# Pixel level data fusion: from algorithm to chip

Bimal P. Mathur, H. Taichi Wang, Shih-Chi Liu, Christof Koch, Jin Luo

Bimal P. Mathur, H. Taichi Wang, Shih-Chi Liu, Christof Koch, Jin Luo, "Pixel level data fusion: from algorithm to chip," Proc. SPIE 1473, Visual Information Processing: From Neurons to Chips, (9 July 1991); doi: 10.1117/12.45548

SPIE.

Event: Orlando '91, 1991, Orlando, FL, United States

# Pixel Level Data Fusion: From Algorithm to Chip

Bimal Mathur, H. Taichi Wang, Shih Chi Liu
Rockwell International Science Center, Thousand Oaks, CA 91360

and

Christof Koch, Jin Luo
California Institute of Technology, Pasadena, CA

## ABSTRACT

Pixel level image processing algorithms have to work with noisy sensor data to extract spatial features. This often requires the use of operators which amplify high frequency noise. One method of dealing with this problem is to perform image smoothing prior to any use of spatial differentiation. Such spatial smoothing results in the spread of object characteristics beyond the object boundaries. Identification of discontinuities and explicit use of these as boundaries for smoothing has been proposed as a technique to overcome this problem. This approach has been used to perform cooperative computations between multiple descriptions of the scene, e.g., fusion of edge and motion field for a given scene. We have extended this approach to multisensor systems. We have used the discontinuities detected in the output of one sensor to define regions of smoothing for a second sensor. For example, the depth discontinuities present in laser radar can be used to define smoothing boundaries for infrared focal plane arrays. We have recently developed a CMOS chip ($28 \times 36$) which performs this task in real time. This chip consists of a resistive network and elements that can be switched ON or OFF, by loading a suitable bit pattern. The bit pattern for the control of switches can be generated from the discontinuities found in the output of sensor #1. The output of sensor #2 is applied to the resistive network for data smoothing. If all the switches are held in conducting state, this chip performs the usual data smoothing. However, if switches along object boundaries are turned OFF, a region for bounded smoothing is created. In our chip, information from a third sensor data (e.g., intensity data from laser radar) can be incorporated in the form of a map of "confidence in data." The results obtained with this chip using synthetic data and other potential applications of this chip are described.

## INTRODUCTION

To deal with the visual environment, the human vision system has developed a computational strategy that relies on a multitude of parallel processes; each of these processes by itself produces a noisy and often ambiguous description of the scene. However, our visual precept is neither noisy nor ambiguous; this is believed to be achieved by the process of fusion of information. As an example, let us consider the computation of motion field from images collected by a single camera. Images shown in Figs. 1(a) and 1(b) were captured a short time apart. A commonly used algorithm[1] for the computation of motion field produces the result shown in Fig. 1(d). This clearly produces a qualitatively correct result, i.e., we can conclude that there is an object moving to right. However, the computed motion field spills well beyond the actual object boundary, and the object definition is lost. If we use one of the edge maps, Fig. 1(c), to set up a boundary of the moving object and perform the optical flow computation within this region, a well-defined optical flow field is obtained. This type of "pixel level fusion," based on "discontinuities," can be carried out between any suitable pair of descriptions of the world, e.g., edge map and depth map or edge map and velocity field.

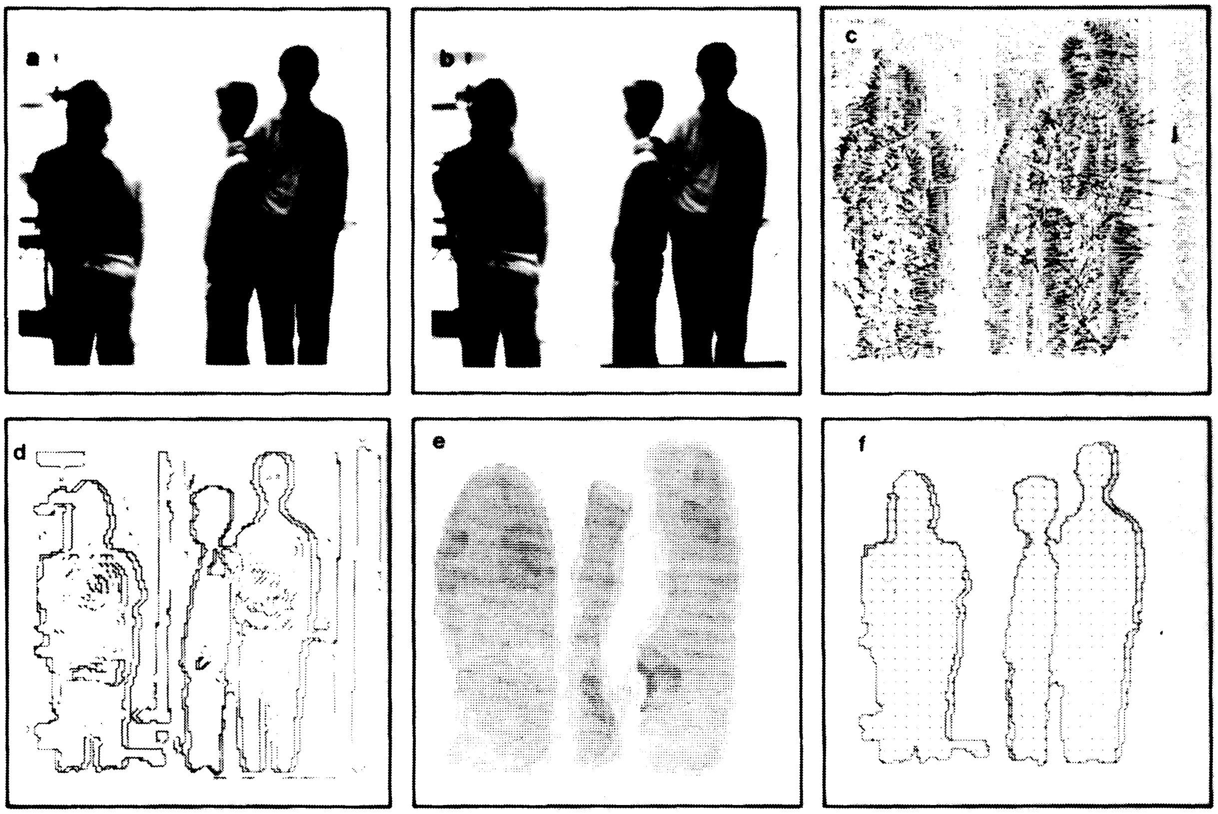


Fig. 1 Fusion of edge map with motion computation improves the computed motion field. (a) and (B) are the input images. (c) Shows the edge of the image. (d) Without fusion with edge. smoothing results in smearing of object boundaries. (e) Use the edge map in (c) to set the boundaries for smoothing results in a much improved and well-defined motion field over the object area.

In this paper, we describe the extension of this type of discontinuity-based data fusion techniques to multi-sensor data processing systems. Our approach is based identifying the location of "discontinuities" in images produced by multiple sensors, and then using this information for data fusion. We assume the sensor data are spatially registered, either through the design of the sensor system or by other registration processes. In the following, we will briefly review some surface interpolation computational models to illustrate the role of discontinuities and then describe our approach to multisensor data processing problems. We will then describe our circuit implementation of this methodology and the performance of the chip.

## INTERPOLATION IN THE PRESENCE OF DISCONTINUITIES

During the past few years, a framework for mathematical modelling of early vision processes,[2,3] has been developed. In these models, early vision processes are cast as energy minimization problems. We have chosen to work with these because the energy minimization problem can be easily mapped to analog VLSI circuits. The basic assumption behind these models is that "Most physical objects occupy a continuous region in space and their observable properties tend to be continuous functions over the extent of the object." Thus, interpolation in the presence of discontinuities, subject to the constraint of parse measured data, is at the heart of many early vision algorithms. We will concentrate on the generic problem of surface interpolation to bring out the essence of this approach, and its relation with sensor data fusion.

In the problem of surface interpolation, we look for a function u which is in reasonable agreement with the available measured data, d, and which yields a low value for a certain smoothness measure:

$$\left(\frac{\partial^p u}{\partial x^p}\right)^2$$

The solution to the surface interpolation problem, after converting from continuous to discrete domain, can be formulated as the minimum of a total energy functional, E, of the form of p = 1:

$$E = \sum_i g_i (u_i - d_i)^2 + \lambda \sum_i (u_{i+1} - u_i)^2 \qquad (1)$$

where $\lambda$ (a constant) is the regularization parameter, and $g_i$ is the confidence level of the data point. This functional, depending on the value of $\lambda$, establishes a balance between the need to be consistent with measured data and the degree of smoothness (specified by the order of the derivative used, p). The selection of the "order of derivative" depends on the nature of the data "d", and the type of information that has to be derived from u. A membrane potential function, p = 1, and the plate potential function, p = 2, are commonly used.

A major drawback of all algorithms which rely on smoothing to reduce noise and to fill in missing data is that these result in smearing of boundaries; i.e., the final solution for u extends beyond the object boundaries. A powerful method to deal with this situation was proposed by Geman and Geman,[4] who introduced the concept of binary line processes which explicitly code for the presence of discontinuities. Koch et al.[5,6] have applied this model of discontinuities to problems encountered in early vision processes.

In surface interpolation, this requires the introduction of additional terms in the energy functional to account for discontinuities. Let l stand for a discontinuity in the final field (f). If the spatial gradient of S is greater than a certain threshold, then l is set to 1; i.e., the discontinuity is "turned on", otherwise it is set to 0; i.e., discontinuity is "off". We can rewrite Equation (1) as follows:

$$E = \sum_i g_i (u_i - d_i)^2 + \gamma \sum_i (1 - l_i)\ (u_{i+1} - u_i)^2 + \alpha \sum_i l_i \qquad (2)$$

The second term causes image smoothing at those locations where $l_i = 0$. The third term is included to impose a "cost," $\alpha$, for introducing a discontinuity. When the discontinuities are turned on at the location of object boundaries, bounded smoothing of data d is performed by the process of minimization of E with respect to u and l.

## AN ALGORITHM FOR PIXEL LEVEL DATA FUSION

In the case of surface interpolation, the binary line processes are set based on the local gradient of the surface itself. The discontinuities can be used to fuse different descriptions of the world from the same sensor, e.g., edges detected from intensity distribution, to define region of smoothing for another map, e.g., surface profile, or motion field, as is shown in Fig. 1. For a laser radar system, we can use this technique to fuse the intensity map with depth map. We have extended this simple, yet powerful strategy, one step further to handle multisensor data fusion. In our algorithm, the two sensors are chosen to provide complementary information. For example we can pick laser radar to provide range data and IR Focal Plane Array (IR FPA) for providing temperature data. The laser radar boundaries can be fused with the IR image by performing the following operations:

(1) Identify boundaries in the laser data,
(2) Use these boundaries as line processes or smoothing boundaries for IR FPA data.
(3) Perform data smoothing on the FPA data for noise reduction and spreading of information.
(4) Read out fused image.

For tactical scenerios, once regions of interest are established (detection of targets), e.g., with a mm-wave radar, we can scan these regions with a laser radar and a high resolution IR FPA. This typically produces target images which are 10-40 pixels in size. We can now use the methodology described above to get a fused description of the target. This process produces a description which has sharp object boundaries and a spatial temperature distribution over the object. This algorithm generates a fused image which has boundaries detected in first sensor and the analog data from second sensor. This description is much richer than what can be achieved by either sensor alone. This methodology is not limited to Laser Radar and FPA alone; it can be applied to other sensor combinations.

We have demonstrated a data fusion chip which implements steps (2)-(4) of the data fusion algorithm described above. This $28 \times 36$ pixel chip performs this computation in real time. Before we present the results of this technique for pixel level sensor fusion, we will first describe some of the implementation issues.

## MAPPING ALGORITHM TO CHIP

This section illustrates the methodology we have used to map the algorithm to a chip. As pointed out, the primary reason for selecting an energy minimization framework is the ease with which these can be mapped to parallel analog circuits. Let us consider the mapping of Equation (1) to hardware:

To ensure that E is minimized by a network we need to have $dE/dt < 0$. We can write $dE/dt = dE/du * du/dt$, and set $dE/du = -du/dt$. This condition gives rise to a network equation, in which S is identified as the node voltages. In this form, the regularization terms appear as connection pattern for each pixel. The network derived from the equation given above is shown in Fig. 2. These networks are a special case of the neural network model proposed by Hopfield.[7] This circuit will settle to a state which will minimize the function E. The steady state node voltages V will be the solution to the problem posed above.

The smoothing part of the computation is implemented as resistive networks. The line processes on the other hand can be visualized as switches which are turned OFF to represent the presence of a discontinuity in the function u. The off state of this switch limits the smoothing process and confines it to the region occupied by the object. For the network shown in Fig. 2 all connections which produce smoothing at a pixel must be disconnected to implement this binary process. To simplify our designs of the network, we have used the membrane potential, $p = 1$, in which the interconnection pattern is limited to nearest neighbors. In this case, each line process can be visualized as a switch in series with a smoothing resistor.

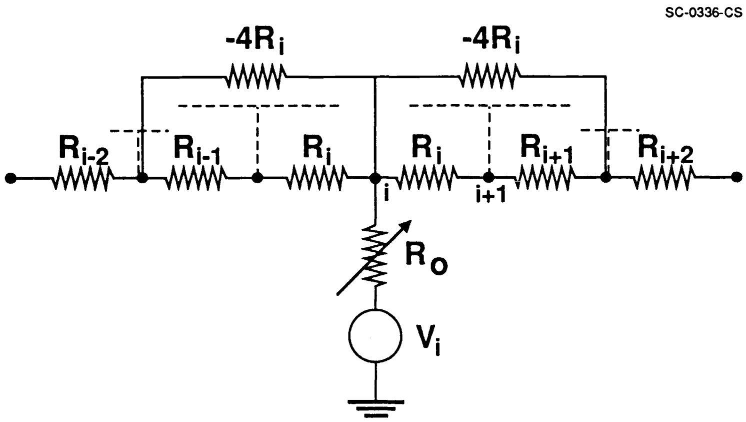


Fig. 2 A resistive network for surface interpolation.

## A CMOS DATA FUSION CHIP

We have designed and tested this pixel level data fusion chip using standard CMOS technology. This chip includes a grid of resistors for data smoothing and an array of switches which implement line processes. The switches are controlled by data loaded into an array of SRAM cells. For a system using IR imager and Laser Radar, we can use the depth discontinuities detected in laser radar data to control the state of the switches. These data can be generated by applying "edge detection algorithms" to laser radar data, and then supplying these to the chip as an array of binary images. A coboresighted IR image can now be fused with the laser radar discontinuity map, by supplying an IR image to the resistive network. The switches define regions for IR data smoothing, and a fused (and smoothed) version of the IR image can be read out, by reading network node voltages.

The schematic diagram for each pixel of our chip is shown in Fig. 3. The resistive grid (Rh) is formed by an array of MOSFETs (M1) that are biased in the subthreshold region. The IR data is supplied through an OpAmp A1. This OpAmp is biased such that it has a finite output resistance (Rv). The IR data is supplied through amplifier A1. The spatial extent of smoothing operator is controlled by the ratio of Rh and Rv. In our chip, Rh is controlled by a global bias, and Rv can be controlled on a pixel by pixel basis, i.e., this bias voltage is scanned and stored on the chip as an image. Each resistor in the smoothing grid, Rh, has a switch Sh (and Sv for two-dimensional case) in series with it. These switches are set by loading a bit in the latch, Lh (and Lv). Each node has an output amplifier A2, which is directly connected to the nodes of the resistive network, and is used to read out the processed version of the image. A two-dimensional, 28 × 36, array of this circuit designed with 2 ) rules occupies 4.5 mm × 6.7 mm of silicon. This was fabricated by DARPA's MOS implementation service.

## PERFORMANCE OF THE CHIP

We have tested this chip with synthetic images. The IR image was constructed by defining an object and assigning a voltage of 1.5 V to some of the pixels, and 1 V to the rest of the object. To this, we have added a random noise in the range of -1.5 to +1.5 V to the image. The composite picture is shown in Fig. 4. A coboresighted, laser radar image is assumed to exist. From this image, we have derived a noisy edge map (shown in Fig. 5) using an edge detection algorithm. If we process the IR image without the laser radar data, a blurred image of the

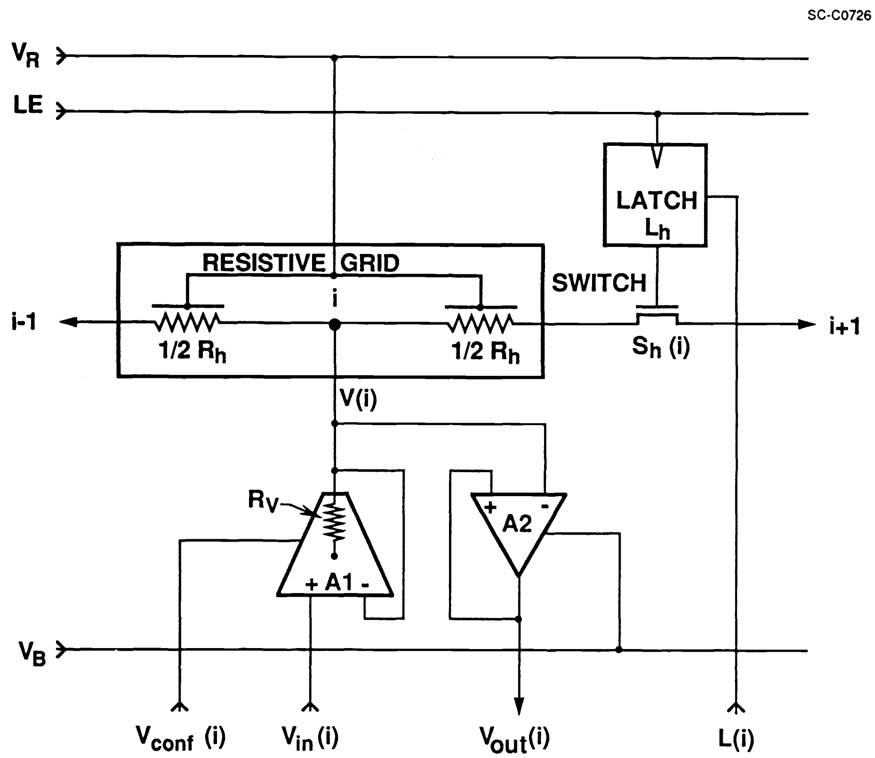


Fig. 3 A schematic diagram of the chip.

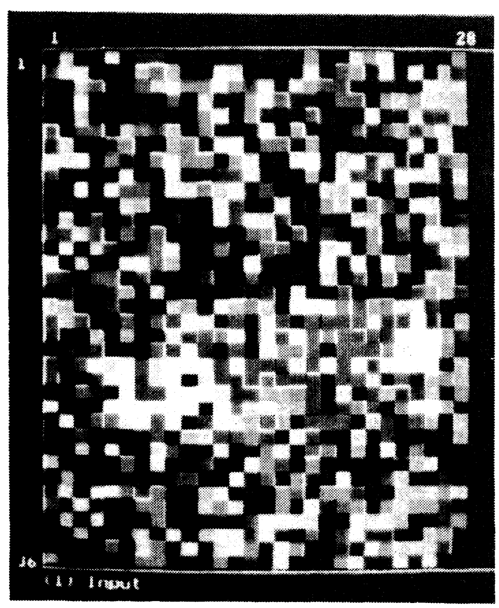

Fig. 4 Synthestic IR image.

object is produced as shown in Fig. 6. This is because the spatial smoothing operation required to get rid of spatial noise produces a spreading of the target signature. If the laser radar discontinuities are supplied to the chip, these set up boundaries for data smoothing which

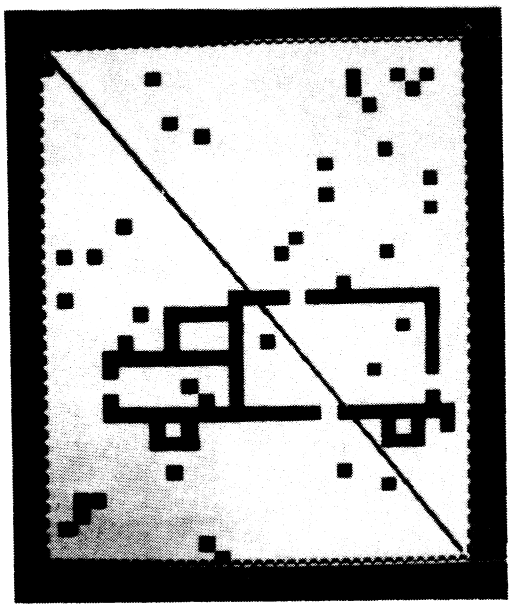

Fig. 5 Edges detected from a noisy laser radar image.

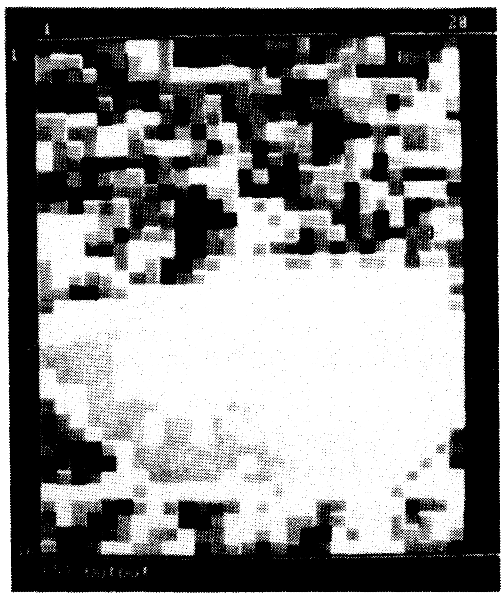

Fig. 6 IR image processed without fision.

restrict the spread of IR signature. As shown in Fig. 7, this operation produces a well-defined IR image of the target. Since this fused image combines the information about object boundaries and corners as contained (LADAR data) and the temperature distribution (IR data), the fused image provides a richer description of the scene, suitable for further feature extraction.

This 28 × 36 pixel chip should perform this computation in real time. The scan registers of this chip can be easily clocked at 5 MHz. So the read in and read out times for the two-dimensional array is less than 1 msec each. The settling time for the chip is expected to be

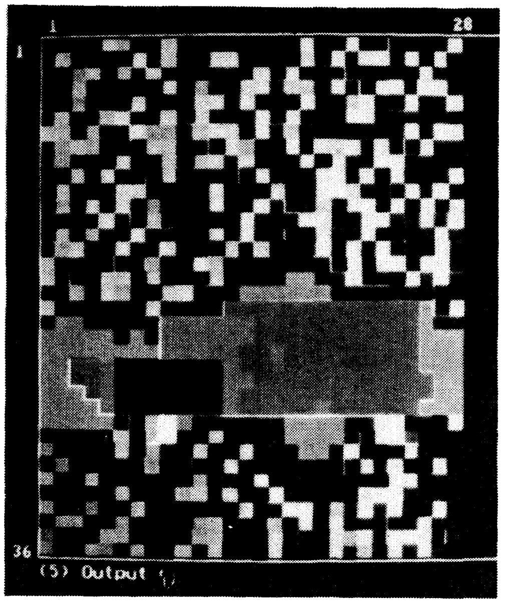

Fig. 7 IR image processed with edges detected in laser radar image.

1 msec, thus this chip should easily perform data fusion at TV frame rates. We are in the process of characterizing the chip at these frame rates. This chip operates with a 5 V power supply, and uses four bias voltage to set the operating points for analog ports of the circuit. During operation this chip draws less than 10 mA of current , i.e., it dissipate less than 50 mW of power.

## SUMMARY

We have developed an algorithm for multisensor data fusion at the pixel-level. This algorithm has been implemented in analog network using standard CMOS technology.

## REFERENCES


1. Horn, B.K.P. and Schunck, B.G., "Determining Optical Flow," Artif. Intell. 17, 141-184 (1981).
2. Poggio, T., Gamble, E.B., and Little, J.J., "Parallel Integration of Vision Modules," Science 242, 436-439 (1988).
3. Poggio, T. and Koch, C., "Ill-Posed Problems in Early Vision: from Computational Theory to Analog Networks", Proc. R. Soc. Lond. B 226, 303-323 (1985).
4. Geman, S., and Geman, D., "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," IEEE Trans. Pattern Anal. Machine Intell. PAMI-5, 721-741 (1984).
5. Poggio, T., Torre, V., and Koch, C., "Computational Vision and Regularization Theory," Nature 317, 314-319 (1985).
6. Harris, J.G., Koch, C., Luo, J., and Wyatt, J., "Resistive Fuses: Analog Hardware for Detecting Discontinuities in Early Vision", in Analog VLSI Implementation of Neural Systems, Kluwer Academic Publishers, 1989.
7. Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proc. Natl. Acad. Sci. USA 79, 2554-2558, (1982).