11. M. Delmar *et al.*, *Circ. Res.* **60**, 780 (1987).
12. A. Lázár, Z. Noszticzius, H. Farkas, H.-D. Försterling, *Chaos* **5**, 443 (1995); A. Lázár, Z. Noszticzius, H.-D. Försterling, Zs. Nagy-Ungvárai, *Physica D* **84**, 112 (1995).
13. T. Yamaguchi, L. Kuhnert, Zs. Nagy-Ungvarai, S. C. Müller, B. Hess, *J. Phys. Chem.* **95**, 5831 (1991).
14. Solutions of the catalyst Fe(batho)$_3^{2+}$ were prepared by dissolving 50.0 mg of 4,7-diphenyl-1,10-phenanthroline (known as bathophenanthroline) and 31.4 mg of $FeSO_4 \cdot 7\ H_2O$ in 10.0 ml of glacial acetic acid. Approximately 5 ml of the filtered catalyst solution (filter pore size, 0.45 μm) were pipetted into a clean, dry printer cartridge (HP51626A), which was resealed. The cartridge was used with a Hewlett Packard DeskJet 520 to print the catalyst patterns onto the polysulphone membranes (Gelman Supor-450).
15. The gel was prepared by mixing 10 ml of liquid agar (Agar-Agar; Fluka, New York) with 10 ml of BZ solution (without catalyst) in a thermostated Petri dish (diameter, 8.5 cm) maintained at 25.0 ± 0.2°C. Resulting initial composition: 1.5% agar, [$NaBrO_3$] = 0.26 M, [malonic acid] = 0.17 M, [NaBr] = 0.10 M, and [$H_2SO_4$] = 0.31 M (Fig. 1) or 0.39 M (Fig. 3).
16. J. J. Tyson and P. C. Fife, *J. Chem. Phys.* **73**, 2224 (1980); J. P. Keener and J. J. Tyson, *Physica D* **21**, 307 (1986).
17. Functions $f$ and $g$ for the catalyst-loaded areas are defined as $f(u, v) = (1/\epsilon)[u - u^2 - fv(u - q)/(u + q)]$ and $g(u, v) = u - v$. Catalyst-free regions were described by Eq. 1 with $f(u, v) = (1/\epsilon)(-u^2)$. The Laplacian was approximated by a nine-point formula; parameters: $\epsilon = 0.05$ and $q = 0.002$.
18. The retarded decay of $HBrO_2$ in catalyst-free regions is an intrinsic feature of the model and is consistent with the Field-Körös-Noyes mechanism of the BZ reaction [R. J. Field, E. Körös, R. M. Noyes, *J. Am. Chem. Soc.* **94**, 8649 (1972)]. It reflects the absence of the catalyzed production of the inhibitor bromide in these regions. One-dimensional simulations with the three-variable Oregonator model [R. J. Field and R. M. Noyes, *J. Chem. Phys.* **60**, 1877 (1974)] showed the same qualitative behavior.
19. J. Jalife, *Pacing Clin. Electrophysiol.* **6**, 1106 (1983).
20. The catalyst-free strips were 0.5 mm in width and the experimental conditions were the same as in Fig. 1 except [$H_2SO_4$] = 0.39 M. Gap penetration by BZ waves has also been observed in an immobilized-catalyst sol-gel system [I. R. Epstein, I. Lengyel, S. Kádár, M. Kagan, M. Yokoyama, *Physica A* **188**, 26 (1992)].
21. The length scale for the calculation corresponds to an approximately threefold enlargement except for the dot-density gradient. This scale was required to achieve sufficient accuracy in feasible computational times. The number of dots per wavelength in simulation and experiment is approximately the same.
22. A. T. Winfree, *J. Theor. Biol.* **138**, 353 (1989).
23. M. Gómez-Gesteira *et al.*, *Physica D* **76**, 359 (1994).
24. G. K. Moe, W. C. Rheinboldt, J. A. Abildskov, *Am. Heart J.* **67**, 200 (1964).
25. K. Agladze *et al.*, *Science* **264**, 1746 (1994); M. Gómez-Gesteira *et al.*, *Phys. Rev. E* **50**, 4646 (1994); A. Babloyantz and J. A. Sepulchre, *Physica D* **49**, 52 (1991); J. A. Sepulchre and A. Babloyantz, in *Chemical Waves and Patterns*, R. Kapral and K. Showalter, Eds. (Kluwer, Dordrecht, Netherlands, 1995), pp. 191–217.
26. The experimental system (Fig. 3A) evolved to the asymptotic behavior in only a few oscillations, whereas the simulated system (Fig. 3, B and C) required many cycles; this difference is likely the result of a greater role played by external fluctuations and imperfections in the experimental system.
27. J. Maselko and K. Showalter, *Physica D* **49**, 21 (1991).
28. F. Mertens and R. Imbihl, *Nature* **370**, 124 (1994).


29. We thank Z. Noszticzius and V. Gáspár for advice on loading polysulphone membranes with bathoferroin. We are grateful to the National Science Foundation (CHE- 9222616), the Office of Naval Research (N00014-95-1-0247), and the Petroleum Research Fund (29565-AC6) for supporting this research. O.S. thanks the Fonds der Chemischen Industrie for a Liebig Fellowship.

22 May 1995; accepted 10 July 1995


# Replicator Neural Networks for Universal Optimal Source Coding


Robert Hecht-Nielsen



**Replicator neural networks self-organize by using their inputs as desired outputs; they internally form a compressed representation for the input data. A theorem shows that a class of replicator networks can, through the minimization of mean squared reconstruction error (for instance, by training on raw data examples), carry out optimal data compression for arbitrary data vector sources. Data manifolds, a new general model of data sources, are then introduced and a second theorem shows that, in a practically important limiting case, optimal-compression replicator networks operate by creating an essentially unique natural coordinate system for the manifold.**


The first serious studies of replicator neural networks were carried out by Kohonen (*1*) during the 1970s. Ackley, Hinton, and Sejnowski (*2*) later considered the "encoder problem" for Boltzmann machines. In 1986, Cottrell, Munro, and Zipser (*3*) developed a replicator network version of the multilayer perceptron neural network (*4*). This report focuses on a three hidden layer version of this last architecture. In the following, a data vector is defined as a point $\mathbf{x} = (x_1, x_2, \ldots, x_m)'$ in $\Re^m$ (for example, the brightnesses of pixels in an $m$-pixel image tile or $m$ regular time samples of a sound stream). A source is a statistical process that supplies data vectors that belong to a compact measurable subset $A$ of $\Re^m$, where the a priori probability of a data vector $\mathbf{x}$ being supplied is governed by a regular probability distribution $\mu$ (*5*) on $\Re^m$ with $\mu(A) = 1$ (a distribution defined on $A$). The case where $m$ is large (say, $m > 1000$) is particularly interesting because many-dimensional data vectors allow high compression ratios.

Although the term replicator neural network is intended to apply to all autoassociative neural networks with compressed internal representations, for simplicity it will be used primarily in the following restricted sense. A replicator neural network is a multilayer perceptron with three hidden layers (see Fig. 1) that meets the following conditions [using the notation of (*6*)]: The input and output layers (layers 1 and 5) of this network each have $m$ units. The inputs to layer 1 (the input layer) are the components of the data vector $\mathbf{x}$ and the outputs of layer 5 are the components of $\mathbf{x}'$ (intended to be an estimate of $\mathbf{x}$). The output $z_{ki}$ of unit $i$ of layer $k$ ($k = 2$, 3, 4, and 5) of the network is given by $s_k(I_{ki})$, where

$$I_{ki} = \sum_{j=0}^{L_{k-1}} w_{kij} z_{(k-1)j} \qquad (1)$$

$w_{kij}$ is the weight of the $i$th unit of layer $k$ that multiplies the output $z_{(k-1)j}$ of unit $j$ ($j = 0, 1, \ldots, L_k$) of the previous layer, and $L_k$ is the number of units on layer $k$. The output of the $i$th unit of layer 1 (the input layer) is $z_{1i} = x_i$ and the bias output $z_{k0} = 1.0$ for $k = 1$, 2, 3, and 4. The activation functions $s_k$ are given by $s_2(\theta) = s_4(\theta) = \tanh(\theta)$ and $s_5(\theta) = \theta$. The activation function for the middle hidden layer is given by

$$s_3(\theta) = \frac{1}{2} + \frac{1}{2(N-1)} \sum_{j=1}^{N-1} \tanh\left[a\left(\theta - \frac{j}{N}\right)\right] \qquad (2)$$

This smooth stairstep activation function with $N$ treadles essentially quantizes the vector of middle hidden layer outputs into $M = N^n$ grid points. As the stairstep gain $a \to \infty$ the middle hidden layer activation function approaches a discontinuous stairstep with exactly $N$ output values: 0, $1/(N-1)$, $2/(N-1)$, ..., $(N-2)/(N-1)$, 1 (Fig. 2). If both $a \to \infty$ and $N \to \infty$, then $s_3(\theta)$ becomes the ramp activation function equal to 0 for $\theta < 0$, $\theta$ for $0 \le \theta \le 1$, and 1 for $\theta > 1$.

A real-time coder with M codes for a data vector source is a pair of mappings $\alpha$ and $\beta$, with $\alpha$: $A \subset \Re^m \to \{1, 2, \ldots, M\}$ and $\beta$: $\{1, 2, \ldots, M\} \to \Re^m$. The compression mapping $\alpha$ maps each data vector $\mathbf{x}$ to its code $\alpha(\mathbf{x})$, an integer between 1 and M [the inverse image $\alpha^{-1}(i)$ of each code is assumed to be measurable]. The decompression mapping $\beta(i)$ maps each code back to a data vector. The mean-squared-error distortion of this coder is

$$D = \int_A |\mathbf{x} - \beta[\alpha(\mathbf{x})]|^2 \, d\mu(\mathbf{x}) \qquad (3)$$

Such a coder is optimal if it minimizes $D$ in comparison with all other M-code coders for the same source. Clearly, in the high-gain case, replicator networks can be viewed as real-time coders with $M = N^n$ codes.

Given any $\varepsilon > 0$, it is straightforward to show that, for real-time coding, we can re-


HNC Software Inc., 5930 Cornerstone Court, San Diego, CA 92121, USA, and Department of Electrical and Computer Engineering and Institute for Neural Computation, University of California, San Diego, La Jolla, CA 92093, USA.

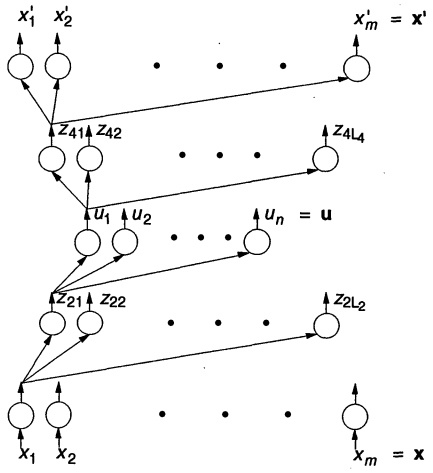


**Fig. 1.** A replicator neural network architecture with $m$ inputs.

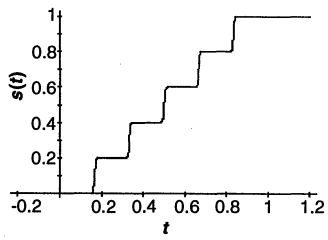


**Fig. 2.** The replicator network middle-hidden-layer unit activation function for $N = 6$ and $a = 200$.

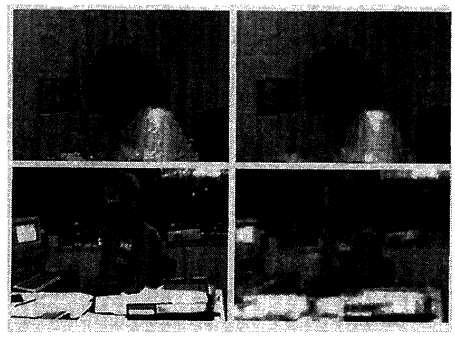

**Fig. 3.** Two panchromatic 192 pixel by 256 pixel test images (left) compressed and decompressed (right) by a partially trained replicator network. Images are first divided into disjoint 64-pixel by 64-pixel tiles. The 4096 eight-bit pixel brightnesses of each tile make up its **x** vector. These **x**'s are used to train or test the replicator network (the images shown were not used during training). The network used had $m = 4096$, $n = 40$, $N = 256$, and $L_2 = L_4 = 410$ (compression ratio = 102.4:1). After reconstruction, the boundaries between the tiles were deemphasized by a smoothing operation. This replicator network has 3,396,476 weights.

place a probability distribution $\mu$ for a source with a continuous probability density function $\rho$ with support $A$ such that the distortion $D$ of any real-time coder $\alpha$, $\beta$ for the source with distribution $\mu$ and the distortion $D'$ of the same coder applied to a source with density $\rho$ satisfies $|D - D'| < \varepsilon$. Thus, we need only consider sources with continuous probability density functions of compact measurable support. For such sources, real-time coders are equivalent to vector quantizers with M codebook vectors (*7*).

Despite more than 30 years of intense study, no practical method for constructing optimal or near-optimal vector quantizer coders for most real-world data vectors has been developed (*8*), although methods for vectors of independent Gaussian or Laplacian random variables do exist (*9*). The basic problem of developing vector quantizers for complicated sources is the construction of codebook vectors that conform to the geometry of the source. A codebook consisting of such a data-conformal set of points must be developed in some compact mathematically defined form to achieve near-optimal performance. (Codebooks consisting of lists of explicit vectors—for example, randomly chosen data vector examples or adaptively adjusted exemplars—might also work, but the required sizes of such codebooks are usually impractically large.) The following theorem shows that replicator neural networks provide a potentially practical means for constructing data-conformal codebooks for complicated data sources.

**Theorem 1.** *Given a data vector source and given any* $\varepsilon > 0$ *and integers* N *and* n *both greater than 1, then there exists a replicator neural network with* $M = N^n$ *codes that has a mean-squared-error distortion within* $\varepsilon$ *of that of an optimum* M*-code real-time coder for that source.*

A replicator network that is functioning as an optimal coder creates one codebook vector as its final output for each of the M combinatorial combinations of the $n$ middle hidden layer unit outputs (each of which takes on $N$ discrete values). In this way, we can manageably define a vast codebook of vectors that are appropriately placed within the data set.

A key question that arises is: Can we actually create optimal-compression replicator neural networks by means of training and architecture adjustment solely on the basis of data vector examples randomly chosen in accordance with $\rho$? What is required is to find a replicator network with $n$ middle hidden layer units (each taking on $N$ discrete stairstep output values) that comes close to minimizing the mean squared reconstruction error that can be obtained with members of this class of networks. In principle, this should be possible (*10*), although considerable additional research will probably be required to obtain practically useful results (see Fig. 3).

Optimal-compression replicator networks operate as follows. Let $\phi$: $[0,1]^n \rightarrow \mathbb{D}$ be a smooth ($C^\infty$) orientation-preserving diffeomorphism (*11*) of the unit cube $[0,1]^n \subset \Re^n$ onto $\mathbb{D} \subset \Re^m$ (typically $n \ll m$), where $\mathbb{D}$ has a fixed smooth nonzero probability density function $\sigma$ defined on it. If we define $\chi$: $[0,1]^n \rightarrow \Re$ by $\chi(\mathbf{u}) \equiv \det[d\phi(\mathbf{u})]$, then $\mathbb{D}$ is a data manifold if and only if $\chi(\mathbf{u}) = \sigma[\phi(\mathbf{u})]$ for all $\mathbf{u}$ in $[0,1]^n$ and is an entropy manifold if and only if

$$\chi(\mathbf{u}) = \frac{\sigma^{\frac{n}{n+2}}[\phi(\mathbf{u})]}{\int_{\mathbb{D}} \sigma^{\frac{n}{n+2}}(\mathbf{x})\, d\mathbf{x}} \qquad (4)$$

Given a point $\mathbf{x}$ on a data manifold (entropy manifold) $\mathbb{D}$, the natural coordinates (entropy coordinates) of $\mathbf{x}$ are the Cartesian coordinates of its preimage $\phi^{-1}(\mathbf{x})$ on the cube. (Natural coordinates are explained below.) Thus, a data manifold $\mathbb{D}$ is an object that can be constructed by taking an $n$-dimensional cube with uniform probability density, embedded in $\Re^m$, and smoothly elastically deforming it without tearing. The probability density function $\sigma$ at a point $\mathbf{x}$ on $\mathbb{D}$ is defined by how much the "material" of the cube is elastically compressed or expanded in density at that point. An entropy manifold is essentially the same thing, except that the elastic compression at a point is proportional to the entropy fraction of Eq. 4. Because for most practical applications $n/(n + 2) \approx 1$, the distinction between data manifolds and entropy manifolds can usually be ignored. Figure 4 illustrates that data manifolds can serve as models for real-world data sources (this can be shown rigorously). Given their universality, only sources that are data (or entropy) manifolds will be considered below.

Most current practical methods of data compression for many dimensional data vectors focus on the construction of a coordinate system that is somehow fitted to the geometry of the data source. For example, local spatial frequency and wavelet compression methods for imagery (*12*) express an image as a linear combination of image basis vectors. Principal components analysis, invented by Gauss 200 years ago (*13*), is a method for constructing data source coordinate systems (the nonsmall principal axes of the covariance ellipsoid). A much studied open problem is to find the proper generalization of principal components analysis to non-Gaussian data sets. Although some interesting ideas have emerged [for example, the "principal curves" heuristic of Hastie and Stuetzle (*14*)], no general solution to this problem has been found. It is proposed that natural coordinates constitute this proper generalization.

Like principal components, natural coordinates depend only on the probability

structure of the data set (infinitesimal volumes constructed of equal infinitesimal natural coordinate increments always have equal probability, although their geometrical volume measures can differ enormously). However, unlike principal components (which are always Cartesian), natural coordinates follow the probabilistic geometry of the data manifold. Further, except for an obvious set of transformations (orientation-preserving cube isometries and probability-preserving data manifold automorphisms), it can be shown that natural coordinates are unique.

**Theorem 2.** *Given an infinite sequence of data manifolds* $\{\mathbb{D}_1, \mathbb{D}_2, \ldots, \mathbb{D}_k, \ldots\}$ *of monotonically increasing dimension* n, *and given, for each set of values* n, N, *and* $\varepsilon > 0$, *a corresponding sequence* $\{\mathbb{N}_1, \mathbb{N}_2, \ldots, \mathbb{N}_k, \ldots\}$ *of replicator networks, each configured to perform (when a is set sufficiently large) to within* $\varepsilon$ *mean squared error of a corresponding data manifold optimal coder with* $M = N^n$ *codes, then as* N, k, *and* $a \rightarrow \infty$ *and* $\varepsilon \rightarrow 0$ *the middle-hidden-layer outputs of* $\mathbb{N}_k$ *will approach natural coordinates for data manifold* $\mathbb{D}_k$.

Thus, in the limit of large problems ($n \rightarrow \infty$), high gain ($a \rightarrow \infty$), and large numbers of quantization levels ($N \rightarrow \infty$), optimal replicator networks compress data vectors by representing them in a natural coordinate system fitted to the data manifold from which the data vectors are drawn. In effect, optimal replicator networks carry out data compression by internally expressing the data vectors of a data manifold in terms of natural coordinates for that manifold.

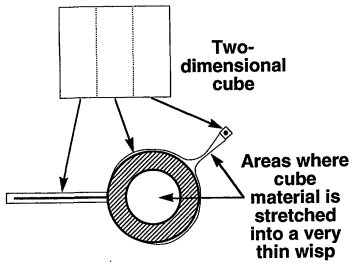


**Fig. 4.** An illustration of how an arbitrary probability distribution can be approximated by a data manifold (even one that has nonsimply connected regions and elements of different dimensionalities). In this example the distribution is composed of three parts: a uniform annulus, a uniform line, and a point, each of total probability 1/3. This distribution (which has topological dimension 2) can be fitted as closely as desired by a two-dimensional data manifold.

The proof of Theorem 2 provides two immediate corollaries. First, for a fixed entropy manifold $\mathbb{D}$ of dimension $n$, in the limit of high gain ($a \rightarrow \infty$) and large numbers of middle hidden layer output quantization levels ($N \rightarrow \infty$), the middle hidden layer unit outputs of members of a sequence of replicator networks that are approaching optimal performance as coders for $\mathbb{D}$ (that is, a sequence for which $\varepsilon \rightarrow 0$) become entropy coordinates for input vectors. Second, a sequence of replicator networks that use the ramp activation function in their $n$ middle hidden layer units and are approaching minimum mean squared reconstruction error for a fixed entropy manifold will also have middle hidden layer unit outputs that become entropy coordinates for $\mathbb{D}$. The second corollary suggests that real-world problems might be solved with the use of the ramp activation function with back-propagation (which might greatly improve the rate of learning), rather than the stairstep activation function of Theorem 2 (which makes learning virtually impossible because of the need to back-propagate errors through a functional form having a derivative close to 0 everywhere). Thus, it might be possible to use the training of replicator networks with ramp activations on raw data as a technique for constructing natural coordinates for practical data sources (see Fig. 5).

In Theorem 2 it is presumed that the correct dimension $n$ of a given data manifold is known. Unfortunately, in most situations $n$ is unknown, but it can be estimated by considering the growth of affine subspace dimensionality of $k$ nearest neighbors around each of a generous selection of points in a large data vector sample. Typically, this function of $k$ grows at one slope, and then a decrease in slope occurs, which persists as $k$ is increased further. The $k$ value at this knee in the curve can be taken as an estimate of the local topological dimension (*15*) at that point; $n$ can then be taken to be the largest of these local dimensions (Fig. 4). Getting $n$ correct is probably the only way of steering between the Scylla of space-filling coordinates ($n$ estimate too low) and the Charybdis of highly elongated natural coordinate grid volumes ($n$ estimate too high). Although the present results have been presented in the context of data compression, replicator networks—and the approximate natural coordinate system representations they produce—will probably be at least as useful for data analysis and exploitation.

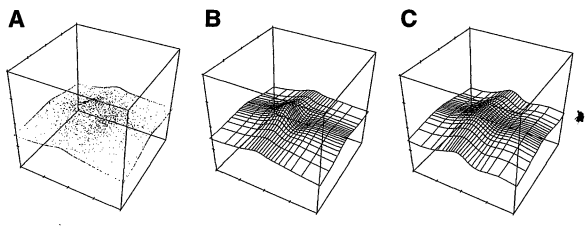


**Fig. 5.** (**A**) A set of 3000 points (chosen at random with respect to the data manifold's probability density $\sigma$) belonging to a two-dimensional data manifold $\mathbb{D}$ in $\Re^3$. (**B**) A natural coordinate system for the data manifold in (A). (**C**) An approximate natural coordinate system developed by training a replicator network ($m = 3$, $n = 2$, and $L_2 = L_4 = 12$) on randomly chosen data vectors from $\mathbb{D}$, then mapping each intersection point of (B) through the network and connecting the outputs with straight lines.

## REFERENCES AND NOTES


1. T. Kohonen *et al.*, *Neuroscience* **2**, 1065 (1976); P. Lehtiö and T. Kohonen, *Med. Biol. (Helsinki)* **52**, 110 (1978); T. Kohonen, E. Oja, P. Lehtiö, in *Parallel Models of Associative Memory*, G. E. Hinton and J. A. Anderson, Eds. (Erlbaum, Hillsdale, NJ, 1981), pp. 105–144; T. Kohonen, *Self-Organizing Maps* (Springer-Verlag, Berlin, 1995).
2. D. H. Ackley, G. E. Hinton, T. J. Sejnowski, *Cognit. Sci.* **9**, 147 (1985).
3. G. W. Cottrell, P. Munro, D. Zipser, in *Proc. Ninth Annual Conference, Cognitive Science Society,* Seattle, WA (Erlbaum, Hillsdale, NJ, 1987), pp. 461–473. Some commentaries, embellishments, and applications of Cottrell *et al.*: P. Baldi and K. Hornik, *Neural Networks* **2**, 53 (1989); R. Hecht-Nielsen, *Neurocomputing* (Addison-Wesley, Reading, MA, 1991), pp. 325–332; M. A. Kramer, *Am. Inst. Chem. Eng. J.* **37**, 233 (1991); D. DeMers and G. Cottrell, in *Advances in Neural Information Processing 5*, S. J. Hanson, J. D. Cowan, C. L. Giles, Eds. (Kaufmann, San Mateo, CA, 1993), pp. 580–587; K. Doya and A. I. Selverston, *Neural Comput.* **6**, 696 (1994). Some related studies of different kinds of replicator neural networks: S. Amari, *IEEE Trans. Comput.* **16**, 1197 (1972); S. Grossberg, *Studies of Mind and Brain* (Reidel, Boston, 1982); E. Oja, *Neural Networks* **5**, 927 (1992); R. Linsker, *Neural Comput.* **4**, 691 (1992); S. Becker and G. E. Hinton, *Nature* **355**, 161 (1992); M. Kawato, H. Hayakama, T. Inui, *Network: Comput. Neural Syst.* **4**, 415 (1993); W. Klimesch, *The Structure of Long-Term Memory* (Erlbaum, Hillsdale, NJ, 1994); G. E. Hinton, P. Dayan, B. J. Frey, R. M. Neal, *Science* **268**, 1158 (1995).
4. D. E. Rumelhart, G. E. Hinton, R. J. Williams, in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. (MIT Press, Cambridge, MA, 1986), vol. I, chap. 8.
5. J. L. Doob, *Measure Theory* (Springer-Verlag, New York, 1994); K. Ito, *Introduction to Probability Theory* (Cambridge Univ. Press, Cambridge, 1984); R. J. Zimmer, *Essential Results of Functional Analysis* (Univ. of Chicago Press, Chicago, 1990).
6. R. Hecht-Nielsen, *Neurocomputing* (Addison-Wesley, Reading, MA, 1991), pp. 124–138.
7. A. Gersho, *IEEE Trans. Inf. Theory* **25**, 373 (1979); R. M. Gray, *IEEE Acoust. Speech Signal Process. Mag.* **1**, 4 (April 1984); R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication* (Addison-Wesley, Reading, MA, 1977); P. Zador, *IEEE Trans. Inf. Theory* **28**, 139 (1982); W. R. Bennet, *Bell Syst. Tech. J.* **27**, 446 (1948).
8. J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups* (Springer-Verlag, New York, ed. 2, 1993); T. R. Fischer, *IEEE Trans. Inf. Theory* **35**, 137 (1989).
9. D. G. Jeong and J. D. Gibson, *IEEE Trans. Inf. Theory* **39**, 786 (1993); P. F. Swaszek, *ibid.* **37**, 1355 (1991).
10. K. Hornik, M. Stinchcombe, H. White, *Neural Networks* **2**, 359 (1989); S. E. Fahlman and C. Lebiere, in *Advances in Neural Information Processing 2*, D. S. Touretzsky, Ed. (Kaufmann, San Mateo, CA, 1990), pp. 524–532; S. Amari, *Neural Networks* **4**, 443 (1991); L. B. Almeida and F. M. Silva, in *Artificial Neural Networks 2*, I. Aleksander and J. Taylor, Eds. (Elsevier, Amsterdam, 1992), pp. 149–156; S. Amari and N. Murata, *Neural Comput.* **5**, 140 (1993); A. R.

Barron, *IEEE Trans. Inf. Theory* **39**, 930 (1993); K. Hornik, M. Stinchcombe, H. White, P. Auer, *Neural Comput.* **6**, 1262 (1994); A. R. Barron, *Mach. Learn.* **14**, 155 (1994).
11. V. Guillemin and A. Pollack, *Differential Topology* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
12. M. Rabbani, Ed., *Selected Papers on Image Coding and Compression* (SPIE Press, Bellingham, WA, 1992).
13. W. K. Bühler, *Gauss: A Biographical Study* (Springer-Verlag, New York, 1981); G. Strang, *Linear Algebra and Its Applications* (Academic Press, New York, ed. 2, 1980).
14. T. Hastie and W. Stuetzle, *J. Am. Stat. Assoc.* **84**, 502 (1989).
15. J. R. Munkres, *Topology* (Prentice-Hall, Englewood Cliffs, NJ, 1975); P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces* (Cambridge Univ. Press, Cambridge, 1995).
16. I thank the Ballistic Missile Defense Organization, the Army Research Office, the Advanced Research Projects Agency, the Office of Naval Research, the Naval Surface Warfare Center, and the Air Force Wright and Rome Laboratories for their support of portions of this work. The comments and suggestions of F. Crick on an early draft of this paper are appreciated, as are those of D. Modha on later drafts. Thanks also to S. Wang, who carried out computer experiments and produced Figs. 3 and 5, and to the referees for several helpful suggestions.

24 January 1995; accepted 2 August 1995

# Circadian Oscillations of Cytosolic and Chloroplastic Free Calcium in Plants

Carl Hirschie Johnson,* Marc R. Knight, Takao Kondo, Patrick Masson, John Sedbrook, Ann Haley, Anthony Trewavas

**Tobacco and *Arabidopsis* plants, expressing a transgene for the calcium-sensitive luminescent protein apoaequorin, revealed circadian oscillations in free cytosolic calcium that can be phase-shifted by light-dark signals. When apoaequorin was targeted to the chloroplast, circadian chloroplast calcium rhythms were likewise observed after transfer of the seedlings to constant darkness. Circadian oscillations in free calcium concentrations can be expected to control many calcium-dependent enzymes and processes accounting for circadian outputs. Regulation of calcium flux is therefore fundamental to the organization of circadian systems.**

The circadian clock controls critical daily behavioral and reproductive activities in plants and animals. This timing mechanism is an oscillator that orchestrates a variety of circadian outputs in processes as diverse as ion fluxes, behavior, and gene expression (*1*). Recognition of the extraordinary diversity of calcium-regulated processes (*2*) has led to speculations that intracellular free calcium ($[Ca^{2+}]_i$) acts as a component of the central circadian clockworks (*1*); that it entrains circadian clocks in plants, molluscs, and mammals (*3*); and that it regulates circadian outputs, including the expression of specific genes (*1*, *4*, *5*). Technical limitations of procedures (*6*) for in vivo $[Ca^{2+}]_i$ measurement have hindered direct tests of these speculations. We have developed a method to monitor $[Ca^{2+}]_i$ in plants genetically transformed with the complementary DNA (cDNA) for apoaequorin (*7*). When reconstituted to aequorin, this calcium-sensitive luminescent protein enables continuous noninvasive reporting of $[Ca^{2+}]_i$ for many weeks in tissues, single cells, or whole seedlings (*7*, *8*).

C. H. Johnson, Department of Biology, Vanderbilt University, Nashville, TN 37235, USA.
M. R. Knight, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.
T. Kondo, Department of Biology, Faculty of Science, Nagoya University, Nagoya 464-01, Japan.
P. Masson and J. Sedbrook, Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA.
A. Haley and A. Trewavas, Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK.
*To whom correspondence should be addressed.

For measurements of the cytosolic free calcium concentration ($[Ca^{2+}]_c$), tobacco (*Nicotiana plumbaginifolia*) was genetically transformed with the cDNA for apoaequorin downstream of the cauliflower mosaic (CaMV) 35S promoter [strain MAQ 2.4, see (*7*)]. After reconstitution of aequorin with the luminophore coelenterazine (*9*), the luminescence of these seedlings (which reports $[Ca^{2+}]_c$) was monitored after transfer to continuous light (LL), and a circadian pattern was observed (Fig. 1). The luminescence peak occurred shortly after dawn and recurred at circadian intervals. This luminescence rhythm could be phase-shifted by an extension of the last dark interval (compare Fig. 1, A and B), which demonstrates that this rhythm is genuinely circadian. The luminescence rhythm could also be entrained by light-dark (LD) cycles; the two rhythms depicted in Fig. 1D come from separate groups of seedlings that were entrained to reversed-phase LD cycles and then were measured simultaneously.

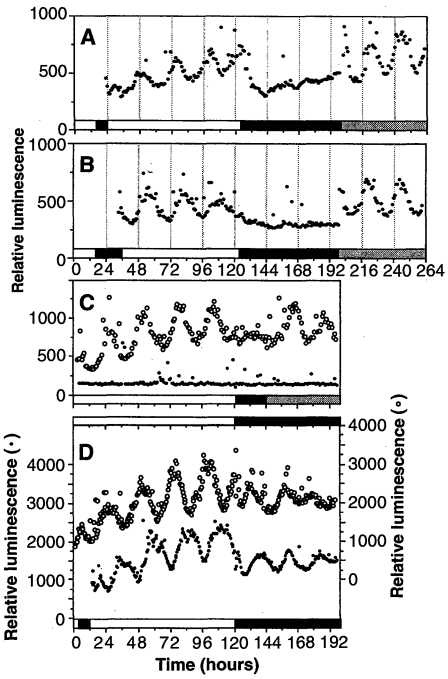


**Fig. 1.** Luminescence of transgenic tobacco seedlings containing cytoplasmic aequorin (MAQ 2.4) under various illumination regimes (*9*). White bars along the abscissas indicate white light background (22 $\mu E\ m^{-2}\ s^{-1}$), black bars indicate darkness, and gray bars indicate red light (18 $\mu E\ m^{-2}\ s^{-1}$). (**A**) Seedlings (two per vial) grown on a light-dark cycle of 16 hours light, 8 hours dark (LD 16:8); the last light cycle is shown. (**B**) Same as in (A), except that the final dark period was extended by 12 hours, resulting in a delay of the subsequent rhythm. (**C**) Open circles, transgenic MAQ 2.4 seedlings; closed circles, nontransgenic wild-type tobacco seedlings. Both groups were treated with coelenterazine (two seedlings per vial, grown on LD 16:8). (**D**) Separate cultures of MAQ 2.4 grown on reversed LD 16:8 cycles (10 seedlings per vial). The last light cycle for each is shown. The upper bar is for the open-circle trace, and the lower bar is for the closed-circle trace. Seedlings were initially in LL, then transferred to DD at hour 120.

Luminescence was not detected in either nontransgenic wild-type seedlings incubated with coelenterazine (Fig. 1C) or transgenic seedlings that were not exposed to coelenterazine (*10*). The rhythms of strain MAQ 2.4 were measured in both white and red LL (Fig. 1, A to C). In constant darkness (DD), the rhythm usually damped rapidly (Fig. 1, A and B), but in a few experiments, an oscillation was observed in DD, albeit at a reduced amplitude (Fig. 1D). The greater amplitude of the luminescence rhythm in LL may reflect the common observation that circadian expression in plants is more robust in LL (*11*), or that the circadian pacemaker modulates the sensitivity of its photoreceptive pathway (*11*), resulting in rhythmic phytochrome-mediated calcium flux (*12*).

Aequorin luminescence is specific for $Ca^{2+}$ and is sensitive to free calcium in the