Intermediate periodic "saddle-splay" nematic phase in the vicinity of a nematic–smectic-A transition

G. Barbero[1,*] and V. M. Pergamenshchik[2,†]
[1]*Departimento di Fizica, Politecnico di Torino, Duca degli Abruzzi, Torino, Italy*
[2]*Institute of Physics, prospect Nauki 46, 03650 Kiev-28, Ukraine*
(Received 7 May 2002; published 19 November 2002)

We consider possible spontaneous modulations of the nematic director induced by the elastic saddle-splay K_{24} term when the value of the elastic constant K_{24} does not satisfy the Ericksen stability condition for the homogeneous ground state. According to the standard formula expressing K_{24} in terms of the twist elastic constant K_{22}, this can be expected close to the nematic–smectic-A transition where K_{22} becomes very large. It is predicted that in a planar nematic layer (or, more generally, if the surface director alignment is sufficiently close to a planar one), a modulated phase with observable long wavelength period can occur in samples considerably thicker than the anchoring extrapolation length. The modulated nematic phase is expected to persist into the smectic phase so that its temperature of the transition to smectic phase has to be lower than that for the homogeneous nematic liquid crystal. Low amplitude short wavelength modulations are predicted for any thickness if the surface director is sufficiently far from a pure homeotropic alignment. At the expense of this mode the temperature of a nematic–smectic-A transition in a planar cell with isotropic surfaces has to be lower than that for a homeotropic cell even if the periodic structure is not accessible for the direct observation.

DOI: 10.1103/PhysRevE.66.051706 PACS number(s): 61.30.Gd, 64.70.Md

I. INTRODUCTION

Different phases in the condensed matter physics are classified by the symmetry and specific structure of their ground states. The ground state structure of a specific phase is not only a formal sign thereof, but, which is most important, is an external manifestation of the principal intrinsic forces and their balance, hidden behind the phase appearance. Therefore, different phases represent different mechanisms of setting the order in an external-field-free condensed matter.

In particular, different phases of liquid crystals (LCs) represent different intrinsic mechanisms of setting LC order that can be described by a spatial distribution of correspondent order parameters in the ground state. One of the specific LC order parameters is the director **n** showing a macroscopic anisotropy axis resulting from averaging the individual orientations of the constituting molecules. Mechanisms of setting director distortions in the ground state play a principal role in the physics of LCs since the long range order related to the director makes some of the LC phases visualized by means of a polarizing microscope. For instance, a nematic phase is characterized by a homogeneous undistorted ground state while cholesteric phase is just a twisted nematic where the director spontaneously rotates about some single direction. The cholesteric order is set by a balance between the chiral force that tends to twist the director and is described by the chiral Lifshits energy term, and the nematic elastic force that resists twist deformations and is described by the positive definite K_{22} term in the nematic energy.

Generally speaking, the director distribution is always set by a balance of distortion-inducing terms that gain energy for finite deformations, and distortion-resisting terms that are minimum for an undistorted state. Therefore, in principle, a nonuniform director ground state is possible when the free energy (FE) functional contains sign indefinite terms capable of decreasing its value at the expense of finite distortions.

A nematic phase of liquid crystals is not an exclusion. The director deformation energy F_d is the sum

$$F_d=\frac{1}{2}\int dV\bigg\{\frac{1}{2}K_{11}(\nabla\cdot\mathbf{n})^2+\frac{1}{2}K_{22}(\mathbf{n}\cdot\nabla\times\mathbf{n})^2+\frac{1}{2}K_{33}(\mathbf{n}\times\nabla\times\mathbf{n})^2-K_{24}\nabla\cdot[\mathbf{n}(\nabla\cdot\mathbf{n})+\mathbf{n}\times\nabla\times\mathbf{n}\bigg]+K_{13}\nabla\cdot[\mathbf{n}(\nabla\cdot\mathbf{n})] \qquad (1)$$

of the terms quadratic in the differentiation operator ∂. Along with the three positive definite splay, twist, and bend terms which resist any deformations, F_d contains two sign indefinite terms, the so-called divergence K_{24} and K_{13} terms, which can be a source of spontaneous distortions. Therefore, the fundamental stability condition of the standard uniform nematic ground state must be derived from the FE functional. Note that the total director dependent FE functional F is a sum of the deformational part F_d and the surface anchoring energy that depends solely on the director orientation on the surface S of the nematic body.

Long ago, Ericksen considered stability of a uniform director ground state, disregarding the anchoring and K_{13} term, and found that the elastic constants K_{22} and K_{24} in a uniform nematic phase cannot be arbitrary. In terms of the dimensional quantities $k_{22}=K_{22}/K_{11}$ and $p_{||}=1-2K_{24}/K_{11}$ the correspondent restriction reduces to the two inequalities [1],

$$|p_{||}|<1, \qquad (2)$$

*Email address: barbero@polito.it
†Email address: pergam@victor.carrier.kiev.ua

 ©2002 The American Physical Society

$$-2k_{22}<1-p_{||}-2k_{22}<0. \tag{3}$$

The constant K_{24} enters these inequalities just in the combination $p_{||}\propto K_{11}-2K_{24}$, which is natural as $p_{||}$ represents the total contribution of the term $\nabla\cdot[\mathbf{n}(\nabla\cdot\mathbf{n})+\mathbf{n}\times\nabla\times\mathbf{n}]$ in F_d (the K_{24} term in Eq. (1) is not the only one of this form: the other contribution $\propto K_{11}$ is hidden in the sum of the positive definite terms, for details see, e.g., Ref. [2]; nevertheless, referring to effects always related to the total contribution of this form, we will use the symbol K_{24}). Therefore, the first inequality means that the energy gain due to the K_{24} term-induced deformations must be smaller than the energy cost due to the splay term. The second inequality will be of no concern in this paper, and we just note that it restricts a combined action of the K_{24} and K_{22} terms setting some lower bound to the twist ratio k_{22}.

The role of the K_{13} term and anchoring in stability of the homogeneous nematic ground state have been considered much later. Two qualitatively different spontaneous modes have been predicted. The first one has the form of a surface director distortion vanishing over a few molecular lengths from the surface [3]. Although this surface mode is predicted for any finite K_{13} and, practically, in any geometry of the director, it cannot be directly observed because its wavelength is of the molecular scale. The second mode can have the form of an observable long wavelength periodic distortion [2]. For instance, in the case of a layer with a planar anchoring at both surfaces, which will be considered in this paper, the homogeneous planar state is unstable if the thickness H is smaller than the critical value H_c given by

$$H_c=-2L_a(1-p_{||}^2+K_{13}/K_{11})>0, \tag{4}$$

where L_a is the anchoring extrapolation length [2]. In this particular geometry, formula (4) generalizes Erisksen inequality (2) to the case of a finite anchoring and K_{13}. Since inequality (2) is assumed to hold, $|K_{13}/K_{11}|\sim 1$, and L_a is of the order of a micrometer, this mode can be expected only in submicrometer thin films (see, e.g., Refs. [4,5]).

However, the situation when at least one of the Ericksen inequalities is violated has never been considered [6]. Presumably, one of the reasons has been that the value of K_{24} required for such a violation seemed to be unrealistic, and the situation when it could actually happen seemed to be difficult to find. Contrary to this, here we show that, according to the standard ideas of the physics of liquid crystals, such a situation should be rather common.

Indeed, on the one hand, the standard elastic approach predicts that the value of the constant K_{24} is given by the formula

$$K_{24}=\frac{K_{11}+K_{22}}{4}, \tag{5}$$

derived in Refs. [7,8] (also, see Ref. [9]). On the other hand, close to a nematic–smectic-A (N-SmA) transition the constants K_{22} and K_{33} grow very large [11,12]. Then Eq. (5) shows that while the second inequality (3) remains satisfied, the first inequality (2) is violated as the constant $p_{||}\approx -k_{22}/2$ is large negative, i.e.,

$$p_{||}\ll -1. \tag{6}$$

The inequality (6) implies that the homogeneous ground state of the nematic director can be spontaneously deformed in the proximity of a smectic phase. In this paper we will explore this possibility which is not so obvious since large K_{22} and K_{33} also imply a very strong resistance to the twist and bend deformations.

Since the K_{24} term identically vanishes if the director depends just on a single Cartesian coordinate, the K_{24} term-induced spontaneous deformations can be very complicated. Following arguments of Ref. [2] and experimental observation of the K_{24} term-induced stripe domains in thin nematic films [4,5], we will consider a planar nematic layer and seek the director ground state in the form of a function periodic in a single direction. The K_{13} term will be omitted in this analysis since close to a N-SmA transition $p_{||}^2$ is expected to be large, whereas the ratio K_{13}/K_{11} remains of the order of one and can be neglected, see Eq. (4).

II. K_{24} TERM-INDUCED PERIODIC INSTABILITY OF THE HOMOGENEOUS NEMATIC GROUND STATE AT A NEMATIC–SMECTIC-A TRANSITION

A. Unboundedness of the functional F_d from below and finding long wavelength spontaneous deformations

First of all we notice that the derivation of inequalities (2) and (3) indicates that when these are not satisfied, the functional F_d has no minimum. Indeed, Ericksen showed that the inequality opposite to Eq. (2) or (3) implies a negative elastic energy density f_d in each spatial point inside the nematic body. As f_d is quadratic in the modulation wave number q, the volume integral over f_d is proportional to q, i.e., $F_d\propto -qK$. Since the maximum value of the anchoring energy does not depend on q, the total energy F can be made unlimitedly large negative for an infinite wave number q. Thus, F has no lower boundary and cannot be directly minimized.

This problem is typical for incorporating the sign indefinite divergence terms into the elastic theory. Indeed, the K_{13} term is known to give rise to a similar problem: for any nonzero K_{13} the functional F_d is unbounded below [13,14]. Nevertheless, it was shown [15] that observable consequences of the presence of the K_{13} term in F_d can be derived solely in terms of this standard functional. Following the same arguments, we will show that observable modulations of the ground state, even when the Ericksen inequality (2) is violated and F_d has no lower bound, can be found from the Euler-Lagrange equations and boundary conditions associated solely with the functional F_d.

Unboundedness of the functional F_d for a nonzero K_{24} formally results in a mode with an infinite wave number q. However, F_d is just the first term in the energy expansion in the director derivatives, i.e.,

$$F_{tot}=F_d+R, \tag{7}$$

where F_{tot} is the total elastic energy and R is the higher-order elastic resistance. The functional F_d describes the standard elasticity linear in the director derivatives ∂n, whereas R describes nonlinear elasticity. For standard weak deformations, $l_M|\partial n|\ll 1$, where l_M is the molecular length, the nonlinear elastic energy R is completely negligible compared to F_d. As a consequence, the linear elastic force is much larger than the nonlinear elastic force, which justifies the very idea of the linear elasticity (note that actually we deal with an elastic torque—the generalized force that corresponds to angular variables). However, starting from some value ξ of $|\partial n|$, which is expected to be not much smaller than 1, the nonlinear term R as a functional of ∂n grows much faster than F_d so that, for derivatives just slightly larger than ξ, the nonlinear elastic resistance attains the level of the linear elastic forces [14]. In particular, this implies that for sufficiently large director derivatives, the nonlinear elastic force can balance the linear elastic force due to the presence of the K_{24} term. In terms of energy this means that R brings the lower boundary to the sum (7) so that F_{tot} attains the minimum value for some large but finite wave number Q_M which is of the order of $1/l_M$.

Further, the total energy F_{tot} is bounded below and determines the director distribution by means of the Euler-Lagrange equations $\hat{L}(F_d+R)=0$ and boundary conditions $\hat{B}(F_d+R)=0$ where $\hat{L}$ and $\hat{B}$ are the known linear operators. To find the solution of these equations in the whole range of distortion strengths, one needs to know the form of the functional R. However, if one is interested in mesoscopic wavelength modes, i.e., modes whose wavelength is much larger than l_M and that correspond to small director derivatives, the problem can be dramatically simplified. Indeed, the contributions of F_d and R in the above equations represent the correspondent generalized elastic forces. Since for small derivatives the linear force dominates, the second term in both of the above equations can be neglected compared to the first one, which reduces to $\hat{L}F_d=0$ and $\hat{B}F_d=0$. But these are exactly the Euler-Lagrange equation and boundary condition formally corresponding to the standard functional F_d for any value of K_{24}. Thus, for any K_{24}, mesoscopic wavelength modes can be found from the standard equations associated solely with F_d in spite of the fact that this functional may have no minimum. This is the essence of the procedure derived in the case of the K_{13} term in Ref. [15].

Now we address the modes with a wavelength of the molecular scale. To find such modes one needs to know the functional R. However, these modes would be just a formal solution and actually are of no interest. As mentioned in Introduction, such modes have a very short wavelength and cannot be directly observed. Indeed, if such a mode has a notable amplitude a then its energy, which can be estimated as $-Q_M K a^2 V$ where V is the system volume, is so low and hence the nematic phase is so stable that the transition to a smectic-A phase is not possible. Moreover, great deformations $\partial n\sim Q_M a$ involved would make impossible the very nematic phase. If, in contrast, its amplitude is very small then such a short wavelength mode cannot be observed. Moreover, a very short wavelength modulation with a small amplitude can be considerably suppressed by thermal fluctuations that are not incorporated in the elastic approach. The only effect of such modes is some drop $-F_M$ of the total energy which is difficult to estimate, but which is definitely negative. In principle, this energy drop due to directly unobservable modes can result in a shift of the temperature of the N-SmA transition which will be discussed in Sec. III.

We arrive at the conclusion that the search for the K_{24} term-induced spontaneous deformations is physically interesting if they have sufficiently long wavelengths or, equivalently, sufficiently small wave numbers. These modes can be found from the standard Euler-Lagrange equations and boundary conditions for the functional F_d. In the following section, this is done in the geometry of a planar nematic layer.

B. Spontaneous modulations in a planar layer

Consider a plane nematic layer of thickness H and assume that at its two surfaces the polar anchoring potentials f_a are the same and favor a planar alignment, whereas the azimuthal anchoring is negligible. The last assumption is adopted both for simplicity and because isotropic surfaces, that modern technology enables one to make even on solid substrates [16], provide the best conditions for the K_{24}, mechanism to come into play. We assume that the director is a periodic function of the coordinate y which is along the layer, and also depends on the coordinate z, which is normal to the layer and has the onset $z=0$ on its midplane. The director components on the upper surface $z=H/2$ and lower surface $z=-H/2$ will be indicated by subscripts 2 and 1, respectively. Then the director-dependent FE of the periodic structure per one period $2\pi/q$ can be written in the form [2]

$$F=\frac{K_{11}q}{4\pi}\int_{-H/2}^{H/2}dz\int_0^L dy[(\partial_i n_j)^2+(k_{22}-1)(\mathbf{n}\cdot\nabla\times\mathbf{n})^2+(k_{33}-1)(\mathbf{n}\times\nabla\times\mathbf{n})^2]+\frac{K_{11}q}{4\pi}\int_0^L dy[2p_{||}(n_{z,2}\partial_y n_{y,2}-n_{z,1}\partial_y n_{y,1})+f_a(n_{z2}^2)+f_a(n_{z1}^2)], \tag{8}$$

where $k_{33}=K_{33}/K_{11}$ is the reduced dimensionless bend constant. We will use the standard director parametrization $\mathbf{n}=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$, with the polar angle θ counted from the z axis, and the azimuthal angle ϕ counted from the y axis. In terms of these angles, the standard uniform planar director ground state is given by $\theta=\pi/2$, $\phi=0$.

As in Ref. [2] we assume small deviations from the planar state to be in the form

$$\theta-\pi/2=f(z)\sin(qy), \tag{9}$$

$$\phi=g(z)\cos(qy).$$

Expanding F in a functional Taylor series of these amplitudes up to quadric terms and performing the y integration, one obtains

$$F = F_2 + F_4. \tag{10}$$

The quadratic part is of the form

$$F_2\{f,g\} = \frac{K_{11}}{2H}\int_{-1/2}^{1/2} dz[f'^2 + \chi^2 g^2 + k_{22}(\chi^2 f^2 + g'^2 + 2\chi g' f)] + \frac{K_{11}}{2H}[p_{||}(f_2 g_2 - f_1 g_1) + h(f_1^2 + f_2^2)], \tag{11}$$

where we introduced following reduced quantities: the dimensionless coordinate $z = z/H$ (for which we leaved the same notation), reduced thickness $h = H/L_a$ (L_a is the anchoring extrapolation length), and dimensionless wave number $\chi = qH$. In the last bulk term in Eq. (11) we neglected 1 compared to a large k_{22}; the quadric energy F_4 will be given below.

Now we show that this functional actually describes two independent modes. To this end we separate symmetric and antisymmetric (with respect to the middle plane $z=0$) parts of the functions $f(z)$ and $g(z)$, i.e., $f = f^+ + f^-$, $g = g^+ + g^-$, where $f^+(z) = f^+(-z)$, $g^+(z) = g^+(-z)$, whereas $f^-(z) = -f^-(-z)$ and $g^-(z) = -g^-(-z)$. Then the functional (11) splits into the sum

$$F_2\{f,g\} = F_2\{f^-, g^+\} + F_2\{f^+, g^-\}, \tag{12}$$

where $F_2\{f^-, g^+\}$ is the energy of the mode f^-, g^+, and $F_2\{f^+, g^-\}$ is the energy of the mode f^+, g^-. As there is no interaction term in the leading order energy (11), the modes f^+, g^- and f^-, g^+ are linearly independent. In particular, the Euler-Lagrange equations for $F_2\{f,g\}$ split into independent Euler-Lagrange equations for the functionals $F_2\{f^-, g^+\}$ and $F_2\{f^+, g^-\}$.

The Euler-Lagrange equations for the functional $F_2\{f^-, g^+\}$ constitute the system

$$f^{-\prime\prime} - k_{22}\chi^2 f^- - \chi k_{22} g^{+\prime} = 0, \tag{13}$$

$$k_{22}(g^{+\prime} + \chi f^-)' - \chi^2 g^+ = 0, \tag{14}$$

which can be readily transformed to the form

$$g^+ = f^{-\prime\prime\prime}/\chi^3, \tag{15}$$

$$t f^{-\prime\prime\prime\prime} - \chi^2 f^{-\prime\prime} + k_{22}\chi^4 f^- = 0. \tag{16}$$

From these equations the mode f^-, g^+ is found to be

$$f^- = a f_a^- + b f_b^-, \tag{17}$$

$$g^+ = a g_a^+ + b g_b^+,$$

where a and b are the integration constants;

$$f_a^-(z) = \cos(\chi\alpha z)\sinh(\chi\beta z),$$

$$f_b^-(z) = \sin(\chi\alpha z)\cosh(\chi\beta z), \tag{18}$$

$$g_a^+(z) = (-\alpha^3 + 3\alpha\beta^2)\cos(\chi\alpha z)\cosh(\chi\beta z) + (\beta^3 - 3a^2\beta)\sin(\chi\alpha z)\sinh(\chi\beta z),$$

$$g_b^+(z) = (\alpha^3 - 3\alpha\beta^2)\sin(\chi\alpha z)\sinh(\chi\beta z) + (\beta^3 - 3a^2\beta)\cos(\chi\alpha z)\cosh(\chi\beta z);$$

χ is an arbitrary positive number, and

$$\alpha = \sqrt{\frac{2k_{22} - 1}{4k_{22}}}, \tag{19}$$

$$\beta = \sqrt{\frac{2k_{22} + 1}{4k_{22}}}.$$

The Euler-Lagrange equations and their reduced form (16), (15) for the mode f^+, g^- can be obtained by replacing f^-, g^+ with f^+, g^-. The mode f^-, g^+ obtains in the form

$$f^+ = c f_c^+ + d f_d^+, \tag{20}$$

$$g^- = c g_c^- + d g_d^-,$$

where c and d are arbitrary constants, and

$$f_c^+(z) = \cos(\chi\alpha z)\cosh(\chi\beta z),$$

$$f_d^+(z) = \sin(\chi\alpha z)\sinh(\chi\beta z), \tag{21}$$

$$g_c^-(z) = (\alpha^3 - 3\alpha\beta^2)\sin(\chi\alpha z)\cosh(\chi\beta z) + (\beta^3 - 3a^2\beta)\cos(\chi\alpha z)\sinh(\chi\beta z),$$

$$g_d^-(z) = (\beta^3 - 3a^2\beta)\sin(\chi\alpha z)\cosh(\chi\beta z) + (-\alpha^3 + 3\alpha\beta^2)\cos(\chi\alpha z)\sinh(\chi\beta z).$$

The above solution of the linear problem allows one to find critical points where the system becomes unstable with respect to nonzero amplitudes a, b, c, and d. This point is determined by those parameters for which the quadratic functional F_2 vanishes for finite values thereof, and can be found from the boundary conditions for F_2.

The boundary conditions to the Euler-Lagrange equations (13) and (14) can be reduced to the form

$$A_a a + A_b b = 0, \tag{22}$$

$$B_a a + B_b b = 0, \tag{23}$$

where

$$A_a = (f_a^{-\prime} + p_{||}\chi g_a^+ + d f_a^-)_2,$$

$$A_b = (f_b^{-\prime} + p_{||}\chi g_b^+ + d f_b^+)_2, \tag{24}$$

$$B_a = [k_{22}(g_a^{+\prime} + \chi f_a^-) + p_{||}\chi f_a^-]_2,$$

$$B_b = [k_{22}(g_b^{+\prime} + \chi f_b^-) + p_{||}\chi f_b^-]_2;$$

the subscript 2 indicates that the correspondent function is calculated for $z=1/2$. The critical condition for the mode f^-,g^+ (17) to occur is

$$D_{ab}=A_aB_b-A_bB_a=0, \tag{25}$$

where D_{ab} is the determinant of the system (22), (23). This equation can be treated analytically only in the limit of a very small χ. In this limit one has the following asymptotic behavior:

$$f_2^- \simeq \frac{\chi}{2\sqrt{2}}\left[a+b+\frac{1}{4k_{22}}(b-a)\right],$$

$$f_2^{-\prime} \simeq \frac{\chi}{\sqrt{2}}\left[a+b+\frac{1}{4k_{22}}(b-a)\right], \tag{26}$$

$$g_2^+ \simeq \frac{1}{\sqrt{2}}\left[a-b+\frac{3}{4k_{22}}(a+b)\right],$$

$$g_2^{+\prime} \simeq -\frac{\chi^2}{2\sqrt{2}}\left[a+b+\frac{5}{4k_{22}}(b-a)\right].$$

Substituting this into Eq. (25) up to terms $O(1/k_{22})$ gives

$$h+2(1-p_{||}^2)=0, \tag{27}$$

or, equivalently,

$$h_{lw}=-2(1-p_{||}^2), \tag{28}$$

where h_{lw} is the critical value of thickness of the instability with a very small wave number, which reproduces Eq. (4) for $K_{13}=0$. We will see that, as in Ref. [2], h_{lw} is actually the upper critical thickness below which the modulation with $\chi\to 0$ appears.

The critical condition for the mode f^+,g^- (20) with the amplitudes c and d can be found similarly. Introducing A_c, A_d, B_c, and B_d by replacing a by c, b by d, f^- by f^+, and g^+ by g^- in the definition (24), this condition can be reduced to the form

$$D_{cd}=A_cB_d-A_dB_c=0, \tag{29}$$

where D_{cd} is the determinant of the linear system for the amplitudes c and d, obtained from the boundary conditions for the mode f^+,g^- (21). However, in contrast to Eq. (25), equality (29) is not possible for small χ, which can be easily seen from the asymptotic value of the functional $F_2\{f^+,g^-\}$. Indeed, one has $f_2^+\simeq c$, $f_2^{+\prime}\sim g_2^-\sim\chi$, and hence the positive term $f_2^+f_2^{+\prime}\sim\chi$, whereas the negative K_{24} term $p_{||}\chi g_2^- f_2^+ \sim\chi^2$, which is much smaller for $\chi\to 0$. Therefore, the mode f^+,g^- cannot appear for very small wave numbers as the energy of such a modulation would have been positive. For $\chi\sim 1$ the critical condition can only be analyzed numerically.

The pure linear problem we have dealt with so far only allows for finding the critical points that correspond to $D_{ab}=0$ or $D_{cd}=0$. We, however, are interested in the behavior of the system for the parameters h, $p_{||}$, k_{22}, and k_{33} below this point, where the determinants are nonzero and the instability is well developed. This behavior and, in particular, small but finite amplitudes of the modes can be found only with regards for the quadric term F_4 in the expansion of F (8). In F_4 we can retain only the bend and twist terms as they have very large coefficients k_{33} and k_{22}, which gives

$$F_4=F_{4bend}+F_{4twist},$$

$$F_{4bend}=\frac{k_{33}}{8H}\int_{-1/2}^{1/2}dz[3(\chi^2g^2+f'^2)f^2+2\chi f^2f'g+(fg'-\chi g^2)^2], \tag{30}$$

$$F_{4twist}=-\frac{k_{22}}{8H}\int_{-1/2}^{1/2}dz[(\chi f-g')(f^2+3g^2)(\chi f-g')+(g'f^2+2ff'g)+3g^2g'+\chi f^3+2g(ff'+3gg')+2\chi f(f^2-g^2)].$$

The procedure of finding small modulation amplitudes below the transition homogeneous state-stripe state was developed in Ref. [17] (to be specific, we will first consider the amplitudes a and b). It assumes that the distance ϵ from the critical point, which is some *a priori* unknown combination of the physical parameters, is sufficiently small as the amplitudes a and b vanish and grow along with this distance. This small parameter ϵ is proportional to the determinant D_{ab} of the system of linearized boundary conditions, and thus the amplitudes are small if the system is sufficiently close to the critical points where D_{ab} vanishes. Describing small amplitudes close to the instability onset involves two general steps: first, finding ϵ as a function of the parameters of the problem and determining intervals thereof where the amplitudes are finite, and, second, calculating values of these amplitudes.

The first step does not involve quadric term (30) and deals solely with the linear equations considered above. Both linear homogeneous equations (13) and (14) cannot be simultaneously satisfied for $D_{ab}\neq 0$. As a result, one can choose one of the two linear boundary conditions (13) and (14) and solve it to find one amplitude as a function of the other, let us say $a=a(b)$. Then this function of the amplitude b is substituted to $F_2\{f^-,g^+\}$ to give $F_2\propto D_{ab}b^2$. The sign of this expression determines the instability intervals: the amplitude is finite when the sign is negative, and zero when it is positive; while the square root of the negative of the coefficient before b^2 plays the role of the small parameter ϵ, as in the former case the amplitude is obviously proportional to this quantity [18].

The second step allows one to find the amplitude value in the intervals of the parameters found in the step one and requires extensive calculations. However, calculating the precise amplitude values would not add any principal information that could facilitate an experimental study of the predicted modulated phase since knowing the magnitude order is sufficient for this purpose [19]. Moreover, measuring the values of k_{22} and k_{33} at a N-SmA transition is a problem in itself, but these are necessary for the precise calculation of the modulation amplitude. In this situation, it is the most important to have a correct order of this quantity, which can be obtained quite easily. To this end the total contribution of the term of the fourth order in the modulation amplitude can be estimated by substituting the solution $a(b)$, obtained from the linear equations, into the quadric term (30) [20]. After that the sum (10) is minimized with respect to b. Finally, smallness of the amplitudes should be verified *a posteriori* which gives the parameter range where the perturbation method described is applicable. Below we follow this procedure.

We choose the first equation (22) and solve it with respect to a, which gives

$$a=-\frac{A_b}{A_a}b. \tag{31}$$

Now the values of F_2 and F_4 have to be calculated for the solution given by Eqs. (17), (18), and (31). To calculate the energy $F_2\{f^-,g^+\}$ of the equilibrium mode f^-,g^+ we make use of the Euler-Lagrange equations and Eq. (22), which is similar to the derivation of virial theorems. Multiplying Eq. (13) by f and Eq. (14) by g, integrating across the layer, adding the results, and then using Eq. (31), one obtains $F_2\{f^-,g^+\}=\epsilon_b b^2$ (the coefficient ϵ_b is given below). A similar expression is obtained for the mode f^+,g^- with the amplitudes c and d, i.e., $F_2\{f^+,g^-\}=\epsilon_d d^2$. As a result, the total quadratic term is the sum

$$F_2=\epsilon_b b^2+\epsilon_d d^2, \tag{32}$$

where ϵ_b and ϵ_d, which play the role of the distance from the critical point, are of the form

$$\epsilon_b=2(g^+_{b2}A_a-g^+_{a2}A_b)\frac{D_{ab}}{A_a^2}, \tag{33}$$

$$\epsilon_d=2(g^+_{d2}A_c-g^+_{c2}A_d)\frac{D_{cd}}{A_c^2}.$$

In accordance with the general expectation, these quantities are proportional to the determinants D_{ab} and D_{cd}, respectively, which justifies considering them as a distance from the critical point where $D_{ab}=0$ or $D_{cd}=0$. It is clear that ϵ, which vanishes in the critical point, is necessarily negative in the modulated phase. A numerical analysis shows that ϵ_b as a function of the wave number χ can be negative for both small χ and large χ, whereas the function ϵ_d is always positive for small χ. For this reason and for brevity, we will consider the amplitudes b and $a(b)$ of the mode f^-,g^+ alone.

Following the method outlined above we now calculate the quadric term (30) as a function of the single amplitude b. Obviously, this function $F_4(b)$ has the form

$$F_4\{a(b),b\}=\tilde{F}_{4,b}b^4,$$

where $\tilde{F}_{4,b}$ is a result of substituting Eqs. (17), (18), and (31) in the functional (30), omitting the common factor b^4 in thus obtained expression, and performing the z integration (in our case, it can be done just numerically). Then up to the fourth order in the amplitude, one has

$$F\{f^-,g^+\}=\epsilon_b b^2+\tilde{F}_{4,b}b^4. \tag{34}$$

Minimization of this expression with respect to the amplitude b gives

$$b_m=\begin{cases}\sqrt{-\dfrac{\epsilon_b}{2\tilde{F}_{4,b}}}, & \epsilon_b\leqslant 0\\ 0, & \epsilon_b>0,\end{cases} \tag{35}$$

$$F_m=F(b_m)=\begin{cases}-\dfrac{\epsilon_b^2}{4\tilde{F}_{4,b}}, & \epsilon_b\leqslant 0\\ 0, & \epsilon_b>0.\end{cases} \tag{36}$$

These are the desired expressions, respectively, for the modulation amplitude and energy which can now be studied as functions of the deformation source $p_{||}$, volume to surface ratio (reduced thickness) h, the bend and twist elastic resistance k_{33} and k_{22}, and the wave number χ. Similar formulas determine the amplitudes d_m and $c(d_m)$ and the energy $F(d_m)$ of the mode f^+,g^-. In the following section, we describe the modulations represented by these formulas.

III. THE MODULATED SADDLE-SPLAY NEMATIC PHASE: APPEARANCE AND OTHER OBSERVABLE EFFECTS

A. Modulation amplitude and spectrum

Numerical calculations by formulas (35), (36), and (33) show that, as expected, the quantities $\tilde{F}_{4,b}$ and $\tilde{F}_{4,d}$ are always positive so that the negativity of ϵ_b and ϵ_d is necessary for the correspondent mode to occur. We fixed values of all the parameters but χ and calculated the amplitudes b and d as functions of χ (spectrum). In line with the analytical analysis of the small χ limit, the spectrum $b(\chi)$ may have a long wavelength branch that exists for $h<h_{lw}$ and starts at $\chi=0$, and a short wavelength branch that appears for any thickness provided $p_{||}<-1$. At the same time, the spectrum $d(\chi)$ has only a short wavelength branch (we emphasize that the short wavelength branch contains a wide range of χ corresponding to modulations with small director derivatives and mesoscopic wavelength; this nomenclature appears here just to label the two spectral branches).

We will describe the spectrum in terms of the supercriticality of the main parameters: $\Delta p_{||}=p_{||}+1$ which shows the distance from the point where the Ericksen inequality (2) is violated (becomes the equality), and $\Delta h=(h-h_{lw})/h_{lw}$ which shows the distance from the critical point of the long wavelength instability. For $\Delta p_{||}>0$, a nematic phase remains homogeneous for any Δh (if K_{13} is neglected [2]). For $\Delta h>0$ and negative $\Delta p_{||}$ somewhat smaller than 0, a long wavelength branch is absent, and both amplitudes b and d

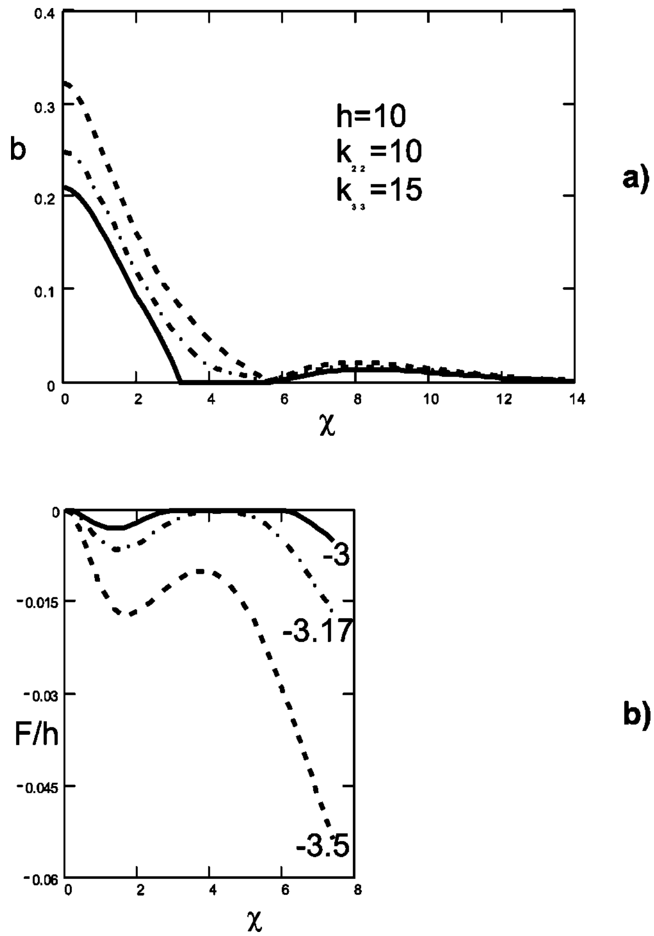

FIG. 1. (a),(b) The amplitude b (in radians) (a) and energy density F/h (in units K_{11}/h^2) (b) as functions of the dimensionless wave number χ for the thickness fixed at $h=10$ and three different values of $p_{||}$: $p_{||}=-3$ (solid), $p_{||}=-3.17$ (dash-dotted), $p_{||}=-3.5$ (dashed). The reduced twist and bend constants are modestly large: $k_{22}=10$, $k_{33}=15$. This might be relevant to a nematic LC relatively far from a nematic–smectic-A transition.

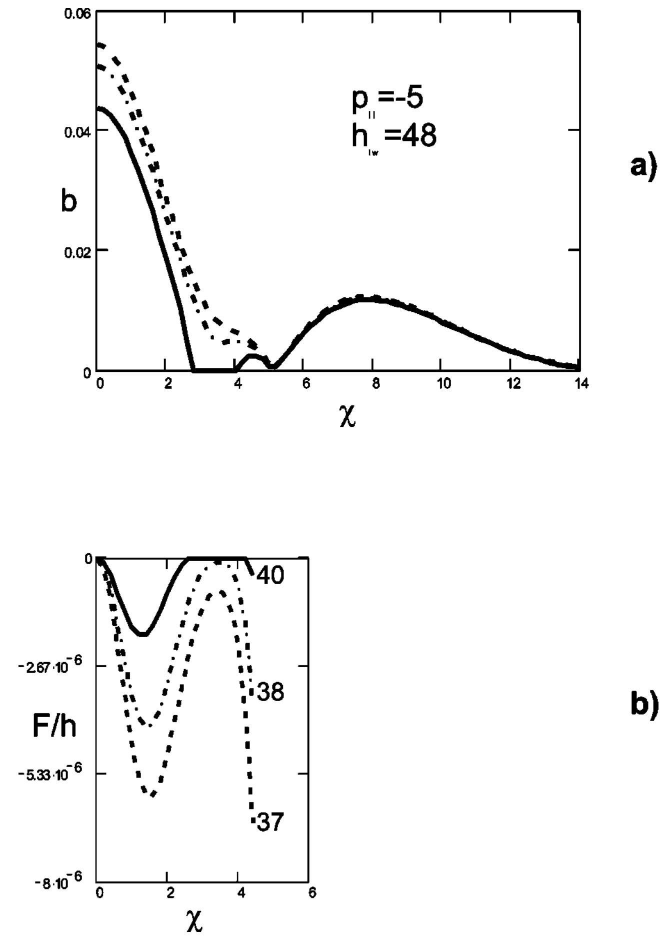

FIG. 2. (a),(b) The amplitude b (in radians) (a) and energy density F/h (in units K_{11}/h^2) (b) as functions of the dimensionless wave number χ for $p_{||}$ fixed at $p_{||}=-5$ (which corresponds to $h_{lw}=48$) and three different values of h: $h=40$ (solid), $h=38$ (dashed), $h=37$ (dotted). The reduced twist and bend constants are $k_{22}=70$, $k_{33}=90$.

have a short wavelength branch for large $\chi > \chi_{sw} \sim 10$, this χ_{sw} being always smaller for b than for d. For a very large χ, both b and d exponentially decrease which can be seen from the formula (35) and the presence of the hyperbolic functions in $f(z)$ and $g(z)$: $b \sim d \sim \exp(-2^{-3/2}\chi)$ when $\chi \to -\infty$. At the same time, the energy of both the modes monotonically drops toward large negative numbers as was predicted in Sec. II A. Obviously, if $\Delta p_{||}<0$, behavior of the energy and amplitude for very large χ has exactly the same character also for $\Delta h<0$, which is illustrated in Figs. 1–4. Formally speaking, the short wavelength spectrum continues to infinity, but, of course, as was discussed in Sec. II A, actually it ends at some very large but finite $\chi_M = Q_M H$.

Here it is in order to note that, although the amplitudes of modulations with an extremely large $\chi \sim \chi_M$ cannot be found without incorporating the higher-order term R, the above asymptotic results, obtained from the standard equations valid only for a sufficiently small χ, can give us an idea about the exact amplitude behavior for $\chi \sim \chi_M$. Indeed, the higher-order term R describes an increasing elastic resistance to large deformations, and it is natural to assume that its incorporating would have resulted in even smaller modulation amplitudes. For small χ it means nothing as the linear elasticity dominates in this range, whereas for modulations with a very large χ it gives more ground to believe that their exact amplitudes remain negligible as predicted by the analysis above. This reasoning allows one to assume that the exponential amplitude decrease at the right side of the short wavelength spectrum provides a natural upper χ bound for physically interesting modulations. In other words, extremely short wavelength modes, that we cannot describe accurately by the standard equations, are of no interest as they have negligible amplitudes; while all modes with non-negligible amplitudes are described accurately by these equations. This remarkable property makes our consideration self-consistent and closed.

The inequality $\Delta h<0$ can take place only if $\Delta p_{||}<0$. For $\Delta h<0$, in addition to the short wavelength branch there appears a long wavelength branch which begins at $\chi=0$ and ends at some χ_{sl} (solid curves in Figs. 1 and 2). The positions of the end χ_{sl} of the long wavelength spectrum and the

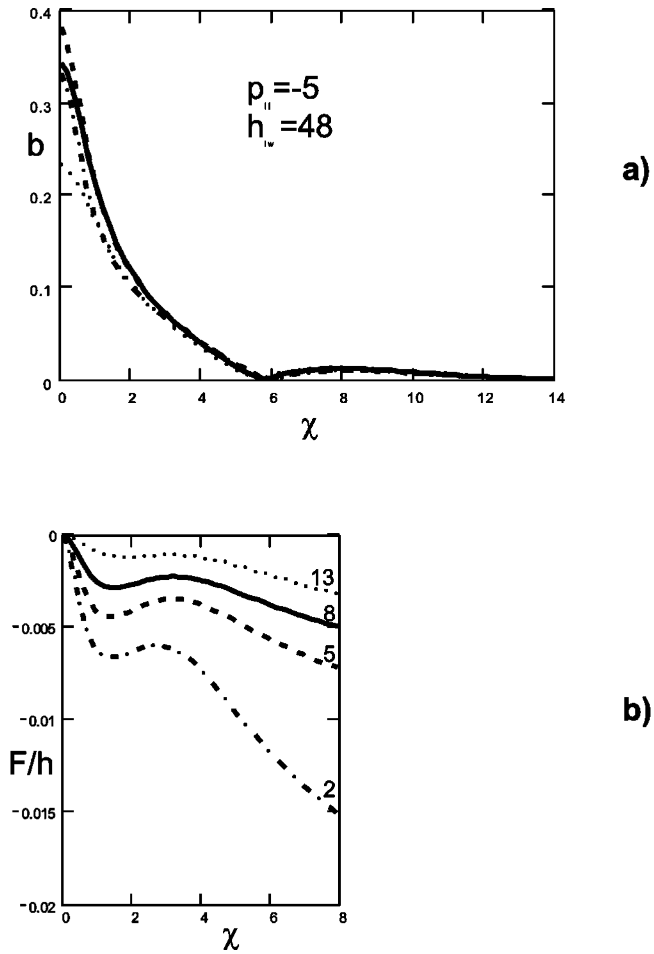

FIG. 3. (a),(b) The amplitude b (in radians) (a) and energy density F/h (in units K_{11}/h^2) (b) as functions of the dimensionless wave number χ for $p_{||}$ fixed at $p_{||}=-5$ (which corresponds to $h_{lw}=198$) and four different values of h: $h=13$ (dotted), $h=8$ (solid), $h=5$ (dashed), and $h=2$ (dash-dotted). The reduced twist and bend constants are $k_{22}=70$, $k_{33}=90$.

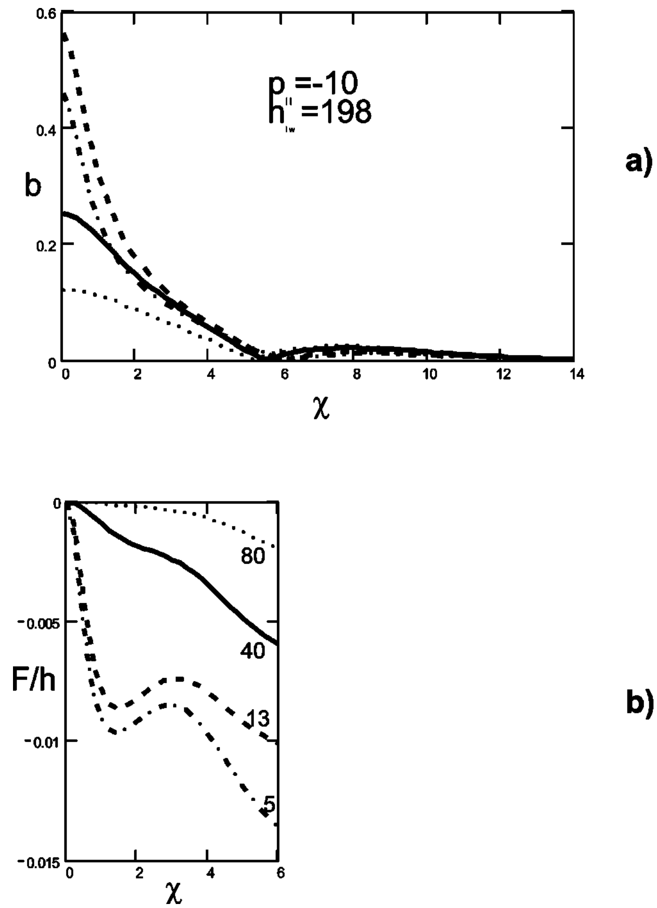

FIG. 4. (a), (b) The amplitude b (in radians) (a) and energy density F/h (in units K_{11}/h^2) (b) as functions of the dimensionless wave number χ for $p_{||}$ fixed at $p_{||}=-10$ (which corresponds to $h_{lw}=198$) and four different values of h: $h=80$ (dotted), $h=40$ (solid), $h=13$ (dashed), and $h=5$ (dash-dotted). The reduced twist and bend constants are $k_{22}=70$, $k_{33}=90$.

beginning χ_{sw} of the short wavelength spectrum depend both on Δh and $\Delta p_{||}$ (Figs. 1 and 2): the closer Δh and $\Delta p_{||}$ to 0, the smaller χ_{sl}, the larger χ_{sw}, and thus the longer the gap $\chi_{sw}-\chi_{lw}$ between the two spectral branches. As Δh and $\Delta p_{||}$ become more and more supercritical (i.e., negative and large in the modulus) these two points first coincide (dash-dot curves in Figs. 1 and 2), and then the gap between the branches disappears (Figs. 1–4).

In previous sections we have restricted the formulas to consideration of a single mode with an arbitrary value χ of the wave number. The spectrum of the instability, however, shows that the periodic modulation can contain many harmonics: the modulation energy is negative for a continuous range of χ. For the periodic wave, this implies that it is a superposition of the form $\Sigma_{n=1}a_n\cos(\chi_0 n/H)$ with some principal wave number χ_0. In principle, this wave number can be found as a minimizer of the sum (34), but this is very difficult. Indeed, different harmonics are independent only in the linear approximation: only the total quadratic energy of a nonmonochromatic wave is the sum of energies (11) of the constituent harmonics. Such an additivity does not take place in the fourth-order terms: the quadric functional (30) describes harmonics interaction which determines the amplitudes a_n via n nonlinear equations. However, the principal wave numbers of the instability and the correspondent amplitudes can be estimated qualitatively from the dependences $b(\chi)$ and $F(\chi)$ without solving this complicated nonlinear problem.

The curves on Figs. 1–4 suggest that, in spite of having in some situations no gap between long and short spatial waves and thus no formal division into two branches, there are two quite different wave numbers that determine the observable modulation. The first one is determined by the local energy minimum that lies at χ between 1.5 and 2; and the second one is determined by the local amplitude maximum that lies close to $\chi\sim 8$. Figs. 1, 3, and 4 show that the amplitude of the principal long wave (~ 0.1) is considerably larger than the amplitude of the principal short wave. However, Fig. 2 suggests that in some situations these waves can have small but comparable amplitudes. Now we note that modulations with very different wave numbers can always be considered quasi-independent: the fast spatial mode with a large wave number follows adiabatically the amplitude change of the slow spatial wave with a small wave number. This drives us to the following plausible picture of the instability appearance.

The spontaneous modulation has a wave number of the order of 2, and its amplitude can be small or as large as a few tens of degrees. This wave can have a fine structure with a wave number of order of 10. The fine structure can be seen or can have an insufficiently large amplitude to be easily observed. In addition, even if the parameters of the problem do not allow for the instability with a directly observable amplitude and spatial period to occur (e.g., the thickness is too large), the energy of the system can nevertheless be lower than the energy of the homogeneous nematic LC at the expense of the instability with a very short wavelength and small amplitude. This can result in an observable effect which is considered below.

B. Spontaneous modulations and the temperature of a nematic–smectic-A transition

Any deformation of a nematic director—spontaneous or induced by an external source—adds some amount of energy to the energy of the uniform state. This can influence conditions that determine transitions of this nematic LC into other phases, and correspondent shift of the transition temperatures can, in principle, be observed. Here we will be concerned with the temperature of a nematic–smectic-A transition.

There is a fundamental difference between the temperature shift of the transition to a uniform SmA phase from a nematic phase distorted spontaneously (N_{sd}), and from a nematic phase with distortions induced by an external force (N_{ind}): these shifts have different signs which gives a principal possibility to distinguish between the spontaneously modulated phase N_{sd} and the standard uniform nematic phase N_{ind} with distortions induced by an external source. Let us show this.

Spontaneous deformations lower the energy because they appear in the ground state of a nematic phase which means that

$$F(N_{sd})<F(N_u). \tag{37}$$

In contrast, if the nematic ground state is uniform then any deformations cost a positive amount of energy, i.e.,

$$F(N_{ind})>F(N_u). \tag{38}$$

The temperature T_0 of a standard transition $N_u \rightarrow \mathrm{Sm}A$ between the two spatially uniform phases can serve an obvious control point, and all other related temperatures will be considered relative to it. The transition temperatures $T(N_u \rightarrow \mathrm{Sm}A) \equiv T_0$, $T(N_{sd} \rightarrow \mathrm{Sm}A)$, and $T(N_{ind} \rightarrow \mathrm{Sm}A)$ are determined by the following obvious equalities:

$$F(N_u)=F(\mathrm{Sm}A) \Rightarrow T_0,$$

$$F(N_{sd})=F(\mathrm{Sm}A) \Rightarrow T(N_{sd} \rightarrow \mathrm{Sm}A), \tag{39}$$

$$F(N_{ind})=F(\mathrm{Sm}A) \Rightarrow T(N_{ind} \rightarrow \mathrm{Sm}A).$$

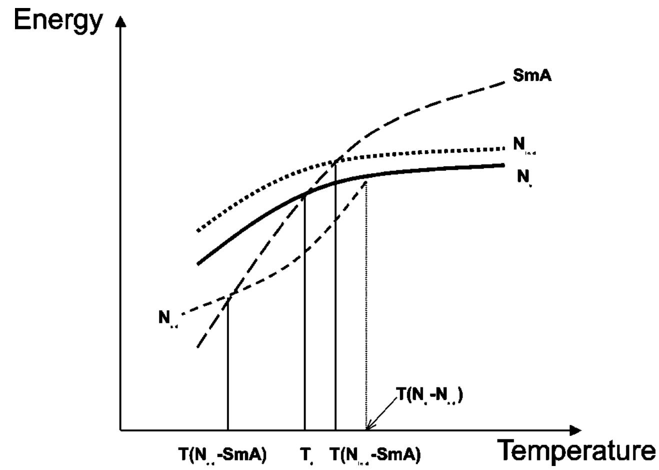

FIG. 5. Shift of the temperature of a nematic–smectic-A transition in a nematic LC with spontaneous and induced deformations. $T(N_u - N_{sd})$ is the temperature at which a uniform nematic gets unstable and a spontaneous modulation appears. Qualitative dependence phase energy vs temperature (both in arbitrary units): solid line, uniform nematic phase; long-dash line, smectic A phase; dot line, nematic phase with deformations induced by an external source; short-dash line, spontaneously modulated nematic phase.

Further, the energy of both N_u and SmA is a decreasing function of the temperature, and, for $T>T(N \rightarrow \mathrm{Sm}A)$, the energy of a nematic phase N is lower than the energy of a smectic-A phase, and vice versa. This picture is illustrated in Fig. 5. It clearly shows that a spontaneously modulated phase transforms into a SmA phase at the temperature which is lower than the temperature T_0 of the standard transition from a uniform nematic phase, whereas deformations induced in a nematic with the uniformed ground state bring the temperature of the transition up. This can be expressed by the following inequalities:

$$T(N_{sd} \rightarrow \mathrm{Sm}A)<T_0, \tag{40}$$

$$T(N_{ind} \rightarrow \mathrm{Sm}A)>T_0.$$

A negative shift of the nematic–smectic-A transition temperature gives a possibility to identify a spontaneously modulated phase. Furthermore, it is easy to see that the results obtained in this and previous sections imply that the temperature of the transition $N_u \rightarrow \mathrm{Sm}A$ in a planar cell can be lower than that in a homeotropic cell even if a director modulation is not observed in the planar cell. Indeed, assume that a planar cell has isotropic surfaces, and, thus, the azimuthal anchoring is negligible, but the cell thickness is larger than the upper critical thickness h_{lw} (28) of the long wavelength modulation. Then sufficiently close to the nematic–smectic-A transition, in a planar cell there appear short wavelength modulations that are not easily observable but produce negative energy $-F_M$, see Sec. II A. At the same time, no modulation appears in a homeotropic cell as the K_{24} term vanishes here. This implies that

$$T_{0,pl}<T_{0,hm},$$

where $T_{0,pl}$ is the temperature of apparently uniform N-SmA transition in a planar cell, and $T_{0,hm}$ is the temperature of a N_u-SmA transition in a homeotropic cell. All the effects are better pronounced for smaller thicknesses. For instance, submicrometer thin planar films seem to be the best for detecting spontaneous modulations. However, as far as the temperature shift is concerned, the cell thickness must be not too small to avoid size effects on the N-SmA transition temperature.

IV. CONCLUSION

We predicted a spontaneously modulated intermediate nematic phase that can be expected in a narrow temperature interval between conventional uniform nematic and smectic-A phases. This possibility is derived from the relation (5) between the saddle-splay and twist elastic constants, predicted by the elastic theory, and the critical growth of the twist elastic constant in the proximity of a N-SmA transition. By virtue of this relation, in this temperature range the saddle-splay term breaks the uniformity of the nematic director ground state. The phase can manifest itself as a spontaneous director modulation in a cell with azimuthally isotropic surfaces with a planar anchoring. The cell thickness should be sufficiently small but can be considerably larger than the anchoring extrapolation length. The temperature of the spontaneously modulated "saddle-splay" nematic–smectic-A transition is predicted to be lower than that of the uniform nematic–smectic-A transition.

Of course, the relation (5) should be considered with certain circumspection as close to a smectic phase it can be strongly influenced by nascent fluctuations of the density. Moreover, the result of Ref. [4] suggests that the value of K_{24} notably deviates from that predicted by Eq. (5) even in a nematic LC which does not have a smectic phase. We hope, however, that this relation shows at least a correct tendency. Experimental observation of a saddle-splay nematic phase can throw a considerable light into this problem related to an intrinsic ability of LCs to spontaneous pattern formation.

ACKNOWLEDGMENTS

This work was supported by an INCO Copernicus Photocom grant and INFM grant (INFM, Calabria UdR) which made possible V.M.P.'s visit to Politecnico di Torino.

[1] J.L. Ericksen, Phys. Fluids **9**, 1205 (1966).
[2] V.M. Pergamenshchik, Phys. Rev. E **61**, 3936 (2000).
[3] G. Barbero, N.V. Madhusudana, and C. Oldano, J. Phys. (Paris) **50**, 2263 (1989).
[4] O.D. Lavrentovich and V.M. Pergamenshchik, Phys. Rev. Lett. **73**, 979 (1994).
[5] O.D. Lavrentovich and V.M. Pergamenshchik, Int. J. Mod. Phys. B **12**, 2389 (1995).
[6] In this concern it is interesting to note that a wide accepted setting $K_{24}=0$ automatically puts nematic phase to the stability edge as then the inequalities (2) and (3) are broken and replaced by the equalities $|p_{||}|=1$ and $-2k_{22}=-2k_{22}<0$.
[7] S. Faetti and M. Riccardi, J. Phys. II **5**, 1165 (1995).
[8] V.M. Pergamenshchik and S. Žumer, Phys. Rev. E **59**, R2531 (1999).
[9] This formula incorporates spatial boundedness of a nematic body. In the formula, originally derived by Nehring and Saupe in Ref. [10], the spatial boundedness was disregarded which resulted in replacing K_{11} with the so-called bare constant $K_{11}^{(0)}$ ($K_{11}=K_{11}^{(0)}-2K_{13}$). In the case of our interest both formulas give the same value of K_{24} as $K_{22}\gg K_{11}$.
[10] J. Nehring and A. Saupe, J. Chem. Phys. **54**, 337 (1971).
[11] P.G. de Gennes and J. Prost, *The Physics of Liquid Crystals* (Oxford University Press, Oxford, 1993).
[12] L. Cheung, R.B. Meyer, and H. Gruler, Phys. Rev. Lett. **31**, 349 (1973).
[13] G. Barbero and C. Oldano, Nuovo Cimento D **6**, 479 (1985).
[14] V.M. Pergamenshchik, Phys. Rev. E **48**, 1254 (1993).
[15] V.M. Pergamenshchik, Phys. Lett. A **243**, 167 (1998); Phys. Rev. E **58**, R16 (1998).
[16] I. Dozov *et al.*, in *the European Conference on Liquid Crystals*, *Hersonissos*, *Greece*, edited by D.P. Photinos (University of Patras, Patras, 1999).
[17] V.M. Pergamenshchik, Phys. Rev. E **47**, 1881 (1993).
[18] G.P. Crawford and S. Žumer, Int. J. Mod. Phys. B **12**, 2469 (1995).
[19] V.M. Pergamenshchik, P.I.C. Teixeira, and T.J. Sluckin, Phys. Rev. E **48**, 1265 (1993); V.M. Pergamenshchik, Ukr. Fiz. Zh. **38**, 59 (1993) (in Russian).
[20] Calculations of the total fourth-order term in Refs. [17,2] show that its restricting solely to F_4 results in slightly underestimated amplitudes.