Thermalized polarization dynamics of a discrete optical-waveguide system with four-wave mixing

S. A. Derevyanko[*]

Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel
(Received 10 July 2013; published 30 September 2013)

Statistical mechanics of two coupled vector fields is studied in the tight-binding model that describes propagation of polarized light in discrete waveguides in the presence of the four-wave mixing. The energy and power conservation laws enable the formulation of the equilibrium properties of the polarization state in terms of the Gibbs measure with positive temperature. The transition line $T = \infty$ is established beyond which the discrete vector solitons are created. Also in the limit of the large nonlinearity an analytical expression for the distribution of Stokes parameters is obtained, which is found to be dependent only on the statistical properties of the initial polarization state and not on the strength of nonlinearity. The evolution of the system to the final equilibrium state is shown to pass through the intermediate stage when the energy exchange between the waveguides is still negligible. The distribution of the Stokes parameters in this regime has a complex multimodal structure strongly dependent on the nonlinear coupling coefficients and the initial conditions.

I. INTRODUCTION

The equilibrium behavior and the equipartition of energy between various degrees of freedom in nonlinear, noninte-grable discrete systems has attracted considerable interest since the seminal study of Fermi, Pasta, and Ulam [1]. In Hamiltonain systems with conserved number of excitations (waves) the maximum entropy principle suggests that in the final state of thermal equilibrium the statistics of the system is given by grand-canonical Gibbs distribution with the effective temperature and chemical potential [2–6]. However unlike the conventional statistical mechanics the effective temperature of this grand-canonical distribution depends on the initial position in the phase space and for certain regions can become negative making the distribution non-normalizable. Such regime is commonly attributed to the emergence of stable, localized, nonlinear structures corresponding to solitons in continuous systems [3] and discrete breathers [4,7,8] in discrete systems. From the point of view of the wave turbulence theory [9,10] the resulting equilibrium distribution provides stationary Rayleigh-Jeans spectra [9]. Also thermalization of light in nonlinear multimode waveguides and cavities has recently attracted attention in the context of classical optical wave condensation [11].

Here we will study the phenomenon of thermalization in the context of light propagation in a system of coupled nonlinear optical waveguides but the results can have wider applicability beyond the scope of the nonlinear optics. When the individual waveguide modes are strongly localized the nonlinear propagation of light is most commonly modeled by the discrete nonlinear Schrödinger equation (DNLSE) [12]. In fact most studies of thermalization in nonlinear *discrete* systems have concentrated on DNLSE in one [4–8,13] or two [14] dimensions. Thanks to the plethora of results in the field of DNLSE thermalization the structure of the final equilibrium state and the thermodynamical conditions for the occurrence of discrete breathers are now well understood. Among numerous discoveries in this area we would like to point the reader's attention to the universal correlations in one-dimensional (1D) systems of optical waveguides predicted in Ref. [13] in the limit when the nonlinearity dominates over the linear coupling. In this limit the effective dimensionless temperature turns out to be a universal constant independent of system parameters (provided that the initial state is characterized by uniform intensities) and the same universality is also manifested in the shape of the field correlation function.

In this paper we would like to focus on a much less studied model, namely, the thermalization of two coupled fields in the presence of four-wave mixing (FWM) [12]. In the context of nonlinear optics the situation corresponds to the propagation of polarized light in a birefringent material [15,16] or mode interaction from different Floquet-Bloch bands [17]. Here we will concentrate on the first case, however, the results presented here are quite general and can be applied to other nonlinearly coupled systems. In order to give reference to the real-world units we use AlGaAs as an common example of a material with cubic symmetry and fused silica as the corresponding example of isotropic crystal.

II. MODEL

The wave dynamics of the two orthogonally polarized fields is given by the following pair of coupled equations [15,16]:

$$i \frac{d\,a_n}{dz} + k\,a_n + C(a_{n-1} + a_{n+1}) + \gamma(|a_n|^2 + \lambda_1 |b_n|^2)a_n \\ + \gamma\,\lambda_2\,b_n^2\,a_n^* = 0 \tag{1a}$$

$$i \frac{d\,b_n}{dz} - k\,b_n + C(b_{n-1} + b_{n+1}) + \gamma(|b_n|^2 + \lambda_1 |a_n|^2)b_n \\ + \gamma\,\lambda_2\,a_n^2\,b_n^* = 0, \quad n = 1, \ldots, N. \tag{1b}$$

In the above equations, a_n and b_n are slowly varying field envelopes of the TE and TM polarized waves, $k = k_0(n_x - n_y)/2$ is the polarization mode dispersion constant, k_0 is the vacuum wave vector, $n_x - n_y$ is the linear birefringence ($n_x - n_y = 1.8 \times 10^{-4}$ for AlGaAs), C is the coupling constant

───────────
[*]Also at Aston Institute of Photonic Technologies, Aston University, Birmingham, United Kingdom; stanislav.derevyanko@weizmann.ac.il

$(C \sim 0.1 \text{ mm}^{-1})$, $\gamma = (k_0/2) n_2 n (\varepsilon_0/\mu_0)^{1/2}$ is the nonlinear coefficient, n_2 is the Kerr coefficient ($n_2 = 1.5 \times 10^{-13} \text{ cm}^2/\text{W}$ for AlGaAs), and $n = n_x$ is the linear refractive index ($n = 3.3$ for AlGaAs). The dimensionless constants λ_1 and λ_2 represent the relative strength of cross-phase modulation and four-wave mixing (XPM and FWM). If one puts $\lambda_1 = \lambda_2 = 0$ the system (1) breaks into two independent scalar DNLSE equations. We can restrict ourselves to the case of positive coupling $C > 0$ since the case of negative coupling can be recovered via a standard staggered transformation $a_n \to (-1)^n a_n$, $b_n \to (-1)^n b_n$. The change of sign in the nonlinearity can be also be compensated via a more complicated transformation, which involves staggering, complex conjugation, and swapping: $a_n \to (-1)^n b_n^*$, $b_n \to (-1)^n a_n^*$. Both transformations only affect the field correlation functions and phase distributions (and not, e.g., the intensity distributions) in a controlled way and here without loss of generality we will also restrict ourselves to the case of positive nonlinearity. In this paper we assume periodic boundary conditions although in the thermodynamic limit $N \to \infty$ this choice is not essential. Note in passing that continuous analogs of system (1) were studied in Ref. [18] with regard to pulse propagation in optical fibers.

In any chosen nonlinear medium the dimensionless XPM and FWM constants, λ_1 and λ_2, are not independent and three possible cases of interest can be envisaged [19]:

(1) Anisotropic cubic medium (e.g., AlGaAs): $\lambda_1 = 2\lambda_2$.

(2) Generic isotropic medium: $\lambda_1 = 1 - \lambda_2$.

(3) Isotropic cubic medium (e.g., fused silica): $\lambda_1 = 2/3$, $\lambda_2 = 1/3$.

We will refer to cases (1) and (2), (3) as anisotropic and isotropic correspondingly.

The system is Hamiltonian with the Hamiltonian function:

$$\mathcal{H} = k \sum_n (|a_n|^2 - |b_n|^2) + C \sum_n (a_n a_{n+1}^* + b_n b_{n+1}^* + \text{c.c.})$$
$$+ \frac{\gamma}{2} \sum_n \left[|a_n|^4 + |b_n|^4 + 2\lambda_1 |a_n|^2 |b_n|^2 + 2\lambda_2 \text{Re}\left(a_n^2 b_n^{*2}\right) \right],$$
(2)

which is a natural conserved quantity in the system while the additional integral of motion is provided by the total pulse power (proportional to the sum of local intensities)

$$\mathcal{P} = \sum_n (|a_n|^2 + |b_n|^2).$$

Additionally in the absence of the four-wave mixing ($\lambda_2 = 0$) the individual powers in each polarization $P_a = \sum_n |a_n|^2$ and $P_b = \sum_n |b_n|^2$ are conserved.

In the state of thermodynamic equilibrium the stationary field distribution $P(\{a_n, a_n^*, b_n, b_n^*\})$ maximizes the entropy $S = -\int P \ln P \prod_n da_n \, da_n^* \, db_n \, db_n^*$. However the nonlinear evolution always takes place on the shell $\mathcal{H} = \text{const}$ and $\mathcal{P} = \text{const}$, which introduces two constraints for the optimization problem. The solution then represents a grand-canonical Gibbs distribution (see, e.g., [4]):

$$P(\{a_n, a_n^*, b_n, b_n^*\}) = \mathcal{Z}^{-1} \exp\left[-\beta(\mathcal{H} - \mu \mathcal{P})\right] \quad (3)$$

where the Lagrange multipliers β and μ play the roles of the inverse temperature and chemical potential, respectively

while the normalizing factor $\mathcal{Z}$ has the familiar meaning of the partition function.

The assumption of field thermalization implies that instead of averaging the dynamics of the system (1) over a long interval of z one can compute the same averages via the equilibrium Gibbs measure (3). However from the point of view of the nonlinear optical waveguides it is impractical to use averaging over large distances since it requires optical waveguides that are much too long. Instead the averaging can be understood as averaging over *disordered initial conditions* that can be experimentally controlled [13,14]. This brings about the notion of the thermalization distance [14] z_{th} after which the information about the initial state of the system is forgotten and the averaging over the initial conditions is equivalent to Gibbs averaging. In our treatment we will assume (unless otherwise specified) that the initial amplitudes for both TE and TM components are constant $|a_n| = a$, $|b_n| = b$ while the phases are uncorrelated uniformly distributed random variables. Such assumption is not necessary but it simplifies the calculation of the ensemble averages of the initial Hamiltonian and power.

The knowledge of the partition function $\mathcal{Z} = \int \prod_n d a_n da_n^* db_n \, db_n^* \exp\left[-\beta(\mathcal{H} - \mu \mathcal{P})\right]$ allows one to calculate an average energy per waveguide $h = \mathcal{H}/N$ and the average intensity per waveguide $p = \mathcal{P}/N$. On the other hand such averages must correspond to their initial values h_0, p_0 (averaged over the disordered initial conditions) since for each realization of the disorder these are conserved integrals of motion. Thus, from (3) it follows that

$$h_0 = -\frac{1}{N} \frac{\partial \ln \mathcal{Z}}{\partial \beta} + \mu \, p_0, \quad p_0 = \frac{1}{\beta N} \frac{\partial \ln \mathcal{Z}}{\partial \mu}. \quad (4)$$

For each set of phase-averaged initial conditions [i.e., given pair (h_0, p_0)] the above two transcendental equations yield the effective inverse temperature β and the chemical potential, μ. As in the scalar case [4] the necessary condition for the Gibbs distribution (3) to be normalizable is the positiveness of the temperature, $\beta > 0$. Thus the curve $\beta = 0$ in the (h_0, p_0) diagram represents a natural boundary between the conventional thermalization region ($\beta > 0$) and the area where $\beta < 0$ and the energy is localized in a form of discrete breathers that have been observed experimentally [15].

It is convenient to introduce a canonical transformation to real intensities and phases: $a_n = \sqrt{A_n} \exp(i \, \phi_{1,n})$, $b_n = \sqrt{B_n} \exp(i \, \phi_{2,n})$ with $\delta_n \equiv \phi_{2,n} - \phi_{1,n}$ being the phase difference between the two field components. The case $\delta_n = m\pi$ (m integer) corresponds to the linear polarization in the nth waveguide, $\delta_n = \pm \pi/2$, $A_n = B_n$ describe circular polarization, etc. This notation corresponds to the Jones description of polarization (see, e.g., [20]). To calculate the partition function $\mathcal{Z}$ one needs to integrate the Gibbs exponential in (3) with the Hamiltonian (2). In the new variables the integral takes the form:

$$\mathcal{Z} = \int \prod_n dA_n \, dB_n \, d\phi_{1,n} \, d\phi_{2,n} \exp\{-\beta[k(A_n - B_n)$$
$$+ H_C(A_n, B_n, \phi_{1,n} - \phi_{1,n+1}, \phi_{2,n} - \phi_{2,n+1})$$
$$+ H_\gamma(A_n, B_n, \delta_n)]\}, \quad (5)$$

where we have collected all the nonlinear coupling terms as well as chemical potential in the nonlinear interaction part $\mathcal{H}_\gamma = \sum_n H_\gamma(A_n, B_n, \delta_n)$ with

$$H_\gamma(A, B, \delta) = \frac{\gamma}{2} \left[A^2 + B^2 + 2\lambda_1 AB + 2\lambda_2 AB \cos(2\delta) \right] - \mu(A + B) \tag{6}$$

while the linear waveguide coupling is given by the Hamiltonian $H_C = \sum_n [\sqrt{A_n A_{n+1}} \cos(\phi_{1,n} - \phi_{1,n+1}) + \sqrt{B_n B_{n+1}} \cos(\phi_{2,n} - \phi_{2,n+1})]$.

The Gibbs distribution (3) is normalizable when the partition function is finite. A close inspection of Eq. (5) reveals that in order to achieve this not only the temperature must be positive $\beta > 0$, but additionally the inequality

$$|\lambda_1 - \lambda_2| < 1 \tag{7}$$

must hold. Already we can see a departure from the scalar case [4] where the Gibbs distribution is always normalizable as long as the temperature remains positive. For the isotropic case $\lambda_2 = 1 - \lambda_1$ this corresponds to possible thermalization for $0 < \lambda_1 < 1$ and for the anisotropic case $\lambda_2 = \lambda_1/2$ this implies $-2 < \lambda_1 < 2$. The borderline case $|\lambda_1 - \lambda_2| = 1$ (which includes the XPM without FWM $\lambda_1 = 1$, $\lambda_2 = 0$) generally requires special treatment and we will not consider it here. The non-normalizable property of the Gibbs distribution indicates the emergence of the localized structures, i.e., the genuine equilibrium state now consist of high-amplitude discrete breather (or several breathers if the system has not yet quite reached the equilibrium) interacting weakly with a small quasilinear background thermalized at infinite temperature, $\beta = 0$ [3,7,8]. So here the statistical mechanics provides us with a clue as to the regions in the parameter space where the localized structures can be observed in principle [4,6].

In this paper we will only study the thermalization regime where the inequality (7) is fulfilled and the temperature is positive so that Gibbs distribution (3) is always normalizable. We leave the analysis of the localized structures for future studies although we will comment on these in the following section. Since it is impossible to evaluate the partition function $\mathcal{Z}$ in the closed form we will study three different limiting regimes, which for the scalar case were already analyzed in Refs. [4,6,13]. These regimes are: (i) low-temperature limit $\beta \to \infty$, (ii) high-temperature limit $\beta \to +0$ (which also serves as a borderline for emergence of localized structures), and (iii) the anticontinuum (high-intensity) regime when the effective nonlinearity parameter Γ defined below in Sec. IV is large, $\Gamma \gg 1$. It turns out that doubling the amount of degrees of freedom as compared to the scalar case has a significant impact on the statistical properties of system, (1). We start our analysis with the first two regimes: $\beta \to \infty$ and $\beta \to 0$.

III. TEMPERATURE BOUNDARIES AND THERMALIZATION REGION

The limit of $\beta \to \infty$ corresponds to the configuration that minimizes the Hamiltonian (2) subject to the given conserved pulse intensity per waveguide p. If we restrict our search to a solution with constant amplitudes, phase shifts, and locked state of polarization (i.e., fixed δ) the minimal configuration is achieved by the following field distribution:
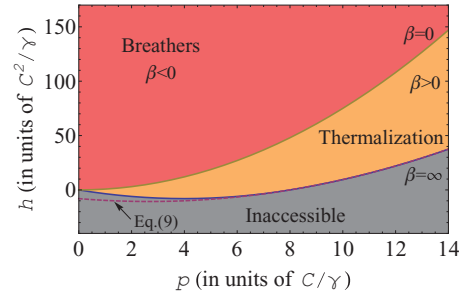

FIG. 1. (Color online) The different regions in (h, p) parameter space. The ratio $k/C = 2$ was assumed.

$a_n = \sqrt{A} \exp(i \pi n)$, $b_n = \sqrt{B} \exp(i\pi n + i\pi/2)$ where the amplitudes A and B as well as chemical potential (i.e., the corresponding Lagrange multiplier), μ, are given by:

$$A = \frac{p}{2} - \frac{k}{\gamma} \frac{1}{1 - \lambda_1 + \lambda_2}, \quad B = \frac{p}{2} + \frac{k}{\gamma} \frac{1}{1 - \lambda_1 + \lambda_2}$$
$$\mu = \frac{1}{2} p \gamma (1 + \lambda_1 - \lambda_2). \tag{8}$$

This solution exists only for not too low intensities, i.e., for $p > (2k/\gamma)(1 - \lambda_1 + \lambda_2)^{-1}$ and provides a low bound for the energy per waveguide:

$$h_{\min} = -2pC + \frac{(1 + \lambda_1 - \lambda_2)\gamma}{4} p^2 - \frac{k^2}{\gamma(1 - \lambda_1 + \lambda_2)}. \tag{9}$$

In the limit of zero birefringence, XPM and FWM, $\lambda_1 = \lambda_2 = k = 0$ we get the energy as a sum of energies of two identical scalar DNLSEs with the average intensity per waveguide $p/2$, which corresponds to the result of Ref. [4]. In Fig. 1 (which is an analog of Fig. 1 of Ref. [4]) we plot a phase diagram in the (h, p) space for a specific case of 16 waveguides with $\lambda_1 = 1$, $\lambda_2 = 0.5$. Both h and p have been rescaled to the dimensionless multiples of C^2/γ and C/γ respectively. One can see that Eq. (9) provides an excellent low bound approximation even below the critical intensity for which the solution of (8) exists (which is $p = 8$ for the chosen parameters). Let us now turn to the opposite case of high temperatures $\beta \to 0$ assuming that the product $\beta\mu < 0$ remains finite. We will use the method similar to that of Ref. [6] applied earlier to the scalar case. In this limit one can neglect the linear coupling energy H_C in the exponent of (5) together with the birefringence contribution $k(A_n - B_n)$. The partition function is then given by $\mathcal{Z} = [4\pi^2 y(\beta, \mu)]^N$ with

$$y(\beta, \mu) = \int_0^\infty dA \int_0^\infty dB \, \mathrm{I}_0 \left[\beta\gamma \lambda_2 AB \right]$$
$$\times \exp \left[-\frac{\beta\gamma}{2}(A^2 + B^2 + 2\lambda_1 AB) - \beta|\mu|(A + B) \right].$$

Next we assume that $\lim_{\beta \to 0}(\beta\gamma) = 0$ and write approximately

$$y(\beta, \mu) = \int_0^\infty dA \int_0^\infty dB \exp[-\beta|\mu|(A + B)]$$
$$\times \left(1 - \frac{\gamma\beta}{2}(A^2 + B^2 + 2AB\lambda_1) \right) + O((\beta\gamma)^2),$$

where we have used the fact that the product $\beta|\mu|$ is fixed while $\beta\gamma$ tends to zero. The answer is

$$y(\beta,\mu) = \frac{1}{(\beta\mu)^2} - \frac{\beta\gamma(2+\lambda_1)}{(\beta\mu)^4} + \cdots.$$

Finally from Eqs. (4) we obtain in the leading approximation:

$$\lim_{\beta\to 0}(\beta\mu) = -2/p, \quad h_{\max} = \frac{\gamma(2+\lambda_1)}{4} p^2. \quad (10)$$

The parabola $h_{\max}(p)$ provides the upper boundary $\beta = 0$ for the thermalization region in the plane of parameters (h,p) (see Fig. 1). Interestingly enough it does not depend on the four-wave mixing constant λ_2.

IV. NONLINEARITY-DOMINATED REGIME

In the following, we will adopt normalized units where the propagation distance is measured in units of coupling length $z \to C z$. We will also assume that the intensities of both TE and TM components are uniform $A_n(0) = A_0$, $B_n(0) = B_0$. We can also normalize the intensities of both components by the half of the initial intensity $p_0/2$ so that $A_0 + B_0 = 2$. In the new dimensionless units one must simply substitute k/C for k, 1 for C in the original coupled equations (1), and instead of the nonlinear coefficient γ one now has a dimensionless nonlinearity parameter $\Gamma = \gamma p_0/(2C)$. Since in most common nonlinear materials and waveguide geometries the ratio k/C is in the order of unity the parameter Γ indeed gauges the relative strength of nonlinearity [13,14].

In this section we will be interested in the highly nonlinear regime, $\Gamma \gg 1$. The motivation for this is twofold. Firstly, this regime permits almost full analytical treatment, which is always helpful when studying the general properties of any nonlinear system and secondly, in Ref. [13] it was shown that in the scalar case this limit corresponds to the reciprocal temperature β that does not depend on the value of parameter Γ and is a universal constant (in dimensionless units). The universality of the temperature in turn gives rise to a universal shape of the field correlation function. Therefore it is interesting to see how this result changes in the vector case. One should expect that the dynamics and statistics in the vector case are much richer due to the doubled number of interacting degrees of freedom. Here we will show that this is indeed the case.

Our main objects of interest will be the statistics of the intensity and polarization state of each waveguide as well as the distribution of the phase differences (i.e., phase gradients in the continuous limit) between the same components (e.g., TE) of the adjacent waveguides together with the field correlation functions. It turns out that in the strongly nonlinear regime the statistics of the phase gradient is decoupled from those of the intensity and polarization. The latter are most conveniently described by using a popular alternative to the Jones description of polarization, namely by introducing the four Stokes parameters $\{S_i\}_{i=0}^3$ [18–20]. Each waveguide has its own set of Stokes parameters, which are related to the Jones

parameters via

$$S_0^n = A_n + B_n$$
$$S_1^n = A_n - B_n$$
$$S_2^n = 2\sqrt{A_n B_n}\cos(\delta_n)$$
$$S_3^n = -2\sqrt{A_n B_n}\sin(\delta_n) \quad \sum_{i=1}^3\left(S_i^n\right)^2 = \left(S_0^n\right)^2. \quad (11)$$

The vector $\vec{S}^n = (S_1^n, S_2^n, S_3^n)$ is called a Stokes vector on a Poincaré sphere of radius S_0^n. The components of the Stokes vector for each waveguide are related to the polarization state of the waveguide while its magnitude provides the total intensity carried by both field components. For example, linear polarization corresponds to the equatorial plane $S_3^n = 0$ while the circular clockwise and anticlockwise polarizations correspond to the north and south poles $\vec{S}^n = (0,0,\pm S_0^n)$. In what follows we will use both Jones and Stokes descriptions whichever is more convenient.

A. Statistical properties of the field in the anticontinuum limit ($k = C = 0$)

We will start our analysis with the anticontinuum limit when one can completely neglect both linear coupling and birefringence in Eqs. (1). As will be seen later this corresponds to the initial stages of evolution of field distribution towards the final state of equilibrium. The absence of linear waveguide coupling makes field dynamics in each waveguide independent from the rest. Therefore instead of considering the system of coupled field equations (1) one can analyze the dynamics in a single waveguide. In particular the first three Stokes parameters obey the following equations (the waveguide index has been omitted for convenience):

$$\frac{dS_1}{dz} = 2\Gamma\lambda_2 S_2 S_3,$$
$$\frac{dS_2}{dz} = -\Gamma(1 + \lambda_2 - \lambda_1) S_1 S_3, \quad (12)$$
$$\frac{dS_3}{dz} = \Gamma(1 - \lambda_2 - \lambda_1) S_1 S_2.$$

This system is completely integrable and its various special cases and generalizations (e.g., the inclusion of the birefringence terms $\sim k$) have been extensively studied in literature both with connection to the nonlinear polarization dynamics [21] in general and the dynamics of the polarization-locked vector solitons [18,19] in particular. The complete integrability of system (12) is due to the existence of two integrals of motion of which the first is just the intensity, i.e., the zeroth Stokes parameter, S_0, and the second, R, is related to the Hamiltonian of the original system [21]:

$$S_0^2 = S_1^2 + S_2^2 + S_3^2,$$
$$R = \begin{cases} (3\lambda_1/2 - 1)S_2^2 - (1 - \lambda_1/2)S_3^2, & \text{anisotropic} \\ S_3, & \text{isotropic} \end{cases}. \quad (13)$$

In other words the dynamics of the system takes place on the intersection of the Poincaré sphere of radius S_0 and a hyperbolic (or elliptic) cylinder in the anisotropic case or a plane in the isotropic case given by the equation $R = \text{const}$.
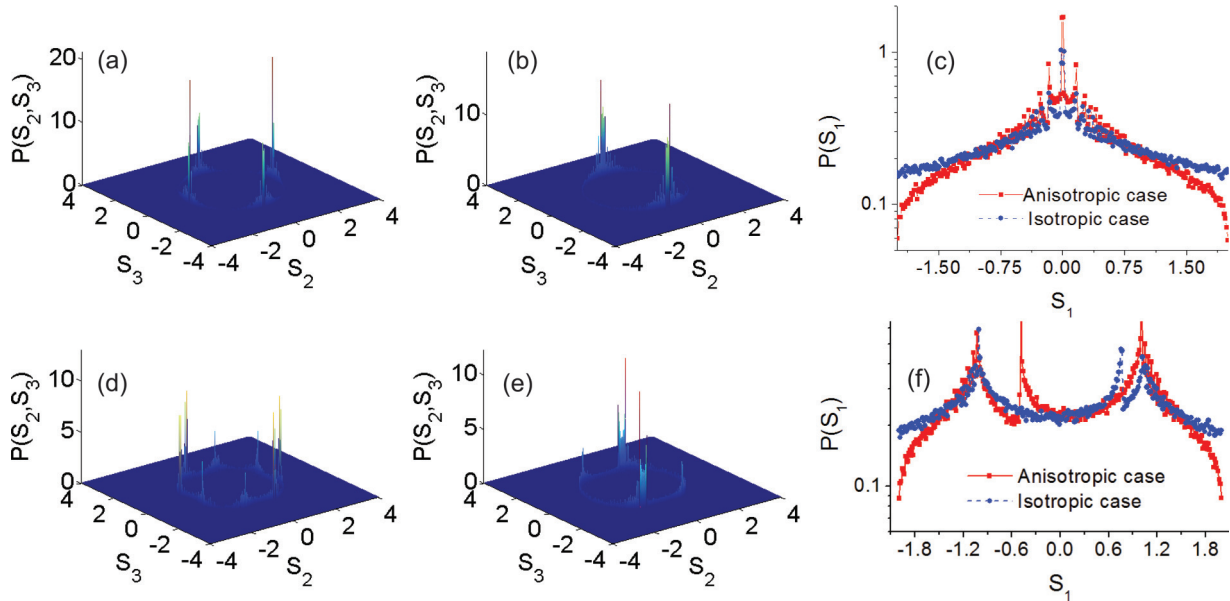
FIG. 2. (Color online) Marginal probability density functions for different Stokes parameters evaluated for the nonlinearity parameter $\Gamma = 50$ at distance $z = 10$ (in units of coupling lengths). Top row corresponds to the initial equipartition of intensity between the components $A_0 = B_0 = 1$ while the bottom row shows the case when $A_0 = 1.5$ and $B_0 = 0.5$. (a) and (d) correspond to anisotropic case with $\lambda_1 = 1$, $\lambda_2 = 1/2$, (b) and (e) assume isotropic case with $\lambda_1 = 2/3$, $\lambda_2 = 1/3$. Marginal PDF $P(S_1)$ is plotted in (c) and (f).

One can obtain an autonomous equation for S_1, which has the form of a Duffing oscillator equation in the anisotropic case [21] and harmonic oscillator equation in the isotropic case [19] (see also Appendix). The other two Stokes parameters are recovered from Eqs. (12) and (13) while the phase of the TE component can be determined by simple integration of the original field equations. Since the system in the anticontinuum limit is completely integrable it does not thermalize, i.e., formula (3) is inapplicable. Instead one must use the exact solution for the Stokes vector $\vec{S}(z)$ and average it directly over the initial conditions.

As mentioned earlier we will assume that all waveguides initially have the same set of intensities, A_0, B_0, $S_0 = A_0 + B_0 = 2$ and random, independent phases. From the definition of the Stokes parameters (11) it follows that that initial value $S_1(0) = A_0 - B_0$ is fixed while the points $[S_2(0), S_3(0)]$ are uniformly distributed on a circle of radius $2\sqrt{A_0 B_0}$. In Fig. 2 we show the evolution of the marginal probability density functions (PDFs) $P(S_2, S_3)$ and $P(S_1)$ for the anisotropic and anisotropic cases obtained by averaging the solution of system (12) over the initial phase distribution. The parameters $\lambda_{1,2}$ were chosen to correspond to the AlGaAs compounds in the anisotropic case and fused silica for the isotropic case. One can see that the structure of the histograms is very sensitive to both the values of λ coefficients and the symmetry of the initial condition. For example when the initial intensity is equally distributed between the components, i.e., $A_0 = B_0 = 1$, $S_1(0) = 0$ the marginal distribution $P(S_2, S_3)$ shown in Fig. 2(a) is four modal. It is largely confined to the initial circle of radius $S_0 = 2$ and the four maxima are at the points $(0, \pm S_0)$, $(\pm S_0, 0)$. At the same time the marginal PDF for the first Stokes parameter S_1 is sharply centered at zero [Fig. 2(c)], i.e., for most initial realizations the symmetry between the components is preserved $A(z) \approx B(z)$. We have also checked this property

for different runs with different distances, z (not shown). This means that in the anticontinuum limit for the anisotropic cubic crystal (e.g., AlGaAs) each given waveguide evolves into either linear or circular state of polarization, despite that no such preference existed in the initial conditions (the phase distributions were uniform on a circle). In the isotropic case one can observe that although the distribution of $P(S_1)$ is still sharply peaked around zero [Fig. 2(c)] the distribution of the two remaining Stokes parameters Fig. 2(b) is only bimodal (and not four modal as in the anisotropic case) with the two maxima at $(0, \pm S_0)$ corresponding to circular polarization only. The positions of the maxima in both cases can be easily explained by considering the two integrals of motion (13). As we have seen in both cases for most realizations one can neglect the value of S_1 and the PDF $P(S_1, S_2)$ therefore remains confined to the circle of radius $S_0 = 2$ for all values of z. Next if one looks at the distribution of the second integral of motion in (13), namely R, then for the initial values $[S_2(0), S_3(0)]$ uniformly distributed on a circle of radius 2 one can see that the distribution of R is bimodal with the two maxima at $R_- = -4(1 - \lambda_1 + \lambda_2)$ and $R_+ = 4(\lambda_1 + \lambda_2 - 1)$ in the anisotropic case and $R_\pm = \pm 2$ in the isotropic case. The intersection points of the pair of curves $R(S_2, S_3) = R_\pm$ with the circle in the (S_2, S_3) plane are exactly the four observed maxima in the anisotropic case and the two in the isotropic one.

In the case when initially the symmetry between the field components is broken (bottom row in Fig. 2) the situation is much more complex owing to the multimodal features of the distribution of the parameter S_1, see Fig. 2(f). Each maximum of the $P(S_1)$ gives rise to at least two maxima in the distribution $P(S_2, S_3)$ producing complex crownlike shapes shown in Figs. 2(d) and 2(e). Each of the maxima now corresponds to a certain elliptically polarized state, which varies from

waveguide to waveguide. Note also that the positions of the maxima remain fixed with the propagation distance z while their widths experience weak periodic oscillations (not shown). The full theoretical analysis of these multimodal distributions requires further study, which is beyond the scope of this paper.

B. Structure of final thermal state: Regime of universal temperature

We have seen above that in the anticontinuum limit when the birefringence and the linear waveguide coupling can be neglected completely the system is integrable and instead of reaching thermal equilibrium it experiences periodic oscillations that can be averaged directly over the initial conditions to produce multimodal distribution for the established state of polarization. The question then arises as to what will happen if both the birefringence and the coupling are taken into account. One can expect that at the initial stages of evolution when the distance is less than the thermal length, z_{th}, the evolution of the system is close to anticontinuum limit and the state of polarization follows the statistics described in the previous subsection. For $z \gtrsim z_{\text{th}}$ however, the nonintegrability of the system becomes essential. The linear coupling between the waveguides leads to energy and power exchange between the waveguides so that eventually most of the initial peaks in the marginal PDF $P(S_2, S_3)$ are destroyed and the final thermal distribution sets in. This final distribution is characterized by the Gibbs statistics (3) with the partition function given by (5).

We now wish to calculate the partition function and the probability distributions in this final state still assuming strong nonlinearity $\Gamma \gg 1$ but now taking into account the linear coupling and birefringence as well. Let us first assume *a priori* that the initial conditions are such that the system thermalizes into a state with finite temperature, i.e., $\beta \to$ const as $\Gamma \to \infty$. We will shortly see that this is only possible for a very special choice of the initial conditions. Then the dominating part of the exponent in (5) is the nonlinear interaction term $\mathcal{H}_\gamma$. Other terms (including the phase-dependent coupling energy) are relatively slow functions of the amplitudes and do not therefore contribute to the partition function. However it is important to retain the chemical potential term since as we shall see below $\mu \sim \Gamma$ always. When calculating the partition function (5) one can resort to the saddle point approximation in $\gamma = \Gamma$, which implies that at large values of Γ the main contribution to the integral comes from absolute minimum of $H_\gamma(A_n, B_n, \delta_n)$. This minimum is achieved for the values

$$A_* = B_* = \frac{\mu}{\Gamma(1 + \lambda_1 - \lambda_2)}, \quad \delta_* = \pm\pi/2. \quad (14)$$

In other words the absolute minimum of the interaction Hamiltonian subject to given total power is achieved by a circularly polarized state, the most symmetric of all. This minimum is also degenerate; for a system of N waveguides there are 2^N possible choices of polarization orientation. This degeneracy is drastically reduced however if one takes into account the coupling term H_c. This term favors the configuration where the phase difference δ_n is uniform across the waveguides, which leaves only two possibilities: either all fields are clockwise polarized or they are all anticlockwise polarized. These states also correspond to the two maxima

of the distribution function $P(A, B, \delta)$ [or $P(S_1, S_2, S_3)$ in Stokes parameters] one for each direction of rotation. As previously we assume here that $|\lambda_1 - \lambda_2| < 1$, which ensures the convergence of the integrals for positive temperature. The minimal value of the interaction energy (at fixed power) is given by

$$H_\gamma(A_*, B_*, \delta_*) = -\frac{\mu^2}{\Gamma(1 + \lambda_1 - \lambda_2)}.$$

If we assume uniform initial amplitudes A_0 and B_0 ($A_0 + B_0 = 2$) and neglect the terms that are not proportional to Γ in the limit $\Gamma \gg 1$ the average energy per waveguide is given by:

$$h_0 = \Gamma\left[\frac{A_0^2 + B_0^2}{2} + A_0 B_0(\lambda_1 + \lambda_2\langle\cos(2\delta_0)\rangle)\right]. \quad (15)$$

The energy is always bound from below by $h_{\min} = \Gamma(1 + \lambda_1 - \lambda_2) > 0$. In the spirit of saddle-point approximation we can now plug $A_n = A_*$, $B_n = B_*$, $\delta_n = \delta_*$ from Eq. (14) into linear coupling prefactor $\exp(-\beta H_C)$ when calculating the partition function. The integration over the phases $\phi_{1,n}$ of the TE component is reduced then to calculation of the partition function of the 1D classical XY model (see Refs. [4] or [22] for details). In the thermodynamic limit ($N \to \infty$) the contribution of this linear coupling term to the logarithm of the partition function (i.e., free energy) is $-N \ln I_0\{4\beta\mu/[\Gamma(1 + \lambda_1 - \lambda_2)]\}$. Note incidentally that in Ref. [22] it was shown that in the same limit the difference between different types of boundary conditions (free boundary or periodic boundary conditions) amounts to terms of the order $O(1)$ in the free energy that can be safely neglected. The remaining Gaussian integration over the fluctuations around the minimum of H_γ Eq. (14) is trivial. Inserting the resulting expression for the partition function $\mathcal{Z}$ into the system of equations (4) after some simple algebra we obtain the following transcendental equation for the inverse temperature β:

$$h_0 - h_{\min} = \frac{3}{2\beta} - \frac{4I_1(4\beta)}{I_0(4\beta)}, \quad \mu = \Gamma(1 + \lambda_1 - \lambda_2). \quad (16)$$

with h_0 given by Eq. (15). For the consistency of our approximation we must demand that the inverse temperature β remains finite as Γ goes to infinity (as is the case in the scalar DNLSE [13]). But this is achieved only when $h_0 = h_{\min}$ so that the left-hand side vanishes, which means that the initial input must be circularly polarized: $A_0 = B_0 = 1$, $\delta_0 = \pm\pi/2$. In other words, in order to obtain a universal (i.e., Γ-independent) temperature constant, similar to the scalar case, the initial state must necessarily be the one that minimizes the nonlinear part of the Hamiltonian, H_γ subject to the intensity constraint $A_0 + B_0 = 2$. The system then becomes effectively scalar and thermalization occurs only in the distribution of the TE phase differences $\theta_n = \phi_{1,n} - \phi_{1,n+1}$ while the field in each waveguide remains locked in its original clockwise or anticlockwise circular state of polarization. As in Ref. [13] the Gibbs distribution can be approximately factorized so that the intensity of each waveguide $P_n = A_n + B_n = S_0^n$ has a narrow distribution close to Gaussian centered around the conserved mean value of 2. The results of the numerical simulations of universal correlations in the vector case of $N = 64$ waveguides are presented in Fig. 3. As prescribed, we have started from a symmetric initial configurations $A_0 = B_0 = 1$ and uniform,
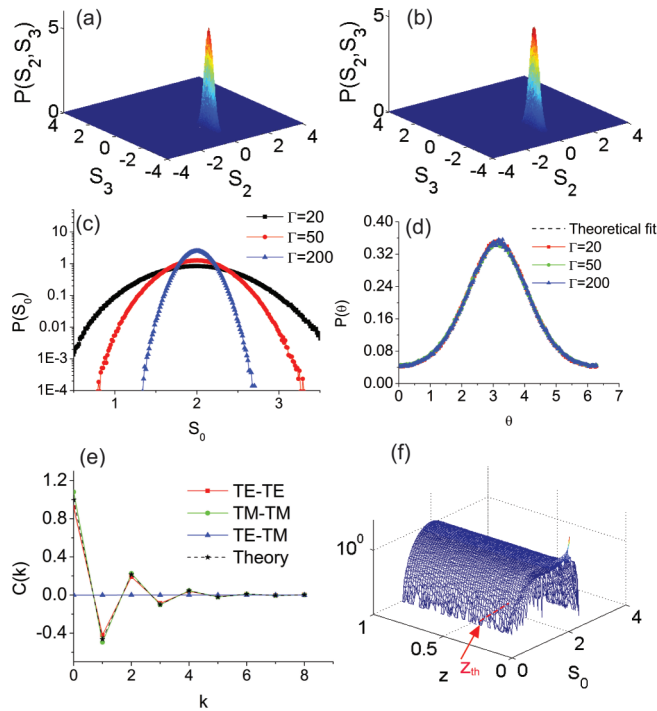
FIG. 3. (Color online) The statistics of the final thermal state in the limit of large nonlinearity when the initial field in *each* waveguide is circular clockwise polarized. (a), (b) The marginal distribution of Stokes parameters $P(S_2, S_3)$ for anisotropic and isotropic case respectively. (c), (d) The statistics of the intensity and phase difference (of the TE component) respectively for different values of the nonlinearity parameter. (e) Cross- and self-field correlation functions for the TE and TM components. (f) The evolution of the intensity PDF with distance. The thermalization length z_{th} is determined as a point where the distribution stabilizes.

uncorrelated phase distribution for the TE components. To ensure constant circular polarization locking the phases of the TM component were obtained by those of the TE by adding $\pi/2$ (corresponding to the clockwise field rotation). In all the figures (unless otherwise specified) the values of the parameters are: $k = 2$, $\Gamma = 50$ and the propagation length is $z = 10$ (all in the normalized units). In Figs. 3(a) and 3(b) we show numerical simulations of the marginal PDF $P(S_2, S_3)$ for the anisotropic ($\lambda_1 = 1, \lambda_2 = 1/2$) and isotropic ($\lambda_1 = 2/3, \lambda_2 = 1/3$) cases respectively. One can see that both distributions have narrow maxima at the point $(0, -2\sqrt{A_* B_*}) = (0, -2)$ corresponding to clockwise circular polarization (as in the initial state). We have found very little qualitative difference between the anisotropic and isotropic cases so Figs. 3(c)–3(f) feature only the former. In Fig. 3(c) we can see that the intensity PDF becomes more and more narrowly centered around the average value of 2 and the distribution is close to Gaussian with the variance proportional to $(\beta\Gamma)^{-1}$. As for the PDF for the TE phase difference, $P(\theta)$, the theoretical prediction $P(\theta) = [2\pi I_0(4\beta)]^{-1} \exp(-4\beta\cos\theta)$ is clearly corroborated by the numerics shown in Fig. 3(d): one can observe a perfect data collapse for different values of Γ. Similar to universal correlations in the scalar case, one can obtain universal filed correlations in the vector case. In Fig. 3(e) we plot the field correlation functions defined as $C_{qq'}(k) = \langle \text{Re}[q_n q_{n+k}'^*] \rangle$

where q, q' denote different components of the field, TE or TM, and the average is taken both over the waveguide position, n, and the initial conditions. Since in the thermal equilibrium the TE and TM components are always $\pi/2$ out of phase the cross-field correlation $C_{ab}(k)$ is always zero while the self-field correlation functions have identical universal shape:

$$C_{aa}(k) = C_{bb}(k) = \eta^k, \quad \eta = \langle \cos\theta \rangle = -\frac{I_1(4\beta)}{I_0(4\beta)}. \quad (17)$$

This is illustrated in Fig. 3(e), where all three components are plotted together with the theoretical fit. Finally, Fig. 3(f) illustrates how quickly the final thermal distribution is achieved. The initial δ peak in intensity distribution quickly broadens and settles to a stationary, almost Gaussian distribution of the type shown in Fig. 3(c) already at the point $z \approx 0.3$ (in units of the coupling length) that naturally serves as the thermalization length z_{th}. One can see that thermalization here occurs over just a fraction of the coupling length, i.e., on the scale of 0.15 mm for a typical AlGaAs waveguide system. Of course one can come up with a more rigorous quantitative definitions of z_{th}, by looking, e.g., at the saturation of the temperature defined from the simulated data via equipartition theorem [14] but for the purposes of this paper we can restrict ourselves to the crude estimate above. All other distributions also stabilize at the same length.

As for the value of the universal temperature constant, β, one might think that it can be obtained from Eq. (16) by putting the left-hand side to zero. Unfortunately this is not so since the resulting value of β differs from the one observed in the numerics ($\beta \approx 0.26$) by the factor of almost 2. The reason for such a discrepancy is that the vanishing left-hand side of Eq. (16) generally represents the main terms of approximation ($\sim\Gamma$). As these terms have now canceled each other the universal temperature $\beta \sim 1$ is given by the balance of the next terms in the expansion that are of the order of unity. But it turns out that the neglected terms in the saddle-point approximation of the partition function being of the order $O(1/\Gamma)$ have nevertheless the contribution to its logarithmic derivative that are in the order of 1, which affects the value of the reciprocal temperature β. Therefore in principle one has to consider the next terms in the saddle-point approximation of (5). However we have opted for a different approach, namely, we have evaluated the logarithmic derivatives of the integral of $\exp(-\beta H_\gamma)$ numerically and the obtained solution of system (4) provided us a value of β, which fitted well the dependencies shown in Figs. 3(d)–3(e) [and also the intensity PDFs in Fig. 3(c)]. From a few separate runs (not shown) it appears also that not only the found universal value $\beta \approx 0.26$ is independent of Γ (when $\Gamma \gg 1$) but it is apparently independent of the material parameters $\lambda_{1,2}$ as well. The latter observation however requires further thorough verification.

C. Structure of final thermal state: General case

In the above section we have shown that if one ensures that the initial field components are locked in the same circular state of polarization for each waveguide, the reciprocal temperature is a universal numerical constant β, independent of the nonlinearity parameter Γ, which implies universal shape of the intensity and phase PDFs as well as the field

correlation functions. Let us now turn to the more general case when the initial field does not have any preferred state of polarization, i.e., although both TE and TM components have constant intensities A_0, B_0 ($A_0 + B_0 = 2$) their phases are always independent and uniformly distributed. It turns out that although the final thermal state is still given by the constrained minimum of the nonlinear energy corresponding to the circular polarization given by (14), the value of the temperature β is no longer a universal constant but depends on the nonlinearity. Also there is a drastic increase in the thermalization length, sometimes by up to three orders of magnitude. However we shall see below that in this general case the universality is not altogether lost. Rather it is the product $\beta\Gamma$ and not the temperature itself that is universal so instead of, e.g., universal field correlations as in the scalar case [13] one has *universal intensity distribution* $P(S_0)$ that is not affected by the value of the nonlinearity parameter Γ. The width of this distribution (i.e., the variance of the intensity fluctuations) is of the order of $\beta\Gamma \sim 1$. It turns out, however, that this new constant $\beta\Gamma$ is less universal than the temperature in the above-considered case of initial circular polarization inasmuch as it exhibits strong dependence on parameters $\lambda_{1,2}$.

We start by noticing that from Eq. (16) it follows that if the initial fields are not circularly polarized, i.e., $(h_0 - h_{\min}) \sim \Gamma \neq 0$ the reciprocal temperature β must necessarily be of the order of $\Gamma^{-1} \ll 1$ to balance the large terms in the left-hand side. Of course (16) was obtained in the saddle-point approximation, which strictly speaking no longer applies when the product $\beta\Gamma$ is in the order of unity. But as we shall presently see the relation $\beta\,\Gamma \sim 1$ also follows from the exact scaling dependence of the partition function valid beyond the saddle-point approximation. We will introduce the new notations for the rescaled temperature $\tilde{\beta} = \beta\Gamma$, and chemical potential $\tilde{\mu} = \mu/\Gamma$ that are now both in the order of 1. One can notice that in this regime since β scales as Γ^{-1} one can neglect the coupling energy, H_C [as well as the birefringence term $k(A_n - B_n)$] in (5), which simplifies the calculations somewhat. In fact this assumption is similar to the one used in Sec. III when we determined the infinite temperature boundary $\beta = 0$ with the only difference being that now one cannot expand $\exp(-\beta H_\gamma)$ in powers of the argument (since the latter is in the order of unity).

Regardless of the applicability of the saddle-point approximation the numerical simulations demonstrate that a single waveguide probability density function factorizes into the product of the PDF of the phase differences in the TE polarization, $P(\theta_n)$ and the PDF for the remaining three Jones parameters, $P(A_n, b_n, \delta_n)$, with:

$$P(\theta) = \frac{1}{2\pi \mathrm{I}_0(4\tilde{\beta}\tilde{\mu}/h_{\min})} \exp(-4\tilde{\beta}\tilde{\mu}\cos\theta/h_{\min}),$$
$$P(A, B, \delta) \propto e^{-\beta H_\gamma(A, B, \delta)}. \tag{18}$$

Note that since $h_{\min} \sim \Gamma$ while $\tilde{\beta} \sim \tilde{\mu} \sim 1$ the phase PDF $P(\theta_n)$ is (i) nonuniversal (i.e., depends on the nonlinearity) and (ii) approaches the uniform distribution as $\Gamma \to \infty$ [unlike the scalar case where it remains fixed, cf. Fig. 3(d)]. The intensity distribution is no longer narrow and has a finite width proportional to $\tilde{\beta}^{-1/2}$.

It is convenient at this stage to change variables from Jones to Stokes parameters $(A, B, \delta) \to (S_1, S_2, S_3)$ according to the

definition (11). The Jacobian of this transformation is equal to $(2S_0)^{-1}$ and the marginal PDF for the Stokes parameters has the form:

$$P(S_1, S_2, S_3) \propto S_0^{-1} \exp\left[-\tilde{\beta}\left(\frac{1}{2}S_1^2 + \frac{1+\lambda_1}{4}\left(S_2^2 + S_3^2\right)\right.\right.$$
$$\left.\left. + \frac{\lambda_2}{4}\left(S_2^2 - S_3^2\right) - \tilde{\mu}\,S_0\right)\right], \tag{19}$$

and $S_0 = \sqrt{S_1^2 + S_2^2 + S_3^2}$.

The marginal probability for the intensity as well as the overall normalization, the temperature and the chemical potential can be established by introducing the spherical coordinates on the Poincaré sphere of radius S_0 and integrating out the angular variables. The result reads:

$$P(S_0) = y_1^{-1}\, S_0\, e^{\tilde{\mu}\,\tilde{\beta}S_0 - \tilde{\beta}S_0^2/2} \int_0^1 \frac{dy}{\sqrt{1-y}}\, \mathrm{I}_0\left[\tilde{\beta}\frac{\lambda_2}{4}\,S_0^2\,y\right]$$
$$\times e^{\tilde{\beta}S_0^2(1-\lambda_1)y/4}. \tag{20}$$

Here y_1 is the normalization constant related to the overall partition function of Eq. (5) via $\mathcal{Z} = (2\pi y_1)^N$ (as both linear coupling and the birefringence can be neglected). The full analytical form of the intensity PDF $P(S_0)$ cannot be obtained in the general case but one can see that it is (i) asymmetric and (ii) has a Gaussian asymptote $P(S_0) \propto S_0^{-1} \exp[\tilde{\beta}\tilde{\mu}S_0 - (\tilde{\beta}/4)(1 + \lambda_1 - \lambda_2)S_0^2]$ as $S_0 \to \infty$. Again the condition $|\lambda_1 - \lambda_2| < 1$ ensures the convergence. The normalized inverse temperature $\tilde{\beta}$ and the chemical potential $\tilde{\mu}$ can then be determined in a standard way from Eqs. (4). By closely inspecting the normalization integral y_1 one can infer that it has a self-similar form (and so does the partition function):

$$y_1(\tilde{\beta}, \tilde{\mu}) = \frac{1}{\tilde{\beta}}\, F\left(\tilde{\mu}\sqrt{\tilde{\beta}}\right),$$

where the explicit form of the function F is given below. Plugging this ansatz into (4) and excluding the terms containing the derivative of $\ln F$ we obtain the relation between the (normalized) chemical potential and the temperature:

$$\tilde{\mu} = \frac{h_0}{\Gamma} - \frac{1}{\tilde{\beta}}. \tag{21}$$

The temperature is then to be obtained from a single transcendental equation:

$$F'(f_0\,\tilde{\beta}^{1/2} - 1/\tilde{\beta}^{1/2}) = 2\tilde{\beta}^{1/2}\, F(f_0\,\tilde{\beta}^{1/2} - 1/\tilde{\beta}^{1/2}), \tag{22}$$

where the derivative is taken with respect to the argument, and $f_0 = h_0/\Gamma$ is determined via Eq. (15). As for the function $F(x)$, it is generally only available in quadratures as

$$F(x) = \frac{1}{2}\int_0^\infty dz\, e^{x\sqrt{z} - z/2} \int_0^1 \frac{dy}{\sqrt{1-y}}\, \mathrm{I}_0\left[\frac{\lambda_2}{4}\,zy\right] e^{(1-\lambda_1)yz/4}.$$

The solution of the transcendental equation (22), $\tilde{\beta} = \beta\Gamma$ is universal inasmuch as it does not depend on the value of nonlinearity, Γ. For given choice of $\lambda_{1,2}$ it depends only on the average normalized initial energy f_0 and as long as the latter is not very close to its minimal value $h_{\min}/\Gamma$ (which corresponds to the universal regime considered in the previous section) the solution always exists an is in the order of 1.

In our numerical simulations shown in Fig. 4 we have chosen a symmetric initial condition with completely randomised
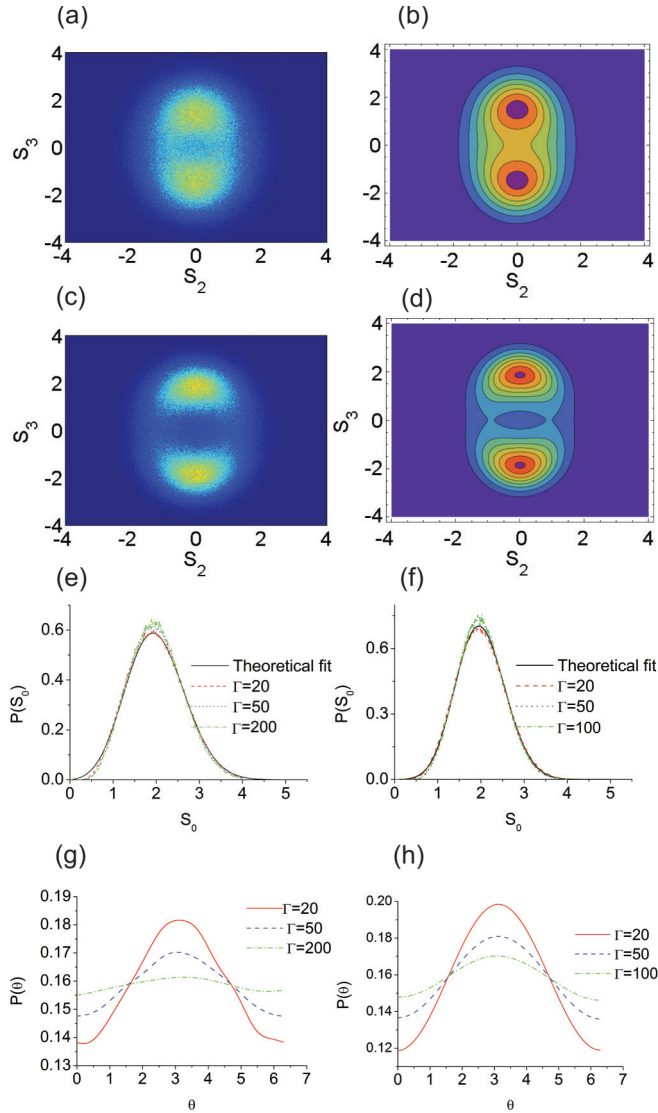
FIG. 4. (Color online) The statistics of the final thermal state in the limit of large nonlinearity and no initial phase correlation. (a)–(d) The marginal distribution of Stokes parameters $P(S_2, S_3)$ for anisotropic (a), (b) and isotropic (d), (e) cases respectively. (a) and (c) show numerical simulations while (b) and (d) demonstrate theoretically calculated profiles. (e), (f) The statistics of the intensity for anisotropic (e) and isotropic (f) cases for different values of nonlinearity parameter Γ. The phase difference distribution $P(\theta)$ for anisotropic (g) and isotropic (h) cases for different values of the nonlinearity parameter.

phases: $A_0 = B_0 = 1$, $\langle \cos \delta \rangle = 0$. From (15) this corresponds to the value $f_0 = 1 + \lambda_1$. For the anisotropic case (AlGaAs) the numerical solution of (22) yields $\tilde{\beta} = 2.32$ while for the isotropic one (fused silica) one gets $\tilde{\beta} = 4.41$. We have found that unlike in the case of universal correlations observed in Sec. IV B (or the scalar case), the thermalization length now depends on the temperature (which is directly proportional to the nonlinearity). Generally, the higher the temperature, the longer it takes for the system to reach equilibrium. For the highest level on nonlinearity achieved in our simulations ($\Gamma = 200$) the thermalization length was of the order $z_{th} \sim 500$ units of coupling length. This is of course an extreme limit

and lower values of nonlinearity (i.e., temperature) produce lower values of the thermalization length (typically ~ 10). In Figs. 4(a)–4(d) we can see marginal distribution of two Stokes parameters $P(S_2, S_3)$ for $\Gamma = 50$. Theoretical profiles [Figs. 4(b) and 4(d)] were obtained by integrating out S_1 in the (normalized) distribution (19) while the parameters $\tilde{\beta}$ and $\tilde{\mu}$ were determined via (21) and (22). One can observe a good agreement between theory and numerics [23]. The distribution is bimodal but the peaks (corresponding to circular clockwise and anticlockwise polarization) are quite wide (the width is in the order of $\tilde{\beta}^{-1/2} \sim 1$). Figures 4(e) and 4(f) compare theoretical prediction given by (20) with the numerics. One can see that for different values of the nonlinearity Γ the product $\tilde{\beta} = \beta \Gamma$ [and hence the intensity PDF $P(S_0)$] indeed remains the same and instead of universal correlation functions observed in the scalar case or the case of circular initial polarization [Fig. 3(c)] one now has universal intensity distribution as given by (20). The opposite case occurs with regards to the angle distribution $P(\theta)$, Figs. 4(g) and 4(h). Whereas in the the scalar case and in the case of circular polarization [Fig. 3(d)] this PDF remains invariant, now, according to the theoretical prediction (18) the distribution becomes closer and closer to uniform as Γ increases since $h_{\min}$ scales as Γ while $\tilde{\beta}$ and $\tilde{\mu}$ remain constant. In some respect the results of [Figs. 4(e)–4(h)] are complementary to those for the circular polarization, Figs. 3(c) and 3(d), and the scalar case (Fig. 2 of Ref. [13]) with the angular and intensity distribution trading places.

The universal nature of the intensity PDF $P(S_0)$ can also be explained by means of energy conservation arguments (similar to those put forward in Ref. [13] for the scalar case). In the limit of large Γ most of the energy concentrates in the nonlinear part $\mathcal{H}_\gamma$, which is a sum of on-site components given by (6). We can introduce the standard deviation for intensities as $\sigma^2(z) = \langle (A + B)^2 \rangle - 4$ (where angular brackets denote both averaging over the initial conditions and over waveguides) and rewrite the nonlinear energy in terms of σ^2. Then the following average integral of motion exists:

$$\Gamma \sigma^2(z) + \Gamma \langle (\lambda_1 - 1)AB + \lambda_2 AB \cos(2\delta) \rangle + O(1) = \text{const}, \tag{23}$$

where the term $O(1)$ collectively denotes the contribution from the linear coupling and birefringence. The value of the constant is determined by the initial conditions and since initially all waveguides have the same amplitudes, $A_0, B_0, (A_0 + B_0 = 2)$ and random uncorrelated phases the initial variance vanishes $\sigma^2(0) = 0$ and one obtains the following relation:

$$\sigma^2(z) = -\langle (\lambda_1 - 1)(AB - A_0 B_0) + \lambda_2 AB \cos(2\delta) \rangle + O(1/\Gamma).$$

This means that beyond the thermalization distance z_{th} the first term in the right-hand side becomes a stationary Γ-independent constant of the order of unity and so does the variance of the intensity distribution as clearly seen in Fig. 4(e). The situation changes however if one starts from the locked circular polarization state for each waveguide as in Sec. IV B. Then the first term in the right-hand side of the above changes to $-\langle (\lambda_1 - 1)(AB - 1) + \lambda_2 AB[1 + \cos(2\delta)] \rangle$. But the distribution in this case centres narrowly on the extremal values $A_* = B_* = 1$, $\delta_* = \pm \pi/2$ so that the average vanishes and in this regime the variance scales inversely proportional to Γ as is indeed seen in Fig. 3(c).
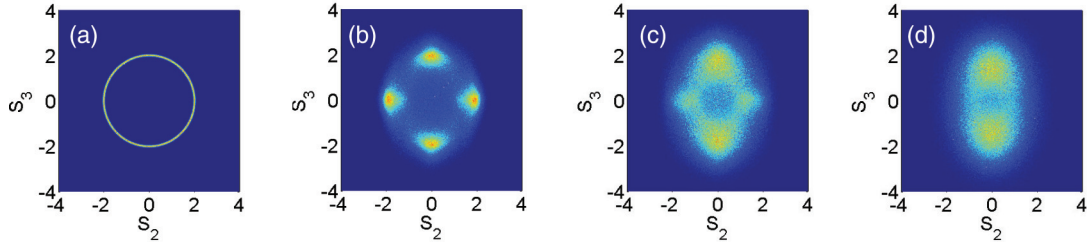
FIG. 5. (Color online) The evolution of the marginal PDF $P(S_2, S_3)$ with distance. Initial amplitudes are equal $A_0 = B_0 = 1$ and all the phases are uniformly distributed and uncorrelated. (a) The initial uniform distribution on a circle $z = 0$ (b) a four-modal distribution is formed initially $z = 1$ and is later gradually destroyed (c) at $z = 9.4$ and the final two-modal Gibbs state sets in (d) ($z = 50$).

Finally it is instructive to see how the distribution of the polarization state $P(S_2, S_3)$ evolves with distance z, passing from the initial isotropic state through the integrable regime described in Sec. IV A and eventually relaxing towards the equilibrium Gibbs distribution. The successive snapshots of such evolution are given in Fig. 5 for the anisotropic case and $\Gamma = 50$. One can see all successive stages: from the uniform distribution on a circle of radius 2 onto the four-modal distribution that occurs in a transient anticontinuum regime when the energy exchange between the waveguides is still negligible [cf. Fig. 2(a)] and finally towards the thermal equilibrium state, which is achieved when the weak linear interwaveguide coupling provides uniform mixing between all the degrees of freedom.

V. CONCLUSIONS

In this paper we have considered field thermalization in discrete birefringent waveguide systems. We have defined the exact boundary in the space of the integrals of motion separating the thermal phase with positive temperature from that corresponding to localized excitations (discrete breathers). We have shown that in the limit of high nonlinearity depending on the choice of the initial conditions the marginal PDF for the second two Stokes parameters, $P(S_2, S_3)$ relaxes either to a universal, broad, bimodal distribution (with the maxima corresponding to clockwise or anticlockwise circular polarization) or (if initially all waveguides are locked in the same circular polarization) a narrow-peaked one, corresponding to the thermal fluctuations around the initial state. In both cases either the effective temperature constant β is universal (i.e., does not depend on the nonlinearity parameter Γ) or, generally, the product $\beta\Gamma$ remains fixed as one increases the value of the nonlinearity.

Also in the limit of strong nonlinearity before reaching the final thermal state the system passes through the state where its dynamic is integrable, corresponding to the anticontinuum limit of the nonlinear polarization dynamics. In this regime the probability distribution of finding a system in a certain polarization state has a complicated multimodal structure. As the thermalization sets in the different modes of the distribution gradually disappear leaving only the two maxima corresponding to the circularly polarized state achieving the global minima of the nonlinear coupling energy subject to fixed total power.

As a possible continuation of this research one may suggest a more detailed study of the material dependence of the final thermal state (here only two specific choices of $\lambda_{1,2}$ were

considered) as well as a full statistical analysis of the breather region of the phase space as done by Rumpf [7,8] for the scalar case.

APPENDIX: DYNAMICS OF STOKES PARAMETERS IN THE ANTICONTINUUM LIMIT

In this section we recall the exact results for the dynamics of the system of Stokes parameters (12) (see Refs. [18,21]). By differentiating the first Eq. in (12) and using the other two we arrive at the following system:

$$S_1'' + \omega^2 S_1 = 0 \quad \text{Isotropic case}$$
$$S_1'' + \alpha S_1 + \beta S_1^3 = 0 \quad \text{Anisotropic case} \tag{A1}$$

with

$$\alpha = -\frac{\Gamma^2}{2} \left[(3\lambda_1 - 2)(\lambda_1 - 2)S_1^2(0) + \lambda_1(2 - 3\lambda_1)S_2^2(0) \right.$$
$$\left. + \lambda_1(\lambda_1 - 2)S_3^2(0) \right]$$
$$\beta = \frac{1}{2}\Gamma^2(2 - \lambda_1)(2 - 3\lambda_1), \quad \omega = 2\gamma(1 - \lambda_1)S_3(0)$$

with the initial values $S_1(0)$, and $S_1'(0) = 2\Gamma\lambda_2 S_2(0)S_3(0)$. The coefficient β should not be confused with the reciprocal temperature as defined in other sections. Thus in the isotropic case we get an equation for a harmonic oscillator while in the anisotropic case the equation is that of the nonlinear Duffing oscillator.

The solution in the isotropic case is simple:

$$S_1(z) = S_1(0) \cos(\omega z) + S_2(0) \sin(\omega z). \tag{A2}$$

In the anisotropic case the form of the solution depends on the sign of the coefficient β as well as the Duffing oscillator energy:

$$E = \frac{S_1'^2}{2} + \frac{\alpha}{2}S_1^2 + \frac{\beta}{4}S_1^4 = \frac{1}{8}\Gamma^2 \left[-S_1^2(-2 + \lambda_1) + 2\lambda_1 S_2^2 \right]$$
$$\times \left[2\lambda_1 S_3^2 + S_1^2(-2 + 3\lambda_1) \right]. \tag{A3}$$

Here we will only consider the case when $2/3 < \lambda_1 \leqslant 2$ so that we have simultaneously $\beta < 0$, $\alpha > 0$, and $E > 0$. For AlGaAs we have $\lambda_1 = 1$, which corresponds exactly to

this regime. Moreover in this case the potential energy has two symmetric maxima with the values $E_* = -\alpha^2/4\beta$ and one can prove that for any choice of the initial conditions $\{S_1(0), S_2(0), S_3(0)\}$ the difference

$$E - E_* = \frac{\Gamma^2\lambda_1^2\left[S_3^2(0)(-2+\lambda_1) + S_2^2(0)(-2+3\lambda_1)\right]^2}{8\left(4 - 8\lambda_1 + 3\lambda_1^2\right)} < 0.$$

The effective particle oscillates in the valley between the two maxima. The solution is expressed via Jacobi elliptic function:

$$S_1(z) = S_1^* \operatorname{sn}\left[F(S_1(0)/S_1^*, \kappa) + \sqrt{\frac{\alpha}{1+\kappa^2}}z, \kappa\right],$$

$$S_1^* = \sqrt{\frac{\alpha}{|\beta|}}\frac{\sqrt{2}\kappa}{\sqrt{1+\kappa^2}}, \quad \kappa = \sqrt{\frac{E_*}{E}} - \sqrt{\frac{E_*}{E} - 1} \leqslant 1. \tag{A4}$$

Here S_1^* is the maximal reachable amplitude of the oscillations, $\operatorname{sn}(x, \kappa)$ is the elliptic sine function and $F(x, \kappa)$ is the incomplete elliptic integral of the first kind:

$$F(x, \kappa) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-\kappa^2 t^2)}}.$$

Now using either the two integrals of motion of the main system, i.e., Eqs. (13) or integrating the system (12) directly, one can restore the remaining Stokes parameters, $S_2(z)$ and $S_3(z)$. In the anticontinuum limit the averaging over the initial conditions amounts to averaging either (A2) or (A4) for S_1 and using the integrals of motion to obtain the marginal PDF $P(S_2, S_3)$.

[1] E. Fermi, J. Pasta, and S. Ulam, Los Alamos Technical Report LA-1940 (1955) (unpublished) [Lect. Appl. Math. **15**, 143 (1974)].

[2] J. L. Lebowitz, H. A. Rose, and E. R. Speer, J. Stat. Phys. **50**, 657 (1988).

[3] R. Jordan, B. Turkington, and C. L. Zirbel, Physica D **137**, 353 (2000).

[4] K. O. Rasmussen, T. Cretegny, P. G. Kevrekidis, and N. Gronbech-Jensen, Phys. Rev. Lett. **84**, 3740 (2000).

[5] B. Rumpf and A. C. Newell, Physica D **184**, 162 (2003); B. Rumpf, Phys. Rev. E **69**, 016618 (2004).

[6] M. Johansson and K. O. Rasmussen, Phys. Rev. E **70**, 066610 (2004).

[7] B. Rumpf, Phys. Rev. E **77**, 036606 (2008).

[8] B. Rumpf, Physica D **238**, 2067 (2009).

[9] V. E. Zakharov, V. S. L'vov, and G. Falkovich, *Kolmogorov Spectra of Turbulence* (Springer, Berlin, 1992).

[10] A. C. Newell, S. Nazarenko, and L. Biven, Physica D **152**, 520 (2001).

[11] P. Aschieri, J. Garnier, C. Michel, V. Doya, and A. Picozzi, Phys. Rev. A **83**, 033838 (2011); C. Michel, M. Haelterman, P. Suret, S. Randoux, R. Kaiser, and A. Picozzi, *ibid.* **84**, 033848 (2011).

[12] F. Lederer, G. I. Stegeman, D. N. Christodoulides, G. Assanto, M. Segev, and Y. Silberberg, Phys. Rep. **463**, 1 (2008); S. Flach and A. V. Gorbach, *ibid.* **467**, 1 (2008).

[13] Y. Silberberg, Y. Lahini, Y. Bromberg, E. Small, and R. Morandotti, Phys. Rev. Lett. **102**, 233904 (2009).

[14] E. Small, R. Pugatch, and Y. Silberberg, Phys. Rev. A **83**, 013806 (2011).

[15] J. Meier, J. Hudock, D. Christodoulides, G. Stegeman, Y. Silberberg, R. Morandotti, and J. S. Aitchison, Phys. Rev. Lett. **91**, 143907 (2003).

[16] R. L. Horne, P. G. Kevrekidis, and N. Whitaker, Phys. Rev. E **73**, 066601 (2006).

[17] D. Mandelik, H. S. Eisenberg, Y. Silberberg, R. Morandotti, and J. S. Aitchison, Phys. Rev. Lett. **90**, 253902 (2003).

[18] N. Akhmediev, A. Buryak, and J. M. Soto-Crespo, Opt. Commun. **112**, 278 (1994); E. A. Ostrovskaya, N. N. Akhmediev, G. I. Stegeman, J. U. Kang, and J. S. Aitchison, J. Opt. Soc. Am. B **14**, 880 (1997).

[19] N. N. Akhmediev and A. Ankiewicz, *Solitons. Nonlinear Pulses and Beams* (Chapman & Hall, London, 1997).

[20] A. Yariv and P. Yeh, *Photonics. Optical Electronics in Modern Communications*, 2nd ed. (Oxford University Press, Oxford, 2007).

[21] B. Daino, G. Gregory, and S. Wabnitz, J. Appl. Phys. **58**, 4512 (1985); G. Gregori and S. Wabnitz, Phys. Rev. Lett. **56**, 600 (1986).

[22] D. C. Mattis, *The Theory of Magnetism II, Thermodynamics and Statistical Mechanics* (Springer-Verlag, Berlin, 1985).

[23] Because of the Jacobian $(2S_0)^{-1}$ there is a narrow integrable singularity at zero both in (19) and in the marginal PDF $P(S_2, S_3)$, but the width of this singularity is much smaller than the computational grid and it does not feature prominently in our simulations.