Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation issues

Philippe Lalanne, Donald Prévost, Pierre Chavel

Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation

Philippe Lalanne, Donald Prévost, and Pierre Chavel

An analogy can be established between image processing and statistical mechanics. Many early- and intermediate-vision tasks such as restoration, image segmentation, and motion detection can be formulated as optimization problems that consist in finding the ground states of an energy function. This approach yields excellent results, but, once it is implemented in conventional sequential workstations, the computational loads are too extensive for practical purposes, and even fast suboptimal optimization approaches are not sufficient. We elaborate on dedicated massively-parallel integrated circuits, called stochastic artificial retinas, that minimize the energy function at a video rate. We consider several components of these artificial retinas: stochastic algorithms for restoration tasks in the presence of discontinuities, dedicated optoelectronic hardware to implement thermal motion by photodetection of speckles, and hybrid architectures that combine optoelectronic, asynchronous-analog, and clocked-digital circuits.

OCIS codes: 100.3020, 100.2080, 100.1160.

1. Introduction

In early and intermediate vision, artificial vision systems are confronted with two major problems. The first problem in machine vision is the sheer amount of input data to be acquired, managed, and processed. The extraordinary volume of data in real-time gray-level images leads to communication bottlenecks in the flow of data among imager, memory, and processor in conventional machines. The second problem arises from the fact that the acquisition system provides incomplete two-dimensional and noisy observations of the three-dimensional scene. The interpretation of the data requires solution of an inverse problem. One has to introduce generic constraints into the problem to force the solution to lie in a subspace of the solution space by incorporating physical *a priori* information at the scene. In this paper we elaborate on dedicated massively-parallel integrated circuits, called stochastic artificial retinas, that rely on regularization methods to solve the inverse problem through video-rate stochastic minimization of appropriate energy functionals. These retinas can be used as single-circuit sensors dedicated to a specialized application or as modular early-vision modules that are assembled at the board level and that are used as front-end processors for a variety of vision applications.

Many problems such as those of edge and motion detection, segmentation, and restoration in early and intermediate vision can be formulated in terms of energy minimization.[1] The energy function combines *a posteriori* knowledge of the data acquisition system and some *a priori* knowledge of image structures. The latter is important since it regularizes the inverse problem by restricting the set of possible solutions. Within the framework of Markov random fields,[2] *a priori* and *a posteriori* knowledge is simply combined through Bayes theory. In general, the Markov random field formulation exhibits two interesting features. First, because only small neighborhood systems are involved for most early- and intermediate-vision tasks, the formulation yields a simple form for the energy function whose gradients involve only local computations. Second, experimental evidence (see, for instance, Refs. 2–4) has shown that, in practice, energy functions with small neighborhood systems can be minimized with manageable, albeit still heavy, computational loads.

The criterion that is most often used for solution of the optimization problem is known as a maximum *a posteriori* (MAP) estimator. Finding the MAP estimate is clearly a formidable computational task: For an image with $N \times N$ pixels and G gray levels, the

The authors are with the Laboratoire Charles Fabry de l'Institut d'Optique, Centre National de la Recherche Scientifique, B.P. 147, F-91403 Orsay Cedex, France.

P. Lalanne's e-mail address is philippe.lalanne@iota.u-psud.fr.

Received 27 November 2000; revised manuscript received 7 May 2001.

0003-6935/01/233861-16$15.00/0

number of possible intensity images is $G^{N\times N}$, which rules out any direct search, even for small values of G and N. A useful computational approach consists in performing a random exploration of the energy landscape following the Gibbs distribution, i.e., visiting configurations with a probability proportional to the exponential of the energy. Several algorithms such as the Gibbs sampler[2] permit such random exploration. The computation of the MAP estimate is provided by annealing.[5] The temperature is continually decreased, so high-energy configurations are less and less frequently visited during the computation. At the end of the annealing, a deep minimum of the energy landscape is obtained in practice.

When the random exploration is implemented in a conventional computer, computational loads are expensive, and suboptimal deterministic algorithms are usually preferred for the MAP estimation. In this paper we focus on very large-scale integrated (VLSI) optoelectronic circuits, hereafter called stochastic artificial retinas, for implementing video-rate minimization in early- and middle-vision tasks by use of stochastic rather than deterministic approaches. More precisely, we try to answer the following two questions: A, Can we design massively-parallel machines that are able to perform video-rate stochastic optimization in early and middle vision? B, How do these machines based on stochastic processing compete with their deterministic counterparts, especially in terms of problem-solving power and complexity of implementation? Several related studies based on deterministic minimization were reported previously; see, for instance, Refs. 6–11. The originality of our approach lies in the deliberate choice to consider stochastic processing. This choice was motivated by the fact that stochastic processing is expected to provide better estimates than deterministic processing over an enlarged set of energy functions or of processing tasks.

Obviously, answering questions A and B is difficult, especially because of the large variety of algorithms and models that are involved in image processing. However, we shall see that a partial answer can be obtained for restricted applications. For several reasons that are discussed below, the classic problem of image restoration appears suitable for studying the feasibility and relevancy of stochastic artificial retinas. Thus in this paper we focus on the so-called compound Gauss–Markov random field models, which can be described by the posterior distribution

$$P(\mathbf{x}, \mathbf{l}/\mathbf{y}) = \frac{1}{Z} \exp - E(\mathbf{x}, \mathbf{l}, \mathbf{y})/E_0, \qquad (1)$$

where

$$E(\mathbf{x}, \mathbf{l}/\mathbf{y}) = \|\mathbf{y} - \mathbf{F}\mathbf{x}\|^2 + \lambda^2[\|\mathbf{M}_\mathbf{l}\mathbf{x}\|^2 + V(\mathbf{l})], \quad (2)$$

Z and E_0 are normalization constants, $\mathbf{y}$ is an $N \times N$ matrix that represents sparse and noisy observation data, $\mathbf{x}$ is a field with continuous variables (pixel process) defined on a regular $N \times N$ lattice, and $\mathbf{l}$ is another field with binary variables that represent the presence or absence of edges (line process). As in many related studies (see, for instance, Refs. 2 and 12), the binary sites are placed midway between the two components of each vertical or horizontal pair of pixels; see Fig. 1 for more details. Notice that sites at or near the boundary have fewer neighbors than interior sites. In what follows, we adopt the natural free-boundary model of Fig. 1, and no periodic boundaries or toroidal lattices are considered. In Eq. (2), λ^2 is the regularization parameter that controls the degree of regularity of the solution or, equivalently, reflects the confidence in the data. This parameter has to be estimated and is related to E_0 and to the variance of the white additive Gaussian noise, which is assumed throughout this paper to be responsible for the noisy observation. $N^2 \times N^2$ matrix $\mathbf{F}$ is sparse when it represents a possible blur of the observed data or can be a more globally interacting projector when middle-vision tasks such as tomographic reconstruction are considered. Although the algorithm contribution presented in this paper is valid for any matrix $\mathbf{F}$, we are more concerned in what follows with blurring because blurring involves more-modest neighborhood sizes compatible with parallel processing. $V(\mathbf{l})$ is the so-called potential. It is simply $V(\mathbf{l}) = \alpha \sum_i l_i$ when noninteracting discontinuities are considered (the energy cost for setting a discontinuity l_i to l is simply α for any i) or can take more-complex forms (see, for instance, Refs. 2 and 12 and Fig. 6) when interacting boundaries that exploit certain physical and geometric constraints on discontinuities, such as smoothness and connection features, are taken into account. $\mathbf{M}_\mathbf{l}$ is a $2N^2 \times N^2$ matrix that represents a smoothing operator that incorporates the line field. With free boundary conditions (as is the case in most early-vision problems), $\mathbf{M}_\mathbf{l}$ has $2N$ useless null lines that can be removed. For simplicity, the sizes of matrices given in what follows will include null lines or columns (boundary effects are omitted for clarity). Specifically, $\mathbf{M}_\mathbf{l}$ is defined

$$\|\mathbf{M}_\mathbf{l}\mathbf{x}\|^2 = \sum_{i\neq kN(k=1,2,3,\ldots)}^{N^2} (x_i - x_{i+1})^2(1 - v_i) + \sum_{i=1}^{N^2-N} (x_i - x_{i+N})^2(1 - h_i). \qquad (3)$$

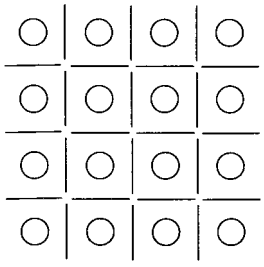

Fig. 1. Field structure for $N = 4$: circles, pixel field; lines, line field.

In Eq. (3), lexicographic notation is used for numbering pixel and line sites, and horizontal and vertical line sites are denoted **h** and **v**, respectively. The energy function of Eq. (2) can be viewed as a generalized form of the classic regularization functional

$$E(\mathbf{x}/\mathbf{y}) = \|\mathbf{y} - \mathbf{F}\mathbf{x}\|^2 + \lambda^2\|\mathbf{M}\mathbf{x}\|^2, \tag{4}$$

incorporating discontinuities. To that end, note that smoothing operator **M** in Eq. (4) is replaced by operator $\mathbf{M}_\mathbf{l}$ in Eq. (2), where **l** denotes the line-site set that results from the union of **h** and **v**.

In attempting to answer questions A and B, we find restoration problems that rely on the energy function of Eq. (2) attractive for several reasons. First, inasmuch as it deals explicitly with two different fields with binary and continuous variables, the energy function of Eq. (2) appears to be appropriate for use in examining issues related to the implementation of stochastic artificial retinas. Second, as a generalization of the classic regularization functional of Eq. (4), the energy function can be applied to many problems in early and middle vision that previously had been solved with standard regularization.[13] Third, the essential problem of preserving discontinuities in image restoration is directly addressed by the energy function of Eq. (2). Finally, although much research has addressed the energy function of Eq. (2), to our knowledge no optimal deterministic algorithm exists for interacting boundaries. For noninteracting discontinuities, one can use nearly optimal fully deterministic algorithms[14,15] to reduce computational costs dramatically. The fact that deterministic approaches are restricted to optimization problems with noninteracting line processes is relevant for access to unique features of stochastic artificial retinas and for answering question B.

In Section 2 an optimal stochastic algorithm for finding the ground state of the energy function of Eq. (2) is described. This algorithm combines different kinds of noise source to produce samples of the intensity and line sites. For a fixed temperature, it achieves thermal equilibrium, sampling the posterior distribution of Eq. (1). It includes a simulated annealing procedure for computing MAP estimates. The algorithm preserves local computation, is parallel and, as we show in Section 4 below, is suitable for VLSI implementation. By considering restoration problems in one- and two-dimensional image formats, in Section 3 we analyze the overall performance of the new algorithm. We perform computational tests to evaluate the ability of the algorithm to sample at thermal equilibrium and to provide MAP estimates. Moreover, the restoration performance obtained for noninteracting and interacting line processes are compared. Comparative visual results for artificial and real images are provided for MAP estimates. In Section 4 we analyze some relevant architectures for the implementation of stochastic artificial retinas that provide video-rate global minimization. These retinas are seen as VLSI circuits that incorporate photodetectors illuminated by the image to be processed and by time-varying speckle patterns. The speckles provide random illumination, which, once it is converted into photocurrents, acts as a source of thermal noise for stochastic relaxation. A retinal architecture dedicated to the minimization of the energy function of Eq. (2) is described. A 24×24 smart VLSI sensor that uses differential speckle detection to minimize Ising-spinlike energy functions at the video rate has been designed and fabricated. Experimental results of the school problem of binary image restoration are presented. Section 5 summarizes the main results obtained throughout the paper and provides a discussion of the capabilities and limitations of stochastic artificial retinas for early vision.

2. Stochastic Algorithm for Compound Gauss–Markov Random Fields

Minimizing the nonconvex energy function of Eq. (2) is not an easy task. The reason for this has to do mainly with the dimension of the problem, which is usually extremely large, and to the existence of many local minima. In the research reported in Ref. 2, the Gibbs sampler algorithm was used to sample the two fields (**x** and **l**), but experimental results were obtained only with multilevel (not continuous) x_i variables and for a small (no more than four) number of gray levels. Several contributions to generalizing the primary work of Ref. 2 to continuous variables of the intensity field were reported previously. It is not our intention to list all these contributions; we restrict ourselves to a brief overview of those that are strongly related to our research and that are useful for its understanding. All these studies rely on the same basic idea and exploit the semiquadratic form of the energy function: For a given line process configuration, the resultant conditional energy functional is a nondegenerate multinormal distribution. In Ref. 16 a mixed-annealing (MA) algorithm is proposed that relies on the facts that the conditional energy functional is convex and that its unique minimum can be found straightforwardly. It uses the Gibbs sampler to update the line process but changes the stochastic relaxation of the intensity field into a deterministic relaxation. The deterministic relaxation removes most of the computational burden, namely, the sampling of the continuous fields, but breaks thermal equilibrium. Thus the MA does not guarantee global minimization. Generalization of the Gibbs sampler to continuous variables[17,18] is not straightforward, and some care must be taken in general. For the specific case of Gaussian distributions, relatively simple stochastic algorithms for sampling the **x** field exist.[19,20] These algorithms use N^2 independent Gaussian delta-correlated noise samples for each full iteration of the **x** field. One iteration corresponds to visiting all pixels and performing a basic sampling step. As for the Gibbs sampler, many full iterations must be repeated a great many times to guarantee convergence toward thermal equilibrium. In the research reported Ref. 10, an algorithm, called Global hereafter, is proposed. It is based on an im-

plicit formulation of the $\mathbf{x}$ field and restricts the stochastic search to subspace $\{\mathbf{x}_l^*, \mathbf{l}\}$, where $\mathbf{x}_l^*$ is the unique intensity configuration that minimizes the conditional energy functional for a given boundary field $\mathbf{l}$. This approach allows global optimization to be performed, but the elimination of the $\mathbf{x}$ field introduces long-range interaction and sacrifices the computational benefits of local neighborhood systems in early vision. Thus the global algorithm is not parallel. By exploiting the semiquadraticity of the energy function, Geman and Yang[21] proposed to sample the multinormal distribution of the $\mathbf{x}$ field directly by using an annealing procedure based on a fast Fourier transform, which provides global minimization. We introduce an algorithm, called the quasi-static relaxation (QSR) algorithm, that provides MAP estimates for the energy function of Eq. (2). This algorithm, which was reported previously in a preliminary study,[22] is suitable for massive parallel implementation, and is especially efficient when small neighboring structures (sparse matrices $\mathbf{F}$) are considered.

A Markov chain $\{[\mathbf{X}(k), \mathbf{L}(k)], k = 0, 1, 2, \ldots\}$ with its associated temperature sequence $\{T_k, k = 0, 1, 2, \ldots\}$ is constructed as follows. For each k, core procedure H is applied as follows:

Subprocedure H1. Generate $\mathbf{X}(k)$ from $\mathbf{L}(k-1)$ by means of the conditional distribution of $\mathbf{x}$, given $\mathbf{l} = \mathbf{L}(k-1)$.

Subprocedure H2. Generate $\mathbf{L}(k)$ from the current configuration $\mathbf{X}(k)$ by sampling the boundary process, using the Gibbs sampler algorithm as described in Ref. 2, by repeatedly programming the Gibbs sampler for local boundary replacement.

Subprocedure H1 directly exploits the fact that the energy function of Eq. (2) is semiquadratic and thus that the conditional distribution of $\mathbf{x}$ given $\mathbf{l}$ is Gaussian, as we discuss in detail below. Subprocedure H2 is more usual; see, for instance, Refs. 2 and 16. In a more general form, n_l full iterations over the line field can be provided, and, for each k, subsequences $\{\mathbf{L}_q(k), q = 1, 2, \ldots n_l\}$ are generated. Clearly, for a given $\mathbf{X}$ configuration, $\mathbf{L}_1(k)$ is obtained from $\mathbf{L}_{n_l}(k-1)$, $\mathbf{L}_q(k)$ $(q \neq 1)$ from $\mathbf{L}_{q-1}(k)$ and $\mathbf{L}(k) = \mathbf{L}_{n_l}(k)$. Although the convergence rate of the core procedure toward thermal equilibrium probably depends on parameter n_l, we did not explore this issue. Throughout the paper, the numerical results are obtained for $n_l = 1$. Thus the core procedure consists in visiting once all the pixel and line sites. Because only short-range interactions have to be considered in updating the line process, the basic updating step is rather elementary and does not require complex computations. Moreover, as the associated chromatic number is small, a high degree of parallelism can be implemented. For the MAP estimation, a slowly decreasing temperature sequence is implemented. General results of simulated annealing guarantee convergence in distribution to a measure concentrated over the global minima of $E(\mathbf{x}, \mathbf{l}/\mathbf{y})$ if the cooling schedule is sufficiently slow. However, because we are concerned mostly with practical implementation, we opt for a faster cooling schedule and construct the inhomogeneous Markov chain with temperature $T(k)$ given by

$$T(k) = cT(k-1), \tag{5}$$

where $c < 1$. In the numerical and experimental results reported below, control parameter c was set equal to 0.9. This annealing schedule no longer guarantees convergence to optimal solutions, but it does return nearly optimal solutions for most problems.[23] At each given temperature, $n$scan full iterations are performed, where the parameter $n$scan represents the length of each homogeneous Markov chain. In summary, the QSR algorithm involves the following steps:

1. Set the initial temperature $T(0)$.
2. Provide $n$scan iterations of the core procedure H at temperature $T(k)$.
3. Set the new temperature according to Eq. (5).
4. Return to step 2 as long as $T(k)$ is larger than a final control temperature.

Let us consider step H1 of procedure H. For a given temperature T, the conditional distribution of $\mathbf{x}$ given $\mathbf{l}$, $P_T(\mathbf{x}|\mathbf{l})$, is normally distributed. Identifying $P_T(\mathbf{x}|\mathbf{l})$ with the canonical form $A \exp[-(1/2)(\mathbf{x} - \mu)^t\mathbf{C}^{-1}(\mathbf{x} - \mu)]$ of a multinormal distribution, we read mean vector μ and covariance matrix $\mathbf{C}$ simply as

$$\mu = [\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]^{-1}\mathbf{F}^t\mathbf{y}, \tag{6a}$$

$$\mathbf{C} = \frac{T}{2}[\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]^{-1}. \tag{6b}$$

We now proceed with generating samples according to the multinormal distribution function defined by Eqs. (6). To do so we solve for $\mathbf{x}$ the following set of linear equations,

$$(\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l)\mathbf{x} = \mathbf{F}^t\mathbf{y} + \mathbf{B}\mathbf{w}^{qs}, \tag{7}$$

where $\mathbf{B}$ is a symmetric matrix defined hereafter and $\mathbf{w}^{qs}$ is a Wiener vector, i.e., a collection of zero-mean independent Gaussian random numbers of unit variance. Clearly, the average solution for $\mathbf{x}$ of Eq. (7) satisfies Eq. (6a). For $\mathbf{x}$ to be normally distributed with the covariance matrix of Eq. (6b), $\mathbf{B}$ has to satisfy the following fluctuation-dissipation relationship (see Appendix A for more details):

$$\mathbf{B}\mathbf{B}^t = (T/2)[\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]. \tag{8}$$

A. General Method

If w^{qs} is an M component vector with $M > N^2$, it is always possible to solve Eq. (8) for $N^2 \times M$ matrix $\mathbf{B}$. For instance, to calculate $\mathbf{B}$ for $M = N^2$ we can use the unit transform matrix $\mathbf{U}$, which reduces $[\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]$ to diagonal form,

$$\mathbf{U}[\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]\mathbf{U}^t = \|k_i(\mathbf{l})\delta_{ij}\|, \tag{9}$$

where $k_i(\mathbf{l})$, $i = 1, 2, \ldots N^2$, are the N^2 eigenvalues of $[\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l]$ and δ_{ij} is the Kronecker symbol: δ_{ij} equals 1 if i equals j and 0 otherwise. As a symmetric matrix, $\mathbf{B}$ is given by

$$\mathbf{B} = \mathbf{U}^t\|\mathbf{k}_i(\mathbf{l})^{1/2}\delta_{ij}\|\mathbf{U}. \tag{10}$$

This approach involves N^2 independent standard normal random variables. In Ref. 21 the fast-Fourier-transform-based annealing algorithm exploits the relationship between block circulant matrices and a two-dimensional discrete Fourier transform to effectively reduce the burdensome computation of the eigenproblem. It is valid if toroidal boundary conditions are used and, in our opinion, is highly efficient for tomographic reconstruction problems but is computationally expensive for problems that involve short-range interactions. More specifically, we note that every $b_{ij}(\mathbf{l})$ element of matrix $\mathbf{B}$ potentially depends on the whole line process, even for a sparse matrix $\mathbf{F}$; the local nature of the field interactions is not reflected in $\mathbf{B}$.

B. Present Method

A more practical solution of Eq. (8) is inspired directly by the form of Eqs. (6). Indeed, it is easy to verify that $[N^2 \times 3N^2]$ matrix $\mathbf{B}$, given by

$$\mathbf{B} = \sqrt{T/2}\,[\mathbf{F}^t, \lambda\mathbf{M}_l{}^t], \tag{11}$$

satisfies the fluctuation-dissipation relationship of Eq. (8). Thus the companion quasistatic vector $\mathbf{w}^{qs}$ required for solving Eq. (7) has $3N^2$ components. The larger dimension of the system is compensated for by the fact that the local structure of interaction with the line process is retained in $\mathbf{B}$. In other words, every $b_{ij}(\mathbf{l})$ element depends at most on a single line site, and parallel calculation strategies can be employed to increase effectiveness. Summarizing, to implement step H1 of the procedure in the short-range interaction case, do the following:

1. Generate $3N^2$ independent Gaussian random variables to form vector $\mathbf{w}^{qs}$.
2. Form matrix $\mathbf{B}$ given by Eq. (11).
3. Solve Eq. (7) for $\mathbf{x}$.

The advantage of using Eq. (7) to update the intensity field lies in the fact that the iterative and burdensome stochastic computation process associated with the intensity field is replaced by a fast and deterministic computation. Basically, the computation requires solving Eq. (7). The largest matrix $\mathbf{B}$ used for the computation has dimensions $N^2 \times 3N^2$. But, when blurring operators with a narrow support are considered, the matrix is sparse, and efficient algorithms can be employed for solving the system of Eqs. (7). The QSR and MA algorithms are similar and are of equivalent computational complexity; both algorithms use the Gibbs sampler algorithm for sampling the line field and require solution of a system of N^2 equations for sampling the intensity field. The difference between the algorithms lies in the fact that, whereas the MA algorithm samples the intensity field at a null temperature and imposes the condition that $\mathbf{x} = \mu$ at every iteration, the QSR algorithm uses independent Gaussian random variables to restore thermal equilibrium.

3. Numerical Results

In this section we report on several tests performed with the QSR algorithm, focusing on its convergence rate and its ability to achieve thermal equilibrium in the coupled field or to provide MAP estimates. Also, the influence of interacting or noninteracting line processes on the quality of the restoration is discussed.

A. Tests Performed on One-Dimensional Signals

We carried out computer simulations to test whether thermal equilibrium was achieved for the coupled field ($\mathbf{x}$, $\mathbf{l}$). The procedure was as follows: For a given data configuration $\mathbf{y}$ and from an arbitrary configuration of the coupled field, the system was slowly annealed ($n$scan = 50) from a high temperature down to a given temperature T. At that temperature, 10,000 configurations ($\mathbf{x}$, $\mathbf{l}$) and their corresponding energy values were collected after each global update of the fields. Then the system was annealed to another temperature, T' ($T' < T$). At that temperature, another set of 10,000 energy samples was collected. Directly testing the Boltzmann–Gibbs distribution of Eq. (1) is not feasible because of the hugeness of the configuration space. Instead, we prefer to deal with the canonical distribution $P(E, T)$ at temperature T,

$$P(E, T)\mathrm{d}E = C_T\Omega(E)\exp(-E/T), \tag{12}$$

where C_T is a normalization constant that depends only on temperature and $\Omega(E)$ is the unknown state density of the coupled fields. $P(E, T)\mathrm{d}E$ is the probability of observing a configuration ($\mathbf{x}$, $\mathbf{l}$) with energy in the interval $[E; E + \mathrm{d}E[$ at thermal equilibrium. From the two sets of samples collected at temperatures T' and T, the two frequency distributions P' and P of observed populations were computed. $P'\Delta E$ and $P\Delta E$ are the number of configurations with energy in the interval $[E; E + \Delta E[$ observed at temperatures T′ and T, respectively. If the sampling is achieved under thermal equilibrium, as $\Omega(E)$ is independent of T, the logarithm of the ratio of the two frequency distributions is given by

$$\ln\left[\frac{P}{P'}\right] = \frac{(T - T')}{TT'}E + C(T, T'), \tag{13}$$

where $C = \ln[C_T] - \ln[C_{T'}]$ is a negative constant that depends on T and T'. According to Eq. (13), the ratio logarithm should vary linearly with energy; a slope $m = (T - T')/TT'$ is expected.

For illustration, we consider the weak-string problem defined by the one-dimensional signal shown in Fig. 2. Uncorrelated Gaussian noise with a standard deviation σ of 16 is added to form the 128×1

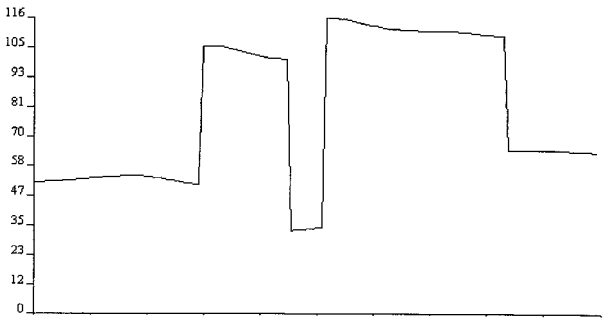

Fig. 2. 128×1 weak string.

input data **y**. No blurring is considered, and matrix **F** is equal to identity matrix **I**. The parameters of the energy function are $\lambda = 4.5$ and $\alpha = 150$. In Fig. 3, the logarithm (P/P') is plotted versus energy E in the overlapping region of the two histograms. Numerical data correspond to black rectangles, and the vertical bars are the uncertainties as estimated from the population's square root. Results were gathered for temperatures $T = 9.41$ and $T' = 8.47$ ($m = 0.0118$). The best least-squares line from the data has a slope equal to 0.012 ± 0.004 and a negative coordinate -13.7 ± 0.1 at the origin. Simulation results and theoretical prediction are in good agreement, with departures from the theoretical curve only at both ends, where frequencies are weak. This result clearly illustrates the achievement of thermal equilibrium for the coupled field. In Section 4 below, we apply this test to experimental data obtained with a fully parallel analog–digital machine. The achievement of thermal equilibrium is important in practice, especially if estimates other than the MAP estimate, such as the maximum of the posterior marginals or the thresholded posterior mean, are chosen.[24]

When a fast cooling schedule such as that of Eq. (5) is used, thermal equilibrium is not perfectly achieved at each temperature $T(k)$. It is intuitively clear that large decrements in $T(k)$ will require longer homogeneous Markov chain lengths to be able to restore quasi equilibrium at the next temperature $T(k + l)$.

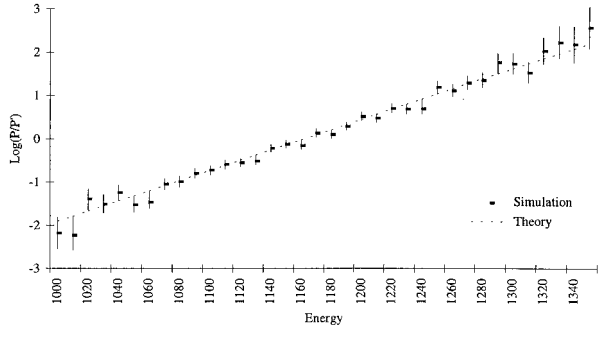

Fig. 3. Numerical test performed on the weak-string problem. At each temperature, 10,000 configurations (**x**, **l**) are collected. The temperatures are $T = 9.41$ and $T' = 8.47$. Vertical bars are statistical uncertaincies.

Table 1. Comparison of Quality of the MAP Estimation for the Weak String of Fig. 2 for Various Algorithms and for Various Values of Parameter *n*scan[a]

Algorithm (*n*scan)	Mean Energy
QSR (1)	90.1
QSR (10)	80.5
QSR (50)	78.0
QSR (100)	77.5
MA (10)	217.9
MA (50)	218.6
Global (5)	77.2
Global (10)	77.2
Global (100)	77.0

[a]The mean energy is obtained by averaging over 50 instances of white Gaussian noise.

Thus there is a trade-off between large decrements and small values of *n*scan. Although this assertion can be stated in more-mathematical terms,[25] we do not pursue the details. To gain a quantitative insight into the influence of parameter *n*scan on the convergence performance of the QSR algorithm, we proceed as follows: From the signal defined in Fig. 2, a set of 50 input data **y** is generated with independent instances of white Gaussian noise, all with a standard deviation of 16. For every instance of noise and for a given value of *n*scan, the QSR algorithm is implemented. The mean values of the energy of the 50 configurations obtained at the end of the annealing are listed in Table 1. As we expect, the mean values decrease with increasing values of *n*scan. For comparison, Table 1 also provides the mean values observed with the MA and the global algorithms. We note that the MA algorithm is clearly not optimal and provides high mean energy values. The lowest mean energy value is obtained for the global algorithm and an *n*scan of 50. Although this algorithm performs well, we must note that it is extremely demanding in terms of computational loads. Because the global algorithm restricts exploration of the configuration space to the set of all the local minima associated with each state of the line process, we need to solve Eq. (6a) for μ for every component of the **x** vector, i.e., N^2 times for each full iteration. For comparison, note that only one inversion per full iteration is required for the QSR algorithm, whatever the number of pixels is.

B. Reconstruction Problems with Two-Dimensional Images

In Subsection 3.A we focused on the weak string as a one-dimensional discontinuity-detecting filter. In this subsection we examine its two-dimensional equivalent, the weak membrane. To qualitatively and quantitatively study the potentialities of the QSR algorithm, we present the results of several numerical computations. The computations were all conducted without any blurring in the image ($\mathbf{F} = \mathbf{I}$).

For the first experiment, the 64×64 input image shown in Fig. 4(a) was considered. This image was synthesized by addition of white Gaussian noise ($\mu =$

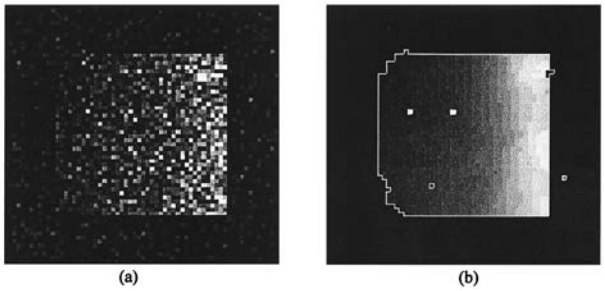

Fig. 4. 64 × 64 weak membrane. (a) Hand-drawn ramp image corrugated by white Gaussian noise ($\mu = 0, \sigma = 65$). (b) Restored image obtained without interaction of the line process ($\lambda = 4$, $\alpha = 500$, $n$scan = 300).

0, $\sigma = 65$) to an original synthetic image formed by a linear ramp with a 39 × 39 square base. The minimum and maximum step heights on the left and the right sides of the original image are 50 and 250, respectively. Figure 4(b) shows a restored image obtained with the QSR algorithm. We delineate the outcome of the line process by painting white any pixel sides at a discontinuity. For the numerical computation, no interaction in the line field was considered, and the free parameters (λ, α) were chosen to equal (4, 500). Basically, the restored image exhibits a large smooth area that corresponds to the square base of the original image. However, we note that there are several inclusions at the pixel size level and that the small step on the left side is badly restored. Although the degradation that is due to white noise is severe and no interaction in the line process is used for the restoration, no ending appears in the line process of Fig. 4(b), and closed regions in the image are obtained. This tendency to form unbroken lines without any need to impose additional cost on line endings is an intrinsic property of membrane elasticity that in Ref. 14 is termed hysteresis. It is worth mentioning that, despite the hysteresis, in other experiments performed with the weak membrane the restored image did not always exhibit closed line processes. In particular, the upper left-hand corner of the ramp was smooth in several observations.

For the second experiment we considered the noisy image of Fig. 5(a). This synthetic image was obtained by addition of white Gaussian noise ($\mu = 0$, $\sigma = 65$) to an original synthetic image, a step with a 39 × 39 square base. The step height is 50. Figure 5(b) shows the restored image obtained with the QSR algorithm for (λ, α) = (3, 250) and $n$scan = 300. Clearly, the restoration quality is poor, but the values (3, 250) of the free parameters are the best values that we could find in several trials. Figure 5(c) shows the restored image obtained with the QSR algorithm for the interacting line process shown in Fig. 6 and for $\alpha = 300$. Basically, the potential favors turns and continuations and penalizes endings. We note that, except for some minor inclusions of one-pixel size, the step is approximately reconstructed. The net benefit of regularization by explicit incorporation of an interacting line process was also ob-

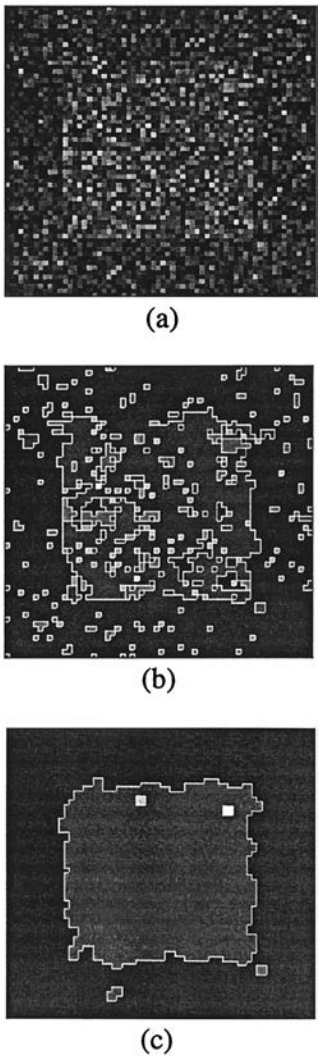

Fig. 5. 64 × 64 weak membrane. (a) Hand-drawn step image corrugated by white Gaussian noise ($\mu = 0, \sigma = 65$). (b) Restored image obtained without interaction of the line process ($\lambda = 3$, $\alpha = 250$, $n$scan = 300). (c) Restored image obtained with $\lambda = 3$, $\alpha = 250$, $n$scan = 300, and the potentials of Fig. 6.

served for real images with or without blurring. To achieve a more-accurate reconstruction, we tried other potentials. As was proposed in Ref. 26, some line configurations, such as ending boundary placements, were forbidden by constrained optimization. We could not achieve a better restoration quality. In our opinion, for the image of Fig. 5(a), better reconstructions could be achieved with higher-order models of the line process that would involve larger

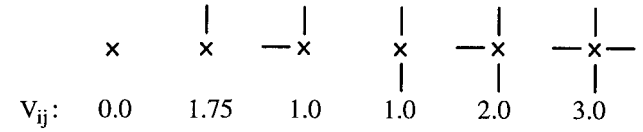

Fig. 6. Numerical values of potentials V_{ij} used for the interacting line process. Crosses, nodes; vertical and horizontal lines, existing edges. Rotational invariance is assumed.

neighborhood structures. We did not explore this method of achieving reconstructions because it might be difficult to implement in dedicated circuitry.

4. Implementation of Stochastic Artificial Retinas

Here we discuss architecture and hardware for implementing stochastic artificial retinas. The following discussion is restricted to early-vision tasks with reasonably local neighborhood systems, for which communication requirements are small and a high degree of parallelism can be achieved by suitable coloring of the sites. These retinas provide video-rate stochastic relaxation operations with which nearly optimal estimates can be computed. The retinas are silicon chips built from a mesh of processing elements (PEs), each with its own photodetectors and some computational abilities. The latter are rather weak, but the computational power comes from all the PEs working in parallel. Clearly, the implementation of massive parallelism is not obvious from a practical point of view because the required silicon area per PE is much too large. Typically, a 256×256 regular array of PEs could be implemented on a 1-cm^2 chip in a 1-μm complementary metal-oxide semiconductor (CMOS) technology, provided that each PE contains fewer than 30 transistors; this of course is a very small size, which can be achieved only if analog operations are implemented.

A. Parallel Generation of Massive Amounts of Random Numbers

Ideally, a stochastic artificial retina that involves 256×256 PEs working in parallel at a global chip frequency of 1 MHz requires approximately 65×10^9 independent random numbers each second. Clearly, such a huge rate deserves particular study. Although noise often appears naturally in systems, paradoxically, when one is attempting to use this noise to generate random numbers it is often difficult to obtain properties that are sufficiently statistically random. Several techniques that exploit natural noise sources[27–30] or pseudorandom-number generators[31,32] are compatible with these requirements.

Our approach relies on speckle. Speckle is a natural noise of coherent light that is observable when the random complex amplitudes of many coherent scatterers are added. The resultant interference pattern is recognizable by its random granularity, which can be described by statistical means.[33] In a preliminary study[34] we argued in favor of the use of speckle created by step-index fibers. By computing values as low as 10^{-3} for some first-order autocorrelation values and by successfully implementing statistical tests of independence, we concluded that fiber speckle offers good statistical properties of space and time independence. More recently, the differential detection of speckles implemented by a phototransistor pair was shown[35] to be an efficient tool for implementing zero-mean Gaussian random current sources for VLSI silicon circuits. Analytical and experimental evidence shows that the photocurrent's

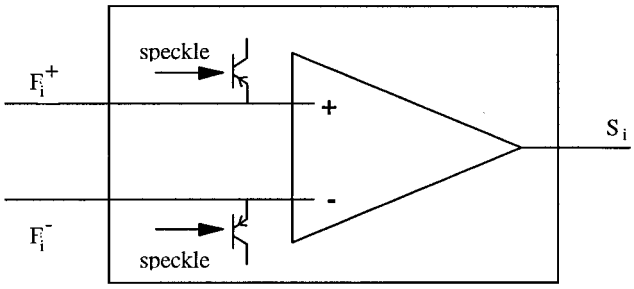

Fig. 7. Clipped differential detection of speckle with a dual-rail comparator; $F_i = F_i^+ - F_i^-$.

standard deviation σ is proportional to the mean illumination $\langle I \rangle$ of each phototransistor:

$$\sigma = A\langle I \rangle, \tag{14}$$

and thus one can set it freely by controlling the laser power. In Eq. (14), proportionality factor A depends only on the average number of speckles incident onto each phototransistor and is fixed by the experimental setup. Although it exploits the analog nature of speckle statistics, the differential detection is robust and accurate; deviations from Gaussian laws are less than 1%.

The quasi-Gaussian laws obtained by the differential detection of speckles can be exploited for sampling binary variables. As has been explained above, binary variables are important in early and middle vision because they may encode the presence or absence of discontinuities or may label sites in motion or at rest, for instance. We denote these variables with on–off states $\{-1, +1\}$ S_i, $i = 1, 2 \ldots N$. The local site replacement used repeatedly for sampling (heat-bath criterion) is

$$p(S_i = 1) = \frac{1}{1 + \exp(F_i/T)}, \tag{15}$$

where force F_i is the energy difference between the on and off states: $F_i = E(S_i = 1 | S_r = s_r, r \neq i) - E(S_i = -1 | S_r = s_r, r \neq i)$.

The implementation of the heat-bath criterion with speckles[36] relies on the fact that the cumulative distribution function of a Gaussian, the error function, never differs by more than 1% percent from the sigmoid probability function of Eq. (15). A possible implementation with a comparator is shown in Fig. 7. Four currents are injected into the comparator: the two photocurrents generated by the two phototransistors and illuminated by independent speckles and two other currents, F_i^+ and F_i^-, related to the force by $F_i = F_i^+ - F_i^-$ and typically representing the positive and negative contributions in F. The stochastic PE depicted in Fig. 7 was integrated in the same CMOS technology as that used in the study reported in Ref. 35. Under time-varying speckle illumination, its operation was tested, and we computed the probability that the output voltage S_i equals 5 V, $p(S_i = 1)$ by averaging for different values of the photocurrents encoding the force. Experimen-

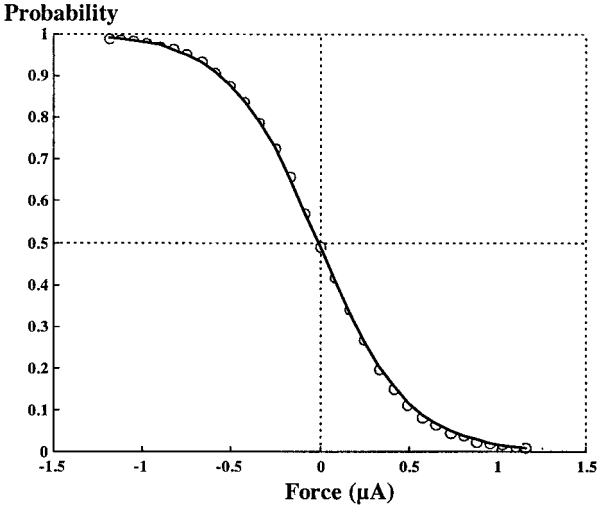

Fig. 8. Circles and solid curve, experimental data and the fitted heat-bath criterion, respectively. Every circle was estimated over 100,000 collected independent measurements.

tal results (circles) are shown in Fig. 8 and are compared with those for the sigmoid heat bath (solid curve). Excellent agreement is obtained. Simple dimensional analysis shows that temperature T is proportional to the average speckle illumination. Thus we can control temperature by changing the laser power. To avoid any confusion, let us note that the S-shaped response shown in Fig. 8 is due only to the probabilistic updating and not to any nonlinear response of the comparator. In the absence of speckle, a Heaviside function is obtained instead of the S-shaped response (null temperature).

B. Stochastic Artificial Retinas for Implementing Ising Spin Models

To gain a more-quantitative insight into hardware problems relative to the implementation of stochastic artificial retinas, we built a prototype that is able to perform stochastic processing of spin-glass models.[37] This choice was also motivated by the fact that spin-glass models exhibit a strong analogy with some simple early-vision tasks.[24,38] According to the analogy, the binary pixels S_i are called spins, and the energy $E(\mathbf{S})$ of a given configuration $\mathbf{S}$ is

$$E(\mathbf{S}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} J_{ij}S_iS_j - \sum_{i=1}^{N} h_iS_i. \quad (16)$$

In Eq. (16), h_i denotes an external local field and J_{ij} is the interaction coefficient between spins i and j. We consider a two-dimensional spin-glass model arranged in an image format, with symmetrical ($J_{ij} = J_{ji}$) and binary (± 1) interaction coefficients. Only horizontal and vertical nearest neighbors are considered; i.e., the system is a first-order Markov field in the i and j directions.

The system that we implemented combines speckle and electronic circuitry. It is composed of two main devices: A silicon integrated circuit is in charge of the parallel computation of the energy gradients (forces) involved in the minimization, and a speckle random-number generator provides the circuit with 10^4 spatially and temporally independent illuminations each microsecond. Following the analogy with statistical mechanics, the integrated circuit alone implements the time evolution of the spin system at a null temperature, whereas the speckle illuminations, converted into photocurrents, act as sources of thermal noise. The temperature is read as the amount of randomness provided by the speckle photocurrents, and we obtain annealing simply by decreasing the laser power.

The integrated circuit implemented with 1-μm CMOS digital technology is composed of a 24×24 array of identical PEs. The total circuit area, including connection pads, is 8 mm $\times$ 8 mm. Every PE includes an analog and a digital block. The digital block of the ith PE stores the spin value S_i and two of the four coupling coefficients J_{ij} in three static memory points. These three static memory points are incorporated into a horizontal shift register, permitting reading and writting of the spin configurations and J_{ij} coefficients. The four bipolar products $J_{ij}S_j$ involved in the computation of the force F_i are implemented with logical XOR gates. The analog block is based on a dual-rail architecture for minimizing on-chip dispersions. This architecture fits nicely with the differential detection of speckles. We compute forces F_i^+ and F_i^- by adding the positive and negative bipolar products $J_{ij}S_j$ through current summing at the nodes of the plus and minus rails, respectively. Further details of the electronic implementation can be found in Ref. 39. A spatially uniform external field can be implemented by injection of a global current into either the plus or the minus rail of every PE, whereas any external field configuration can be simulated by imaging of a gray-level mask illuminated by an incoherent source onto the integrated circuit. According to the sign of h_i, the light that issues from a square subdomain of the mask is placed such as to illuminate the phototransistor connected to the positive or to the negative rail. Finally, the photocurrents generated by two phototransistor speckle illuminations are injected into their associated rails. The same phototransistors are used for the optical inputs of the speckle and of the nonuniform external field. We update spin i according to the heat-bath criterion by latching the voltage that results from comparison of the two rail currents into the S_i memory point. We implement the speckle illumination of the integrated circuit by imaging the exit face of a multimode fiber onto the circuit. The rotation speed of a diffuser inserted between the laser diode and the fiber input face allows the time correlation of the speckle illuminations to be controlled. This duration is adjusted to the inverse of main clock frequency of the chip used for successively updating black-and-white spin sites alternately. The whole system is controlled by a personal computer, enabling one to read and write spin values, to write interactive-coefficient values, and to set the global and local external fields, the laser power, and the clock signals.

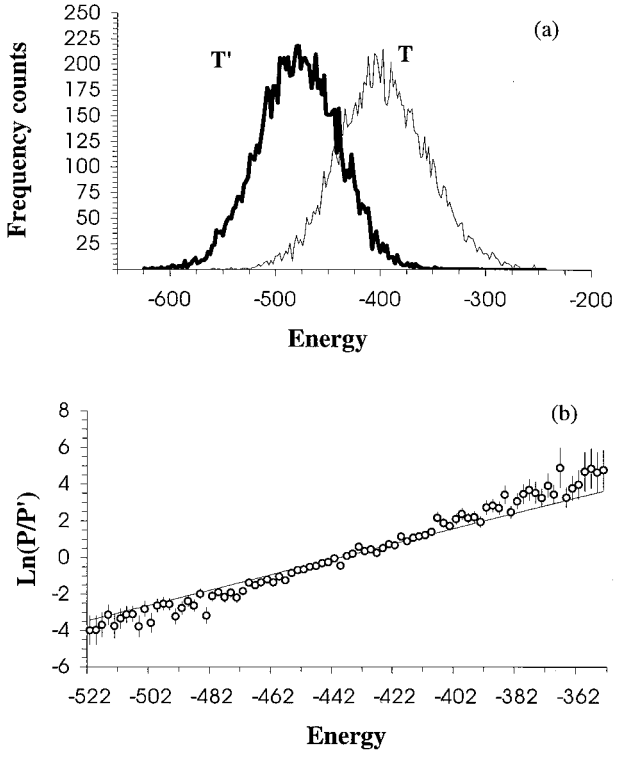

Fig. 9. Test of thermal equilibrium with the implemented system. (a) Experimental data obtained by collection of 10,000 energies at temperatures $T = 3.02$ and $T' = 2.68$. (b) Logarithm of the ratio P/P' as a function of energy. Circles, experimental data. The slope of the solid line is $T - T'/TT'$.

Many local tests were performed with the prototype to validate the digital–analog operation. We also performed global tests to study the influence of inaccuracies in analog computation on the overall system performance, for instance, on the ability to reach deep minima of the energy by annealing. Two such tests are now described.

The first experimental test evaluates the ability of the implemented system to sample at thermal equilibrium. The J_{ij} coefficients were randomly set to 1 with a probability equal to 0.87, and no external field was applied (spin-glass model without a magnetic field). As the configuration space is very large (2^{576} configurations), the test was performed on the canonical distribution; see Subsection 3.A. For two values of the laser power corresponding to two temperatures, T and T', statistics were collected during time evolution of the system. Figure 9(a) shows the two histograms obtained by collection of 10,000 configurations. In the overlapping region of the two histograms ($E \in [-520; -350]$), the ratio of the histograms is computed, and its logarithm is plotted as a function of energy [Fig. 9(b)]. Because of spatial dispersions that are due to the analog implementation, temperature is defined only locally, at every spin site. For a fixed laser power, the observation of the probability that spin i takes the value 1 as a function of F_i results in a stochastic updating function that differs slightly from the heat-bath criterion. By fitting the observed and expected probabilities, we derive the local temperature T_i associated with spin i. The global temperature is defined by averaging over all sites. For the two histograms, the global temperatures T and T' are 3.02 and 2.68, respectively. The slope $m = (T - T')/TT'$ is thus 0.042. In Fig. 9(b) the solid line represents the least-squares fit of slope m to the experimental data. Clearly, a deviation exists. In fact, the best least-squares line fitted from the experimental data has a slope $m' = 0.82$ m. This deviation is due to inaccuracies of the analog implementation, as was confirmed by numerical simulation. In general, it was found that this test is sensitive and reveals slight deviations from thermal equilibrium. This conclusion also holds for numerical simulations, and in our opinion this test is a good candidate procedure for monitoring the decreasing rate of temperature in annealing.

The second test concerns a binary-image restoration problem. The interaction coefficients are all equal to 1. Experimental results are shown in Fig. 10. Figure 10(a) shows a letter L degraded with 25% white channel noise. In the system, this noisy image represents the external field and is encoded through a binary-level mask that is imaged onto the integrated circuit. The restored image, shown in Fig. 10(b), is obtained by annealing in the presence of speckle. Figure 10(c) represents a typical result obtained without speckle. In this case the system is

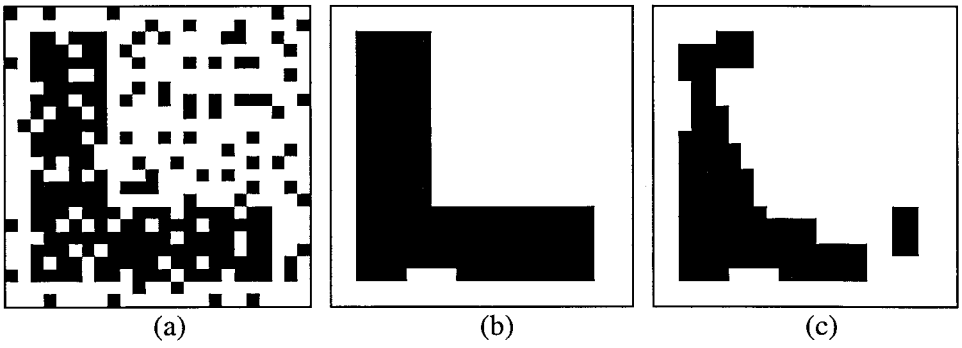

Fig. 10. Experimental results. (a) 24×24 synthetic image corrugated by white channel noise; the original image is a letter L. (b) Restoration with speckle (annealing). (c) Restoration without speckle (gradient descent).

operating at a null temperature and implements a gradient descent algorithm. In a few iterations, it is trapped in a local minimum that depends on the initial configuration. Comparison with numerical results obtained with a personal computer indicates that, despite the inaccuracies of the analog implementation, the system has good overall performance and provides deep minima. For instance, the average energy value obtained over 50 annealings with the implemented system is only 19 larger than that found by numerical computations performed on a personal computer. For comparison, the energies found with a gradient descent algorithm are on the average approximately 200 larger.

For the design of the silicon chip our primary goal was to develop a robust prototype incorporating several test modes to better evaluate the performance of the stochastic processing and not to beat records in terms of integration density or computational speed. With respect to PE integration density, there is room for improvement in both the logical and the analog blocks. For a silicon VLSI chip dedicated to the restoration of binary images, the $J_{ij}s$ coefficients need not be reconfigurable, and no global external field has to be implemented. It is reasonable to expect that a moderately risky design of a 1-cm $\times$ 1-cm silicon chip fabricated with the same technology will incorporate approximately 100 $\times$ 100 PEs. With up-to-date technologies, this density can be further increased. However, note that the phototransistor area (at present 30 μm $\times$ 30 μm) does not scale down well with decreasing resolution. Further integration will therefore imply other approaches such as the use of focal plane arrays to concentrate light onto the active areas. At present, the global clock frequency of the chip is limited by the phototransistor's cutoff frequency ($\approx$250 kHz), and there is a full iteration over all black and white spins every 10 μs. Thus video-rate annealings are possible if no more than 2000 full iterations are performed. This is not the ultimate limit. In our opinion, 1 or 2 orders-of-magnitude increases in the phototransistor response speed can be achieved with a slightly different design. A 1-MHz global clock frequency seems a reasonable prediction for stochastic machines composed of an array of PEs operating on binary variables with local neighborhood systems.

C. Architecture for Image Restoration Problems

In this subsection we elaborate on stochastic artificial retinas for video-rate implementation of the QSR algorithm. A hybrid parallel architecture that comprises analog, digital, and optoelectronic circuits is proposed for sampling the $\mathbf{x}$ and $\mathbf{l}$ fields under thermal equilibrium. Of course, as we are concerned with VLSI silicon circuits, only short-range neighborhood systems are considered. For simplicity we restrict the following discussion to early-vision tasks for which matrix $\mathbf{F}$ is identity matrix $\mathbf{I}$. Our approach is directly inspired by the well-known fact that, according to Maxwell's heat theorem, the steady state of an electrical network composed of sources and linear resistances is the global minimum of a convex quadratic form.[40] The mapping between analog networks and the solution of variational problems has been extensively studied in the context of regularization in early vision. Analog linear networks[1] are seen as a natural computational model for finding the global minimum of the classic regularization functional of Eq. (4). For solving nonconvex variational problems, nonlinear networks have to be considered. In Refs. 6 and 24 the use of a hybrid architecture that relies on a sequence of alternate probabilistic and deterministic steps was proposed for implementing the MA algorithm. In Ref. 10, the implementation of the global algorithm was discussed by use of basically the same hybrid architecture made from a grid of digital PEs interacting with a linear neural network. In Refs. 41–43, resistor-with-fuse networks are proposed as a means for minimizing the energy function of Eq. (2) for noninteracting boundaries. By tuning of the voltage control of the fuse resistances, a minimization is performed through a sequence of convex functionals that are free of spurious local minima. Clearly, the architecture presented herein is directly inspired from those related papers. Its originality comes from the fact that rigorous stochastic relaxation schemes are investigated. For this, we introduce the concept of resistive networks corrugated by Gaussian noises. This noise is called quasi-static to emphasize the fact that its temporal correlation length is much larger than the relaxation time of the network. This situation has to be opposed to that of the well-known Johnson noise, which automatically generates voltage fluctuations in an electric resistance. As was explained by Nyquist,[44] this noise is white in the sense that its spectrum is almost flat up to frequencies much higher than the inverse of the network relaxation time. To our knowledge, the use of quasi-static noises in linear resistive networks for sampling multinormal distributions was not proposed earlier.

Before going to a description of a stochastic artificial retina for implementing the QSR algorithm, we need to introduce a property of linear resistive networks corrupted by quasi-static Gaussian noise. Let us consider a network with $n + 1$ nodes. Node i ($i = 1, 2 \ldots n$), whose voltage is denoted V_i, is shown in Fig. 11. The resistance that links node i to node j is denoted R_{ij} ($R_{ij} = R_{ji}$), and the grounded resistance at node i is R_{ii}. A current of magnitude equal to J_i is injected into node i. The associated current generator is also connected to the common ground (node 0) of the resistive network. Moreover, we suppose that, at every link between nodes i and j, a Gaussian-noise current source with a magnitude equal to θ_{ij} ($\theta_{ij} = -\theta_{ji}$) is associated in parallel with resistance R_{ij}. The property (P1) of linear resistive networks corrupted by quasi-static Gaussian noise sources can be described as follows:

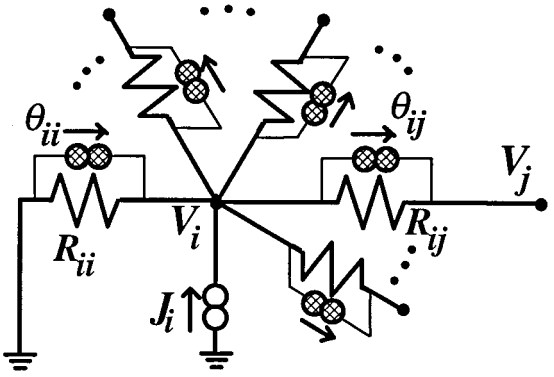

Fig. 11. Node i of a resistive network described with the generalized voltage coordinates. R_{ij} and V_i denote the resistance of the resistor linking nodes i and j and the voltage at node i, respectively. θ_{ij} and J_i are current source generators.

If all Gaussian sources are independent with zero mean and variance v_{ij} given by

$$v_{ij} = \frac{T}{2} R_{ij}^{-1}, \tag{17}$$

and if the temporal correlation length of the sources is much longer than the RC relaxation time of the resistive network, then the distribution function of the potentials is independent of the node capacitances and is given by

$$P(\mathbf{V}) = \frac{1}{Z} \exp\left[-\frac{K(\mathbf{V})}{T} \right], \tag{18}$$

where the pseudoenergy $K(\mathbf{V})$ is equal to

$$K(\mathbf{V}) = \sum_{i=1}^{n} \frac{V_i^2}{R_{ii}} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(V_i - V_j)^2}{R_{ij}} - 2 \sum_{i=1}^{n} I_i V_i. \tag{19}$$

$K(\mathbf{V})$ is the total power dissipated in the resistances minus twice the total power taken from the grounded current generators. For $T = 0$, i.e., in the absence of noise, property P1 is simply reduced to Maxwell's heat-dissipation theorem: The stationary generalized voltages are those that minimize $K(\mathbf{V})$. The demonstration of Eqs. (18) and (19) comes mainly from simple electrical and statistical considerations and is given in Appendix 2.

A possible machine for the implementation of the QSR algorithm is shown in Fig. 12. It is composed of two interacting meshes, a four-connected-analog noisy resistive network, and a locally interconnected array of digital–analog PEs. These two meshes, respectively, implement the intensity and the boundary fields. The analog's resistive network works in a purely asynchronous mode. It features switches that set or break connections between adjacent nodes to implement the absence or presence of discontinuities. Its principle of operation can be understood

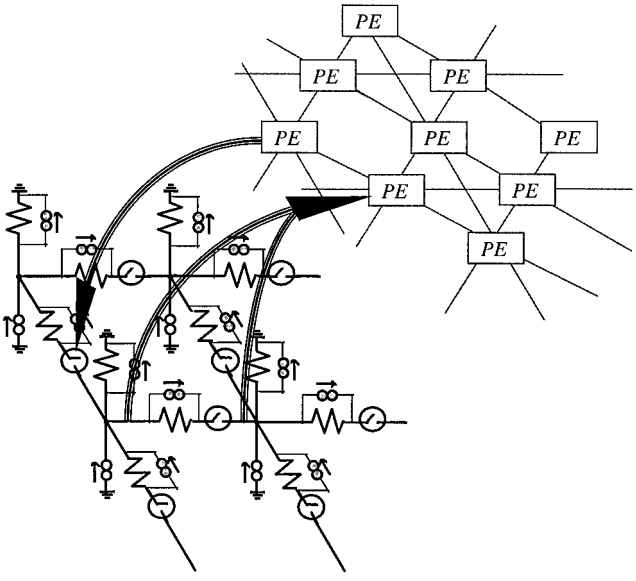

Fig. 12. Possible hybrid machine for implementation of the QSR algorithm for $\mathbf{F} = \mathbf{I}$. It features two interacting meshes, an analog noisy resistive network and a locally interconnected array of stochastic PEs. The resistive network, which is mapped onto a four-connected lattice of nodes, samples the intensity field at thermal equilibrium; the mesh of PEs uses the Gibbs sampler to sample the boundary field. The hybrid machine has two basic cycles. At fixed time intervals, the update of the binary variables l_i encoding the presence or absence of boundaries requires the measurement (large upward arrows) of the voltage difference between the adjacent nodes of the resistive network and some local computation involving the current states of neighboring PEs. The updated binary output is injected into the resistive network (large downward arrow) through the new state of the corresponding switch, which sets or breaks the resistive connection between nodes.

straightforwardly if we rewrite the pseudoenergy $K(\mathbf{V})$ as

$$K(\mathbf{V}) = \sum_{i=1}^{n} \frac{(V_i - R_{ii} I_i)^2}{R_{ii}} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(V_i - V_j)^2}{R_{ij}} - \sum_{i=1}^{n} R_{ii} I_i^2. \tag{20}$$

The last term, which does not depend on the potentials, is not relevant and can be incorporated into the normalizing constant Z of Eq. (18). By reading the generalized potentials in Eq. (20) as node voltages encoding the continuous intensity field $\mathbf{x}$, it is easily shown that the steady-state potentials $\mathbf{V}$ of the resistive network of Fig. 12 are samples of the multinormal distribution function defined by Eqs. (2) and (3) for $\mathbf{F} = \mathbf{I}$. This holds, provided that

$$R_{ii} = 1, R_{ij} = \lambda^{-2} \quad (i \neq j), \tag{21a}$$

$$J_i = y_i, \tag{21b}$$

and independent Gaussian noise sources with variances $T/2$ and $\lambda^2(T/2)$ are attached to the grounded (vertical) and the lateral resistors, respectively.

These noise sources are represented in Fig. 12 as

shaded current generators. Following the results of Subsection 4.A, we can implement them by attaching a differential pair of phototransistors to every resistor and by illuminating the resistive network with speckle. According to Eq. (14), the mean illumination of the phototransistors attached to vertical resistors has to be λ^2 times smaller than that of the phototransistors attached to lateral resistors. In practice, under a spatially uniform speckle, this can be achieved by integration of phototransistors of different sizes, with λ^2 as the ratio of their area. The data to be processed are incorporated into the resistive network through current generators; see Eq. (21b). This incorporation can be achieved by imaging of a spatial light modulator that is encoding the data onto photocurrent generators $\mathbf{J}$. Or, as was described in Subsection 4.B, the same phototransistors may be used for the inputs of the speckle and of the image to be processed. In this approach, generators $\mathbf{J}$ are useless, and the data are directly imaged onto one of the two phototransistors used for the implementation of the Gaussian random sources attached to the vertical resistors.

The sampling of the line field is obtained on a locally interconnected mesh of digital–analog PEs, each implementing the heat-bath criterion. Once again, the use of clipped differential detection of speckles may advantageously replace area-demanding silicon digital random-number generators. The operation and implementation of this mesh are similar to those of the stochastic artificial retina described in Subsection 4.B. The main difference comes from the force computation, which includes contributions from the resistive network (voltage differences between adjacent nodes) and from neighboring PEs (the force that results from potentials V_{ij}). If an interacting line process such as that shown in Fig. 6 is considered, every PE has to be connected to six neighbors, and the chromatic number for parallel computation is four. Unlike for the Ising-spin model, the PE's contribution to the force does not depend linearly on the line variables. Its implementation requires Boolean operations, which may limit the PE's integration density. The implementation of noninteracting line processes is much easier because the force computation reduces to the voltage differences between adjacent nodes and a constant bias equal to α. In this case, no coloring is required, and all the PEs can be updated in parallel.

With this type of design, we end up with a hybrid machine that involves asynchronous-analog and clocked-digital operations for the intensity and the line fields, respectively. All random-number generation can be performed optically, producing three differential phototransistor pairs per intensity site and an additional pair per line site. PEs are updated synchronously at clock time $t_i = 1, 2, \ldots$. This updating implies computing the forces, which in turn involves the resistive network and neighboring PEs. For a successful operation, the time interval $\Delta t = t_{i+1} - t_i$ has to be much larger than the RC relaxation time of the resistive network. This ensures that the voltages read by the digital–analog mesh of PEs will effectively correspond to steady states of the resistive network. The relaxation time of the resistive network is difficult to evaluate, as it depends on the line process configuration and on the temperature. Basically it is proportional to the area of the largest smooth part of the network that has no discontinuities. At high temperature, this area is rather small, as many switches break connections between adjacent nodes. At low temperature, larger areas are involved, but the voltage swings are lower. For all practical purposes, the voltage changes induced by the switches typically have to propagate a few nodes away before the decay that is due to the grounded resistors swamps them out, and the voltages at nodes farther away will remain relatively unchanged. A typical value for the relaxation time is a few hundreds of nanoseconds.[45] Thus a reasonable global clock frequency for the system of Fig. 12 is 1 MHz. If video-rate operation for the minimization is undertaken, approximately 10^3–10^4 full iterations can be performed every 20 ms, depending on the coloring of the edge field.

Clearly, the implementation of such a machine with up-to-date silicon technology is challenging. If noninteracting line processes are considered, a monochip integration is preferable, as it favors the integration density. For interacting line processes, a monochip integration is risky, as the large voltage swings of the digital electronic required for the edge field are likely to contaminate the nearly analog computation. Thus a hybrid approach that relies on the integration of several chips in digital and analog CMOS technologies is more appropriate. The difficulty in such a hybrid system lies in the implementation of the large number of connections required for the interaction of the two meshes. With conventional chip-to-substrate contacts along the perimeter of a chip, time multiplexing has to be implemented, and the speed of the system is likely to be limited by the inevitable bottleneck that one usually has to deal with in vision systems. With forthcoming packaging technologies developed for multichip modules, controlled-collapse chip connection processes will connect the chip directly to the substrate. Thus the contact density is proportional to the chip area. Based on this board-scale integration, more than 10,000 connections/cm^2 are available,[46] and the prospects for implementing our proposal multichip system may therefore improve.

5. Concluding Remarks

In this paper we have considered stochastic artificial retinas that are able to implement stochastic relaxations on Markov random fields in early and middle vision. These retinas are hybrid parallel machines that mix optics and digital and analog electronics and provide global optimizations at a video rate. They can be used as single-circuit sensors dedicated to a specialized application or as modular early-vision modules that are assembled at the board level and that are used as front-end processors for a variety of

vision applications. As an aid to better understanding of the advantages and drawbacks of building such dedicated machines, several aspects of their construction, including algorithms, computational efficiency, parallelism, hardware, and architectures, have been considered. Although all the numerical and experimental examples covered in the paper are obtained for the maximum *a posteriori* estimator, the results for issues of algorithm and implementation are not restricted to this estimator. Other, more local, estimates that might provide better performance,[14,47] for instance, the maximizer of the posterior marginals, can be straightforwardly considered.

Throughout the paper we have focused on the problem of image restoration in the presence of discontinuities. In the more-general case, our results may be applied to optimization problems defined by a semiquadratic energy function. This function explicitly depends on a continuous field related to the observed data and on an additional binary field that explicitly encodes the presence or absence of some qualitative property such as the presence of a boundary. For this energy function we presented a rigorous algorithm, called the quasi-static relaxation algorithm, that induces stochastic relaxation in coupled fields. For a fixed temperature, it samples configurations at thermal equilibrium, i.e., under the Gibbs distribution. When it is used with simulated annealing, it provides maximum *a posteriori* estimates. Its main attractive feature is that, when local neighborhood structures are considered, as is the case in most early-vision tasks, it fully exploits the short-range interactions. Thus it can be implemented with a high degree of parallelism. Moreover, the algorithm is efficient for conventional workstations because sampling of the continuous fields requires only the inversion of a sparse linear system for which efficient inversion routines exist.

Considering restoration problems in one- and two-dimensional image formats, we performed computations to test the algorithm's overall performance. More specifically, we successfully tested the ability of the algorithm to perform sampling under conditions of thermal equilibrium at a fixed temperature and to compute global minima of the energy function. To our knowledge, for interacting boundary fields that reflect some basic constraints such as the absence of ending, no rigorous deterministic algorithms exist, and the QSR algorithm is one of the most efficient algorithms at providing global estimates. The effect on restoration performance of incorporating interacting boundary fields into the energy function was studied. Visual comparative results for synthetic images clearly showed that models with interacting boundaries enhance the restoration quality. For simple reconstruction operators ($\mathbf{F} = \mathbf{I}$), enhancement was observed for images strongly corrugated by noise. We believe that more-complex problems, such as restoration with blur or tomographic reconstruction, may benefit from interacting boundaries at an intermediate noise level.

Implementing stochastic artificial retinas that provide video-rate operation was discussed. We introduced the imaging of time-varying speckle patterns onto silicon chips as an efficient tool for generating random numbers in VLSI circuits. Once they are converted into photocurrent, the speckles act as sources of thermal noise for stochastic relaxation. These sources are nearly Gaussian, and, when they are used with comparators, they provide an accurate implementation of the heat-bath criterion for sampling binary Markov random fields. The feasibility of building analog–digital machines that mix electronics and speckle random-number generators was demonstrated, and the successful video-rate operation of a 24×24 smart VLSI sensor to minimize Ising-spin-like energy functions was validated.

An architecture for the implementation of the QSR algorithm was proposed. It relies on dedicated parallel circuits that use speckle random-number generators to implement the stochastic relaxation of both fields. It guarantees the rigorous optimization of several estimators for both interacting and noninteracting line processes. This generality is obtained at the expense of the integration of additional noise sources to implement the stochastic samplings. In our opinion, the extra work involved in using this architecture does not present a serious difficulty; efficient solutions that are compatible with massive integration and good statistical properties, such as the differential detection of speckle, exist.

Stochastic artificial retinas do have two major drawbacks with respect to conventional numerical processors. As their successful integration depends on dedicated analog computations, they are sensitive to the inevitable drift and imprecision of analogs; they are not versatile, inasmuch as every implementation is tailored to a specific processing task. Clearly, the architecture proposed in this paper is not exempt from these two drawbacks; for instance, taking into account an eventual blur with a narrow support would lead to a somewhat different and more-complex design. Not only would a network with longer connections be required, but also integration of resistors with negative resistances and control of foreseeable network stability problems would have to be achieved. Hybrid architectures that combine digital and analog computation, such as that of Fig. 12, probably represent a good compromise that exploits the flexibility of digital electronics without sacrificing the net benefit to integration density of analog computations. Although the successful operation of a hybrid machine that implements the QSRelaxation algorithm is clearly a continuing challenge, perhaps one of the most important results of this paper is to show that solutions exist for the implementation of optimal stochastic processing techniques in early-vision problems. Stochastic artificial retinas represent a generic platform for building efficient vision machines to operate over a large class of energy functions and estimators. Moreover, they contribute to reducing the gap between sophisticated models developed by researchers on the one hand and dedicated machines designed by engineers on the other, and

they present promising prospects for the implementation of video-rate Monte Carlo–like computations.

Appendix A: Demonstration of Eq. (8)

For simplicity, we denote by $\mathbf{G}$ the matrix $\mathbf{F}^t\mathbf{F} + \lambda^2\mathbf{M}_l{}^t\mathbf{M}_l$. From Eq. (7) we have found that $\mathbf{x}$ is equal to $\mathbf{G}^{-1}\mathbf{F}^t\mathbf{y} + \mathbf{G}^{-1}\mathbf{B}\,\mathbf{w}^{\mathrm{qs}}$. Because the mean vector μ of $\mathbf{x}$ is equal to $\mathbf{G}^{-1}\mathbf{F}^t\mathbf{y}$ [see Eq. (6a)], the covariance matrix $(\mathbf{x} - \mu)(\mathbf{x} - \mu)^t$ is simply $\mathbf{G}^{-1}\mathbf{B}\mathbf{w}^{\mathrm{qs}}(\mathbf{w}^{\mathrm{qs}})^t\mathbf{B}^t(\mathbf{G}^{-1})^t$. Because $\mathbf{w}^{\mathrm{qs}}$ is a Wiener vector, the product $\mathbf{w}^{\mathrm{qs}}(\mathbf{w}^{\mathrm{qs}})^t$ is the identity matrix. Moreover, as $\mathbf{G} = \mathbf{G}^t$ and consequently $(\mathbf{G}^{-1})^t = \mathbf{G}^{-1}$, covariance matrix $(\mathbf{x} - \mu)(\mathbf{x} - \mu)^t$ reduces to $\mathbf{G}^{-1}\mathbf{B}\mathbf{B}^t\mathbf{G}^{-1}$. Clearly, if $\mathbf{B}$ satisfies Eq. (8), the covariance matrix becomes $T/2\,\mathbf{G}^{-1}$, which is exactly the desired covariance matrix $\mathbf{C}$ of Eq. (6b).

Appendix B: Demonstration of Eqs. (18) and (19)

We obtain the steady state of the network of Fig. 12 by writing Kirchhoff's current law. At node i we have

$$\sum_{m=1}^{n} \Lambda_{im} V_{\mathrm{m}} = J_i + q_i, \tag{B1}$$

where the total random current q_i is equal to $\Sigma_{m=1}^{n}\,\theta_{mi}$ and Λ_{im} equals $-(1/R_{im})$ for $m \neq i$ and $\Sigma_{p=1}^{n}\,1/R_{ip}$ for $m = i$. Note that, in Eq. (B1), capacitances at all nodes are not incorporated because they participate in the transient response and not in the steady state. This amounts to considering that the correlation time of the Gaussian noise is much longer than the RC response time of the network. In compact form, Eq. (B1) can be written as $\Lambda\mathbf{V} = \mathbf{J} + \mathbf{q}$, where Λ is a real symmetric matrix formed by the coefficients Λ_{im} and $\mathbf{V}$, $\mathbf{J}$, and $\mathbf{q}$ are $n \times 1$ vectors with coefficients V_i, J_i, and q_i. Let us denote by $\mathbf{Q}$ the covariance matrix of the random variables q_i. From Eq. (17) and noting that $\theta_{ij} = -\theta_{ji}$, we find from simple linear algebra that

$$\mathbf{Q} = (T/2)\Lambda. \tag{B2}$$

From Eq. (B1) it follows that the potentials $\mathbf{V}$ obey a multinormal distribution with mean vector $\boldsymbol{\mu} = \Lambda^{-1}\mathbf{J}$ and covariance matrix $(T/2)\Lambda^{-1}$. Thus the distribution function $P(\mathbf{V})$ of the potentials is given by

$$P(\mathbf{V}) = \frac{1}{Z'} \exp\left[-\frac{1}{T} (\mathbf{V} - \mu)^t \Lambda (\mathbf{V} - \mu) \right], \tag{B3}$$

where Z' is a normalizing constant and the superscript t denotes matrix transposition. It is easily shown that the distribution function of Eq. (B3) is equal to that of Eq. (18) for $J(\mathbf{V})$ given by Eq. (19).

We thank our colleagues A. Dupret and E. Belhaire of the Institut d'Electronique Fondamentale, Centre National de la Recherche Scientifique (CNRS), Orsay, and P. Garda of the Université Pierre et Marie Curie, Paris, who designed for us the analog part of the integrated circuit. Most of the experimental work was carried out by Jean Claude Rodier at the Institut d'Optique Théorique et Appliquée and was supported by the Direction de la Recherche et des Etudes Techniques under contract DRET#92-139. The algorithmic contribution to this study benefitted from the expertise of Line Garnero of the Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Centre National de la Recherche Scientifique, Hôpital de la Salpétrière, Paris.

References

1. T. Poggio and C. Koch, "Ill-posed problems in early vision: from computational theory to analogue networks," Proc. R. Soc. London Ser. B **226,** 303–323 (1985).
2. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6,** 721–741 (1984).
3. B. Chalmond, "Image restoration using an estimated Markov model," Signal Process. **15,** 115–129 (1988).
4. P. Bouthemy and E. François, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," Int. J. Comput. Vision **10,** 157–182 (1993).
5. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220,** 671–680 (1983).
6. C. Koch, J. Marroquin, and A. Yuille, "Analog neuronal networks in early vision," Proc. Natl. Acad. Sci. USA **83,** 4263–4267 (1986).
7. D. W. Murray, A. Kashko, and H. Buxton, "A parallel approach to the picture restoration algorithm of Geman and Geman on an SIMD machine," Image Vision Comput. **4,** 133–142 (1986).
8. J. G. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," Int. J. Comput. Vision **4,** 211–222 (1990).
9. A. Lumsdaine, J. L. Wyatt, and I. M. Elfadel, "Nonlinear analog networks for image smoothing and segmentation," J. VLSI Signal Process. **3,** 53–68 (1991).
10. L. Bedini and A. Tonazzini, "Image restoration preserving discontinuities: the Bayesian approach and neural networks," Image Vision Comput. **10,** 108–118 (1992).
11. P. C. Yu, S. J. Decker, H. S. Lee, C. G. Sodini, and J. L. Wyatt, "CMOS resistive fuses for image smoothing and segmentation," IEEE J. Solid-State Circ. **27,** 545–553 (1992).
12. F. C. Jeng and J. W. Woods, "Compound Gauss–Markov random fields for image estimation," IEEE Trans. Signal Process. **39,** 683–697 (1991).
13. M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," Proc. IEEE **76,** 869–889 (1988).
14. A. Blake and A. Zisserman, *Visual Reconstruction* (MIT, Cambridge, Mass., 1987), Chap. 4.
15. P. Charbonier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," IEEE Trans. Image Process. **6,** 298–311 (1997).
16. J. Marroquin, "Surface reconstruction preserving discontinuities," A. I. Memo 792 (MIT, Cambridge, Mass., 1984).
17. D. Vanderbelt and S. G. Louie, "A Monte Carlo simulated annealing approach to optimization over continuous variables," J. Comput. Phys. **56,** 259–271 (1984).
18. M. A. Styblinsky and T. S. Tang, "Experiments in nonconvex optimization: stochastic approximation with function smoothing and simulated annealing," Neural Networks **3,** 467–483 (1990).
19. R. L. Stratonovitch, *Markov Processes and Related Processes*, Vol. 1 of Topics in the Theory of Random Noise (Gordon & Breach, New York, 1967), Chap. 4.
20. C. W. Gardiner, "The Ito calculus and stochastic differential equations," in *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1990), Chap. 4, pp. 106–113.
21. D. Geman and C. Yang, "Nonlinear image recovery with half-

quadratic regularization," IEEE Trans. Image Process. **4,** 932–945 (1996).
22. D. Prévost, Ph. Lalanne, L. Garnero, and P. Chavel, "Quasi-static algorithm for image restoration preserving discontinuities," in *Neural and Stochastic Methods in Image and Signal Processing III*, S.-S. Chen, ed., Proc. SPIE **2304,** 156–164 (1994).
23. E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (Wiley, New York, 1989), Chap. 3, pp. 33–52.
24. J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," J. Am. Statist. Assn. **82,** 76–89 (1987).
25. Ref. 23, Chap. 4, pp. 57–75.
26. D. Geman, S. Geman, C. Graffine, and P. Dong, "Boundary detection by constrained optimization," IEEE Trans. Pattern Anal. Mach. Intell. **12,** 609–628 (1990).
27. J. Alspector, J. W. Gannett, S. Haber, M. B. Parker, and R. Chu, "A VLSI-efficient technique for generating multiple uncorrelated noise sources and its application to stochastic neural networks," IEEE Trans. Circ. Syst. **38,** 109–123 (1991).
28. A. J. Martino and G. M. Morris, "Optical random number generator based on photoevents locations," Appl. Opt. **30,** 981–989 (1991).
29. G. M. Morris, "Optical computing by Monte Carlo methods," Opt. Eng. **24,** 86–90 (1985).
30. J. Marron, A. J. Martino, and G. M. Morris, "Generation of random arrays using clipped laser speckle," Appl. Opt. **25,** 26–30 (1987).
31. A. Dupret, E. Belhaire, and P. Garda, "Scalable array of Gaussian white noise sources for analogue VLSI implementation," Electron. Lett. **31,** 1457–1458 (1996).
32. S. Wolfram, "Random sequence generation by cellular automata," Adv. Appl. Math. **7,** 123–169 (1986).
33. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985), Chap. 4.
34. Ph. Lalanne, H. Richard, J. C. Rodier, P. Chavel, J. Taboury, K. Madani, P. Garda, and F. Devos, "2D generation of random numbers by multimode fiber speckle for silicon arrays of processing elements," Opt. Commun. **76,** 387–394 (1990).
35. Ph. Lalanne, E. Belhaire, J. C. Rodier, A. Dupret, P. Garda, and P. Chavel, "Gaussian random number generation by differential detection of speckles," Opt. Eng. **34,** 1835–1837 (1995).
36. G. Prémont, Ph. Lalanne, P. Chavel, M. Kuijk, and P. Heremans, "Generation of sigmoid probability functions by clipped differential speckle detection," Opt. Commun. **129,** 347–352 (1996).
37. K. H. Fisher and J. A. Hertz, *Spin Glasses*, D. Edwards, ed. (Cambridge U. Press, Cambridge, 1991).
38. A. J. Gray, J. W. Kay, and D. M. Titterington, "On the estimation of noisy binary Markov random fields," Pattern Recogn. **25,** 749–768 (1992).
39. A. Dupret, E. Belhaire, J. C. Rodier, Ph. Lalanne, D. Prévost, P. Garda, and P. Chavel, "An optoelectronic CMOS circuit implementing a simulated annealing algorithm," IEEE J. Solid-State Circ. **31,** 1046–1050 (1996).
40. W. Millar, "Some general theorems for non-linear systems possessing resistance," Phil. Mag. **42,** 1150–1160 (1951).
41. J. Harris, C. Koch, and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision," Science **248,** 1209–1211 (1990).
42. J. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," Int. J. Comput. Vision **4,** 211–223 (1990).
43. A. Lumsdaine, J. L. Wyatt, and I. M. Elfadel, "Nonlinear analog networks for image smoothing and segmentation," J. VLSI Signal Process. **3,** 53–68 (1991).
44. H. Nyquist, "Thermal agitation of electric charge in conductors," Phys. Rev. **32,** 110–113 (1928).
45. H. Kobayashi, J. L. White, and A. A. Abidi, "An active resistor network for Gaussian filtering of images," IEEE J. Solid-State Circ. **26,** 738–747 (1991).
46. R. A. Nordin, A. F. J. Levi, R. N. Nottenburg, J. O'Gorman, T. Tanbun-Ek, and R. A. Logan, "A systems perspective on digital interconnection technology," J. Lightwave Technol. **10,** 811–827 (1992).
47. J. Besag, "On the statistical analysis of dirty pictures," J. R. Statist. Soc. B **48,** 259–302 (1986).