



**HAL**  
open science

## Entropy of partially polarized light and application to statistical processing techniques

Philippe Réfrégier, François Goudail, Pierre Chavel, Ari Friberg

► **To cite this version:**

Philippe Réfrégier, François Goudail, Pierre Chavel, Ari Friberg. Entropy of partially polarized light and application to statistical processing techniques. *Journal of the Optical Society of America. A Optics, Image Science, and Vision*, 2004, 21 (11), pp.2124-2134. hal-00079555

**HAL Id: hal-00079555**

**<https://hal.science/hal-00079555>**

Submitted on 6 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Entropy of partially polarized light and application to statistical processing techniques

Philippe Réfrégier and François Goudail

*Physics and Image Processing Group, Fresnel Institute, Unité Mixte de Recherche 6133, Ecole Généraliste d'Ingénieurs de Marseille, Domaine Universitaire de Saint-Jérôme, 13397 Marseille Cedex 20, France*

Pierre Chavel

*Laboratoire Charles Fabry de l'Institut d'Optique, Ecole Supérieure d'Optique, Université Paris Sud et Centre National de la Recherche Scientifique, Batiment 503, Centre Scientifique d'Orsay, 91403 Orsay Cedex, France*

Ari Friberg

*Kungliga Tekniska Högskolan, Electrum 229, 164 40 Kista, Sweden*

Received February 25, 2004; revised manuscript received June 10, 2004; accepted June 17, 2004

We have analyzed entropy properties of coherent and partially polarized light in an arbitrary number of spatial dimensions. We show that for Gaussian fields, the Shannon entropy is a simple function of the intensity and of the Barakat degree of polarization. In particular, we provide a probabilistic interpretation of this definition of the degree of polarization. Using information theory results, we also deduce some physical properties of partially polarized light such as additivity of the entropy and depolarization effects induced by mixing partially polarized states of light. Finally, we demonstrate that entropy measures can play an important role in segmentation and detection tasks. © 2004 Optical Society of America

*OCIS codes:* 260.5430, 030.0030, 030.4280, 100.1000.

## 1. INTRODUCTION AND BACKGROUND

### A. Introduction to the Problem

The polarization of light is a fundamental concept in optics and is of interest in several active technological fields such as imagery, telecommunications, medicine, and instrumentation. Many applications use active coherent illumination and analyze the variation of the polarization state. In this paper, we analyze the entropy properties of Gaussian partially polarized light and discuss some of its applications to statistical image processing. Our main goal is to show that the entropy of Gaussian, partially polarized light is a central concept in engineering applications and can be easily generalized to an arbitrary number of spatial dimensions.

From a practical point of view, the case of polarimetric images is of prime importance because of the present interest of the image science community in active polarimetric imagery.<sup>1-4</sup> Polarimetric imagers are powerful tools for improving the information content of images, since they can measure the whole polarimetric state, i.e., the Stokes parameters, of the light backscattered from each pixel of the scene.<sup>4-7</sup> Polarimetric images can thus reveal differences between objects that have different polarizing properties, which differences would not appear in conventional intensity images.<sup>2,8</sup>

One can note that in a coherent polarimetric image, the state of polarization of each speckle can be perfectly de-

finied, since it corresponds to a given realization of the random coherent electric field. However, this polarization state may vary from one speckle to another, so that from a spatial or ensemble-averaging point of view, the light may be totally or partially depolarized.

Because of the speckle effect, when one uses coherent light polarimetric imaging necessitates specific processing techniques. We shall discuss two kinds of problems: estimation and decision tasks. To illustrate the first case, object segmentation with a polygonal snake technique will be discussed. This technique can be useful for target recognition or shape measurement, for example. We shall then analyze detection of inhomogeneities (e.g., target, defect, or edge detection) with the generalized likelihood ratio test (GLRT). We shall show that this test is related in a simple way to the estimated mixing entropy of the different polarization states. Furthermore, we will see that these entropy-based algorithms depend on the data only through the determinant of the estimated coherency matrix, which can be written simply as a function of the estimated mean intensity and of the estimated Barakat degree of polarization.

Although we shall consider mainly the case of polarimetric imagery in the following, the generalization of these concepts to temporal signals is straightforward as long as a Gaussian, partially polarized light model is used.

To provide a rigorous description of the properties that

we have mentioned, various definitions have to be clarified; this will be the purpose of Subsection 1.B. In Section 2 we will introduce the definition of the entropy of partially polarized light and discuss its relation to different information-theoretic measures. We will then be able to discuss additivity properties of entropy and to analyze entropy increase induced by mixing several states of polarized light in different ways. These results will be useful in Section 3, where two examples of image processing problems will be discussed. We will conclude in Section 4 and propose some perspectives.

## B. Background on Gaussian Optical Fields

Classically, in dimension two the electric field  $\mathbf{E} = (E_1, E_2)^T$  of a speckled, coherent electromagnetic radiation is represented by a two-dimensional (2D) complex random vector<sup>9</sup> with covariance matrix  $\Gamma$ , defined by<sup>10</sup>

$$\Gamma = \begin{bmatrix} \langle |E_1|^2 \rangle & \langle E_1 E_2^* \rangle \\ \langle E_2 E_1^* \rangle & \langle |E_2|^2 \rangle \end{bmatrix} = \begin{bmatrix} \mu_1 & \rho \\ \rho^* & \mu_2 \end{bmatrix}, \quad (1)$$

where the angle brackets denote ensemble averaging. Assuming suitable ergodicity properties, this ensemble averaging can correspond to different types of physical averaging depending on the application considered. For coherent images, it can correspond to spatial average or to averaging over different diffusers of the same kind. For slowly varying temporal signals, it can correspond to temporal averages. This covariance matrix  $\Gamma$  defines the 2D polarization state of the electric vector and is frequently called the ‘‘coherency matrix.’’<sup>9</sup> Since it is Hermitian,  $\Gamma$  is completely defined by four real-valued parameters: the average intensities  $\mu_1$ ,  $\mu_2$ , and the complex-correlation coefficient  $\rho$ .

The well-known fully developed speckle model<sup>11,12</sup> leads to an electric field vector represented by a 2D circular Gaussian random vector with a probability-density function (PDF) equal to

$$P_{\Gamma}(\mathbf{E}) = \frac{1}{\pi^2 \det(\Gamma)} \exp(-\mathbf{E}^{\dagger} \Gamma^{-1} \mathbf{E}), \quad (2)$$

where the symbol  $\dagger$  denotes the transpose conjugate and  $\det(\Gamma)$  is the determinant of  $\Gamma$ .

One can generalize the previous discussion to  $d$ -dimensional electric vectors. In this case, the PDF is

$$P_{\Gamma}(\mathbf{E}) = \frac{1}{\pi^d \det(\Gamma)} \exp(-\mathbf{E}^{\dagger} \Gamma^{-1} \mathbf{E}). \quad (3)$$

Considering dimensions greater than two can be useful for nonplane optical waves.<sup>13</sup> One can also consider the situation in which the reflected electric field is analyzed when a scene is illuminated with two orthogonal polarization states. For each incident polarization state one measures the backscattered field in two orthogonal polarizations states, thus leading to a four-component vector. In some cases and because of symmetry properties, one can consider that there are only three independent measures that form a three-dimensional random electric vector.<sup>14</sup> One can also define  $d$ -dimensional electric fields when simultaneous measurements at different frequencies are performed. For example, let us consider

that a partially polarized light with an electric field  $\mathbf{E}(t)\exp(i\omega t)$  is split into two waves, one of them undergoing a frequency shift  $\delta\omega$ . Assume that the polarization states of these two parts are independently modified so that one obtains the electric field  $\mathbf{E}^{(1)}(t)\exp(i\omega t) + \mathbf{E}^{(2)}(t)\exp[i(\omega + \delta\omega)t]$ . It can be useful to define a four-dimensional vector  $\mathbf{E}(t) = [E_1^{(1)}(t), E_2^{(1)}(t), E_1^{(2)}(t), E_2^{(2)}(t)]^T$  and to study its entropy and degree of polarization. Indeed, postulating a four-dimensional vector is the standard approach in mathematics to considering possible correlations between the different measurements.

## 2. ENTROPY OF PARTIALLY POLARIZED LIGHT AND INFORMATION-THEORETIC MEASURES

### A. Shannon Entropy in Dimension $d$ for Gaussian Light

In this paper we consider light with a Gaussian PDF, which will also be denoted Gaussian light. We shall come back to this assumption in Subsection 2.C.

In dimension  $d$  the Shannon entropy is defined by<sup>15</sup>  $S = -\int \ln[P_{\Gamma}(\mathbf{E})]P_{\Gamma}(\mathbf{E})d\mathbf{E}$ , where  $\int d\mathbf{E}$  stands for complex  $d$ -dimensional integration. In the 2D case one thus has

$$S_2 = \log[\pi^2 \det(\Gamma)] + \int (\mathbf{E}^{\dagger} \Gamma^{-1} \mathbf{E}) P_{\Gamma}(\mathbf{E}) d\mathbf{E},$$

but since  $\int (\mathbf{E}^{\dagger} \Gamma^{-1} \mathbf{E}) P_{\Gamma}(\mathbf{E}) d\mathbf{E} = 2$ , one gets  $S_2 = \log[\pi^2 e^2 \det(\Gamma)]$ . The degree of polarization is defined as  $\mathcal{P}_2 = [1 - 4 \det(\Gamma)/\text{tr}(\Gamma)^2]^{1/2}$ ,<sup>9</sup> where  $\text{tr}(\Gamma)$  is the trace of  $\Gamma$ , which represents the total intensity  $I_0$ . One thus has  $S_2 = \log[\pi^2 e^2 I_0^2 (1 - \mathcal{P}_2^2)/4]$ .

In dimension  $d$  the PDF of the electric field is given by Eq. (3), and one can easily obtain the following property<sup>6</sup>:

*Property A:* The entropy of partially polarized light in dimension  $d$  is

$$S_d = \log[\pi^d e^d \det(\Gamma)]. \quad (4)$$

Defining the dimensionless parameter  $\delta = \det(\Gamma)/\text{tr}(\Gamma)^d$ , one gets  $S_d = \log(\pi^d e^d I_0^d \delta)$ . One can develop this expression to get three terms:

$$S_d = d \log(\pi I_0) + \log(\delta) + d. \quad (5)$$

The first term on the right-hand side is representative of the dependence of the entropy on the disorder in each channel, which is a function of the energy. The second one is linked to disorder due to decorrelation of the various components of the electric field vector.

Different alternatives exist<sup>13,16,17</sup> for the definition of the degree of polarization in dimension  $d$  and are still a subject of studies and discussion. We will show in this paper that the Barakat definition<sup>17</sup> is linked in a simple way to the statistical properties of the Gaussian vector field. Let us first write the entropy as a function of the eigenvalues of the covariance matrix  $\lambda_1, \lambda_2, \dots, \lambda_d$ :

$$S_d = d + d \log(\pi I_0) + \log \left[ \frac{\lambda_1 \lambda_2 \dots \lambda_d}{(\lambda_1 + \lambda_2 + \dots + \lambda_d)^d} \right].$$

By analogy with the 2D case, let us consider a  $d$ -dimensional definition of the degree of polarization:  $1 - \mathcal{P}_d^2 = u \lambda_1 \lambda_2 \dots \lambda_d / (\lambda_1 + \lambda_2 + \dots + \lambda_d)^d$ . Since one wishes that  $\mathcal{P}_d = 0$  when  $\lambda_1 = \lambda_2 = \dots = \lambda_d$ , it is clear that one needs to choose  $u = d^d$ . Since  $\det[\Gamma]$  and  $\text{tr}[\Gamma]$

are invariant by any unitary transform,  $\mathcal{P}_d$  is also invariant by any unitary transform. At that level, no simple choice of  $n$  has been proposed and one may arbitrarily choose  $n = 2$ . When  $d = 3$  one obtains the Barakat definition<sup>17</sup> of the degree of polarization  $\mathcal{P}_3^2 = 1 - 27 \det(\Gamma)/\text{tr}(\Gamma)^3$ ; this denomination will also be used for arbitrary dimensions. In dimension  $d$  the entropy is thus obtained with the following corollary of Property A.

*Corollary A.1:* The expression of the entropy of partially polarized light in dimension  $d$  as a function of the intensity and of the Barakat degree of polarization is

$$S_d = d \left[ 1 + \log\left(\frac{\pi}{d}\right) \right] + d \log(I_0) + \log(1 - \mathcal{P}_d^2). \quad (6)$$

In dimension three the entropy is thus  $S_3 = 3[1 + \log(\pi/3)] + 3 \log(I_0) + \log(1 - \mathcal{P}_3^2)$ .

### B. Additive Property of Entropy

Let us consider the three-dimensional vector field model in the case where the actual vector field lies within a plane, i.e., is (2D). One eigenvalue is then equal to 0 (let us assume that it is  $\lambda_3$ ) and entropy goes to  $-\infty$ , which can be a problem in practical cases. This divergence is due to the fact that one considers a PDF instead of a probability law. In other words, this divergence is due to the representation of the electric field with continuous variables. In reality, several arguments lead to a nonzero lower bound  $\epsilon$  for  $\lambda_3$ . One can, for example, consider quantum indetermination for the spatial frequency vectors  $k$  or residual noise on detectors. One thus gets

$$S_3 \approx 3 + 3 \log(\pi I_0) + \log\left[\frac{\epsilon}{(\lambda_1 + \lambda_2)}\right] + \log\left[\frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}\right],$$

and finally  $S_3 \approx 1 + \log(\pi\epsilon) + S_2$ , which corresponds to the additive property of the entropy of independent subsystems.

Let us discuss this property a little further. Let us assume that we have a pure, vertically polarized light with intensity  $\alpha_1 I_0$ . Then its entropy is  $S_1 = 1 + \log(\pi\alpha_1 I_0)$ . Now if one considers a pure, horizontally polarized light with intensity  $\alpha_2 I_0$ , one obtains the entropy  $S'_1 = 1 + \log(\pi\alpha_2 I_0)$ . Since both cases are assumed independent, the entropy of the union of both systems should be the sum of the entropies; thus  $S_2 = 2 + 2 \log(\pi I_0) + \log(\alpha_1 \alpha_2)$ . One can note that the union of both systems is nothing but a 2D partially polarized light. If we add a third orthogonal component, we obtain  $S_3 = 3 + 3 \log(\pi I_0) + \log(\alpha_1 \alpha_2 \alpha_3)$ . In other words, one finds the classical result that for independent systems, our definition of the global entropy corresponds to an additive combination of the entropies of each subsystem, which is not always the case with other definitions of the entropy.<sup>18</sup> In particular, when  $\lambda_3$  reaches its inferior limit value  $\epsilon$ , the entropy reduces to  $S_3 \approx 1 + \log(\pi\epsilon) + S_2$ .

### C. Probabilistic Interpretation

Many different definitions of entropy have been proposed.<sup>19</sup> One can thus wonder why the Shannon definition should be preferable to others, such as the Renyi entropy, for example, which has also been found useful for image processing problems.<sup>20</sup> In fact, the Shannon entropy has a simple probabilistic interpretation that is the origin of its great success in information theory.<sup>15</sup> Let us consider an experiment in which we observe  $N$  independent measurements of the electric field of a partially polarized light. Let us also assume that the measurements are performed with a precision  $q$  with  $q^{2d} \ll \det(\Gamma)$ , so that one can consider PDFs instead of the probability laws obtained after quantification; or, in other words, one can consider integration instead of summation.

We thus obtain a series of  $N$  complex vectors from which we can determine the number of times  $N(\mathbf{E})$  the value  $\mathbf{E}$  is observed with precision  $q$  on each component. The ratio  $N(\mathbf{E})/N$  is the frequency  $F(\mathbf{E})$  of the realization of the value  $\mathbf{E}$  in the observed series of length  $N$ . We will write this frequency  $F(\mathbf{E})$  with the density  $Q(\mathbf{E})$  defined by  $Q(\mathbf{E}) = F(\mathbf{E})/q^{2d}$ . It is well known that  $Q(\mathbf{E})$  converges to the probability density  $P(\mathbf{E})$  when  $N \rightarrow +\infty$ . Let us address two questions when  $N$  becomes large but is still finite:

1. What is the probability of observing the frequency  $F(\mathbf{E}) = Q(\mathbf{E})q^{2d}$ ?
2. How many different series exist with  $Q(\mathbf{E}) = P(\mathbf{E})$ ?

The answer to the first question is easily obtained if we introduce the Kullback measure:

$$K(Q\|P) = \int \log\left[\frac{Q(\mathbf{E})}{P(\mathbf{E})}\right] Q(\mathbf{E}) d\mathbf{E}. \quad (7)$$

When  $N$  is large and  $q^{2d} \ll \det(\Gamma)$ , the probability  $P_N(Q\|P)$  of observing the frequency  $F(\mathbf{E})$  is<sup>21</sup>

$$P_N(Q\|P) \approx \exp[-NK(Q\|P)]. \quad (8)$$

Since  $K(Q\|P)$  is positive if  $Q \neq P$ <sup>15</sup> and  $K(Q\|P) = 0$  if  $Q = P$ , Eq. (8) is of course in agreement with the fact that when  $N \rightarrow +\infty$ , series for which  $Q = P$  have a probability of unity of being observed, while other series have a negligible probability.

A direct calculus<sup>22</sup> allows one to show that the Kullback measure between two Gaussian distributions  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  is given by

$$\begin{aligned} K(P_{\Gamma_a}\|P_{\Gamma_b}) &= \int \log\left[\frac{P_{\Gamma_a}(\mathbf{E})}{P_{\Gamma_b}(\mathbf{E})}\right] P_{\Gamma_a}(\mathbf{E}) d\mathbf{E} \\ &= \log\left[\frac{\det(\Gamma_b)}{\det(\Gamma_a)}\right] + \text{tr}(\Gamma_a \Gamma_b^{-1}) - d, \end{aligned} \quad (9)$$

When  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  correspond to PDFs of states of light with the same intensities, and when  $P_{\Gamma_b}(\mathbf{E})$  is the PDF of a totally depolarized light, one gets

$$K(P_{\Gamma_a}\|P_{\Gamma_b}) = -\log(1 - \mathcal{P}_a^2) \quad (10)$$

where  $\mathcal{P}_a$  is the Barakat degree of polarization of the light of PDF  $P_{\Gamma_a}(\mathbf{E})$ . One thus obtains the following property:

*Property B:* In a series of  $N$  measurements, the probability of observing the frequency  $P_{\Gamma_a}(\mathbf{E})q^{2d}$  that corresponds to a light with intensity  $I_0$  and Barakat degree of polarization  $\mathcal{P}_a$  when the light is in fact totally depolarized with intensity  $I_0$  is

$$P_N(P_{\Gamma_a}|P_{\Gamma_b}) = (1 - \mathcal{P}_a^2)^N. \quad (11)$$

Let us now address the second question, which pertains to series for which  $Q(\mathbf{E}) = P(\mathbf{E})$ . When  $N$  is large and  $q^{2d} \ll \det(\Gamma)$ , the number  $\mathcal{M}_N$  of different series is simply<sup>21</sup>  $\mathcal{M}_N \approx \exp\{N[S_d - 2d \log(q)]\}$ , where  $S_d$  is the entropy of  $P(\mathbf{E})$ . This interpretation makes it clear that Shannon entropy is a measure of the potential disorder of partially polarized light. Let us compare two states of light with the same intensity  $I_0$  but with respective degrees of polarization  $\mathcal{P}_a$  and  $\mathcal{P}_b$ . One gets from Eq. (6) the following property:

*Property C:* Considering two lights of the same intensity and with degrees of polarization  $\mathcal{P}_a$  and  $\mathcal{P}_b$ , one has

$$\frac{\mathcal{M}_N^{(a)}}{\mathcal{M}_N^{(b)}} = \left( \frac{1 - \mathcal{P}_a^2}{1 - \mathcal{P}_b^2} \right)^N, \quad (12)$$

where  $\mathcal{M}_N^{(a)}$  (respectively  $\mathcal{M}_N^{(b)}$ ) is the number of different series of length  $N$  of the light of degree of polarization  $\mathcal{P}_a$  (resp.  $\mathcal{P}_b$ ) for which the frequency of observing  $\mathbf{E}$  is  $P_{\Gamma_a}(\mathbf{E})q^{2d}$  [resp.  $P_{\Gamma_b}(\mathbf{E})q^{2d}$ ] when  $q^{2d} \ll \det(\Gamma)$ .

Since to assign a number to  $M$  objects, one needs at least  $\log_2(M)$  bits (where  $\log_2$  is the base-2 logarithm), one finds that the minimum number of bits to encode  $N$  independent measurements of the electric vector  $\mathbf{E}$  will be  $\log_2(\mathcal{M}_N^{(a)})$ , and thus  $N[\tilde{S}_d - 2d \log_2(q)]$ , where  $\tilde{S}_d$  is the entropy defined with the base-2 logarithm. This well-known result in information theory can be rigorously proved and is the subject of one of the Shannon theorems.<sup>23</sup> Property C thus allows one to see that encoding  $N$  realizations of a partially polarized light with a degree of polarization equal to  $\mathcal{P}_d$  will require  $N \log_2(1 - \mathcal{P}_d^2)$  fewer bits than to encode  $N$  realizations of a totally unpolarized light with the same intensity.  $1 - \mathcal{P}_d^2$  is thus a measure of the information disorder of the light or, equivalently, the Barakat degree of polarization is a measure of the information order of the light.

In the previous sections we assumed that the PDF of the electric field of the light was Gaussian. This assumption was made following the initial arguments of Goodman and Dainty on the theory of speckle.<sup>12</sup> The previous discussion allows us to obtain a new insight into this assumption. We have seen that the number of different series  $\mathcal{M}_N$  that can be observed with a frequency of observing the electric vector  $\mathbf{E}$  equal to  $P_{\Gamma}(\mathbf{E})q^{2d}$  diverges as  $\exp\{N[S_d - 2d \log(q)]\}$ .<sup>21</sup> Thus for a given coherency matrix  $\Gamma$ , the PDF of the light that maximizes the entropy is the most complex one in the sense that it leads to the largest number of different possible series  $\mathcal{M}_N$  when  $N$  becomes large.

The problem at hand is thus to determine  $P_{\Gamma}(\mathbf{E})$  for which  $S = -\int \ln[P_{\Gamma}(\mathbf{E})]P_{\Gamma}(\mathbf{E})d\mathbf{E}$  is maximal with the constraints that  $\langle E_i^* E_j \rangle = \Gamma_{ij}$  and  $\int P_{\Gamma}(\mathbf{E})d\mathbf{E} = 1$ . The solution to this problem is well known<sup>24</sup> and in fact leads to the PDF of Eq. (3). The light that maximizes the en-

$$\mathbf{E}^{(\alpha)} = \begin{cases} \mathbf{E}^{(a)} & \text{with proba. } \alpha \\ \mathbf{E}^{(b)} & \text{with proba. } 1 - \alpha \end{cases}$$

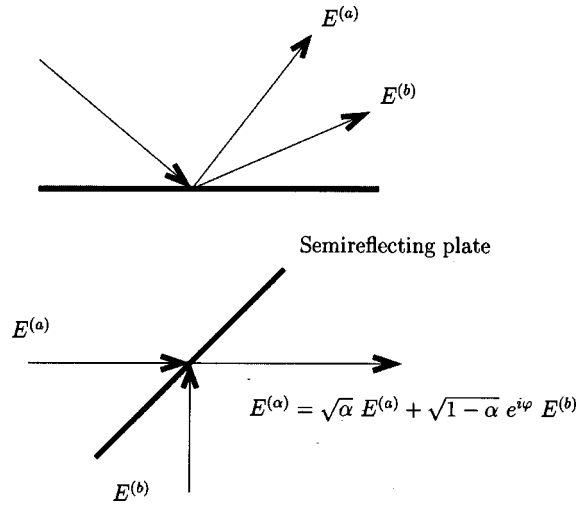


Fig. 1. Illustration of two different types of mixing: top, random choice between  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$ ; bottom additive mixing of  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$ .

tropy with a given coherency matrix thus corresponds to Gaussian light; in other words, the Gaussian assumption is equivalent to assuming maximal complexity from the entropic point of view.

#### D. Mixing Entropy and Chernoff Distance

We propose in this subsection to analyze the evolution of the entropy in two kinds of experiments involving mixing of Gaussian, partially polarized states of light. The first one corresponds to mixing due to random choice between Gaussian partially polarized states of light, the second to the addition of independent Gaussian partially polarized light. In particular, we will show that the mixing entropy of the addition of independent Gaussian partially polarized states of light is equal to the Chernoff distance, which is a well-known information measure in statistics. We will show in Section 3 that these results are useful in practical engineering problems.

Let us first consider a mixing that corresponds to disorder introduced by a random choice between electric fields  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  [see Fig. 1(a)]. More precisely let us assume that the measured electric field  $\mathbf{E}$  is  $\mathbf{E}^{(a)}$  with probability  $\alpha$  and  $\mathbf{E}^{(b)}$  with probability  $1 - \alpha$ . The covariance matrix of  $\mathbf{E}$  is  $\alpha\Gamma_a + (1 - \alpha)\Gamma_b$ , but  $\mathbf{E}$  is not a Gaussian field even if  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  are Gaussian. Since among random fields of covariance matrix  $\alpha\Gamma_a + (1 - \alpha)\Gamma_b$  Gaussian fields have the largest entropy, the entropy  $S_{\alpha}^R$  of  $\mathbf{E}$  satisfies the inequality

$$S_{\alpha}^R \leq S_{\alpha} = \log\{\pi^d e^d \det[\alpha\Gamma_a + (1 - \alpha)\Gamma_b]\}. \quad (13)$$

Noting that the entropy of  $\mathbf{E}^{(a)}$  (respectively  $\mathbf{E}^{(b)}$ ) is  $S(a) = \log[\pi^d e^d \det(\Gamma_a)]$  (respectively  $S(b) = \log[\pi^d e^d \det(\Gamma_b)]$ ), one can define the mixing entropy as  $\Delta S_{\alpha}^R = S_{\alpha}^R - [\alpha S(a) + (1 - \alpha)S(b)]$ . Let us establish the following property:

*Property D:* The mixing entropy  $\Delta S_{\alpha}^R = S_{\alpha}^R - [\alpha S(a) + (1 - \alpha)S(b)]$  of a random choice between electric fields  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  is positive.

To demonstrate this inequality, let us introduce the PDF  $P_{\alpha}(\mathbf{E})$  of  $\mathbf{E}$ , which is equal to  $\alpha P_{\Gamma_a}(\mathbf{E}) + (1 - \alpha)P_{\Gamma_b}(\mathbf{E})$ , where  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  are the PDFs corresponding to the Gaussian fields  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$ . The entropy of  $\mathbf{E}$  is  $S_{\alpha}^R = -\int H[P_{\alpha}(\mathbf{E})]d\mathbf{E}$  where  $H[z] = z \log z$ .  $H[z]$  is convex, and thus  $-H[z]$  is concave. So one can deduce that

$$\begin{aligned} -H[\alpha P_{\Gamma_a}(\mathbf{E}) + (1 - \alpha)P_{\Gamma_b}(\mathbf{E})] \\ \geq -\alpha H[P_{\Gamma_a}(\mathbf{E})] - (1 - \alpha)H[P_{\Gamma_b}(\mathbf{E})]; \end{aligned}$$

thus  $S_{\alpha}^R \geq \alpha S(a) + (1 - \alpha)S(b)$ .

This property means that if a partially polarized light is diffused without energy absorption, its entropy cannot decrease, which corresponds to the following corollary:

*Corollary D.1:* If partially polarized light with Barakat degree of polarization  $\mathcal{P}_d$  is diffused without energy absorption, then the diffused light cannot have a higher Barakat degree of polarization  $\mathcal{P}'_d$  (i.e.,  $\mathcal{P}'_d \leq \mathcal{P}_d$ ).

Let us now consider the second example, in which two independent Gaussian electric fields  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  with respective PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  are merged so as to obtain  $\mathbf{E}^{(\alpha)} = (\alpha)^{1/2}\mathbf{E}^{(a)} + (1 - \alpha)^{1/2}\exp(i\varphi)\mathbf{E}^{(b)}$ , where  $\varphi$  accounts for a possible phase difference between the two terms [see Fig. 1(b)]. The entropy  $S_{\alpha}^A$  of this mixed field  $\mathbf{E}^{(\alpha)}$  is easily obtained since it is a Gaussian field with covariance matrix  $\alpha\Gamma_a + (1 - \alpha)\Gamma_b$ , and thus  $S_{\alpha}^A = \log\{\pi^d e^d \det[\alpha\Gamma_a + (1 - \alpha)\Gamma_b]\}$ . One can now establish the following property:

*Property E:* The mixing entropy  $\Delta S^A$  corresponding to

$$\mathbf{E}^{(\alpha)} = (\alpha)^{1/2}\mathbf{E}^{(a)} + (1 - \alpha)^{1/2}\exp(i\varphi)\mathbf{E}^{(b)},$$

where  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  are independent and have respective PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$ , is

$$\Delta S_{\alpha}^A = \log\left\{\frac{\det[\alpha\Gamma_a + (1 - \alpha)\Gamma_b]}{\det(\Gamma_a)^{\alpha}\det(\Gamma_b)^{(1-\alpha)}}\right\}. \quad (14)$$

One can see that if  $\Gamma_a = \Gamma_b$ , then  $\Delta S_{\alpha}^A = 0$ . It is now easy to show the following property:

*Property F:* The mixing entropy  $\Delta S_{\alpha}^A$  corresponding to

$$\mathbf{E}^{(\alpha)} = \sqrt{\alpha}\mathbf{E}^{(a)} + (1 - \alpha)^{1/2}\exp(i\varphi)\mathbf{E}^{(b)},$$

where  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  are independent and have respective PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$ , is nonnegative.

Indeed, one has  $S_{\alpha}^A \geq S_{\alpha}^R$  and thus  $\Delta S_{\alpha}^A \geq \Delta S_{\alpha}^R$ .

The particular case of this property when  $\det(\Gamma_a) = \det(\Gamma_b)$  shows that the entropy of the mixed field  $\mathbf{E}^{(\alpha)}$  is higher than the entropy of each individual field  $\mathbf{E}^{(a)}$  or  $\mathbf{E}^{(b)}$ , which leads to the following corollary:

*Corollary F.1:* When one adds independent Gaussian partially polarized lights with the same Barakat degree of polarization  $\mathcal{P}_d$ , the Barakat degree of polarization  $\mathcal{P}'_d$  of the mixed light cannot increase (i.e.,  $\mathcal{P}'_d \leq \mathcal{P}_d$ ).

The previous probabilistic interpretation of entropy is interesting for information processing applications. Let us consider a two-hypothesis testing (or detection) problem between two different partially polarized states of the light with PDF given by Eq. (3) but with different covari-

ance matrices  $\Gamma_a$  and  $\Gamma_b$ . It has been shown<sup>21</sup> that the probability of error can be approximated by the Chernoff bound. Indeed, for decision problems on series of  $N$  measurements, only some specific series will have nonnegligible influence on the probability of error. These correspond to series for which the frequency of realization of the electric field vector  $\mathbf{E}$  is equal to  $\mathbf{Q}^*(\mathbf{E})q^{2d}$  such that  $P_N(\mathbf{Q}^*|P_{\Gamma_a}) = P_N(\mathbf{Q}^*|P_{\Gamma_b})$ . Since  $P_N(\mathbf{Q}|P_{\Gamma_a}) \approx \exp[-NK(\mathbf{Q}|P_{\Gamma_a})]$  and  $P_N(\mathbf{Q}|P_{\Gamma_b}) \approx \exp[-NK(\mathbf{Q}|P_{\Gamma_b})]$ ,  $\mathbf{Q}^*$  is obtained by writing  $K(\mathbf{Q}^*|P_{\Gamma_a}) = K(\mathbf{Q}^*|P_{\Gamma_b})$ . It can be shown<sup>21</sup> that

$$\mathbf{Q}^*(\mathbf{E}) = \frac{[P_{\Gamma_a}(\mathbf{E})]^{1-s^*}[P_{\Gamma_b}(\mathbf{E})]^{s^*}}{\int [P_{\Gamma_a}(\mathbf{E})]^{1-s^*}[P_{\Gamma_b}(\mathbf{E})]^{s^*}d\mathbf{E}},$$

where  $s^*$  is the value of  $s$  that minimizes  $\int [P_{\Gamma_a}(\mathbf{E})]^{1-s}[P_{\Gamma_b}(\mathbf{E})]^s d\mathbf{E}$ . The Chernoff distance between the PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  is defined<sup>21</sup> as a function of  $s$  by

$$C(s) = -\log\left\{\int [P_{\Gamma_a}(\mathbf{E})]^{1-s}[P_{\Gamma_b}(\mathbf{E})]^s d\mathbf{E}\right\}.$$

It can be proved (see Appendix A) that  $C(s) \geq 0$ ,  $dC/ds(0) = K(P_{\Gamma_a}||P_{\Gamma_b})$ , and  $dC/ds(1) = -K(P_{\Gamma_b}||P_{\Gamma_a})$ . Furthermore,  $K(\mathbf{Q}^*|P_{\Gamma_a}) = K(\mathbf{Q}^*|P_{\Gamma_b}) = C(s^*)$ .

One can obtain (see Appendix B) a physical interpretation of  $C(s)$  as a function of the entropy of the different partially polarized states:

*Property G:* The Chernoff distance  $C(\alpha)$  between the Gaussian PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  is equal to the mixing entropy  $\Delta S_{\alpha}$  corresponding to  $\mathbf{E}^{(\alpha)} = (\alpha)^{1/2}\mathbf{E}^{(a)} + (1 - \alpha)^{1/2}\exp(i\varphi)\mathbf{E}^{(b)}$ , where  $\mathbf{E}^{(a)}$  and  $\mathbf{E}^{(b)}$  are independent and have respective PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$ .

Let us now analyze the particular case where partially polarized light with a degree of polarization  $\mathcal{P}$  is diffused into a new state of light so that a part  $\alpha$  of that light has been totally depolarized. In this case the final state of the light, with a degree of polarization  $\mathcal{P}(\alpha)$ , can be considered as an additive mixing of the initial state with totally depolarized light with coefficient  $\alpha$ . The entropy of this light can be denoted  $S_{\alpha}$ ; let  $S_i$  correspond to the entropy of the initial light. Let us assume that only a small amount of light is totally depolarized so that  $\alpha \ll 1$ . One has  $S_{\alpha} = \Delta S_{\alpha} + [\alpha S_u + (1 - \alpha)S_i]$ , where  $S_u$  is the entropy of the totally depolarized (or unpolarized) part of the light. Since  $\Delta S_{\alpha} = C(\alpha)$  and  $\alpha \ll 1$ , one has  $S_{\alpha} \approx \partial C/\partial \alpha(0)\alpha + [\alpha S_u + (1 - \alpha)S_i]$ . We have seen that  $\partial C/\partial \alpha(0) = K(P_{\Gamma_a}||P_{\Gamma_b})$ , where  $\Gamma_a$  is the coherency matrix of the initial light and  $\Gamma_b$  is the coherency matrix of the totally depolarized part of the diffused light with the same intensity  $I_0$  for both states. We have also seen, in Subsection 2.C, that  $K(P_{\Gamma_a}||P_{\Gamma_b}) = -\log(1 - \mathcal{P}^2)$ , where  $\mathcal{P}$  is the Barakat degree of polarization of the initial light. Furthermore since  $S_u = d[1 + \log(\pi/d)] + d \log(I_0)$  and  $S_i = d[1 + \log(\pi/d)] + d \log(I_0) + \log(1 - \mathcal{P}^2)$ , one gets

$$S_{\alpha} \approx S_i - 2\alpha \log(1 - \mathcal{P}^2), \quad (15)$$

which can also be written with the degrees of polarization as

$$\log[1 - \mathcal{P}(\alpha)^2] \simeq (1 - 2\alpha)\log(1 - \mathcal{P}^2), \quad (16)$$

which is a simple relation for the evolution of the Barakat degree of polarization.

### 3. APPLICATION TO STATISTICAL PROCESSING

In the present section we discuss statistical algorithms that can solve a variety of problems in coherent polarimetric images.

Automatic detection and recognition of objects by means of passive imaging systems, such as passive IR sensors, suffer from limitations, and it can be interesting to increase the contrast between targets and their backgrounds. Polarimetric imaging is one of the possible techniques to increase this contrast. Since our purpose in this paper is to illustrate applications of the previous concept of entropy of partially polarized light, and not to analyze the general problem of image processing and automatic target recognition, the reader may refer for example to Ref. 25 for a general introduction to this topic and to Ref. 26 and references therein for applications to polarization imaging.

We will define an image model and consider two processing tasks, segmentation and detection, which correspond, respectively, to an estimation and a decision problem.

#### A. Image Model and Maximum-Likelihood Processing

Image segmentation consists of partitioning an image into a number  $K$  of regions having homogenous properties. For the sake of simplicity, we will consider here only the case of  $K = 2$  regions. We thus assume that the considered image, or subimage,  $\mathbf{E}(x, y)$  with  $(x, y) \in [1, N_x] \times [1, N_y]$  can be divided into two homogeneous regions  $\Omega_a^\theta$  and  $\Omega_b^\theta$  modeled as follows:

$$\mathbf{E}(x, y) = \mathbf{E}_a(x, y)w_a^\theta(x, y) + \mathbf{E}_b(x, y)w_b^\theta(x, y) \quad (17)$$

where the  $w_u^\theta$  ( $u = a$  or  $b$ ) are the binary masks that define the regions  $\Omega_u^\theta$ , that is,  $w_u^\theta(x, y) = 1$  in region  $\Omega_u^\theta$  and 0 elsewhere. The polarimetric vectors representative of the electric field in each region  $\Omega_u^\theta$  are assumed to form a sample  $\mathbf{E}_u(x, y)$  distributed with a spatially uncorrelated, homogeneous, Gaussian circular PDF as defined in Eq. (3). Please note that for the sake of simplicity, we define the image  $\mathbf{E}(x, y)$  in terms of the electric field, although it is not the electric field, but quadratic values corresponding to the coefficients of the coherency matrix that are measured in optical polarimetric imaging systems. However, we will see in the following that the only needed information to process the images can be determined with the coefficients of the coherency matrix.  $\theta$  is the vector of parameters that depend on the task to be performed. For segmentation of a single object over a uniform background,  $\theta$  specifies the shape of the mask and can be defined in several ways depending on the model of the object contour. In the following a polygonal model will be considered and  $\theta$  will thus contain the coordinates of each node of the polygon. For target detection applications,  $\theta$  is a binary value that is 1 if a target of

shape  $w(x, y)$  is present and zero if it is absent. In other words, in hypothesis 1,  $w_a^1(x, y) = w(x, y)$  and  $w_b^1(x, y) = 1 - w(x, y)$ , and in hypothesis 0,  $w_a^0(x, y) = 0$  and  $w_b^0(x, y) = 1$ .

In the following statistical algorithms, it will be necessary to determine the likelihood of each region  $\Omega_u^\theta$  with  $u = a, b$  of  $N_u^\theta$  pixels. Its mathematical expression is

$$\begin{aligned} \mathcal{L}(\Gamma_u) &= -N_u^\theta \log \pi^d - N_u^\theta \log[\det(\Gamma_u)] \\ &\quad - \sum_{(x,y) \in \Omega_u^\theta} \mathbf{E}(x, y)^\dagger \Gamma_u^{-1} \mathbf{E}(x, y). \end{aligned} \quad (18)$$

In most applications the coherency matrices  $\Gamma_a$  and  $\Gamma_b$  are unknown. One thus estimates them from the sample in the maximum-likelihood sense and injects these estimates back into the expression of the log-likelihood. A nontrivial but classic calculus shows that the ML estimate of the coherency matrix in each region is simply the empirical covariance matrix of the sample<sup>27</sup>:

$$\widehat{\Gamma}_u^\theta = \frac{1}{N_u^\theta} \sum_{(x,y) \in \Omega_u^\theta} \mathbf{E}(x, y) \mathbf{E}(x, y)^\dagger. \quad (19)$$

This estimate of the covariance matrix is sufficient statistics for the problem at hand.<sup>28</sup> This means that it gathers all the information about the sample that is useful to estimate the polarization state. Injecting this estimate of  $\Gamma_u$  in Eq. (18) one obtains the following ‘‘profile’’ likelihood:

$$\widehat{\mathcal{L}}_u(\theta) = -N_u^\theta \log[\det(\widehat{\Gamma}_u^\theta)] - dN_u^\theta(1 + \log \pi). \quad (20)$$

Only the first member on the right-hand side of Eq. (20) is of interest, the second member being a constant. It is important to note that what is needed to determine the profile likelihood is just the determinant of the estimate of the coherency matrix of the sample, which can be computed from the measures provided by polarimetric imaging systems. In other words, if one denotes  $\widehat{\mathcal{P}}_d(u)$  as the estimated Barakat degree of polarization in region  $\Omega_u^\theta$  with  $\widehat{\mathcal{P}}_d^2(u) = 1 - d^d \det(\widehat{\Gamma}_u^\theta) / (\text{tr} \widehat{\Gamma}_u^\theta)^d$ , one obtains

$$\begin{aligned} \widehat{\mathcal{L}}_u(\theta) &= -N_u^\theta \left\{ d \log(\widehat{I}_u) + \log[1 - \widehat{\mathcal{P}}_d^2(u)] \right. \\ &\quad \left. + d \log\left(\frac{\pi e}{d}\right) \right\}, \end{aligned} \quad (21)$$

where  $\widehat{I}_u = \text{tr} \widehat{\Gamma}_u^\theta$  is the empirical mean value of the intensity in region  $\Omega_u^\theta$ . This result shows that the profile likelihood is a simple function of two empirical values  $\widehat{I}_u$  and  $\widehat{\mathcal{P}}_d(u)$ . One can also remark, with Eq. (6), that the likelihood is proportional to the entropy but with estimated values of the intensity and of the Barakat degree of polarization.

#### B. Minimum Entropy-Based Snake Segmentation

Let us first consider object segmentation applications. As seen previously, the parameter vector  $\theta$  parametrizes the shape of the object, and we will use a polygonal model for the contour of the object: The parameter vector  $\theta$  is

constituted of the coordinates of each node of the polygon. A technique based on the minimum description length (MDL) principle introduced by Rissanen<sup>29</sup> has been developed in Ref. 30 and consists of finding the shape parameter vector  $\theta$  that leads to a description of the image with a minimum number of bits. This technique is analogous to those developed in Refs. 31 and 32 and allows one to estimate the shape and the number of nodes of the polygon used to perform the segmentation. This approach leads to a segmentation technique based on the minimization of a criterion without free parameters. Since the image is composed of three parts (the target, the background, and the contour),  $\Delta$  is the sum of three terms: the length  $\Delta_a$  of the description of the target gray levels, the length  $\Delta_b$  of the description of the background gray levels, and the length  $\Delta_\theta$  of the description of the parameter  $\theta$  that defines the polygonal shape. Let us first provide an approximation of  $\Delta_\theta$ . The number of possible locations for one node is  $N$ . Thus for  $k$  nodes, the number of different locations is  $N^k$ , and we will consider that it is an approximation of the number of different polygons. The number of bits sufficient to describe the polygon (if they are assumed equally likely) is thus approximately  $\log_2(N^k)$  (where  $\log_2$  is the base-2 logarithm). As mentioned in subsection 2.C, the mean number of bits needed to describe  $N_a$  random variables distributed with PDF  $P_{\Gamma_a}(x)$  is  $\Delta_a \approx N_a[S_a - 2d \log_2(q)]$ , where  $S_a$  is the entropy in bits of the PDF and is given by  $S_a = -\int P_{\Gamma_a}(\mathbf{E}) \log_2[P_{\Gamma_a}(\mathbf{E})] d\mathbf{E}$ , and where  $q$  is the quantization precision. Since the contribution of  $\log_2(q)$  will consist only of adding a constant term to the description length, it will not be taken into account in the following. The mean number of bits sufficient to describe the background region is thus  $\Delta_b \approx N_b S_b$ , and the total description length (in nats; i.e., using natural logarithms) is

$$\Delta \approx N_a S_a + N_b S_b + k \log(N). \quad (22)$$

The entropy can be approximated by using the empirical mean instead of the statistical mean; one obtains  $N_a S_a \approx -\sum_{i \in \Omega_a^\theta} \log[P_{\Gamma_a}(\mathbf{E}_i)]$ , where the pixel set  $\Omega_a^\theta$  is defined by  $w_i = 1$ , i.e., it is the interior of the polygon  $\mathbf{w}$ . Furthermore when the coherency matrices  $\Gamma_a$  and  $\Gamma_b$  are unknown, one can estimate them from the sample in the maximum-likelihood sense as discussed previously. It is thus easy to see that an approximation of  $-N_a S_a - N_b S_b$  is given by the profile log-likelihood  $\hat{\mathcal{L}}_a(\theta) + \hat{\mathcal{L}}_b(\theta)$  of the hypothesis that, in the image, the shape of the target is defined by the parameters  $\theta$ . This is nothing but the statistical expression that was minimized in Ref. 33. One can thus see that the MDL principle leads to the minimization of

$$\Delta_{\text{app}} = N_a^\theta \log\{\hat{I}_a^d[1 - \hat{\mathcal{P}}_d^2(a)]\} + N_b^\theta \log\{\hat{I}_b^d[1 - \hat{\mathcal{P}}_d^2(b)]\} + k \log(N) - Cst, \quad (23)$$

where  $Cst = dN \log(e\pi/d)$ . It can be noted that since it is based on the estimation of the entropy of the image, this segmentation technique mainly relies on two statistics: the empirical Barakat degrees of polarization  $\hat{\mathcal{P}}_d^2(u)$  and the mean intensities  $\hat{I}_u$  estimated in each region.

The optimal value of shape parameters  $\theta^{\text{opt}}$  is thus obtained by simultaneously determining the values of  $k$  and  $\theta$  that minimize  $\Delta_{\text{app}}$ . This double optimization problem is nontrivial, and the adopted strategy may have strong influence on the quality and relevance of the application of the MDL principle. The simplest strategy will consist of determining the segmentation  $\theta^{(k)}$  that optimizes the

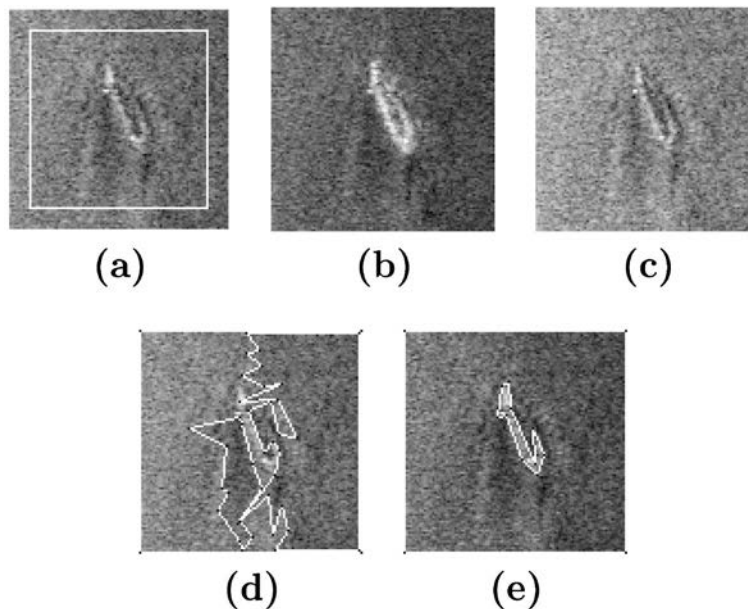


Fig. 2. Segmentation of a polarimetric image extracted from an image acquired with the NASA-JPL AIRSAR system: (a), (b), (c) modulus squares of each channel of the polarimetric image; (d) segmentation result on the intensity channel (a) with the MDL polygonal snake (a gamma PDF has been assumed); (e) segmentation result on the 3-complex channel polarimetric image with the MDL polygonal snake. For both segmentation examples, the initial contour was the white square represented in (a).



MDL criterion  $\Delta_{\text{app}}$  for different fixed values of  $k$  and then selecting the value of  $k$  that leads to the minimum value of  $\Delta_{\text{app}}$ .

An efficient technique has been proposed in Ref. 30, and it can be generalized to polarimetric data. We have used it to segment a polarimetric synthetic-aperture radar image in Fig. 2. Such images consist of three channels ( $d = 3$ ) with complex values, in which the noise due to speckle is distributed in first approximation with a Gaussian circular PDF as in Eq. (3).<sup>34</sup> We have represented in Fig. 2 the modulus squares of the three channels and the result of the segmentation of the object of interest in the center with the snake. We have applied the snake to a single intensity channel [see Fig. 2(d)] and to the fully polarimetric, 3-complex channel image [see Fig. 2(e)]. It can be seen in this example that using the polarimetric information can help in distinguishing the object of interest from the background.

### C. Maximum-Likelihood-Based Detection

In image processing applications, one often needs to detect the presence of a target at any position in an image: This is a combined detection–localization problem. In practice, the scene can be scanned with a binary mask  $F$ , for example a square region of dimensions  $M_x \times M_y$  that is larger than the mask  $w$  defining the target, to detect–localize. Let  $\bar{w}$  denote the complementary of  $w$  in  $F$ . The generalized likelihood ratio test (GLRT) is a standard technique for detecting features such as targets, inhomogeneities, and edges. For each position of the mask, the GLRT consists of computing the ratio of the profile likelihoods of the following two hypotheses:

- $H_1$ : There is a target in  $\mathbf{F}$ , and the samples in  $w$  and  $\bar{w}$  have different coherency matrices  $\Gamma_a$  and  $\Gamma_b$ .
- $H_0$ : There is no target (only background) in  $\mathbf{F}$ , and the samples in  $w$  and  $\bar{w}$  have the same statistical distribution.

It can be shown that the expression of the GLRT is<sup>35</sup>

$$\hat{\mathcal{R}} = -N_a \log[\det(\hat{\Gamma}_a)] - N_b \log[\det(\hat{\Gamma}_b)] + N_F \log[\det(\hat{\Gamma}_F)], \quad (24)$$

where  $N_F = N_a + N_b$  and where  $\hat{\Gamma}_a$ ,  $\hat{\Gamma}_b$ , and  $\hat{\Gamma}_F$  are the estimates of the covariance matrices in  $w$ ,  $\bar{w}$ , and  $F$ . Detection is performed by comparing the value of the test statistic  $\hat{\mathcal{R}}$  with a threshold for deciding if the target is present or not at the current location.

Since  $\hat{\Gamma}_F = N_a/N_F \hat{\Gamma}_a + N_b/N_F \hat{\Gamma}_b$  one can write

$$\hat{\mathcal{R}} = (N_a + N_b) \log \left\{ \frac{\det[\alpha \hat{\Gamma}_a + (1 - \alpha) \hat{\Gamma}_b]}{\det(\hat{\Gamma}_a)^\alpha \det(\hat{\Gamma}_b)^{(1-\alpha)}} \right\}, \quad (25)$$

with  $\alpha = N_a/(N_a + N_b)$ .  $\hat{\mathcal{R}}$  can thus be interpreted as the empirical mixing entropy  $\Delta \hat{S}_\alpha$  of  $\Delta S_\alpha$  defined in Eq. (14).  $\hat{\mathcal{R}}$  is thus also related to the Chernoff distance  $C(\alpha)$  between the PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$  through Property E. Since the profile log-likelihoods of the gray levels in these regions are given by Eq. (21), one finally gets

$$\begin{aligned} \hat{\mathcal{R}} &= -N_a d \log[\hat{I}_a] - N_b d \log[\hat{I}_b] + N_F d \log[\hat{I}_F] \\ &\quad - N_a \log[1 - \hat{\mathcal{P}}_d^2(a)] - N_b \log[1 - \hat{\mathcal{P}}_d^2(b)] \\ &\quad + N_F \log[1 - \hat{\mathcal{P}}_d^2(F)], \end{aligned} \quad (26)$$

where  $\hat{I}_u = \text{tr} \hat{\Gamma}_u^\theta$  and  $\hat{\mathcal{P}}_d^2(F) = 1 - d^d \det(\hat{\Gamma}_F^\theta) / (\text{tr} \hat{\Gamma}_F^\theta)^d$ . Thus here again the detection technique is based mainly on two statistics: the empirical Barakat degree of polarization  $\hat{\mathcal{P}}_d^2(u)$  and the mean intensities  $\hat{I}_u$  estimated in each region.

In practice it is useful to characterize the detection performance as a function of the contrast between the target and the background. This contrast obviously depends on the difference between the statistical distributions of the gray levels of the two regions. There are several ways to express this difference; one can use, for example, the Kulback–Leibler measure defined above [see Eq. (7)]. Through the Stein’s Lemma,<sup>21</sup> the Kulback–Leibler measure describes the asymptotic behavior (for large  $N$ ) of the probability of detection or of false alarm. However, the distributions  $P_{\mathbf{p}^a}(x)$  and  $P_{\mathbf{p}^b}(x)$  do not play symmetric roles in the Kulback–Leibler measure, which may be an undesirable property for a contrast parameter. We will thus consider the Kullback divergence, which is a symmetrized version of the Kulback–Leibler measure and is defined as

$$\mathcal{D} = K[P_{\mathbf{p}^a} \| P_{\mathbf{p}^b}] + K[P_{\mathbf{p}^b} \| P_{\mathbf{p}^a}]. \quad (27)$$

We have chosen dimension  $d = 2$ , and we have considered many different configurations of  $\Gamma_a$ ,  $\Gamma_b$ ,  $N_a$ , and  $\alpha = N_a/(N_a + N_b)$ . For each configuration we have estimated the detection performance in terms of the AUC, which is the area under the receiver operating characteristic.<sup>36</sup> The AUC varies between 0.5 and 1, and it increases with the detection performance: AUC = 0.5 corresponds to the worst possible detection performance (random choice) and 1 to the best possible one (probability of detection of 1 whatever the probability of false alarm). We have then plotted the AUC corresponding to each configuration of  $\Gamma_a$ ,  $\Gamma_b$ ,  $N_a$ , and  $\alpha$  as a function of the corresponding Kullback divergence multiplied by the number of pixels of the target; that is,  $N_a \mathcal{D}$ . It can be seen in Fig. 3(a) that the relation between the AUC and  $N_a \mathcal{D}$  obviously depends on  $N_a$  and on  $\alpha$ . For a given value of  $N_a \mathcal{D}$ , the AUC is larger both when  $N_a$  is larger and when  $\alpha$  is smaller (that is,  $N_b$  becomes larger). Since different values of the AUC can be obtained with the same value of  $N_a \mathcal{D}$ , the Kullback divergence does not represent the contrast between the target and the background.

Now we can use the fact that the GLRT is proportional to an estimation of the Chernoff distance between the PDFs  $P_{\Gamma_a}(\mathbf{E})$  and  $P_{\Gamma_b}(\mathbf{E})$ , as seen in Eq. (25). Thus in first approximation, the detection performance of the GLRT should depend mainly on the actual Chernoff distance, or, more precisely, on  $(N_a + N_b)C(\alpha)$ . This conjecture is verified in Fig. 3(b), where we have represented the AUC for the same configurations as in Fig. 3(a) as a function of  $(N_a + N_b)C(\alpha)$ . It can be seen that all the points gather around a master curve. In other words, when one knows the value of  $(N_a + N_b)C(\alpha)$ , one can approximately deduce the detection performance that will

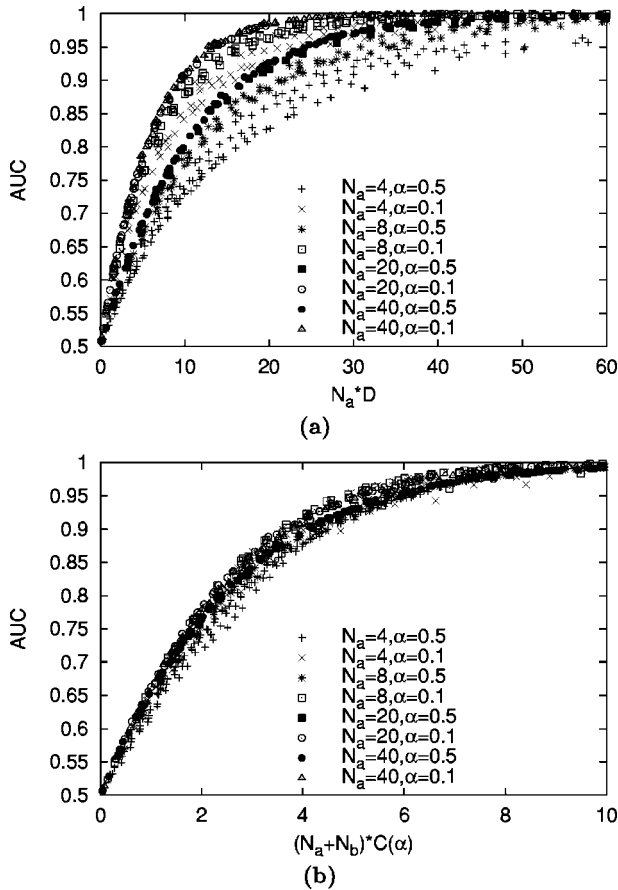


Fig. 3. AUC in function of (a)  $N_a D$  and (b)  $(N_a + N_b)C(\alpha)$ , where  $C(\alpha)$  is the Chernoff distance and  $\alpha = N_a/(N_a + N_b)$ . Each point corresponds to a different configuration of parameters ( $\Gamma_a, \Gamma_b, N_a, \alpha$ ). The AUC for each configuration has been estimated on  $10^4$  Monte Carlo trials.

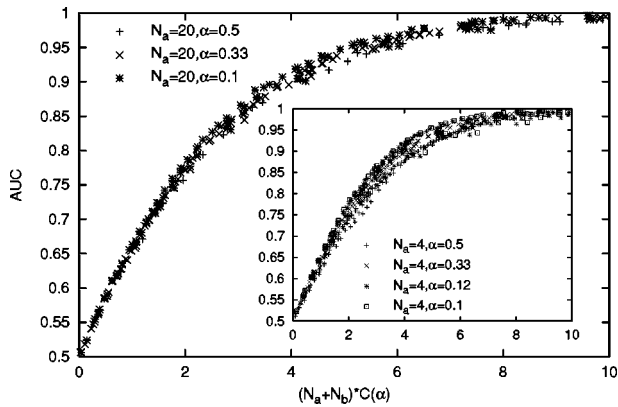


Fig. 4. AUC in function of  $(N_a + N_b)C(\alpha)$  for different values of  $\alpha$  and  $N_a = 20$ . Inset, same results for  $N_a = 4$ .

be obtained. This result is, of course, quite interesting in practice, although the relation is not completely bijective, since the GLRT is only an approximation of the Chernoff distance based on the data. The approximation is better when  $N_a$  and  $N_b$  are larger. This is verified in Fig. 4, where we have considered the cases of  $N_a = 20$  for different values of  $\alpha$  and of  $N_a = 4$  (see inset). It can be seen

that the relation between the AUC and  $(N_a + N_b)C(\alpha)$  is better described by a bijective relation for the larger target.

#### 4. CONCLUSION

In this paper we have analyzed entropy properties of coherent and partially polarized light in an arbitrary number of spatial dimensions. We have shown that the Shannon entropy is a simple function of the intensity and of the Barakat degree of polarization in the case of Gaussian light, which corresponds to the maximum complexity, i.e., which leads to the maximum number of different measurement series. We have provided probabilistic and information-theoretic interpretations of the entropy and thus of the Barakat definition of the degree of polarization. We have deduced physical properties of partially polarized light, such as additivity of the entropy and depolarization effects of mixing partially polarized states of light. Finally we have demonstrated how the entropy measure can be applied to segmentation and detection tasks.

Further developments can be envisaged. In particular it would be interesting to study how one can establish links with image processing techniques based on the Mueller matrix. It would also be useful to analyze the relations of the present study with other definitions of the entropy<sup>14</sup> and of the degree of polarization.<sup>13</sup>

#### APPENDIX A

Let us introduce  $U(s) = \int P_{\Gamma_a}(\mathbf{E})^{1-s} P_{\Gamma_b}(\mathbf{E})^s d\mathbf{E}$ . One has

$$\frac{d}{ds} U(s) = \int \log \left[ \frac{P_{\Gamma_b}(\mathbf{E})}{P_{\Gamma_a}(\mathbf{E})} \right] P_{\Gamma_a}(\mathbf{E})^{1-s} P_{\Gamma_b}(\mathbf{E})^s d\mathbf{E};$$

thus

$$\frac{d}{ds} U(0) = \int \log \left[ \frac{P_{\Gamma_b}(\mathbf{E})}{P_{\Gamma_a}(\mathbf{E})} \right] P_{\Gamma_a}(\mathbf{E}) d\mathbf{E} = -K(P_{\Gamma_a} \| P_{\Gamma_b}),$$

$$\frac{d}{ds} U(1) = \int \log \left[ \frac{P_{\Gamma_b}(\mathbf{E})}{P_{\Gamma_a}(\mathbf{E})} \right] P_{\Gamma_b}(\mathbf{E}) d\mathbf{E} = K(P_{\Gamma_b} \| P_{\Gamma_a}).$$

Furthermore

$$\frac{d}{ds} C(s) = -\frac{d}{ds} \ln U(s) = -\frac{1}{U(s)} \frac{d}{ds} U(s),$$

and  $U(0) = U(1) = 1$ ; thus  $d/ds C(0) = K(P_{\Gamma_a} \| P_{\Gamma_b})$  and  $d/ds C(1) = -K(P_{\Gamma_b} \| P_{\Gamma_a})$ . One can also obtain

$$\frac{d^2}{ds^2} U(s) = \int \log \left[ \frac{P_{\Gamma_b}(\mathbf{E})}{P_{\Gamma_a}(\mathbf{E})} \right]^2 P_{\Gamma_a}(\mathbf{E})^{1-s} P_{\Gamma_b}(\mathbf{E})^s d\mathbf{E},$$

which shows that  $(d^2/ds^2)U(s) \geq 0$ ; one thus sees that  $U(s)$  is convex on  $[0, 1]$ . Since  $U(0) = U(1) = 1$  and  $U(s)$  is convex on  $[0, 1]$ , one has  $U(s) \leq 1$  and thus  $C(s) \geq 0$  since  $C(s) = -\log[U(s)]$ .

## APPENDIX B

One has

$$P_{\Gamma_a}(\mathbf{E}) = \frac{1}{\pi^d \det(\Gamma_a)} \exp(-\mathbf{E}^\dagger \Gamma_a^{-1} \mathbf{E}),$$

$$P_{\Gamma_b}(\mathbf{E}) = \frac{1}{\pi^d \det(\Gamma_b)} \exp(-\mathbf{E}^\dagger \Gamma_b^{-1} \mathbf{E}).$$

The mixing entropy is

$$\Delta S_\alpha = \log \left\{ \frac{\det[\alpha \Gamma_a + (1 - \alpha) \Gamma_b]}{\det(\Gamma_a)^\alpha \det(\Gamma_b)^{(1-\alpha)}} \right\}.$$

Let us determine the Chernoff distance  $C(s) = -\log\{ \int [P_{\Gamma_a}(\mathbf{E})]^{1-s} [P_{\Gamma_b}(\mathbf{E})]^s d\mathbf{E} \}$ . If one defines  $U(s) = \int [P_{\Gamma_a}(\mathbf{E})]^{1-s} [P_{\Gamma_b}(\mathbf{E})]^s d\mathbf{E}$ , one has

$$U(s) = (1/A) \int \exp\{-\mathbf{E}^\dagger [(1-s)\Gamma_a^{-1} + s\Gamma_b^{-1}] \mathbf{E}\} d\mathbf{E}$$

with  $A = \pi^d \det(\Gamma_a)^{1-s} \det(\Gamma_b)^s$ . One can easily obtain

$$U(s) = \frac{\det[(1-s)\Gamma_a^{-1} + s\Gamma_b^{-1}]^{-1}}{\det(\Gamma_a)^{1-s} \det(\Gamma_b)^s}.$$

Using standard properties of determinants, one gets

$$U(s) = \frac{\det(\Gamma_a)^s \det(\Gamma_b)^{1-s}}{\det[(1-s)\Gamma_b + s\Gamma_a]},$$

and thus, since  $C(s) = -\log[U(s)]$ ,

$$C(s) = \log \left\{ \frac{\det[(1-s)\Gamma_b + s\Gamma_a]}{\det(\Gamma_a)^s \det(\Gamma_b)^{1-s}} \right\},$$

which proves that  $C(\alpha) = \Delta S_\alpha$ .

## ACKNOWLEDGMENTS

Philippe Réfrégier and François Goudail thank the members of the Phyti team, and Ari Friberg thanks the Swedish Research Council, for their support.

Corresponding author Philippe Réfrégier may be reached by phone, 33-4-91-28-8809, fax, 33-4-91-28-8201, or e-mail, philippe.refregier@fresnel.fr.

## REFERENCES

1. M. Floch, G. Le Brun, C. Kieleck, J. Cariou, and J. Lotrian, "Polarimetric considerations to optimize lidar detection of immersed targets," *Pure Appl. Opt.* **7**, 1327–1340 (1998).
2. S. Breugnot and Ph. Clémenceau, "Modeling and perfor-

- mances of a polarization active imager at  $\lambda=806$  nm," in *Laser Radar Technology and Applications IV*, G. W. Kamerman and C. Werner, eds., Proc. SPIE **3707**, 449–460 (1999).
3. A. Gleckler and A. Gelbart, "Multiple-slit streak tube imaging lidar MS-STIL applications," in *Laser Radar Technology and Applications V*, G. W. Kamerman, U. N. Singh, C. H. Werner, and V. V. Molebny, eds., Proc. SPIE **4035**, 266–278 (2000).
4. L. B. Wolff, "Polarization camera for computer vision with a beam splitter," *J. Opt. Soc. Am. A* **11**, 2935–2945 (1994).
5. J. E. Solomon, "Polarization imaging," *Appl. Opt.* **20**, 1537–1544 (1981).
6. W. G. Egan, W. R. Johnson, and V. S. Whitehead, "Terrestrial polarization imagery obtained from the space shuttle: characterization and interpretation," *Appl. Opt.* **30**, 435–442 (1991).
7. J. L. Pezzaniti and R. A. Chipman, "Mueller matrix imaging polarimetry," *Opt. Eng.* **34**, 1558–1568 (1995).
8. J. S. Tyo, M. P. Rowe, E. N. Pugh, and N. Engheta, "Target detection in optical scattering media by polarization-difference imaging," *Appl. Opt.* **35**, 1855–1870 (1996).
9. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985), pp. 116–156.
10. C. Brosseau, *Fundamentals of Polarized Light—A Statistical Approach* (Wiley, New York, 1998), pp. 138–164.
11. Ref. 9, pp. 237–285.
12. J. C. Dainty, *Laser Speckle and Related Phenomena* (Springer-Verlag, Heidelberg, Germany, 1975).
13. T. Setälä, M. Kaivola, and A. T. Friberg, "Degree of polarization in near fields of thermal sources: effects of surface waves," *Phys. Rev. Lett.* **88**, 123902 (2002).
14. R. S. Cloude and E. Pottier, "Concept of polarization entropy in optical scattering," *Opt. Eng.* **34**, 1599–1610 (1995).
15. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991), pp. 12–49.
16. J. C. Samson, "Descriptions of the polarization states of vector processes: applications to ULF magnetic fields," *Geophys. J. R. Astron. Soc.* **34**, 403–419 (1973).
17. R. Barakat, "N-fold polarization measures and associated thermodynamic entropy of N partially coherent pencils of radiation," *Opt. Acta* **30**, 1171–1182 (1983).
18. Ref. 10, pp. 165–175.
19. M. D. Esteban and D. A. Morales, "A summary of entropy statistics," *Kybernetika* **31**, 337–346 (1995).
20. R. Baraniuk, P. Flandrin, and O. Michel, "Measuring time frequency information content using the Renyi entropies," *IEEE Trans. Inf. Theory* **47**, 1391–1409 (2001).
21. Ref. 15, pp. 279–335.
22. C. W. Therrien, *Decision Estimation and Classification* (Wiley, New York, 1989), pp. 139–155.
23. C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
24. Ref. 15, pp. 266–278.
25. A. Firooz and A. Sadjadi, "Passive infrared automatic target recognition," in *Image Recognition and Classification: Algorithm, System and Applications*, B. Javidi, ed., (Marcel Dekker, New York, 2002), pp. 37–60.
26. A. F. Sadjadi and C. S. L. Chun, "Automatic detection of small objects from their infrared state-of-polarization vectors," *Opt. Lett.* **28**, 531–533 (2003).
27. R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
28. T. S. Ferguson, *Mathematical Statistics, a Decision Theoretic Approach* (Academic, New York, 1967), pp. 112–119.
29. J. Rissanen, *Stochastic Complexity in Statistical Inquiry* (World Scientific, Singapore, 1989).
30. O. Ruch and Ph. Réfrégier, "Minimal-complexity segmentation with a polygonal snake adapted to different optical noise models," *Opt. Lett.* **41**, 977–979 (2001).
31. S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 884–900 (1996).

32. M. Figueiredo, J. Leitão, and A. K. Jain, "Unsupervised contour representation and estimation using B-splines and a minimum description length criterion," *IEEE Trans. Image Process.* **9**, 1075–1087 (2000).
33. C. Chesnaud, Ph. Réfrégier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 1145–1157 (1999).
34. L. Ferro-Famil, E. Pottier, and J. S. Lee, "Unsupervised classification of multifrequency and full polarimetric SAR images based on the H/A/alpha–Wishart classifier," *IEEE Trans. Geosci. Remote Sens.* **39**, 2332–2342 (2001).
35. F. Goudail, F. Galland, and Ph. Réfrégier, "A general framework for designing image processing algorithms for coherent polarimetric images," in *Proceedings of IEEE 2003 International Conference on Image Processing* (IEEE Press, Piscataway, N.J. 2003), pp. 153–156.
36. S. M. Kay, *Fundamentals of Statistical Signal Processing—Volume II: Detection Theory* (Prentice Hall, Upper Saddle River, N.J., 1998), pp. 186–247.