Sine-Gordon mean field theory of a Coulomb gas

Alexandre Diehl, Marcia C. Barbosa, and Yan Levin
Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, 91 501-970 Porto Alegre, Rio Grande do Sul, Brazil
(Received 6 November 1996)

Sine-Gordon field theory is used to investigate the phase diagram of a neutral Coulomb gas. A variational mean-field free energy is constructed and the corresponding phase diagrams in two and three dimensions are obtained. When analyzed in terms of chemical potential, the sine-Gordon theory predicts the phase diagram topologically identical to the Monte Carlo simulations and a recently developed Debye-Hückel-Bjerrum theory. In two dimensions we find that the infinite-order Kosterlitz-Thouless line terminates in a tricritical point, after which the metal-insulator transition becomes first order. However, when the transformation from the chemical potential to the density is made the whole insulating phase is mapped onto zero density. [S1063-651X(97)05206-9]

PACS number(s): 05.70.Fh, 64.70.−p, 64.60.−i

The Coulomb gas provides a paradigm for the study of various models of critical phenomena [1]. In particular it is well known that the two-dimensional (2D) Coulomb gas (CG) can be directly used to study the superfluidity transition in ^{4}He films, arrays of Josephson junctions [2], melting of two-dimensional crystals [3], roughening transition [4], etc. Not withstanding its versatility, our full understanding of the most basic model of Coulomb gas, namely, an ensemble of hard spheres carrying either positive or negative charges at their center, is still lacking.

It is now well accepted that at low density the two-dimensional plasma of equal numbers of positive and negative particles undergoes a Kosterlitz-Thouless (KT) metal-insulator transition [5]. This transition is of an infinite order and is characterized by a diverging Debye screening length. Thus, in the low-temperature phase (insulator) all the positive and negative particles are associated into the dipolar pairs, while in the high-temperature phase (conductor) there exists a finite fraction of unassociated, free charges. As the density of particles increases the validity of the KT theory becomes questionable and the possibility of the KT transition being replaced by some kind of first-order discontinuity has been speculated for a long time [6]. The idea that there can exist a discontinuous transition between the insulating and conducting phases has gained further credence in view of the increasing computational power and an improving algorithm design needed for running large-scale simulations of the particles interacting by a long-range potentials [7]. Thus it has been demonstrated quite convincingly that at high densities the KT infinite-order line becomes unstable and is replaced by a first-order coexistence between the low-density insulating vapor and the high-density conducting fluidlike phase. From the theoretical perspective, however, the nature of this metamorphosis is far from clear. At the moment there appear to exist two competing views of what happens to the two-dimensional plasma at higher densities. The first of these theories, presented by Minnhagen *et al.* [8], predicts that the KT line will terminate in a *critical end point*, while the critical point of the coexistence curve separating the low- and the high-density phases lies in the conducting region. It is important to note that within Minnhagen's theory the portion of the coexistence curve in between the critical point and the critical end point has both vapor (low-density phase) and liquid (high-density phase) that are conducting. Although Minnhagen's approach is often characterized as a version of renormalization group (RG) theory, this is a misnomer. Its basis, which lies in a clever combination of a linear response formalism with some aspects of sine-Gordon (SG) field theory, is much closer to the integral equations of liquid state theory than to the RG theory. The RG methodology is used more as a tool in studying the solutions of the integral equation found by Minnhagen.

An alternative approach suggested by Levin *et al.* [9] is based on a recently developed Debye-Hückel-Bjerrum (DHBj) theory [10]. This method, which is intrinsically of mean-field type, relies on calculating the full electrostatic free energy of the ionic solution based on a linearized Poisson-Boltzmann equation. The effects of linearization are then corrected by allowing for the presence of dipolar pairs, the density of which is determined through the law of mass action. This theory has proven extremely powerful in elucidating the critical properties of a 3D electrolyte solution [10]. In particular, the coexistence curve obtained on its basis was found to be in an excellent agreement with the recent Monte Carlo (MC) simulations [11]. The application of this theory to the 2D plasma has lead to a stark disagreement with the work of Minnhagen. Whereas Minnhagen has found that the KT line terminates in a *critical end point*, the DHBj theory predicts that it will terminate in a *tricritical* point, after which the vapor *insulating* phase will coexist with a liquid *conducting* phase [9] (see the inset in Fig. 2).

Since the DHBj theory is intrinsically of mean-field type, one might argue that the fluctuations, such as a variation in dipolar sizes, might modify the phase diagram. This, however, is not very likely. It is well known that a properly constructed mean-field theory almost always retains the topology of the phase diagram upon inclusion of fluctuations. One of the few exceptions is when the volume of fluctuations is extremely large, such as in the case of transition between disordered and lamellar phases [12] in magnets or amphiphilic systems. This, however, is not the case here. Furthermore, the scaling of dipolar sizes can be included in a

straightforward way into the DHBj theory, leaving the topology of the phase diagram unchanged [13]. The metal-insulator line then becomes in exact agreement with the KT theory, in particular, giving the correct critical exponent $\nu=1/2$ for the divergence of the screening length upon the approach to the transition. The tricritical point persists, while the first-order coexistence curve remains extremely narrow in the vicinity of the tricritical point.

Comparing the predictions of the Minnhagen and DHBj theories to MC simulations, we find that neither one is in quantitative agreement with MC theory, which finds that the first-order transition appears at a temperature that is significantly lower and a density that is significantly higher than the prediction of either one of the above theories [7]. Nevertheless, the topology of the phase diagram observed on the basis of DHBj theory is the same as that found in MC theory. Furthermore, the location of the tricritical point obtained in the MC theory corresponds closely to the region of the phase diagram where the narrow DHBj coexistence is found to swell significantly [13].

The current impasse lead us to reexamine some of the foundations on which our understanding of the CG is based. Most of the rigorous theorems concerning the nature of interactions inside the neutral plasma are based on the isomorphism between the CG and the sine-Gordon field theory [2(b),14]. The mapping is exact only for the *point* Coulomb gas in the grand canonical ensemble. The short-range repulsion is included *post facto* by introducing a suitable cutoff on all momentum space integrals. To what extent this procedure is valid is far from clear. Nevertheless, if the sine-Gordon field theory is renormalized, one obtains *exactly* the KT flow equations in terms of renormalized fugacity and temperature [15]. This equation, however, remains valid only for low density (small fugacity).

The attempts to construct a sine-Gordon-based mean-field theory go back to the work of Saito [16], who observed that at the mean-field level the sine-Gordon Hamiltonian accounts for the metal-insulator transition. In particular Saito was able to show that at low fugacity the Debye screening length diverged as $\xi_D=e^{c/t^\nu}$, where $t=(T-T_c)/T_c$ and $\nu=1$. This should be compared with an equivalent expression obtained by Kosterlitz and Thouless, but with $\nu=1/2$. Using field-theoretic methodology, Zhang *et al.* [17] extended the mean-field type of calculations of Saito and found that above a critical fugacity the screening length has a discontinuous jump from a finite to an infinite value. Zhang *et al.* then interpreted this point as a tricritical point terminating the continuous line of the metal-insulator transition.

To compare the results of the sine-Gordon-based theory to the MC theory, one must be able to come up with a transformation from the fugacity, which is a natural variable in the field-theoretic description, to the density, which is what the MC simulations measure. In the following we present a simple variational mean-field theory that accomplishes just that. It is in the process of transforming the phase diagram from the temperature fugacity to temperature-density plane that the surprising results were found.

Our starting point is the grand canonical partition function for point particles of charge $\pm q$,

$$\mathcal{Z}=\sum_{N_+=0}^{\infty}\sum_{N_-=0}^{\infty}\frac{z_+^{N_+}}{N_+!}\frac{z_-^{N_-}}{N_-!}\mathcal{Q}(N_+,N_-), \tag{1}$$

where

$$\mathcal{Q}(N_+,N_-)=\int\prod_{i=1}^{N}\frac{d^2r_i}{\lambda^2}\exp\left[-\frac{\beta}{2D}\sum_{i\neq j}^{N}q_iq_jU(r_{ij})\right]. \tag{2}$$

Here $N=N_++N_-$ is the total number of particles immersed in a homogeneous medium of dielectric constant D and $\lambda=(h^2/2\pi mk_BT)^{1/2}$ is the thermal wavelength; the two-dimensional interaction term is $U(r_{ij})=\ln r_{ij}/a$, where a is an arbitrary scale and r_{ij} is the distance between particles i and j. The fugacity is related to the chemical potential through $z_\pm=e^{\beta\mu_\pm}$ and, along with the temperature ($\beta=1/k_BT$), determines all thermodynamic characteristics of the 2D CG.

To explore the thermodynamic properties of the above partition function it is convenient to map it onto the sine-Gordon field theory [2(b),14]. Thus the partition function $\mathcal{Z}$ can be expressed as a functional integral over a real field ϕ,

$$e^{-\beta\mathcal{G}}\equiv\mathcal{Z}=\frac{\int\mathcal{D}\phi e^{-H_{SG}}}{\int\mathcal{D}\phi\,\exp\left(-\int d^2r\frac{1}{2}(\nabla\phi)^2\right)}, \tag{3}$$

where

$$H_{SG}=\int d^2r\left[\frac{1}{2}(\nabla\phi)^2-\frac{2\overline{z}}{a^2}\cos\left(\sqrt{\frac{2\pi\beta}{D}}q\phi\right)\right] \tag{4}$$

is the effective Hamiltonian for a neutral Coulomb gas and $\overline{z}=z(a/\lambda)^2e^{\beta q^2U(0)/2D}$ is the fugacity renormalized by a self-energy term. It is interesting to note that the saddle point of the sine-Gordon field theory corresponds to the familiar Poisson-Boltzmann equation. In this paper, however, we will not use this analogy but instead construct a variational bound for the free energy. To this end we shall rely on the Gibbs-Bogoliubov-Feynman inequality exploring the convexity of free energy $\mathcal{G}\leqslant G=\mathcal{G}_0+\langle H-H_0\rangle_0$, where $\mathcal{G}_0$ is the free energy associated with an arbitrary trial Hamiltonian H_0. The angular brackets indicate averaging over H_0. It is particularly convenient to choose as a trial Hamiltonian one having a Gaussian form

$$H_0=\int d^2r\left[\frac{1}{2}(\nabla\phi)^2+\frac{m^2}{2a^2}\phi^2\right]. \tag{5}$$

In this case the free energy $\mathcal{G}_0$ and the average $\langle H-H_0\rangle_0$ are easily calculated and we find

$$\frac{\beta G}{V}=\frac{1}{8\pi a^2}\ln(1+m^2)-\frac{2\overline{z}}{a^2}\left(1+\frac{1}{m^2}\right)^{-1/4T^*}. \tag{6}$$

To perform the momentum space integrals the ultraviolet cutoff ($\Lambda=1/a$), corresponding to the effect of the hard core, was introduced [18]. The optimum upper bound is found by

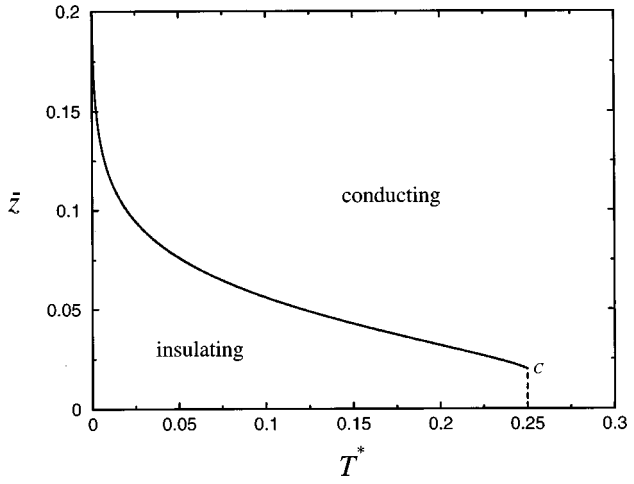

FIG. 1. Phase diagram of the 2D CG in the $\overline{z}$-T^* plane. The solid line corresponds to the first-order transition, while the dashed line is the KT infinite-order metal-insulator transition. The tricritical point C is located at $\overline{z}=1/16\pi$ and $T^*=1/4$.

minimizing the free energy G over all possible m^2. We find that the value of $m^2=m_0^2$ that leads to the optimal approximation to the real free energy $\mathcal{G}$ satisfies

$$m_0^2=\frac{4\pi\overline{z}}{T^*}\left(\frac{m_0^2}{1+m_0^2}\right)^{1/4T^*}, \tag{7}$$

where $T^*=k_BTD/q^2$ is the reduced temperature. The parameter m_0^2 is inversely proportional to the Debye screening length ξ_D inside the electrolyte solution since it can easily be shown that the effective potential between two test particles separated by a distance r is $V_{\text{eff}}(r)\sim\langle\phi(0)\phi(r)\rangle_0$. Furthermore, if $m_0^2=0(\xi_D=\infty)$ signifies that there is no screening which means that all the ions have paired up forming dipolar pairs. If this is the case the presence of an insulating phase is ensured.

Indeed, from Eqs. (6) and (7) we find that the free energy possesses two minima, one of which is for $m_0^2=0$ and the other for $m_0^2\neq 0$. The first-order phase transition occurs when the free energies corresponding to the two local minima become equal. The phase diagram in the $\overline{z}$-T plane for the 2D Coulomb gas is presented in Fig. 1. It is essentially divided into two regions, each of which is characterized by a specific value of m_0^2: a conducting phase with a finite value of m_0^2 and an insulating phase with $m_0^2=0$. Separating these two phases, there is a first-order transition line that ends at a tricritical point C ($T_c^*=1/4$ and $\overline{z}_c=1/16\pi$). Below $\overline{z}_c$ and close to T_c^*, Eq. (7) can be approximated by

$$\xi_D=\frac{a}{m_0}\sim a\,\exp\left[\frac{1}{2t^{\nu}}\left(\ln\frac{T^*}{4\pi\overline{z}}\right)\right], \tag{8}$$

where $t\equiv 1-1/4T^*$ and $\nu=1$. When $T^*\to 1/4$ Eq. (8) defines a line of a critical point ($\xi_D=\infty$) that separates the conducting and insulating phases. This corresponds to the usual KT line of metal-insulator transitions. Contrary to the appearance, the first-order line and the critical line join smoothly at the tricritical point, with the tangency of the first-order line ensured by the divergence of $d\overline{z}/dT^*\simeq\ln|t|/4\pi$ when $t\to 0^-$.

As emphasized earlier, in order to compare the results of our variational treatment with those of MC simulations it is essential to perform a transformation from the fugacity-temperature plane to the density-temperature domain. To this end we note that $\mathcal{G}$ is related to pressure and volume through $\mathcal{G}=-PV$, while the density is $\rho=\overline{z}\partial(\beta P)/\partial\overline{z}$. The transformation is then easily achieved and we find the coexistence curve presented in Fig. 2. This curve is topologically identical to that obtained on the basis of pure linearized Debye-Hückel (DH) theory [9]. In particular we find that the high-density conducting phase coexists with the *zero-density* insulating phase. Namely, although the SG theory accounts for the metal-insulator transition, it cannot give a proper account of the low-density phase. Instead of producing dipoles the oppositely charged ions self-annihilate on contact.

From our treatment it is not clear whether this is a true property of the SG model or an artifact of the mean-field

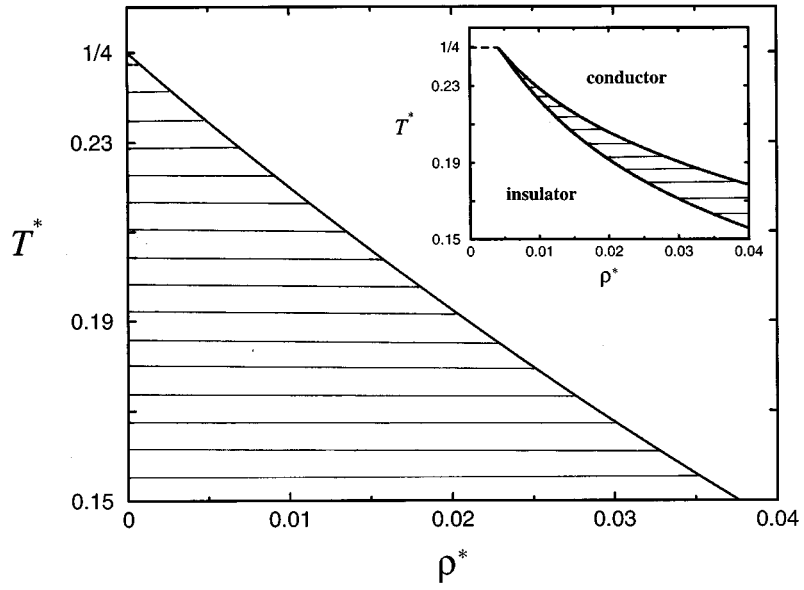

FIG. 2. Coexistence curves for the 2D SG field theory. The tricritical point is at $T_c^*=1/4$ and $\rho_c^*=0$. The inset represents the phase diagram for $d=2$, according to DHBj theory [9], with the KT line ending at a tricritical point (dashed line) localized at $\rho_c^*\simeq 0.003\,954$ and $T_c^*=1/4$.

treatment, or maybe a result of the artificial way in which the hard core was introduced into the model. Whichever the case, it is interesting to compare this result with the recently developed DHBj theory [9], which, although also is of a mean-field type, does predict a finite density for the insulating phase. The coexistence curve for the DHBj theory is presented in the inset of Fig. 2.

The inability of the SG model (at least at the mean-field level) to give a proper account of dipoles is also confirmed in $d=3$. In this case a first-order phase transition at low temperatures between a low-density (vapor) and a high-density (liquid) phase is found, as expected. Indeed, based on the DHBj theoretical [10] studies and simulations [11], the critical point is localized at $T_c^* \simeq 0.057$ and $\rho_c^* \simeq 0.025$. The resulting coexistence curve that emerges from our variational treatment predicts a critical point at $T_c^* \simeq 0.0565$ and $\rho_c^* \simeq 0.00135$. While the critical temperature T_c^* is in agreement with previous results, the critical density ρ_c^* is too small. This value should once again be compared with the pure linearized DH theory, which does not account for the existence of dipoles; in that case it was found that $T_c^* = 0.0625$ and $\rho_c^* = 1/64\pi \simeq 0.005$. The underestimate of the critical density clearly indicates that at least at the mean-field level the SG theory, just like the pure DH theory, does not give a proper account of nonlinear effects such as the formation of dipoles.

How can these nonlinearities be included is far from obvious. Why should the SG theory predict a metal-insulator transition in the temperature-fugacity plane only to later map the whole insulating phase onto the zero density? What is the proper class of diagrams that would have to be summed to produce a finite density for the insulating phase? These questions require serious attention if we wish to have a complete theory. The inability of the sine-Gordon model, at least at the mean-field level, to give a proper account of the low-density phase might also be responsible for the distinct predictions between the Minnhagen and DHBj theories. In particular, it can be shown that the variational method that we have used corresponds to the summation to all orders of a certain class of diagrams. In the case of the standard scalar field theory this is the familiar Hartree-Fock approximation [19]. This class of diagrams is obviously insufficient if we are to believe that the SG theory can give a realistic account for the phase structure of the Coulomb gas. In his approach Minnhagen also relied on the SG theory to calculate the charge-charge correlation function. To this end he summed another set of diagrams. If that set was incomplete it could lead to some undesirable effects such as, for example, the wrong topology of the phase diagram. At the moment, however, this is only a speculation and a renewed theoretical effort is needed to study the sine-Gordon field theory now that it is evident that this model, besides the KT transition, also contains a first-order discontinuity.

[1] J. L. Cardy, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1987), Vol. 11.

[2] Good reviews are available in (a) B. Nienhuis, in *Phase Transitions and Critical Phenomena* (Ref. [1]); (b) P. Minnhagen, Rev. Mod. Phys. **59**, 1001 (1987).

[3] D. R. Nelson and B. I. Halperin, Phys. Rev. B **21**, 5312 (1980).

[4] S. T. Chui and J. D. Weeks, Phys. Rev. B **14**, 4978 (1976); Y. Levin, *ibid.* **43**, 10 876 (1991).

[5] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); A. P. Young, *ibid.* **11**, L453 (1978); Phys. Rev. B **19**, 1855 (1979).

[6] A. N. Berker and D. R. Nelson, Phys. Rev. B **19**, 2488 (1979).

[7] J. M. Caillol and D. Levesque, Phys. Rev. B **33**, 499 (1986); J.-R. Lee and S. Teitel, Phys. Rev. Lett. **64**, 1483 (1990); Phys. Rev. B **46**, 3247 (1992); J. P. Valleau, J. Chem. Phys. **95**, 584 (1991); J. M. Caillol, *ibid.* **100**, 2161 (1994); G. Orkoulas and A. Z. Panagiotopoulos, *ibid.* **104**, 7205 (1996); P. Gupta and S. Teitel, Report No. cond-mat/9609031.

[8] P. Minnhagen, Phys. Rev. Lett. **54**, 2351 (1985); P. Minnhagen and M. Wallin, Phys. Rev. B **36**, 5620 (1987); **40**, 5109 (1989); J. M. Thijssen and H. J. F. Knops, *ibid.* **38**, 9080 (1988).

[9] Y. Levin, X. Li,and M. E. Fisher, Phys. Rev. Lett. **73**, 2716 (1994).

[10] M. E. Fisher and Y. Levin, Phys. Rev. Lett. **71**, 3826 (1993); M. E. Fisher, J. Stat. Phys. **75**, 1 (1993); G. Stell, *ibid.* **78**, 197 (1994); Y. Levin and M. E. Fisher, Physica A **225**, 164 (1996); B. Guillot and Y. Guissani, Mol. Phys. **87**, 37 (1996).

[11] A. Z. Panagiotopoulos, Fluid Phase Equilibria **76**, 97 (1992); G. Orkoulas and A. Z. Panagiotopoulos, J. Chem. Phys. **101**, 1452 (1994).

[12] S. A. Brazovskii, Zh. Éksp. Teor. Fiz. **68**, 175 (1975) [Sov. Phys. JETP **41**, 85 (1975)].

[13] X.-J. Li, Y. Levin, and M. E. Fisher (unpublished).

[14] S. F. Edwards, Philos. Mag. **4**, 1171 (1959); see also D. Brydges, Commun. Math. Phys. **58**, 313 (1978); D. Brydges and P. Federbush, *ibid.* **73**, 197 (1980).

[15] H. J. F. Knops and L. W. J. den Ouden, Physica A **103**, 597 (1980); D. J. Amit, Y. Y. Goldschmidt,and G. Grinstein, J. Phys. A **13**, 585 (1980); T. Ohta and D. Jasnow, Phys. Rev. B **20**, 139 (1979); T. Ohta, Prog. Theor. Phys. **60**, 968 (1978).

[16] Y. Saito, Prog. Theor. Phys. **62**, 927 (1979).

[17] G. M. Zhang, H. Chen,and X. Wu, Phys. Rev. B **48**, 12 304 (1993); B. W. Xu and Y. M. Zhang, *ibid.* **50**, 18 651 (1994).

[18] There is no unique relationship between the cutoff Λ and the particle size. We adopt an isotropic sharp cutoff $\Lambda = 1/a$. It should be remembered that the mapping of the Coulomb gas onto the sine-Gordon field theory is only rigorously valid for point particles.

[19] G. Parisi, *Statistical Field Theory* (Addison-Wesley, Reading, MA, 1988).