Sum rule analysis of umklapp processes and Coulomb energy: Application to cuprate superconductivity

Misha Turlakov[1] and Anthony J. Leggett[2]
[1]*Cavendish Laboratory, University of Cambridge, Cambridge, CB3 OHE, United Kingdom*
[2]*Loomis Lab, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
(Received 17 July 2002; published 31 March 2003)

The third moment frequency sum rule for the density-density correlation function is rederived in the presence of umklapp processes. Upper and lower bounds on the electron-electron Coulomb energy are derived in two-dimensional and three-dimensional media, and the umklapp processes are shown to be crucial in determining the spectrum of the density fluctuations (especially for the two-dimensional systems). This and other standard sum rules can be used in conjunction with experimental spectroscopies (electron energy-loss spectroscopy, optical ellipsometry, etc.) to analyze changes of the electron-electron Coulomb energy at the superconducting transition in cuprates.

DOI: 10.1103/PhysRevB.67.094517 PACS number(s): 74.25.−q, 78.20.−e

I. INTRODUCTION

Theoretical progress in the understanding of the microscopic origin of high-temperature superconductivity appears to be ambiguous. The novelty and difficulty is to describe the strong effects of the electron-electron interactions which determine the strongly correlated phases of cuprates at various dopings: the Mott antiferromagnetic insulator, the ''anomalous'' metallic state, and the superconductor. In fact, the fundamentally new microscopic origin of superconductivity (SC) in these materials is, perhaps, due to electron-electron interactions unlike standard phonon-mediated superconductivity.[1,2] We explore the general aspects of such a scenario with the help of sum rules for the density-density correlation function.

The particular question of interest is the origin of the condensation energy, the difference in the energy between the ''normal'' state extrapolated to $T=0$ K, and the superconducting ground state. The famous BCS theory of superconductivity[1] is based on the assumption that the attractive interaction between electrons arises from the lattice vibrations. Here we would like to investigate a general alternative to the BCS phonon-mediated superconductivity, namely, that the superconducting state (e.g., the condensation energy) is promoted either by the long-range part of electron-electron Coulomb interaction or by the short-range part of electron-electron and static electron-ion interactions. The identification of the part of the full electron-ion Hamiltonian responsible for the condensation energy would be an important step towards a complete and consistent theory of high-temperature superconductivity.

Many proposals for the condensation energy have been suggested,[3,4] most of them (except Ref. 3) based on phenomenological Hamiltonians. Experimental confirmation of the origin of the condensation energy based on a phenomenological Hamiltonian would still require the validation of the phenomenological Hamiltonian (by various other experiments) in order to construct a self-consistent theory. Alternatively, the essential terms of the full[5] original electron-ion Hamiltonian (and changes of the expectation values thereof upon transitions) can be established consistently from experiments with the use of sum rules at the outset. Such an approach, taken in this paper, can potentially identify the origin of the condensation energy and the phenomenological Hamiltonian sufficiently to describe superconductivity and other strongly correlated phases. For a much more explicit exposition of the general philosophy adopted in this paper, see the Introduction in Ref. 3.

If the ion kinetic and ion-ion Coulomb energies are assumed irrelevant (or in other words, these terms do not change upon the phase transitions and do not determine the important correlated phases), the full[5] electron-ion Hamiltonian can be reduced to the following form:

$$\hat{H}=\sum_{p,\sigma}\frac{p^2}{2m}c^{+}_{p,\sigma}c_{p,\sigma}+\frac{1}{2\Omega}\sum_{q\neq 0}V_q[\hat{\rho}_q\hat{\rho}_{-q}-N]+\sum_{\kappa\neq 0}U_{-\kappa}\hat{\rho}_{\kappa}, \tag{1}$$

where $\hat{\rho}_q=\Sigma_{k,\sigma}c^{+}_{k-q,\sigma}c_{k,\sigma}=\Sigma_{\vec{r}_i}e^{i\vec{q}\vec{r}_i}$ is the total density operator. N is the number of electrons, and Ω is a total volume. The first and second terms are the kinetic and Coulomb interaction energies of the electrons. The third term describes the interaction of the electrons with the periodic potential of the lattice, which can be represented by the umklapp pseudopotential $U_{-\kappa}$ with the sum over corresponding wave vectors κ of the reciprocal lattice.[6] The term of interaction between electrons and the positive homogeneous ion background is omitted. In spite of making the ''static lattice'' assumption [the dynamic lattice effects (e.g., phonons) are neglected], the Hamiltonian (1) is quite general. For instance, the Hubbard model (and multiband versions of it) is only a reduced version of the Hamiltonian (1) which neglects the long-range part of the Coulomb interaction. Presumably, the Hamiltonian (1) is sufficient not only to describe the Mott insulating state and the metallic state at high densities (far away from the half-filled band) but also all other important phases of the cuprates.

We can analyze the electron Coulomb energy in situations of different dimensionalities. In an isotropic medium the three-dimensional (3D) Coulomb potential is $V_q = e^2/\epsilon_0\epsilon_\infty q^2$, where ϵ_∞ is the high-frequency dielectric constant due to the screening by the core electrons.[7] In a two-dimensional (2D) plane, the Coulomb potential is $V_q = e^2/2\epsilon_0\epsilon_\infty q$. For a layered electron gas with interplane distance d (relevant for the discussion of single-layer cuprates) the Coulomb interaction is

$$V_{q,q_z} = \frac{e^2 d}{2\epsilon_0\epsilon_\infty q}\frac{\sinh(qd)}{\cosh(qd)-\cos(q_z d)} \tag{2}$$

with different dependencies on the wave numbers q parallel and q_z perpendicular to the planes.[8] The Coulomb potential of a layered gas becomes three-dimensional, $V_q = e^2/\epsilon_0\epsilon_\infty q^2$, in the long-wavelength limit ($q_z d \ll 1$ and $qd \ll 1$) and two-dimensional, $V_q = e^2/2\epsilon_0\epsilon_\infty q$, for short wavelengths ($qd \gg 1$, independently of q_z momentum).

II. SUM RULES.

The electron-electron Coulomb energy can be related to the density response function $\chi(q,\omega)$, and thus it is instructive to analyze various sum rules for this response function.[9] In particular, the expectation value $\langle V_c\rangle$ of the Coulomb energy can be written in the form

$$\langle V_c\rangle \equiv \frac{1}{2\Omega}\sum_q V_q[\langle\hat\rho_q\hat\rho_{-q}\rangle - N] = \sum_q\left[\langle V_{c,q}\rangle - \frac{1}{2}V_q n\right],$$

$$\langle V_{c,q}\rangle \equiv \frac{1}{2}V_q\int\frac{\hbar d\omega}{2\pi}\,\mathrm{Im}\,\chi(q,\omega)\coth\left(\frac{\hbar\omega}{2kT}\right), \tag{3}$$

where $n = N/\Omega$ is the density of the electron system, and the density-density correlation function[10] $\chi(q,\omega)$ is defined in the standard way:[11,12]

$$\chi(q,\omega) \equiv \frac{i}{\hbar\Omega}\int_0^{+\infty} dt e^{i(\omega+i\delta)t}\langle[\hat\rho(q,t),\hat\rho^+(q,0)]\rangle. \tag{4}$$

$\langle V_{c,q}\rangle$ is the expectation value of the partial Coulomb energy corresponding to a particular value of momentum q. The structure factor S_q,

$$S_q = \frac{1}{N}\langle\hat\rho_q\hat\rho_{-q}\rangle - N\delta_{q,0}, \tag{5}$$

can be expressed again through the imaginary part $\mathrm{Im}\,\chi(q,\omega)$ using the fluctuation-dissipation theorem

$$nS_q = \int_{-\infty}^{+\infty}\mathrm{Im}\,\chi(q,\omega)\coth\left(\frac{\hbar\omega}{2kT}\right)\frac{\hbar d\omega}{2\pi},$$

$$\langle V_{c,q}\rangle = \frac{1}{2}V_q n S_q. \tag{6}$$

Various sum rules for the imaginary part of the susceptibility $\mathrm{Im}\,\chi(q,\omega)$ (valid for an arbitrary dimensional system with the appropriate form of Coulomb potential V_q) can be derived by calculating the commutators of the density operator and the Hamiltonian. To calculate the well-known f-sum rule (or the first moment sum rule) it is necessary to calculate the expectation value of the operator $[[\hat\rho_q,\hat H],\hat\rho_{-q}]$. Another additional sum rule is the causality (Kramers-Kronig) relation. These two well-known sum rules are[12]

$$J_{-1} \equiv \frac{2}{\pi}\int_0^{+\infty}\frac{\mathrm{Im}\,\chi(q,\omega)}{\omega}d\omega = \chi(q,0), \tag{7}$$

$$J_1 \equiv \frac{2}{\pi}\int_0^{+\infty}\omega\,\mathrm{Im}\,\chi(q,\omega)d\omega = \frac{nq^2}{m}. \tag{8}$$

To calculate the ω^3-moment sum rule[13] the expectation value of the operator $[[[[\hat\rho_q,\hat H],\hat H],\hat H],\hat\rho_{-q}]$ must be calculated. The result of a long calculation is

$$\begin{aligned} J_3 &\equiv \frac{2}{\pi}\int_0^{+\infty}\omega^3\mathrm{Im}\,\chi(q,\omega)d\omega \\ &= \frac{1}{m^2}\left\langle\frac{1}{\Omega}\sum_\kappa(\vec\kappa\vec q)^2(-U_{-\kappa}\hat\rho_\kappa)\right\rangle \\ &\quad + q^4\frac{n^2}{m^2}V_q + q^4\frac{3}{m^2}\langle\hat T_{pr}\rangle + q^6\frac{2n\hbar^2}{(2m)^3} \\ &\quad + \frac{n}{m^2}\frac{1}{\Omega}\sum_{p\neq -q}[V_{|\vec p+\vec q|}(\vec p\vec q+q^2)^2 - V_p(\vec p\vec q)^2]S_p, \end{aligned} \tag{9}$$

where $\hat T_{pr} = 1/\Omega\Sigma_p[(\vec p\vec{\tilde q})^2/2m]c_p^+c_p = \int[d^3p/(2\pi\hbar)^3]\times[(\vec p\vec{\tilde q})^2/2m]c_p^+c_p$ is the projected kinetic-energy operator, and $\vec{\tilde q} = \vec q/|q|$ is a unit vector along the direction of $\vec q$. This higher-order sum rule [Eq. (9)] is convenient for analysis of high-frequency transitions in the density response, because it weights higher frequencies by a factor ω^3. The existence (or convergence) of the third moment sum rule can be demonstrated by showing that all terms on the right-hand side of Eq. (9) are finite. The first and the third terms are expected to be finite, because they are essentially related to the finite expectation values of the electron lattice and the kinetic energy in the ground state. The convergence and q dependence of the last term will be discussed in detail separately for 2D and 3D cases.

The three sum rules [Eqs. (7)–(9)] allow us to derive upper and lower bounds[15] on the electron Coulomb energy, and hence to discuss the possible changes of Coulomb energy due to the phase transition. Since the imaginary part of the susceptibility $\mathrm{Im}\,\chi(q,\omega)$ is a real positive definite function, the two Cauchy-Schwartz inequalities can be written for the partial Coulomb energy $\langle V_c\rangle_q$ at $T=0$ K,

$$\frac{1}{2}(V_q^2 J_{-1}J_1)^{1/2} \geqslant \langle V_{c,q}\rangle \geqslant \frac{1}{2}\left(V_q^2\frac{J_1^3}{J_3}\right)^{1/2}. \tag{10}$$

It is convenient to introduce the notional "plasma frequency" $\omega_p(q)$ defined by $\omega_p(q) = (nq^2V_q/m)^{1/2}$. In fact,

defined as above, the ''plasma frequency'' has the right asymptotics for the corresponding plasma waves in the three-dimensional case $[\omega^2_{p,3D}(q)=e^2n_{3D}/(\epsilon_0 m)]$ and the two-dimensional case $[\omega^2_{p,2D}(q)=e^2n_{2D}q/(2\epsilon_0 m)]$. Another relation, which is useful in order to rewrite the J_{-1} sum rule for $q\to 0$, can be derived in the long-wavelength limit ($q \ll q_{TF}$, where q_{TF} is the inverse of the Thomas-Fermi screening length) for the full susceptibility $\chi(q,0)$, if we express $\chi(q,\omega)$ through the ''bare''(or local) susceptibility $\chi_0(q,\omega)$:

$$\chi(q,\omega)=\frac{\chi_0(q,\omega)}{1+V_q\chi_0(q,\omega)} \tag{11}$$

(this is an exact result, not a random-phase approximation, provided χ_0 is defined in terms of the relevant irreducible diagrams). If ''bare'' susceptibility $\chi_0(q,0)$ of the electron liquid is assumed to be finite (see Sec. 4.1 of Ref. 12 for the compressibility sum rule), then for $q \ll q_{TF}$,

$$\chi(q,\omega=0)=\frac{1}{V_q\left[1+\dfrac{1}{V_q\chi_0(q,0)}\right]}\approx\frac{1}{V_q}. \tag{12}$$

The discussion up to this point is valid for any dimensionality of the system (with the corresponding form of Coulomb potential V_q). In what follows we analyze the two-dimensional and three-dimensional cases separately and find important differences.

We consider first the 3D case. The leading terms at small q for the third moment sum rule are

$$J_3\approx\frac{q^2}{m^2}\langle\hat{A}\rangle+q^4\frac{n^2}{m^2}V_q, \tag{13}$$

where $\hat{A}=1/\Omega\Sigma_\kappa(\vec{\kappa}\vec{\hat{q}})^2(-1)U_{-\kappa}\hat{\rho}_\kappa$.[16] The third and fourth terms in Eq. (9), being proportional to q^4 and q^6 powers, are subdominant. The last term in Eq. (9),

$$\frac{1}{m^2}\int d^Dp[V_{|\vec{p}+\vec{q}|}(\vec{p}\vec{q}+q^2)^2-V_p(\vec{p}\vec{q})^2]nS_p, \tag{14}$$

requires careful analysis of its convergence and q dependence. The q dependence (and convergence), due to the part of the integral over small momenta p, is evident, since the upper bound on the pair-correlation function S_p [see Eqs. (6) and (10)] is $S_p\leqslant Bp^{(D+1)/2}$ (where B is a constant and D is dimensionality)[17] is

$$\int d^3pV_p(\vec{p}\vec{q})^2[S_{p+q}-S_p]\sim\int d^3pV_p(\vec{p}\vec{q})^2S_q \tag{15}$$

for small momenta $p<q$. Since in the 3D case the upper bound $S_q\leqslant Bq^2$, the contribution from small momenta p integration is at least of order q^4 (or higher power of q for small q). To analyze the q dependence of the part of the integral [Eq. (14)] from the integration over large momenta p, we use the ''cusp theorem,''[18] which gives the asymptotic behavior at large momenta p (the 3D case):

$$S_p=1-\frac{C_{3D}}{p^4}+o\left(\frac{1}{p^5}\right), \tag{16}$$

where C_{3D} is some constant. The integral of Eq. (14) with $S_p=1$ is identically zero (this is why S_p can be substituted by S_p-1 if convenient). It is easy to see that the integral over p with the second term of expansion C_{3D}/p^4, is convergent. By expanding in powers of (q/p) for large p, the leading q dependence of the considered term

$$\int d^3pV_p(\vec{p}\vec{q})^2[S_{p+q}-S_p]$$

$$=\int d^3p\frac{V_p(\vec{p}\vec{q})^2C_{3D}}{p^4}\left(\frac{1}{[1+(q/p)^2+2(q/p)\cos\theta]^2}-1\right) \tag{17}$$

is found to be proportional to q^4. The terms, proportional to q^3 and other odd powers of q, equal zero after the integration over angle θ [$\cos\theta=(\vec{p}\vec{q})/pq$ is the angle between $\vec{p}$ and $\vec{q}$)]. It can be seen that for large q the last term [Eq. (14)] grows no faster than q^4 as well. What is important for the ensuing discussion is that, both for small and large q, the last term of Eq. (9) has *subleading* q^4 dependence on the wave vector q.

The upper and lower limits on the partial Coulomb energy can be conveniently written in terms of the 3D ''plasma frequency'' as

$$\frac{\hbar}{2}\omega_{p,3D}(q)+o(q^2)\geqslant\langle V_{c,q}\rangle,$$

$$\langle V_{c,q}\rangle\geqslant\frac{\hbar}{2}\frac{\omega_{p,3D}(q)}{\sqrt{1+\dfrac{\langle\hat{A}\rangle}{nm\omega^2_{p,3D}(q)}}}+o(q^2). \tag{18}$$

The inequalities (18) have an interesting significance for the possibility of gaining energy, in the SC condensation, from small-q modes. We first notice that in the 3D case if $\langle\hat{A}\rangle=0$ in both normal and superconducting states, which is certainly the case if there is no crystalline potential, the terms proportional to q^4 [the third and last terms of the right-hand side of Eq. (9)] determine the difference between the upper and lower bounds. Therefore if $\langle\hat{A}\rangle=0$, then for given q the maximum possible saving is proportional to $(\hbar/2)\omega_{p,3D}(q/q_0)^2$ (where $q_0\sim q_{TF,3D}$ and $q_{TF,3D}=\sqrt{e^2k_Fm^*/\epsilon_0\pi^2\hbar^2}$ in three dimensions, k_F is the Fermi wave vector), which is a negligible portion of the partial Coulomb energy $\hbar/2\omega_{p,3D}$ at long wavelengths. In fact, in the absence of the umklapp processes the sum rules [Eqs. (8) and (9)] essentially fix the density spectrum at long wavelengths $q<q_{TF}$ to the plasma pole contribution: $\mathrm{Im}\,\chi(q,\omega)\sim[\pi\hbar\omega_{p,3D}(q)/2V_q]\delta[\omega-\omega_{p,3D}(q)]$, which satisfies completely all three sum rules at $q\to 0$. Other terms in the J_3 sum rule become comparable with the dominant term $q^4V_qn^2/m^2$ only at $q\geqslant q_{TF}$. Therefore, a large saving of

Coulomb energy *in the 3D case in the absence of the lattice* due to the phase transition is possible only for short wavelengths $q \geqslant q_{TF,3D}$.

In the presence of the periodic lattice potential ($\langle \hat{A} \rangle \neq 0$), it can be seen from the inequalities (18) that for $q < q_{TF,3D}$ the maximum theoretical saving is a finite fraction of the zero-point plasma energy $\hbar/2\omega_{p,3D}(q)$. Therefore, *substantial saving of the Coulomb energy in the long-wavelength limit is possible only in the presence of a strong crystalline potential*. In three dimensions (where the Coulomb interaction $V_q = e^2/\epsilon_0 q^2$), the umklapp term $q^2/m^2\langle \hat{A} \rangle$ contributes in the same leading order of powers of q into the J_3 sum rule as the Coulomb term $q^4(n^2/m^2)V_q$ [see Eq. (13)]. The umklapp term is then responsible for the finite width (or ''lifetime'') of the plasmon peak.

In the 2D case, the leading terms are again

$$J_3 = \frac{q^2}{m^2}\langle \hat{A} \rangle + q^4 \frac{n^2}{m^2} V_{q,2D}. \tag{19}$$

Other terms are subleading and proportional to q^4 and q^6 powers. The last term can be analyzed similarly to the 3D case and shown proportional to q^4 at small q. Using the ''cusp theorem'' for the 2D case,[18]

$$S_p = 1 - \frac{C_{2D}}{p^3} + o\left(\frac{1}{p^4}\right), \tag{20}$$

we can expand in powers of q/p for large momenta p and keep the leading term as a function of q/p:

$$\int d^2p V_{p,2D}(\vec{p}\vec{q})^2 [S_{p+q} - S_p] \simeq \int d^2p \frac{V_{p,2D}(\vec{p}\vec{q})^2 C_{2D}}{p^3} \times \left(\frac{1}{[1 + (q/p)^2 + 2(q/p)\cos\theta]^{3/2}} - 1 \right). \tag{21}$$

Due to the angle θ integration, terms proportional to q^3 and other odd powers vanish. For large q, the last term grows no faster than q^4 also [due to the first term in the bracket of Eq. (14)]. In the absence of umklapp scattering ($\langle \hat{A} \rangle = 0$), the density spectrum is given by the expression $\mathrm{Im}\,\chi(q,\omega) \sim [\pi\hbar\omega_{p,2D}(q)/2V_{q,2D}]\delta[\omega - \omega_{p,2D}(q)]$, which satisfies all three sum rules at small q. The maximum possible saving of partial Coulomb energy at long wavelengths is of order $\hbar/2\omega_{p,2D}(q)(q/q_{TF,2D})$ [where $q_{TF,2D} = e^2 m^\star/2\pi\epsilon_0\hbar^2$ is a 2D Thomas-Fermi screening wave vector].

The presence of the umklapp term $q^2/m^2\langle \hat{A} \rangle$ has a much more dramatic effect on the density spectrum in two dimensions, because this term, proportional to q^2, has a leading power of q in the third moment sum rule at small q dominating over the Coulomb term (unlike the 3D case, where the umklapp term has the same power-q dependence as the Coulomb term). The density spectrum cannot even be approximated by a plasma pole expression, and *the plasmon is never really a well-defined excitation in two dimensions in the presence of umklapp scattering*. This means that the umklapp scattering modifies strongly (or ''nonperturbatively'') the spectrum of the density fluctuations, and the spectrum is dominated by multipair and pair excitations rather by a plasmon. Of course, the mere existence of a large value of $\langle \hat{A} \rangle$ is in itself perfectly compatible with a traditional textbook picture, in which the sum rules are satisfied by taking proper account of interband transitions; in such a case there is no *a priori* reason why a plasmon associated with the excitations of the conduction band must automatically be ill defined. However, in a system where umklapp (quasi-momentum-nonconserving) interactions between the Bloch quasiparticles are strong it seems natural that these alone could give rise to a substantial value of $\mathrm{Im}\,\chi$ even below the first band gap; particularly in view of the above remark about the enhanced effect in the 2D case, it is tempting to view the so-called midinfrared peak in the cuprates in this light.

The upper bound on the partial Coulomb energy is still half of the plasmon energy [$\omega_{p,2D}(q) \sim \sqrt{q}$], but the lower bound at small q if $\langle \hat{A} \rangle \neq 0$ is essentially given by

$$\langle V_{c,q} \rangle \geqslant \frac{\hbar}{2} \frac{\omega_{p,2D}(q)}{\sqrt{1 + \dfrac{\langle \hat{A} \rangle}{nm\omega_{p,2D}^2(q)}}}, \tag{22}$$

and so in the limit $q \to 0$,

$$\langle V_{c,q} \rangle \geqslant \frac{\hbar}{2}\omega_{p,2D}^2(q)\left(\frac{nm}{\langle \hat{A} \rangle}\right)^{1/2} \sim q. \tag{23}$$

Therefore, bounds, based on sum rules, are compatible with saving of almost all Coulomb energy $(\hbar/2)\omega_{p,2D}(q)$ in two dimensions when $\langle \hat{A} \rangle \neq 0$.

It is necessary to mention extensive literature (for instance, Ref. 14) using sum rules (in particular, a third moment sum rule) in order to analyze and derive various local-field corrections and approximations of the density response, whereas our goal in this paper is to analyze general constraints on the electron Coulomb energy at small q without relying on any approximation.

It is also interesting to discuss briefly, for comparison, the interaction energy of a many-particle system interacting via a short-range potential ($V_q \to V_0 = \mathrm{const}$, for $q \to 0$). The upper bound on the partial interaction energy $\langle V_{int,q} \rangle$ is given by the ''acoustic mode'' $\hbar/2[V_0^2(nq^2/m)\chi_0(q,0)]^{1/2}$,

$$\frac{\hbar}{2}\left(\frac{V_0 n}{ms}\right) q \geqslant \langle V_{int,q} \rangle, \tag{24}$$

where $\chi_0(q,0)$ is given by the compressibility sum rule[12] $\chi_0(q,0) = n/ms^2$, where s is the velocity of sound. Therefore, the maximum available interaction energy at a long wavelength is insignificant (especially when weighted by the phase volume). For instance, it implies that in most phase transitions in neutral systems (i.e., many-particle systems in-

teracting via a short-range potential) the interaction energy is saved predominantly at short distances.

III. EXPERIMENTAL PROBES

The remainder of this paper is devoted to a brief discussion of the experimental spectroscopies which should, at least in principle, be able to shed light on the origin of the condensation energy in the SC transition, and the inferences which we may currently draw from them (cf. also Ref. 3, Sec. 4.2). For simplicity we will consider explicitly a single-plane cuprate such as Tl-2201, so that in the normal phase there is only one large characteristic length (other than, possibly, the electron mean free path) compared to the quantity q_{TF}^{-1}, namely, the interplane spacing d (typically ~ 10 Å, i.e., $\sim 10-20 q_{TF}^{-1}$) (note notational differences from Ref. 3). The case of a multilayer cuprate such as Bi-2212 is more complicated, since there is now a second ''large'' characteristic length, namely, the intrabilayer spacing ($\sim 3-5$ Å); however, the general pattern of the results is unchanged. In addition, we will assume tetragonal symmetry.

The two spectroscopies which most directly probe the Coulomb energy, or something closely related to it, are electron energy-loss spectroscopy[19] (EELS) and optical reflectivity;[20] in the latter case we shall assume that ellipsometric measurements are possible in an interesting frequency regime[21,22] so that we may deduce the relevant complex dielectric constant without the use of Kramers-Kronig relations. In a bulk isotropic 3D metal the situation is very simple: to the approximation that we neglect multiple scattering and the effect of the ionic cores, the transmission EELS cross section $\sigma(q,\omega)$ is a direct measure of the quantity $V_q^2 \mathrm{Im}\,\chi(q,\omega)$, where $\chi(q,\omega)$ is the ''true'' density susceptibility as defined in Eq. (4). Since in the 3D case the longitudinal dielectric constant $\epsilon_\parallel(q,\omega)$ is identically equal to $1+(e^2/\epsilon_0 q^2)\chi_0(q,\omega)$ with the help of Eq. (11), we find the simple result

$$\sigma(q,\omega) = \mathrm{const}\frac{1}{q^2}\mathrm{Im}\left[-\frac{1}{\epsilon_\parallel(q,\omega)}\right], \tag{25}$$

where the constant is of purely geometrical origin and can be calculated, and the quantity $\mathrm{Im}[-1/\epsilon_\parallel(q,\omega)]$ is usually known as the loss function. This formula is valid for arbitrary q, including values larger than the inverse lattice spacing in the approximation of neglect of the ionic cores direct-scattering effect. If the latter is taken into account, the effect is to multiply the formula (25) by a factor which is, in general, a function of q but not of ω, provided that the latter is small compared to typical core excitation energies (~ 20 eV), and expected to be unaffected by the superconducting transition. Writing out the integrand of Eq. (6) explicitly in terms of $\chi_0(q,\omega)$ [see Eq. (11)] and using the 3D relation between the latter and $\epsilon_\parallel(q,\omega)$, we see that apart from a function of q *the transmission EELS cross section is a direct measure of the Coulomb energy locked up in* $d\vec{q}\,d\omega$.

It is well known that the condensation energy due to the SC transition is extremely small [of order 10^{-4} eV per electron (or per unit cell)] in comparison with the atomic energies (10 eV). It implies stringent requirements on experimental techniques; nevertheless, changes associated with the SC transition were observed by optics at midinfrared frequencies[21,22] of a magnitude sufficient to provide the condensation energy (measured directly by specific-heat measurements). It is worth noting that we do not discuss the changes at frequencies comparable to or lower than a superconducting gap (although these changes, of course, are most remarkable consequences of superconductivity), because the change of Coulomb energy associated with this region of frequencies is negligible (if limited to small momenta $q \ll q_{TF}$). The optical reflectivity measurements[21,22] have enough precision to explore the type of questions discussed in this paper, while it is hoped that transmission EELS can achieve the required accuracy in the near future.

In a 3D bulk metal, ellipsometric optical measurements can measure the complete *transverse* dielectric constant $\epsilon_\perp(q,\omega)$, in the limit $\vec{q}\to 0$, and hence the corresponding ''transverse'' loss function $\mathrm{Im}[-1/\epsilon_\perp(q,\omega)]$. Since in the normal phase, at least, there should be no distinction, in the limit $\vec{q}\to 0$, between $\epsilon_\parallel(q,\omega)$ and $\epsilon_\perp(q,\omega)$, it follows that in this phase the $\vec{q}\to 0$ limit of ''loss functions'' measured by EELS and by optics should coincide. It is a somewhat delicate question, once one renounces reliance on some specific model such as the Fermi-liquid one, which is to count as ''the $\vec{q}\to 0$ limit''; in addition to the obvious scale q_{TF} or q_F, it is not immediately clear whether the inverse electron mean free path $1/l$ is a relevant quantity. However, it is plausible that this quantity may not play a major role for ω in the midinfrared region, thus we shall tentatively take ''$\vec{q}\to 0$'' to mean in the 3D case $q \ll q_{TF}, q_F$.

Some care is needed in adapting the above results to the case of a layered material such as cuprates, even if we specialize (as we shall) in the limit $q_z \ll 1/d$, which is automatically fulfilled in optical experiments and may be satisfied in transmission EELS by a suitable choice of geometry. In the context of EELS experiments, we now have to distinguish between the cases $qd \ll 1$ and $qd \gg 1$ (where q is the ab-plane component of the momentum loss). In the former case the 3D bulk formula [Eq. (25)] applies unchanged, provided that $\epsilon_\parallel(q,\omega)$ is defined to be the tensor component of the longitudinal dielectric constant corresponding to current flow in the ab plane; note, in particular, that, at least in the normal phase, we expect that $\mathrm{Im}[-1/\epsilon_\parallel(q,\omega)]$ is nearly independent of $|\vec{q}|$ in the limit $\vec{q}\to 0$. In the opposite limit $qd \gg 1$, we could choose to continue to use the 3D formula [Eq. (25)], but we would then find that the $\epsilon_\parallel(q,\omega)$ so defined has a strong explicit dependence on q. A much more natural convention in this limit is to treat the scattering as occurring independently from the different CuO_2 planes, and to define a two-dimensional (''per-plane'') version $\chi_0^{(2)}(q,\omega)$ of $\chi_0(q,\omega)$, or equivalently a quantity (cf. Ref. 3), Sec. 4.1)

$$K(q,\omega) \equiv \frac{1}{2\epsilon_0 q^2}\chi_0^{(2)}(q,\omega) = \frac{d}{2}[\epsilon_\parallel(q,\omega)-\epsilon_b], \tag{26}$$

where $\epsilon_\parallel(q,\omega)$ is the ''natural'' definition of the 3D bulk ab-plane dielectric constant, i.e. the quantity which relates

the local polarization to the local field, and is expected to be nearly constant over a range $q \gg 1/d$, and ϵ_b is its ''background'' (off-plane) contribution (cf. Ref. 3). With this definition we find that, apart from factors depending only on q, both the transmission EELS cross section and the (single-plane) Coulomb energy locked up in the range $d\vec{q}d\omega$ are proportional to the quantity

$$-\mathrm{Im}\left[\frac{1}{1+qK(q,\omega)/\epsilon_{sc}}\right], \tag{27}$$

where ϵ_{sc} is the dielectric constant which screens the Coulomb interactions of the in-plane electrons (note that, in general, ϵ_{sc} is not equal to ϵ_b). Thus, just as in the bulk 3D case, the transmission EELS cross section is a direct measure of the Coulomb energy locked up in the relevant region of (q,ω) space; note, however, that Eq. (27) introduces an extra explicit q dependence that is absent in the bulk case. This subtlety seems to have been overlooked in the analysis of existing normal-state EELS data[19] on the cuprates, where it seems to be assumed that even in the regime $qd \gg 1$ EELS experiments measure the ''bulk'' $\epsilon_{\parallel}(q,\omega)$. In the intermediate case $(qd \sim 1)$ a similar analysis using Eq. (2) is possible, but will not be given here.

In the case of optical experiments on the cuprates (with the sample surface assumed to lie in the ab plane) we always have $qd \ll 1$, $q_z d \ll 1$, and thus at first sight we would expect the complex dielectric constant inferred from reflectivity measurements to be identical to the $\epsilon_{\parallel}(q,\omega)$ inferred, via Eqs. (25) and (27), from EELS experiments in the limit $\vec{q} \to 0$. Existing measurements in the normal phase do appear to be consistent with this prediction. In the superconducting state, however, there are three complications: first, it is not clear that even in the midinfrared regime the Cooper-pair radius ξ_0 is not a relevant length scale, so that it may be illegitimate to use the ''true'' $\vec{q} \to 0$ $(q\xi_0 \ll 0)$ behavior observed in the optics to infer the behavior in the regime $1/\xi_0 \ll q \ll q_{TF}$. Secondly, it is not completely obvious, particularly in the former limit $q\xi_0 \ll 1$, that the finite-frequency longitudinal and transverse dielectric constants must be equal in the superconducting state. This latter complication may be somewhat mitigated by a third consideration, namely, that at the non-normal angles of incidence necessary in the ellipsometric technique, in layered material the measured quantity is not simply $\epsilon_{\perp}$ but a combination of $\epsilon_{\perp}$ and $\epsilon_{\parallel}$. We will not attempt to develop those points further, but rather use them to draw the conclusion that, while the spectacular changes observed[21,22] in the optically measured dielectric constants of the cuprates at and below the superconducting transition are strongly suggestive, a quantitative test of any scenario (such as the midinfrared one of Ref. 3) that attributes the energy saving largely to a regime of q, small compared to q_{TF} but large compared to ξ_0 (and $1/d$), will require accurate transmission EELS data taken across the transition, something which (as regards the midinfrared regime of frequencies) does not, to our knowledge, exist at present.

We finally and directly address the question of which of the three terms in Eq. (1) is (are) reduced in the superconducting transition. If it is the second (Coulomb) one, in what regime(s) of q and ω does the saving predominantly occur? Part of the interest in these questions is that as we have seen above, a conjectured answer can be tested directly in transmission EELS experiments.

We start by recalling a well-known result: since the original (''true'') Hamiltonian of the N-body system (the nonrelativistic limit of the Dirac Hamiltonian) is composed exclusively of kinetic-energy terms and (unscreened) Coulomb interactions, the virial theorem immediately tells us that the change in total kinetic energy (of electrons and ions) must be exactly minus half that of the total Coulomb energy (electron-electron, electron-ion, and ion-ion), and thus Coulomb energy must be saved in the superconducting transition (and indeed in any other phase transition into a lower-energy state). While this conclusion is very generic and rigorous, it is not usually regarded as providing much insight into the ''mechanism'' of superconductivity in the cuprates (or for that matter in the classic superconductors) since the term ''mechanism'' is often held to refer to a low-energy effective Hamiltonian in which the separation of the original kinetic and potential energies may no longer be explicit. However, the ''intermediate-level'' effective Hamiltonian (1) is sufficiently close to the original truly first-principles one in which the virial-theorem result for the latter might at least suggest that it is one or both of the last two terms of Eq. (1) that are saved.[23]

If we assume for the sake of argument that Coulomb energy is indeed saved, then where in the space of q and ω is it saved? It is at this point that the sum-rule arguments of Sec. II come into their own. For convenience we reproduce here the three relevant sum rules with terms of relative order q^4 and h,igher omitted on the right-hand sides, with notation as in Eqs. (7)–(9):

$$J_{-1} = \frac{1}{V_q}, \tag{28}$$

$$J_1 = \frac{nq^2}{m}, \tag{29}$$

$$J_3 = \frac{n^2 q^4}{m^2} V_q + \frac{q^2}{m^2}\langle \hat{A} \rangle, \tag{30}$$

where in the case of a layered system, V_q is given by Eq. (2) and tends to $e^2/(\epsilon_0 \epsilon_\infty q^2)$ for $qd \ll 1$ (and $q_z d \ll 1$, see above) and to $e^2/(2\epsilon_0 \epsilon_\infty q)$ for $qd \gg 1$, and $\hat{A}$ is defined below Eq. (13). The $(T=0\text{-K})$ contribution $\langle V_{c,q} \rangle$ to the expectation value of the Coulomb energy from wave vector $\vec{q}$ is, up to a factor, just J_0 [see Eq. (6)].

The arguments of Sec. II now show that the maximum change of $\langle V_{c,q} \rangle$ in the ''essentially 3D'' regime $qd \ll 1$ is proportional to a finite fraction of the ''3D plasmon energy'' but weighted by q^2 of the phase space (d^3q), and hence the maximum total saving possible from this regime is very small. On the other hand, the contribution from the regime $1/d \ll q \ll q_{TF}$ [where the truncated forms, Eqs. (28)–(30), are still a good approximation] can be of order $q^{1/2}$ [cf. the con-

clusion after Eq. (23)], provided the quantity $\langle \hat{A} \rangle$ is substantial, while the phase space allows a significant saving of Coulomb energy since $dq_z d^2q \sim (2\pi/d) 2\pi q dq$ and $1/d \ll q \ll q_{TF}$. Thus, in a quasi-2D system with a large value of $\langle \hat{A} \rangle$, substantial energy is, in principle, available for saving in this small-q regime. To estimate the value of $\langle \hat{A} \rangle$ we return for a moment to the limit $qd \ll 1$ and refer to the normal-state optical loss-function data: using Eqs. (28)–(30) with the appropriate form $[e^2/(\epsilon_0 \epsilon_\infty q^2)]$ of V_q, we see that the quantity $\langle \hat{A} \rangle_n$ is given, in natural units of $n^2e^2/(\epsilon_0 \epsilon_\infty)$, by the expression ("$n$" is the normal-state value)

$$\frac{\langle \hat{A} \rangle_n}{n^2e^2/(\epsilon_0 \epsilon_\infty)} = J_3 - J_1^2/J_{-1}. \tag{31}$$

Although a strict evaluation of the right-hand side of Eq. (31) from the optical loss function requires us to know the effective frequency cutoff [since at high frequencies there will be contributions to $\epsilon(q,\omega)$ from "core" processes not described by Eq. (1)], it is clear that the mere existence of a midinfrared (MIR) peak extending over an order of magnitude in frequency already implies that it at least of order of one. Thus, a very appreciable fraction of the Coulomb energy locked up, in the normal state, in the low-q, MIR-frequency regime is, in principle, available for saving in the SC transition (or indeed in other possible phase transitions). Whether it is saved and to what extent, as is, in fact, postulated in the "MIR scenario" of Ref. 3, depends of course on the cost of the formation of the Cooper pairs in kinetic and/or static lattice energies. Actually, rather than asking as above for the fraction of the Coulomb energy which is, in principle, available for saving (something which is not that significant if the original value is itself small), it may be more informative to estimate the relative contribution of the small-q regime in two and three dimensions *for a given change in* χ_0 due to the phase transition. Taking into account both the phase-space factor and the extra factor of q in the denominator of the expression (27) in the 2D case, we find that in the regime where (qK) dominates, the contribution of small q is proportional to q^2 in three dimensions but to a constant in two dimensions, so that the relative importance of the long-wavelength regime is vastly enhanced in the 2D case.

On the experimental front, it has to be said that as noted in Ref. 21, the optical data, if extrapolated into the relevant $(q\xi_0 \gg 1)$ regime with several other assumptions, indicate rather the opposite, i.e., that the Coulomb energy associated with the MIR regime actually increases in the SC state. However, because of the various considerations noted above, this extrapolation may be problematic, and a definitive test of the MIR hypothesis must await quantitative transmission EELS measurements across the superconducting transition [or a better theoretical understanding of the generic q dependence of $\epsilon(q,\omega)$ in the SC state].

In sum, we analyzed the electron-electron Coulomb energy in the presence of the periodic lattice potential using various sum rules for the density-density response function. We believe that in this paper we have made it plausible that two specific properties of cuprates, namely, (i) the layered (two-dimensional) structure of the CuO_2 planes and (ii) the occurrence of a broad and strong peak in the optical loss function can be essential ingredients in the occurrence of high-temperature superconductivity in these materials by conspiring to save the small-q $(q < q_{TF})$ part of electron-electron Coulomb energy.

ACKNOWLEDGMENTS

We are grateful for discussions with Dirk van der Marel and Neil Ashcroft. A.J.L. acknowledges the support from Grant No. NSF-DMR-99-86199. M.T. was supported by EPSRC.

[1] J. R. Schrieffer, *Theory of Superconductivity* (Benjamin, New York, 1964).

[2] Although the phonon mechanism for superconductivity is not ruled out irrevocably for the cuprate superconductors, we explore the general scenario of an electronic mechanism, at least, as a matter of principle.

[3] A. J. Leggett, Proc. Natl. Acad. Sci. U.S.A. **96**, 8365 (1999).

[4] P. W. Anderson, *The Theory of Superconductivity in the High T_c Cuprates* (Princeton University, Princeton, NJ, 1997); D. J. Scalapino and S. R. White, Phys. Rev. B **58**, 8222 (1998); E. Demler and S.-C. Zhang, Nature (London) **396**, 733 (1998); V. Emery, S. Kivelson, and O. Zachar, Physica C **282-287**, 174 (1997); M. R. Norman, M. Randeria, B. Janko, and J. C. Campuzano, Phys. Rev. B **61**, 14 742 (2000).

[5] The separation into conduction and core electrons is assumed, where the core electrons contribute to the screening constant ϵ_∞ and to the effective periodic potential. For simplicity, we do not consider here possible effects of anisotropic (or q-dependent) screening by the core electrons. We also neglect the effects of disorder.

[6] Throughout this paper we do not use the reduced zone scheme (in order not to be based on this one-electron approximation in the presence of strong electron-electron interaction), and all momenta are real (not quasimomenta). Here the terms umklapp and lattice scattering are used interchangeably to denote the static effects of the lattice potential. In the reduced zone scheme, one can distinguish between "lattice" and "umklapp" effects, the former being the real momentum-nonconserving effects of the static lattice, while the latter are due to effects of electron-electron Coulomb interaction (rigorously conserving real momentum, but in general not conserving quasimomentum).

[7] In the analysis of upper and lower bounds on the electron Coulomb energy in the text of the paper we omit the constant ϵ_∞ in order to keep expressions simple. The modifications due to the dielectric constant ϵ_∞ are trivial.

[8] A. L. Fetter, Ann. Phys. (N.Y.) **88**, 1 (1974).

[9] The sum-rule analysis is applicable only to homogeneous phases of cuprates.

[10] Of course, in a system with crystalline periodicity the coordinate-

space form of χ is a function of the center-of-mass coordinate as well as the relative one, and the momentum-space form is correspondingly a matrix quantity $\chi(q+\kappa_i, q+\kappa_j)$, where $\kappa_{i,j}$ are reciprocal-lattice vectors. The quantity discussed everywhere in the text is in this notation $\chi(q,q)$.

[11] We take here a definition of the density correlation function such that for positive frequencies ($\omega>0$) the imaginary part of the susceptibility is positive $\mathrm{Im}\,\chi(q,\omega)>0$, and also $\mathrm{Re}\,\chi(q,0)>0$. This definition is different by way of a minus sign from the definition in Ref. 12, Sec. 4.1. See also below that χ_0, defined in Eq. (11) is different only by a minus sign from χ_{sc}, defined in Ref. 12, Sec. 4.1.

[12] P. Nozieres and D. Pines, *The Theory of Quantum Liquids* (Perseus Books, Cambridge, MA, 1999).

[13] N. Mihara and R. D. Puff, Phys. Rev. **174**, 221 (1968).

[14] N. Iwamoto, Phys. Rev. A **30**, 3289 (1984); K. Sturm, Phys. Rev. B **52**, 8028 (1995).

[15] In different contexts, upper and lower bounds on the structure factor for classical and quantum liquids were discussed in the literature, for instance, see Ref. 13; N.D. Mermin, Phys. Rev. **171**, 272 (1968), and references therein.

[16] Here we give arguments that the expectation value of $\langle\hat{A}\rangle$ is positive. Namely, if $U(x)\sim U_0\cos(\kappa x)$, the electron density minimizes the interaction energy by taking the form $\rho(x)\sim-\rho_0\cos(\kappa x)$. In other words, the electron-lattice energy $\langle\Sigma_\kappa U_{-\kappa}\widehat{\rho_\kappa}\rangle$ (without the zero-momentum "background" harmonic $\kappa=0$) is always negative, because the harmonics of the density ρ_κ are shifted by the phase π to the harmonics of the lattice potential $U_{-\kappa}$. In addition, from inequalities of Eqs. (10), (18), and (23) it follows that the expectation value $\langle\hat{A}\rangle$ is positive.

[17] It can be convenient for the analysis to rewrite Eq. (14) as $(n/m^2)\int d^Dp\,V_p(\vec{p}\vec{q})^2[S_{p+q}-S_p]$ by making a change of the variable $p'=p+q$ in the first term in the bracket of Eq. (14).

[18] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957); J. C. Kimball, Phys. Rev. A **7**, 1648 (1973); A. K. Rajagopal and J. C. Kimball, Phys. Rev. B **15**, 2819 (1977).

[19] N. Nucker, H. Romberg, S. Nakai, B. Scheerer, J. Fink, Y. F. Fan, and Z. X. Zhao, Phys. Rev. B **39**, 12 379 (1989); Yun-Yu Wang, Goufu Feng, and A. L. Ritter, *ibid.* **42**, 420 (1990).

[20] Raman scattering is also relevant, but is a less direct probe and will not be discussed here.

[21] H. J. A. Molegraaf, C. Presura, D. van der Marel, P. H. Kes, and M. Li, Science **295**, 2239 (2002).

[22] M. Rubhausen, A. Gozar, M. V. Klein, P. Guptasarma, and D. G. Hinks, Phys. Rev. B **63**, 224514 (2001).

[23] Of course we cannot apply the virial theorem directly to the Hamiltonian [Eq. (1)], because the lattice potential U_κ is not, in general, a sum of simple Coulomb terms.