Phase diagram of electronic systems with quadratic Fermi nodes in $2 < d < 4$: $2 + \epsilon$ expansion, $4 - \epsilon$ expansion, and functional renormalization group

Lukas Janssen[1] and Igor F. Herbut[2]

[1]*Institut für Theoretische Physik, Technische Universität Dresden, 01062 Dresden, Germany*
[2]*Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6*

Several materials in the regime of strong spin-orbit interaction such as HgTe, the pyrochlore iridate $Pr_2Ir_2O_7$, and the half-Heusler compound LaPtBi, as well as various systems related to these three prototype materials, are believed to host a quadratic band touching point at the Fermi level. Recently, it has been proposed that such a three-dimensional gapless state is unstable to a Mott-insulating ground state at low temperatures when the number of band touching points N at the Fermi level is smaller than a certain critical number N_c. We further substantiate and quantify this scenario by various approaches. Using ϵ expansion near two spatial dimensions, we show that $N_c = 64/(25\epsilon^2) + O(1/\epsilon)$ and demonstrate that the instability for $N < N_c$ is towards a nematic ground state that can be understood as if the system were under (dynamically generated) uniaxial strain. We also propose a truncation of the functional renormalization group equations in the dynamical bosonization scheme which we show to agree to one-loop order with the results from ϵ expansion both near two as well as near four dimensions, and which smoothly interpolates between these two perturbatively accessible limits for general $2 < d < 4$. Directly in $d = 3$ we therewith find $N_c = 1.86$, and thus again above the physical $N = 1$. All these results are consistent with the prediction that the interacting ground state of pure, unstrained HgTe, and possibly also $Pr_2Ir_2O_7$, is a strong topological insulator with a dynamically generated gap—a topological Mott insulator.

I. INTRODUCTION

Solid matter is commonly classified by electrical transport behavior. In a sufficiently pure metal (or semimetal) the conductivity diverges when temperature T goes to zero, or witnesses a superconducting transition at finite (though usually small) T. On the other hand, we call a material an insulator (or semiconductor) if its conductivity vanishes when the system is cooled down. Scanning tunneling spectroscopy measurements correlate with this behavior: In a metal the density of states near the Fermi level is finite, while there is a finite gap in the spectrum of an insulator (or semiconductor).

Lately, a new class of materials that do not quite fit into this scheme has moved into the focus of attention: Systems in which the Fermi surface shrinks to a few isolated points in the Brillouin zone live right on the edge between metals and semiconductors. In the purest graphene samples, for instance, the conductivity at charge neutrality decreases with decreasing temperature (as in a semiconductor), but approaches at the lowest temperatures a finite value on the order of the conductance quantum e^2/h [1]. In three dimensions, materials with strong spin-orbit coupling can host pairs of Weyl nodes, which may be thought of as three-dimensional (3D) analogs of graphene's linear band crossing points, albeit with nontrivial topology [2]. Although in 3D the density of states as function of energy ε vanishes as $D(\varepsilon) \sim \varepsilon^2$ near the linear band crossing point, the low-temperature conductivity is large as long as disorder is weak—an unusual (semi)metallic state [3]. Such a *Weyl semimetal* state has recently been identified experimentally, together with its characteristic Fermi-arc surface states, in TaAs [4,5].

In this work we focus on three-dimensional systems with *quadratic* Fermi nodes, as realized, for instance, in α-Sn or HgTe. These materials feature band inversion due to strong spin-orbit coupling, ensuring that the degeneracy at the quadratic band touching point (QBT) is protected by the crystal's cubic symmetry, and in the undoped situation the Fermi energy is right at the touching point [6]. Furthermore, no other states turn out to cross the Fermi level away from this point in these materials. The systems may thus be viewed as 3D analogs of *bilayer* graphene, however, with just a single QBT that is located at the center Γ of the Brillouin zone. Such low-energy band structure has also been measured in the pyrochlore iridate $Pr_2Ir_2O_7$, therewith making it a strongly correlated analog of HgTe [7]. Recently, a quadratic Fermi node in 3D has also been identified in the related compound $Nd_2Ir_2O_7$ [8]. In these materials, it is believed that the main physics is predominantly driven by the iridium electronic structure, while the interaction with the local rare-earth moments playing a role only at very low temperatures [9]. Another family of materials which may host a QBT at the Fermi level are given by ternary half-Heusler compounds, with LaPtBi as a prototype system [10–12].

In the noninteracting case, the conductivity in a Fermi system with QBT vanishes at low temperatures with a power law [13] (in contrast to the Weyl semimetal), and the system hence should properly be regarded as "gapless semiconductor" [6]. However, as the density of states has the square-root form as function of energy ε, $D(\varepsilon) \sim \sqrt{\varepsilon}$, the long-range Coulomb interaction is only marginally screened and the low-temperature behavior of these systems might decisively depend on the role played by the interactions. In fact, using an ϵ expansion around the upper critical spatial dimension of $d = 4$, as well as in the limit of large number N of QBTs at the Fermi level, Abrikosov and Beneslavskii [14,15] found a scale-invariant interacting ground state with unusual power laws in various thermodynamic observables—a 3D non-Fermi liquid (NFL) state. This scenario has recently been reviewed and put forward as an explanation for the anomalous low-temperature

behavior measured in $Pr_2Ir_2O_7$ [13]. In an ensuing series of papers [16–18], we have argued, on the other hand, that the expansion around the upper critical dimension, and also the strict large-N limit itself, might possibly even qualitatively fail to describe the ground-state behavior in the physical situation for $d = 3$ and $N = 1$. From a renormalization-group (RG) perspective, the chief culprit is the negligence of those short-range components of the Coulomb interaction, which one is tempted to discard as irrelevant at large N or small $\epsilon = 4 - d$, which however may become important away from these limits. By employing a simple one-loop analysis for fixed $N = 1$, we found the Abrikosov-Beneslavskii scale-invariant ground state to be unstable in $d = 3$, indicated by a runaway flow of short-range interactions [16]. The situation is similar to $(2 + 1)$-dimensional relativistic quantum electrodynamics (QED_{2+1}), which has a scale-invariant ("conformal") ground state at large number of fermions N, but is believed to suffer from a quantum phase transition towards a symmetry-broken ground state if N drops below a critical number N_c [19–24]. In a calculation that resembles the inaugural work in QED_{2+1}, we showed that an analogous critical fermion number N_c exists in the nonrelativistic system with QBT in three spatial dimensions, by deriving and exploiting the nonperturbative solution of the Dyson-Schwinger equations within the $1/N$ expansion [18].

In the present work we add to the picture the results of various complementary approaches to the problem. We first demonstrate that above and near two spatial dimensions, the instability of the NFL state occurs at a large value of N. This allows us to prove the existence of a phase boundary in the d-N plane, and to calculate its shape to leading order in $\epsilon = d - 2$. Within this limit, we also show that the ground state for $N < N_c$ is a *nematic* insulator with spontaneously broken rotational symmetry and full, but anisotropic gap. We furthermore revisit the expansion near the upper critical dimension with control parameter $\epsilon = 4 - d$ by formulating the corresponding Gross-Neveu-Yukawa theory, and establish the existence of another, quantum critical fixed point (QCP), alongside the Abrikosov-Beneslavskii NFL fixed point. The existence of this QCP turns out to be responsible for the nematic instability. We demonstrate that both fixed points, and their nontrivial interplay as a function of N, can also be assessed in a simple perturbative expansion in fixed dimension $d = 3$, thereby extending our previous results [16] to general fermion number N. If N is lowered from infinity towards the physical $N = 1$, the QCP and the NFL fixed point approach each other in coupling space and eventually merge at some critical N_c. For $N < N_c$ they disappear into the complex-coupling plane, leaving behind the runaway flow of short-range couplings. Computing the precise value for N_c in $d = 3$ beyond simple approximations is, of course, a challenging task. In order to gain yet another estimate we employ the functional renormalization group (FRG) in the so-called dynamical bosonization scheme. Although this scheme *a priori* lacks an obvious control parameter, we discover *a posteriori* that our FRG predictions coincide precisely with the results from the ϵ expansions both near two and near four dimensions. For general $2 < d < 4$, it smoothly interpolates between these two perturbatively accessible limits [25]. This leads to a phase diagram in the d-N plane with the insulating nematic state at
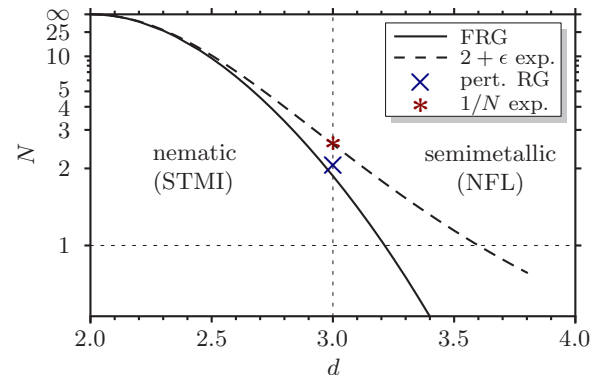

FIG. 1. Phase diagram of electronic systems with N quadratic Fermi nodes in $2 < d < 4$ spatial dimensions from $2 + \epsilon$ expansion (Sec. III), perturbative RG in fixed dimension (Sec. IV), and functional RG (Sec. VI). For comparison: $1/N$-expansion result from Ref. [18]. For visualization purposes, the vertical axis has been rescaled by $N/(1 + N)$. For $N < N_c$ the system is unstable towards a nematic ground state with full, but anisotropic gap. For the systems with the band structure equivalent to that of HgTe, this state corresponds to a strong topological Mott insulator (STMI). For $N > N_c$ the systems remains semimetallic, but exhibits unusual exponents in various observables—a non-Fermi liquid state (NFL) [13]. All these results place the physical situation for $N = 1$ and $d = 3$ (dotted lines) on the Mott insulating side of the transition.

small N and/or near $d = 2$, and the scale-invariant non-Fermi liquid state at large N and/or near $d = 4$; see Fig. 1. In full agreement with our previous results, we find all these approaches to place the physical situation for $d = 3$ and $N = 1$ on the insulating side of the transition. The resulting nematic ground state can be understood as if the material were under, in this case "dynamically generated," uniaxial strain. In the systems with the band structure equivalent to that of HgTe or α-Sn, it corresponds to a strong topological insulator with dynamically generated band gap—a strong topological *Mott* insulator.

The paper is organized as follows: We describe our model in Sec. II. In Secs. III and IV we derive and discuss the $(2 + \epsilon)$-expansion results and the results from perturbative expansion in fixed $d = 3$. The properties of the nematic QCP are examined within an effective Gross-Neveu-Yukawa theory in Sec. V, within the expansion around the upper critical dimension $d = 4$. Section VI discusses our FRG approach to the problem. Concluding remarks are given in Sec. VII.

II. MODEL

Consider binary II-VI compounds crystallizing in the zinc blende structure, such as CdTe and HgTe. CdTe is a semiconductor with a direct band gap of $\varepsilon_g \simeq 1.6$ eV at the center Γ of the Brillouin zone [6]. In the nonrelativistic limit the valence-band upper edge at crystal momentum $\vec{k} = 0$ consists of six degenerate p states with orbital angular momentum quantum number $\ell = 1$. Spin-orbit coupling partially lifts this degeneracy, giving rise to an energetically lower-lying doublet with total angular momentum $j = 1/2$ ("$p_{1/2}$" or "Γ_7" states) and an energetically higher-lying quadruplet with total angular momentum $j = 3/2$ ("$p_{3/2}$" or "Γ_8" states). The $p_{3/2}$ states at
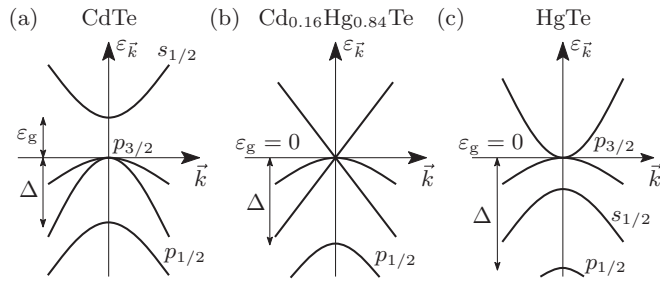
FIG. 2. Schematic electron band structure of II-VI zinc blende crystals near Γ point. In the finite-gap semiconductor CdTe (a) the spin-orbit coupling splits the three p-type bands into a lower-lying $j = 1/2$ band ($p_{1/2}$) and an energetically higher multiplet with total angular momentum $j = 3/2$ ($p_{3/2}$). In the solid solution $Cd_{1-x}Hg_xTe$ spin-orbit coupling increases for increasing x. The $p_{3/2}$ multiplet is therewith shifted towards the conduction s band, culminating in a linear band crossing at $x \approx 0.84$ (b). For even larger spin-orbit coupling the s band exchanges roles with one of the $p_{3/2}$ bands ("band inversion"), and the band gap consequently remains identically zero, with the conduction and valence bands touching *quadratically* at the Fermi level. This is the situation in the gapless semiconductor HgTe (c).

$\vec{k} = 0$ form the upper edge of the heavy- and light-hole bands. They are shifted by the spin-orbit interaction towards the energetically lowest conduction-band states which originate from the s states of the next shell ("$s_{1/2}$" or "Γ_6" states), thereby reducing the band gap; see Fig. 2(a). Experimentally, the size of the band gap can be tuned by gradually substituting Cd atoms by Hg atoms in the solid solution $Cd_{1-x}Hg_xTe$. While Cd and Hg have the same valence shell electron configuration, the main difference is the influence of the relativistic effects due to the higher nuclear charge of Hg. Upon increasing x in $Cd_{1-x}Hg_xTe$ the gap between the $s_{1/2}$ electron band and the $p_{3/2}$ light- and heavy-hole bands shrinks; simultaneously, the effective masses of the electron band and the light-hole band decrease with increasing x. Eventually, at $x \approx 0.84$, the bands touch and the dispersion of both the electron band and the light-hole band becomes linear. By contrast, the effective mass of the heavy-hole band remains finite; see Fig. 2(b). For $x > 0.84$ the $s_{1/2}$ electron states drop below the $p_{3/2}$ multiplet and the curvature of the s-type band turns negative: it converts into a valence band. At the same time the curvature of the $p_{3/2}$ light-hole band becomes positive, thus now representing a conduction band; see Fig. 2(c). The band structure is *inverted* and exhibits nontrivial topology with protected Dirac surface states [26]. The system, however, is not (yet) a topological insulator: The degeneracy of the bulk $p_{3/2}$ states at the Γ point is protected by the crystal symmetry. The band gap is therefore identically zero as long as the (discrete) rotational symmetry is not explicitly (e.g., by external strain) or spontaneously (e.g., by interactions) broken. Moreover, in the undoped system, the Fermi level is locked right at the QBT. The touching point is fourfold degenerate and consists of states with total angular momentum quantum number $j = 3/2$. Consequently, the effective low-energy Hamiltonian can be written in terms of 4×4 angular momentum representation matrices $J_x, J_y, J_z \in \mathbb{C}^{4 \times 4}$. Taking the crystal's cubic symmetry into account, the

only possible invariants that can be constructed out of these matrices and which are quadratic in momentum $\vec{k}$ are given by

$$\mathbb{1}_{4 \times 4}, \quad (\vec{k} \cdot \vec{J})^2, \quad k_x^2 J_x^2 + k_y^2 J_y^2 + k_z^2 J_z^2. \tag{1}$$

In the above, the first two invariants respect the full spherical $O(3)$ symmetry, while the third one breaks the rotational symmetry down to the discrete cubic $C_4 \times C_4$ symmetry. Experimentally, the degree of nonsphericity is usually relatively small [6]. In fact, the full spherical symmetry is expected to be *emergent* at low energies [15].

A low-energy model for the Fermi systems with quadratic band touching is therefore given by the spherically symmetric Luttinger Hamiltonian [27]

$$H_0(\vec{k}) = \frac{1}{2m_0}\left[\left(\alpha_1 + \frac{5}{2}\alpha_2\right)p^2\mathbb{1}_{4\times4} - 2\alpha_2(\vec{k} \cdot \vec{J})^2\right], \tag{2}$$

with electron mass m_0 and phenomenological parameters α_1 and α_2. The spectrum of H_0 is

$$\varepsilon_{\vec{k}} = (\alpha_1 \pm 2\alpha_2)\frac{\vec{k}^2}{2m_0}, \tag{3}$$

and describes a QBT at $\vec{k} = 0$ if $|\alpha_1| < 2|\alpha_2|$. The parameters α_1 and α_2 can be extracted experimentally by fitting, for instance, magnetoabsorbtion measurements to the predictions of the Luttinger model, yielding

$$\alpha_1 \simeq 12.8, \quad \alpha_2 \simeq 8.4, \tag{4}$$

at temperature $T = 4.4$ K [28]. Such values correspond to effective masses of the electron and hole bands near the Γ point as $m_e \simeq \frac{m_0}{30}$ and $m_h \simeq \frac{m_0}{4}$. The electron effective mass is significantly smaller than the hole effective mass. This is because at the band inversion point the former vanishes while the latter remains finite, see Fig. 2(b). Nevertheless, at low temperatures and in pure samples, it is theoretically expected that electron and hole effective masses renormalize and eventually approach a common value towards $T \to 0$—emergent particle-hole symmetry [13,29]. Plasma reflection measurements in p-doped HgTe indeed show a systematic decrease of the hole effective mass with decreasing charge carrier concentration, and this has been attributed to the effect of electron-electron interactions [30]. To the best of our knowledge, however, emergent particle-hole symmetry has not yet been experimentally verified so far.

To simplify the problem, we here assume both particle-hole and spherical symmetry from the outset. The intricate effects of symmetry-breaking terms will be discussed in a separate paper [31]. We thus set $\alpha_1 \equiv 0$ and absorb $2\alpha_2$ in the definition of the effective mass $m \equiv m_0/(2\alpha_2)$. The spectrum hence simply becomes $\varepsilon_{\vec{k}} = \pm k^2/(2m)$. The Luttinger Hamiltonian can then be written in a form that allows its immediate generalization to arbitrary dimension d [17]:

$$H_0(\vec{k}) = \sum_{a=1}^{(\frac{d}{2}+1)(d-1)} d_a(\vec{k})\gamma_a, \tag{5}$$

with Hermitian Dirac matrices γ_a, $a = 1, \ldots, (d/2 + 1)(d - 1)$, satisfying the Clifford algebra $\{\gamma_a, \gamma_b\} = 2\delta_{ab}$. They have dimension $d_\gamma = 2^{\lfloor (d+2)(d-1)/4 \rfloor}$ with $\lfloor \cdot \rfloor$ symbolizing the floor function. The functions $d_a(\vec{k})$ denote real hyperspherical

harmonics for the angular momentum of two on the $(d-1)$ sphere:

$$d_a(\vec{k}) = \sqrt{\frac{d}{2(d-1)}} \sum_{i,j=1}^{d} k_i \Lambda_{a,ij} k_j, \qquad (6)$$

with the generalized real Gell-Mann matrices Λ_a as constructed in Ref. [17]. In $d=3$, they read

$$\Lambda_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\Lambda_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \Lambda_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\Lambda_5 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \qquad (7)$$

From this we obtain, for $d=3$, $d_1 + i d_2 = \frac{\sqrt{3}}{2} k^2 e^{2i\varphi} \sin^2 \vartheta$, $d_3 + i d_4 = \frac{\sqrt{3}}{2} k^2 e^{i\varphi} \sin 2\vartheta$, and $d_5 = \frac{1}{2} k^2 (3 \cos^2 \vartheta - 1)$, with ϑ and φ as spherical angles in $\vec{k}$ space. The five Dirac matrices $\gamma_1, \ldots, \gamma_5$ in this case have dimension $d_\gamma = 4$. This form of the three-dimensional Luttinger Hamiltonian was previously put forward in the context of finite-gap semiconductors with zinc blende structure in Ref. [32]. Setting $d=2$, we recover the known two-dimensional QBT Hamiltonian [33–35]. For $d=4$, Eqs. (5) and (6) become equivalent to the four-dimensional QBT Hamiltonian proposed by Abrikosov [15].

In the continuum limit, the noninteracting Lagrangian that describes the electronic system with N independent quadratic nodes at the Fermi level is then given by

$$L_0 = \psi_i^\dagger [\partial_\tau + H_0(-i\nabla)] \psi_i, \qquad i = 1, \ldots, N. \qquad (8)$$

Each Grassmann field $\psi_1, \ldots, \psi_N$ has d_γ components. τ denotes imaginary time. In Eq. (8) and from now on, where unambiguous, we assume summation convention over repeated indices. The physical situation that describes HgTe and α-Sn, as well as the pyrochlore iridates and half-Heusler compounds, is given by $N=1$. For generality and computational clarity, however, we allow an arbitrary number N of fermion species, i.e., Fermi nodes.

As the Coulomb interaction is only partially screened [18], its effects are crucial. We take them into account by introducing a scalar field a which mediates the long-range interaction via

$$L_a = \frac{1}{2} (\nabla a)^2 + i e a \psi_i^\dagger \psi_i. \qquad (9)$$

Here e is the effective charge, into which all numerical constants of the system have been absorbed, $e^2 = 2m e_0^2/(4\pi \hbar^2 \kappa)$. e_0 is the electron charge and κ denotes the "background" dielectric constant arising from energy bands away from the Fermi level [36]. Integrating out the Coulomb field a in the partition function would generate a nonlocal density-density interaction $\propto 1/k^2$ in Fourier space, which is $\propto 1/r^{d-2}$ in real space, and thus of the usual form in $d=3$.

FIG. 3. Feynman diagrams that generate local four-fermion interactions from the long-range Coulomb interactions at one-loop order. Straight (wiggly) inner lines correspond to fermion (Coulomb) propagators.

III. $2 + \epsilon$ EXPANSION

In a RG treatment, new effective interactions which are not present microscopically may be generated by loop corrections. Most of them are power-counting irrelevant and can be safely neglected. Local four-fermion interactions, however, are marginal in $d=2$ and could thus become relevant at an interacting fixed point in $d=2+\epsilon$. We therefore have to take them into account as well. In fact, the diagrams shown in Fig. 3 generate a local four-fermion term of the form $\propto e^4 (\psi_i^\dagger \gamma_a \psi_i)^2$ from the long-range interaction. Once generated, this term generates further contact interactions. At one-loop order, we find that all possible four-fermion interactions can be written as a linear combination of the following three basis terms:

$$L_\psi = g_1 (\psi_i^\dagger \psi_i)^2 + g_2 (\psi_i^\dagger \gamma_a \psi_i)^2 + g_3 (\psi_i^\dagger \gamma_{ab} \psi_i)^2, \qquad (10)$$

with $\gamma_{ab} = \frac{i}{2} [\gamma_a, \gamma_b]$ and couplings g_α, $\alpha = 1,2,3$.

The four-fermion couplings have mass dimension $[g_\alpha] = 2 - d$, suggesting that an ϵ expansion around $d=2$ may be feasible. The charge e, however, has dimension $[e^2] = 4 - d$ and is thus strongly RG relevant towards the infrared. In order to gain full control over the perturbative expansion, we have to take the limit of both small ϵ and large N. In this double-expansion limit proper fixed points become weakly coupled in all interaction channels. Fortunately, as we shall see below, the fixed-point annihilation we are after automatically happens at large N as long as ϵ is small. The scenario is therefore under perturbative control with a single control parameter $\epsilon = d - 2$.

The RG flow in the full theory space, given by

$$L = L_0 + L_a + L_\psi, \qquad (11)$$

is obtained by integrating out a thin momentum shell from the ultraviolet cutoff Λ to Λ/b with $b > 1$. At one-loop order, we find the flow equations

$$\frac{de^2}{d \ln b} = (2 + z - d - \eta_a) e^2, \qquad (12)$$

$$\frac{dg_1}{d \ln b} = (z - d) g_1 - (e^2 + 2g_1) g_2 - 24 g_3^2, \qquad (13)$$

$$\frac{dg_2}{d \ln b} = (z - d) g_2 + \frac{4(e^2 + 2g_1) g_2}{5} - \frac{(e^2 + 2g_1)^2}{20}$$
$$- \frac{37 + 16N}{5} g_2^2 + \frac{112}{5} g_2 g_3 - \frac{136}{5} g_3^2, \qquad (14)$$

$$\frac{dg_3}{d \ln b} = (z - d) g_3 - \frac{1}{5} (e^2 + 2g_1) g_3 + g_2^2 - 6 g_2 g_3$$
$$+ \frac{4(11 - 4N)}{5} g_3^2, \qquad (15)$$

with the anomalous dimension of the Coulomb field η_a and dynamical exponent z as

$$\eta_a = Ne^2, \quad z = 2 - \frac{4}{15}e^2. \quad (16)$$

In order to arrive at Eqs. (12)–(16) we have rescaled the couplings as $e^2\Lambda^{d+\eta_a-z-2}S^d/(2\pi)^d \mapsto e^2$ and $g_\alpha\Lambda^{d-z}S^d/(2\pi)^d \mapsto g_\alpha$ with S^d as the surface area of the $(d-1)$ hypersphere in d dimensions. We have performed the angular integration directly in $d = 3$, and have employed the 4×4 representation of the Dirac matrices γ_a, with $a = 1, \ldots, 5$. The dimension of the couplings e^2 and g_α is counted in general dimension d. These flow equations generalize those previously published [16] to general fermion number $N > 1$. To see that they coincide with the latter in the limit $N \searrow 1$ we note that in this limit one of the terms in Eq. (10) can be eliminated in favor of the other two by making use of the Fierz identity [16,37]

$$\sum_{a,b=1}^{5}(\psi^\dagger\gamma_{ab}\psi)^2 = -10(\psi^\dagger\psi)^2 - 2\sum_{a=1}^{5}(\psi^\dagger\gamma_a\psi)^2. \quad (17)$$

For $N = 1$ we can thus rewrite Eq. (10) as

$$L_\psi^{N=1} = (g_1 - 10g_3)(\psi^\dagger\psi)^2 + (g_2 - 2g_3)\sum_{a=1}^{5}(\psi^\dagger\gamma_a\psi)^2, \quad (18)$$

which is the form employed in Ref. [16]. If we now shift the couplings in Eqs. (13)–(15) appropriately,

$$g_1 - 10g_3 \mapsto g_1, \quad g_2 - 2g_3 \mapsto g_2, \quad (19)$$

we find that the flow equations for the shifted couplings g_1 and g_2 indeed become independent of the third coupling g_3 when $N = 1$:

$$\frac{dg_1}{d\ln b} = (z-d)g_1 - (e^2 + 2g_1)g_2 - 10g_2^2, \quad (20)$$

$$\frac{dg_2}{d\ln b} = (z-d)g_2 - \frac{(e^2 + 2g_1)^2}{20} + \frac{4(e^2 + 2g_1)g_2}{5} - \frac{63}{5}g_2^2. \quad (21)$$

Moreover, Eqs. (20) and (21) in conjunction with Eqs. (12) and (16) for $N = 1$ are evidently equivalent to the flow equations published in Ref. [16]. Our results for general N are therefore continuously connected to the $N = 1$ case, the presence of a third independent coupling for $N > 1$ notwithstanding.

Some comments on the structure of the flow equations are in order:

(1) If we start the RG for small (but finite) initial charge $e^2 > 0$ and vanishing contact interactions $g_\alpha = 0$ (as relevant for HgTe and α-Sn), e^2 will flow to larger values towards the infrared until the anomalous dimension η_a becomes of the order $2 + z - d$ when the flow of e^2 slows down and eventually stops for $\eta_a = 2 + z - d$ [see Eq. (12)].

(2) The beta function for the charge e^2 has an especially simple form [38]; in particular, there is no vertex correction $\propto e^2 g_\alpha$ in $(de^2)/(d\ln b)$. That this happens here at one loop is not a coincidence, but basically a consequence of the Ward

identity associated to the gauge symmetry $\psi \mapsto e^{ie\lambda(\tau)}\psi, a \mapsto a - \partial_\tau\lambda$. We therefore expect the form of Eq. (12) [but not Eq. (16)] to hold at arbitrary loop order. In this way we obtain an *exact* relation for the Coulomb anomalous dimension η_a at a putative charged fixed point,

$$\eta_a = 2 + z - d, \quad (22)$$

with z being the (presumably nontrivial) dynamical exponent at the fixed point. This resembles the analogous situation in the Abelian Higgs model and in QED$_{2+1}$, where similar exact relations for the gauge anomalous dimensions are known [23,39]. The form of the (marginally) screened Coulomb potential at a charged fixed point is therefore $V(r) \propto 1/r^z$ exactly. This is in agreement with the large-N result of Ref. [18].

(3) Using the exact relation (22) together with the one-loop formulas for η_a and z in Eq. (16) we obtain the fixed-point value for the charge: $e_*^2 = (4-d)/N + O(1/N^2)$, which is under perturbative control in the limit of large N. In this limit we find

$$\eta_a = (4-d)\left[1 - \frac{4}{15N} + O(1/N^2)\right], \quad (23)$$

$$z = 2 - \frac{4(4-d)}{15N} + O(1/N^2), \quad (24)$$

at a putative charged fixed point, in agreement with Ref. [13].

In the following we describe the fixed-point structure in the double-expansion limit $1/N \ll \epsilon \ll 1$ with $\epsilon = d - 2$. It will prove convenient to consider a finite and fixed, but arbitrary, product $n \equiv N\epsilon^2$. In the limit of small ϵ with fixed n, the flow equations decouple and no contact interactions $\propto g_1$ and $\propto g_3$ will be generated by the charge, if absent initially. The coupling g_2, however, will be generated according to the flow equation

$$\frac{dg_2}{d\ln b} = -\epsilon g_2 - \frac{16g_2^2}{5} - \frac{ne^4}{20\epsilon^2}, \quad (25)$$

where we have rescaled $g_\alpha \mapsto g_\alpha/N$ for convenience. The charge sector has a stable fixed point at $e_*^2 = 2\epsilon^2/n + O(\epsilon^4)$. With this value for the charge, the above flow equation has fixed points at

$$g_{2*}^\pm = -\frac{5\epsilon}{32}\left(1 \pm \sqrt{1 - \frac{64/25}{n}}\right) + O(\epsilon^2). \quad (26)$$

They are located at real values of the coupling if and only if $n \geqslant 64/25$, i.e.,

$$N \geqslant N_c(\epsilon) = \frac{64}{25\epsilon^2} + O(1/\epsilon), \quad (27)$$

to the leading order in $\epsilon = d - 2$.

Higher-order loop corrections indeed contribute only to subleading order $O(1/\epsilon)$ to N_c, as anticipated in the above equation. This can be argued as follows: To two-loop order, there are three classes of diagrams that contribute to the beta function of g_2 as $\propto g_2^3$, $\propto e^2 g_2^2$, and $\propto e^6$, respectively. Examples for each class are given in Fig. 4. Without explicitly evaluating the diagrams, we can deduce their leading-order scaling with N by counting the number of closed fermion loops in each diagram. The examples in Fig. 4 are representatives of those
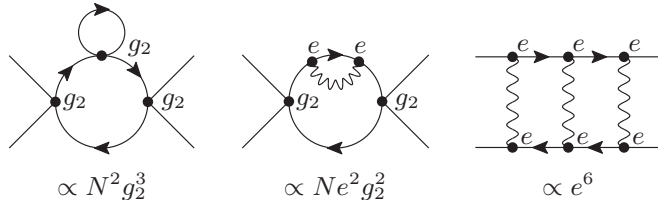
FIG. 4. Two-loop Feynman diagrams contributing to the flow of g_2 at large N. From the number of closed fermion loops we can deduce the scaling of these terms with N to leading order in $1/N$, which after the rescaling [as below Eq. (25)] yield the contributions as displayed in Eq. (28).

diagrams that contribute to the leading order for large N. From this we obtain the form of the two-loop corrections to the flow of g_2 within our double-expansion limit as [after the rescaling as below Eq. (25)]

$$\left.\frac{dg_2}{d\ln b}\right|_{\text{two loop}} = c_1 g_2^3 + c_2 e^2 g_2^2 + \frac{c_3 n e^6}{\epsilon^2}, \qquad (28)$$

with *fixed* coefficients $c_{1,2,3}$. These terms evidently contribute only to order $O(\epsilon^3)$ to the flow of g_2 when $g_2 = O(\epsilon)$ and $e^2 = O(\epsilon)$, while the tree-level and one-loop terms given in Eq. (25) are of order $O(\epsilon^2)$. We expect a similar hierarchy to hold also beyond the two-loop order. Consequently, the one-loop value for N_c as displayed in Eq. (27) is the correct leading-order value for small ϵ.

The presence (absence) of the two charged fixed points for N above (below) N_c has striking implications for the structure of the RG flow. Consider first $N > N_c$: The fixed point at g_{2*}^- is fully infrared attractive, representing a scale-invariant phase with gapless fermions but nontrivial exponents as given in Eqs. (23) and (24). The weakly interacting regime with $e^2 \ll 1$ and $g_\alpha = 0$ indeed lies in this fixed point's basin of attraction. The fixed point represents a conformal phase and is nothing but the Abrikosov-Beneslavskii non-Fermi-liquid fixed point previously found at large N [13,15]. The fixed point at g_{2*}^+, on the other hand, is a quantum critical point with precisely one RG relevant direction. It was found earlier within a simple perturbative RG analysis for $N = 1$ [16]. The strongly interacting regime for $g_2 < g_{2*}^+ < 0$ is no longer in the basin of attraction of the Abrikosov-Beneslavskii NFL fixed point, but exhibits an instability towards divergent coupling $g_2 \to -\infty$ at finite RG scale. The transition is governed by the QCP at g_{2*}^+. In order to elucidate the nature of the instability and the corresponding infrared phase we add to the Lagrangian various types of infinitesimally small symmetry-breaking bilinears

$$L_\Delta = \Delta_{\text{nem}} \psi_i^\dagger \gamma_5 \psi_i + \Delta_{\text{sc}} \psi_i^\dagger \gamma_{45} \psi_i^* + \Delta_{\text{sc}}^* \psi_i^T \gamma_{45} \psi_i$$
$$+ \Delta_{\text{ch}} \psi_i^\dagger \psi_i + \Delta_{\text{mag}} \psi_i^\dagger \gamma_{45} \psi_i, \qquad (29)$$

where $\gamma_{45} \equiv i\gamma_4\gamma_5$. The *nematic* order parameter Δ_{nem} breaks the rotational symmetry and can be understood as originating from uniaxial strain in z direction [13,40]. Finite Δ_{nem} opens a full, but anisotropic gap in the spectrum and converts the semimetal into a three-dimensional topological insulator [26,41]. In the related uncharged system with the long-range Coulomb interaction neglected, nematic quantum

criticality was previously extensively investigated by devising and exploiting the corresponding Gross-Neveu-Yukawa theory in $d = 4 - \epsilon$ dimensions [17]. The superconducting s-wave order parameter Δ_{sc}, on the other hand, breaks $U(1)$ charge symmetry. A corresponding superconducting quantum critical point for strong attractive contact interactions was also recently investigated [29]. Δ_{ch} induces a finite charge density, and represents a finite chemical potential. Finally, Δ_{mag} breaks time reversal. The corresponding magnetic quantum critical point, which in the pyrochlore iridates governs a transition towards an all-in-all-out antiferromagnet, was previously studied at large N [42]. To the leading order in the present $(\epsilon, 1/N)$ double expansion with fixed $n = N\epsilon^2$ the flow of these "mass parameters" reads

$$\frac{d\Delta_{\text{nem}}}{d\ln b} = \left(z - \frac{16g_2}{5} + \frac{2e^2}{5}\right)\Delta_{\text{nem}} + O(\Delta_\alpha^2), \qquad (30)$$

$$\frac{d\Delta_{\text{sc}}}{d\ln b} = \left(z - \frac{e^2}{2}\right)\Delta_{\text{sc}} + O(\Delta_\alpha^2), \qquad (31)$$

$$\frac{d\Delta_{\text{ch}}}{d\ln b} = z\,\Delta_{\text{ch}} + O(\Delta_\alpha^2), \qquad (32)$$

$$\frac{d\Delta_{\text{mag}}}{d\ln b} = \left(z + \frac{e^2}{5}\right)\Delta_{\text{mag}} + O(\Delta_\alpha^2), \qquad (33)$$

where we have applied the same rescalings as below Eqs. (16) and (25). Near the QCP (e_*^2, g_{2*}^+) the free energy density has a scaling form [43]

$$f(\delta g, \Delta_\alpha) = |\delta g|^{(d+z)/y}\, F_\alpha^\pm\!\left(\frac{\Delta_\alpha}{|\delta g|^{x_\alpha/y}}\right), \qquad (34)$$

where $\delta g \equiv g_2 - g_{2*}$ defines the distance to criticality and $F_\alpha^\pm$ is a scaling function. The exponents x_α and y are given by the linearized flow of Δ_α and δg,

$$\frac{d\Delta_\alpha}{d\ln b} = x_\alpha \Delta_\alpha + O(\Delta^2), \quad \frac{d\delta g}{d\ln b} = y\,\delta g + O(\delta g^2), \quad (35)$$

with $\alpha \in \{\text{nem,sc,ch,mag}\}$. The corresponding susceptibilities χ_α therefore scale as

$$\chi_\alpha = -\frac{\partial^2 f}{\partial \Delta_\alpha^2} \propto |\delta g|^{-\gamma_\alpha} \quad \text{with} \quad \gamma_\alpha = \frac{2x_\alpha - d - z}{y} \quad (36)$$

near the QCP. χ_α diverges if $\gamma_\alpha > 0$. From Eqs. (26) and (30)–(33) we find for $n > 64/25$ to the leading order

$$\gamma_{\text{nem}}/\nu = \epsilon\sqrt{1 - \frac{64/25}{n}} + O(\epsilon^2), \qquad (37)$$

$$\gamma_{\text{sc}}/\nu = -\epsilon + O(\epsilon^2), \qquad (38)$$

$$\gamma_{\text{ch}}/\nu = -\epsilon + O(\epsilon^2), \qquad (39)$$

$$\gamma_{\text{mag}}/\nu = -\epsilon + O(\epsilon^2), \qquad (40)$$

with the correlation-length exponent $\nu = 1/y = 1/[\epsilon\sqrt{1 - 64/(25n)}]$. At the QCP, there is therefore a unique ordering tendency that corresponds to a positive susceptibility exponent: the nematic instability with order parameter $\langle \psi_i^\dagger \gamma_5 \psi_i \rangle$ and exponent $\gamma_{\text{nem}} = 1 + O(\epsilon)$. Note also that Δ_{ch} (chemical potential) couples to the conserved
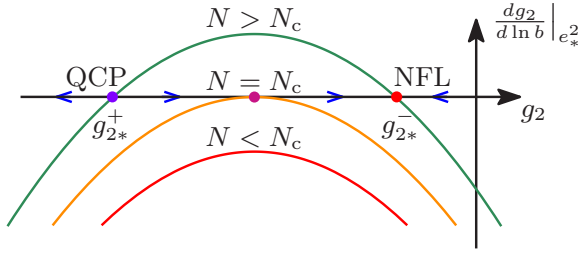
FIG. 5. Schematic beta function for g_2 with e^2 at its infrared fixed-point value e_*^2. Blue arrows on horizontal axis indicate RG flow towards infrared. For $N \searrow N_c$ the QCP and the NFL fixed point merge and annihilate for $N < N_c$, leaving behind the runaway flow towards negative g_2.

charge in the theory and as such did not receive perturbative corrections in Eq. (32).

Consider now N near N_c: The two fixed points at g_{2*}^+ and g_{2*}^- then approach each other and eventually merge at $N \searrow N_c$. Below N_c, they disappear into the complex-coupling plane. The beta function $(dg_2)/(d \ln b)$ [Eq. (25)] for $e^2 = e_*^2$ is then always negative, see Fig. 5.

The flow for $N < N_c$ is therefore towards divergent negative g_2 for *all* ultraviolet starting values for g_2, i.e., even in the weakly interacting limit with $g_2|_{UV} \approx 0$, relevant for HgTe and α-Sn.

When the fixed-point annihilation takes place, the correlation-length exponent ν at the QCP diverges, which is consistent with the infinite-order transition at N_c, see below. The susceptibility exponent γ_{nem}, on the other hand, remains finite. Note that the structure of the flow diagram changes only locally near $g_{2*}^+ = g_{2*}^-$ when N crosses N_c, with the behavior away from this merging point remaining unchanged. By continuity, we therefore expect that the nature of the infrared phase for N near and below N_c with small $g_2|_{UV} \approx 0$ is the *same* as the infrared phase for N near and above N_c with $g_2|_{UV} < g_{2*}^+$.

This way we conclude that the electronic systems with quadratic Fermi nodes and (even weak) long-range Coulomb interaction is unstable towards the nematic ordering when $N < N_c$, with $N_c = 64/(25\epsilon^2) + O(1/\epsilon)$ to the leading order in the $2 + \epsilon$ expansion. This is in agreement with the previous result for $N = 1$ [16]. Naive extrapolation of the critical fermion number to the physical dimension $d = 3$ leads to $N_c(d = 3) \simeq 2.56$, and thus above the physical case for $N = 1$. The low-N system appears as if under, in this case dynamically generated, uniaxial strain, and represents a topological Mott insulator.

Near and below N_c, the RG flow effectively slows down at the merging point. Upon integrating the flow equations we find that the RG "time" b it takes the flow of g_2 to diverge is

$$b_0 = \exp\left(\frac{\pi/\epsilon}{\sqrt{\frac{N_c}{N} - 1}} + O[(N_c/N - 1)^0]\right). \quad (41)$$

For $N \lesssim N_c$, the dynamically generated gap Δ is hence exponentially suppressed,

$$\Delta \propto b_0^{-z}. \quad (42)$$

The above scaling law has an essential singularity at $N \nearrow N_c$, much like the thermal Berezinskii-Kosterlitz-Thouless transition [43], and quite typical for the present type of conformal phase transition [44]. The result is consistent with the form previously derived for the 3D QBT system by solving the Dyson-Schwinger equations directly in $d = 3$ within the $1/N$ expansion [18]. Here this infinite-order transition as function of N follows as a direct consequence of the fixed-point annihilation mechanism. Analogous scaling laws for N-dependent transitions are known to hold also for conformal phase transitions in QED$_3$ [19,23,24], the Abelian Higgs model [39,43,45], and many-flavor quantum chromodynamics [46], and in all of these cases are due to an analogous mechanism.

IV. PERTURBATIVE RG IN FIXED $d = 3$

Now that the *existence* of a finite critical fermion number N_c in $d > 2$ is established within a controlled $2 + \epsilon$ expansion, a natural next step is to predict its *value* in the physical case for $d = 3$. This is a difficult strong-coupling problem. Similar to classical critical phenomena, a reliable theoretical estimate can only be obtained by employing and comparing various different approaches to the problem. While the large-N theory in fixed $d = 3$ has been devised recently [18], another simple approach is the perturbative renormalization group in fixed dimension. This is the subject of the present section, thereby generalizing the $N = 1$ results of Ref. [16] to $N > 1$. Yet another approach to the problem will be employed in Sec. VI.

The downside of this perturbative approach is the lack of a small control parameter, as the fixed-point annihilation will take place in a strong-coupling regime. Ignoring this reservation, we may evaluate the flow equations (12)–(16) directly in $d = 3$. Although the structure of the flow considerably gains in complexity as the different short-range couplings no longer decouple near N_c, the physical conclusions drawn within the $2 + \epsilon$ expansion entirely carry over to the present approach:

Small initial charge e^2 flows to strong coupling towards the infrared with the fixed-point value being $e_*^2 = 15/(15N + 4)$. At a putative charged fixed point the dynamical exponent is $z = 2 - 4/(15N + 4)$ and the Coulomb anomalous dimension is $\eta_a = 15N/(15N + 4)$, in agreement with the exact relation, Eq. (22). For $N > N_c$ with

$$N_c = 2.095 \quad (43)$$

we find an RG attractive NFL fixed point located at real couplings. In the large-N limit it is located at

$$\text{NFL:} \quad (g_{1*}, g_{2*}, g_{3*}) = \left(0, -\frac{1}{20N^2}, 0\right) + O(1/N^3). \quad (44)$$

It governs the infrared behavior of the weakly interacting theory with ultraviolet values $g_\alpha|_{UV} \approx 0$, $\alpha = 1,2,3$, and $e^2|_{UV} > 0$. There is also a QCP with one RG relevant direction. It is located at strong repulsive short-range coupling $g_{2*} < 0$ and weak $|g_{1,3*}| \ll |g_{2*}|$. In the large-N limit its fixed-point values read

$$\text{QCP:} \quad (g_{1*}, g_{2*}, g_{3*}) = \left(0, -\frac{5}{16N}, 0\right) + O(1/N^2). \quad (45)$$

For $N \searrow N_c$ the QCP and the NFL fixed point merge at

$$\text{QCP/NFL:} \quad (g_{1*}, g_{2*}, g_{3*}) \simeq (9.4, -23.7, 0.5) \times 10^{-3}, \tag{46}$$

and annihilate for $N < N_c$, leaving behind the runaway flow towards divergent short-range coupling. In contrast to the situation in $d = 2 + \epsilon$, the interaction channels now do not decouple, and all short-range couplings g_α therefore diverge at the same RG time. Their ratio, however, remains finite, and we find that $|g_1/g_2|$ and $|g_3/g_2|$ always remains small. Such hierarchy is usually taken as an indication that the dominant ordering tendency is the one that corresponds to the strongest short-range coupling [47]. In the present system, strong $g_2 < 0$ leads to nematic order [17]. This suggests to associate the runaway flow with a nematic instability, as done in our previous work [16]. The argument can be solidified by comparing susceptibilities in analogy to the analysis in the preceding section. We determine the flow of infinitesimally small "mass parameters" Δ_α given in Eq. (29),

$$\frac{d\Delta_{\text{nem}}}{d\ln b} = \left(z + \frac{2(e^2 + 2g_1)}{5} - \frac{4(4N+3)g_2}{5} \right) \Delta_{\text{nem}}, \tag{47}$$

$$\frac{d\Delta_{\text{sc}}}{d\ln b} = \left(z - \frac{e^2 + 2g_1}{2} - 5g_2 \right) \Delta_{\text{sc}}, \tag{48}$$

$$\frac{d\Delta_{\text{ch}}}{d\ln b} = z\Delta_{\text{ch}}, \tag{49}$$

$$\frac{d\Delta_{\text{mag}}}{d\ln b} = \left(z + \frac{e^2 + 2g_1}{5} + \frac{2g_2}{5} \right) \Delta_{\text{mag}}, \tag{50}$$

where we have neglected for simplicity terms $\propto g_3 \Delta_\alpha$ as those should give only small corrections of the order of $|g_{3*}/g_{2*}| \simeq 2\%$ to the flow of the Δ_α's near the QCP, see Eqs. (45) and (46). In the large-N limit, we find the susceptibility exponents at the QCP as

$$\gamma_{\text{nem}}/\nu = 1 + O(1/N), \tag{51}$$

$$\gamma_{\text{sc}}/\nu = -1 + O(1/N), \tag{52}$$

$$\gamma_{\text{ch}}/\nu = -1 + O(1/N), \tag{53}$$

$$\gamma_{\text{mag}}/\nu = -1 + O(1/N), \tag{54}$$

where $\nu = 1 + O(1/N)$. The QCP therefore governs the continuous transition towards the nematic state, in agreement with the previous mean-field result [16]. At the QCP-NFL merging point (g_{1*}, g_{2*}, e_*^2) for $N \searrow N_c$ the expansion is no longer under perturbative control, and the one-loop approximation does not necessarily lead to a unique positive susceptibility exponent. We find

$$\gamma_{\text{nem}}/\nu \simeq -0.33, \tag{55}$$

$$\gamma_{\text{sc}}/\nu \simeq -1.32, \tag{56}$$

$$\gamma_{\text{ch}}/\nu \simeq -1.11, \tag{57}$$

$$\gamma_{\text{mag}}/\nu \simeq -0.95. \tag{58}$$

Although the one-loop result for γ_{nem}/ν is no longer positive, it still represents the largest value among the four examined here. This leads us to conclude that the runaway flow we find for $N < N_c$ signals the onset of the nematic instability, in agreement with our results within the $2 + \epsilon$ expansion.

V. $4 - \epsilon$ EXPANSION

It has previously been shown that the properties of the Abrikosov-Beneslavskii NFL fixed point can be assessed by employing an ϵ expansion in $d = 4 - \epsilon$ [13,15,16]. Here we demonstrate that the charged QCP that we found within the $2 + \epsilon$ expansion (Sec. III) as well as the perturbative RG in fixed $d = 3$ (Sec. IV) can similarly be examined in a controlled way within a $4 - \epsilon$ expansion. To this end, we now focus on the nematic channel $\propto g_2$ in L_ψ [Eq. (10)] alone, which in both above approaches turned out to be the most dominant ordering tendency.

The quartic fermionic interaction can be traded for the corresponding Yukawa-type interaction by means of a Hubbard-Stratonovich transformation [17],

$$L_{\psi\phi} = h\phi_a \psi_i^\dagger \gamma_a \psi_i, \tag{59}$$

where ϕ_a, $a = 1, \ldots, (d/2+1)(d-1)$, represents the tensorial [48] nematic order-parameter field and h is the Yukawa coupling. RG loop corrections will generate a kinetic term for ϕ as well as bosonic self-interactions. We thus include these terms from the outset,

$$L_\phi = \tfrac{1}{2}\phi_a\left(-c\partial_\tau^2 - \nabla^2 + r\right)\phi_a + \lambda\phi_a\phi_b\phi_c \operatorname{Tr}(\Lambda_a\Lambda_b\Lambda_c), \tag{60}$$

where Λ_a are the generalized real Gell-Mann matrices introduced in Eq. (6). The form of the cubic interaction parametrized by the coupling λ is dictated by the rotational symmetry of the model [17]. The flow of the parameter c in front of the frequency term in L_ϕ is in general nontrivial and cannot be fixed to unity by simple rescaling. The boson mass r can be understood as a tuning parameter for the nematic transition. The transition is signaled by a nonzero vacuum expectation value $\langle\phi_a\rangle \neq 0$, which is equivalent to $\langle\psi_i^\dagger\gamma_a\psi_i\rangle \neq 0$ for some a. The energetically favored direction of $\phi \equiv (\phi_a)$ leads to a uniaxial nematic state with a full gap in the fermionic spectrum [17]. In $d = 3$, it is given within our conventions by $\langle h\phi_5\rangle > 0$ and $\langle\phi_1\rangle, \ldots, \langle\phi_4\rangle = 0$ [modulo $O(3)$ rotations]. The resulting Gross-Neveu-Yukawa-type field theory is defined by the Lagrangian

$$L = L_0 + L_a + L_{\psi\phi} + L_\phi. \tag{61}$$

The theory is equivalent to the four-fermion model introduced in Eq. (11) upon identifying

$$g_2 \equiv -\frac{h^2}{2r} \tag{62}$$

and setting $g_1 \equiv 0$ and $g_3 \equiv 0$ in the four-fermion theory, as well as setting $\lambda \equiv 0$ and taking the limit of $r \to \infty$ with fixed h^2/r in the order-parameter theory.

The cubic coupling λ, the Yukawa coupling h, and the charge e^2 have engineering dimensions

$$[h^2] = [\lambda^2] = [e^2] = 4 - d. \tag{63}$$

They thus become *simultaneously* relevant below four dimensions. This suggests that the theory's critical behavior can be assessed within an ϵ expansion with small control parameter $\epsilon = 4 - d$. Higher-order interactions are perturbatively irrelevant near $d = 4$ and have for this reason been omitted in L. The analogous Gross-Neveu-Yukawa theory with the long-range Coulomb interaction neglected was investigated previously in Ref. [17]. Here we generalize these results to the case in which $e^2 > 0$.

Integrating the momentum shell from Λ to Λ/b leads to the flow equations

$$\frac{de^2}{d\ln b} = (z + 2 - d - \eta_a)e^2, \tag{64}$$

$$\frac{dc}{d\ln b} = (2 - 2z - \eta_\phi)c + \frac{2N}{5}h^2 + \frac{21}{4}\frac{\sqrt{c}\lambda^2}{(1+r)^{5/2}}, \tag{65}$$

$$\frac{dr}{d\ln b} = (2 - \eta_\phi)r - \frac{8N}{5}h^2 - 21\frac{\lambda^2}{\sqrt{c}(1+r)^{3/2}}, \tag{66}$$

$$\frac{dh^2}{d\ln b} = (6 - d - z - \eta_\phi - 2\eta_\psi)h^2 + \frac{12}{5}\frac{h^4}{1+r} + \frac{4}{5}h^2e^2, \tag{67}$$

$$\frac{d\lambda^2}{d\ln b} = (6 - d - z - 3\eta_\phi)\lambda^2 - \frac{27}{2}\frac{\lambda^4}{\sqrt{c}(1+r)^{5/2}} - \frac{2\sqrt{3}N}{35}\lambda h^3, \tag{68}$$

with the anomalous dimensions and the dynamical exponent as

$$\eta_a = Ne^2, \tag{69}$$

$$\eta_\phi = \frac{44}{35}Nh^2 + \frac{21}{4}\frac{\lambda^2}{\sqrt{c}(1+r)^{5/2}}, \tag{70}$$

$$\eta_\psi = \frac{4}{5}\frac{h^2}{(1+r)^3} + \frac{4}{15}e^2, \tag{71}$$

$$z = 2 - \eta_\psi. \tag{72}$$

Here we have again employed the usual rescalings

$$e^2\Lambda^{d+\eta_a-z-2}S_d/(2\pi)^d \mapsto e^2, \tag{73}$$

$$h^2\Lambda^{d+z+\eta_\phi+2\eta_\psi-6}S_d/(2\pi)^d \mapsto h^2, \tag{74}$$

$$\lambda^2\Lambda^{d+z+3\eta_\phi-6}S_d/(2\pi)^d \mapsto \lambda^2, \tag{75}$$

and $c\Lambda^{2z+\eta_\phi-2} \mapsto c$, $\Lambda^{\eta_\phi-2}r \mapsto r$. Again we have kept the general counting of dimensions in the couplings but have performed the angular integrations and the traces over spinor indices directly in $d = 3$. N consequently counts the number of four-component fermions ψ_i. Similarly, we have fixed the number of components of the nematic order-parameter field $\phi = (\phi_a)$ to be five in all dimensions, $a = 1, \ldots, 5$. An alternative prescription to analytically continue the theory to noninteger dimension, in which the number of components of (ϕ_a) depends on d leads to equivalent critical behavior, at least when $e^2 = 0$ and to the leading order in $\epsilon = 4 - d$ [17].

We have also assumed c to be small at the putative fixed point, $c_* = O(\epsilon^\alpha)$ with $\alpha \geqslant 1$, which turns out to be consistent with the fixed-point values derived below. Equations (65)–(70) and (72) reduce to the flow equations listed in [17] when setting $e^2 \equiv 0$. Equations (64), (71), (69), and (72) also agree with Ref. [13] when setting $h \equiv 0$.

Similarly to the uncharged case [17], the QCP can readily be identified by introducing the new variables

$$u = \frac{\lambda}{c_*^{1/4}}, \quad v = \frac{h}{c_*^{1/12}}, \tag{76}$$

with c_* chosen such that it satisfies the fixed-point equation for c,

$$0 = (2 - 2z)c_* + \left(\frac{2}{5} - \frac{44}{35}c_*\right)Nc_*^{1/6}v^2. \tag{77}$$

We therewith find an interacting charged fixed point to the leading order in $\epsilon = 4 - d$ at

$$e_*^2 = \frac{15\epsilon}{15N + 4} + O(\epsilon^2), \tag{78}$$

$$c_* = O(\epsilon^{6/5}), \tag{79}$$

$$r_* = \frac{6(5N + 4)\epsilon}{15N + 4} + O(\epsilon^{6/5}), \tag{80}$$

$$u_*^2 = \frac{4(5N + 4)\epsilon}{7(15N + 4)} + O(\epsilon^{6/5}), \tag{81}$$

$$v_*^2 = \sqrt[3]{\frac{16(5N + 4)(600N + 515)^2}{21N}}\frac{\epsilon}{15N + 4} + O(\epsilon^{6/5}), \tag{82}$$

with u_* and v_* having opposite signs. Note that the two fixed points at $u_* > 0$, $v_* < 0$ and $u_* < 0$, $v_* > 0$ are physically equivalent as the partition function is invariant under simultaneous sign change of both h and λ. As can be easily checked, the above fixed point is infrared attractive in the (e^2, c, h, λ) coupling space. The only relevant direction is given by the tuning parameter r, and the fixed point hence indeed represents a QCP. The corresponding critical exponents read

$$z = 2 - \frac{4\epsilon}{15N + 4} + O(\epsilon^{6/5}), \tag{83}$$

$$\eta_a = \frac{15N\epsilon}{15N + 4} + O(\epsilon^{6/5}), \tag{84}$$

$$\eta_\psi = \frac{4\epsilon}{15N + 4} + O(\epsilon^{6/5}), \tag{85}$$

$$\eta_\phi = \frac{(15N + 12)\epsilon}{15N + 4} + O(\epsilon^{6/5}). \tag{86}$$

The correlation-length exponent is obtained from the flow of the tuning parameter as

$$1/\nu = 2 + \frac{15(5N + 4)\epsilon}{15N + 4} + O(\epsilon^{6/5}). \tag{87}$$

The nematic QCP that we have found within the $2 + \epsilon$ expansion at large N can hence be shown to exist also

within the $4 - \epsilon$ expansion by using a Gross-Neveu-Yukawa reformulation of the theory, which also allows us to study the *qualitative* properties of the nematic instability. This represents the main result of this section. We do not expect, however, that the *quantitative* predictions when extrapolating our leading-order results to $\epsilon \to 1$ will accurately describe the physics in $d = 3$. We therefore refrained from displaying the next-to-leading order corrections $\propto \epsilon^{6/5}$, which would be straightforwardly computable from the present one-loop flow equations. The reason is that the QCP will interfere with the fully attractive NFL fixed point once we go sufficiently away from the upper critical dimension. This interplay is suppressed for small ϵ, but not necessarily in $d = 3$, as we have seen in Secs. III and IV, and as we will demonstrate within the Gross-Neveu-Yukawa formulation in Sec. VI.

We now show that the NFL fixed point can also be rediscovered within the present formulation. This is achieved by employing a change of variables according to Eq. (62),

$$g = -\frac{h^2}{2r}, \quad \alpha = \frac{r}{1+r}, \qquad (88)$$

with $g \leqslant 0$ and $0 \leqslant \alpha \leqslant 1$. (Here we suppress the index of $g \equiv g_2$ for simplicity.) The upper bound $\alpha \nearrow 1$ corresponds to the limit of large boson mass $r \to \infty$, in which the order-parameter field decouples. $\alpha \ll 1$ defines the quantum critical region with strong order-parameter fluctuations. From Eqs. (66) and (67) we find the flow equations in the new variables as

$$\frac{d\alpha}{d \ln b} = (1 - \alpha) \left[(2 - \eta_\phi)\alpha + \frac{16N}{5} g\alpha \right.$$
$$\left. - 21(1 - \alpha)^{5/2} \frac{\lambda^2}{\sqrt{c}} \right], \qquad (89)$$

$$\frac{dg}{d \ln b} = (z - d)g - \frac{24}{5}\alpha g^2 + \frac{4}{5}e^2 g - \frac{16N}{5}g^2$$
$$+ 21 \frac{\lambda^2}{\sqrt{c}} \frac{(1 - \alpha)^{5/2}}{\alpha} g, \qquad (90)$$

which exhibit a charged fixed point at $\alpha_* = 1$, $g_* = 0$, and $e_*^2 = 15\epsilon/(15N + 4)$. The fixed point is fully attractive in *all* directions in coupling space, in particular, both $g - g_*$ and $\alpha - \alpha_*$ are irrelevant perturbations near the fixed point. The dynamical exponent is $z = 2 - 4\epsilon/(15N + 4)$ and the Coulomb anomalous dimension is $\eta_a = 15N\epsilon/(15N + 4)$, in precise agreement with the $4 - \epsilon$ expansion results for the Abrikosov-Beneslavskii NFL fixed point given in Ref. [13].

We conclude that the present Gross-Neveu-Yukawa formulation of the theory allows us to study the charged QCP and the NFL fixed point on an equal footing within an ϵ expansion around the upper critical spatial dimension of four. To the leading order in $\epsilon = 4 - d$, we find that both fixed points persist at real couplings for all $N \geqslant 1$. However, as we shall see in the next section, once we go sufficiently away from the upper critical dimension, the two fixed points will approach each other upon lowering N with d held fixed (or equivalently upon lowering d with N held fixed). Eventually, at some critical fermion number $N_c(d)$ [equivalently, some critical dimension $d_c(N)$] the fixed points will merge and
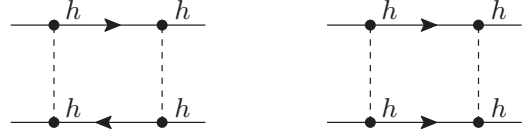

FIG. 6. Feynman diagrams that generate four-fermion interactions from the Yukawa interaction at one-loop order. Solid (dashed) lines correspond to fermion (order-parameter) propagators.

annihilate. As a consequence, the flow from the weak-coupling regime with small $e^2 > 0$ and $h^2/(2r) \approx 0$, $\lambda \approx 0$ is "bended" from the NFL regime towards the symmetry-broken regime in which $r < 0$, leading to a nonvanishing vacuum expectation value $\langle \phi_a \rangle \neq 0$.

VI. FRG WITH DYNAMICAL BOSONIZATION

In order to arrive at the Gross-Neveu-Yukawa field theory in the previous section, we have traded the four-fermion term $g_2(\psi_i^\dagger \gamma_a \psi_i)$ for a Yukawa vertex $h\phi_a \psi^\dagger \gamma_a \psi$. While this allowed us to eliminate the four-fermion interaction at the ultraviolet scale, such terms will inevitably again be generated during the RG flow. At one loop, this happens by means of the box diagrams displayed in Figs. 3 and 6. Within the $4 - \epsilon$ expansion, these terms can be safely neglected as irrelevant, as done in the above. However, if we want to make contact with the results from $2 + \epsilon$ expansion, these terms have to be taken into account, as they become marginal in $d = 2$ and therefore potentially relevant at an interacting fixed point in $2 < d < 4$. Fortunately, their influence can effectively incorporated into the present formulation by means of the so-called "dynamical bosonization scheme" [49]. The idea is to perform a Hubbard-Stratonovich transformation *after every RG step*, such that newly generated four-fermion terms are always again converted into Yukawa interactions at each scale. In this section we implement this strategy entirely on the level of the standard Wilsonian momentum-shell RG. We understand it as a functional RG approach in the sense that perturbatively irrelevant operators are taken into account. In very much the same way, it can be implemented on the level of the flowing effective action [50].

To be explicit, let us write down the effective action after integrating out a thin momentum shell between Λ and Λ/b,

$$S_< = \int_{\vec{k},\omega} \frac{1}{2}(r + \delta r)\phi_a^2 + \int_{\vec{k}_1,\vec{k}_2,\omega_1,\omega_2} (h + \delta h)(\phi_a \psi_i^\dagger \gamma_a \psi_i)$$
$$+ \int_{\vec{k}_1,\vec{k}_2,\vec{k}_3,\omega_1,\omega_2,\omega_3} \delta g(\psi_i^\dagger \gamma_a \psi_i)^2 + \cdots, \qquad (91)$$

where $\int_{\vec{k},\omega} \equiv \int_0^{\Lambda/b} \frac{d\vec{k}}{(2\pi)^d} \int_{-\infty}^\infty \frac{d\omega}{2\pi}$, etc. The ellipsis represents the kinetic terms of ψ, a, and ϕ, as well as the other interaction terms discussed in the previous sections, all of which do not play a direct role in the present discussion and as such are not explicitly displayed for notational simplicity. $\delta r = O(\ln b)$ and $\delta h = O(\ln b)$ denote the explicit loop corrections to the tuning parameter and the Yukawa vertex. They lead to the standard loop contributions to the flows of r and h, as shown at one-loop order in Eqs. (66) and (67). $\delta g = O(\ln b)$ describes the newly generated four-fermion term, which arises from

the diagrams in Figs. 3 and 6 and is assumed to be of the type corresponding to the proposed nematic instability. The partition function reads

$$\mathcal{Z} = \int \mathcal{D}\psi \mathcal{D}\psi^\dagger \mathcal{D}a \mathcal{D}\phi \, e^{-S_<(\psi,\psi^\dagger,a,\phi)}. \tag{92}$$

For any given configuration of ψ, $\psi^\dagger$, and a we can shift ϕ_a in the inner functional integral $\int \mathcal{D}\phi \, e^{-S_<}$ as

$$\phi_a \mapsto \phi_a + \delta\omega \, \psi^\dagger \gamma_a \psi, \tag{93}$$

with arbitrary $\delta\omega = O(\ln b)$. Then, for $\ln b \ll 1$ the partition function becomes

$$\mathcal{Z} = \int \mathcal{D}\psi \mathcal{D}\psi^\dagger \mathcal{D}a \mathcal{D}\phi \, \exp\left\{-\int_{\vec{k},\omega,\dots}\left[\frac{1}{2}(r+\delta r)\phi_a^2\right.\right.$$
$$+ (h + \delta h + r\delta\omega)(\phi_a \psi_i^\dagger \gamma_a \psi_i)$$
$$\left.\left.+ (\delta g + h\delta\omega)(\psi_i^\dagger \gamma_a \psi_i)^2 + \cdots\right]\right\}. \tag{94}$$

Thus, if we choose

$$\delta\omega \equiv -\frac{\delta g}{h}, \tag{95}$$

the newly generated four-fermion terms will be exactly canceled in Eq. (94). At the same time, the Yukawa-coupling flow is modified as

$$\frac{dh^2}{d\ln b} = (6 - d - z - \eta_\phi - 2\eta_\psi)h^2 + 2h\frac{\partial \delta h}{\partial \ln b} - 2r\frac{\partial \delta g}{\partial \ln b}. \tag{96}$$

In the above equation, the second term represents the standard contribution from the explicit vertex renormalization, whereas the last contribution arises from the dynamical bosonization. By computing the box diagrams in Figs. 3 and 6 we find the modified flow

$$\frac{dh^2}{d\ln b} = (6 - d - z - \eta_\phi - 2\eta_\psi)h^2 + \frac{12}{5}\frac{h^4}{1+r} + \frac{4}{5}h^2 e^2$$
$$+ \frac{13}{10}\frac{rh^4}{(1+r)^2} + \frac{re^4}{10}, \tag{97}$$

while the flow equations for e^2, c, r, and λ, as well as the anomalous dimensions η_a, η_ϕ, η_ψ and the dynamical exponent z remain the same as in Eqs. (64)–(72).

Near the upper critical dimension of $d = 4$, the additional two terms in Eq. (97) as compared to Eq. (67) are of subleading order both at the charged QCP as well as the NFL fixed point. Consequently, the fixed-point structure found in the previous section carries over completely to the present FRG approach when $d \nearrow 4$. Let us demonstrate that the dynamically bosonized flow coincides also with the flow as obtained within the $2 + \epsilon$ expansion (Sec. III) in the limit $d \searrow 2$. To this end, we again introduce the variables g and α as in Eq. (88), leading to the modified flow equation

$$\frac{dg}{d\ln b} = (z - d)g - \frac{24}{5}\alpha g^2 + \frac{4}{5}e^2 g - \frac{16N}{5}g^2$$
$$+ 21\frac{\lambda^2}{\sqrt{c}}\frac{(1-\alpha)^{5/2}}{\alpha}g - \frac{13}{5}\alpha^2 g^4 - \frac{e^4}{20}, \tag{98}$$

TABLE I. Fixed-point values and critical exponents at the QCP for different N from functional RG.

N	Ne_*^2	r_*	Nh_*^2	η_a	η_ϕ	$2-z$	$1/\nu$
1.856	0.87	12.75	1.45	0.87	1.82	0.13	0.00
2	0.88	8.24	1.38	0.88	1.73	0.12	0.26
3	0.92	4.12	1.22	0.92	1.53	0.08	0.58
4	0.93	3.08	1.12	0.93	1.41	0.07	0.68
5	0.95	2.58	1.06	0.95	1.34	0.05	0.74
10	0.97	1.81	0.93	0.97	1.17	0.03	0.86
25	0.99	1.46	0.85	0.99	1.07	0.01	0.94
100	1.00	1.32	0.81	1.00	1.02	0.00	0.99
∞	1	$\frac{14}{11}$	$\frac{35}{44}$	1	1	0	1

and $(d\alpha)/(d\ln b)$ as in Eq. (89). For $\alpha \to 1$ we find that the above flow equation for g agrees with Eq. (14) upon setting $g \equiv g_2$ and $g_1 = g_3 = 0$. We therefore conclude that the approximation scheme is under perturbative control both near $d = 2$ and near $d = 4$.

We now turn to the physically interesting case of $d = 3$. In the limit of large N, the fixed-point equations can be solved analytically. In this limit, we recover both the QCP and the NFL fixed point. The former is located at

$$\text{QCP:} (Ne_*^2, c_*, r_*, Nh_*^2, N\lambda_*^2) = \left(1, \frac{7}{22}, \frac{14}{11}, \frac{35}{44}, \frac{105}{85\,184}\right)$$
$$+ O(1/N). \tag{99}$$

From this we find $-h_*^2/(2r_*) = -5/(16N) + O(1/N^2)$, which agrees with the result of the fermionic formulation [Eq. (45)] upon recalling the identification given in Eq. (62). The above fixed point has exactly one RG relevant direction, with the mass r of the order-parameter field ϕ being the tuning parameter for the nematic transition. The NFL fixed point is located in the large-N limit at

$$\text{NFL:} \left(Ne_*^2, c_*, \frac{r_*}{N}, Nh_*^2, N\lambda_*^2\right) = \left(1, \frac{7}{44}, \frac{175}{11}, \frac{35}{22}, \frac{21}{13\,310}\right)$$
$$+ O(1/N). \tag{100}$$

Note the different scaling of r_* with N as compared to the QCP: While the QCP is located at some finite $r_* = O(1)$, the mass parameter at the NFL fixed point becomes large, $r_* = O(N)$. Consequently, fluctuations of the order-parameter field are suppressed at the NFL fixed point for large N. (However, we shall see below that this is not necessarily the case for small N.) At the NFL fixed point we have $-h_*^2/(2r_*) = -1/(20N^2) + O(1/N^2)$, which again precisely agrees with the large-N result of the fermionic formulation, Eq. (44). The NFL fixed point is RG attractive in all directions, with the corrections-to-scaling exponent ω that is determined by the RG flow in the "least-irrelevant" direction being $\omega = 1 + O(1/N)$.

For finite N we have solved the coupled system of fixed-point equations for e^2, c, r, h, and λ numerically. The results for the location of the fixed points and the corresponding universal scaling exponents are given for various N in Tables I and II. Note that the Coulomb anomalous dimension η_a in each case fulfills the exact relation, Eq. (22), as it should be. With decreasing N, we find that the mass of the order-parameter field at the NFL fixed point rapidly decreases, thereby progressively enhancing fluctuations in the nematic channel. The QCP and

TABLE II. Fixed-point values and scaling exponents at the NFL fixed point for different N from functional RG.

N	Ne_*^2	r_*/N	Nh_*^2	η_a	η_ϕ	$2-z$	ω
1.856	0.87	6.87	1.45	0.87	1.82	0.13	0.00
2	0.88	9.81	1.49	0.88	1.88	0.12	0.27
3	0.92	13.33	1.54	0.92	1.94	0.08	0.62
4	0.94	14.29	1.56	0.94	1.96	0.06	0.74
5	0.95	14.74	1.56	0.95	1.97	0.05	0.80
10	0.97	15.43	1.58	0.97	1.98	0.03	0.91
25	0.99	15.74	1.59	0.99	1.99	0.01	0.96
100	1.00	15.87	1.59	1.00	2.00	0.00	0.99
∞	1	$\frac{175}{11}$	$\frac{35}{22}$	1	2	0	1

NFL fixed point approach each other in coupling space, and eventually merge when $N \searrow N_c$ with

$$N_c = 1.856. \qquad (101)$$

For $N < N_c$ and small initial couplings $h^2/(2r)|_{UV} \approx 0$, $\lambda|_{UV} \approx 0$, and $0 < e^2|_{UV} \ll 1$, the RG flow is always towards

the regime in which $r < 0$, signaling the nematic transition and the spontaneous breakdown of the rotational symmetry. The RG flow for different values of N above, at, and below N_c is visualized in Fig. 7.

The fixed-point equations can be solved numerically for all N in any given dimension $2 < d < 4$. We have displayed the result in terms of the critical fermion number $N_c(d)$ in Fig. 1. As evident there, the FRG estimate approaches the result from the $2 + \epsilon$ expansion for $d \searrow 2$, as expected. With the FRG, we can also estimate the critical dimension d_c below which the fixed-point annihilation occurs for fixed $N = 1$. This way we find $d_c(N = 1) = 3.21$ and thus close to the value of $d_c = 3.26$ found within the perturbative RG approach for $N = 1$ [16]. For $d \nearrow 4$, we have explicitly checked that the scaling exponents for both the QCP and the NFL fixed point numerically coincide with their counterparts from the leading-order $4 - \epsilon$ expansion. We reiterate that the FRG approach within the dynamical bosonization scheme becomes one-loop exact both in $d = 2 + \epsilon$ as well as in $d = 4 - \epsilon$, and smoothly interpolates between these two perturbatively accessible limits for intermediate dimension.
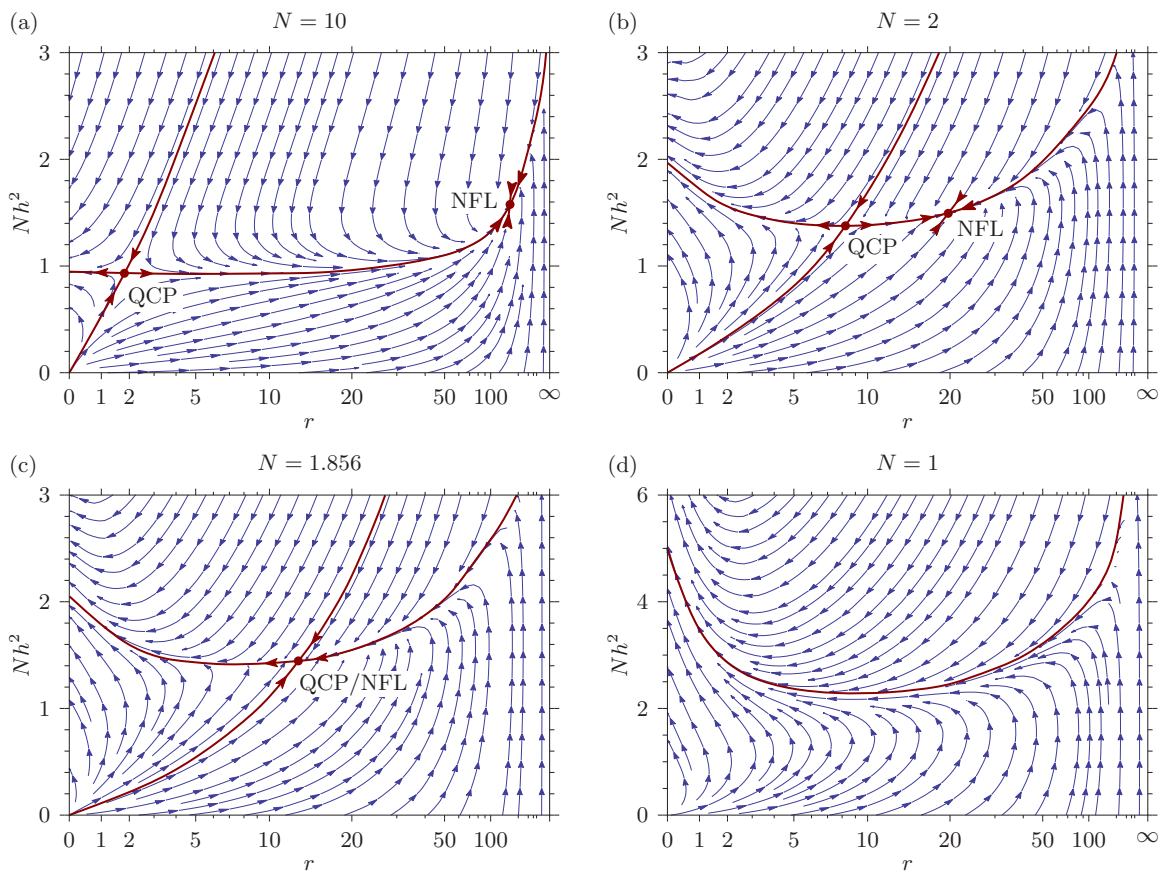

FIG. 7. RG flow diagram in $d = 3$ for $N = 10, 2, 1.856$, and 1 in r-h^2 plane from functional renormalization group, displaying the quantum critical fixed point (QCP) and the fully attractive non-Fermi liquid fixed point (NFL). Arrows point towards infrared. To visualize the flow, the remaining couplings c, e^2, and λ have been fixed at their values at the QCP. (For $N = 1$, for which the QCP does not exist, we have chosen their large-N predictions.) The horizontal axis has been rescaled by $r/(1 + r)$ for reasons of clarity. For $N > N_c$ [(a) and (b)] a weakly correlated material with initially small coupling $h^2/(2r)|_{UV} \ll 1$ flows to the fully attractive non-Fermi liquid (NFL) fixed point. A (hypothetical) strongly correlated material with $h^2/(2r)|_{UV} > [h^2/(2r)]_c$ would flow to negative r, indicating a continuous [17] phase transition towards a nematic state, with the critical behavior governed by the QCP. At $N \searrow N_c = 1.856$ (c) the NFL fixed point merges with the QCP, such that for $N < N_c$ (d) the flow is *always* towards the nematic state, even for small initial coupling.

VII. CONCLUSIONS

In conclusion, we have studied the zero-temperature ground state of gapless semiconductors with quadratic Fermi nodes and weak long-range Coulomb interaction in 3D. We have confined ourselves to a model in which the full rotational symmetry and the particle-hole symmetry is imposed from the outset, leaving the discussion of the subtle effects of these perturbations for a separate publication [31]. The present model serves as the minimal effective low-energy description of the electronic behavior of weakly correlated 3D gapless semiconductors such as HgTe and α-Sn [6], and possibly also certain pyrochlore iridates of the form $R_2Ir_2O_7$ (with R being a rare-earth element) [7,8,13], as well as some half-Heusler compounds [10–12]. In order to gain analytical control over the low-temperature physics, we have extended the model by allowing a general number N of fermion species (which may be understood as the number of QBTs at the Fermi level), as well as by generalizing to arbitrary spatial dimensionality $2 < d < 4$. Our main results are the following:

(1) At large N, the system has a scale-invariant gapless ground state which is characterized by anomalous low-temperature power laws of various thermodynamic observables—a 3D non-Fermi liquid. The specific heat, for instance, would scale in this state as

$$C_V \propto T^{d/z}, \tag{102}$$

with nontrivial dynamical exponent $1 < z < 2$. The emergence of the scale-invariant ground state can be traced back to the existence of a fully infrared stable NFL fixed point. We have computed the universal exponents in this state by employing $2 + \epsilon$ expansion, perturbative RG in fixed $d = 3$, $4 - \epsilon$ expansion, and functional RG in the dynamical bosonization scheme. The results are consistent with the earlier works [13,15].

(2) Upon lowering N, the NFL fixed point approaches another, quantum critical, fixed point. At some critical N_c, the NFL fixed point and the QCP merge and eventually disappear for $N < N_c$ into the complex-coupling plane. As a consequence, the electronic system with a small number of QBTs at the Fermi level is unstable to weak long-range interactions. Such fixed-point annihilation scenario was proposed earlier in the context of a $1/N$ expansion in fixed $d = 3$ [18] as well as a one-loop RG for fixed $N = 1$ and varying dimensionality [16]. In both these earlier approaches, however, the fixed-point annihilation occurs outside the regime in which the expansion is under control, and one may wonder whether higher loop orders may qualitatively change the conclusion. In the present work, we have demonstrated that the scenario can be described in a fully controlled way by employing an expansion around two spatial dimensions. On that point, we have exploited the fortunate fact that for small $\epsilon = d - 2$ the fixed-point annihilation occurs at large N, thereby pushing the interesting physics entirely into the perturbative domain. This proves the existence of a phase boundary between the NFL phase at large N and a novel symmetry-broken phase at small N in the d-N plane, see Fig. 1.

(3) Using a susceptibility analysis in $d = 2 + \epsilon$ dimensions, we find that the instability for $N < N_c$ is towards a nematic state in which the rotational symmetry is spontaneously

TABLE III. Critical fermion number N_c in $d = 3$ spatial dimensions from different approaches.

Method	Reference	$N_c(d = 3)$
$2 + \epsilon$ expansion	Sec. III	2.56
RG in fixed $d = 3$	Sec. IV	2.10
Functional RG	Sec. VI	1.86
$1/N$ expansion in $d = 3$	Ref. [18]	$\geqslant 2.6(2)$

broken. This result is in agreement with all other approaches employed here, as well as with the previous work [16]. The low-N quantum ground state has a full, but anisotropic gap and converts the semimetal into a topological Mott insulator.

(4) We have employed a variety of different approaches to gain a reasonable estimate of the critical fermion number N_c in the limit of $d = 3$. The results are summarized in Table III. All estimates obtained so far consistently place the physical situation for $N = 1$ into the Mott insulating regime. Gapless semiconductors with one quadratic Fermi node in 3D, such as clean HgTe and α-Sn, should therefore suffer from a transition towards a nematic state in which an interaction-induced gap is dynamically generated. For such weakly correlated materials the relevant energy scale at which interaction effects become important is $\varepsilon_* = 1$–10 meV [6,16]. From this and Eq. (42) we estimate the size of the Mott gap Δ for $N_c \approx 2$ and $z \approx 1.7$ as

$$\Delta \approx 10^{-2}\varepsilon_* \approx 0.01\text{–}0.1 \text{ meV}. \tag{103}$$

As function of temperature, we correspondingly expect the Mott transition to occur at a critical temperature T_c of the order $T_c \approx 0.1$–1 K, in agreement with [16]. Besides the opening of the Mott gap, the transition reveals itself experimentally through a thermodynamic singularity at T_c, as measurable, for instance, in the specific heat. Below the jump at T_c, the specific heat is exponentially suppressed,

$$C_V \propto e^{-\Delta/(2k_B T)}, \tag{104}$$

while for $T > T_c$ it is expected to resemble the NFL behavior with nontrivial exponents as in Eq. (102), see Ref. [16]. The Hall coefficient should exhibit a similar temperature dependence [18]. The effects are experimentally accessible if the sample can be prepared sufficiently pure.

To approach the complex physics in the pyrochlore iridates, the nontrivial interplay of the (itinerant) iridium electrons with the (local) rare-earth magnetic moments should be investigated [8]. A separate paper [31] will present a detailed discussion of the effects of deviations from the spherical and particle-hole symmetries which may also be important for this class of materials [42,51].

[1] A. K. Geim and K. S. Novoselov, The rise of graphene, Nat. Mater. **6**, 183 (2007).

[2] O. Vafek and A. Vishwanath, Dirac fermions in solids: From high-T_c cuprates and graphene to topological insulators and Weyl semimetals, Annu. Rev. Condens. Matter Phys. **5**, 83 (2014).

[3] A. A. Burkov and L. Balents, Weyl Semimetal in a Topological Insulator Multilayer, Phys. Rev. Lett. **107**, 127205 (2011).

[4] S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, S.-M. Huang, H. Zheng, J. Ma, D. S. Sanchez, B. Wang, A. Bansil, F. Chou, P. P. Shibayev, H. Lin, S. Jia, and M. Z. Hasan, Discovery of a Weyl fermion semimetal and topological Fermi arcs, Science **349**, 613 (2015).

[5] B. Q. Lv, H. M. Weng, B. B. Fu, X. P. Wang, H. Miao, J. Ma, P. Richard, X. C. Huang, L. X. Zhao, G. F. Chen, Z. Fang, X. Dai, T. Qian, and H. Ding, Experimental Discovery of Weyl Semimetal TaAs, Phys. Rev. X **5**, 031013 (2015).

[6] I. M. Tsidilkovski, *Electron Spectrum of Gapless Semiconductors* (Springer, Berlin, 1997).

[7] T. Kondo, M. Nakayama, R. Chen, J. J. Ishikawa, E.-G. Moon, T. Yamamoto, Y. Ota, W. Malaeb, H. Kanai, Y. Nakashima, Y. Ishida, R. Yoshida, H. Yamamoto, M. Matsunami, S. Kimura, N. Inami, K. Ono, H. Kumigashira, S. Nakatsuji, L. Balents, and S. Shin, Quadratic Fermi node in a 3D strongly correlated semimetal, Nat. Commun. **6**, 10042 (2015).

[8] M. Nakayama, T. Kondo, Z. Tian, J. J. Ishikawa, M. Halim, C. Bareille, W. Malaeb, K. Kuroda, T. Tomita, S. Ideta, K. Tanaka, M. Matsunami, S. Kimura, N. Inami, K. Ono, H. Kumigashira, L. Balents, S. Nakatsuji, and S. Shin, Slater to Mott Crossover in the Metal to Insulator Transition of $Nd_2Ir_2O_7$, Phys. Rev. Lett. **117**, 056403 (2016).

[9] W. Witczak-Krempa, G. Chen, Y. B. Kim, and L. Balents, Correlated quantum phenomena in the strong spin-orbit regime, Annu. Rev. Condens. Matter Phys. **5**, 57 (2014).

[10] S. Chadov, X. Qi, J. Kübler, G. H. Fecher, C. Felser, and S. C. Zhang, Tunable multifunctional topological insulators in ternary Heusler compounds, Nat. Mat. **9**, 541 (2010).

[11] H. Lin, L. A. Wray, Y. Xia, S. Xu, S. Jia, R. J. Cava, A. Bansil, and M. Z. Hasan, Half-Heusler ternary compounds as new multifunctional experimental platforms for topological quantum phenomena, Nat. Mat. **9**, 546 (2010).

[12] D. Xiao, Y. Yao, W. Feng, J. Wen, W. Zhu, X.-Q. Chen, G. M. Stocks, and Z. Zhang, Half-Heusler Compounds as a New Class of Three-Dimensional Topological Insulators, Phys. Rev. Lett. **105**, 096404 (2010).

[13] E.-G. Moon, C. Xu, Y. B. Kim, and L. Balents, Non-Fermi-Liquid and Topological States with Strong Spin-Orbit Coupling, Phys. Rev. Lett. **111**, 206401 (2013).

[14] A. A. Abrikosov and S. D. Beneslavskii, Possible existence of substances intermediate between metals and dielectrics, Sov. Phys. JETP **32**, 699 (1971).

[15] A. A. Abrikosov, Calculation of critical indices for zero-gap semiconductors, Sov. Phys. JETP **39**, 709 (1974).

[16] I. F. Herbut and L. Janssen, Topological Mott Insulator in Three-Dimensional Systems with Quadratic Band Touching, Phys. Rev. Lett. **113**, 106401 (2014).

[17] L. Janssen and I. F. Herbut, Nematic quantum criticality in three-dimensional Fermi system with quadratic band touching, Phys. Rev. B **92**, 045117 (2015).

[18] L. Janssen and I. F. Herbut, Excitonic instability of three-dimensional gapless semiconductors: Large-N theory, Phys. Rev. B **93**, 165109 (2016).

[19] T. Appelquist, D. Nash, and L. C. R. Wijewardhana, Critical Behavior in (2+1)-Dimensional QED, Phys. Rev. Lett. **60**, 2575 (1988).

[20] C. S. Fischer, R. Alkofer, T. Dahm, and P. Maris, Dynamical chiral symmetry breaking in unquenched QED_3, Phys. Rev. D **70**, 073007 (2004).

[21] J. Braun, H. Gies, L. Janssen, and D. Roscher, Phase structure of many-flavor QED_3, Phys. Rev. D **90**, 036002 (2014).

[22] L. Di Pietro, Z. Komargodski, I. Shamir, and E. Stamou, Quantum Electrodynamics in $d = 3$ from the ϵ Expansion, Phys. Rev. Lett. **116**, 131601 (2016).

[23] L. Janssen, Spontaneous breaking of Lorentz symmetry in $(2 + \epsilon)$-dimensional QED, Phys. Rev. D **94**, 094013 (2016).

[24] I. F. Herbut, Chiral symmetry breaking in three-dimensional quantum electrodynamics as fixed point annihilation, Phys. Rev. D **94**, 025036 (2016).

[25] A similar observation was made for the relativistic Gross-Neveu universality classes relevant for the various possible transitions on graphene's honeycomb lattice; see L. Janssen and I. F. Herbut, Antiferromagnetic critical point on graphenes honeycomb lattice: A functional renormalization group approach, Phys. Rev. B **89**, 205403 (2014).

[26] L. Fu and C. L. Kane, Topological insulators with inversion symmetry, Phys. Rev. B **76**, 045302 (2007).

[27] J. M. Luttinger, Quantum theory of cyclotron resonance in semiconductors: General theory, Phys. Rev. **102**, 1030 (1956).

[28] Y. Guldner, C. Rigaux, M. Grynberg, and A. Mycielski, Interband $\Gamma_6 \rightarrow \Gamma_8$ magnetoabsorption in HgTe, Phys. Rev. B **8**, 3875 (1973).

[29] I. Boettcher and I. F. Herbut, Superconducting quantum criticality in three-dimensional Luttinger semimetals, Phys. Rev. B **93**, 205138 (2016).

[30] V. I. Ivanov-Omskii, A. S. Mekhtiev, E. N. Ukraintsev, and S. A. Rustambekova, Plasma reflection and hole optical effective mass in p-HgTe, Phys. Status Solidi B **119**, 159 (1983).

[31] I. Boettcher and I. F. Herbut, Anisotropy induces non-Fermi liquid behavior and nemagnetic order in three-dimensional Luttinger semimetals, arXiv:1611.05904.

[32] S. Murakami, N. Nagaosa, and S.-C. Zhang, SU(2) non-Abelian holonomy and dissipationless spin current in semiconductors, Phys. Rev. B **69**, 235206 (2004).

[33] K. Sun, H. Yao, E. Fradkin, and S. A. Kivelson, Topological Insulators and Nematic Phases from Spontaneous Symmetry Breaking in 2D Fermi Systems with a Quadratic Band Crossing, Phys. Rev. Lett. **103**, 046811 (2009).

[34] O. Vafek and K. Yang, Many-body instability of Coulomb interacting bilayer graphene: Renormalization group approach, Phys. Rev. B **81**, 041401(R) (2010).

[35] B. Dóra, I. F. Herbut, and R. Moessner, Nematic, topological and Berry phases when a flat and a parabolic band touch, Phys. Rev. B **90**, 045310 (2014).

[36] B. I. Halperin and T. M. Rice, Possible anomalies at a semimetal-semiconductor transition, Rev. Mod. Phys. **40**, 755 (1968).

[37] I. F. Herbut, V. Juričić, and B. Roy, Theory of interacting electrons on honeycomb lattice, Phys. Rev. B **79**, 085116 (2009).

[38] I. F. Herbut, Quantum Critical Points with the Coulomb Interaction and the Dynamical Exponent: When and Why $z = 1$, Phys. Rev. Lett. **87**, 137004 (2001).

[39] I. F. Herbut and Z. Tešanović, Critical Fluctuations in Superconductors and the Magnetic Field Penetration Depth, Phys. Rev. Lett. **76**, 4588 (1996).

[40] B. J. Roman and A. W. Ewald, Stress-induced band gap and related phenomena in gray tin, Phys. Rev. B **5**, 3914 (1972).

[41] C. Brüne, C. X. Liu, E. G. Novik, E. M. Hankiewicz, H. Buhmann, Y. L. Chen, X. L. Qi, Z. X. Shen, S. C. Zhang, and L. W. Molenkamp, Quantum Hall Effect from the Topological Surface States of Strained Bulk HgTe, Phys. Rev. Lett. **106**, 126803 (2011).

[42] L. Savary, E.-G. Moon, and L. Balents, New Type of Quantum Criticality in the Pyrochlore Iridates, Phys. Rev. X **4**, 041027 (2014).

[43] I. Herbut, *A Modern Approach to Critical Phenomena* (Cambridge University Press, Cambridge, England, 2007).

[44] D. B. Kaplan, J.-W. Lee, D. T. Son, and M. A. Stephanov, Conformality lost, Phys. Rev. D **80**, 125005 (2009).

[45] B. I. Halperin, T. C. Lubensky, and S.-K. Ma, First-Order Phase Transitions in Superconductors and Smectic-A Liquid Crystals, Phys. Rev. Lett. **32**, 292 (1974).

[46] H. Gies and J. Jaeckel, Chiral phase structure of QCD with many flavors, Eur. Phys. J. C **46**, 433 (2006).

[47] F. Gehring, H. Gies, and L. Janssen, Fixed-point structure of low-dimensional relativistic fermion field theories: Universality classes and emergent symmetry, Phys. Rev. D **92**, 085046 (2015).

[48] Note that ϕ_a can be written as $\phi_a = \sum_{ij}^{d} T_{ij} \Lambda_{a,ij}$ with T_{ij} transforming as a traceless symmetric tensor under spatial rotations, see Ref. [17].

[49] H. Gies and C. Wetterich, Renormalization flow of bound states, Phys. Rev. D **65**, 065001 (2002).

[50] See e.g., L. Janssen and H. Gies, Critical behavior of the (2+1)-dimensional Thirring model, Phys. Rev. D **86**, 105007 (2012).

[51] P. Goswami, B. Roy, and S. Das Sarma, Itinerant spin ice order, Weyl metal, and anomalous Hall effect in $Pr_2Ir_2O_7$, arXiv:1603.02273.