

Theory of many-fermion systems. II. The case of Coulomb interactions

D. Belitz and F. Evers*

Department of Physics and Materials Science Institute, University of Oregon, Eugene, Oregon 97403

T. R. Kirkpatrick

Institute for Physical Science and Technology, and Department of Physics, University of Maryland, College Park, Maryland 20742

(Received 26 March 1998)

In a recent paper a general field-theoretical description of many-fermion systems with short-range interactions has been developed. Here we extend this theory to the case of disordered electrons interacting via a Coulomb potential. A detailed discussion is given of the Ward identity that controls the soft modes in the system, and the generalized nonlinear σ model for the Coulombic case is derived and discussed. [S0163-1829(98)10239-4]

I. INTRODUCTION

The theory of many-fermion systems is central to our understanding of condensed matter, as well as of nuclei and certain astrophysical systems. Of particular interest in a condensed-matter context are the universal properties of many-electron systems, i.e., phenomena at long wavelength and small-frequency scales that are independent of the material's detailed microscopic structure. Historically, there have been two important avenues to this problem: Landau's phenomenological Fermi-liquid theory,¹ and the microscopic many-body perturbation theory or Feynman diagram approach.² The latter soon was generalized to include the scattering of electrons by static impurities,³ and it has led to many important insights concerning the nature of interacting disordered electron systems. For instance, it was used to show that the combined effects of disorder and interactions lead to the nonanalyticities in the frequency and wave-number dependence of both thermodynamic and transport properties of disordered metals that have become known as "weak-localization effects."^{4,5}

There are, however, many interesting phenomena for which an approach based on many-body perturbation theory is not feasible. An example is the Anderson-Mott metal-insulator transition that the electrons undergo with increasing disorder strength.⁶ This quantum phase transition is best studied by means of effective field theories and the use of the renormalization group, an approach that was pioneered by Wegner.⁷ Also, many-body perturbation theory is ultimately unsatisfactory as a tool even in the metallic phase, despite its impressive successes. One major problem is that, within perturbation theory, it is not clear whether the weak-localization effects are actually the leading nonanalyticities. Furthermore, universal phenomena such as the weak-localization effects arise from the presence of soft modes in the problem, which many-body perturbation theory is not well suited to deal with. The softness or masslessness of these modes is the result of symmetries that are often not explicit in the usual perturbative formalism, and therefore soft modes arise apparently accidentally as a result of complicated cancellations, rather than being manifestly built into the formalism. Finally, it is unsatisfactory to have entirely different approaches to

the universal properties of the various phases on the one hand, and to those of the transitions between these phases on the other.

The concepts for overcoming these problems exist; they consist of a systematic application of renormalization-group ideas. Historically, the renormalization group has been mainly associated with critical phenomena at continuous phase transitions, for which the importance of symmetries and soft modes has been universally appreciated. While it has always been clear in principle and occasionally has been emphasized⁸ that the renormalization group, rather than just being a tool for studying critical phenomena, allows for a unified description of both phases and phase transitions, these ideas have not been widely appreciated. Only very recently has there been the beginning of a paradigm shift in this respect. For instance, starting with Shankar's work,⁹ there has been much activity recently on the derivation of clean Fermi-liquid theory as a stable renormalization-group fixed point.¹⁰ These methods, however, have proven very hard to generalize to the case of quenched disorder.

In a previous paper, to be referred to as I,¹¹ two of us have developed an effective field theory for many-electron systems that is particularly suitable for dealing with disordered systems. This theory, which is formulated in terms of classical matrix fields, allows for a systematic separation of soft and massive modes, and the latter can be integrated out in a simple approximation to yield an effective theory for the soft modes. The theory also allows for the clean limit to be taken, but the soft-mode structure in that case turns out to be more complex, which gives the effective theory fewer advantages over traditional approaches than is the case in the presence of disorder. This theory has been used, *inter alia*, to prove that the well-known weak-localization effects are indeed the leading nonanalyticities, to provide a technically satisfactory derivation of Finkel'stein's generalization¹² of Wegner's nonlinear σ model,⁷ and to derive a previously unknown nonanalyticity in the spin susceptibility of clean Fermi liquids.^{13,11}

In I, the effective classical field theory was developed for fermions that interact via a short-range potential, assuming that the underlying Coulomb potential had been screened at the level of the basic fermionic theory. In the present paper,

we show how to handle a long-range interaction entirely within the framework of the classical matrix field theory, focussing on the disordered case. We give a detailed discussion of the Ward identity that controls the structure of the soft modes, and we derive Finkel'stein's generalized nonlinear σ model¹² for the Coulombic case.

The outline of our paper is as follows. In Sec. II we recall the matrix field theory of I, and generalize it slightly to allow for a Coulomb interaction. We show that the saddle point considered in I remains valid, and that an expansion about the saddle point to Gaussian order produces RPA-type screening. In Sec. III we perform a symmetry analysis, and derive and discuss the Ward identity that controls the soft modes in the system. In Sec. IV we show that integrating out the massive modes in the simplest approximation that respects the Ward identity leads to Finkel'stein's model. In Sec. V we conclude by discussing our results, and give, in particular, a discussion of the accuracy of the σ model.

II. MATRIX FIELD THEORY

A. Grassmannian field theory

Our model and basic theoretical setup is the same as in I. We will therefore restrict ourselves to discussing the changes that are necessary to accommodate a Coulomb interaction.

The action is given by

$$S = - \int dx \sum_{\sigma} \bar{\psi}_{\sigma}(x) \partial_{\tau} \psi_{\sigma}(x) + S_0 + S_{\text{dis}} + S_{\text{int}}. \quad (2.1a)$$

Here the $\bar{\psi}$ and ψ are Grassmann valued fields, and we use a $(d+1)$ -vector notation, with $x = (\mathbf{x}, \tau)$, and $\int dx = \int_V d\mathbf{x} \int_0^{\beta} d\tau$. \mathbf{x} denotes position, τ imaginary time, V the system volume, $\beta = 1/T$ the inverse temperature, σ the spin label; and we use units such that $\hbar = k_B = e^2 = 1$. S_0 (together with the time derivative term) and S_{dis} describe free fermions and their interaction with a static random potential, respectively, and have been defined in I. The random potential we assume to be Gaussian distributed, and we employ the replica trick to handle it. S_{int} describes a Coulomb interaction, which in the replicated theory (which we denote by a tilde) takes the form

$$\begin{aligned} \tilde{S}_{\text{int}} = \sum_{\alpha} \tilde{S}_{\text{int}}^{\alpha} = & - \frac{1}{2} \sum_{\alpha} \int dx_1 dx_2 \sum_{\sigma_1, \sigma_2} v(\mathbf{x}_1 - \mathbf{x}_2) \\ & \times \delta(\tau_1 - \tau_2) \bar{\psi}_{\sigma_1}^{\alpha}(x_1) \bar{\psi}_{\sigma_2}^{\alpha}(x_2) \psi_{\sigma_2}^{\alpha}(x_2) \psi_{\sigma_1}^{\alpha}(x_1). \end{aligned} \quad (2.1b)$$

Here α is the replica index, and

$$v(\mathbf{x}) = 1/|\mathbf{x}| \quad (2.1c)$$

is the Coulomb potential. For the dimensionalities of interest, $d=2,3$, its Fourier transform is

$$v(\mathbf{q}) = (1 - \delta_{\mathbf{q},0}) \frac{2^{d-1} \pi}{|\mathbf{q}|^{d-1}}, \quad (2.1d)$$

where the factor $1 - \delta_{\mathbf{q},0}$ represents a uniform positive background charge that ensures charge neutrality.

We now introduce a momentum cutoff λ , and rewrite the interacting part of the action as

$$\tilde{S}_{\text{int}}^{\alpha} = \tilde{S}_{\text{int}}^{\alpha(1)} + \tilde{S}_{\text{int}}^{\alpha(2)} + \tilde{S}_{\text{int}}^{\alpha(3)}, \quad (2.2a)$$

where

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(1)} = & - \frac{T}{2} \sum_{\sigma_1 \sigma_2} \sum_{k,p} \sum_q' v(\mathbf{q}) \bar{\psi}_{\sigma_1}^{\alpha}(k) \bar{\psi}_{\sigma_2}^{\alpha}(p+q) \\ & \times \psi_{\sigma_2}^{\alpha}(p) \psi_{\sigma_1}^{\alpha}(k+q), \end{aligned} \quad (2.2b)$$

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(2)} = & - \frac{T}{2} \sum_{\sigma_1 \sigma_2} \sum_{k,p} \sum_q' v(\mathbf{p}-\mathbf{k}) \Theta(|\mathbf{p}-\mathbf{k}|-\lambda) \\ & \times \bar{\psi}_{\sigma_1}^{\alpha}(k) \bar{\psi}_{\sigma_2}^{\alpha}(p+q) \psi_{\sigma_2}^{\alpha}(k+q) \psi_{\sigma_1}^{\alpha}(p), \end{aligned} \quad (2.2c)$$

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(3)} = & - \frac{T}{2} \sum_{\sigma_1 \neq \sigma_2} \sum_{k,p} \sum_q' v(\mathbf{k}+\mathbf{p}) \Theta(|\mathbf{p}+\mathbf{k}|-\lambda) \\ & \times \bar{\psi}_{\sigma_1}^{\alpha}(-k) \bar{\psi}_{\sigma_2}^{\alpha}(k+q) \psi_{\sigma_2}^{\alpha}(-p+q) \psi_{\sigma_1}^{\alpha}(p). \end{aligned} \quad (2.2d)$$

As in real space, we use a $(d+1)$ -vector notation with $k = (\mathbf{k}, \omega_n)$, where $\omega_n = 2\pi T(n+1/2)$ is a fermionic Matsubara frequency. The prime on the q summation indicates that only momenta up to the momentum cutoff λ are integrated over. While the long-wavelength, small-frequency phenomena we are interested in do not in general depend on λ , the choice of this cutoff will be important for the range of validity of the final effective theory. We will come back to this point in Secs. III and V below. Equations (2.2) represent the same phase-space decomposition as in Eqs. (2.8) of I, except that we have explicitly inserted the step functions in Eqs. (2.2c,2.2d), because the small-wave-number part of v is already contained in Eq. (2.2b). In I the overcounting that resulted from not having the step functions explicitly present was of no consequence, since v in that case was not singular in the small-wave-number limit. For a Coulomb interaction, more care must be taken.

We next introduce spinors

$$\psi_n^{\alpha}(\mathbf{x}) = \begin{pmatrix} \psi_{n\uparrow}^{\alpha}(\mathbf{x}) \\ \psi_{n\downarrow}^{\alpha}(\mathbf{x}) \end{pmatrix}, \quad (2.3a)$$

and their Fourier transforms

$$\psi^{\alpha}(k) \equiv \psi_n^{\alpha}(\mathbf{k}) = \begin{pmatrix} \psi_{n\uparrow}^{\alpha}(\mathbf{k}) \\ \psi_{n\downarrow}^{\alpha}(\mathbf{k}) \end{pmatrix}, \quad (2.3b)$$

as well as their adjoints, $\bar{\psi}^{\alpha}(k)$, and a scalar product in spinor space, $(\bar{\psi}, \psi) = \bar{\psi} \cdot \psi$, where the dot denotes the matrix product. Then we can write the interaction term as

$$\tilde{S}_{\text{int}}^{\alpha} = \tilde{S}_{\text{int}}^{\alpha(s)} + \tilde{S}_{\text{int}}^{\alpha(t)} + \tilde{S}_{\text{int}}^{\alpha(3)}, \quad (2.4a)$$

with

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(s)} = & -\frac{T}{2} \sum_{k,p} \sum'_q \Gamma_{k,p}^{(s)}(q) (\psi^\alpha(k), s_0 \psi^\alpha(k+q)) \\ & \times (\psi^\alpha(p+q), s_0 \psi^\alpha(p)), \end{aligned} \quad (2.4b)$$

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(t)} = & -\frac{T}{2} \sum_{k,p} \sum'_q \Gamma_{k,p}^{(t)}(q) \sum_{i=1}^3 (\psi^\alpha(k), s_i \psi^\alpha(k+q)) \\ & \times (\psi^\alpha(p+q), s_i \psi^\alpha(p)). \end{aligned} \quad (2.4c)$$

Here $s_j = i\sigma_j$, with $\sigma_{1,2,3}$ the Pauli matrices, and $s_0 = \sigma_0$ is the 2×2 identity matrix. We have also defined the singlet (s) and triplet (t) interaction amplitudes

$$\Gamma_{k,p}^{(t)}(q) = \frac{1}{2} v(\mathbf{p}-\mathbf{k}) \Theta(|\mathbf{p}-\mathbf{k}|-\lambda), \quad (2.5a)$$

and

$$\Gamma_{k,p}^{(s)}(q) = v(\mathbf{q}) - \Gamma_{k,p}^{(t)}(q). \quad (2.5b)$$

In addition we define the Cooper channel or $2k_F$ -scattering amplitude,

$$\Gamma_{k,p}^{(c)}(q) = v(\mathbf{k}+\mathbf{p}) \Theta(|\mathbf{k}+\mathbf{p}|-\lambda). \quad (2.5c)$$

These expressions are the same as the corresponding ones in I, with the exception of the momentum restrictions in the effective interaction potentials $\Gamma_{k,p}^{(s,t,c)}$ discussed above. Next we project the modes in S_{int} onto density modes in the particle-hole and particle-particle channels that were defined in I. The result of this procedure, which was explained in Appendix A of I, is

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(s)} = & -\frac{T}{2} \sum_{k,p} \sum'_q \Gamma^{(s)}(\mathbf{q}) (\psi^\alpha(k), s_0 \psi^\alpha(k+q)) \\ & \times (\psi^\alpha(p+q), s_0 \psi^\alpha(p)), \end{aligned} \quad (2.6a)$$

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(t)} = & -\frac{T}{2} \Gamma^{(t)} \sum_{k,p} \sum'_q \sum_{i=1}^3 (\psi^\alpha(k), s_i \psi^\alpha(k+q)) \\ & \times (\psi^\alpha(p+q), s_i \psi^\alpha(p)), \end{aligned} \quad (2.6b)$$

$$\begin{aligned} \tilde{S}_{\text{int}}^{\alpha(3)} = & -\frac{T}{2} \Gamma^{(c)} \sum_{\sigma_1 \neq \sigma_2} \sum_{k,p} \sum'_q \bar{\psi}_{\sigma_1}^\alpha(k) \bar{\psi}_{\sigma_2}^\alpha(-k+q) \\ & \times \psi_{\sigma_2}^\alpha(p+q) \psi_{\sigma_1}^\alpha(-p). \end{aligned} \quad (2.6c)$$

Here

$$\Gamma^{(s)}(\mathbf{q}) = v(\mathbf{q}) - \Gamma^{(t)}, \quad (2.6d)$$

and $\Gamma^{(t)}$ and $\Gamma^{(c)}$ are numbers that result from integrating over the wave vectors in Eqs. (2.5) as explained in Appendix A of I. Notice that as a result of this procedure, $\Gamma^{(t)}$ and $\Gamma^{(c)}$ for clean electrons depend logarithmically on the cutoff λ , and diverge as $\lambda \rightarrow 0$. For the disordered case, the logarithmic singularity is protected both by λ and by the disorder. This singularity, which is a consequence of the Coulomb interaction, is the reason for our modification of the procedure employed in I.

B. Composite variables: matrix field theory

After completing the phase-space decomposition and projecting onto densities we are in a position to reformulate the theory in terms of composite variables. This proceeds in exact analogy to I, and we therefore only quote the result. The partition function \tilde{Z} for the replicated theory is written as an integral with respect to two matrix fields, Q and $\tilde{\Lambda}$,

$$\tilde{Z} = \int D[Q] D[\tilde{\Lambda}] e^{\mathcal{A}[Q, \tilde{\Lambda}]}, \quad (2.7)$$

with an effective action

$$\begin{aligned} \mathcal{A}[Q, \tilde{\Lambda}] = & \mathcal{A}_{\text{dis}}[Q] + \mathcal{A}_{\text{int}}[Q] + \frac{1}{2} \text{Tr} \ln(G_0^{-1} - i\tilde{\Lambda}) \\ & + \int d\mathbf{x} \text{tr}[\tilde{\Lambda}(\mathbf{x})Q(\mathbf{x})]. \end{aligned} \quad (2.8)$$

The matrix field Q corresponds to expressions that are bilinear in the fermionic fields ψ and $\bar{\psi}$. Correspondingly, Q carries two frequency indices n and m , and two replica indices α and β . Each matrix element $Q_{nm}^{\alpha\beta}$ is an element of $Q \times Q$, with Q the quaternion field. $\tilde{\Lambda}$ is an auxiliary field the technical role of which is to constrain the products of fermionic fields to the Q . It is convenient to expand the $Q_{nm}^{\alpha\beta}$ in a spin-quaternion basis,

$$Q_{nm}^{\alpha\beta} = \sum_{r,i=0}^4 i_r Q_{nm}^{\alpha\beta} \tau_r \otimes s_i, \quad (2.9)$$

and analogously for $\tilde{\Lambda}$. Here $\tau_0 = s_0 = \mathbb{1}_2$ with $\mathbb{1}_2$ the 2×2 unit matrix, and $\tau_j = -s_j = -i\sigma_j$ ($j=1,2,3$), where the σ_j are the Pauli matrices. The properties of Q and $\tilde{\Lambda}$ have been derived in I, and for completeness we list them again in Appendix A. In Eq. (2.8) and in what follows, Tr is a trace over all degrees of freedom, including an integral over \mathbf{x} , while tr is a trace over all discrete degrees of freedom that are not shown explicitly.

$$G_0^{-1} = -\partial_\tau + \nabla^2/2m + \mu, \quad (2.10)$$

is the inverse of the free-electron Green operator, and it is clear from the structure of the $\text{Tr} \ln$ term in Eq. (2.8) that the physical interpretation of the field $\tilde{\Lambda}$ is that of a self-energy. The contributions \mathcal{A}_{dis} and \mathcal{A}_{int} to the action read

$$\mathcal{A}_{\text{dis}}[Q] = \mathcal{A}_{\text{dis}}^{(1)}[Q] + \mathcal{A}_{\text{dis}}^{(2)}[Q], \quad (2.11a)$$

$$\mathcal{A}_{\text{dis}}^{(1)}[Q] = \frac{-1}{2\pi N_F \tau_1} \int d\mathbf{x} [\text{tr} Q(\mathbf{x})]^2, \quad (2.11b)$$

$$\mathcal{A}_{\text{dis}}^{(2)}[Q] = \frac{1}{\pi N_F \tau_{\text{rel}}} \int d\mathbf{x} \text{tr}[Q(\mathbf{x})]^2, \quad (2.11c)$$

with τ_{rel} the single-particle relaxation time and τ_1 a related scattering time defined in I, and

$$\mathcal{A}_{\text{int}}[Q] = \mathcal{A}_{\text{int}}^{(s)} + \mathcal{A}_{\text{int}}^{(t)} + \mathcal{A}_{\text{int}}^{(c)}, \quad (2.12a)$$

$$\begin{aligned} \mathcal{A}_{\text{int}}^{(s)} &= \frac{T}{2} \int d\mathbf{x} d\mathbf{y} \sum_{r=0,3} (-1)^r \sum_{n_1, n_2, m} \sum_{\alpha} \Gamma^{(s)}(\mathbf{x}-\mathbf{y}) \\ &\times \{\text{tr}[(\tau_r \otimes s_0) \mathcal{Q}_{n_1, n_1+m}^{\alpha\alpha}(\mathbf{x})]\} \\ &\times \{\text{tr}[(\tau_r \otimes s_0) \mathcal{Q}_{n_2+m, n_2}^{\alpha\alpha}(\mathbf{y})]\}, \end{aligned} \quad (2.12b)$$

$$\begin{aligned} \mathcal{A}_{\text{int}}^{(t)} &= \frac{T\Gamma^{(t)}}{2} \int d\mathbf{x} \sum_{r=0,3} (-1)^r \sum_{n_1, n_2, m} \sum_{\alpha} \\ &\times \sum_{i=1}^3 \{\text{tr}[(\tau_r \otimes s_i) \mathcal{Q}_{n_1, n_1+m}^{\alpha\alpha}(\mathbf{x})]\} \\ &\times \{\text{tr}[(\tau_r \otimes s_i) \mathcal{Q}_{n_2+m, n_2}^{\alpha\alpha}(\mathbf{x})]\}, \end{aligned} \quad (2.12c)$$

$$\begin{aligned} \mathcal{A}_{\text{int}}^{(c)} &= \frac{T\Gamma^{(c)}}{2} \int d\mathbf{x} \sum_{r=1,2} \sum_{n_1, n_2, m} \\ &\times \sum_{\alpha} \{\text{tr}[(\tau_r \otimes s_0) \mathcal{Q}_{n_1, -n_1+m}^{\alpha\alpha}(\mathbf{x})]\} \\ &\times \{\text{tr}[(\tau_r \otimes s_0) \mathcal{Q}_{-n_2, n_2+m}^{\alpha\alpha}(\mathbf{x})]\}. \end{aligned} \quad (2.12d)$$

Here N_F is the density of states at the Fermi level in saddle-point approximation, as defined in Eq. (2.13c) below. We have written the action in real space, but one should remember that all of the fields are restricted to Fourier components with wave numbers $|\mathbf{k}| < \lambda$.

C. Saddle-point solution, Gaussian approximation, and physical correlation functions

It is easy to see that the Fermi-liquid saddle-point discussed in I remains a valid saddle-point in the presence of a long-range interaction, with $\Gamma^{(s)}$ in I replaced by $\Gamma^{(s)}(\mathbf{q} = 0)$. The single-particle Green function in saddle-point approximation is therefore given by the same expression as in I, viz.,

$$G_{\text{sp}}(\mathbf{p}, \omega_n) = [i\omega_n - \mathbf{p}^2/2m + \mu - \Sigma_n]^{-1}, \quad (2.13a)$$

with μ the chemical potential, and the self-energy Σ the solution of the equation

$$\begin{aligned} \Sigma_n &= \frac{1}{\pi N_F \tau_{\text{rel}} V} \sum_{\mathbf{p}} [i\omega_n - \mathbf{p}^2/2m + \mu - \Sigma_n]^{-1} \\ &+ 2\Gamma^{(s)}(\mathbf{q}=0) T \sum_m e^{i\omega_m 0} \frac{1}{V} \\ &\times \sum_{\mathbf{p}} [i\omega_m - \mathbf{p}^2/2m + \mu - \Sigma_m]^{-1}. \end{aligned} \quad (2.13b)$$

This is the standard Hartree-Fock result, with the disorder treated in the self-consistent Born approximation;

$$N_F = \frac{-2}{\pi} \frac{1}{V} \sum_{\mathbf{p}} \text{Im} G_{\text{sp}}(\mathbf{p}, i\omega_n \rightarrow i0) \quad (2.13c)$$

is the density of states at the Fermi level in saddle-point approximation, which is used for normalization purposes throughout.

Similarly, in an expansion about the saddle-point to Gaussian order, all of the expressions derived in I remain valid if we substitute $\Gamma^{(s)}(\mathbf{p})$ for $\Gamma^{(s)}$ in all propagators taken at wave vector \mathbf{p} . In particular, we obtain for the low-frequency, long-wavelength limit of the density susceptibility χ in the disordered case,

$$\chi(\mathbf{p}, \Omega_n) = \chi_{\text{st}}(\mathbf{p}) \frac{d(\mathbf{p})\mathbf{p}^2}{|\Omega_n| + d(\mathbf{p})\mathbf{p}^2}. \quad (2.14a)$$

Here Ω_n is a bosonic Matsubara frequency,

$$\chi_{\text{st}}(\mathbf{p}) = \frac{-N_F}{1 + N_F \Gamma^{(s)}(\mathbf{p})}, \quad (2.14b)$$

is the static density susceptibility, and

$$d(\mathbf{p}) = D[1 + N_F \Gamma^{(s)}(\mathbf{p})], \quad (2.14c)$$

with D the Boltzmann diffusion coefficient. In the clean limit we obtain the usual RPA expression,

$$\chi(\mathbf{p}, \Omega_n) = \frac{\chi_0(\mathbf{p}, \Omega_n)}{1 + \Gamma^{(s)}(\mathbf{p})\chi_0(\mathbf{p}, \Omega_n)}, \quad (2.15)$$

with χ_0 the Lindhard function. We see that in Gaussian approximation, the field theory describes screening, and the existence of plasmons, in RPA and its disordered generalization, respectively.

III. SYMMETRY ANALYSIS

From Sec. II C it follows, in conjunction with I, that at the level of the Gaussian approximation the Q_{nm} with $nm < 0$ are soft modes. In this section, we perform a symmetry analysis in order to show that they are indeed the exact soft modes of the theory. This will allow us to explicitly separate the soft modes from the massive ones, and to formulate an effective field theory for which the soft modes remain manifestly soft to all orders in perturbation theory.

A. Basic transformation properties, and Ward identity

Let us start with a symmetry analysis of our field theory that is a slight generalization of the procedure followed in I, which in turn was a generalization of the work on noninteracting electrons by Schäfer and Wegner, and Pruisken and Schäfer.¹⁴ We consider an infinitesimal simultaneous rotation in frequency and replica space given by

$$\begin{aligned} {}_r \hat{T}_{nm}^{\alpha\beta} &= \delta_{i0} \delta_{r0} [\delta_{\alpha\alpha_1} \delta_{nn_1} \delta_{\beta\alpha_2} \delta_{mn_2} - \delta_{\alpha\alpha_2} \delta_{nn_2} \delta_{\beta\alpha_1} \delta_{mn_1}] \theta \\ &\equiv \delta_{i0} \delta_{r0} \hat{t}_{nm}^{\alpha\beta} + O(\theta^2), \end{aligned} \quad (3.1)$$

which transforms the Q matrices according to $Q \rightarrow TQT^{-1}$, with $T = C\hat{T}C^T$, where $C = i\tau_1 \otimes s_2$. ($n_1 > 0, n_2 < 0$) and (α_1, α_2) are fixed pairs of frequency and replica indices that characterize the transformation. Under such an infinitesimal rotation, the Q matrices transform like

$$Q_{nm}^{\alpha\beta} \rightarrow Q_{nm}^{\alpha\beta} + \delta Q_{nm}^{\alpha\beta}, \quad (3.2a)$$

with

$$\delta Q_{nm}^{\alpha\beta} = [\delta_{\alpha\alpha_1} \delta_{nn_1} Q_{n_2 m}^{\alpha_2 \beta} + \delta_{\beta\alpha_1} \delta_{mn_1} Q_{nn_2}^{\alpha\alpha_2} - (1 \leftrightarrow 2)] \theta + O(\theta^2). \quad (3.2b)$$

Here we have shown only the frequency and replica indices, since all other degrees of freedom are unaffected by the transformation. The $\tilde{\Lambda}$ matrices transform accordingly. The symbol $(1 \leftrightarrow 2)$ indicates the same terms as written previously, but with the indices 1 and 2 interchanged.

Of the terms in the action, Eq. (2.8), only $\text{Tr} \ln(G_0^{-1} - i\tilde{\Lambda})$ and \mathcal{A}_{int} are not invariant under the above transformation. Their transformation behaviors are easily determined by an explicit calculation. We find

$$\text{Tr} \ln(G_0^{-1} - i\tilde{\Lambda}) \rightarrow \text{Tr} \ln(G_0^{-1} - i\tilde{\Lambda}) + \theta \text{Tr}(G \delta i\omega), \quad (3.3a)$$

with $G \equiv (G_0^{-1} - i\tilde{\Lambda})^{-1}$, and

$$i_r(\delta i\omega)_{nm}^{\alpha\beta} = \delta_{i0} \delta_{r0} (\delta_{\alpha\alpha_1} \delta_{nn_1} \delta_{\beta\alpha_2} \delta_{mn_2} + \delta_{\alpha\alpha_2} \delta_{nn_2} \delta_{\beta\alpha_1} \delta_{mn_1}) i\Omega_{n_1 - n_2}, \quad (3.3b)$$

and

$$\mathcal{A}_{int}^{(s)} \rightarrow \mathcal{A}_{int}^{(s)} + \delta \mathcal{A}_{int}^{(s)}, \quad (3.4a)$$

with

$$\delta \mathcal{A}_{int}^{(s)} = 32 \int d\mathbf{x} d\mathbf{y} \Gamma^{(s)}(\mathbf{x} - \mathbf{y}) \sum_{r=0,3} T \times \sum_{n_a n_b} [{}^0_r Q_{n_a n_b}^{\alpha_1 \alpha_1}(\mathbf{x}) {}^0_r Q_{n_2, n_2 - (n_a - n_b)}^{\alpha_2 \alpha_1}(\mathbf{y}) - (1 \leftrightarrow 2)] \theta. \quad (3.4b)$$

For our purposes, we concentrate on a discussion of the particle-hole spin-singlet interaction; the discussion of the remaining interaction channels proceeds analogously, but is less interesting since in these channels the interaction is short range. Proceeding analogously to I, we obtain from the above transformation properties the Ward identity

$$W_{int} + 8\Omega_{n_1 - n_2} \int d\mathbf{y} \langle {}^0_0 Q_{12}(\mathbf{y}) {}^0_0 Q_{34}(\mathbf{x}) \rangle = \delta_{13} \delta_{24} (\langle {}^0_0 Q_{11}(\mathbf{x}) \rangle - \langle {}^0_0 Q_{22}(\mathbf{x}) \rangle), \quad (3.5a)$$

where

$$W_{int} = -32 \int d\mathbf{y} d\mathbf{z} \Gamma^{(s)}(\mathbf{y} - \mathbf{z}) \sum_{r=0,3} T \times \sum_{n_a n_b} [\langle {}^0_0 Q_{34}(\mathbf{x}) \rangle_r {}^0_r Q_{n_a n_b}^{\alpha_1 \alpha_1}(\mathbf{y}) {}^0_r Q_{n_2, n_1 - (n_a - n_b)}^{\alpha_2 \alpha_1}(\mathbf{z}) - (1 \leftrightarrow 2)]. \quad (3.5b)$$

To simplify the notation we have adopted the convention $1 \equiv (n_1, \alpha_1)$. For the special case of a short-range interaction, $\Gamma^{(s)}(\mathbf{x} - \mathbf{y}) = \Gamma^{(s)} \delta(\mathbf{x} - \mathbf{y})$, we recover Eqs. (3.14) of I.

B. Solution of the Ward identity

We now solve the Ward identity by employing a method that is more transparent than the one used in I, and more suitable for a generalization to the long-range case.

We first split up the Q -matrices into their averages and fluctuations: $Q_{nm} = \langle Q_{nm} \rangle + (\Delta Q)_{nm} = \delta_{nm} \langle Q_{nn} \rangle + (\Delta Q)_{nm}$. Since $n_1 \neq n_2$, and $n_3 \neq n_4$, we have $\langle Q_{n_1 n_2} \rangle = \langle Q_{n_3 n_4} \rangle = 0$. Furthermore, if we put $n_a = n_b$ in the three-point functions, then the expression in square brackets in Eq. (3.5b) vanishes due to Eq. (A3a), so effectively we also have $\langle Q_{n_a n_b} \rangle = 0$. Equation (3.5a) then takes the form

$$W_{int} + 8\Omega_{n_1 - n_2} C_{12,34} = \delta_{13} \delta_{24} N_{12}, \quad (3.6a)$$

where

$$C_{12,34} = \int d\mathbf{y} \langle {}^0_0(\Delta Q)_{12}(\mathbf{y}) {}^0_0(\Delta Q)_{34}(\mathbf{x}) \rangle, \quad (3.6b)$$

and

$$N_{12} = \langle {}^0_0 Q_{11}(\mathbf{x}) \rangle - \langle {}^0_0 Q_{22}(\mathbf{x}) \rangle. \quad (3.6c)$$

W_{int} can be written as the sum of two terms, $W_{int} = W_{int}^{(1)} + W_{int}^{(2)}$, with

$$W_{int}^{(1)} = -32 \delta_{\alpha_1 \alpha_2} \Gamma^{(s)}(\mathbf{k} \rightarrow 0) N_{12} T \times \sum_{n_a, n_b} \delta_{\alpha_1 \alpha_a} \delta_{a-b, 1-2} C_{ab,34}, \quad (3.6d)$$

and

$$W_{int}^{(2)} = -32 \sum_{r=0,3} \sum_{\mathbf{q}}' \Gamma^{(s)}(\mathbf{q}) T \times \sum_{n_a n_b} [\langle {}^0_r(\Delta Q)_{n_1, n_2 - (n_a - n_b)}^{\alpha_1 \alpha_2}(-\mathbf{q}) \rangle \times \langle {}^0_r(\Delta Q)_{n_a n_b}^{\alpha_1 \alpha_1}(\mathbf{q}) {}^0_0(\Delta Q)_{34}(\mathbf{x}) \rangle - (1 \leftrightarrow 2)], \quad (3.6e)$$

Here we have chosen a mixed representation for $W_{int}^{(2)}$, with some ΔQ in real space, and some in Fourier space, in order to make the cutoff on the \mathbf{q} integration explicit.

To proceed, let us ignore $W_{int}^{(2)}$ for the time being. Its effect will be analyzed later. Equation (3.6a) then turns into a closed integral equation for the homogeneous correlation function C . At this point we note that our global symmetry transformation, Eq. (3.1), produces a Ward identity for homogeneous correlation functions. For a short-range interaction, this is sufficient to capture the important structural restrictions imposed on the theory by the symmetry of the action. However, in the long-range case the homogeneous limit is singular, and we have to be more careful. For instance, it is obvious that a local symmetry transformation would generate a wave-number-dependent $\Gamma^{(s)}$ in $W_{int}^{(1)}$, and in the long-range case $\Gamma^{(s)}(\mathbf{k} \rightarrow 0) \neq \Gamma^{(s)}(\mathbf{k} = 0)$. This is the reason why we have written $\Gamma^{(s)}(\mathbf{k} \rightarrow 0)$ in Eq. (3.6d). Furthermore, it is known from perturbation theory that the dispersion of the soft modes that are controlled by the Ward

identity is diffusive, $\Omega \sim \mathbf{k}^2$.¹⁵ The equation for the wave-number-dependent C then takes the form

$$\begin{aligned} 8(\Omega_{n_1-n_2} + D\mathbf{k}^2)C_{12,34}(\mathbf{k}) &= \delta_{13}\delta_{24}N_{12}(\mathbf{k}) + 32\Gamma^{(s)}(\mathbf{k}) \\ &\times \delta_{\alpha_1\alpha_2}N_{12}(\mathbf{k})T \\ &\times \sum_{ab} \delta_{\alpha_a\alpha_1}\delta_{1-2,a-b}C_{ab,34}(\mathbf{k}), \end{aligned} \quad (3.7)$$

with D the (exact) diffusion constant, and $N_{12}(\mathbf{k})$ the wave-vector-dependent generalization of N_{12} as defined in Eq. (3.6c). It is useful to define

$$P_n(\mathbf{k}) = T \sum_{n_1n_2} \delta_{n,n_1-n_2}N_{n_1n_2}(\mathbf{k}). \quad (3.8a)$$

By performing appropriate summations over the Ward identity, we find that

$$P_n(\mathbf{k}) = \frac{1}{4}(\Omega_n + D\mathbf{k}^2)\chi_{\text{sc}}(\mathbf{k}, \Omega_n), \quad (3.8b)$$

with χ_{sc} the screened density susceptibility. The latter is defined as

$$\chi_{\text{sc}}(\mathbf{k}, \Omega_n) = \frac{\chi(\mathbf{k}, \Omega_n)}{1 + \Gamma^{(s)}(\mathbf{k})\chi(\mathbf{k}, \Omega_n)}, \quad (3.8c)$$

where

$$\chi(\mathbf{k}, \Omega_n) = 32T \sum_{1234} \delta_{n,n_1-n_2}C_{12,34}(\mathbf{k}) \quad (3.8d)$$

is the full density susceptibility. Notice that this relation between P_n and χ_{sc} is exact, since $T\sum_{12} \delta_{n,n_1-n_2}W_{12,34}^{(2)} = 0$.

The integral equation, Eq. (3.7), can be solved in terms of P_n or χ_{sc} by means of the same methods that were employed to discuss the Gaussian Q -field theory in I. We find

$$\begin{aligned} C_{12,34}(\mathbf{k}) &= \frac{1}{16}[\delta_{13}\delta_{24}D_{12}(\mathbf{k}) + \delta_{1-2,3-4}\delta_{\alpha_1\alpha_3}\delta_{\alpha_1\alpha_2} \\ &\times 2T\Gamma^{(s)}(\mathbf{k})D_{12}(\mathbf{k})D_{34}^{(s)}(\mathbf{k})], \end{aligned} \quad (3.9a)$$

where

$$D_{12}(\mathbf{k}) = 2N_{12}(\mathbf{k})[\Omega_{n_1-n_2} + D\mathbf{k}^2]^{-1}, \quad (3.9b)$$

and

$$D_{12}^{(s)}(\mathbf{k}) = D_{12}(\mathbf{k})[1 - \Gamma^{(s)}(\mathbf{k})\chi_{\text{sc}}(\mathbf{k}, \Omega_{n_1-n_2})]^{-1}. \quad (3.9c)$$

To appreciate the difference between this structure and the analogous one in the short-range case, it is instructive to consider the limit of small wave numbers and frequencies with $D\mathbf{k}^2 \ll \Omega_{n_1-n_2}$. In this limit we have $2N_{12}(\mathbf{k}) \rightarrow \pi N(\epsilon_F)$ with $N(\epsilon_F)$ the exact density of states at the Fermi level, and $\chi_{\text{sc}}(\mathbf{k}, \Omega_{n_1-n_2}) \sim \mathbf{k}^2/\Omega_{n_1-n_2}$. In $d=3$, we obtain

$$\begin{aligned} 16C_{12,34}(\mathbf{k} \rightarrow 0) &= \delta_{12}\delta_{34} \frac{\pi N(\epsilon_F)}{\Omega_{n_1-n_2}} \\ &+ \delta_{1-2,3-4}\delta_{\alpha_1\alpha_2}\delta_{\alpha_1\alpha_3} \frac{\pi N(\epsilon_F)2\pi T}{\Omega_{n_1-n_2}D\mathbf{k}^2}. \end{aligned} \quad (3.10)$$

We see that the long-range Coulomb interaction causes the part of C that is nondiagonal in frequency to diverge like $T/\Omega\mathbf{k}^2$, rather than like T/Ω^2 in the short-range case. This structure is responsible for the well-known log-squared terms in the density of states and in the sound attenuation coefficient that appear in perturbative calculations of weak-localization effects in 2D systems with Coulomb interactions.⁶

We now consider the remaining contribution to W_{int} , $W_{\text{int}}^{(2)}$. Without the restriction on the wave-number integral in Eq. (3.6e), one could relabel wave vectors to show that $W_{\text{int}}^{(2)}$ contains terms that have the same structure as $W_{\text{int}}^{(1)}$. However, due to the cutoff contained in the definition of our Q -field theory this is not the case, and we resort to perturbation theory to analyze the structure of $W_{\text{int}}^{(2)}$. By reexpressing ΔQ in fermion fields, and using Wick's theorem to write the correlation function in Eq. (3.6e) in terms of Green functions, we find

$$W_{\text{int}}^{(2)} \sim \Gamma^{(s)}(\mathbf{k}) \frac{T}{\Omega} \text{Min}[\lambda^2, \lambda^2(\lambda l)], \quad (3.11)$$

with λ the cutoff introduced in Sec. II A, and l the elastic scattering mean free path. Based on this information, we now choose the cutoff. There are three obvious possible choices: One can make λ a fixed, small fraction of either the Fermi wave number k_F , or the mean free path l , or the screening wave number κ . The first choice would not allow us to control the terms in $W_{\text{int}}^{(2)}$, and is therefore undesirable. The second choice is possible, but means foregoing the option to take the clean limit, for reasons discussed after Eq. (2.6d). We therefore choose the third option, which makes $W_{\text{int}}^{(2)}$ of higher order in the interaction than the other terms in the Ward identity. To linear order in the interaction the solution given above in Eqs. (3.9), (3.10), is then exact. To higher order, while $W_{\text{int}}^{(2)}$ cannot make the modes Q_{nm} ($nm < 0$) massive, it will change the prefactors of the diffusive singularity at small frequencies and wave numbers. We conclude that an effective theory that respects Eqs. (3.9) will have the correct soft-mode structure as it follows from the symmetry of the action. It will further exactly reproduce perturbation theory to first order in the interaction. In Sec. IV we will show that the generalized nonlinear σ model of Ref. 12 is an effective theory with these properties.

C. Separation of soft and massive modes

From the previous subsection we know that the correlation functions of the Q_{nm} with $nm < 0$ remain soft, while those with $nm > 0$ remain massive even in the presence of long-range interactions. Therefore the mode separation for this case is the same as it is for short-range interactions, and we can restrict ourselves to summarizing the results of I.

One consequence of the symmetry properties of the Q matrices (see Appendix A) is that the set of Q is isomorphic

to the set of anti-Hermitian $8Nn \times 8Nn$ matrices. As shown in I, from this it follows that the most general Q can be written as a product,

$$Q = SPS^{-1}. \quad (3.12)$$

Here P is a matrix that is blockdiagonal in frequency space,

$$P = \begin{pmatrix} P^> & 0 \\ 0 & P^< \end{pmatrix}, \quad (3.13)$$

with elements $P^>$ ($P^<$) for $nm > 0$ ($nm < 0$) that are isomorphic to the set of anti-Hermitian $4Nn \times 4Nn$ matrices, and S an element of the homogeneous space $\text{USp}(8Nn, \mathcal{C})/\text{USp}(4Nn, \mathcal{C}) \times \text{USp}(4Nn, \mathcal{C})$, i.e., the set of all cosets of $\text{USp}(8Nn, \mathcal{C})$ with respect to $\text{USp}(4Nn, \mathcal{C}) \times \text{USp}(4Nn, \mathcal{C})$.¹⁶

This achieves the desired separation of our degrees of freedom into soft and massive ones. The massive degrees of freedom are represented by the matrix P , while the soft ones are represented by the transformations $S \in \text{USp}(8Nn, \mathcal{C})/\text{USp}(4Nn, \mathcal{C}) \times \text{USp}(4Nn, \mathcal{C})$.

In order to formulate the field theory in terms of the soft and massive modes, one also needs the invariant measure $I[P]$, or the Jacobian of the transformation from the Q to the P and the S , defined by

$$\int D[Q] \cdots = \int D[P] I[P] \int D[S] \cdots \quad (3.14)$$

We will not need the measure explicitly for our purposes, and refer to I, where it has been constructed in terms of the eigenvalues of P .

IV. EFFECTIVE FIELD THEORY FOR DISORDERED INTERACTING FERMIONS

Having achieved a separation of soft and massive modes, we are now in a position to formulate an effective theory for electrons with a long-range interaction that focuses on the soft modes. In the short-range case, this was done by integrating out the massive modes in tree approximation. This led to the nonlinear σ model in I, and it was shown that this procedure preserves the structure of the Ward identity. As a result, the σ model contains the same Fermi-liquid fixed point, as well as the leading corrections to scaling near it, as the underlying full model, and it also contains a critical fixed point that describes an Anderson-Mott metal-insulator transition.

We cannot simply repeat this procedure for the present case of long-range interactions, since integrating out the massive modes in tree approximation would lead to a theory that violates the Ward identity. We therefore must treat the massive modes more carefully, and our aim is to find the simplest approximation that will still guarantee the correct structure of the Ward identity, and hence lead to an effective theory that has the correct symmetry.

The first steps are the same as in I: We define a transformed auxiliary field Λ by

$$\Lambda(\mathbf{x}) = S^{-1}(\mathbf{x}) \tilde{\Lambda}(\mathbf{x}) S(\mathbf{x}), \quad (4.1a)$$

and a new field \hat{Q} by

$$\hat{Q}(\mathbf{x}) = \frac{4}{\pi N_F} \mathcal{S}(\mathbf{x}) \langle P \rangle \mathcal{S}^{-1}(\mathbf{x}). \quad (4.1b)$$

We then expand about the expectation values of Λ and P ,

$$P = \langle P \rangle + \Delta P, \quad \Lambda = \langle \Lambda \rangle + \Delta \Lambda. \quad (4.2)$$

As explained in I, it is sufficient to replace $\langle P \rangle$ and $\langle \Lambda \rangle$ by the respective saddle-point values, and to further replace $\langle P \rangle$ in the definition of \hat{Q} , Eq. (4.1b), by the simple approximation

$$\langle P \rangle \approx \frac{\pi}{4} N_F \pi, \quad (4.3a)$$

with

$$\pi_{12} = \delta_{12} \text{sgn } \omega_{n_1}. \quad (4.3b)$$

We mention that there is no obvious small parameter that controls these approximations. Rather, they will be justified *a posteriori* by the fact that the resulting effective field theory, the nonlinear σ model, respects the Ward identity, Eqs. (3.9), (3.10), that was derived in the previous section. This in turn shows that the approximations, Eqs. (4.3), are consistent with neglecting the term $W_{\text{int}}^{(2)}$ in the Ward identity, which itself is perturbatively controlled for small interaction strengths. If the theory is renormalizable, this implies that the effective theory resulting from the above approximations will have the same structure as the full one, albeit with different coefficients. We will come back to this point at the end of this section, and in Sec. V below. With these approximations, \hat{Q} has the properties

$$\hat{Q}^2(\mathbf{x}) = 1, \quad \hat{Q}^\dagger = \hat{Q}, \quad \text{Tr } \hat{Q}(\mathbf{x}) = 0, \quad (4.4a)$$

and can be parametrized by

$$\hat{Q} = \begin{pmatrix} \sqrt{1 - qq^\dagger} & q \\ q^\dagger & -\sqrt{1 - q^\dagger q} \end{pmatrix}, \quad (4.4b)$$

where the matrix q has elements q_{nm} whose frequency labels are restricted to $n \geq 0$, $m < 0$. S can also be expressed in terms of q ,¹⁶

$$S = \begin{pmatrix} \sqrt{1 - bb^\dagger} & b \\ -b^\dagger & \sqrt{1 - b^\dagger b} \end{pmatrix}, \quad (4.4c)$$

where

$$b(q, q^\dagger) = \frac{-1}{2} q f(q^\dagger q), \quad (4.4d)$$

with

$$f(x) = \sqrt{\frac{2}{x}} (1 - \sqrt{1-x})^{1/2}. \quad (4.4e)$$

Unlike I, where we just dropped the fluctuations of P and Λ , here we next expand to second order in ΔP and $\Delta \Lambda$. The reason for this change of procedure compared to the short-

range case is that here the ΔP fluctuations are multiplied by a divergent Coulomb potential, and therefore must be retained. The part of the action that is quadratic in these massive fluctuations reads

$$\begin{aligned} \mathcal{A}_m^{(2)} = & \frac{1}{4} \int d\mathbf{x} d\mathbf{y} \operatorname{tr} G_{\text{sp}}(\mathbf{x}-\mathbf{y}) \Delta\Lambda(\mathbf{y}) G_{\text{sp}}(\mathbf{y}-\mathbf{x}) \Delta\Lambda(\mathbf{x}) \\ & + \int d\mathbf{x} \Delta\Lambda(\mathbf{x}) \Delta P(\mathbf{x}) + \mathcal{A}_{\text{int}}^{(s)} [\mathcal{S}\langle P \rangle \mathcal{S}^{-1} + \mathcal{S} \Delta P \mathcal{S}^{-1}], \end{aligned} \quad (4.5)$$

with $\mathcal{A}_{\text{int}}^{(s)}[Q]$ the spin-singlet interaction term from Eq. (2.12b). \mathcal{A}_{dis} , and the invariant measure, expanded to second order in ΔP , also contribute to $\mathcal{A}_m^{(2)}$. However, their net effect is to add a constant to the singular interaction potential in the $\mathcal{A}_{\text{int}}^{(s)}$ contribution to $\mathcal{A}_m^{(2)}$, and hence they can be neglected. Likewise, there are terms linear in $\Delta\Lambda$ that couple to the soft modes \mathcal{S} . These always contain at least one frequency or gradient squared, and therefore are unimportant for the leading structure imposed by the Ward identity. They also turn out to be renormalization-group irrelevant at the Fermi-liquid fixed point that we will discuss in Sec. V below. We therefore neglect all of these terms.

Because of the coupling between ΔP and \mathcal{S} , $\mathcal{A}_m^{(2)}$ still represents a complicated quadratic form. To handle it, we expand \mathcal{S} , Eq. (4.4c), in powers of q . To lowest order, we just have $\mathcal{S}=1$. It turns out that this lowest-order approximation is sufficient to ensure the correct structure of the Ward identity. We have also explicitly checked that higher-order terms in this q expansion lead to corrections that are irrelevant near the Fermi-liquid fixed point. With $\mathcal{S}=1$, the massive Gaussian fluctuations are easily integrated out. Neglecting terms that are of first or higher order in the external frequency, the result is a change of the interaction term $\mathcal{A}_{\text{int}}^{(s)}$ to a term of the same structure, but with $\Gamma^{(s)}$ replaced by its screened counterpart,

$$\Gamma_{\text{sc}}^{(s)}(\mathbf{p}) = \frac{\Gamma^{(s)}(\mathbf{p})}{1 + N_F \Gamma^{(s)}(\mathbf{p})}, \quad (4.6)$$

with N_F from Eq. (2.13c) (here we have neglected a wave-number dependence that is subleading compared to that of $\Gamma^{(s)}$). We see that integrating out the massive fluctuations in the approximation we have chosen leads to static screening of the Coulomb interaction. Analogous screening effects occur in the remaining interaction channels. However, they are uninteresting there since they just renormalize the numbers $\Gamma^{(t)}$ and $\Gamma^{(c)}$.

The remaining steps in the derivation of the nonlinear σ model are the same as in I. We thus obtain the σ model action

$$\begin{aligned} \mathcal{A}_{\text{NL}\sigma\text{M}} = & \frac{-1}{2G} \int d\mathbf{x} \operatorname{tr} (\nabla \tilde{Q}(\mathbf{x}))^2 + 2H \int d\mathbf{x} \operatorname{tr} (\Omega \tilde{Q}(\mathbf{x})) \\ & + \mathcal{A}_{\text{int}}[\tilde{Q}], \end{aligned} \quad (4.7)$$

where $\tilde{Q} = \hat{Q} - \pi$, with π from Eq. (4.3b). $G = 8/\pi\sigma_0$ with σ_0 the conductivity in the self-consistent Born approximation, and $H = \pi N_F/8$. \mathcal{A}_{int} is given by Eqs. (2.12), but with

$\Gamma^{(s)}$ replaced by $\Gamma_{\text{sc}}^{(s)}$, Eq. (4.6). This is the generalized nonlinear σ model for disordered electrons with a Coulomb interaction, as proposed and discussed by Finkel'stein,¹² and the above procedure represents a technical derivation of this model. Its properties have been reviewed and discussed in detail in Ref. 6. Using the explicit results of that reference, it is straightforward to check that the model indeed obeys the Ward identity, Eq. (3.10).¹⁷ This justifies, *a posteriori*, the approximations that have entered our derivation. We will further discuss the merits and limitations of the model in the next section.

V. DISCUSSION

We finally discuss our results, and the procedures used to derive them.

A. Role of the phase-space decomposition

Let us start with a discussion of the phase-space decomposition in Sec. II A, which writes the interaction term (and also the disorder term) in the action as a sum of terms of different structures, with a cutoff to avoid double counting. As a result, the definition of the effective action contains this cutoff, which is *a priori* unspecified. For instance, without the cutoff λ the three terms $\tilde{S}_{\text{int}}^{\alpha(1,2,3)}$, Eqs. (2.2), would all be equal and equal to $\tilde{S}_{\text{int}}^{\alpha}$. A superficial consideration might conclude that this decomposition of the action introduces an unnecessary ambiguity into the theory. In fact, however, the phase-space decomposition is necessary in order to derive a theory that allows for a well-behaved perturbation theory. This can be explained most easily by using the disorder term \mathcal{A}_{dis} , Eqs. (2.11), as an example. As explained in I, the two contributions $\mathcal{A}_{\text{dis}}^{(1,2)}$ result from a phase space decomposition analogous to the one performed on the interaction, and the matrix $Q(\mathbf{x})$ is therefore to be understood as containing only Fourier components with wavenumbers $|\mathbf{k}| < \lambda$. Now suppose we had not performed the decomposition. Then \mathcal{A}_{dis} would consist of $\mathcal{A}_{\text{dis}}^{(2)}$ only, with τ_{rel} replaced by $2\tau_{\text{rel}}$, and $Q(\mathbf{x})$ containing all Fourier components. The saddle-point Green function for this action would then contain a disorder part of the self-energy that is half the Born value. In a perturbation expansion in powers of the disorder, higher orders would then have to make up for the missing factors of 2 at *zeroth* order, i.e., the perturbation expansion would be singular. This is precisely what happens in standard many-body perturbation theory,³ where singular integrals make it impossible to easily determine the order of a contribution from its diagrammatic structure, and infinite resummations are in general necessary to obtain all contributions of a given order. In contrast, perturbation theory for the nonlinear σ model is much better behaved, with the number of loops determining the order to which a given diagram contributes. This is a consequence of a judicious choice of the starting point for the loop expansion, which in turn depends crucially on the phase-space decomposition. The fact that the theory depends on a cutoff is the price paid for the controlled nature of perturbation expansions.

Similarly, the phase-space decomposition performed on the interaction part of the action allows us to perturbatively

control the complicated contribution $W_{\text{int}}^{(2)}$ to the Ward identity, Eq. (3.11). This guarantees that the nonlinear σ model correctly reproduces perturbation theory in the interaction to first order, as is well known from comparing results obtained by either method.^{18,12,6}

Conversely, the above discussion makes it clear that the nonlinear σ model, owing to its derivation, is perturbative in nature with respect to the electron-electron interaction. Indeed, its restriction in that respect is more serious than with respect to disorder: Since the loop expansion is an expansion in powers of the disorder, going to higher order in perturbation theory will always include higher order disorder effects. With respect to the interaction, the analogous statement is not true, since some effects of higher order in the interaction are left out of the model, although the loop expansion resums certain classes of interaction terms to all orders. Of course, a complete renormalization of the model would, in principle, supply all of the effects that might have been left out of the bare model, but this will in general not be captured by the standard perturbative renormalization based on low orders in the loop expansion. This is a restriction that is important to keep in mind in the context of discussions about possible exotic effects of a strong effective interaction, such as, e.g., a metallic non-Fermi liquid ground state. It may also be relevant for understanding the observation¹⁹ that renormalization group calculations based on the σ model in high dimensions ($d > 6$) reveal relevant terms of a structure that is not seen in low-order $2 + \epsilon$ expansions. Physically, the ‘‘standard’’ generalized σ model approach is valid if the physics one is interested in is determined by the two-particle diffusive modes. If, for example, there were also soft single-particle excitations, then in general the coefficients in the σ model would be singular and this approach would break down.

Finally, we note that already in the Gaussian approximation the perturbative nature of the σ model approach with respect to interactions is apparent. In I we pointed out that before the massive modes were integrated out, the Gaussian field theory explicitly contained the Stoner theory for ferromagnetism. However, after the σ model approximation was made, the interaction terms that lead to the Stoner theory were absent.

B. Screening, and the disordered Fermi-liquid fixed point

A characteristic feature of the long-range Coulomb interaction is that it leads to the system being incompressible: The wave-number-dependent thermodynamic derivative $(\partial n / \partial \mu)(\mathbf{k}) = \chi(\mathbf{k}, \Omega = 0)$, which is proportional to the compressibility, is for small wave numbers given by

$$\left(\frac{\partial n}{\partial \mu}\right)(\mathbf{k}) = \left(\frac{\partial n}{\partial \mu}\right)_{\text{sc}} \frac{\mathbf{k}^2}{\mathbf{k}^2 + \kappa^2}, \quad (5.1)$$

with κ the screening wave number, and $(\partial n / \partial \mu)_{\text{sc}} = \chi_{\text{sc}}(\mathbf{k} \rightarrow 0, \Omega = 0)$ the screened density susceptibility, which is a nonzero number. This structure follows from, and is controlled by, the Ward identity, as can be seen from Sec. III B above. It is instructive to check explicitly that the nonlinear σ model respects the compressibility sum rule, Eq. (5.1). Within the framework of the σ model, one has⁶

$$\left(\frac{\partial n}{\partial \mu}\right)(\mathbf{k}) = \frac{\pi}{8} [H + K_s(\mathbf{k})], \quad (5.2)$$

with H as defined after Eq. (4.7), and $K_s(\mathbf{k}) = -\pi \Gamma^{(s)}(\mathbf{k})/8$. We see that the σ model indeed respects the compressibility sum rule, with the Gaussian approximation for the screened compressibility, $(\partial n / \partial \mu)_{\text{sc}} \approx N_F$. Notice that in the model originally proposed by Finkel’stein,¹² Fermi-liquid corrections had been put in to make $(\partial n / \partial \mu)_{\text{sc}}$ the exact screened density susceptibility. These are missing here, since we have integrated out the massive modes, which account for the screening, in a Gaussian approximation. We emphasize that, while it is of course always possible to put in Fermi-liquid corrections by hand, nothing is really gained by doing so: Such a procedure only amounts to a partial resummation of some terms that are of higher order in the interaction, which does not change the fact that the effective theory has a perturbative character with respect to the interaction, as was discussed in Sec. V A above. Furthermore, the point of any effective theory is that it correctly captures the *structure* of the full theory, while the coefficients can be represented by some approximation in the bare theory. Upon renormalizing the bare effective theory, the coefficients will be renormalized by fluctuation effects.

C. Renormalization-group properties of the effective field theory

We finally mention that the renormalization-group properties of the σ model, Eq. (4.7), are well known. The theory possesses a critical fixed point that describes an Anderson-Mott metal-insulator transition.^{20,6} Also, due to the short-range nature of the effective, screened, interaction [Eq. (4.6)], the discussion of the stable Fermi-liquid fixed point given in Sec. III B 2 of I still applies. The Fermi-liquid ground state is stable for $d > 2$ in the presence of quenched disorder, and for $d > 1$ in the clean limit. The corrections to scaling yield the weak-localization nonanalyticities and their clean counterparts as discussed in I, modified by the log-squared singularities in the density of states and the sound attenuation that are induced by the Coulomb interaction^{4,6} as mentioned after Eq. (3.10) above. We emphasize again, however, that due to the perturbative nature of the effective field theory, our considerations do not in any sense constitute a proof that the Fermi-liquid fixed point will be stable for arbitrary strengths of the bare interaction constant. What we have shown is that the fixed point is perturbatively stable for weak interactions. It is easy to see that an interaction that is of longer range than the Coulombic one, $v(\mathbf{x}) \sim 1/|\mathbf{x}|^\alpha$ with $\alpha < 1$, will destroy the screening process, and hence lead to a relevant operator that destroys the Fermi-liquid fixed point, at least close to its lower critical dimension. If an effective interaction of such long range were generated by the renormalization group acting on the σ model, or if it were present in the bare action for a different effective theory that is not subject to the perturbative restriction of weak interactions, then this could lead to a non-Fermi-liquid ground state. These points may be important in the context of the ongoing discussion about possible non-Fermi-liquid ground states in 2D (clean) electron systems.²¹

ACKNOWLEDGMENTS

This work was supported by the NSF under Grants Nos. DMR-95-10185, DMR-96-32978, and DMR-98-70597, and by the Deutsche Forschungsgemeinschaft. Part of this work was performed at the Aspen Center for Physics. We thank the Center for its hospitality, and E. Abrahams for a discussion concerning the perturbative nature of the σ model.

APPENDIX: PROPERTIES OF THE Q MATRICES

Here we list again the symmetry properties of the Q matrices that were defined in I, and add some additional remarks. Q is self-adjoint under the adjoint operation $Q^\dagger = C^T Q^T C$, with $C_{nm}^{\alpha\beta} = \delta_{nm} \delta_{\alpha\beta} i \tau_1 \otimes s_2$,

$$Q = C^T Q^T C. \quad (\text{A1a})$$

In addition, the Hermitian conjugate Q^\dagger of Q is related to Q by²²

$$Q^\dagger = (\tau_3 \otimes s_0) \Gamma Q \Gamma^{-1} (\tau_3 \otimes s_0), \quad (\text{A1b})$$

where the similarity transformation denoted by Γ has the property

$$(\Gamma Q \Gamma^{-1})_{nm} = Q_{-n-1, -m-1}. \quad (\text{A1c})$$

We now expand our matrix fields in the spin-quaternion basis defined after Eq. (2.9),

$$Q_{12}(\mathbf{x}) = \sum_{r,i=0}^3 (\tau_r \otimes s_i) {}^i Q_{12}(\mathbf{x}), \quad (\text{A2a})$$

$$\tilde{\Lambda}_{12}(\mathbf{x}) = \sum_{r,i=0}^3 (\tau_r \otimes s_i) {}^i \tilde{\Lambda}_{12}(\mathbf{x}), \quad (\text{A2b})$$

where again $1 \equiv (n_1, \alpha_1)$, etc. In this basis, we have the following symmetry properties:

$${}^0 Q_{12} = (-)^{r_0} {}^r Q_{21}, \quad (r=0,3), \quad (\text{A3a})$$

$${}^i Q_{12} = (-)^{r+1} {}^i Q_{21}, \quad (r=0,3; i=1,2,3), \quad (\text{A3b})$$

$${}^0 Q_{12} = {}^0 Q_{21}, \quad (r=1,2), \quad (\text{A3c})$$

$${}^i Q_{12} = -{}^i Q_{21}, \quad (r=1,2; i=1,2,3). \quad (\text{A3d})$$

Together with the behavior under Hermitian conjugation, Eq. (A1b), this further implies

$${}^i Q_{12}^* = -{}^i Q_{-n_1-1, -n_2-1}^{\alpha_1 \alpha_2}, \quad (\text{A4})$$

for all i and r .²² Analogous relations hold for $\tilde{\Lambda}$ by virtue of the linear coupling between Q and $\tilde{\Lambda}$.

We further note that as a result of Eqs. (A3d) and (A4) the matrix elements ${}_{1,2}^{1,2,3} Q_{11}$ in the particle-particle spin-triplet channel are real (not imaginary, as erroneously stated in I). As a result, the Gaussian theory [Eqs. (2.36) in I] is formally unstable in that channel, and the formally diverging Gaussian integral needs to be interpreted. We have ascertained that a rotation of the relevant Q -integration contour onto the imaginary axis, which effectively makes ${}_{1,2}^{1,2,3} Q_{11}$ imaginary, provides an interpretation that guarantees agreement with the well-known results of conventional perturbation theory.⁴ Also note that within the σ model, the matrix elements of Q in the particle-particle spin-triplet channel are imaginary.⁶ This follows from Eq. (A1a) in conjunction with the Hermiticity of Q within the σ model, Eq. (4.4a).

*Present address: Institut für Theorie der kondensierten Materie, Universität Karlsruhe, 76128 Karlsruhe, Germany.

¹See, e.g., G. Baym and C. Pethick, *Landau Fermi-Liquid Theory* (Wiley, New York, 1991).

²See, e.g., A. L. Fetter and D. L. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).

³See, e.g., A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).

⁴B. L. Altshuler, A. G. Aronov, D. E. Khmel'nitskii, and A. I. Larkin, in *Quantum Theory of Solids*, edited by I. M. Lifshits (Mir, Moscow, 1982), p. 130; B. L. Altshuler and A. G. Aronov, in *Electron-Electron Interactions in Disordered Systems*, edited by A. L. Efros and M. Pollak (North-Holland, Amsterdam, 1985), p. 1.

⁵As in I, we use the term ‘‘weak localization’’ to refer to the nonanalytic behavior of electronic correlation functions in the limit of zero momentum and/or frequency that is induced by quenched disorder, or by a combination of interactions and quenched disorder, and that occurs even if the system is far from any kind of phase transition. The physical reason for these nonanalyticities is the diffusive motion of the electrons in the presence of quenched disorder. For a detailed discussion see, e.g., Ref. 4.

⁶For a review, see, D. Belitz and T. R. Kirkpatrick, *Rev. Mod. Phys.* **66**, 261 (1994).

⁷F. Wegner, *Z. Phys. B* **35**, 207 (1979).

⁸P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin, Menlo Park, CA, 1984), Chap. 5; M. E. Fisher (unpublished).

⁹R. Shankar, *Rev. Mod. Phys.* **66**, 129 (1994).

¹⁰N. Dupuis and G. Y. Chitov, *Phys. Rev. B* **54**, 3040 (1996); N. Dupuis, cond-mat/9604189 (unpublished); A. Houghton and J. B. Marston, *Phys. Rev. B* **48**, 7790 (1993); A. Houghton, H. J. Kwon, and J. B. Marston, *ibid.* **50**, 1351 (1994); H. J. Kwon, A. Houghton, and J. B. Marston, *ibid.* **52**, 8002 (1995); A. Houghton, H. J. Kwon, J. B. Marston, and R. Shankar, *J. Phys.: Condens. Matter* **6**, 4909 (1994); J. Fröhlich and R. Götschmann, *Phys. Rev. B* **55**, 6788 (1997); P. Kopietz, J. Hermisson, and K. Schönhammer, *ibid.* **52**, 10 877 (1995); P. Kopietz and K. Schönhammer, *Z. Phys. B* **100**, 259 (1996); P. Kopietz, *J. Phys.: Condens. Matter* **8**, 10 483 (1996).

¹¹D. Belitz and T. R. Kirkpatrick, *Phys. Rev. B* **56**, 6513 (1997).

¹²A. M. Finkel'stein, *Zh. Éksp. Teor. Fiz.* **84**, 168 (1983) [*Sov. Phys. JETP* **57**, 97 (1983)]; *Z. Phys. B* **56**, 189 (1984).

¹³D. Belitz, T.R. Kirkpatrick, and T. Vojta, *Phys. Rev. B* **55**, 9452 (1997).

¹⁴L. Schäfer and F. Wegner, *Z. Phys. B* **38**, 113 (1980); A. M. M. Pruisken and L. Schäfer, *Nucl. Phys. B* **200** [FS4], 20 (1982). See also A. J. McKane and M. Stone, *Ann. Phys. (N.Y.)* **131**, 36 (1981).

- ¹⁵It would be hard to actually *prove* that the Ward identity has a diffusive structure, as a local symmetry transformation would be much harder to analyze than our global one. Since there is no reasonable doubt that our model leads to diffusive electron dynamics, we therefore *assume* a diffusive structure. Note that this assumption pertains to the dispersion relation of the soft modes only, *not* to their existence and identification in terms of Q propagators.
- ¹⁶For a discussion of these algebraic structures, see, e.g., R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
- ¹⁷Strictly speaking, this statement assumes that the σ model is renormalizable to all orders. It has been checked explicitly to one-loop order, and partially to two-loop order, see Ref. 6.
- ¹⁸C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984).
- ¹⁹D. Belitz and T. R. Kirkpatrick, Z. Phys. B **98**, 513 (1995).
- ²⁰A. M. Finkel'stein, Zh. Éksp. Teor. Fiz. **86**, 367 (1984) [Sov. Phys. JETP **59**, 212 (1984)].
- ²¹See, e.g., Papers presented at the Institute for Theoretical Physics Conference on Non-Fermi Liquid Behavior in Metals [J. Phys.: Condens. Matter **8**, 48 (1996)].
- ²²Equations (2.26b) and (2.29b) in I were not correct.