to Q_-, Q_+, and may be derived by the same method.